

Spatial Venomics - Cobra Venom System Reveals Spatial Differentiation of Snake Toxins by Mass Spectrometry Imaging

Benjamin-Florian Hempel^{1, 2, ‡, *}, Maik Damm^{2, ‡}, Daniel Petras³, Taline D. Kazandjian⁴, Claudia A. Szentiks⁵, Guido Fritsch⁵, Grit Nebrich¹, Nicholas R. Casewell⁴, Oliver Klein^{1, #} and Roderich D. Süssmuth^{2, #, *}

¹ BIH Center for Regenerative Therapies BCRT, Charité - Universitätsmedizin Berlin, 13353 Berlin, Germany.

² Institut für Chemie, Technische Universität Berlin, 10623 Berlin, Germany.

³ CMFI Cluster of Excellence, Interfakultäres Institut für Mikrobiologie und Infektionsmedizin Tübingen, Universität Tübingen, 72076 Tübingen, Germany.

⁴ Centre for Snakebite Research & Interventions, Liverpool School of Tropical Medicine, Liverpool, L3 5QA, United Kingdom.

⁵ Department of Wildlife Diseases and Reproduction Management, Leibniz-Institut für Zoo- und Wildtierforschung (IZW) im Forschungsverbund Berlin e.V., 10315 Berlin, Germany.

ABSTRACT: Among venomous animals, toxic secretions have evolved as biochemical weapons associated with various highly specialized delivery systems on many occasions. Despite extensive research, there is still limited knowledge of the functional biology of most animal toxins, including their venom production and storage, as well as the morphological structures within sophisticated venom producing tissues that might underpin venom modulation. Here we report on the spatial exploration of a snake venom gland system by matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI), in combination with standard proteotranscriptomic approaches, to enable in situ toxin mapping in spatial intensity maps across a venom gland sourced from the Egyptian cobra (*Naja haje*). MALDI-MSI toxin visualization on the elapid venom gland reveals high spatial heterogeneity of different toxin classes at the proteoform level, which may be the result of physiological constraints on venom production and/or storage that reflects the potential for venom modulation under diverse stimuli.

INTRODUCTION

Venoms are sophisticated and complex mixtures consisting of low molecular weight compounds, peptides and proteins, which have evolved for use as defensive and/or foraging adaptations in various animal lineages.^{1,2} Venom research is the focus of various scientific disciplines due to their great medical importance, i.e. the generation of effective immunological therapies against envenoming and the discovery of novel drugs from venoms.^{3,4} The structural and functional diversity of venom systems has been triggered by a variety of adaptive evolutionary processes, such as convergence², co-option of single copy genes⁵, gene duplication⁶, accelerated rates of molecular evolution⁷, and protein neofunctionalization⁸. Thus, collectively venoms have proven to be valuable models for examining the origins of adaptations and the link between genotype and functional phenotypes.

Among venomous animals, serpents have received the most scientific attention due to their frequently lethal encounters with human.³ Snake venoms are among the most lethal bioweapons in nature that are produced, stored and released by a sophisticated venom delivery system in the upper jaw.⁹ This secretion, produced in specialized tissues and embedded in the venom gland apparatus, causes a cascade of physiological and biochemical perturbations once delivered into the target (e.g. prey, predator or aggressor).² Despite the diverse variety of venomous species with highly complex and multifunctional venom cocktails, in particular the toxic arsenal in venomous snakes is a characteristic trait and knowledge of its molecular composition has aid to categorize major serpent families, as venom toxin variation is ubiquitous across multiple taxonomic levels.¹⁰⁻¹² Venoms are energetically expensive commodities and the degree of toxin diversity in relation to the morphological attributes of the venom system substantially dependent on the degree to which an animal relies on biochemical versus physical means for overpowering prey.¹³ Although mass spectrometry-based (MS) venom proteomics and sequencing-based methods facilitate detailed insight into venom diversity, classical proteomics and transcriptomics analyses using venom secretions or even tissue homogenates are not able to correlate and spatially resolve single toxins within cellular environments.¹⁴

Matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI) can act at the interface of venom diversity and spatial localization to bypass existing limitations of classical venomics approaches. It allows unsupervised and simultaneous analyses of molecules (e.g., metabolites, proteins, peptides, lipids and glycans) in situ on a single tissue section, preserving their spatial coordinates and generating a molecular intensity map reflecting the relative molecule abundance.^{15,16,14} Since Caprioli et al. introduced MALDI-MSI for biological samples in the late 90s, continual technical improvements has raised its suitability and applicability to become a valuable tool for use in the field of venomics.^{16,17} Hence, it facilitates mapping of individual toxins on-tissue and provides deeper insights into the spatial venom distribution as well as heterogeneity and its effects on venom regulation or modulation.¹⁸ The localization of different venom toxins within the context of morphological structures has previously been described by MALDI-MSI in a multidimensional manner.¹⁹⁻²² However, previous MSI experiments are restricted in their venom identification, which allows intact profiling of selected toxin classes and excludes the identification of higher molecular weight toxins, like snake venom serine proteases (svSP), snake venom metalloproteases (svMP), or L-amino acid oxidases (LAAO).

Here we demonstrate the feasibility and potential of MALDI-MSI, in combination with a global venom analysis via a proteotranscriptomics approach, as the basis for the spatial identification of different toxins associated with the venom system of the Egyptian cobra (*Naja haje*, formerly *N. h. legionis*, locale: Morocco).¹⁸ Our spatial venom analysis helps to place the simple and centralized elapid venom system in an evolutionary context, gaining insight into the complexity of toxin compartmentalization, which may be the result of physiological constraints on venom production and/or storage, or reflect the potential for venom modulation under different stimuli.

EXPERIMENTAL SECTION

Specimen, venom collection and venom gland dissection. The venom was obtained from a single Moroccan male Egyptian *Naja haje* (*legionis*) individual, lyophilized and stored at +4 °C (**Figure S1**). The adult specimen of unknown age (>10 years old; snout to vent length of 164 cm, total length 194 cm) was kept at the Tierpark Berlin (Germany) since 2007 and euthanized in August 2017, due to tumorous occurrences. The healthy and intact venom glands were removed by veterinary supervision, snap frozen in liquid nitrogen, and stored at -80 °C until further processing (**Figure S1**).

