State-of-the-Art Mass Spectrometry for the Discovery and Identification of Biodiverse Natural Toxins

vorgelegt von

M. Sc.

Benjamin-Florian Hempel

ORCID: 0000-0002-1998-4033

an der Fakultät II – Mathematik und Naturwissenschaften

der Technischen Universität Berlin

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften

Dr. rer. nat.

genehmigte Dissertation

Berlin 2021

dedicated to my family

Abstract

Nowadays natural toxins, fascinating and sophisticated biochemical products in all forms of life, play an exceptional role in human health and peptide drug discovery. Natural toxins and their biochemical features reflect two sides of a coin, either associated to historical terrible disasters responsible for innumerable death and permanent health issues, or chemical lead structures for the development of lifesaving drugs against some of major global burdens.

The introduction of mass spectrometry, a major key player of analytical spectrometric instruments in mid of the 20th century, resulted in many well-funded, large-scale screening programs and gained momentum for the analysis of medically important toxins of various origin. Down to the present day, several technical improvements of analytical sciences helped to get a deeper insight to complex biological systems and understanding of the mode of action for toxigenic agents. The investigations resulted in manifold clinical drugs based on natural-derived toxins.

The present thesis is dealing with the identification and biochemical biosynthesis investigation by different state-of-the-art mass spectrometry technologies of biodiverse natural toxins, both exotoxins and zootoxins. In addition, several toxic components from highly complex snake venoms were evaluated for potential application as cytoxic agents against cancerous cell lines.

The first part of the thesis is dedicated to biosynthesis investigations of the nonribosomal-polyketide hybrid albicidin. The phytotoxin is produced by the xylem-invading sugarcane pathogen, *Xanthomonas albilineans*, causing chlorosis by interrupt the plant DNA replication. Albicidin blocks the plant and prokaryotic specific topoisomerase IV by interference in the DNA binding subunit A. The biosynthesis of albicidin revealed the incorporation of different para-aminobenzoic acid (pABA) variants, para-coumaric acid (pCA), and β-cyanoalanine (β-Cya). In chapter 2 we found, besides the main albicidin metabolite, an N-terminal modified derivative by untargeted mass spectrometry metabolomics experiments from wild strain cultures of *X. albilineans*. The carbamoylated albicidin derivative was characterized by *in vivo* and *in vitro* biochemical investigations. The post-NRPS modification by the carbamoyltransferase Alb15 in the biosynthesis cluster of albicidin was confirmed by gene inactivation and *in vitro* experiments by the heterologously expressed enzyme. Conclusively, chemical synthesis of carbamoyl-albicidin enabled us to test bioactivities by means of *in vitro* gyrase and antibacterial inhibition assays. Chapter 3 gives a closer insight in the inter-NRPS modification of incorporated pABA derivatives. Interestingly, substrate specificity experiments for respective adenylation domains revealed a preferential incorporation of the substrate para-3-hydroxy-aminobenzoic acid (pABA-3OH)

Abstract

rather than previously mentioned para-2-hydroxy-3-methoxy-benzoic acid (pMBA). The biosynthesis gene *alb02* encodes a putative S-adenosylmethionine (SAM)-dependent *O*-methyltransferase, which was *in vitro* characterized by comprehensive substrate screening and substrate-dependent kinetics. Top-down mass spectrometry experiments on *crypto*-PCP domains loaded with precursor pABA-3OH were carried out to track the *in situ* modification to para-amino-3-methoxybenzoic acid (pABA-3OMe). In combination with NMR spectroscopy and crosslinking mass spectrometry, we examined a transient substrate-controlled interaction that so far has not been described and impact our understanding of the assembly line logic.

In the second part of the thesis, we combined standard and alternatively introduced mass spectrometry-guided workflows to depict venom variations on the level of population or within different species. The detailed venom analysis of postulated subspecies by different venom proteomics strategies allowed us to show a remarkable agreement of the venom compositions that can help to overcome the controversial question of the taxonomic status of *Vipera ammodytes transcaucasiana*. In addition, we tested venom constituents in chapter 4 for a cytotoxic bioactivity screening and found exceptional results for the human breast adenocarcinoma epithelial MDA-MB-231 cell line. In chapter 5, we introduced a combined approach that enabled us for a faster and more detailed comparison of venom proteomes from multiple specimens. We identified intraspecies venom variations of *Vipera kaznakovi* individuals, and find these were mainly driven by the age of the animals. In chapter 6, we report on the venom proteome of *Vipera anatolica senliki* and an alternative in-source decay (ISD)-based proteomics workflow. Top-down ISD venomics allowed for disulfide bond counting and effective *de novo* sequencing of high-molecular-weight venom proteoforms, both of which are difficult to achieve by the commonly established top-down approach.

Finally, in chapter 7 perspectives for future directions of state-of-the-art mass spectrometry techniques are briefly discussed and exemplary experiments outlined.

Zusammenfassung

Heutzutage spielen natürliche Toxine in einer Vielzahl von verschiedensten Lebensformen, als hoch spezialisierte und faszinierende biochemische Produkte, eine außergewöhnliche Rolle für die Entdeckung neuer peptidbasierter Arzneimittel und die menschliche Gesundheit. Natürliche Toxine und ihre biochemischen Eigenschaften spiegeln zwei Seiten einer Medaille wider, denn entweder werden sie mit historischen Katastrophen sowie unzähligen Todesfällen und dauerhaften Gesundheitsproblemen assoziiert, oder als chemische Leitstrukturen für die Entwicklung lebensrettender Medikamente gegen einige der größten globalen Herausforderungen.

Die Einführung der Massenspektrometrie, eine Schlüsseltechnologie in der analytisch-quantitativen Lebenswissenschaft, spielte Mitte des 20. Jahrhunderts eine wichtige Rolle und führte zu vielen gut finanzierten sowie groß angelegten Screening-Programmen welche für die Analyse von medizinisch relevanten Toxinen verschiedener Herkunft schnell an Dynamik gewannen. Bis heute haben verschiedene technische Verbesserungen der analytischen Wissenschaften dazu beigetragen einen tieferen Einblick in komplexe biologische Systeme zu bekommen und ein besseres Verständnis der Wirkungsweise von Toxigenen zu erhalten. Die Untersuchungen ergaben vielfältige klinische Medikamente, die auf natürlichen Toxinen basieren. Die vorliegende Arbeit befasst sich mit der Identifizierung und Untersuchung der biochemischen Biosynthese von biodiversen natürlichen Toxinen, sowohl Exotoxinen als auch Zootoxinen, mittels unterschiedlicher hochmoderner Massenspektrometrie-Technologien. Zusätzlich wurden verschiedene toxische Komponenten von hochkomplexen Schlangengiften auf den potenziellen Einsatz als zytotoxische Wirkstoffe gegen Krebszelllinien untersucht.

Der erste Teil dieser Arbeit untersucht die Biosynthese des nonribosomalen Sekundärmetaboliten Albicidin. Das Phytotoxin wird durch das Xylem-befallenden Zuckerrohr-Pathogen *Xanthomonas albilineans* produziert, welcher die Chlorose durch Störung der pflanzlichen DNA Replikation verursacht. Albicidin blockiert die für Pflanzen und Prokaryoten spezifische Topoisomerase IV durch Interaktion in der DNA-Bindungsuntereinheit A. Biochemische Untersuchungen zur Biosynthese von Albicidin zeigten den Einbau von verschiedenen para-Aminobenzoesäure (pABA) Varianten, para-Cumarsäure (pCA) und β-Cyanoalanin (β-Cya). In Kapitel 2 fanden wir neben dem Albicidin-Hauptmetaboliten ein N-terminal modifiziertes Derivat mittels ungezielter massenspektrometrischer und metabolomischer Experimente in *X. albilineans* Wildtyp-Kulturen. Im Folgenden wurde die post-NRPS Modifizierung durch die Carbamoyltransferase Alb15 mittels Geninaktivierung und *in vitro*

Experimente mit dem heterolog exprimierten Alb15 Enzym bestätigt. Abschließend ermöglichte uns die chemische Synthese von Carbamoyl-Albicidin die Testierung der Bioaktivitäten mittels *in vitro* Gyraseinhibierung und dessen antibakteriellen Eigenschaften. In Kapitel 3 konnten wir einen tieferen Einblick in die *in situ* Modifikation der eingebauten pABA Derivate gewinnen. Interessanterweise offenbarten Substratspezifitäts-Experimente für die entsprechenden Adenylierungsdomänen einen selektiven Einbau des Substratmonomers para-3-Hydroxy-Aminobenzoesäure (pABA-3OH) anstelle der im Albicidin identifizierten para-2-Hydroxy-3-Methoxybenzoesäure (pMBA). Das Biosynthese-Gen *alb02* kodiert für eine putative *S*-Adenosylmethionin (SAM)-abhängige *O*-Methyltransferase, welche durch ein umfassendes Substrat-Screening und Substrat-abhängige Kinetik *in vitro* charakterisiert wurde. Top-down Massenspektrometrie-Experimente an *crypto*-PCP Domänen beladen mit dem Precursor pABA-3OH wurden durchgeführt um die *in situ* Modifikation zu para-Amino-3-Methoxybenzoesäure (pABA-3OMe) zu beobachten. Mithilfe einer Kombination aus zweidimensionaler NMR Spektroskopie und ´cross-linking` Massenspektrometrie geben wir Einblicke in eine bisher unbeschriebene basale und substrat-gesteuerte *in situ* Interaktion, die unser Verständnis über die NRPS-gesteuerte Syntheselogik neu definiert.

Im zweiten Teil der Arbeit kombinieren wir standardisierte und alternative Massenspektrometrie-Verfahren um Venomvariationen auf der Ebene von Populationen oder innerhalb verschiedener Spezies abzubilden. Die detaillierte Venomanalyse von postulierten Subspezies, mittels unterschiedlicher Venomproteom-Strategien, erlaubte uns eine bemerkenswerte Übereinstimmung der Giftzusammensetzung darzustellen, die dazu beitragen kann die umstrittene Frage des taxonomischen Status von *Vipera ammodytes transcaucasiana* zu klären. Zusätzlich haben wir in Kapitel 4 verschiedene Venombestandteile auf zytotoxische Bioaktivität getestet, welche vielversprechende Resultate für humane Brustadenokarzinom-Epithelzelllinien zeigen. In Kapitel 5 etablieren wir einen kombinierten Ansatz zur schnelleren und genaueren Analyse von Venomproteomen mehrerer Individuen. Das Verfahren erlaubte die Identifikation von Venomvariationen in einer *Vipera kaznakovi* Population, die hauptsächlich auf dem Alter der Tiere basieren. In Kapitel 6 berichten wir von dem Venomproteom der anatolischen Wiesenotter, *Vipera anatolica senliki*, und die Einführung einer alternativen Identifzierungsmethode für Venomproteome basierend auf dem ´in-source` Zerfall (ISD). Top-down ISD Untersuchungen erlaubt die Identifizierung durch Reduktion und anschließender Zählung der Disulfid-Bindungen, sowie effektive *de novo* Sequenzierung von hochmolekularen Venom-Proteoformen, was durch den bereits etablierten Top-down Ansatz schwierig ist.

Abschließend diskutieren wir in Kapitel 7 kurz Perspektiven für zukünftige Richtungen moderner Massenspektrometrie-Techniken und zeigen beispielhafte Experimente für die Projekte.

Acknowledgment

First of all, deepest gratitude goes to my supervisor Prof. Roderich D. Süssmuth, who gave me his trust and welcomed me into his group. I thank him for his patience despite all my personal and professional problems in this time. He always takes care for the problems in the group and excites an open dialogue to foster the scientific and personal development of his co-workers despite his many other tasks and responsibilities. I am grateful for his steady motivation and ideas to my research, particularly giving me the freedom to realize various projects in the context of natural toxins. I also would like to thank him for the financial support that allowed me to attend different conferences in Europe and to present my results to the scientific community. The high scientific demand and your steady support made me to the scientist I am today; thank you for all.

I would like to thank Prof. Dr. Martina Marchetti-Deschmann (TU Wien) for accepting to examine this dissertation and Prof. Dr. Maria A. Mroginski for being the chairperson of the defense session.

Furthermore, I would like to express my gratitude to all the collaborators working with me on different fruitful projects of natural toxins: Ayse Nalbantsoy, Bayram Göçmen, Nicholas R. Casewell, Mrinalini, Oliver Klein, Chris Weise, and R. Manjunatha Kini.

All group members in the past and today has been a great support for me over the past few years, and there has always been an open ear for scientific as well as private questions. The naming of all members would go beyond the scope of this acknowledgment, so I would like to briefly mention the most important people to me:

Strong gratitude goes to Maik Damm, my former master student, who has become a fantastic colleague and great buddy over time. I will miss our many constructive and funny conversations in the office and laboratory.

