Review

Snake venomomics - from low-resolution toxin-pattern recognition
to toxin-resolved venom proteomes with absolute
quantification

Juan J. Calvete

Evolutionary and Translational Venomomics Laboratory, CSIC, Jaime Roig 11, 46010
Valencia, Spain

Email: jcalvete@ibv.csic.es

Phone: +34-96-339-1778

Fax: +34-96-369-0800Abstract

Introduction: Venoms are integrated phenotypes used by a wide range of organisms for predatory and defensive purposes. The study of venoms is of great interest in diverse fields, such as evolutionary ecology and biotechnology. Omics technologies have contributed to understanding the evolutionary mechanisms that molded snake venoms to their present-day structural and functional variability landscape.

Areas covered: This review article reflects on two recent implementations in venomomics: absolute quantification of intact proteins by elemental mass spectrometry, and top-down molecular mass spectrometry.

Expert commentary: Leveraging on a new way of polyatomic interference removal, a triple quadrupole ICP-MS configuration has proven feasible for the absolute quantification of venom toxins via sulfur detection. A major advantage of this approach over quantitative molecular MS techniques is that only a generic S-standard is required to quantify all the chromatographically separated sulfur-containing fractions. Top-down venomomics is in its infancy but, due to recent hardware and software developments, is gaining momentum. Proteoform-resolved venom proteomes are needed to understand the spatio-temporal variability landscape underlying the adaptations that drive intraspecific venom evolution. Integrating top-down venomomics and absolute proteoform quantification into a novel elemental and molecular mass spectrometry configuration will represent a quantitative leap in the study of individual venoms.

Keywords: Snake venom, venomomics, top-down mass spectrometry, elemental mass spectrometry, proteome-resolved venom proteome, ICP-MS, absolute quantification

1. Introduction. The big bang of snake venomomics

The term “venomics” was coined in 2004 to describe an analytical strategy to unveil the global proteome composition of snake venoms [1]. As proof-of-concept, a venomics platform, consisting of the combination of cysteine mapping, N-terminal Edman degradation, and de novo sequencing by tandem mass spectrometry of internal tryptic peptides gathered by in-gel digestion of SDS-PAGE bands of reverse-phase HPLC separated venom proteins, was applied to characterize the venom of the dusky pigmy rattlesnake, *Sistrurus miliarius barbouri*, a pit viper subspecies endemic to the southeastern United States [1]. That was the first time that my laboratory was faced with the challenge of portraying the compositional complexity of a snake venom. Until then, our research had focused on the study of structure-function correlations of disintegrins, low molecular mass, cysteine-rich peptides isolated from the venom of Viperidae and Crotalidae snake species, and which antagonize the binding of ligands to integrin receptors [2-4], thereby impairing the biochemical signals that these adhesive receptors bi-directionally transduce through the cell plasma membrane [5]. In the course of the development of chromatographic methods for the isolation and quantification of disintegrins from different venoms we set out to analyze the "other components of the chromatogram", in order to get a detailed view of the molecular environment of disintegrins in the crude venom. The interest this work aroused among the toxinology community encouraged us to adopt venomics as a main research avenue in our laboratory. Conceptual and technological alliances, eg. standalone de novo bottom-up peptide-centered approaches or, more and more frequently the integration of homologous venom gland transcriptomics, bioinformatic analysis and LCMS/MS workflows [6, 7], that contributed to the transition from the description of the inventory of toxins present in a venom to a full biological discipline, have been reviewed in previous papers [8-12]. Starting from this base, this review focuses on recent (2015-to date) advances towards the development of integrative venomics and discusses foreseeable near future developments in the field.

An important challenge in the characterization of venoms (and other complex biological samples) is the inability of any single method to provide a complete view of its proteome. Proteome decomplexation and toxin identification and quantification

represent major pillars of current state-of-the-art snake venomomics protocols. As any novel field of research, venomomics is bound to evolve. This circumstance provides the golden opportunity to standardize next-generation experimental protocols to render

Accepted Manuscript

research in this field comparable among the growing number of laboratories joining the field of molecular toxinology.

2. The evils of quantification in a venomomics workflow

The development of quantitative proteomics techniques represents a major advance in the field over the first decade of the XXI century [13]. Among the multiple hyphenated approaches proposed by different research groups for integrating venom proteome decomplexation and the identification and quantification of the individual toxins into a single workflow [14], the most successful combination has turned out to be the one that takes advantage of venom fractionation to simultaneously quantitate the relative abundances of its different venom components [10, 15] (Fig.1). This venomomics strategy, which is usually referred to as "snake venomomics", relies on monitoring the reverse-phase column eluate at the absorbance wavelength of the peptide bond to estimate the relative abundances of the chromatographically separated fractions by peak area integration [10, 12]. The relative abundances of the different venom proteins can then be calculated through a three-levels quantification protocol. The first level of quantification provides an initial estimate of the representation of the venom's toxin families from the ratio of the sum of the areas of the reverse-phase chromatographic peaks containing proteins from the same family to the total area of venom protein peaks in the reverse-phase chromatogram. The relative abundances of proteins coeluting in the same reverse-phase fraction are estimated, in a second level of quantification, by densitometry of Coomassie-stained SDS-polyacrylamide gels. In the third level of quantification, the relative abundances of different proteins co-migrating in the same SDS-PAGE band are estimated based on the relative ion intensities of the three more abundant peptide ions assigned by MS/MS analysis to each the comigrating proteins. In contrast to most protein multidimensional separations approaches, an added value of this three-step quantification workflow is that the hierarchical relationship between the quantification levels guarantees that the

sequential venom subfractionation steps do not alter the relative proportions of the components in the original venom.

Each level of the three-levels quantification workflow provides a different degree of quantitative information, whose importance is hierarchical, in the sense that a lower level refines the value obtained in a higher level. This means that the densitometric analysis of an SDS-PAGE separation of the components coeluting in a chromatographic peak, which represented 20% of the total area of the reverse-phase chromatogram of a venom proteome, serves to refine this figure without altering it. Let's assume that this chromatographic fraction contained 4 electrophoretic bands (b1-b4) that corresponded, respectively, to 63%, 22%, 13% and 2% of the total area of the densitogram of the SDS-PAGE lane. This allows us to calculate that b1-b4 comprised, respectively, relative abundances of 12.6%, 4.4%, 2.6%, and 0.4% of the original venom proteome. Let's now assume that MS/MS analysis of b2 revealed 2 components, B2a and b2b, whose ratio of the average intensity of an equal number of ions assigned to each component was 3:1, suggesting that the relative abundances of b2a and b2b in the original venom proteome were 3.3% and 1.1%, respectively. This distribution of relative abundances did not affect the values calculated for the other components of the original proteome. In addition, in most cases, snake venom components coeluting in the same electrophoretic band usually correspond to structural homologs, and thus the third level of quantification does not influence the calculation of the relative abundances of the venom's toxin families derived through the combined RP-HPLC and SDS-PAGE quantification steps.

This property is also shared by 2DE-based and, to a certain extent, also by shotgun approaches. 2DE reflects the venom protein complexity in a single image better than any other single approach. However, and notwithstanding the limitations of mass and pI ranges in its capacity to resolve a complex proteome, the quantification of 2DE protein maps can be complex because of the challenge of assigning spot boundaries due to variable background, spots overlapping and spot saturation [16]. Shotgun proteomic workflows rely on using peptide surrogates to infer protein amounts, assuming that peptide quantity represents a reliable read-out for the parent protein amount. However, the relationship of the identified peptides to their intact

parent protein is (except for proteotypic peptides) lost, or very difficult to reconstruct. On the other hand, in shotgun proteomics, protein quantification is based on the relative normalized number of peptide spectra that are identified for a particular protein (spectral counting), or on a protein abundance index calculated as the median intensities of extracted ion chromatographic (XIC) peaks matched to the three more intense reference peptides per protein [17]. In short, label-free quantification is based on the assumption that the likelihood of data-dependent precursor ion selection and identification is higher for abundant precursor ions. Together, these premises suggest that the number of successful peptide identifications will be correlated with the abundance of the parent protein. However, label-free approximations present two inherently important limitations. On one side, peptide properties, such as mass, amino acid length and composition, solubility, net charge, etc., can impact ionization, transmission, and peptide detection efficiency [18], leading to errors in assessing the abundance of the parent protein producing the tryptic peptides. On the other hand, inference of protein abundance based on the frequency of identification of surrogate peptides is only applicable in those cases in which MS/MS spectral matching to a reference database does not represent a limiting factor of the experiment [19, 20].