Computer tomography (CT). CT scans were performed with the 16 cm z-axis coverage and 0.5 mm resolution computer tomography system Aquilion ONE Genesis Edition (Canon Medical Systems, Zoetermeer, Netherlands) at the Leibniz-Institut für Zoo- und Wildtierforschung (IZW, Berlin, Germany). The postmortem *in situ* positioning on skeletal and soft tissue was documented using virtual coneXact double slice section technology with 640 slices every rotation (scan parameter: 120.0 kV, 200.0 mA, slice thickness 0.25 mm). Three-dimensional reconstructions were established by vitrea 2 software (Canon Vital Images, Veenendaal, Netherlands).

Histology and Histochemistry. Venom gland was fixed using 4% formalin in phosphate-buffered solution for 48 hours at 4 °C in the dark with a fixative volume five times of the tissue volume at least. The formalin-fixed tissue was placed in an embedding cassette and routinely processed by a fully-automated tissue processor Excelsior AS (Thermo Scientific, Bremen, Germany) with a solution series for stepwise dehydration and paraffin embedding. Venom gland sections were cut at 3 µm thickness using a microtome HM 340E (Thermo Fisher Scientific, Bremen, Germany) and directly placed on superfrost microscope glass slides (Thermo Fisher Scientific, Bremen, Germany). Subsequently, glass slides were incubated at least 3 h at 37 °C for adherence and dewaxed by a reversed solution series (10 min; RT), with the exception of two cycles 100% xylene and distilled water instead of 50% EtOH. Histological staining was performed with haematoxylin and eosin using standard protocols.²³

Venom gland transcriptome. A separate specimen of *Naja haje*, from Uganda, was used for the construction of the venom gland transcriptome. The construction of the assembled transcriptome has previously been reported.^{24,25} For annotation, assembled contigs were batch annotated with BLAST2GO Pro v3²⁶ using the BLASTx-fast algorithm with a significance threshold of 1e-5 against the NCBI non-redundant (NR) protein database. BLAST2GO annotations were inspected manually and preliminary annotation was performed by assigning contigs to toxins based on a combination of the BLAST2GO output and BLASTx searching sequences against the non-redundant protein database (nr). We built a species-specific assembled transcriptome database, extended by various *Naja* species²⁵ and the NCBI protein database (taxid: 8602).

Bottom-up venom proteomics. Proteomic analyses were guided by published protocols and are briefly summarized.^{27–29} Briefly, crude venom was separated using an Agilent 1260 High-pressure Gradient System (Agilent, Waldbronn, Germany) and dried fractions subsequently analyzed by SDS-PAGE (12% polyacrylamide). Afterwards, an in-gel trypsin digestion was performed and extracted peptides analyzed by LC-MS/MS analysis with an Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Bremen, Germany).

Further, complementary protein identification was performed with a specific peptide reference library to appropriate *m/z* values of the MSI measurements. Therefore, crude venom (50 µg) was digested in solution with or without the sample preparation kit (Biognosys, Schlieren, Switzerland) according to the manufacturer's protocol. In addition, we performed on-tissue trypsin digestion with prior paraffin removal and antigen retrieval as described in detail below. Sample was acidified by adding 0.1% (v/v) TFA solution and centrifuged at 10,000 x g for 10 min. Afterwards, all samples were submitted to nUPLC-MS/MS analysis using an analytical UPLC System (Thermo Dionex Ultimate 3000, Acclaim PepMap RSLC C18 column 75 µm x 15 cm; flow rate 400 nL/min, 70 min) coupled to an Impact II (QTOF-MS, Bruker Daltonik) in the data-dependent instant expertise mode.

Top-down mass profiling. Native and denaturing top-down proteomic experiments were performed as described previously and are briefly summarized.^{28,29} In short, venom samples were dissolved in ultrapure water (10 mg/mL) and half of the dissolved sample was mixed with 0.5 M tris(2-carboxyethyl) phosphine

(TCEP), and 0.1 M citrate buffer (pH 3) to reduce disulfide bonds. After incubation, samples were mixed with acetonitrile/formic acid/H₂O (10:1:89, v/v/v) and supernatant of native and reduced sample were injected for LC-MS/MS analyses of two technical replicates to a Q-Exactive hybrid quadrupole orbital ion trap (Thermo Fisher Scientific, Bremen, Germany).

MS data analysis. All raw LC-MS files were converted to open data format files (.mgf, .mzXML) via MSConvert GUI of the ProteoWizard³⁰ cross-platform package.

The bottom-up analysis was performed with search engine PEAKS DB³¹ (version 10.5, Bioinformatics Solutions Inc., Waterloo, Canada). The detailed parameter settings are stated in the supplementary. Identification of MALDI-MSI *m/z* values using the aforementioned peptide reference library requires the accordance of more than one peptide (mass differences ≤ 0.2 Da) to subsequently correctly assign the corresponding protein. Peptides with the lowest mass difference and highest $-\log P$ value compared to the LC-MS/MS reference list value were assumed to be a match in accordance with Cillero-Pastor et al. (2014) guidelines.³² The subsequent peptide assignment ('PAssT', <https://github.com/benjaminhempel/PAssT>) and multivariate analyses were performed with our in-house R scripts.

The top-down proteomics data were deconvoluted to a msalign file using TopFD software. The final sequence annotation was performed by TopPIC³³. The spectra were matched against a combined database of an in-house *Naja* transcriptome (532 entries), *Naja* protein database (1834 entries, NCBI taxid 8638) and cRAP protein sequences (116 entries), manually validated and graphically visualized using the MS and MS/MS spectra by Qual Browser (Thermo Xcalibur 2.2 SP1.48).

The intact mass profile was inspected with the Qual Browser (Thermo Xcalibur 2.2 SP1.48) and FreeStyle (Thermo Xcalibur 1.6.75.20). Deconvolution of isotopically resolved spectra was carried out by using the XTRACT algorithm of Thermo Xcalibur.

Mass spectrometry imaging (MSI) tissue preparation. Venom gland sections with 3 μm thickness were placed directly onto an indium-tin oxide (ITO) glass slide and heated at least 3 h at 37 °C for adherence. Afterwards, section slides were heated for 15 min at 80 °C and dewaxed by repetitive washes of various solvents. Antigen retrieval (AR) was performed for 1 h at 95 °C in deionized water. On-tissue trypsin digestion and subsequent matrix overlay was performed using an automated spraying device (HTX TM-Sprayer, HTX Technologies LLC, Riemerling, Germany). The buffered trypsin solution (20 μg , 20 mM ammonium bicarbonate, 0.1% glycerol) was applied onto the section and after enzymatic incubation of tissue sections (2 h at 50 °C in a moist chamber with a saturated potassium sulfate solution), matrix solution (7 g/L α -cyano-4-hydroxycinnamic acid in 50% acetonitrile and 1% trifluoroacetic acid) was overlaid using the aforementioned HTX Sprayer (75 °C, 120 $\mu\text{L}/\text{min}$).