I would also like to thank Dr. Daniel Petras, my former supervisor and good friend, introducing me into the field of natural toxins and in the fascinating world of science.

Furthermore, I would like to thank Dr. Andi Mainz and Dr. Agnes Mühlenweg for helpful advices in the meetings and taking care for the everyday problems in the laboratory. I thank Andi for all his help and time to finalize my project with the help of interaction studies using NMR spectroscopy.

My sincere appreciation goes to my companions, Tam Dang, Vincent Wiebach, and Lida Rosotock, who came with me to the group and enabled a great scientific atmosphere in the working group.

Publications

All parts of chapter 2-6 of this thesis have been published in the articles listed below or are in preparation. A list with detailed author contributions is stated on the following pages.

1. D. Petras, D. Kerwat, A. Pesic, **BF. Hempel**, L. von Eckardstein, S. Semsary, J. Arasté, M. Marguerettaz, M. Royer, S. Cociancich, RD. Süssmuth[+] "The O-Carbamoyl-Transferase Alb15 is Responsible for the Modification of Albicidin" *ACS Chem. Biol.* **2016**, 11(5), 1198-1204.
(*accepted manuscript*)
DOI: https://doi.org/10.1021/acschembio.5b01001

2. **BF. Hempel***, D. Petras*, A. Mainz, L. von Eckardstein, S. Cociancich, M. Royer, RD. Süssmuth[+] "The Biosynthesis of Albicidin Involves On-line Enzymatic Methylation During Nonribosomal Peptide Assembly" **2020**, in preparation.

3. **BF. Hempel**, M. Damm, B. Göçmen, M. Karis, MA. Oguz, A. Nalbantsoy[+], RD. Süssmuth[+] "Comparative Venomics of the *Vipera ammodytes transcaucasiana* and *Vipera ammodytes montandoni* from Turkey Provides Insights into Kinship" *Toxins* **2018**, 10(1), 23.
(*accepted manuscript*); CC BY (http://creativecommons.org/licenses/by/4.0)
DOI: https://doi.org/10.3390/toxins10010023

4. D. Petras*,[+], **BF. Hempel***, B. Göçmen, M. Karis, G. Whiteley, SC. Wagstaff, P. Heiss, NR. Casewell, A. Nalbantsoy[+], RD. Süssmuth[+] "Intact Protein Mass Spectrometry reveals Intraspecies Variations in Venom Composition of a Local Population of *Vipera kaznakovi* in Northeastern Turkey" *J Prot* **2019**, 199, 31-50.
(*accepted manuscript*); CC BY-NC-ND (http://creativecommons.org/licenses/by/4.0)
DOI: https://doi.org/10.1016/j.jprot.2019.02.004

5. **BF. Hempel***, M. Damm*, Mrinalini*, B. Göçmen, M. Karis, A. Nalbantsoy, RM. Kini and RD. Süssmuth[+] "Extended Snake Venomics by Top-Down In-Source Decay: Investigating the Newly Discovered Anatolian Meadow Viper Subspecies, *Vipera anatolica senliki*" *J Prot Res* **2020**, 19(4), 1731-1749.
(*accepted manuscript*)
DOI: https://doi.org/10.1021/acs.jproteome.9b00869

Publications

Publications prepared during my doctoral studies and not part of this thesis.

6. A. Nalbantsoy*, **BF. Hempel***, D. Petras, P. Heiss, B. Göçmen, N. Iğci, MZ. Yildiz, RD. Süssmuth[+] "Combined venom profiling and cytotoxicity screening of the Radde's mountain viper (*Montivipera raddei*) and Mount Bulgar Viper (*Montivipera bulgardaghica*) with potent cytotoxicity against human A549 lung carcinoma cells" *Toxicon* **2017**, 135, 71-83.
   DOI: https://doi.org/10.1016/j.toxicon.2017.06.008.

7. M. Damm*, **BF. Hempel***, A. Nalbantsoy, RD. Süssmuth[+] "Comprehensive Snake Venomics of the Okinawa Habu Pit Viper, *Protobothrops flavoviridis*, by Complementary Mass Spectrometry-Guided Approaches" *Molecules* **2018**, 23(8), 1893.
   DOI: https://doi.org/10.3390/molecules23081893.

8. CS. Ozverel*, M. Damm*, **BF. Hempel**[+], B. Göçmen[c], R. Sroka[b], RD. Süssmuth, A. Nalbantsoy[+] "Investigating the cytotoxic effects of the venom proteome of two species of the *Viperidae* family (*Cerastes cerastes* and *Cryptelytrops purpureomaculatus)* from various habitats" *Comp Biochem Physiol C* **2019**, 220, 20-30.
   DOI: https://doi.org/10.1016/j.cbpc.2019.02.013.

9. NR. Casewell[+], D. Petras, DC. Card, V. Suranse, AM. Mychajliw, D. Richards, I. Koludarov, LO. Albulescu, J. Slagboom, **BF. Hempel**, NM. Ngum, RJ. Kennerley, JL. Brocca, G. Whiteley, RA. Harrison, FMS. Bolton, J. Debono, FJ. Vonk, J. Alföldi, J. Johnson, E. Karlsson, K. Lindblad-Toh, I. Mellor, RD. Süssmuth, BG. Fry, S. Kuruppu, WC. Hodgson, J. Kool, TA. Castoe, I. Barnes, K. Sunagar, EAB. Undheim and ST. Turvey "Solenodon genome reveals convergent evolution of venom in eulipotyphlan mammals" *PNAS* **2019**, 116(51), 25745-25755.
   DOI: https://doi.org/10.1073/pnas.1906117116.

Author contributions

"The O-Carbamoyl-Transferase Alb15 is Responsible for the Modification of Albicidin"

D. Petras, D. Kerwat, A. Pesic, BF. Hempel, L. von Eckardstein, S. Semsary, J. Arasté, M. Marguerettaz, M. Royer, S. Cociancich, RD. Süssmuth

J.A., S.C., and A.P. performed the cultivation and isolation of carbamoyl-albicidin. D.P. and A.P. performed the MS/MS experiments. D.P. performed the bioinformatic analysis. D.P. and **B-F.H.** performed the heterologous production of Alb15 and performed the *in vitro* carbamoylation. M.M. and M.R. produced the gene inactivation mutant. D.K. and L.v.E. synthesized carbamoyl-albicidin. D.P. and S.S. performed the antibacterial and gyrase *in vitro* assays. D.P., A.P., and R.D.S. designed the study and analyzed the data. D.P. and R.D.S. wrote the manuscript. All authors read, discussed, and approved the manuscript.

"The Biosynthesis of Albicidin Involves On-line Enzymatic Methylation During Nonribosomal Peptide Assembly"

BF. Hempel, D. Petras, A. Mainz, L. von Eckardstein, RD. Süssmuth

B.-F.H., D.P. and R.D.S. planned the study. R.D.S. acquired funding and provided materials and instruments for the study. **B.-F.H.**, A.M. and D.P. conceived and designed the experiments. **B.-F.H.** and A.M. performed the experiments. **B.-F.H.** and L.v.E. performed chemical synthesis. **B.-F.H.** and A.M. performed data analysis. **B.-F.H.**, A.M. and D.P. wrote the manuscript. All contributors critically read, discussed, and approved the manuscript.

"Comparative Venomics of the *Vipera ammodytes transcaucasiana* and *Vipera ammodytes montandoni* from Turkey Provides Insights into Kinship"

BF. Hempel, M. Damm, B. Göçmen, M. Karis, MA. Oguz, A. Nalbantsoy, RD. Süssmuth

B.G., M.K. and M.A.O. catched and milked the snakes. **B.-F.H.**, A.N. and R.D.S. conceived and designed the experiments, **B.-F.H.** and A.N. performed the experiments, **B.-F.H.**, A.N. and M.D. analyzed the data. All authors contributed reagents/materials/analysis tools. **B.-F.H.**, A.N. and M.D. wrote the paper. All contributors critically read and revised the manuscript.

Table of Contents

Table of Contents

I   Figures

II   Tables

1   Motivation

Natural toxins are fascinating and sophisticated biochemical products synthesized by all forms of life and have been known to humans for many centuries.[1,2–4] Today, more than a thousand species in the kingdom Animalia, and even more microorganisms, are known to produce manifold kinds of toxins in all ecosystems around the world.[5,6] The term "toxin" was first described in 1890 by the organic chemist Ludwig Brieger (1849-1919), in the context of bacterial pathogens. The focus of his research was metabolic and infectious diseases with particular attention to *Salmonella typhimurium*, the causative agent of typhoid. Brieger demonstrated the generation of a toxic material (called typhotoxin) and coined the term – toxin.[7] The definition at that time was very simple because essential basics of toxicology had not been investigated yet: „*Toxins are poisonous substances, which in defined doses cause death of injected animals after a certain time of incubation.*"[8] In the following years toxins were differentiated between endo- and exotoxins based on the observations by Robert Koch and his colleague, Richard Pfeiffer, on their research of *Vibrio cholera*, a causal agent of cholera.[9] Endotoxins are part of the outer cell membrane of Gram-negative bacteria or cyanobacteria. Chemically, they are lipopolysaccharides that are composed of a hydrophilic polysaccharide and a lipophilic lipid portion. Endotoxins are heat-stable and even survive sterilization.[10] On the other hand, exotoxins are compounds secreted by bacterial pathogens and cause damage to the host by destroying cells or disrupting normal cellular metabolism.[11] A well-known representative of exotoxins is botulinum toxin (Botox) produced by the bacterium *Clostridium botulinum*. It causes flaccid paralysis by preventing the release of the neurotransmitter acetylcholine from axon endings at the neuromuscular junction.[12]

Besides toxins produced by microorganisms, there are many other excretions for example from fungi (mycotoxins), higher plants (phytotoxins), and animals (zootoxins) that fit into the definition of toxin.[2–4] Mycotoxins and phytotoxins are small natural products synthesized as various classes of secondary metabolites, largely consisting of alkaloids, flavonoids, glycosides, terpenoids, or peptides.[5,13,14] Historically, the 'gold rush' of searches for toxigenic agents derived from plants or fungi arose in the mid of 20th century (1960's) starting well-funded, large-scale screening programs, which were also favored due to the implementation of new instrumental analytics.[15,16] Interesting toxic agents that have emerged from these programs are aflatoxin (myotoxin, *Aspergillus flavus*) and digitoxin (phytotoxin, phytosteroid, *Digitalis purpurea*).[13,17]

However, among all natural-derived toxins, zootoxins have played an exceptional role in our understanding of toxins and cause enormous fascination to humans in all ages.[18] Zootoxins can

be differentiated into different categories: (1) poison – indirect by contact through touch, ingestion, or inhalation; (2) venoms – direct by injection through a bite or sting; and (3) crinotoxins – indirect by release into the environment, usually by means of a pore.[5] In 1843, Charles Lucien Bonaparte, nephew of Napoleon Bonaparte, penned a treatise on the composition of snake secretions describing its proteinaceous nature. Zootoxins are composed of complex cocktails containing high molecular weight components and therefore differ to natural-derived toxins of other domains of life. Biologically, zootoxins can be further distinguished by the method of delivery into venom and poison, meaning that poisons are ingested, while venom is delivered in a bite, sting, or similar action.[4,19] Nowadays, based on the knowledge of toxins gained in the 19th century, an updated and generally applicable terminology was published: Toxins can be small molecules, peptides, or proteins of natural or recombinant origin that are capable of causing disturbance to any organism, such as interaction with biological macromolecules or cellular receptors, completely independent of the route of administration. Toxins vary greatly in their toxicity, ranging from usually minor (bee sting) to almost immediately deadly (botulinum toxin) and unlike biological agents, toxins are inanimate and not capable of reproducing themselves.[20]

Friend or foe - Two sides of the same coin

Historically, all these toxins are often associated with negative examples as political intrigues, epidemic diseases, unethical experiments, scientific accidents, hunting/defense, or executions in human tradition.[21] Famous representatives of assassination by naturally occurring toxins are either belladonna, produced by the deadly nightshade (*Atropa belladonna*) and responsible for the death of various statesmen in the Roman Principate, or a poisonous hemlock cocktail (*Conium maculatum*), which is often associated to the execution of Socrates in ancient Greece.[22,23] One of the most famous but also devastating examples of natural toxins in human history was caused by the bacterium *Yersinia pestis*, a highly infectious disease better known as Pestilence or Black Death, that resulted in more than 50 million deaths in the 14th century by release of different exo- and endotoxins in the host.[24] However, mycotoxins were another common cause of epidemic diseases as a result of food infestation.[25] Ergotism, caused by the ingestion of the alkaloids from *Claviceps purpurea*, is the earliest recorded example of mycotoxicosis in Europe in the Middle Ages.[26] In the 1960s, another plague triggered by the carcinogenic aflatoxin from *Aspergillus flavus*, called Turkey X disease, caught attention to the population of England.[27] In the past, zootoxins were often used as military weapons, as for example by indigenous people, who used it to hunt or defend against intruders by applying it to weapons.[28] For example, an extremely poisonous alkaloid (batrachotoxin) from skin secretions of the terrible poison dart frog (*Phyllobates terribilis*), found in Central and South America, were used by the Chocó Indians of Colombia for blowpipe arrows.[29]

Even today, there exist prominent examples of assassinations with poison by naturally-derived toxins, such as cyanid, diamorphine or ricin.[23] However, envenomation by zootoxins is a far greater global burden nowadays, especially for women, children and farmers in many tropical and subtropical countries of poor rural communities.[30–32] Envenomation, and particularly snakebite envenoming, have the consequence of around 2.7 million tremendous socio-economic impacts and even 81,000 to 138,000 deaths each year.[33–35] For this reason, the world health organization (WHO) classified snakebite envenoming in 2017 as a major neglected tropical disease (NTD), along with dengue fever or Ebola, and introduced a strategy to prevent and control snakebite accidents.[36]

There is no kingdom of life which has not evolved toxins at a certain point of evolution for the purpose of hunting, defense or competition to interfere or disrupt the physiological processes of other organisms.[13,37,38] However, besides pathogenic effects there is also a positive side of the coin following the old adage that my enemy's enemy is my friend.