When a comprehensible homolog reference database (transcriptome or genome) is not available, deriving reliable quantitative information from peptide-centric MS/MS data is challenging because the unpredictable extent of missing data. Hence, relative quantification of the set of proteins of a given proteome by label-free methods relies in converting spectral counts or peptide intensity information into protein abundance ratios. To this end, the sum of the total number or intensity of the peptide spectra matched to a given protein (P_i) are normalised by a) the total number of peptides theoretically expected from its amino acid sequence (spectral count method) or b) the total intensity of all the peptide ions assigned to P_i (spectral ion intensity method). Normalised figures are then transform into percentages of the total proteome's ion spectral counts or peptide ion intensities. In the absence of a comprehensive specific database, the calculated normalised spectral count or ion intensity ratios for proteins present (or represented by a highly homologous protein) in the used search database are overestimated with respect to those that being in a similar concentration

produce a smaller subset of identified peptide ions.

In peptide centric approaches, quantification should be based on an unbiased procedure, e.g. one that does not depend on the sequence coverage of the database used for identifying the individual components of a sample's proteome. Currently, in the case of reference database orphaned venoms, surpassing the incompleteness of database information requires a good deal of de novo tryptic peptide ion sequencing combined to derived amino acid sequence similarity searches to either identify a close homolog protein or at least to place the query de novo deduced sequence as a member in a particular venom protein family.

3. Absolute quantification of intact proteins through a hybrid molecular and elemental mass spectrometry configuration

The variable ionization efficacy of peptides [18] and the "missing values problem" [21, 22], which are both inherent to peptide-centric proteomics analyses, particularly of nonmodel organisms, can be circumvented through the determination of absolute intact protein amounts based on database-independent mass spectral data. Absolute protein quantification can be achieved using isobaric reagents (such as Isobaric Tag for Relative and Absolute Quantification, iTRAQ, and Tandem Mass spectrometry Tags, TMT) [23, 24]. However, these workflows are limited by the need of proteotypic internal standards, reproducible and complete or known stoichiometry of protein labeling at each possible derivatization site, as well as efficient (ideally complete) protein digestion [25]. To date, only two venomomics studies have used iTRAQ-tagged venom proteins of neonate and adult Brazilian arrowhead viper, *Bothrops jararaca* [26], and the Chinese short-tailed pit viper [27], *Gloydus brevicaudus*, to quantitate MS2-based intrascan relative venom changes during neonate-to-adult transition.

Other quantitative techniques involve the use of isotope labeled proteins as internal standards in conjunction with isotope dilution strategies for absolute quantification by proteomic strategies [28, 29]. Stable isotope-labeled (SIL) proteins closely mimic the behavior of their natural counterparts and constitute the internal standards of choice for quantitative proteomics. However, such PSAQ (Protein Standard for Absolute Quantification) are costly and difficult to produce, whether through ex vivo (chemical) or in vivo (metabolic) labeling approaches. An alternative to

overcome the need of spiking a synthetic SIL standard for each target protein is a wellknown technique in the field of bioinorganic analysis, namely inductively coupled plasma mass spectrometry (ICP-MS) combined with stable-isotope dilution. ICP-MS is a type of elemental mass spectrometry introduced by Houk et al. in 1980 [30] and commercially introduced in 1983 [31]. ICP-MS has the capability of simultaneously screening and absolute quantifying multiple heteroatoms (any element in biological material other than C, H, N, O and F), including metals, semimetals, and biologically relevant non-metals naturally occurring in proteins (e.g., S, P, I, Se) [32-34]. Detection of the non-metallic elements typically present in proteins has been traditionally hampered to ICP-MS, and only very recent instrumental advances, mostly based on the new concept of using a triple quadrupole configuration (ICPQQQ) for efficient polyatomic interference removal [35], have enabled their detection with high selectivity and sensitivity (11 fmol). This approach was originally applied to the simultaneous absolute quantification of S- and P-peptides without the need for specific standards [35].

The potential of sulfur measurement for the quantitative analysis of proteins and peptides has been pinpointed [36]. Of particular relevance in venomomics, the sulfurcontaining aminoacid cysteine (and in most cases also methionine) is present in virtually all snake venom toxin classes, and is particularly abundant in small proteins whose global folds are primarily stabilized by disulfide bonds [37]. Taking advantage of the fact that the binomial (number of sulfhydryl groups and disulfide bonds per molecule; molecular mass) represents a useful molecular feature for the preliminary classification of toxins into protein families [38], Calderón-Celis and colleagues have reported a proof-of-concept showing the feasibility of applying RP- μ HPLC-ICP-QQQ and on-line ^{34}S isotope dilution analysis for the absolute quantitative analysis of the major toxins comprising elapid venom proteomes [39, 40]. A major advantage of this approach over other quantitative techniques requiring SIL peptides or proteins as internal standards is that only one certified single heteroatom-containing species, such as a generic Sstandard (e.g. BOC-methionine), is required to quantify all the chromatographically separated sulfur-containing venom fractions, in the femtomole range and with accuracy $\geq 99\%$, and excellent linear dynamic range of up to 10 orders of magnitude [34, 41].

Identification of the toxins eluting along the chromatographic (RP- μ HPLC) separation can be achieved via ESI-QToF mass profiling carried out in parallel to the ICP-QQQ MS measurements, by matching the isotope-averaged molecular masses to the calculated values for mature toxins recorded in previous venomomics analyses or computed from homologous venom gland transcriptomic databases (Fig.2). Intrinsic limitations of the combined hybrid molecular and elemental mass spectrometry configuration are the strict requirement for good (ideally baseline) chromatographic peak resolution and knowledge on the full-length amino acid sequence of the proteins present in the sample to derive, from the absolute sulfur amount quantified in each peak and the sulfur (cysteine + methionine residues)/protein stoichiometry, the absolute amounts (moles) of toxins per gram of venom. This way of expressing the data is biologically more meaningful than the usually reported "g toxin/g of total venom proteins" [10, 12], since the number of inoculated molecules, rather than their % by weight, is responsible for the pathological effects of the venom.

4. The challenge of proteoform-resolved quantitative venomomics

The trend towards hybrid configurations of mass analyzers has dominated recent advances in instrumentation for mass spectrometry [42, 43]. Current hybrid mass spectrometry systems use various designs of beam-type and in-time ion-trapping spectrometers to combine the different performance characteristics offered by the individual in-space and in-time analyzers into one instrument. The incorporation of ICPMS into current mass spectrometry configurations, would represent an instrumental complication that, in addition, affects the economic cost of the mass spectrometry system. However, this drawback would be more than compensated if the system would have the capability of determining in the same experiment both the identity and the absolute concentration of each and every one of the components of a complex biological system, such as a snake venom. Venoms are complex phenotypic traits composed of mixtures of proteins belonging to a few toxin families, each venom formulation exhibiting varying complexity (e.g. different expansion of individual protein families and combinations of families) among venomous snakes [9, 44-46]. A key element to achieve full qualitative and quantitative venom proteome coverage is the complete pre-MS separation of all the protein species (iso- and proteoforms [47]) that

constitute the venom's toxin arsenal. Although conceptually it might seem less complicated to separate a few hundred proteoforms than thousands of their proteolytic fragments, proteoform-resolved venomomics is more challenging than peptide-centric approaches, in part due to low resolving power of the current front-end fractionation technologies toward closely related sequence variants of members of the same toxin family [48]. A possible theoretical solution could be provided by the high resolving power ($>1,000,000$ resolution) afforded by Fourier transform (FT) Orbitrap and ion cyclotron resonance (ICR) mass analyzers [49, 50], which offers the unparalleled potential of confine, manipulate, and measure with high accuracy the exact mass-to-charge (m/z) ratios of hundreds of proteoform ions with molecular masses in the range of 5-35 kDa [51]. However, the coupling and efficient transmission of the output of an Fourier transform ion trap to an ionization system presents technological challenges that future advancements in ion trap technology may solve. In addition, the low ion storage capacity of FT ion traps (0.7–3.0 attomole to zeptomole for 8-29 kDa proteins) [52, 53] hampers the productive mass-selective transfer of the ion trap content (20,000–500,000 ions) to current ICP-MS systems, which operate in the femtomole limit-of-detection range, thus precluding the online absolute quantification of individual proteoforms isolated in the ion trap. Consequently, most of the few studies that, to date, have applied quantitative top-down MS to portray the inventory of snake venom proteomes at proteoform level [54-59] have used label-free approaches to quantify the relative abundances of the venom components from integrated peak areas or peak intensities of extracted ion chromatograms of full-scan spectra (MS1), with proteoform intensity based on the sum of all the relevant isotopic peak heights over the elution time of the proteoform [60].