MSI acquisition. MALDI-Imaging data acquisition was performed on a rapifleX MALDI Tissue typer system (Bruker Daltonik GmbH, Bremen, Germany) operating in reflector mode, *m/z* 600–3200 detection range, 500 laser shots per spot, sampling rate of 1.25 GS/s, and raster width 50 μm . MSI was coordinated by FlexControl 3.0 and FlexImaging 4.0 (Bruker Daltonik GmbH, Bremen, Germany) was used to establish the geometry and location of the section on the slide based upon the optical image and call upon FlexControl to acquire individual spectra, accumulating 200 shots per raster point. External calibration was performed using a peptide calibration standard (Bruker Daltonik GmbH, Bremen, Germany). After the MALDI-imaging experiments, the matrix was removed with 70% ethanol and the tissue sections were stained with haematoxylin and eosin (HE) as histological overview staining.

MSI data analysis and visualization. Data analysis and visualization was performed using SCiLS Lab software (Version2015b, SCiLS GmbH, Bremen, Germany). MALDI-MSI raw data was imported into the SCiLS Lab software and converted to the SCiLS Lab file format and pre-processed by convolution

baseline removal (width: 20) and total ion count (TIC) normalization. Segmentation pipelines as published previously were performed for regions of interest (ROIs) identification, specific tissue distinction, and peak finding. For external peptide library adjustment, m/z values from regions of interest (ROIs) were exported from SCiLS Lab SW as csv files.

RESULTS AND DISCUSSION

Morphology of the Egyptian cobra (*N. haje*) venom gland. To visualize the cobra venom production and storage system and to obtain a better understanding of the physiological structures and relative position of an elapid venom gland, we performed three-dimensional computed tomography (3D-CT) and histochemical staining (**Figure 1**).

Morphological and phylogenetic biology studies on the serous-secreting venom glands in different caenophidian snake families suggest a single origin at the base of the colubrid radiation between 60-80 Mya.¹² Since this early origin, associated morphological characteristics of the venom system have evolved independently on multiple occasions in different caenophidian taxons, resulting in particular structural and topological characteristics, including fangs and glandular system.^{34–36} Advanced snakes have centralized venom systems to store their ready-to-use venom cocktail produced within. The simple glandular system might permit quantitative regulation of secreted venom depending on the ecological stimuli a snake is presented with. The 3D-CT reconstruction gave insights into the anatomy reflected by different layers of depth and revealed extensions of the venom glands and their general architecture (**Figure 1A-F**).

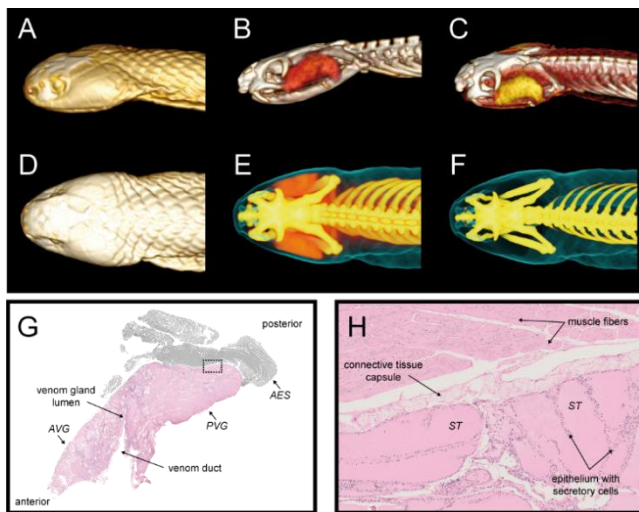


Figure 1. Morphology, topography and histochemistry of the venom delivery system of *Naja haje*.

(**A-F**) Three-dimensional computer tomography (3D-CT) reconstruction showing the relative position (lateral and top view) of the cobra head (**A, D**), morphological location of the venom production system (**B, E**), and skeleton with and without venom gland system (**C, F**). (**G, H**) Longitudinal section and histochemical staining by eosin and hematoxylin (H/E) of the *N. haje* venom gland. (**G**) The histochemical staining of a sagittal section in posteroanterior direction within the venom gland system. The venom duct, associated to a narrow lumen, divides the inner gland structure in an anterior (AVG) and a posterior venom gland (PVG) regime. (**H**) The venom duct region is formed by a high proportion of connective tissue and some strong elongated follicle, so-called secretory tubules (ST), AVG and RVG are composed of large and oval STs and thin connective tissue (zoom-in for rectangle in G).

In the first level, a typical cobra head is shown covered with lateral hoods and small, round pupils in lateral or top view, respectively (**Figure 1A** and **D**). Furthermore, the detailed topographical location of the venom gland system and skeleton with and without surrounded muscular tissue in lateral (**Figure 1B** and **C**) or top view (**Figure 1E** and **F**) is shown, respectively. The slender anterior ends of the elapid venom glands are positioned postorbital associated on each side of the head with specialized venom-conducting fangs located on the maxilla, as in all other advanced snakes (**Figure 1B** and **E**).⁹ The basic structure shows an oval shape and is more compact in contrast to large and triangular-shaped venom glands of viperids (**Figure 1C** and **E**).³⁷ The main function of the venom glands is the production and storage of a specialized toxic secretion prior to its delivery via the fangs into prey or aggressors.³⁷ Its postorbital position and the roughly oval shape is consistent with other previously described studies for closely related species of *Elapidae* and is a typical representative of a centralized and simple venom system.³⁶ However, it is worth mentioning that venom glands differ considerably in form and internal fine structure amongst taxa, sometimes even between species within the same genus.³⁵

The histochemical section in sagittal orientation reveals the closely associated *adductor externus superficialis* (AES), dorsally and caudally attached to elapid venom glands, and responsible for venom release after compression (**Figure 1G**).^{38,35,36} Furthermore, this compressor muscle is attached to a thick connective tissue capsule (*capsula fibrosa*) that encloses the venom gland and clearly delimits it from the AES (**Figure 1H**).^{39,36} The roughly centered venom duct, which is directly associated to a narrow lumen above, runs in a posteroanterior direction and divides the inner gland structure in an anterior (AVG) and a posterior venom gland (PVG) regime (**Figure 1G**). While the venom duct region is formed by a high proportion of connective tissue and some strong elongated follicle, so called secretory tubules (ST), AVG and PVG are mainly composed of large and oval STs (~260 $\mu\text{m} \pm 70 \mu\text{m}$ in diameter) surrounded by thin connective tissue (**Figure 1H**).³⁶ Due to the relatively small size, the venom gland lumen is unsuitable for long-term storage of venom secretion. In consequence, the venom is stored within the STs rather than the lumen. The outer surface of the STs are formed by a single layer of epithelium with secretory cells, which enclose the colloid-like lumen and shapes the venom gland fine structure analogous to other oral secretory systems, also known for secretion and long term storage (**Figure 1H**).^{40,38} This underlying morphological data of longitudinal venom gland sections can help to put the localization and visualization of numerous venom peptides into a morphological context.