A prime example is the discovery and development of drugs against bacterial infectious diseases – antibiotics (based on Greek, (ἀντι-) anti: "against" and (βίος-) bios: "life"). In history, infectious bacterial diseases are responsible for some of the worst epidemics from Europe to Asia, like plague, cholera or tuberculosis that cost millions of lives in different periods of time.[39] Antibiotics are low-molecular natural secondary metabolites of fungi or bacteria that either inhibit the growth or kill competing microorganisms even in low concentrations.[40] The early stages of antibiotic discovery had already started in the late 18th century by the French military doctor Ernest Duchesne. In his doctoral thesis, he described the therapeutic capabilities of molds like *Penicillium glaucum* but failed to report a connection between the fungus and a substance that had antibacterial properties, so his research went unnoticed.[41] In Germany, research on treatments against infectious diseases started with the development of synthetic antibiotics derived from dyes by Paul Ehrlich at about the same time. He discovered that certain dyes could stain human, animal or even bacterial cells, whereas other dyes did not, and began to propose that it might be possible to use chemicals as a selective drug, colloquially named 'magic bullet'.[42] In 1907, Paul Ehrlich and his coworker Alfred Bertheim synthesized the first chemotherapeutic antibacterial drug, Salvarsan, a medically important organoarsenic drug against one of the largest public health burdens, "the great imitator", syphilis.[43] In the following years, several other scientists worked on potential naturally- or synthetically-derived antibacterial drugs. Alexander Fleming, a Scottish pharmacologist and director of the department of systematic bacteriology at St. Mary's London, found an invading fungus on a contaminated *Staphylococcus* culture plate. The colonies of staphylococci directly surrounded by the fungus were killed, while other staphylococcal colonies farther away showed normal growth.[44–46] Fleming grew the mold in a pure culture and identified it as being from the genus *Penicillium*. He found that it produced a

substance that killed a number of disease-causing, Gram-positive bacteria (e.g., *Staphylococcus* or *Streptococcus*), and after some months of calling it "mould juice", he named the substance "penicillin".[44,45] In 1929, Fleming published his discovery in the *British Journal of Experimental Pathology,* but little attention was paid to his research.[47] Fleming continued his investigations, but found that cultivating *Penicillium* was quite difficult, and that after having grown the mold, it was even more difficult to isolate the antibiotic substance. It took another 14 years before Ernst Chain, Howard Florey and Edward Abraham succeeded to efficiently purifying the first penicillin derivative (penicillin G) in bulk.[44,45] These events heralded the triumph of antibiotics in medicine and led to a rapid rise in natural product research. From 1935 to 1968, twelve new classes of antibiotic substances for medical application were launched in the so-called 'Golden Age of Antibiotics'. The discovery and development of antibiotics helped humanity to eradicate large epidemic plagues (e.g., the Pestilence or Cholera) and is one of the most important events in the history of medicine.[48]

On the other hand, indigenous people have known for centuries about the healing attributes of plants or animal secrets in small doses, which can have fatal consequences by incorrect dosage.[49] For example, *Atropa belladonna*, commonly known as belladonna or deadly nightshade, is a perennial herbaceous plant with long thick roots (mandrake) and has a long history of use as a medicine or even cosmetic.[50,51] The name "belladonna" originates from its historic use by fashionable women in the Renaissance era as cosmetic eye drops prepared by extracts of nightshade plants, as *"bella donna"* is Italian for "beautiful lady". The first written heritage of deadly nightshade plants as medicine goes back to the Hellenistic period, when Theophrastus recommended treatment of wounds, gout, and sleeplessness by liquid extracts of the plants. Cleopatra used extracts from a related Egyptian nightshade plant for cosmetic purpose of dilating the pupils of her eyes.[51,52] Later, it was recognized that wine of mandrake could be used as an anesthetic for treatment of pain or sleeplessness, to be given prior surgery. The application of atropine-rich medication for anesthesia, in combination with opium, persisted throughout the Roman and Islamic Empires until it was superseded in the 20th century by modern anesthetics.[51,53] Today, natural-derived belladonna alkaloids are used as pharmaceutical anticholinergics to provide peripheral anticholinergic or antispasmodic action by blocking the action of the neurotransmitter acetylcholine at synapses in the central and peripheral nervous system.[54] Other plant-derived medical drugs are cannabinoid chemicals that were already used in ancient China for millennia as fiber, food, medicine, and to induce spiritual moods.[55,56] Historical Chinese texts going back to the Eastern Han Dynasty report a manifold of applications for the herbaceous flowering plant *Cannabis sativa*, as clothing, production of paper and ropes, and continuous use in traditional folklore medicine.[57,58] From there, cannabis distributed to central Asia and Middle East became also available in Europe and Africa for religious ceremonies or funeral rituals found in pharmacological cults around the globe.[55,57] In the 19th century,

chemists and pharmacologists became interested in 'cannabinoid research' for clinical utility in a range of various disorders, such as cholera, rheumatic diseases, delirium tremens and infantile convulsions. Thereafter, several tinctures and fluids on the basis of cannabis launched the market for treatment of all kinds of diseases, which have often led to improper and abused use due to the psychoactive effect.[55,57] Since the beginning of the 20th century, most countries have criminalized cannabis and enacted laws against the cultivation, possession or transfer of cannabis.[59] Nowadays, cannabis use has increasingly been seen as a health issue instead of criminal behavior and in many countries is legalized or decriminalized under certain restrictions. In addition, different pharmacological prescription drugs made with cannabis are now marketed, as for example, Nabiximols, which is sold as mouth spray against several symptoms of multiple sclerosis.[60]

In contrast to phytotoxins, animal-derived secretions have been used to treat diseases for millennia. The exact historical moment, when humanity learned to use zootoxins for medical purposes, is difficult to determine. The earliest written scientific text of medical potential using venomous secretions belongs to Aristotle in ancient Greece, describing the production of antidotes. At the time of the Roman Empire animal venoms were used to produce drugs for the treatment of smallpox, leprosy, fever, and wound healing. A mixture containing snake venom, called theriac, was produced at this time and manufactured on a larger scale in pharmacies in Europe until the 18th century.[61] Today, ten venom-derived drugs are approved by the Food and Drug Administration (FDA) and freely available on the market from some of the world's most venomous creatures.[62] One of the first and most promising example of a venom-derived drug is captopril, found in the venom of pit vipers, *Bothrops jararaca*, in Brazil. Captopril is an angiotensin-converting enzyme (ACE) inhibitor, a type of drug used to treat hypertension, which reduces the risk of heart failure after a heart attack.[63]

In summary, toxins are evolutionary specialized natural products that harm competing organisms and gain their user an advantage. The fine line between deadly poison and highly effective medication is in many cases a question of dose. The Swiss-German alchemist Paracelsus already recognized this as early as the 16th century and stated that all things are toxic and only the doses decide if something might be a poison or not: „Alle Dinge sind Gift, und nichts ist ohne Gift; allein die Dosis machts, daß ein Ding kein Gift sei."[64] However, the close connection between dosage and mode of action for highly potent toxins was understood much later and culminated in the formation of an autonomous scientific discipline, named toxicology. Toxicology involves the study of adverse effects of chemical or natural-derived substances on living organisms and the practice of diagnosing and treating exposures to toxins and toxicants.[65]

Mode of toxic action – Adverse biological response to toxic diversity

The effect and dose of different toxins from various origins has been known for a long time and are partly well documented.[66] The beginning of pharmacodynamics and the scientific problem of what happens in the body after administration of the toxin has been a fundamental discipline in modern history. The origins of pharmacology date back to the Sumerians in Mesopotamia, focusing on herbalism and natural substances, mainly plant extracts.[67] Pharmacology as an academic discipline did not further advance until the mid-19th century, when the principles of scientific experimentation were applied to therapeutic contexts.[68] For a deeper understanding of how therapeutic drugs and toxins produced their effects, the first German pharmacology department was set up by Rudolf Buchheim in 1847.[69] The development of the ligand binding assay in 1960 by Rosalyn S. Yalow and Solomon Berson was a milestone in directed pharmacological sciences and allowed quantification of the binding affinity of drugs to chemical targets.[70] Today, modern clinical pharmacologists apply genetics, molecular biology, and biochemistry tools to gather specific information about molecular mechanisms and targets from various diseases, defects or pathogens.

The modes of toxic action are as varied as the toxins and their producers throughout the entire natural kingdom. An accurate classification is difficult but can be accomplished by separating toxins into defined categories according to the anatomical location where their effects are most notable: (I) cytotoxins – specific or non-specific harmful effects on certain cells; (II) hemotoxins - destroy red blood cells, disrupt blood clotting, and/or cause generalized tissue damage; (III) myotoxins - non-enzymatic process that leads to severe muscle necrosis; (IV) necrotoxins – destroy tissue cells; (V) neurotoxins - destructive to nerve tissue; (VI) phototoxins - cause allergic reactions and/or dangerous photosensitivity.[65,71,72] In the following, these classes are briefly discussed and examples shown to emphasize differences in microorganisms, plants or venomous animals.

Exemplary prominent cytotoxic agents (category I) are the immune response of T lymphocytes, which play a major role in host defense against intracellular pathogens, but also some types of venoms have cytotoxic effects. In addition, some naturally-derived substances, produced by various microorganisms, can be assigned as a kind of cytotoxic agents affecting specific to bacterial cellular processes, which are essential for their survival and result in cell death. These processes include DNA replication, RNA transcription, protein translation, cell wall synthesis, the cell membrane and a few other metabolic pathways.[73] The bicyclic octapeptide α-amanitin, most toxic of all amatoxins, is produced by several species of the mushroom genus *Amanita*. Amatoxins selectively inhibit eukaryotic RNA polymerase II, resulting in a dramatic decrease of transcription and protein synthesis, which is directly linked to its high toxicity, particularly to hepatocytes.[74] In the liver, α-amanitin is taken up by organic anion transporting polypeptides (OATP1B3),

specific hepatic transporters that mediate the uptake of many clinically important drugs.[75] However, the potent cytotoxic product ricin is, different to small molecular mass natural metabolites, a carbohydrate-binding protein belonging to the class of lectins.[76,77] The glycoprotein is produced in the seeds of the castor oil plant, *Ricinus communis*, and is classified as a type 2 ribosome-inactivating protein (RIP).[77,78] Type 2 RIPs consist of a functional active A chain, covalently connected by a single disulfide bond to a B chain that is catalytically inactive, but serves to mediate transport of the heterodimeric protein complex from the cell surface, via vesicle carriers, to the lumen of the endoplasmic reticulum (ER).[79] The amyloid-like cardiotoxin III (CTX III) is a basic polypeptide toxin of approximately 60 amino acids and derived from the Chinese cobra (*Naja atra*), a species in the family *Elapidae*.[80] The cytotoxic properties of CTX III are mediated by their interaction with the cell membrane disrupting the mitochondrial membrane integrity.[81]

Hemotoxins (category II) are frequently employed by venomous animals, especially vipers and pit vipers, and cause hemorrhagic and/or haemostatic effects triggered by highly precise enzymes or other proteins. For example, snake venom metalloproteinases (svMPs) show hemorrhagic effects by increasing vascular permeability via the degradation of capillary basement membranes, resulting in leakage and reduction in blood pressure.[82,83] However, in contrast to svMPs, the majority of snake venom serine proteases (svSPs) effectively degrade fibrinogen into fibrinopeptides via proteolytic cleavage, which result in the polymerisation of fibrin monomers and lead to unstable blood clots.[84] The envenomation of such functional and bioactive venom components plays a relevant role in the pathogenesis of venom-induced local tissue damage.[85]