5. Top-down venomomics is gaining momentum

Soon after John B. Fenn, inspired by Malcolm Dole's work on molecular beams of macroions published 20 years earlier [61], provided "ESI wings to molecular elephants" [62, 63], the potential of ESI-FT-ICR MS for the structural analysis of intact proteins was appreciated in studies showing the detection of multiple charge states and isotopic resolution on proteins [64-66]. The high resolving power of the same instrument was also utilized to determine the identity of collision-induced dissociation fragment ions of

ubiquitin [67]. These pioneering studies precipitated the “electrospray revolution” that represented the foundation that helped implement some 20 years later top-down MS, which now represents a rapidly growing proteomics discipline with its own idiosyncrasies [68-72].

Top-down mass spectrometry (MS) strategies allow in-depth characterization of proteins by sequential isolation and fragmentation of the entire ionized molecules inside a mass spectrometer without requiring prior proteolytic digestion [71, 72]. Topdown MS was first applied only a few years ago to unveil the molecular biodiversity of the venom of King Cobra (*Ophiophagus hannah*, Elapidae) from Indonesia [54] and Malaysia [55]. The choice of this snake species was motivated by the fact that, at that time, this species (although from Bali) was the only venomous snake whose genome had been sequenced [73]. This election was also fortunate because most elapid venoms are dominated mostly by α - and β -neurotoxins of the three-finger toxin (3FTx; 6-9 kDa), Kunitz-type inhibitor-like (KUN; 7 kDa), and phospholipase A2 (PLA2; 14 kDa) protein fold families [74], and these small protein classes can easily be detected with good sensitivity at standard pressure, whereas larger proteins (>30 kDa) are often reduced in apparent abundance. In addition, 3FTx [75], KUN [76], and PLA2 [77, 78] proteoforms are essentially devoid of post-translational modifications other than disulfide bonds. This structural feature means that they can be easily identified at MS1 level (mass profiling) by comparing the high-resolution experimentally determined isotope-averaged molecular masses and those calculated from an homologous database (Fig.1, workflow 1b-2a-4a). Arguably, if the chemical formula and covalent bonds of a protein are known, and its molecular mass and isotopic envelope are determined with sufficient resolution and accuracy, its identification by mass spectrometry is unambiguous. It is the same principle used in peptide mass fingerprinting (PMF), except that in the top-down experiment a single ion contains more information (100% of the ion's structure) than in the bottom-up approach, where a set of peptide masses, which together usually represent <100% of the parent structure, are required for the identification of a protein becomes statistically indisputable. Nevertheless, although I am not aware that the occurrence of isobaric proteins have been documented in any snake venom proteome, it is always advisable to validate,

whenever possible, MS1-based assignments by MS/MS analysis.

Fifteen 3FTx from the five major 3FTx functional classes, 3 natriuretic peptides, 2 KUN, 1 ohanin, 1 cystatin, 1 ILGF-1, 1 cysteine-rich secretory protein (CRISP) MS1-assigned unique proteins were verified by data-dependent high-resolution top-down mass spectrometry using a hybrid LTQ Orbitrap XL instrument equipped with an Agilent 1260 HPLC system for sample decomplexation [54]. Twenty to 22 venom proteins that could not be identified in the top-down approach (e.g. because they were not represented in the non-redundant GenBank/EMBL/DDBJ database), including 3-5 PLA2 and 2-3 snake venom metalloproteinase (PIII-SVMP) and 1 L-amino acid oxidase (LAO) proteoforms were characterized using bottom-up tryptic peptide-centric venomomics. However, due to the lack of intact mass information and unique locus-specific peptides, some identifications remained ambiguous. In any case, these results highlighted the complementarity between the top-down and bottom-up approaches, particularly when a comprehensive reference database for matching the top-down MS output is missing. Dealing with this limitation does not seem to be a trivial matter given the well-documented inter- and intra-populational intraspecific individual venom variability, underlying the accelerated evolution driven by positive selection operating on venom toxin multigene families [9, 44, 79]. And, as it is easy to understand, it is neither practical nor ethical to perform for each venomomic study a transcriptomic analysis of the venom glands of the same snake specimen.

In another proteoform-resolved interrogation of the venom proteome from two individuals of Malaysian King Cobra, Melani and co-workers [55] applied off-line sample fractionation through denaturing and native GELFrEE [80] and solution isoelectric focusing (sIEF) prior to LC- (top-down and bottom-up) MS/MS analyses on a hybrid quadrupole-Orbitrap (Q-Exactive HF) mass spectrometer. This high-throughput strategy increased proteoform identifications from 17 (direct analysis of whole venom by LC-MS/MS) to 64 (GELFrEE) to 131 proteins, with sIEF/LC-MS/MS providing 63 unique protein identifications not found using the other two separation strategies. Proteins identified by absolute mass searching were manually validated to precisely annotate the amino acid substitutions, alternative processing cleavage sites, and the presence of extra post-translational modifications (PTMs), resulting in the identification

of 53 additional proteoforms, mapped to genes coding for 3FTx, PLA2, KUN, and CRISP family members. Nearly 70% of the 53 distinct proteoforms arose from either amino acid changes compared to the canonical sequence present in the nonredundant database, and/or PTMs not related to protein proteolytic processing. In total, 184 unique proteoforms were identified [55]. In addition, off-line fractionation of whole venom by native GELFrEE followed by quadrupole-isolation and subsequent fragmentation by high-collision dissociation (HCD) was employed to access large proteins present in King Cobra venom, resulting in the identification of two peaks with average masses centered at $130,014.2 \pm 0.7$ Da and $126,271.0 \pm 2.3$ Da as glycoforms of LAO with p-scores of 3×10^{-53} and 2×10^{-51} , respectively, and a cobra venom factor (CVF) of $146,565 \pm 63$ Da [55]. This study highlighted the importance of prefractionation and high-resolution measurements for a thorough top-down MS characterization of a venom proteome. It also demonstrated that high molecular mass and post-translationally modified toxins are really accessible for characterization by top-down strategies.

Whether as a standalone workflow or in combination with bottom-up approaches, top-down venomomics is gaining momentum, particularly for the analysis of venoms from species for which homologous transcriptomic reference database is available. A recent transcriptomics-guided bottom-up and top-down comparative venomomics study of neonate and adult specimens of the arboreal rear-fanged Brown Treesnake, *Boiga irregularis*, from Guam [59] illustrates this assertion. The venom proteomes of neonate (N) and adult (A) *B. irregularis* were both overwhelmingly (78% and 84%, respectively) dominated by monomeric and dimeric 3FTxs, followed by moderately abundant (21% (N) and 13% (A)) CRISPs, low abundance (1% (N), 3% (A)) PIII-SVMPs, and very low abundance ($< 0.01\%$) PLA2 and SVMP inhibitors. Topdown experiments were performed using a ultra-high performance liquid chromatography (uHPLC) system coupled to a Q-Exactive hybrid quadrupole orbital ion trap. Neonate and adult venoms shared 30 3FTx proteoforms and, in addition, the neonate venom expressed 7 unique 3FTxs. In both age classes, these proteoforms belonged to the same 10 3FTxs transcripts. Noteworthy, despite the overlapping sets of venom components expressed in both neonate and adult snakes, the relative

abundances of the toxins differed between age classes. The most abundant components in the neonate venom proteome were the proteoforms of 3FTx-8b/d (16.1% of the total venom proteome), Irditoxin subunits A (6.5%) and B (7.6%), 3FTx-3 (4.7%), and 3FTx-4 (3%), whereas in the adult venom proteome the most abundant toxins were the proteoforms of Irditoxin chains A (17.1%) and B (13.2%), 3FTx-4 (9.4%), 3FTx-3 (7.4%), 3FTx-8b/d (6.6%), 3FTx-5 (3.9%), and 3FTx-7b (3.0%). This detailed study would have been difficult to achieve through a purely bottom-up strategy.

Venom is an adaptive trophic trait, and detailed ontogenetic venom shifts may help establish correlates with differences in prey consumption. Brown Treesnakes undergo a pronounced ontogenetic shift in prey preference, from a diet consisting almost exclusively of ectothermic prey (small lizards, particularly *Carlia* skinks and *Hemidactylus* geckos) in neonates, to both ectothermic and endothermic vertebrates (birds and mammals) being consumed by adults [81]. In line with this observation, neonates produce more toxic venom toward house geckos (*Hemidactylus frenatus*; LD50: 1.1 g/g) than adults (LD50: 2.5 g/g) [82], whereas the most abundant toxins of adult *B. irregularis* venom, the dimeric 3FTx iriditoxin proteoforms, exhibited taxonspecific lethality toward birds (LD50 0.55 g/g chick) and lizards, and was nontoxic to mice (up to 25 g/g (i.p.)) [83]. The lack of toxic venom components targeting mammalian prey is striking. However, this apparent paradox can be reconciled with the sense of the ontogenetic venom change, taking into account the multiple predatory modalities employed by adult *B. irregularis* to subdue prey; envenomation to subdue bird and lizard prey, and constriction to subdue mammalian prey [84].