Proteotranscriptomics of Egyptian cobra venom. To examine the secretory output of the elapid venom gland, we utilized a set of transcriptomics and complementary bottom-up and top-down proteomics data, thereby providing a holistic overview of venom composition (**Table S1-5**).

The *N. haje* venom gland transcriptome resulted in 4,418 assembled contigs, of which 58 exhibited gene annotations relating to 18 venom toxin families (**Figure 2A** and **Table S1**).²⁵ Of these contigs, six were full-length toxin sequences that encoded toxin isoforms relating to three-finger toxins (3FTx) family. This toxin family exhibited the highest expression levels by far of all toxin families identified, in combination accounting for >86% of the total toxin expression. In particular, long neurotoxins (INTx) exhibited the highest expression level (41%) within the 3FTx family, followed by short neurotoxins (sNTx, 39%), weak neurotoxins (wNTx, 5%) and cytotoxins (CTx, 2%). The Kunitz-type inhibitor (KUN), snake venom metalloproteinases (svMP) and serine proteases (svSP) were the second most abundant toxin families representing each ~3% of the toxin gene expression. Lower expression levels were identified for cysteine-rich secretory protein (CRISP), phospholipase A₂ inhibitor (PLA₂-i), cobra venom factor (CVF), venom nerve growth factor (NGF), L-amino acid oxidase (LAO), 5'-nucleotidase (5NUC), and phosphodiesterase (PDE), which account each $\geq 1\%$ of total toxin expression (**Figure 2A**; secondary). A number of other

potential venom toxin families (**Figure 2A**; minor) were identified with very low expression levels (in total <1%) (**Table S1**).

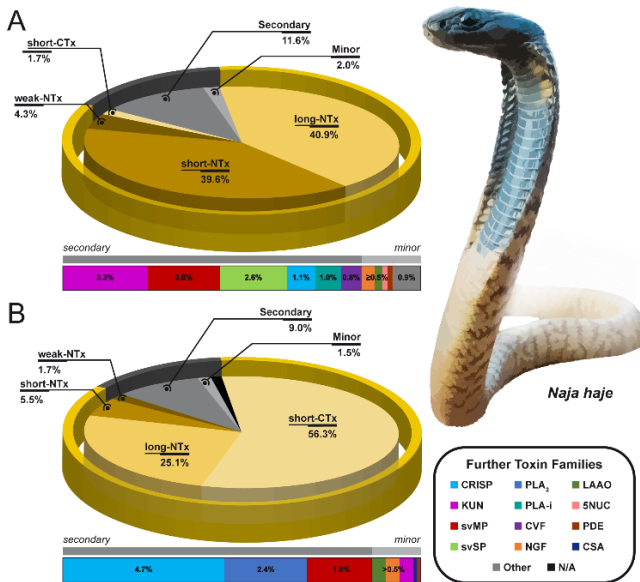


Figure 2. Proteotranscriptomics analysis of the Egyptian cobra (*Naja haje*). The relative gene expression levels (**A**) and protein abundance (**B**) of toxin families in percentage values identified in the *N. haje* venom gland transcriptome and venom proteome, respectively (**Table S1** and **S2**). (**A**) The relative gene expression levels (transcriptome) of the major three-finger toxin family (3FTx) and subgroups, secondary (>5%) as well as minor (>0.5%) toxin families represented by a pie chart. The corresponding bar chart reflects the relative expression levels of the secondary and minor categorized toxin families, which in combination account for 13.6% of residual toxins encoded in the venom gland. (**B**) The relative venom protein abundance (proteome) of the major 3FTx toxin family and subgroups, secondary (>5%) as well as minor (>0.5%) toxin families represented by a pie chart. The corresponding bar chart reflects the relative protein abundance of the secondary and minor categorized toxin families, which in combination account for 10.5% of residual toxins present in the venom proteome.

Next, we applied shotgun proteomics to broadly characterize the composition of secreted venom, including the qualitative assessments of main toxin families and the generation of a peptide reference database for later annotation of MALDI-MSI data (**Table S2**). Therefore, we pursued a dual strategy by applying in-solution digestion of crude venom combined with an on-top tissue digestion of an adjacent section. Overall, the combined shotgun approach resulted in 648 peptide spectrum matches (PSM) at a false discovery rate (FDR) of 0.1% to our species-specific assembled transcriptome database, extended by various *Naja* species²⁵ and the NCBI protein database (taxid: 8602). The toxin-encoding contigs covered 16 different toxin classes. We identified a wide range of typical elapid toxin families, such as 3FTx, CRISP, PLA₂, svSP, svMP, LAAO, NGF, KUN, PDE, 5NUC, CVF, and other lowly abundant venom toxins (**Table S2**).

In addition, we performed a semi-quantitative bottom-up analysis by reversed phase-HPLC (**Figure S2A**) and subsequent SDS-PAGE (**Figure S2B**) separation, as already described in detail elsewhere.^{27–29} The decomplexation of the *N. haje* venom resulted in 44 characteristic fractions, which separated into 151 protein bands covering a mass range of 5-100 kDa. Based on the comprehensive transcriptome database, we annotated respective MS/MS data to assign venom toxin families (**Figure 2B**). The 595 annotated PSMs resulted in the identification of various proteins from 12 different toxin families. The most abundant

toxin family is represented by 3FTx (88%), with its subgroups: sNTx (6%), INTx (25%), wNTx (2%) and CTx (56%). Secondary toxin families including CRISP (5%), PLA₂ (3%), and svMP (2%) are followed by low abundant toxin families, such LAAO (0.5%), KUN (0.4%), NGF (0.4%), cobra serum albumin (CSA; 0.1%), and PDE (0.1%). Further, PLA₂-i, 5NUC and CVF were only present in trace amounts, alongside a minor amount of unannotated proteins (N/A; 1.0%) (**Table S3** and **S4**). Employing both proteomics and transcriptomics, rather than only one of these techniques, improved the generation of a comprehensive and accurate venom database. When comparing the abundance of venom toxins (**Figure 2B**) with transcriptomic predictions of expression levels (**Figure 2A**), we observed an overall positive correlation, but noted some considerable differences, particular for the subgroups of the 3FTx family. The observed discrepancies in proteomic abundance and transcriptomic expression levels (e.g. sNTx and CTx; **Figure 2**) might be influenced by a number of biological and experimental factors, as previously outlined.^{41–44} Perhaps the most pressing reason is that we compared toxin transcription levels from a Ugandan individual with toxin abundance calculations from a Moroccan specimen.^{41,45} Thus, while it is possible that these differences are predominantly due to the abovementioned regulatory processes, it seems likely that intra-species variations influence our proteotranscriptomics results (**Figure 2**).⁴⁶