The first myotoxin (category III) to be identified and isolated was crotamine, from the venom of *Crotalus durissus terrificus*. The small, basic protein acts extremely rapidly, inducing spastic paralysis to prevent prey from escaping and leads to severe muscle necrosis. The biological functions of crotamine are increase of the cation Na^+ by acting on the sodium channels of plasmatic membrane of skeletal muscle and increase of basal release of acetylcholine and dopamine.[86]

Necrosis, caused by necrotoxins, is a form of cell injury which results in an uncontrolled rupture and autolysis for all types of tissue. In humans, necrotoxins spread through the bloodstream and primarily attack skin and muscle tissue.[87] External factors, such as infections, trauma, or toxins are responsible for unregulated and fatal autolysis, which is in contrast to apoptosis, a programmed cell death providing beneficial effects.[88] The *Viperidae* (vipers) family of venomous snakes is a popular representative possessing various necrotoxins, such as phospholipases A_2 (PLA$_2$) and svSPs.[89] Other necrotic agents are multiple types of bacteria or occasionally fungi which lead to necrotizing fasciitis (NF), well-known as flesh-eating disease. The disease is classified into four types, whereas almost every patient is diagnosed with one of the first two forms.[90] Type I infections are the most common type of infection and is caused by a mixture of bacterial types.[91,92] The disease combines species of Gram-positive cocci, (*Staphylococcus aureus*,

Streptococcus pyogenes, and enterococci), Gram-negative rods (*Escherichia coli*, *Pseudomonas aeruginosa*), and anaerobes (*Bacteroides* and *Clostridium* species). The *Clostridium* species involved in the infection process cause excessive platelet aggregation which blocks blood vessels and deprives the vital organs of oxygen supply by producing two deadly toxins: alpha-toxin and theta-toxin.[93] The emerging acidic, oxygen-deficient environment is a good breeding ground for the proliferation of further bacterial pathogens.[91,92]

A fascinating class of exogenous toxins in human history is substances that are destructive to nerve tissue (causing neurotoxicity). Many well-known examples of neurotoxins have been reported, such as ethanol[94], glutamate[95], nitric oxide[96], botulinum toxin[97] (e.g. Botox), tetanus toxin[98], tetrodotoxin[99], or other animal-derived toxins adversely affecting the nervous system.[100] Neurotoxic processes range from inhibition of neuron control over ion concentrations across the cell membrane to blocking of communication between neurons across a synapse, by a number of different targets and mechanisms.[101] The membrane depolarization is caused by blocking various kinds of voltage-gated ion channels, such as sodium (Na^+)[99,102], potassium (K^+)[103,104], chloride (Cl^-)[105], or calcium (Ca^{2+})[103,104], due to injection of animal-derived proteinaceous toxins. On the other hand, bacterial neurotoxins (botulinum and tetanus toxin) inhibit vesicular neurotransmitter (e.g. acetylcholine, Ach) release at neuromuscular junctions and thus finally cause flaccid paralysis.[106] Once bacterial neurotoxins are bound to the presynaptic nerve terminals, neurons take up the toxin into a vesicle by receptor-mediated endocytosis. The continuous acidification of the vesicle through progressive transport into the cell, activate a part of the toxin which trigger the release into the cell cytoplasm. Inside the cytoplasm, the toxin starts to cleave SNARE proteins, responsible for vesicle fusions, which finally avoid binding of acetylcholine vesicles to the intracellular membrane.[107]

Phototoxins are a class, mainly produced by a variety of plants as biological defense. Many citruses or herbal remedies contain photoreactive chemicals that are activated by exposure to near ultraviolet light, such as sunlight, and can cause systemic photosensitivity.[108] For example, the common marigold plant produces alpha-terthienyl, a phytotoxin that functions as a nematicide. The phototoxin generates the toxic singlet oxygen when it is exposed to near ultraviolet light and result in damage to the respiratory, digestive and nervous system of larvae, making it a natural insecticide.[109]

In summary, modes of toxic action can reach from alteration or breakdown of enzyme activity, interference with the binding of toxic agents to proteins, intercalation with nucleic acids (RNA, DNA), electrolyte imbalance, and the disorganization or breakdown of cellular membrane lipids.[72,110] The mode of action of toxins may only set a small ball rolling in the thereby triggering various disastrous cascades that finally ends in death of the targeted organism. Interestingly,

many organisms show various types of resistances as evolutionary response to continued exposure of toxins.[111]

Resistances against toxins from microorganisms to higher vertebrates

Venom, as an example of biotoxins, has evolved into a wide variety throughout animal kingdom as a highly specialized weapon for predation. However, several examples of prey species show coevolutionary adaption in response to selective pressure imposed by the complex chemical weapons of predators and the development of chemical/physiological capacity to prevent or hinder the pathologic consequences of envenomation.

The mongoose (*Herpestidae* family) impressively proves to be resistant to the venom of the Chinese cobra (*Naja atra*), one of the most dangerous snakes in world, and shows how the hunter becomes the hunted. The highly neurotoxic venom of the elapid Chinese cobra contains a high amount of α-bungarotoxins (α-BTX), small proteins that competitively bind in a relatively irreversible manner to the α-subunit of the muscle nicotinic acetylcholine receptor (AChR).[112] The corresponding α-subunit of the mongoose AChR contains a number of substitutions in the ligand-binding domain, which markedly reduce the ability to bind α-BTX.[113] Interestingly, similar conformational changes in acetylcholine receptors have been documented in the Chinese cobra itself, which most likely protect against auto-envenomation or may allow evasion from cannibalism or predation by other sympatric elapid snakes.[114] Moreover, the woodrats (genus *Neotoma*) show a greatly elevated tolerance to the venom from the Western diamondback rattlesnake (*Crotalus atrox*) and related species. The rodents are able to significantly decrease the hemorrhagic effects of related venoms. It was found that one candidate, an antihemorrhagic resistance molecule, was able to bind and neutralize *C. atrox* toxins. Unfortunately, no further descriptive studies have been completed on this resistance molecule nor has any other phylogenetic information been revealed so far.[115]

Another well-described example of resistance against snake venoms are endogenous snake venom metalloprotease inhibitors (svMP-i), best documented as an auto-protection process against self-digestion in a reversible manner under physiological conditions[116–118] and in a number of squirrel species in the genus *Otospermophilus* (formerly *Spermophilus*).[119,120] The rock squirrel (*Otospermophilus variegatus*) serum was able to neutralize the pathological effects of the Western diamondback (*C. atrox*) as well as the prairie rattlenake (*C. viridis viridis*).[120] Further, it was shown that resistance could be ineffective against distant populations of the Northern Pacific rattlesnake (*C. oreganus oreganus*), indicating that resistance is geographically localized and requires offensive pressure from the co-localized rattlesnake population.[119]

The resistances that arise for self-protection or selective pressure by various vertebrates against venomous secretions could help to address snakebite envenoming, which is still an underrated

problem especially in rural and poor populations that costs many lives and causes permanent disabilities of several hundred thousand people a year.[30–33,35]

In turn, the constant evolution of antimicrobial resistance (AMR) in pathogens against actual antibiotics is a major global health problem of unprecedented magnitude.[121–123] Some microorganisms are naturally resistant to certain types of anti-infectives but can also become resistant by genetic mutation or acquiring resistance by horizontal gene transfer.[124] The background for the faster emergence of resistance is the much higher reproduction rate of unicellular microorganisms, called prokaryotes, in contrast to more complex organism (eukaryotes) and associated high frequency of new mutations.[125] Anti-infectives increase selective pressure in pathogen populations by causing non-resistant microorganisms to die, and allow increasing the percentage of resistant individuals which can continue to multiply. In addition, rising drug resistance is mainly favored by improper overuse in humans or other animals and ongoing the spread of resistant microorganism due to insufficient clinical healthcare and sanitation.[126] The constant generation of genetic mutations in bacterial pathogens yields to different types of resistance.[127] Some mutations enable bacteria to produce potent enzymes that actively modify antibiotics by inactivation (tetracyclines, monooxygenase)[128], modification (aminoglycosides, acetyl or phosphoryl attachment)[129], or cleavage (penicillins, β-lactamases)[130], while other mutations reduce antibiotic uptake (macrolides, permeability).[131] Still others eliminate the antibiotic cell target (vancomycin, peptidoglycan)[132] or manufacture pumping mechanisms to export the antibiotic back outside (fluoroquinolone, multidrug efflux pump).[133] The ongoing development of antimicrobial resistances and the lack of novel antibiotic classes, due to neglected investigations and strictly controlled clinical approvals, are heading towards a global post-antibiotic crisis.[123] Therefore, it is essential to intensify research into new antibiotic classes by tackling new antimicrobial targets.[134]

Novel therapeutical drugs or just more derivatives?

In the next years the world is running out of effective antibiotics. In the golden era of antibiotics, between 1949 and 1962, more than 20 new classes of antibiotics were marketed. In parallel, resistant antimicrobials emerged many times faster by mutation or horizontal gene transfer than in the majority of organisms, so that the development of antibiotics kept pace until 10-20 years ago. Therefore, an inadequate number of analogues are admitted to the market to stem the tide of antibiotic resistance of life threatening ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp.).[135] Today, a variety of potential anti-infectives are in the development or clinical phases.[136] However, these products are mostly derivatives of existing anti-infectives that

bypass evolved resistances. Nevertheless, many researchers and pharmaceutical companies have tackled this problem and intensified their investigations into new antibiotic classes.[137]

For instance, penicillin-binding proteins (PBPs) have been the focus of a great deal of research since penicillin was discovered as the first antibacterial drug.[47] The mechanism of action and selective toxicity for antibiotics targeting PBPs has always been a well-studied drug target and resulted in a variety of other modern antibacterial drugs. PBPs are members of a subgroup of enzymes called transpeptidases involved in the final stages of peptidoglycan synthesis, which is the major component of bacterial cell walls.[138,139] The class of β-lactam antibiotics binds to the active site of PBPs and the β-lactam amide bond is ruptured to form an irreversible covalent bond with the catalytic serine residue to finally inactivate the enzyme.[139] However, methicillin-resistant *S. aureus* (MRSA) isolates were observed within a few years after introduction to the clinic, due to production of an alternative penicillin-binding protein (PBP2a) that is resistant to currently available β-lactam antibiotics.[140] Based on the resistance mechanism, there are essentially two options to allow the continued application of β-lactam antibiotics. The design of new β-lactam antibiotics through side chain modification, that are not affected by the above-mentioned bacterial resistance mechanisms, or a combinatorial drug application of current β-lactam antibiotics with a drug that disables the resistance mechanisms.[141] Both approaches have been extensively investigated and several new drugs are currently in different clinical phases; however, the past has shown novel β-lactam compounds not keeping pace with clinical resistance development.[142] One key to overcome existing resistances is to identify other prokaryote-specific cellular processes essential for survival and develop new antibacterial drugs to kill pathogens.[143] A promising lead structure with a novel antibacterial drug target, which is part of this thesis, is active against Gram-negative and Gram-positive pathogens up to the nanomolar range.[144] The phytotoxin, called albicidin, is produced by the xylem-invading sugarcane pathogen *Xanthomonas albilineans*, which induces chlorosis by inhibition of the plant DNA replication.[145] Albicidin blocks the plant DNA replication by inhibition of DNA gyrase subunit A, different to other formerly known gyrase inhibitors, such as coumarins and quinolones. The bactericidal agent interferes in the covalent DNA binding pocket of the GyrA subunit, while other gyrase inhibitors act at the ATP binding site of the GyrB subunit.[146] The structure elucidation of albicidin revealed a number of unusual non-canonical amino acids, which are assembled by a polyketide-nonribosomal peptide synthetase (PKS-NRPS) hybrid.[147,148] The biosynthesis of albicidin showed the incorporation of different para-aminobenzoic acid (pABA) variants and β-cyanoalanine (β-Cya).[149] Afterwards, other PKS-NRPS-derived natural products, termed cystobactamide and coralmycin, were discovered and show a high similarity in both, molecular structure and biosynthetic gene cluster.[150,151] Interestingly, substrate specificity experiments for respective adenylation (A) domains revealed a preferential incorporation of substrate para-3-hydroxy-aminobenzoic acid (pABA-3OH) rather than previously noted para-2-hydroxy-3-methoxy-benzoic acid (pMBA).[149]

However, in spite of remarkable bioactivity for this lead structure, there are several resistance mechanisms already known. Initial investigations for biological control agents against the phytotoxin by Walker et al.[152] revealed an albicidin resistance gene (*albA*) in the organism *Klebsiella oxytoca*, resulting in the production of the AlbA protein that was later suggested to be responsible for trapping albicidin in a non-covalent complex.[153] Recently, our group disclosed the AlbA-albicidin binding complex by co-crystallization and reported on the ligand promiscuity by structure-guided NMR binding studies.[154] Another resistance factor against the leaf scald pathogenic agents described by Zhang et al.[155], showed the identification of a resistance mechanism from *Pantoea dispersa*, a Gram-negative Enterobacterium. The corresponding enzyme AlbD was discussed as a potential hydrolase enzyme, but exact mechanism remained unknown. The knowledge about the molecular structure of albicidin allowed us to identify the detoxification mechanism and characterize AlbD as a novel type of endopeptidase that catalyzes the cleavage of albicidin at the β-Cya peptide backbone amide.[156] Nevertheless, the advanced biosynthetic knowledge and combinatorial synthesis route for albicidin enabled us to be one step ahead for potential resistances evolving in pathogens.[157,158,159] The biochemical investigation of a highly-active and soluble albicidin derivative and the unusual construction of the pMBA motive by mass spectrometry-guided techniques is part of this work and is explained in more detail in chapters 2 and 3.