For the past few years, top-down venomomics has benefited from developments of computational tools for aiding in the successful identification of a proteins/proteoforms from complex production spectra of large multi-charged isotopomer parent ions. Such software is particularly useful when only a fragmentary species-specific database is available. The Consortium for Top-Down Proteomics website [201] hosts links to the main programs available for visualization and analysis of tandem top-down data.

TopPIC (Top-Down mass spectrometry-based Proteoform Identification and Characterization) [85] and MASH [86] are software tools that provide statistical

estimates for the confident inference of unknown primary structure alterations, such as amino acid mutations and post-translational modifications, in known protein entries by searching top-down tandem mass spectra against a protein database. Petras and co workers [57] used TopPIC version 0.9.1. for matching LC-MS/MS spectra of green and black mamba (genus *Dendroaspis*) venom toxins, obtained in data-dependent acquisition (DDA) mode in an LTQ Orbitrap XL mass spectrometer against a customized dataset comprising 165 non-redundant protein entries compiled from the NCBI and private *Dendroaspis* spp. databases. This dataset included, respectively, 20 and 27 venom toxin sequences from, respectively, the cognate eastern green mamba (*D. angusticeps*) and black mamba (*D. polylepis*). Mass profiling showed ≥ 200 proteoforms in green and black mamba venoms. Mass shifts between native and reduced components enabled the calculation of the number of disulfide linkages present in the corresponding RP-HPLC separated toxin molecule. The binomial mass and number of disulfide bonds (SS) is a distinctive feature of many toxin folds, and thus represents a useful proxy for the preliminary classification of toxins into putative protein families [1, 38]. Both *Dendroaspis* venom arsenals were dominated by 3FTxs (Da, 113; Dp, 125 proteoforms) and KUN (Da, 74; Dp, 66 protein proteoforms), which corresponded roughly 62-70% of the respective total ion current (TIC). TopPIC-assisted interpretation of good quality fragmentation spectra (Fig.3) enabled the identification of 71 (*D. angusticeps*) and 70 (*D. polylepis*) toxin sequences of which 38 (*D. angusticeps*) and 61 (*D. polylepis*) harboured unexplained mass discrepancies. Thirty-six (*D. angusticeps*) and 3 (*D. polylepis*) minor ($\leq 0.85\%$ TIC) proteoforms had a mass shift of +42 Da, suggesting the presence of monoacetylated lysine residues (Fig.4A). These assignments were fully validated (albeit with higher sequence coverage and lower p-values) and extended (including increased identifications and assignments of unexplained mass discrepancies) (Fig.4B) in a follow-up genus-wide top-down MS analysis across *Dendroaspis* venoms, using TopPIC for matching the MS/MS data to species-specific transcriptomics datasets [58].

6. Venom patterns in evolutionary and translational venomomics

Research on venoms has been continuously enhanced by advances in analytical technologies. And conversely, the growing interest in animal venoms as both model

systems for studying the molecular bases of adaptations [9, 44, 45], but also as a rich source for lead compounds that can drive forward the development of biotechnological tools and novel biotherapeutics [87, 88], has catalyzed the application of omics methodologies to unravel the idiosyncrasies of venoms. An increasing trend in venom analysis is the identification of evolutionary trends within and among whole genera, as unveiling the spatial and temporal distribution of venom variation within and between species and genera is key for understanding the evolutionary processes and the ecological constraints that molded snake venoms to their present-day variability [58, 77, 89-93]. Mapping the predominant toxin families across Elapidae venoms revealed a mosaic pattern of 3FTx- PLA2-predominant, and to a minor extent KUN-rich, venoms (Fig. 5A). This jumbled compositional distribution points to divergent lineage-specific toxin family expansions during the cladogenesis of Elapidae, and the fact that divergent venom phenotypes cluster in monophyletic clades suggested that elusive genomic and/or post-genomic regulatory mechanisms may have shaped the composition of extant elapid venoms.

Mamba venoms, dominated by dendrotoxins (Kunitz-type inhibitor fold) and 3FTxs, represent a striking deviation from the 3FTx/PLA2 dichotomic general trend across elapid venoms (Fig.5A). Genus-wide top-down venomomics analysis of the five species and subspecies within genus *Dendroaspis* revealed their proteoform-resolved venom proteomes (Fig. 5B). This detailed information laid the foundations for rationalising the notably different potency and pharmacological profiles of *Dendroaspis* venoms at locus resolution, and suggested a plausible causal dietary correlation between the divergent KUN-predominant venom of terrestrial mammals specialist *D. polylepis* and the 3FTx-rich venoms shared by the other four, mainly arboreal, mamba species, whose diet is dominated by arboreal taxa, mainly birds. Deploying highly toxic venom in conjunction with a strike-release predatory strategy, thereby reducing potential risk of injury posed by mammalian prey, has also evolved in other mammalspecialist snakes [94].

Identifying venom patterns with more and more resolution at the molecular level may also shed light on the mutually enlightening relationship between evolutionary and clinical toxinology. Arguably, toxins bearing the highest mammalian

prey incapacitation activity are often also the most medically important molecules in the context of a human envenoming. Hence, identifying the molecular basis of venomous snake adaptation to their natural ecosystems may assist in the identification of those toxins that need to be neutralized to reverse the deleterious effects of venom, thereby guiding in the rational development of next generation snakebite therapeutic antivenoms [95, 96].

7. Expert comment

Ultimately, locus-resolved venom proteomes are critically needed to understand the spatial and temporal landscape of venom variability underlying the local adaptations that drive intraspecific venom evolution. However, unveiling genotype-phenotype correlations for complex polygenic traits, such as venom systems, is most times not feasible using peptide-centric bottom-up approaches. Top-down mass spectrometry has the potential to eliminate the shortcomings of the "protein inference problem" intrinsic of bottom-up MS workflows [97]. Although there is the common belief that topdown MS approaches are several years behind bottom-up workflows in terms of its implementation throughput, and versatility in proteomics in general and venom research in particular, the incontestable advantages offered by the possibility of qualitatively and quantitatively characterizing the proteome of individual venom samples, automatically and in a single experiment, represents an incentive for the development of relevant technology for whole protein MS. Improvements needed for making top-down mass spectrometric analysis a routine protocol for confident proteoform identification and quantification, relate to almost every stage of the analytical process, including front-end sample decomplexation, mass spectrometer's source and duty cycle conditions, and the physical environment inside the instrument (eg, ion accumulation, acquisition time, signal-to-noise ratio and mass accuracy of large protein ions), and computational tools for interpreting of the complex production spectra resulting from the fragmentation of multicharged ions. Nevertheless, recent coinciding advances in all these areas are routinely allowing the complete characterization of high molecular mass proteins up to a level of detail that was not possible just a few years ago [98-107].

Snake venoms, particularly those from Elapidae taxa, are moderately complex

proteomes dominated by proteoforms from a few toxin fold families, chiefly 3FTx, PLA2, and KUN. Members within these toxin classes are essentially devoid of PTMs other than disulfide linkages, and therefore the measurement with sufficient precision of their masses is, in most cases, sufficiently informative to reveal their identities through mass profiling against an homologous venom gland transcriptomic dataset. This particularity means that, in addition to its contributions to aspects of evolutionary and translational biology, venomomics also represents an excellent testing ground to check the applicability of non-conventional technologies in the field of proteomics. In this regard, Calderón et al. [39, 40] have recently demonstrated a proof-of-principle use of a novel hybrid molecular and elemental mass spectrometry configuration for the MS1-based identification of snake venom toxins and their absolute quantification through parallel species-unspecific data-independent acquisition (DIA)-based RP- μ HPLC-ICP-QQQ MS and on-line ^{34}S isotope dilution analysis. Top-down venomomics is in its infancy, and the challenge remains to incorporate inductively coupled plasma mass spectrometry (ICP MS) as a detector into a compact and automated top-down MS platform (Fig.6) for the routine absolute quantitative analysis of all the toxin proteoforms present in a venom, and in a single run.

Venom biology is a multidimensional field, spanning the molecules and genes of the phenotype to the ecological consequences of its use [108]. Research of venoms from other sources, particularly those that are rich in small toxins of molecular masses less than 30 kDa and characterized for having heavily disulfide crosslinked structures, such as those of scorpions, sea anemones, spiders, or cone snails, will also greatly benefit from advances in top-down venomomics. On the other hand, hybrid elemental and molecular mass spectrometry systems will have a strong impact in heteroatom-targeted proteomics, most notably phosphoproteomics, metallomics, and metabolomics [109, 110]. Taking advantage of the multi-elemental (-isotopic) character of ICP-MS, heteroatom-targeted proteomics, in multiplexed formats, can bring about new opportunities for creative biological work and interdisciplinary developments.