Top-down mass profiling by high-resolution electrospray ionization (HR-ESI) MS of the native Egyptian cobra venom, located to Morocco, generated an overview of 64 unique intact toxin isoforms, including low abundant compounds and small peptides (**Figure S3** and **Table S3**). The intact mass profile (IMP) is dominated by 43 molecular masses in the range of 6,377–7,897 Da, which reflects a considerable variety of 3FTx, as shown in detail by the bottom-up approach. Within the 3FTx family, we identified the major subclasses, such as CTx with ‘Cytotoxin 2’ (6,845.25 Da; CTx-2), followed by ‘Cytotoxin 5’ (6,761.21 Da; CTx-5), ‘Cytotoxin 6’ (6,813.30 Da; CTx-6) and ‘Cytotoxin 11’ (6,829.55 Da; CTx-11) and the neurotoxins (NTx) that are dominated by ‘long Neurotoxin 1’ (7,896.50 Da; INTx-1). Furthermore, KUN were detected in three fractions, showing only a marginal part with two different proteoforms (6,377 Da and 6,392 Da). CRISP could be associated with nine different proteoform masses between 23–25 kDa. The largest observed intact mass of 49 kDa was identified as a single svMP proteoform, the main representative of proteases in this venom.

Colocalization of molecular features to morphological structures. In order to prove complex morphological structures by discriminative molecular pattern and identify subregions across a venom system section of the Egyptian cobra (*N. haje*), we performed unsupervised multivariate analyses of acquired *m/z* values (peak features) without any need to extract or label samples. (**Figure 3**).

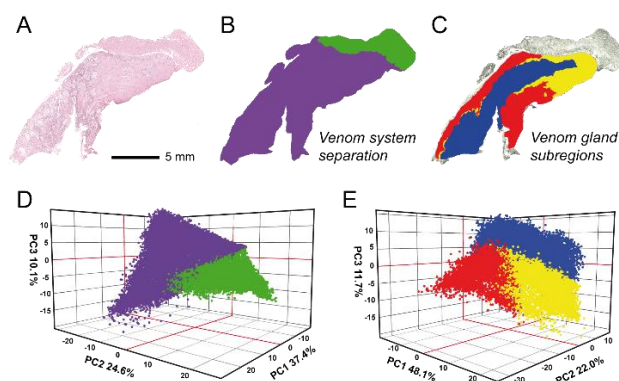


Figure 3. Multivariate statistical analyses of the venom gland system. Spatial segmentation generated by bisecting k-means clustering represents different hierarchical clusters (indicated by different

colors) based on discriminative m/z values (peak features). **(A)** The histochemical staining with hematoxylin and eosin (H&E) for tissue section orientation in segmentation maps of MALDI-MSI analysis. **(B)** The clustering separates the adductor muscle (green) from the venom gland (purple) based on the molecular pattern. **(C)** The inner gland structure is subdivided into a posterior (yellow), central (blue), and peripheral (red) sub-region (**SI Table S4**). **(A)** Spatial intensity map for a multivariate PCA within the full venom gland system. Highest variance (76%, PC1-3) is achieved for the first three principal components (PCs). **(D, E)** Principal component analysis (PCA) of respective tissue spots for the complete venom gland system and the various sub-regions (**Figure S4-5**).

Peak picking was applied to filter for relevant statistical signatures from the analyzed venom gland tissue sample, and resulted in 329 aligned peak features (m/z value range: 600–3200). An overlay of representative average spectra for respective muscle or venom gland regions highlighted highly associated peak features (**Figure S4**). To allow a detailed correlation of mass spectral imaging data and morphological structures within the full venom gland system, we performed bisecting k-means clustering, an unsupervised multivariate segmentation analysis, to detect characteristic peak features clustered by correlation distance (**Figure 3**). The segmentation analysis revealed two distinct clusters of selected peak features and directly separated the closely associated AES muscle from the main elapid venom gland by specific molecular pattern (**Figure 3B**). Considering the specific nature and morphological structures of reptile muscles, which are complex organs made up of multinucleated long and cylindrical cells, called myofibrils, composed of typical myofilament proteins, like actin and myosin or other high-abundant muscle proteins, this clearly different segmentation is not unanticipated.⁴⁷ Thus, we next performed a multivariate segmentation analysis solely on the venom gland and achieved a classification into three sub-regions along the venom gland based on the segmentation clusters (**Figure 3C**). Interestingly, unsupervised segmentation of spectral peak features showed a contrary segmentation to the aforementioned histological morphology (**Figure 3A**). While the histological morphology arranges the inner gland structure in antero-posterior regimes, peak features clearly show centrifugal segmentation (**Figure 3**).

To confirm multivariate clustering of spatial peak features within the venom gland system, unsupervised principle component analysis (PCA), a dimensionality-reduction method, was applied to visualize the most discriminative principal components (PCs) for respective morphological regions (**Figure 3D-E** and **Figure S5**). While principal component 2 (PC-2, 25%) shows high discriminative power between the venom gland and the AES muscle due to specific peak features presumably associated to toxin peptides, PC-1 (40%) and PC-3 (11%) supported sub-areas within the venom gland beside the muscle region (**Figure S5**). Overall, peak features of muscle and gland tissue spots allow for a high variance (76%, PC1-3) in respective regions (**Figure 3D** and **Figure S5**). In addition, targeted PCA analysis (82%, PC1-3) specific to the venom gland supported a classification into three sub-regions along the venom gland (**Figure 3E** and **Figure S5**). Since the high variance can be associated to characteristic molecular signatures, in context to the morphological location, these findings demonstrate that both unsupervised and supervised multivariate methods support discriminative peptide toxin signatures within the venom gland system (**Figure 3** and **Figure S4-5**).