Another antibacterial drug target that has received increasing attention are the two-component signal transduction systems of bacteria, which consist of two proteins, a histidine kinase (HK) and response regulators.[160] The WalK/WalR (formerly known as YycG/YycF) two-component signal transduction system has appeared as a promising candidate, as it is indispensable in the signal transduction pathway for the cell-wall metabolism of *B. subtilis* or *S. aureus* and allow them to rapidly adapt to physical, chemical and biological stresses from outside the cell. The membrane-linked HK (WalK) shows a highly conserved C-terminal cytoplasmic region, consisting of an ATP binding site and phosphoacceptor domains, which makes it to an ideal candidate for inhibitor screening. Interesting resources for potential anti-infectives that target the two-component system are natural products produced from actinomycetes.[160–162] A panel of different chemical-related compounds were isolated and characterized to be a potent inhibitor for the two-component signal transduction system. The inhibitor walkmycin B and related compounds (walkmycin A and C) has a bactericidal effect in a lower micromolar range and showed specific dose-dependent binding to prevent autophosphorylation of WalK.[160] The natural product waldiomycin, belonging to the family of angucycline antibiotics, also showed remarkable inhibition of the autophosphorylation of WalK.[162] Site-directed mutagenesis and NMR experiments showed that waldiomycin directly binds to two conserved regions in the dimerization-inducing and histidine-containing phosphotransfer domain (H-box and X-

region).[163] Similar results could be shown for the actinomycete-derived compound signermycin B by cross-linking experiments.[164]

Further actinomycete-derived secondary metabolites, such as platensimycin and platencin, show bacteriostatic inhibition in a low micromolar range without cross-resistance to other classes of resistant Gram-positive pathogens. The drug target is the bacterial fatty acid synthesis type II (FASII), which is among other functions, required for energy storage.[165] The fact that the FASII system is only present in prokaryotes makes it to an excellent target for antibacterial drug discovery.[166] While platensimycin targets the elongation-condensing enzyme FabF, which is responsible to elongate the growing fatty acid chain by acetate units, platencin is a dual inhibitor of FabF and the initiation-condensing enzyme FabH.[167] Platensimycin compete with malonyl-ACP for the malonate-binding site of FabF and the ability to bind to FabH is highly dependent upon the conformation of the cyclohexenone ring.[168] Unfortunately, these inhibitors are only active against Gram-positive pathogens and poor pharmacokinetics prevented them from entering clinical trials.[167,169] In contrast, afabicin, which targets prokaryotic enoyl-acyl carrier protein reductase (FabI), showed excellent pharmacokinetic and bioavailability properties and is currently in clinical trials. The *in vivo* efficacy of afabicin in multiple staphylococcal-infected animal models demonstrated significant activity and high bone-to-plasma ratios of its active moiety.[170]

Not only microorganisms produce a large number of live-saving drugs. Animal venoms from invertebrates to vertebrates are composed of varieties of proteins and peptides, which also have a major role in therapeutical drug development. These toxins, fine-tuned by millions of years of evolution, attack multiple targets, such as ion channels, receptors, and enzymes with high potency and selectivity.[171,172] Captopril, an anti-hypertensive drug targeting the angiotensin-converting enzyme (ACE), derived from a bradykinin-potentiating peptide (BPP) from the highly venomous pit viper *Bothrops jararaca*, that is endemic to the tropical and subtropical forests in South America.[173] Another approved and life-saving drug, derived from a venomous animal, also help to bypass a major global public health problem that affects more than 300 million people worldwide.[174] Today, obesity and associated insulin resistance are crucial health problems and key contributors for type 2 diabetes.[175] Complications associated with diabetes manifest are coronary heart disease, stroke and peripheral vascular disease that lead to an increased mortality in economies of all nations, particularly developing countries.[176] The Gila monster (*Heloderma suspectum*), one of the very few venomous lizards in the world, produce a peptidic component named exendins, which have high sequence homology to glucacon-like peptides (GLP).[177] In contrast to GLP-1, a common diabetes drug that has a very short plasma half-life (~2 min), exendin-4 (exenatide) was developed with a greatly increased *in vivo* serum half-life (2.4 h).[178] In addition to approved venom-derived drugs, such as aforementioned examples, several potential toxin peptides are currently in clinical or preclinical development.[171,172] Recent

research indicates that animal venoms are rich in neuroactive molecules, which could help against neurodegenerative diseases and disorders including Parkinson's disease, Alzheimer's disease, Huntington's disease, epilepsy, multiple sclerosis, and amyotrophic lateral sclerosis.[179] The α-conotoxins, originally isolated from the venom of the marine cone snail (*Conus geographus*) as well as dendrotoxins, extracted from the venom of the African mambas (genus *Dendroaspis*), have potent analgesic activity by blocking potassium channels and therefore increase acetylcholine release. [180] The α-conotoxin Vc1.1, derived from cone snail *Conus victoriae,* was taken into clinical trials or the treatment of neuropathic pain, but unfortunately failed in phase II trials owing to a lack of efficacy in humans.[181] In a more recent study, Vc1.1 was synthetic engineered by a backbone cyclization, which showed a greatly improved stability and reduced susceptibility to proteolysis.[182] Perhaps one of the most intensively investigated diseases in human history is cancer, but to date there is no miracle cure to defeat the scourge of humanity.[183] Nevertheless, screening of the death stalker scorpion venom, *Leiurus quinquestriatus*, revealed a peptide toxin that blocks chloride ion channels.[184] A unique feature of the 36-residue peptide, named chlorotoxin, is high affinity binding to matrix metalloproteinase (MMP-2) subtypes that are upregulated on the surfaces of glioma and other cancer cells but are not present in normal cells.[185] For this reason, chlorotoxin has been undergoing clinical development as both an *in vivo* diagnostic tool for cancer as well as a potential therapeutic delivery system for radiochemical treatment of malignant cells.[186] The I^{131}-labeled chlorotoxin showed in phase I and II trials with intracavitary administration for glioma no dose-limited toxicity and long-term retention to the tumor periphery. In addition, coupling a cyanine dye to chlorotoxin as an imaging agent helps to locate cancer cells. The fluorescent dye Cy5.5 emits photon in the near infrared spectrum and enables visualization for the surgeon.[187] The Northern short-tailed shrew (*Blarina brevicauda*), found in North America, is one of very few venomous mammals and another fascinating example for a toxic component with potential therapeutic application. The venom is mainly composed of kallikrein-like serine proteases, but also contains a paralytic peptide named soricidin.[188] C-terminal truncation of soricidin (SOR-C13), while not being paralytic, blocks calcium ion uptake by ovarian cancer cells via inhibition of the transient receptor potential vanilloid calcium channel subtype six (TRPV6). Recent phase I studies with patients of advanced epithelial tumors showed no drug-related serious adverse events.[189]

In summary, venom-derived peptides and proteins from venomous animals are natural ligands of membrane ion channels or receptors with excellent specificity and high potency. Though the market of protein- and peptide-based drugs is constantly growing, venom-derived drugs are not without obstacles. Peptide therapeutics are still far behind small molecules, due to the increased proteolytic instability, oral bioavailability or passing of various biological barriers. Nevertheless, the therapeutic potential of peptide-based drugs is increasingly appreciated and their development is both strong and growing rapidly.[190] Parts of chapter 4 deal with bioactivity

screenings against several human cancerous and non-cancerous cell lines of potential antitumor candidates, which demonstrated strong cytotoxic effects on specific breast cancer cells.

The method of choice - Mass-spectrometry combined to big data science

For a long time, analyses of complex natural toxins was quite challenging and required tremendous effort and time even for small organic molecules that were characterized with a number of complex chemical experiments, like derivatization or degradation.[191] Natural products, such as morphine, strychnine or tetrodotoxin (puffer fish toxin) for example, required several decades to more than a century for complete three-dimensional structure elucidation.[192] In the mid-20th century, structure elucidation of complex molecules with isosteric derivatives has been revolutionized step by step with the introduction of various spectrometric instrumental methods. In the 1960s, mass spectrometry (MS) and nuclear magnetic resonance (NMR), two major key players in future analytical sciences, were introduced to the analytical field and gained momentum for the analysis of medically important natural toxins of various origin.[15,16,193] For the analysis of highly complex mixtures, MS represented the method of choice, because gas chromatographic (GC) or later liquid chromatographic (LC) systems could be front-end coupled and therefore allowed a rapid on-line separation of multiple components within one sample.[194] Tandem MS, also known as MS/MS or MS^n, was introduced shortly thereafter and even enabled the identification of isomeric derivatives by so-called mass fingerprints. The tandem mass spectrometer ionizes molecules and analyzes their mass-to-charge (m/z) ratio, called "precursor" or "parent" ion, in a first mass analyzer (MS1). Selected ions of a particular m/z-ratio coming from MS1 are trapped and accumulated in an ion trap and then split into fragment ions, so called "product" or "daughter" ion by collision-induced dissociation, ion-molecule reaction, or photodissociation. These fragment ions are then again introduced into the MS1 or a second mass analyzer (MS2), which in turn detects the m/z-ratio to assign precursor molecules by unique mass fingerprints.[195]

The latest achievements of mass spectrometry enabled life sciences to enter the field of analytic sciences, also called bioanalytics. The development of electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI), soft alternatives to electron ionization (EI), in combination with the improvement of time-of-flight (TOF) mass analyzer as well as the development of the orbitrap mass analyzer, opened the door for the analysis of high molecular weight molecules, such as peptides and proteins, to map full proteomes even up of intact cells, e.g. viruses, spores, bacteria or human cell lines.[196] Today, a major goal in the analytical field of natural toxins is to turn coming from an one-off process to an optimized high-throughput discovery pipeline of complex mixtures with various molecular mass ranges that show the identification of a variety of novel lead structures with potential new therapeutical modes of

action.[197] Thus, in addition to new experimental technologies, future high-throughput screenings of natural products will also require new computational tools.[198]

A recent powerful computational technology makes use of the unique mass fingerprints of various natural metabolites to distinguish not only between known and unknown metabolites through high-throughput dereplication, but also to aid in helping the identification of "*known unknowns*" metabolites by comprehensive LC-MS/MS reference libraries.[199] The term coined in 2011 in context of LC-based MS libraries by Little et al.[200], should illustrate that the identification of an *unknown* chemical substance for an experimental user might actually already be *known* in the chemical literature or any other resources. A follow-up bioinformatic tool groups sets of tandem spectra from related molecules by their spectral similarity (cosine score), even when spectra themselves do not match to any spectral library hit, and thereby visualizes so-called molecular families as molecular networks.[201] Recently, molecular networking helped to identify additional albicidin derivatives from the wild-type producer *X. albilineans*.[159]

A major drawback of mass spectrometry in contrast to other analytical techniques was the lack of insight into structural system biology and spatial localization of biological processes for broad spectrum of analytes ranging from small molecules, drugs and metabolites over peptides, single proteoforms or full proteomes.[202] Biological processes are orchestrated by highly dynamic mechanisms, such as structural transitions or fluctuated interactions, that are achieved in a very short time scale. The fluctuated and transient events of functional quaternary biomolecules cannot be trapped by standard MS approaches. In recent decades, several MS-based techniques came up that allowed the analysis of native complexes by MS.[203]

The cross-linking MS (XLMS) technology is able to map transient structural details and model functional biomolecule assemblies at physiological conditions.[204] Quaternary structures or dynamic complexes of biomolecules are preserved by chemical reagents that introduce covalent bonds at functional amino or hydroxy groups of amino acid side chains (Lys, Thr, Ser) allowing subsequent MS identification of artificially fused biomolecules under denaturing conditions.[205] Nevertheless, acquisition of the direct linkage sites would help to increase the resolution of the method for structure determination from proteins to domains or even smaller sections, dubbed as "peptide-level resolution".[206] The next-generation cross-linker was designed with defined spacers of different length acting as a kind of 'molecular ruler' and chemical groups allowing cleavage of artificial introduced bonds by dissociation in MS.[207,208] After cross-linking and subsequent enzymatic digestion, three different types of cross-linked peptides are observed. In addition to interpeptide cross-links (type 2), which give the most valuable structural information, intrapeptide (type 1) and dead end (type 0) cross-links can be identified via pseudo-MS3 fragmentation.[208,209] Today, cross-linking MS has been turning from *in vitro* studies of multiprotein complexes towards *in vivo* studies of cells or tissues on a global scale.[210] The application of this advanced methodology was used in Chapter 3, to prove the protein interaction

of the methyltransferase Alb02 and peptidyl carrier protein in module 4 and 5 (PCP-4 and PCP-5) of the albicidin biosynthesis assembly line.