8. Five-year view

The application of omic technologies, and particularly mass spectrometry-based methodologies, to study animal venoms has grown steadily in recent years.

Quantitative locus-resolved venom proteomes will increasingly be a goal in venom research during the next five years. In particular, the integration of top-down venomomics and absolute proteoform quantification via elemental mass spectrometry into a novel mass spectrometry configuration will represent a quantitative leap in the study of individual venoms, which will reveal the spatial and temporal patterns of intra- and interspecific molecular variability. Top-down snake venomomics is starting to take off, and only a handful of papers have been reported since the first publication in 2015 [54]. However, recent hardware and software developments allow us to predict its growing importance in a near future [111, 112]. Additionally to providing unparalleled molecular resolution, a practical advantage of absolute quantitative top-down venomomics relies in the possibility of automating the whole process (Fig.6), thereby reducing the analysis time from weeks (bottom-up venomomics) to hours. Reducing the analytical time, while increasing the depth of the analysis, will result in a paradigm shift, from studying pooled venoms as a proxy of the average venom phenotype of a species' ecotype to studying individual venom diversity across populations [113]. Individuals can not evolve. Only populations can evolve using the mechanism of natural selection. Individuals only accumulate adaptations that help them survive in a given environment. Populations evolve because individuals with beneficial traits for survival may have more reproductive success, increasing the likelihood of that particular beneficial trait in the collective gene pool of the population. Given that venom represents a key adaptation that has played an important role in the diversification of venomous creatures, assessing the patterns of toxin variation within and between populations need to be considered in mechanistic and evolutionary explanations for venom variation among conspecific animals at the individual, population and species levels. The rise of top-down venomomics is going to be catalyzed by the imminent advent of comparative snake genomics, fostered by the increasing availability in the next few coming years of genomic resources from species across all the major branches of the evolutionary tree of advanced snakes (Serpentes: Caenophidia) [114]. Genome sequencing represents a fundamental step towards setting snakes up to become model organisms. Integrating comparative snake genomics and proteoform-resolved functional venomomics (toxicovenomics [115, 116]) into a comprehensive evolutionary

framework will revolutionize the field of molecular toxinology in the coming years. Understanding the natural history and evolutionary pressures that shaped the complexity of extant snake venoms is relevant in basic toxinology for unveiling the molecular mechanisms that underlie venom variability and to explore the enormous potential of venoms as sources of chemical and pharmacological novelty. In turn, this basic knowledge is of applied importance for the design of novel, safer and more effective toxin-specific therapeutic antivenoms of broader therapeutic use [74, 117]. Clearly, basic and translational venom research represent two sides of the same coin. The analytical innovations recently implemented in venomomics workflows, as well as others that will presumably be incorporated over the next 5 years, hold promises of an exciting future for the emerging field of integrative venomomics.

Key issues

- Established in the 1990s, as a powerful analytical technique, biological mass spectrometry (proteomics) has revolutionized the scope and methods of modern biological investigations.
- The development of quantitative proteomics techniques has represented a major advance in the field over the first decade of the XXI century.
- After more than three decades of bottom-up biology in the gas phase dominating the field, top-down proteomics is gaining momentum.
- Top down venomomics is in its infancy. However, recent hardware and software developments allow us to predict the growing importance of top-down strategies in venom research in the near future.
- The trend towards hybrid configurations of mass analyzers, to combine the different performance characteristics offered by the individual in-space and in-time analyzers into one compact instrument, has dominated recent advances in biological mass spectrometry instrumentation.
- Since its introduction in the 1980s, inductively coupled plasma mass spectrometry (ICP-MS) has evolved to become a versatile element-specific detection technique for absolute quantification of heteroatom isotopes.
- Coinciding developments in top-down venomomics workflows and in the application of absolute protein quantification through ICP-MS/MS interference-free

measurement of sulfur, offer the possibility of parallel identification and absolute quantification of proteoform-resolved venom proteomes.

- The development of a novel molecular and elemental mass spectrometry configuration will represent a quantitative leap in the study of individual venoms.
- Heteroatom-targeted proteomics in multiplexed formats may also bring about new opportunities for absolute quantitative biological workflows and interdisciplinary developments.

Funding

This paper was supported by grant BFU2013-42833-P, from the Ministerio de Economía y Competitividad, Madrid, Spain.

Declaration of interest

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

References

Papers of special note have been highlighted as:

* of interest

** of considerable interest

1. Juárez P, Sanz L, Calvete JJ. Snake venomomics: characterization of protein families in *Sistrurus barbouri* venom by cysteine mapping, N-terminal sequencing, and tandem mass spectrometry analysis. *Proteomics* 4, 327-38 (2004)
2. Calvete, J.J. Snake venom disintegrins and disintegrin-like domains: soluble antagonists and cellular ligands of integrin receptors. *Integrin-Ligand Interactions* (Eble, J. & Kühn, K., eds), Landes Co. Biomedical Publishers, Austin, Texas, (1997), pp. 157-73. ISBN: 3-540-62539-9
3. Juárez P, Comas I, González-Candela F, et al. Evolution of snake venom disintegrins by positive Darwinian selection. *Mol. Biol. Evol.* 25, 2391-407 (2008)

4. Carbajo RJ, Sanz L, Perez A, et al. NMR structure of bitistatin – a missing piece in the evolutionary pathway of snake venom disintegrins. *FEBS J.* 282, 341-60 (2015)
5. Horton ER, Humphries JD, James J, et al. The integrin adhesome network at a glance. *J. Cell Sci.* 129, 4159-63 (2016)
6. Brahma RK, McCleary RJ, Kini RM, et al. Venom gland transcriptomics for identifying, cataloging, and characterizing venom proteins in snakes. *Toxicon* 93, 1-10 (2015)
7. Kaas Q, Craik DJ. Bioinformatics-Aided Venomics. *Toxins* 7, 2159-87 (2015)
8. Calvete JJ. Proteomic tools against the neglected pathology of snake bite envenoming. *Expert Rev. Proteomics* 8, 739-58 (2011).
9. Calvete JJ. Snake venomics: from the inventory of toxins to biology. *Toxicon* 75, 44-62 (2013)
10. Calvete JJ. Next-generation snake venomics: protein-locus resolution through venom proteome decomplexation. *Expert Rev. Proteomics* 11, 315-29 (2014)
11. Zelanis A, Tashima AK. Unraveling snake venom complexity with 'omics' approaches: challenges and perspectives. *Toxicon* 87, 131-4 (2014)
12. Eichberg S, Sanz L, Calvete JJ, et al. Constructing comprehensive venom proteome reference maps for integrative venomics. *Expert Rev. Proteomics* 12, 557-73 (2015)
- * Describes a snake venomics platform that has contributed to the transition from the inventory of a venom's toxins to an integrative venom research discipline.
13. Evers CE, Gaskell SJ. (eds.) *Quantitative proteomics*, Royal Society of Chemistry, 2014. ASIN: B00QCKOJ68
14. Fox JW, Serrano SM. Exploring snake venom proteomes: multifaceted analyses for complex toxin mixtures. *Proteomics* 8, 909-20 (2008)
15. Tasoulis T, Isbister GK. A Review and Database of Snake Venom Proteomes. *Toxins* 9, E290 (2017)
16. Natale M, Caiazzo A, Bucci EM, et al. A novel Gaussian extrapolation approach for 2D gel electrophoresis saturated protein spots. *Genomics Proteomics*

Bioinformatics 10, 336-44 (2012)

17. Blein-Nicolas M, Zivy M. Thousand and one ways to quantify and compare protein abundances in label-free bottom-up proteomics. *Biochim. Biophys. Acta Proteins Proteomics* 1864, 883–95 (2016)

18. Jarnuczak AF, Lee DC, Lawless C, et al. Analysis of Intrinsic Peptide Detectability via Integrated Label-Free and SRM-Based Absolute Quantitative Proteomics. *J. Proteome Res.* 15, 2945-59 (2016)

19. Bantscheff M, Lemeer S, Savitski MM, et al. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem.* 404, 939-65 (2012)

20. Old WM, Meyer-Arendt K, Aveline-Wolf L, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* 4, 1487-502 (2005)

21. Goeminne LJE, Gevaert K, Clement L. Experimental design and data-analysis in label-free quantitative LC/MS proteomics: A tutorial with MSqRob. *J. Proteomics* 171, 23-36 (2018)