Spatial venomics by MALDI-MSI. Despite the clustering of spectral and spatial peak features within the gland, we were particularly interested in identifying and spatially localizing various distinct toxin classes, named ‘spatial venomics’. To preserve spatial toxin resolution, tissue samples were fixed in formalin and embedded in paraffin in a biocompatible way to facilitate morphology preservation and MSI sample preparation. Alternative workflows include the application of other fixation solutions, i.e. ethanol-based fixative KinFix, as already described for venom gland fixation.^{21,22} A clear advantage of ethanol-

based fixation protocols is the possibility to measure intact toxins, which in turn is a disadvantage in terms of laborious downstream identification and inaccessibility of high molecular weight toxins. Further, immediate tissue fixation by shock freezing, which can be performed without sample embedding, presents an alternative option, though sample preparation and handling of frozen sections is difficult and leads to a decrease in peptide ion signals due to ionization suppression. As a first step, we excluded all aligned peaks present in both, the muscle and venom gland region, to eliminate false-positives. The remaining 263 peaks (m/z value range: 600–3200) were aligned against our extensive peptide library thus permitting the identification of 150 peptide markers from 14 toxin families (**Table S5**). To cover various toxin classes within the entire venom gland system, we spatially localized them with at least two, and up to five, specific peptide markers. To verify the most discriminative toxin peptide marker, we performed statistical analysis by means of receiver operating characteristic (ROC) calculations for the different sub-regions, respectively. The area under the curve (AUC) can assume values between 0 and 1 and expresses the discrimination power of each toxin peptide (**Table S5**). The univariate procedure enabled identification of a large number of toxin families, such as 3FTx, 5NUC, CRISP, CVF, LAAO, NGF, PDE, PLA₂, and svMP by a set of highly discriminative peptide masses (AUC >0.7 or <0.3), and thus enabled the various toxin classes to be spatially visualized within the different venom gland sub-regions (**Figure 4** and **Table S6**). In the central sub-region, which is oriented as a posteroanterior canal, the most discriminative peptide marker resulted from the toxin families PLA₂, svMP, and CTx with high intensities in the anterior region (**Table S6**). The peripheral sub-region is partially associated with PDE toxins, although a spatial distribution map of specific PDE peptide markers shows a denser population in the posterior region of the venom gland. Interestingly, major toxin families are associated with high intensities in the posterior region of the venom gland, whereas none or low intensities of these toxin families were observed in the anterior parts of the venom gland (**Figure 4**). We detected a heterogeneous distribution of different toxin classes mainly along the posteroanterior canal with highly variable intensities within distinct regions. Some toxins were predominantly found in the posterior region (e.g., PDE and CRISP in **Figure 4**), others were found throughout the section but at higher levels in certain regions (e.g., PLA₂, CTx, svMP in **Figure 4**), and some toxins exhibit a consistent distribution across the venom gland (e.g., 5NUC in **Figure 4**). The dominant abundance of 3FTx and the associated high number of peptide markers allowed us to spatially interrogate toxins at the level of various NTx sub-classes or CTx isoforms (**Figure 5**). While the NTx toxin family shows a rather heterogeneous distribution, CTx can be uniformly detected throughout the venom gland via a combination of different specific peptide markers (**Figure 5A**). Visualization of localization at the proteoform level could only be accomplished by individual, specific peptide markers, which allowed accurate mapping and identified a heterogeneous distribution to distinct regions (**Figure 5B**). These analyses provide detailed insights into the fine distribution of different sub-classes of the 3FTx superfamily and show clearly delimited areas, e.g. sharp separation of INTx and wNTx in the posterior region of the venom gland or the strictly heterogeneous distribution of different CTx isoforms (**Figure 5B**). The identification and spatial localization of various sub-classes and isoforms was achieved due to the very high abundance of this toxin family, though the combination of low abundance and reduced ionization efficiency preclude such a granular analysis for other toxin types.

However, lower mass accuracy and limited on-tissue MS/MS identification, due to a large number of isobaric ions⁴⁸ and the presence of chimera spectra⁴⁹, makes it more susceptible to false-positive protein assignments and direct identification (acquired by MALDI-MSI) remains limited to only the most abundant toxins. Nevertheless, MALDI-MSI is preferable in contrast to other MSI instrumentation, such as desorption electrospray ionization (DESI), secondary ion mass spectrometry (SIMS) or liquid extraction

surface analysis (LESA). It requires less time to perform measurements, and provides better spatial resolution for a larger mass range.⁵⁰

In summary, these results improve our knowledge of the basic biology of the snake venom gland as we demonstrate the identification and spatial localization of numerous toxin families in parallel. Our findings reveal that distinct toxin families, and functionally distinct sub-classes, exhibit spatial heterogeneity across the venom production system. To prevent proteolytic processes, spatial differentiation is a simple and efficient tool. On the other hand, homogeneous distribution within the glandular tissue can help to increase antagonistic effects, such as inactivation of svMP by tripeptide inhibitors. We expect that these results will contribute to address the fundamental question as to how snake venom constituents are produced and stored within the venom glandular system.

CONCLUSIONS

Herein MALDI-MSI was combined with a global venom analysis via proteotranscriptomics analyses to identify and spatially map various toxin families associated to the venom system of the Egyptian cobra (*Naja haje*). This particular approach allowed the identification of small and high-molecular mass toxin classes in near-cellular spatial resolution. Our findings highlight the molecular biology of centralized venom systems and the need for more holistic investigations of venom storage and delivery mechanisms. According to their morphological limitations and ecological context, the indirect venom modulation is of great importance to be considered in antivenom efficacy. Future studies on the multifunctionality of snake venoms can help to improve envenomation treatment. Thus, the simple venom-delivery system of an elapid with a complex venom arsenal might be a useful model system for gaining insight into the limitations placed on venom evolution by morphological constraints. Our results provide evidence for the venom optimization concept that toxins are non-uniformly abundant throughout simple venom systems of venomous animals, and that the link between ecology and toxin evolution may be more complex than previously assumed.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Supporting Information: SI Figures and Material/Methods (PDF)

Supporting Information: SI Tables (xlsx)

AUTHOR INFORMATION

Corresponding Author

* Benjamin-Florian Hempel - BIH Center for Regenerative Therapies BCRT, Charité - Universitätsmedizin Berlin, 13353 Berlin, Germany; 0000-0002-1998-4033; Email: benjamin.hempel@fu-berlin.de.