Limited proteolysis-coupled mass spectrometry (LiP-MS) allows the identification internal perturbation-induced biomolecule structural alterations, such as interactions, chemical derivatization, or mutations with peptide-level resolution.[211] LiP experiments on a proteome-wide scale rely first on proteases with broad specificity applied for a short time to a proteome extract under native conditions and is followed by a second digestion step that generates peptides amenable to bottom-up proteomic analysis. Unfortunately, LiP-MS does not allow identification of a direct interaction partner *per se* and requires a high effort of additional control experiments. Nevertheless, LiP-MS is an excellent complementary tool for the identification of protein structural changes upon specific environmental perturbations in a targeted analysis or on a global scale.[212]

The spatial localization of biological compounds, such as biomarkers, metabolites, peptides or proteins, in a multidimensional manner can be achieved by mass spectrometry imaging (MSI).[213,214] The spatial distribution of biomolecules is scanned from surfaces of biological samples in a two-dimensional distribution without any chemical labels.[215] The most common ionization methods for MSI are desorption electrospray ionization (DESI), secondary ion mass spectrometry (SIMS), or MALDI. The mass analyzer of choice for highly complex and high molecular weight samples, however, is a TOF system that achieves a good speed, sensitivity and a broad mass range detection ($m/z\sim1$–100,000).[214,215] The numerous technological advances in recent years allowed MSI to become a robust tool for biomarkers in clinical practice and the pharmaceutical industry.[216] A set of several two-dimensional biological samples analyzed by MSI imaging can be stacked by complex and computationally intensive programs and enable to go into a third dimension of biological samples.[217] The application of MSI will be discussed later in more detail in Chapter 7. We used this approach to localize various toxin families as well as toxin proteoforms in the venom gland apparatus of the Egyptian cobra (*Naja haje*), one of the most medically important snakes in Africa with regard to snakebite envenoming.

The entry of mass spectrometry in the field of toxinology enabled researcher to decipher the biological composition and understand the mechanism of action of venoms from manifold origins and design more effective antidotes for the treatment of envenomation.[218,219,220] The inter- and intraspecies heterogeneity in venom composition and associated different clinical symptoms observed in human victims of envenoming are a major concern in proteomic examination of the venom. Understanding the variation of venoms and their different antigenic constituents related to distinct geographic origin represents thus a key challenge towards the design of novel, toxin-specific antivenoms.[220,221,222] Since the first comprehensive analysis of snake venom by bottom-up MS, several other methods have been established and coined under the term "venomics".[223–226] Recent advances in the field of snake venomics applied high-resolution mass spectrometer

(HRMS) and facilitated the absolute quantification of venom compositions or identification of toxin families by so-called top-down venomics combining intact toxin masses and fragmentation. The approach is a fast alternative to established bottom-up approaches, due to the coupled front-end separation and subsequent analysis by fragmentation in a collision cell.[227–229] However, the different established venomics approaches suffer from different drawbacks. The bottom-up strategies can be divided in in gel-based and liquid chromatographic (LC)-based approaches or a combined workflow from both strategies that are very extensive and time-consuming processes. The proteolytic step in bottom-up proteomics often obscures the differentiation of toxin proteoforms and prevents the identification of post-translational modifications (PTMs).[225,226] These reasons make the application of bottom-up strategies for larger comparative inter- and intraspecies investigations unfavorable.[230] The recently introduced intact mass profiling in combination with top-down proteomics in turn are an excellent alternative to comprehensively annotate venom proteomes with toxin families up to 30 kDa, such as snakes in the genus of *Elapidae*.[227,231,232] In cases of higher-molecular-weight compounds (>30 kDa), which are typically strongly represented in the genus of *Viperidae*, the top-down analysis only provides a partial characterization and is still challenging due to inefficient ionization by denaturing ESI and too few observable fragments in tandem MS.[233,234] The comparative venom profiling with our aforementioned methods at the level of individual subspecies or larger populations are part of this work and further discussed in chapters 4 and 5. In addition, we establish in chapter 6 a *de novo* in-source decay-driven (ISD) venomics workflow using MALDI as an alternative top-down approach to characterize high molecular mass venom components for the newly discovered Anatolian meadow viper subspecies, *Vipera anatolica senliki*.

In summary, mass spectrometry becomes a fundamental and powerful analytical tool that entered into life sciences and is gaining significant impact in the '-omics' fields, such as metabolomics or proteomics. Additional (bio)chemical applications helped to expand MS to structural system biology. The present dissertation and the publications in Chapter 2 and 3 deals with state-of-the-art mass spectrometry methods in combination with NMR spectroscopy to fully elucidate decisive parts of the biosynthetic pathway of albicidin that are achieved by modification enzymes acting either post-NRPS (Chapter 2, Alb15) or as on-line (Chapter 3, Alb02) assembly.

Furthermore, publications in chapter 4 and 5 apply alternative MS workflows to fully characterize venom proteomes of closely related subspecies or populations by a fast and quantitative workflow. In chapter 6, an alternative top-down venomics approach is described to become aware of the drawbacks of the aforementioned top-down venomic workflow.

Finally, in chapter 7 results are summarized and discussed in context of state-of-the-art MS. Furthermore, future perspectives are given for the enlightenment of the complete albicidin biosynthesis pathway as well as evolutionary insights into the venom gland apparatus by MSI.

42

2 The O-Carbamoyl-Transferase Alb15 is Responsible for the Modification of Albicidin

Introduction

Sugarcane leaf scald, a lethal disease of sugarcane, is caused by the xylem-invading bacterium *Xanthomonas albilineans*[235] which belongs to the class of γ-*Proteobacteria*. Transmission of the pathogen usually occurs through human practices during the reproduction of the sugarcane plants by preparing cuttings of already infected plants or by using contaminated cutting tools.[236] However, potential aerial transmission has been reported as well.[237] The plant colonization occurs through the xylem vessels of the plants where *X. albilineans* induces chlorosis which becomes apparent by characteristic white pencil-like lines along the leaves. As it has been shown recently, in more advanced stages of sugarcane leaf scald, *X. albilineans* also spreads into the surrounding tissue of the xylem tubes.[238] A characteristic feature of *X. albilineans* is its ability to synthesize the phytotoxin albicidin, which is encoded in the *alb* gene cluster, a polyketide synthase-non-ribosomal-peptide synthase hybrid (PKS-NRPS). Albicidin, which was first reported decades ago, is a small molecule showing a high inhibitory activity against plant and bacterial DNA-gyrase[148,149,239] and is one of the pathogenic factors of sugarcane leaf scald.[240] Because albicidin is only produced in minute amounts (∼10 μg/L) in cultures of *X. albilineans*, its structure remained elusive for more than 30 years. The high antibacterial activity against a wide range of gram-positive and gram-negative strains[241] combined with a new inhibition mechanism of gyrase makes albicidin a potential candidate for the development of new anti-infectives.[146] Unlike fluoroquinolones and coumarins, which target the ATP-binding pocket[242], albicidin stabilizes the covalent DNA-gyrase complex.[146]

As new antibacterial drugs are urgently needed for the treatment of arising antibiotic resistant strains, we put a significant effort into the increase of albicidin production by using a heterologous *Xanthomonas* strain[243] which finally enabled us to solve the structure of albicidin by a combination of extensive MS/MS and NMR experiments.[149] Having already established the structure of albicidin, we performed its total synthesis, aiming at structure activity relation (SAR) studies to assess albicidin derivatives for drug development purposes.[244] On the other hand, the structure set the basis for our biochemical investigations on the biosynthesis of albicidin in which we introduced para-amino benzoic acids (pABA) as a new class of substrates activated by non-ribosomal peptide synthetases (NRPS).[149] Next to pABA as a characteristic structural feature, albicidin (**Figure 2.1A**) has at the *N*-terminus a coumaric acid derivative and L-cyanoalanine as

the central amino acid, which is most likely produced at an insertional module of a *trans*-acting NRPS domain based on L-asparagine as a precursor.[149] By the combination of *in silico* analysis of the proteins present in the albicidin cluster and *in vitro* testing of the adenylation domain (A-domain) substrate specificity, we proposed a comprehensive biosynthesis model for albicidin.[149] Nevertheless, some gene functions present in the cluster remained unclear in terms of PKS-NRPS assembly as well as the tailoring reactions involved in post-NRPS processing of albicidin.

A putative *O*-carbamoyl transferase gene is part of the albicidin biosynthesis gene cluster (*alb* cluster)[149,239] which finds it precedence in other natural products, e.g. cephalomycin[245], novobiocin[246] and tobramycin[247]. Here we show the structural elucidation and biosynthesis investigation of the putative carbamoylated albicidin derivative through a combination of tandem mass spectrometry, *in vitro*- and gene knockout experiments. Finally, we confirmed the structure by the total synthesis of carbamoyl-albicidin, which furthermore enabled us to study the impact on bioactivity and pharmacological properties, which are part of our ongoing synthetic optimization of the albicidin structure as a new antibacterial drug.

Results and Discussion

During the purification of natural albicidin by preparative HPLC and LCMS analysis of the chromatographic fractions, we identified another minor abundant compound with a delta mass of +43.0030 Da compared to albicidin ([M+H]$^+$ = 843.2636 Da). The exact mass of the pseudomolecular ion of this new compound [M+H]$^+$ = 886.2666 Da corresponds to a molecular formula of $C_{45}H_{40}O_{13}N_7$ (mass error Δm = -2.0 ppm, calculated exact mass [M+H]$^+$ = 886.2684 Da). This differs from the molecular formula of albicidin by a fragment corresponding to CONH which could be interpreted as a carbamoyl group. This assumption is in line with the presence of a putative carbamoyltransferase gene (*alb15*) in the albicidin gene cluster. Alb15 shows high similarity (~30% identity) to TobZ, an ATP-dependent *O*-carbamoyltransferase involved in the carbamoylation of tobramycin of which structural data is available.[247] Interestingly, all amino acid residues in TobZ, which are involved in complexation of a metal ion and ATP binding, are highly conserved in Alb15 as well (alignment is shown in **Appendix Figure 2.1**). This indicates that Alb15 is most likely an ATP-dependent carbamoyltransferase.