22. O'Connell JD, Paulo JA, O'Brien JJ, et al. Proteome-Wide Evaluation of Two Common Protein Quantification Methods. *J. Proteome Res.* 17, 1934-42 (2018)

23. Ross PL, Huang YN, Marchese JN, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 3, 1154-69 (2004)

24. Quaglia M, Pritchard C, Hall Z, et al. Amine-reactive isobaric tagging reagents: requirements for absolute quantification of proteins and peptides. *Anal. Biochem.* 379, 164–9 (2008)

25. Wiese S, Reidegeld KA, Meyer HE, et al. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* 7, 340-50 (2007). Erratum in: *Proteomics* 7, 1004 (2007)

26. Zelanis A, Tashima AK, Pinto AF, et al. Bothrops jararaca venom proteome rearrangement upon neonate to adult transition. *Proteomics* 11, 4218-28 (2011)

27. Gao JF, Qu YF, Zhang XQ, et al. Neonate-to-adult transition of snake venomomics in the short-tailed pit viper, *Gloydus brevicaudus*. *J. Proteomics* 84, 148-57

(2013)

28. Brun V, Masselon C, Garin J, et al. Isotope dilution strategies for absolute quantitative proteomics. *J. Proteomics* 72, 740-9 (2009)
29. Villanueva J, Carrascal M, Abian J. Isotope dilution mass spectrometry for absolute quantification in proteomics: concepts and strategies. *J. Proteomics* 96, 184-99 (2014)
30. Houk RS, Tassel VA, Flesch GD, et al. Inductively coupled argon plasma as an ion source for mass spectrometric determination of trace elements. *Anal. Chem.* 53, 2283-9 (1980)
31. Becker JS. *Inorganic Mass Spectrometry. Principles and Applications*. John Wiley & Sons Ltd., Chichester, UK, pp 118-176 (2007). ISBN 978-0-470-01200-0
32. Bettmer J, Montes-Bayón M, Ruiz Encinar J, et al. The emerging role of ICP-MS in proteomic analysis. *J. Proteomics* 72, 989-1005 (2009)
33. Sanz-Medel A, Montes-Bayón M, Bettmer J, et al. ICP-MS for absolute quantification of proteins for heteroatom-tagged, targeted proteomics. *Trends Anal. Chem.* 40, 52-63 (2012)
34. Calderón-Celis F, Sanz-Medel A, Encinar JR. Universal absolute quantification of biomolecules using element mass spectrometry and generic standards. *Chem. Commun.* 54, 904-7 (2018)
- * A comprehensive review of the power of elemental mass spectrometry for the absolute quantification of heteroatoms in proteomics studies.
35. Díez Fernández S, Sugishama N, Ruiz Encinar J, et al. Triple quad ICPMS (ICPQQQ) as a new tool for absolute quantitative proteomics and phosphoproteomics. *Anal Chem.* 84, 5851-7 (2012)
- ** Reports a new way of polyatomic interference removal that enabled the absolute quantification of non-metallic elements present in proteins.
36. Rappel C, Schaumlöffel D. The role of sulfur and sulfur isotope dilution analysis in quantitative protein analysis. *Anal. Bioanal. Chem.* 390, 605-15 (2008)
37. Reeks TA, Fry BG, Alewood PF. Privileged frameworks from snake venom. *Cell. Mol. Life. Sci.* 72, 1939-58 (2015)
38. Calvete JJ, Juárez P, Sanz L. Snake venomomics. Strategy and applications. *J.*

Mass Spectrom. 42, 1405-14 (2007)

39. Calderón-Celis F, Díez-Fernández S, Costa-Fernández JM, et al. Elemental Mass Spectrometry for Absolute Intact Protein Quantification without Protein-Specific Standards: Application to Snake Venomics. *Anal. Chem.* 88, 9699-706 (2016)

40. Calderón-Celis F, Cid-Barrio L, Encinar JR, Sanz-Medel A, Calvete JJ. Absolute venomics: Absolute quantification of intact venom proteins through elemental mass spectrometry. *J. Proteomics* 164, 33-42 (2017)

41. Calderón-Celis F, Encinar JR, Sanz-Medel A. Standardization approaches in absolute quantitative proteomics with mass spectrometry. *Mass Spectrom. Rev.* doi: 10.1002/mas.21542, 1-23 (2017)

42. Calvete JJ. The Expanding Universe of Mass Analyzer Configurations for Biological Analysis. *Meth. Mol. Biol.* 1072, 61-81 (2014)

43. Senko MW, Remes PM, Canterbury JD, et al. Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. *Anal. Chem.* 85, 11710-4 (2013)

44. Calvete JJ. Venomics: integrative venom proteomics and beyond. *Biochem J.* 474, 611-34 (2017)

45. Casewell NR, Wüster W, Vonk FJ, et al. Complex cocktails: the evolutionary novelty of venoms. *Trends Ecol Evol.* 28, 219-29 (2013)

46. Fry BG, Vidal N, Norman JA, et al. Early evolution of the venom system in lizards and snakes. *Nature* 439, 584-8 (2006)

47. Jungblut PR, Schlüter H. Towards the analysis of protein species: an overview about strategies and methods. *Amino Acids* 41, 219-22 (2011)

48. Patrie SM. Top-Down Mass Spectrometry: Proteomics to Proteoforms. *Modern Proteomics-Sample Preparation, Analysis and Practical Applications* (Mirzaei H, and Carrasco M, eds.). *Advances in Experimental Medicine and Biology* 919 (2016). DOI 10.1007/978-3-319-41448-5_8.

49. Denisov E, Damoc E, Lange O, et al. Orbitrap mass spectrometry with resolving powers above 1,000,000. *Int. J. Mass Spectrom.* 27, 80-5 (2012)

50. Scheffler K, Viner R, Damoc E. High resolution top-down experimental strategies

on the Orbitrap platform. *J. Proteomics* 175, 42-55 (2018)

51. Lorenzatto KR, Kim K, Ntai I, et al. Top Down Proteomics Reveals Mature Proteoforms Expressed in Subcellular Fractions of the *Echinococcus granulosus* Preadult Stage. *J. Proteome Res.* 14, 4805-14 (2015)

52. Valaskovic GA, Kelleher NL, McLafferty FW. Attomole protein characterization by capillary electrophoresis-mass spectrometry. *Science* 273, 1199-202 (1996)

53. Bogdanov B, Smith RD. Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass Spectrom. Rev.* 24, 168-200 (2005)

54. Petras D, Heiss P, Süssmuth RD, et al. Venom Proteomics of Indonesian King Cobra, *Ophiophagus hannah*: Integrating Top-Down and Bottom-Up Approaches. *J. Proteome Res.* 14, 2539-56 (2015)

* First report of the application of top-down MS in snake venomomics

55. Melani RD, Skinner OS, Fornelli L, et al. Mapping Proteoforms and Protein Complexes From King Cobra Venom Using Both Denaturing and Native Topdown Proteomics. *Mol Cell Proteomics* 15, 2423-34 (2016)

56. Göçmen B, Heiss P, Petras D, et al. Mass spectrometry guided venom profiling and bioactivity screening of the Anatolian Meadow Viper, *Vipera anatolica*. *Toxicon* 107, 163-74 (2015)

57. Petras D, Heiss P, Harrison RA, et al. Top-down venomomics of the East African green mamba, *Dendroaspis angusticeps*, and the black mamba, *Dendroaspis polylepis*, highlight the complexity of their toxin arsenals. *J. Proteomics* 146, 148-64 (2016)

58. Ainsworth S, Petras D, Engmark M, et al. The medical threat of mamba envenoming in sub-Saharan Africa revealed by genus-wide analysis of venom composition, toxicity and antivenomics profiling of available antivenoms. *J. Proteomics* 172, 173-89 (2018)

59. Pla D, Petras D, Saviola AJ, et al. Transcriptomics-guided bottom-up and topdown venomomics of neonate and adult specimens of the arboreal rear-fanged Brown Treesnake, *Boiga irregularis*, from Guam. *J. Proteomics* 174, 71-84 (2018)

60. Ntai I, Kim K, Fellers RT, et al. Applying label-free quantitation to top down proteomics. *Anal. Chem.* 86, 4961-8 (2014)