* Roderich D. Süßmuth - Institut für Chemie, Technische Universität Berlin, 10623 Berlin, Germany; 0000-0001-7027-2069; Email: roderich.suessmuth@tu-berlin.de.

Present Addresses

† Benjamin-Florian Hempel – Tiermedizinisches Zentrum für Resistenzforschung (TZR), Freie Universität Berlin, 14163 Berlin, Germany.

Author Contributions

B.-F.H., and M.D. planned the study. B.-F.H. and C.A.S. prepared tissue for morphology and imaging studies. T.D.K. and N.R.C. constructed and analyzed the transcriptome. B.-F.H., M.D., and D.P. performed the protein separation and acquired the mass spectrometry data. B.-F.H. and G.N. performed MALDI Imaging preparation and acquisition. B.-F.H., O.K., and M.D. performed the data analysis. B.-F.H. and M.D. wrote the manuscript with assistance of R.D.S and O.K. All authors read, discussed, and approved the manuscript. ‡These authors contributed equally.

ACKNOWLEDGMENT

In memory and thanks to Dr. Frank Mutschmann for the assistance of the gland removal. We would like to thank Marion Biering for the assistance of the venom gland preparation and Mrinalini for the assistance of the histological annotation. Furthermore, we thank Laura Ruysseveldt from the Herpetological Education & Research Project and Klaus Rudloff for providing images of the living specimen as well as Tierpark Berlin-Friedrichsfelde GmbH for providing the animal. D.P. was supported by the German Research Foundation (DFG) through the CMFI Cluster of Excellence (EXC 2124).

REFERENCES

- (1) Casewell, N. R.; Wüster, W.; Vonk, F. J.; Harrison, R. A.; Fry, B. G. *Trends Ecol. Evol.* **2013**, *4*, 219–229.
- (2) Fry, B. G.; Roelants, K.; Champagne, D. E.; Scheib, H.; Tyndall, J. D. A.; King, G. F.; Nevalainen, T. J.; Norman, J. A.; Lewis, R. J.; Norton, R. S.; Renjifo, C.; La Vega, R. C. R. de. *Annu. Rev. Genomics Hum. Genet.* **2009**, , 483–511.
- (3) Gutiérrez, J. M.; Calvete, J. J.; Habib, A. G.; Harrison, R. A.; Williams, D. J.; Warrell, D. A. *Nat. Rev. Dis. Primers* **2017**, , 17063.
- (4) King, G. F. *Expert Opin. Biol. Ther.* **2011**, *11*, 1469–1484.
- (5) Martinson, E. O.; Mrinalini; Kelkar, Y. D.; Chang, C.-H.; Werren, J. H. *Curr. Biol.* **2017**, *13*, 2007-2013.e8.
- (6) Wong, E. S. W.; Belov, K. *Gene* **2012**, *1*, 1–7.
- (7) Nakashima, K.; Nobuhisa, I.; Deshimaru, M.; Nakai, M.; Ogawa, T.; Shimohigashi, Y.; Fukumaki, Y.; Hattori, M.; Sakaki, Y.; Hattori, S. *Proc. Natl. Acad. Sci. USA* **1995**, *12*, 5605–5609.
- (8) Brust, A.; Sunagar, K.; Undheim, E. A. B.; Vetter, I.; Yang, D. C.; Yang, D. C.; Casewell, N. R.; Jackson, T. N. W.; Koludarov, I.; Alewood, P. F.; Hodgson, W. C.; Lewis, R. J.; King, G. F.; Antunes, A.; Hendrikx, I.; Fry, B. G. *Mol. Cell. Proteomics* **2013**, *3*, 651–663.
- (9) Vonk, F. J.; Admiraal, J. F.; Jackson, K.; Reshef, R.; Bakker, M. A. G. de; Vanderschoot, K.; van den Berge, I.; van Atten, M.; Burgerhout, E.; Beck, A.; Mirtschin, P. J.; Kochva, E.; Witte, F.; Fry, B. G.; Woods, A. E.; Richardson, M. K. *Nature* **2008**, *7204*, 630–633.
- (10) Casewell, N. R.; Jackson, T. N. W.; Laustsen, A. H.; Sunagar, K. *Trends Pharmacol. Sci.* **2020**, *8*, 570–581.
- (11) Chippaux, J. P.; Williams, V.; White, J. *Toxicon* **1991**, *11*, 1279–1303.
- (12) Fry, B. G. *Genome Res.* **2005**, *3*, 403–420.
- (13) Fry, B. G.; Scheib, H.; van der Weerd, L.; Young, B.; McNaughtan, J.; Ramjan, S. F. R.; Vidal, N.; Poelmann, R. E.; Norman, J. A. *Mol. Cell. Proteomics* **2008**, *2*, 215–246.
- (14) Cornett, D. S.; Reyzer, M. L.; Chaurand, P.; Caprioli, R. M. *Nat. Methods* **2007**, *10*, 828–833.
- (15) Angel, P. M.; Caprioli, R. M. *Biochemistry* **2013**, *22*, 3818–3828.
- (16) Caprioli, R. M.; Farmer, T. B.; Gile, J. *Anal. Chem.* **1997**, *23*, 4751–4760.
- (17) Norris, J. L.; Caprioli, R. M. *Chemical reviews* **2013**, *4*, 2309–2342.