In order to unambiguously identify the carbamoylation site and due to only minute amounts of carbamoyl-albicidin from fermentations, which were insufficient for NMR experiments, we performed mass spectrometric product ion scans of both albicidin and carbamoyl-albicidin (**Figure 2.1**), respectively. Albicidin ([M+H]$^+$ = 843.3 Da) shows a characteristic b-ion series at 160.6 Da (b$_1$); 280.0 Da (b$_2$); 495.0 Da (b$_4$) and 660.0 Da (b$_5$). Additionally, we observed two characteristic y-ions at 330.6 Da (y$_2$-H$_2$O) and 468.0 Da (y$_3$).
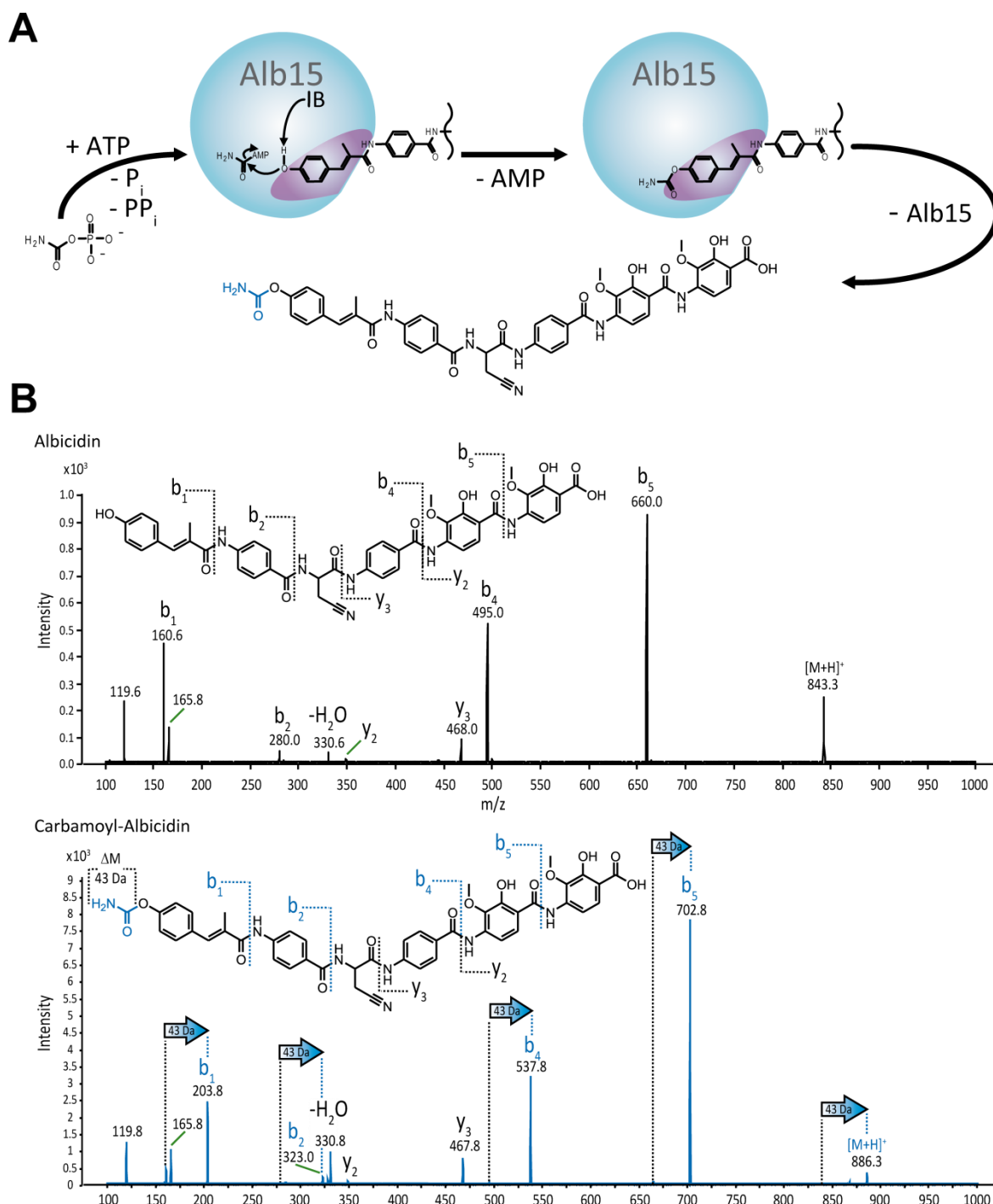
Figure 2.1. (**A**) **Reaction scheme of post-NRPS carbamoylation of albicidin.** As an ATP-dependent *O*-carbamoyltransferase, Alb15 catalyzes first the reaction from carbamoyl-phosphate and ATP to carbamoyl-AMP followed by the carbamoylation of albicidin. (**B**) **MS/MS structural elucidation and comparison of albicidin and carbamoyl-albicidin.** The observed b-ion and y-ion fragments and the mass shift through carbamoylation are indicated in the spectra and structure.

For the b-ion series of the putative carbamoyl-albicidin, we observed a characteristic pattern of ions at 203.8 Da (b_1), 323.0 Da (b_2), 537.8 Da (b_4) and 702.8 Da (b_5) which is shifted throughout the spectrum by 43.0 Da. The y-ions observed for albicidin could be seen for carbamoyl-albicidin at 330.8 Da (y_2-H_2O) and 467.8 Da (y_3) as well. This indicates that the carbamoyl group must be

located at the *N*-terminus, attached to the *para*-hydroxy group of the methyl-coumaric acid. The ultimate proof for such a structural arrangement came from the total synthesis of carbamoyl-albicidin. The main synthesis strategy is based on an orthogonal protecting group at the *para*-hydroxy function that allows site-specific carbamoylation before the global deprotection step (shown in **Appendix Scheme 2.1**), according to the recently published total synthesis of albicidin by our group.[244] Due to the fact that acidic, basic and hydrogenolytic conditions could not be applied for the total synthesis, we decided to use a *tert*-butylsilyl (TBS) protecting group, which is mildly cleavable with a fluorine source. The carbamoyl moiety was introduced by using chlorosulfonyl isocyanate (CSI) followed by global deprotection according to Dauvergne et al.[248]. To this end, synthetic carbamoyl-albicidin had an identical retention time in analytical HPLC runs as the product isolated from the host organism. The exact mass of synthetic carbamoyl-albicidin was determined as $[M-H]^- = 884.2554\,Da$ (mass error $\Delta m = -3\,ppm$, calculated mass 884.2527 Da). The pattern of the MS/MS spectrum of synthetic carbamoyl-albicidin exactly matches with the pattern of the natural product, thus ultimately giving proof of the proposed structure (**Appendix Figure 2.2**).

In order to characterize the gene function of *alb15*, a gene inactivation mutant was generated. Mass spectrometric characterization by multiple reaction monitoring (MRM) of *X. albilineans-Δalb15* compared to *X. albilineans* wild type showed the absence of carbamoyl-albicidin in the mutant strain, while albicidin still could be detected in both cultures (**Figure 2.2**). Subsequently, we cloned the *alb15* gene into the expression vector pETtrx_1c and heterologously expressed the gene as a thioredoxin-Alb15-His$_6$ fusion protein in *E. coli BL 21 gold*. After protein purification through Nickel-affinity chromatography and gel filtration, yields obtained for Alb15 were ~5 mg/L$_{culture}$ (**Appendix Figure 2.3**). The purity and identity of the fusion protein was verified by SDS-PAGE (**Appendix Figure 2.3**), in-gel trypsin digestion and LC-MS/MS analysis of the tryptic peptides (**Appendix Figure 2.4**). We then reconstituted the carbamoylation by incubating purified Alb15 with ATP, albicidin and carbamoyl-phosphate. The detection of carbamoylated albicidin was performed by ESI-MRM mass spectrometry. The MRM chromatograms in figure 2.2A show four points of time (1, 10, 20 and 120 min), including negative controls (no carbamoylphosphate and no ATP at 120 min). We observed peaks at a characteristic retention time for carbamoyl-albicidin of $R_t = 3.0\,min$ from approximately 20 min incubation onwards. These results indicate that the formation of carbamoyl-albicidin occurs with albicidin as a substrate, most likely post-NRPS, which is in line with other known carbamoylation reactions, as for example the post-PKS carbamoylation of ansamitocin.[249]

To assess the effect of carbamoylation on the physicochemical properties and antibacterial activity of albicidin which may be of significance for further medicinal chemistry studies, we performed a simple agar halo assay against *E. coli* DH5α (**Appendix Figure 2.5**).
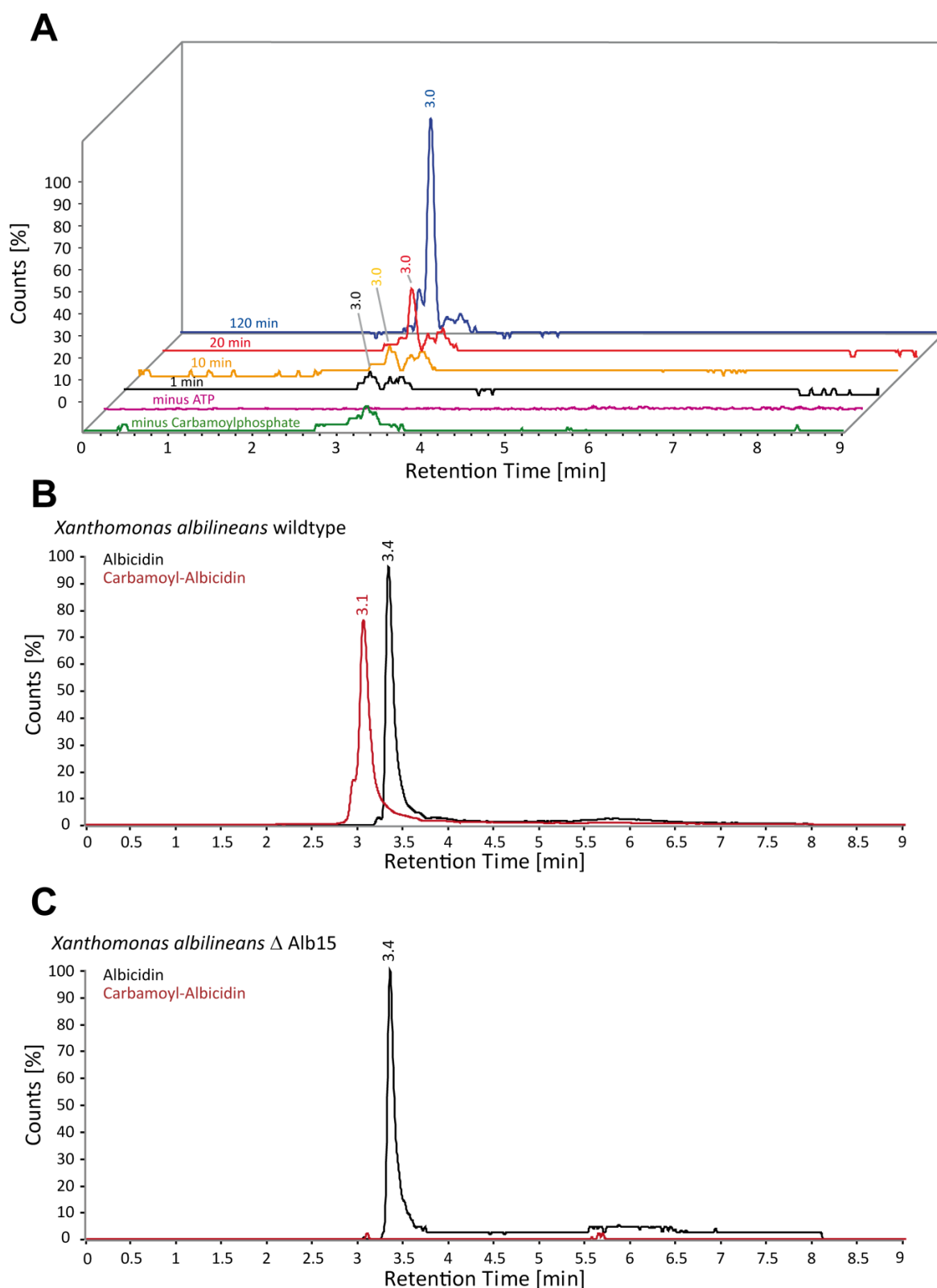
Figure 2.2. (**A**) *In vitro* **carbamoylation of albicidin.** Multiple reaction monitoring (MRM) chromatograms, of carbamoyl-albicidin (886 $m/z \rightarrow$ 703 m/z), at different time points after incubation with Alb15, ATP and carbamoyl-phosphate and negative controls of the *in vitro* conversion assay are shown. (**B**) **MRM chromatogram of XAD extract of *X. albilineans* wild type.** (**C**) **MRM chromatogram of XAD extract of *X. albilineans* Δ*alb15* gene inactivation mutant.** MRM transitions were 843 $m/z \rightarrow$ 660 m/z for albicidin (black) and 886 $m/z \rightarrow$ 703 m/z for carbamoyl-albicidin (red). The presence of carbamoyl-albicidin (R_t = 3.1 min) can only be seen in the wild type whereas albicidin (R_t = 3.4 min) is present in both extracts.