61. Dole M, Mach LL, Hines RL, et al. Molecular beams of macroions. *J. Chem. Phys.* 49, 2240-9 (1968)
62. Fenn JB, Mann M, Meng CK, et al. Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246, 64-71 (1989)
- ** The development of ESI-MS opened the door to addressing biological studies in gas phase, a new paradigm that has since revolutionized the field.
63. Fenn JB. Electrospray wings for molecular elephants (Nobel lecture). *Angew. Chem. Int. Ed. Engl.* 42, 3871-94 (2003)
64. Henry KD, Williams ER, Wang BH, et al. Fourier-transform mass spectrometry of large molecules by electrospray ionization. *Proc. Natl. Acad. Sci. U.S.A.* 86, 9075-8 (1989)
65. Henry KD, McLafferty FW. Electrospray ionization with Fourier-transform mass spectrometry. Charge state assignment from resolved isotopic peaks. *Org. Mass Spectrom.* 25, 490-2 (1990)
66. Henry KD, Quinn JP, McLafferty FW. High-resolution electrospray mass spectra of large molecules. *J. Am. Chem. Soc.* 113, 5447-9 (1991)
67. Loo JA, Quinn JP, Ryu SI, et al. High resolution tandem mass spectrometry of large biomolecules. *Proc. Natl. Acad. Sci. U.S.A.* 89, 286-9 (1992)
68. McLafferty F. High-resolution tandem FT mass spectrometry above 10 kDa. *Acc. Chem. Res.* 27, 379-86 (1994)
69. McLafferty FW. A century of progress in molecular mass spectrometry. *Annu. Rev. Anal. Chem.* 4, 1-22 (2011)
70. Fenn JB. Electrospray ionization mass spectrometry: How it all began. *J. Biomol. Tech.* 13, 101-18 (2002)
71. Dang X, Scotcher J, Wu S, et al. The first pilot project of the consortium for topdown proteomics: a status report. *Proteomics* 14, 1130-40 (2014)
72. Toby TK, Fornelli L, Kelleher NL. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* 9, 499-519 (2016)
73. Vonk FJ, Casewell NR, Henkel CV, et al. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc. Natl.*

Acad. Sci. USA 110, 20651-6 (2013)

* This first report of the sequence of a venomous snake genome provided insights into the origin and mechanisms of venom toxins evolution under strong selective pressure.

74. Gutiérrez JM, Calvete JJ, Habib AG, et al. Snakebite envenoming. Nat. Rev. Dis. Primers 3, 17079 (2017)

75. Utkin Y, Sunagar K, Jackson T, et al. Three-finger toxins (3FTxs). In: Venomous reptiles and their toxins: evolution, pathophysiology and biodiscovery, Fry BG (ed.), Oxford University Press, Oxford, 215-27 (2015)

76. Harvey AL, Robertson B. Dendrotoxins: structure-activity relationships and effects on potassium ion channels. Curr. Med. Chem. 11, 3065-72 (2004).

77. Jackson TN, Koludarov I, Ali SA, et al. Rapid Radiations and the Race to Redundancy: An Investigation of the Evolution of Australian Elapid Snake Venoms. Toxins 8, E309, (2016)

78. Lomonte B, Rey-Suárez P, Fernández J, et al. Venoms of *Micrurus* coral snakes: Evolutionary trends in compositional patterns emerging from proteomic analyses. Toxicon 122, 7-25 (2016)

79. Sunagar K, Jackson TN, Undheim EA, et al. Three-fingered RAVeRs: Rapid Accumulation of Variations in Exposed Residues of snake venom toxins. Toxins 5, 2172-208 (2013)

80. Skinner OS, Do Vale LHF, Catherman AD, et al. Native GELFrEE: A New Separation Technique for Biomolecular Assemblies. Anal. Chem. 87, 3032-8 (2015)

81. Savidge JA. Food habits of *Boiga irregularis*, an introduced predator on Guam. J. Herpetol. 22, 275-82 (1988)

82. Weinstein SA, Chiszar D, Bell RC, et al. Lethal potency and fractionation of Duvernoy's secretion from the Brown Treesnake, *Boiga irregularis*. Toxicon 29 401-7 (1991)

83. Pawlak J, Mackessy SP, Sixberry NM, et al. Irditoxin, a novel covalently linked heterodimeric three-finger toxin with high taxon-specific neurotoxicity. FASEB J. 23, 534-45 (2009)

84. Mackessy SP, Sixberry NM, Heyborne WH, et al. Venom of the brown treesnake, *Boiga irregularis*: ontogenetic shifts and taxa-specific toxicity. *Toxicon* 47, 537-48 (2006)
85. Kou, Q, Xun L, Liu X. TopPIC: a software tool for top-down mass spectrometry based proteoform identification and characterization. *Bioinformatics* 32, 3495–7 (2016)
86. Cai W, Guner H, Gregorich ZR, et al. MASH suite pro: a comprehensive software tool for top-down proteomics. *Mol. Cell. Proteomics* 15, 703–14 (2016)
- * TopPIC and MASH represent a new generation of software tools that provide statistical estimates for the confident inference of unknown primary structure alterations in known protein entries.
87. Harvey AL. Toxins and drug discovery. *Toxicon* 92, 193-200 (2014)
88. King GF. (editor). *Venoms to Drugs: Venom as a Source for the Development of Human Therapeutics*, Royal Society of Chemistry Publishing, Cambridge (UK), 2015. ISBN 978-1-84973-663-3
89. Proteomic analysis of venom variability and ontogeny across the arboreal palmpitvipers (genus *Bothriechis*). *J. Proteomics* 152, 1-12 (2017)
90. Arbuckle K, Rodríguez de la Vega RC, Casewell NR. Coevolution takes the sting out of it: Evolutionary biology and mechanisms of toxin resistance in animals. *Toxicon* 140, 118-31 (2017)
91. Panagides N, Jackson TN, Ikonopoulou MP, et al. How the Cobra Got Its Flesh-Eating Venom: Cytotoxicity as a Defensive Innovation and Its Co-Evolution with Hooding, Aposematic Marking, and Spitting. *Toxins* 9, E103 (2017)
92. Cipriani V, Debono J, Goldenberg J, et al. Correlation between ontogenetic dietary shifts and venom variation in Australian brown snakes (*Pseudonaja*). *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* 197, 53-60 (2017)
93. Goldenberg J, Cipriani V, Jackson TNW, et al. Proteomic and functional variation within black snake venoms (Elapidae: *Pseudechis*). *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* 205, 53-61 (2018)
94. Shine R, Covacevich J. Ecology of highly venomous snakes: the Australian genus *Oxyuranus* (Elapidae). *J. Herpetol.* 17, 60-69 (1983)

95. Laustsen AH, Lohse B, Lomonte B, et al. Selecting key toxins for focused development of elapid snake antivenoms and inhibitors guided by a Toxicity Score. *Toxicon* 104, 43-5 (2015)
96. Calvete JJ, Rodríguez Y, Quesada-Bernat S, et al. Toxin-resolved antivenomics guided assessment of the immunorecognition landscape of antivenoms. *Toxicon* 148, 107-22 (2018)
97. Huang T, Wang J, Yu W, et al. Protein inference: a review. *Brief Bioinform.* 13, 586-614 (2012)
98. Park J, Piehowski PD, Wilkins C, et al. Informed-Proteomics: open-source software package for top-down proteomics. *Nat. Methods* 14, 909-14 (2017)
99. Fornelli L, Durbin KR, Fellers RT, et al. Advancing Top-down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer. *J. Proteome Res.* 16, 609–18 (2017)
- ** Description of an optimized top-down proteomics platform with updated metrics and tools for the high-throughput identification and reporting of proteoforms >30 kDa.
100. Anderson LC, DeHart CJ, Kaiser NK, et al. Identification and Characterization of Human Proteoforms by Top-Down LC-21 Tesla FT-ICR Mass Spectrometry. *J. Proteome Res.* 16, 1087-96 (2017)
101. Greer SM, Brodbelt JS. Top-Down Characterization of Heavily Modified Histones Using 193 nm Ultraviolet Photodissociation Mass Spectrometry. *J Proteome Res.* 17, 1138-45 (2018)
102. Gargano AFG, Roca LS, Fellers RT, et al. Capillary HILIC-MS: A New Tool for Sensitive Top-Down Proteomics. *Anal. Chem.* 90, 6601-9 (2018)
103. McCool EN, Lubeckj RA, Shen X, et al. Deep Top-Down Proteomics Using Capillary Zone Electrophoresis-Tandem Mass Spectrometry: Identification of 5700 Proteoforms from the Escherichia coli Proteome. *Anal. Chem.* 90, 5529-33 (2018)
104. Lermyte F, Valkenborg D, Loo JA, et al. Radical solutions: Principles and application of electron-based dissociation in mass spectrometry-based analysis of protein structure. *Mass Spectrom. Rev.* doi: 10.1002/mas.21560, (2018)