- (18) Walker, A. A.; Robinson, S. D.; Hamilton, B. F.; Undheim, E. A. B.; King, G. F. *Proteomics* **2020**, *17-18*, e1900324.
- (19) Ghezellou, P.; Heiles, S.; Kadesch, P.; Ghassempour, A.; Spengler, B. *J. Am. Soc. Mass Spectrom.* **2021**, *4*, 1105–1115.
- (20) Hamilton, B. R.; Marshall, D. L.; Casewell, N. R.; Harrison, R. A.; Blanksby, S. J.; Undheim, E. A. B. *Angew. Chem. Int. Ed.* **2020**, *10*, 3855–3858.
- (21) Undheim, E. A. B.; Hamilton, B. R.; Kurniawan, N. D.; Bowlay, G.; Cribb, B. W.; Merritt, D. J.; Fry, B. G.; King, G. F.; Venter, D. J. *Proc. Natl. Acad. Sci. USA* **2015**, *13*, 4026–4031.
- (22) Undheim, E. A. B.; Sunagar, K.; Hamilton, B. R.; Jones, A.; Venter, D. J.; Fry, B. G.; King, G. F. *J. Proteomics* **2014**, *1*, 1–10.
- (23) Fischer, A. H.; Jacobson, K. A.; Rose, J.; Zeller, R. *CSH Protoc.* **2008**, , pdb.prot4986.
- (24) Albulescu, L.-O.; Kazandjian, T.; Slagboom, J.; Bruyneel, B.; Ainsworth, S.; Alsolaiss, J.; Wagstaff, S. C.; Whiteley, G.; Harrison, R. A.; Ulens, C.; Kool, J.; Casewell, N. R. *Front. Pharmacol.* **2019**, , 848.
- (25) Kazandjian, T. D.; Petras, D.; Robinson, S. D.; van Thiel, J.; Greene, H. W.; Arbuckle, K.; Barlow, A.; Carter, D. A.; Wouters, R. M.; Whiteley, G.; Wagstaff, S. C.; Arias, A. S.; Albulescu, L.-O.; Plettenberg Laing, A.; Hall, C.; Heap, A.; Penrhyn-Lowe, S.; McCabe, C. V.; Ainsworth, S.; Da Silva, R. R.; Dorrestein, P. C.; Richardson, M. K.; Gutiérrez, J. M.; Calvete, J. J.; Harrison, R. A.; Vetter, I.; Undheim, E. A. B.; Wüster, W.; Casewell, N. R. *Science* **2021**, *6527*, 386–390.
- (26) Conesa, A.; Götz, S.; García-Gómez, J. M.; Terol, J.; Talón, M.; Robles, M. *Bioinformatics* **2005**, *18*, 3674–3676.
- (27) Hempel, B.-F.; Damm, M.; Göçmen, B.; Karis, M.; Oguz, M. A.; Nalbantsoy, A.; Süßmuth, R. D. *Toxins* **2018**, *1*.
- (28) Hempel, B.-F.; Damm, M.; Mrinalini; Göçmen, B.; Karış, M.; Nalbantsoy, A.; Kini, R. M.; Süßmuth, R. D. *J. Proteome Res.* **2020**, *4*, 1731–1749.
- (29) Petras, D.; Hempel, B.-F.; Göçmen, B.; Karis, M.; Whiteley, G.; Wagstaff, S. C.; Heiss, P.; Casewell, N. R.; Nalbantsoy, A.; Süßmuth, R. D. *J. Proteomics* **2019**, , 31–50.
- (30) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. *Nat. Biotechnol.* **2012**, *10*, 918–920.
- (31) Zhang, J.; Xin, L.; Shan, B.; Chen, W.; Xie, M.; Yuen, D.; Zhang, W.; Zhang, Z.; Lajoie, G. A.; Ma, B. *Mol. Cell. Proteomics* **2012**, *4*, M111.010587.
- (32) Cillero-Pastor, B.; Heeren, R. M. A. *J. Proteome Res.* **2014**, *2*, 325–335.
- (33) Kou, Q.; Xun, L.; Liu, X. *Bioinformatics* **2016**, *22*, 3495–3497.
- (34) Jackson, K. *Zool. J. Linn. Soc.* **2003**, *3*, 337–354.
- (35) Jackson, T. N. W.; Young, B.; Underwood, G.; McCarthy, C. J.; Kochva, E.; Vidal, N.; van der Weerd, L.; Nabuurs, R.; Dobson, J.; Whitehead, D.; Vonk, F. J.; Hendrikx, I.; Hay, C.; Fry, B. G. *Zoology* **2017**, *1*, 107–130.
- (36) Rosenberg, H. I. *J. Morphol.* **1967**, *2*, 133–155.
- (37) Kerckamp, H. M. I.; Casewell, N. R.; Vonk, F. J., Evolution of the Snake Venom Delivery System. In *Evolution of Venomous Animals and Their Toxins*; Gopalakrishnakone, P.; Malhotra, Anita, Eds.; Springer Netherlands: Dordrecht, 2015.

- (38) Fry, B. G.; Vidal, N.; Norman, J. A.; Vonk, F. J.; Scheib, H.; Ramjan, S. F. R.; Kuruppu, S.; Fung, K.; Hedges, S. B.; Richardson, M. K.; Hodgson, W. C.; Ignjatovic, V.; Summerhayes, R.; Kochva, E. *Nature* **2006**, *7076*, 584–588.
- (39) Emery, C. *Arch. Für Mikrosk. Anat.* **1875**, *51*, 561–568.
- (40) Barua, A.; Mikheyev, A. S. *Proc. Natl. Acad. Sci. USA* **2021**, *14*.
- (41) Casewell, N. R.; Wagstaff, S. C.; Wüster, W.; Cook, D. A. N.; Bolton, F. M. S.; King, S. I.; Pla, D.; Sanz, L.; Calvete, J. J.; Harrison, R. A. *Proc. Natl. Acad. Sci. USA* **2014**, *25*, 9205–9210.
- (42) Durban, J.; Pérez, A.; Sanz, L.; Gómez, A.; Bonilla, F.; Rodríguez, S.; Chacón, D.; Sasa, M.; Angulo, Y.; Gutiérrez, J. M.; Calvete, J. J. *BMC genomics* **2013**, , 234.
- (43) Ruggles, K. V.; Krug, K.; Wang, X.; Clauser, K. R.; Wang, J.; Payne, S. H.; Fenyö, D.; Zhang, B.; Mani, D. R. *Mol. Cell. Proteomics* **2017**, *6*, 959–981.
- (44) Vogel, C.; Marcotte, E. M. *Nat. Rev. Genet.* **2012**, *4*, 227–232.
- (45) Daltry, J. C.; Wüster, W.; Thorpe, R. S. *Nature* **1996**, *6565*, 537–540.
- (46) Damm, M.; Hempel, B.-F.; Süßmuth, R. D. *Toxins* **2021**, *6*.
- (47) Nagumantri, S. P.; Banu, S.; Idris, M. M. *Sci. Rep.* **2021**, *1*, 3675.
- (48) Schober, Y.; Schramm, T.; Spengler, B.; Römpf, A. *Rapid communications in mass spectrometry : RCM* **2011**, *17*, 2475–2483.
- (49) Houel, S.; Abernathy, R.; Renganathan, K.; Meyer-Arendt, K.; Ahn, N. G.; Old, W. M. *J. Proteome Res.* **2010**, *8*, 4152–4160.
- (50) Hermann, J.; Noels, H.; Theelen, W.; Lellig, M.; Orth-Alampour, S.; Boor, P.; Jankowski, V.; Jankowski, J. *Analytical and bioanalytical chemistry* **2020**, *6*, 1263–1275.