105. Li H, Nguyen HH, Ogorzalek Loo RR, et al. An integrated native mass spectrometry and top-down proteomics method that connects sequence to structure and function of macromolecular complexes. *Nat. Chem.* 10, 139-48 (2018)
106. Kou Q, Wu S, Liu X. Systematic Evaluation of Protein Sequence Filtering Algorithms for Proteoform Identification Using Top-Down Mass Spectrometry. *Proteomics* 18, doi: 10.1002/pmic.201700306, (2018)
107. Chen B, Brown KA, Lin Z, et al. Top-Down Proteomics: Ready for Prime Time? *Anal. Chem.* 90, 110-27 (2018)
108. Daly M, Gibbs HL. Integrating Perspectives on Animal Venom Diversity: An Introduction to the Symposium. *Integr. Comp. Biol.* 56, 934-7 (2016)
109. Szpunar J. Advances in analytical methodology for bioinorganic speciation analysis: metallomics, metalloproteomics and heteroatom-tagged proteomics and metabolomics. *Analyst* 130, 442-65 (2005)
110. Sanz-Medel A. ICP-MS for multiplex absolute determinations of proteins. *Anal. Bioanal. Chem.* 398, 1853-9 (2010)
111. Melani RD, Nogueira FCS, Domont GB. It is time for top-down venomomics. *J. Venom Anim. Toxins Incl. Trop. Dis.* 23, 44 (2017)
112. Calvete JJ, Petras D, Calderón-Celis F, et al. Protein- species quantitative venomomics: Looking through a crystal ball. *J. Venom. Anim. Toxins Incl. Trop. Dis.* 23, 1 (2017)
113. Amazonas DR, Portes-Junior JA, Nishiyama-Jr MY, et al. Molecular mechanisms underlying intraspecific variation in snake venom. *J. Proteomics* 181, 60-72 (2018)
114. Kerkkamp HM, Kini RM, Pospelov AS, et al. Snake Genome Sequencing: Results and Future Prospects. *Toxins* 8, E360 (2016)
115. Lauridsen LP, Laustsen AH, Lomonte B, et al. Toxicovenomics and antivenom profiling of the Eastern green mamba snake (*Dendroaspis angusticeps*). *J. Proteomics* 136, 248-61 (2016)
116. Lauridsen LP, Laustsen AH, Lomonte B, et al. Exploring the venom of the forest cobra snake: Toxicovenomics and antivenom profiling of *Naja melanoleuca*. *J.*

Proteomics 150, 98-108 (2017)

117. Calvete JJ, Sanz L, Pla D, et al. Omics meets biology: application to the design and preclinical assessment of antivenoms. *Toxins* 6, 3388-405 (2014)

118. Rodríguez-González P, Marchante-Gayón JM, García Alonso JI, et al. Isotope dilution analysis for elemental speciation: a tutorial review. *Spectrochim Acta Part B At. Spectrosc.* 60, 151-207 (2005)

119. Pyron RA, Burbrink FT, Wiens JJ. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evol. Biol.* 13, 93 (2013)

201. The Consortium for Top-Down Proteomics website

<http://www.topdownproteomics.org/resources/software>

Figures legends

Figure 1. Scheme of different configurations that combine analytical decomplexation, quantification, and toxin identification strategies used in snake venomomics projects developed in different laboratories. Our bottom-up peptide-centric snake venomomics approach involves the simultaneous venom fractionation and relative quantification of the venom components (combination 1c-2b-3c) identified from surrogate tryptic peptides (3a) using homologous transcriptomic dataset-aided (4a) or de novo (4c) interpretation of tryptic peptide MS/MS spectra. In top-down venomomics, we select configuration 1b-2a (online-LC-MS separation and mass profiling of venom components) followed by transcriptomic-aided identification of the fragmentation pattern of the selected topoisomers (4a) and label-free quantification (4b) through the integration of normalized peak areas or peak intensities of extracted ion chromatograms of full-scan spectra (MS1).

Figure 2. Schematic workflow of the parallel hybrid elemental and molecular mass spectrometry configuration developed for absolute quantitative venomomics (adapted from [34]). The venom sample is spiked with a generic S-containing internal standard, and separated by RP- μ HPLC. Heteroatom (^{32}S) absolute content of standard and analytes is accomplished by on-line species-unspecific ^{34}S isotope dilution analysis. Parallel ESI-QToF mass profiling, along with sequence information

gathered from transcriptomic or peptide-centric venomic analysis, was used to identify the venom proteins across the RP- μ HPLC separation. This structural information provides the sulfur:protein stoichiometry. In order to maintain the sulfur response factor constant along all the chromatogram, ^{32}S and ^{34}S signal intensities were related and transformed into a mass flow applying the mass flow equation, derived from the isotope dilution equation. The absolute amount of sulfur in the different protein fractions can be computed using the mass flow equation [118], where MFS and MFS_p are the mass flows of the sample (S) and the spike (Sp), respectively; WS and WSp are the weighted sample and spike masses, respectively; AwS and AwSp, the sulfur atomic weight in the sample and spike, respectively; AS_p 34, the ^{34}S abundance in the spike; AS₃₂, the ^{32}S abundance in the sample; R_b 32/34, the $^{32}\text{S}/^{34}\text{S}$ ratio in the mixture; RSp 32/34, the $^{32}\text{S}/^{34}\text{S}$ ratio in the spike; and RS 32/34 is the $^{34}\text{S}/^{32}\text{S}$ ratio in the sample. The mass of S for the internal standard (mstd, known) and the mass of S for the species of interest (ms, unknown) are related to the areas under the corresponding peaks (Astd and As, respectively), which must be integrated in the same mass-flow chromatogram.

Figure 3. Scheme of the top-down MS workflow employed for the identifications of proteoforms in the venom proteome of *Dendroaspis viridis* [58]. Reduced venom components were fractionated by RP-HPLC (a) on an Agilent 1260 HPLC system and the molecular masses of the eluate recorded in an Orbitrap LTQ XL mass spectrometer with a mass resolution (R) of 100,000 (at m/z 400) (b). For MS/MS the two most abundant ions of the survey scan with known charge were selected. MS/MS spectra were obtained in data-dependent acquisition (DDA) mode (c). Representative TopPIC annotated top-down MS/MS spectra of the quasi-isobaric monoisotopic 7+ topoisomers of reverse-phase HPLC peak 9 of *D. viridis* venom (m/z 6570,008) (upper panel), and reverse-phase HPLC peak 7 of *D. viridis* venom (m/z 6569,302) (lower panel). Mambalgins-3 [P00982] and delta-dendrotoxin [C0HJB0] were identified using TopPIC with p -values of $2.65e-34$ and $5.88e-43$, respectively. Reproduced from Ainsworth et al. [58] under Creative Commons CCBY-NC-ND license attributes.

Figure 4. Identification of proteoforms by top-down MS/MS. Panels A and B, deconvoluted top-down MS/MS spectra of the monoisotopic 8+ topoisomers of the protein ions at m/z 6561.3 and 6603.3, identified as reduced Dendrotoxin K [P00981] and its Lys-24 modified proteoform, respectively. Sequence-specific product ions bearing the mass increment of +42 Da, that allowed assignment of the specific site of modification, are highlighted in panel A with arrows. Panel C, topdown MS/MS fragmentation pattern of the monoisotopic 6+ topoisomer of *D. viridis* m/z 1127.55 ion was assigned by TopPIC to an unknown proteoform of short - neurotoxin S5C10 [P01419] (p-value of $9.15\text{e-}20$) bearing an unexplained mass discrepancy of +242,16 Da [56]. When the top-down data were re-analysed against the species-specific transcriptome [57], D instead of N at position 9 and two additional residues (KI) at positions 50-51 fully explained the mass difference of +242,16 Da. Panels A-C are reproduced from Petras et al. [57] and Ainsworth et al. [58] under Creative Commons CC-BY-NC-ND license attributes.

Figure 5. Panel A, phylogenetic framework of extant Elapidae [119] highlighting lineages expressing venoms rich in 3FTxs (red-filled circles) and group I PLA2s (green-filled circles). Clades where dichotomous venom distribution has been

documented are labeled with bicolored circles: pale brown/red, dendrotoxin (KUN)/3FTx; red/green, 3FTx/PLA₂. Panel B, Bayesian inference phylogenetic tree of the relationships among the five taxa within genus *Dendroaspis* highlighting the distribution of the major 3FTxs and KUN toxin proteoforms assigned by top-down MS [57].

Figure 6. Scheme of a hypothetical hybrid mass spectrometry configuration for topdown locus-resolved proteoform identification and simultaneous absolute quantification. The elementary modules of such proposed hybrid configuration exist as independent functional units. The challenge for the next future is to arrange the pieces together into a single functional configuration capable of identifying, and in parallel quantify, the individual components of a (venom) proteome.