We observed inhibition zones that were approximately 15% larger for carbamoyl-albicidin in comparison to those of albicidin (at 1, 2 and 10 ng/spot). To quantify the antibacterial activity in a more detailed way, we determined the minimal inhibitory concentration (MIC) against two gram-negative (*E. coli* K-12 and *Salmonella typhimurium*) and two gram-positive (*Bacillus subtilis* and *Mycobacterium phlei*) strains (**Appendix Table 2.2**). We observed MICs of ~0.2 ng/µL for albicidin and carbamoyl-albicidin against both gram-positive strains. The MIC of carbamoyl-albicidin against *Salmonella typhimurium* was 3.1 ng/µL compared to 6.3 ng/µL for albicidin, which corresponds to an approximately 50% stronger inhibition. For *E. coli* K-12 we observed MIC values of ~0.1 ng/µL for carbamoyl-albicidin and an approximately 50% better value for albicidin (0.06 ng/µL). In comparison to the apramycin, these MICs were more than one order of magnitude lower. To investigate the influence of the carbamoyl-group on the inhibitory potential of gyrase, we performed *in vitro* gyrase DNA supercoiling assays.
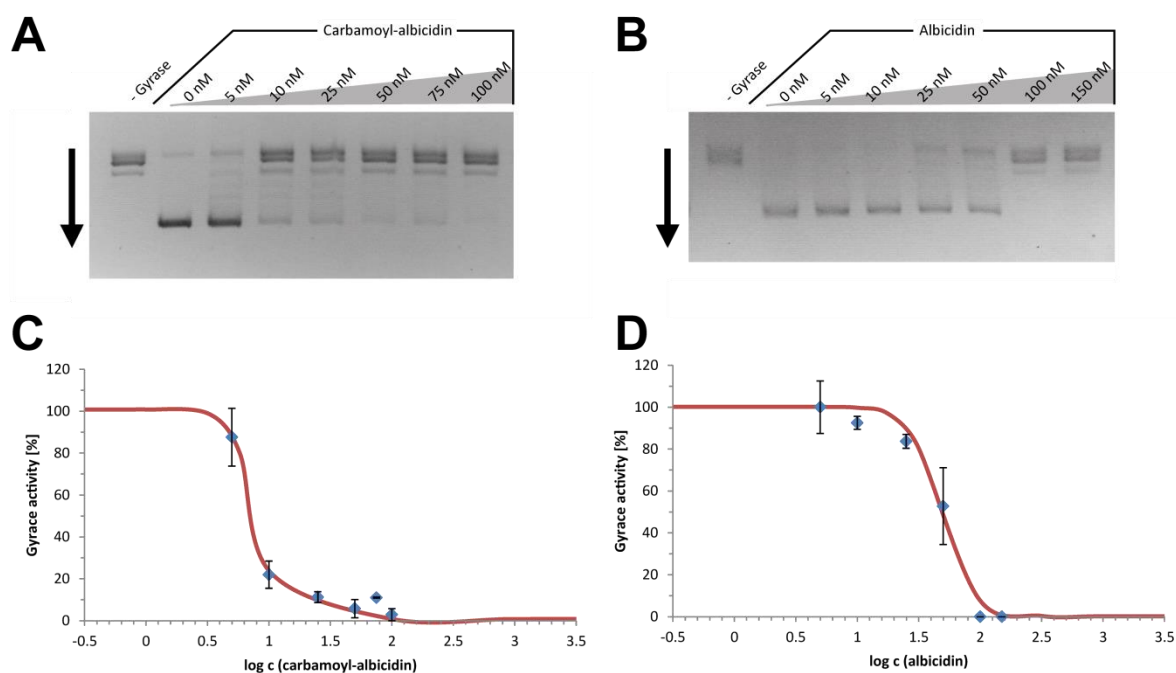
Figure 2.3. (A/B) *In vitro* **determination of the half maximal inhibitory concentration (IC$_{50}$) of carbamoyl-albicidin and of albicidin against bacterial gyrase.** (C/D) **Densitomentric analysis of the gyrase assay shown in A and B.** The intensity of the lower supercoiled lane is plotted against the concentration of carbamoyl-albicidin or albicidin. Error bars represent the standard deviation of triplicates. For the determination of the IC$_{50}$, the density of relaxed DNA at 0 nM carbamoyl-albicidin or albicidin was set as 100% gyrase activity e.g. no inhibition. A regression curve was fitted, based on the following function: f(x)=100/(1+10^((x-LOG(IC$_{50}$))*slope)). IC$_{50}$ values were determined to be ~ 49 nM for albicidin and ~ 8 nm for carbamoyl-albicidin.

The assay is based on different migration of DNA-topoisomers generated in an ATP-dependent manner by gyrase in agarose gels. Hence, we incubated relaxed DNA with bacterial gyrase at different concentrations of carbamoyl-albicidin and albicidin, respectively. As shown in the representative gels in figure 2.3A and B and in the densitometric plots in figures 2.3C and D, no supercoiling occurs in the reaction without gyrase (w/o gyrase) and in a dose-dependent

inhibition (0-100 nM carbamoyl-albicidin) which reaches from inhibition of supercoiling activity (100 nM) to almost complete conversion (0 nM). Through a regression curve and numeric fitting, we calculated the IC_{50} of carbamoyl-albicidin to be ~8 nM, which is nearly a six times higher inhibition compared to the albicidin control of ~ 49 nM, the latter being in a similar range as the results from our previous study.[244]

In summary, it was shown that Alb15 is a carbamoyltransferase that transfers a carbamoyl moiety to the hydroxy-group of the coumaric acid residue at the N-terminus of albicidin, most likely as a post-NRPS reaction. Compared to albicidin, carbamoyl-albicidin exhibits higher inhibition of bacterial gyrase, while the antibacterial effects are strain-dependent. Besides the biological relevance, these findings are important for the deepened understanding of gyrase inhibition by albicidin. Hence, the higher *in vitro* activity strongly suggests that the N-terminal region of albicidin is important for gyrase interaction. Based on these findings, we will further investigate the influence of *N*-terminal modifications in our synthetic studies on albicidin as a potentially new anti-infective drug.

Experimental Section

Fermentation, isolation, and purification of carbamoyl-albicidin

Liquid cultures of a heterologous albicidin-producing strain (strain Xves-alb of *Xanthomonas axonopodis* pv. *vesicatoria*)[243] were prepared in plastic tubes (Corning Inc.). Cultures were grown under agitation (100 rpm) for 5 d at 28 °C in 72 tubes each containing 200 mL (total = 14.4 L) of the optimized XVM3B medium (K_2HPO_4 0.24 mM; KH_2PO_4 0.12 mM; $(NH_4)_2SO_4$ 10 mM; $MgSO_4$*$7H_2O$ 5 mM; casaminoacids 0.015%; $FeSO_4$ 0.01 mM; $CaCl_2$ 1 mM; NaCl 20 mM; glycerol 6 g/L). The isolation of carbamoylated albicidin was adapted from our previous procedure.[149] Briefly, XAD-7 Amberlite was added to the fermentation broth of strain Xves-alb to adsorb albicidin from the supernatant. After eluting the XAD material with 100% MeOH, the fraction was evaporated in a rotary evaporator and re-dissolved in MeOH. After centrifugation, the supernatant was purified by preparative HPLC on an Agilent 1100 system (Agilent) at a detection wavelength of 308 nm on a C_{18} reversed-phase column (GromSil 120 ODS 5 ST, 10 μm; 250 × 20 mm, Grace) using a linear MeOH gradient starting from 35% MeOH plus 0.1% HCOOH to 80% MeOH plus 0.1% HCOOH for 40 min at flow rate of 15 mL/min. The collected fraction was freeze-dried and re-dissolved in aqueous tetrahydrofuran (44%) with acetic acid (1%) and subsequently purified using an analytical Agilent 1200 HPLC system (Agilent) at a detection wavelength of 308 nm, using isocratic conditions (THF 44%/HCOOH 1%; 1 mL/min) on a polymeric reversed-phase (PRP-1, 5 μm; 305 × 7 mm, Hamilton). An additional step of purification was performed with reversed-phase HPLC on an analytical Agilent 1200 HPLC system (Agilent) at a detection wavelength of 308 nm using an Agilent Zorbax RX-C_{18} column (250 x 4.6 mm; 5 μm; Agilent) operated at a flow

rate of 1 mL/min. Elution was performed *via* a gradient of H_2O/ACN/TFA 0.05% from 10 to 80% ACN in 60 min. All fractions were checked by bioactivity test (*E. coli* halo assay) and LC-MS analysis.

Construction of the alb15 deletion mutant of *X. albilineans*

The preparation of the *alb15* deletion mutant in *X. albilineans* was performed according to Rott et al.[250]. All primers used are listed in the appendix (**Appendix Table 2.1**). This method is based on the principle of double recombination. Fragments flanking the *alb15* deletion were amplified from the plasmid pYOAAB23CA09 (Genomic library of strain GPE PC73 of *X. albilineans*[251]) with primers AalbXV/BalbXV and CalbXV/DalbXV, respectively. Resulting PCR fragments were joined at the level of the complemented 24 bp-sequence of primers BalbXV and CalbXV, respectively. The resulting fragment was cloned into Strataclone vector (Stratagene, La Jolla, USA), yielding plasmid pStrata-alb15Xa. Transformants were screened with primers CriblA/CriblB. *Bcl*I insert of pStrata-alb15Xa was then cloned into the pUFR080 (*sacB*) suicide vector[252], digested by *Bam*HI, yielding pUFR080-alb15. Plasmid pUFR080-alb15 was subsequently introduced into *X. albilineans* strain GPE PC73 R5 by electroporation. Transformants were plated on Wilbrink agar plates without sucrose and supplemented by 1% glucose and 20 µg/mL kanamycin. Plates were incubated at 28 °C for five to seven days until isolated colonies appeared, which correspond to mutants in which a first recombination occurred. At this point, to ensure the recombination occurred in the target gene (*alb15*), transformants were screened with primers CriblC/CriblB. Selected transformants were then transferred on classic Wilbrink medium to allow the second recombination to occur. To ensure that the deletion occurred, resulting colonies were screened with primers CriblC/CriblB, CriblA/CriblB and CriblD/CriblB. The PCR product obtained with primers CriblC/CriblB was sequenced using primers CriblC and CriblB.

MS experiments

Full-Scan measurements were routinely performed on an Exactive ESI-Orbitrap-MS (Thermo Fisher Scientific GmbH) coupled to an analytical HPLC 1200 system (Agilent) using a Thermo Hypersil-Gold (5 µm, 100 x 2.1 mm) column with a linear gradient at 0.3 mL/min from 5% B to 100% B (A= water + 0.1% formic acid (HFO), B= acetonitrile + 0.1% HFO) over 6 min followed by a 4 min washout phase at 100% B and a 3 min re-equilibration phase at 5% B.
MS/MS experiments were performed on a triple-quadrupole mass spectrometer coupled to an analytical UHPLC 1290 system (Agilent) using a Grace C_{18} column (3 µm, 50 x 2.1 mm) with a linear gradient from 5% B to 100% B (A= water + 0.1% HFO, B= acetonitrile + 0.1% HFO) over 6 min followed by a 4 min washout phase at 100% B and a 3 min re-equilibration phase at 5% B. For product ion scans, *m/z* 843 and *m/z* 886 were selected as precursors with a unit mass selection window. Normalized collision energy was set to 8%.

For the detection of albicidin and carbamoyl-albicidin, the same triple-quadrupole mass spectrometer with identical chromatographic conditions was used. Detection was performed through multiple reaction monitoring (MRM) mass spectrometry using the transitions m/z 843 → m/z 660 and m/z 886 → m/z 703 as typical product ions for albicidin and carbamoyl-albicidin, respectively. Normalized collision energy was set as well to 8%.

Cloning, expression, and purification of Alb15

For detailed procedure, see appendix. In short, the *alb15* gene was amplified by PCR using the cosmid pALB540[239] as template and cloned into pETtrx_1c and subsequently transformed into *E. coli* BL21-Gold. Protein expression was carried out in Terrific Broth medium at 37 °C and 200 rpm for 2 h followed by 16 h at 18 °C and 180 rpm. After lysis, the protein was purified by Ni-affinity chromatography and size exclusion chromatography using an Äkta purification system (GE Healthcare). The identity of the Alb15 fusion protein was verified by SDS-PAGE and in-gel trypsin digestion (**Appendix Figure 2.3 and 2.4**). Finally, the protein solution was concentrated to a final concentration of 3.3 mg/mL in 20% glycerin/phosphate buffer, shock frozen in liquid N_2 and stored at -20 °C to -80 °C.

In vitro carbamoylation assay

A 25-µL reaction mix containing 4 mM DTT, 400 µM $MnCl_2$, 4 mM $MgCl_2$, 40 mM Carbamoyl-phosphate and 20 mM ATP) was mixed with 135 µL HEPES buffer (100 mM; pH 6.7) and 5 µL albicidin (2.0 mg/mL). After addition of 30 µL Alb15 (3.3 mg/mL), samples were incubated at 30° C and reaction was stopped at different time points (1 min, 10 min, 20 min and 120 min) by adding 20 µL formic acid (100%). Carbamoylated albicidin was extracted with 300 µL ethyl acetate. The organic supernatant was removed and evaporated in a vacuum centrifugation (Thermo Scientific), redissolved in aqueous ACN (20% + 1% HFO) and measured *via* MRM.

Total synthesis of carbamoyl-albicidin

The synthesis of carbamoyl-albicidin was performed according to Kretz et al.[244]. In short, we modified (detailed description can be found in the appendix) the synthesis scheme by making use of the tert-butyldimethylsilyl (TBS) protecting group for the *para* hydroxy group of the cinnamic acid. After coupling of the pentapeptide to the cinnamic acid moiety according to Kretz et al.[244] and cleavage of the TBS ether with tetra-n-butylammonium fluoride (TBAF), the carbamoyl moiety was introduced by using chlorosulfonyl isocyanate (CSI) reagent, subsequently followed by the final de-protection of the allyl protection groups.

In vivo bioactivity

Halo assays were performed on 1.5% LB-agar plates (15 mL) with 0.75% LB top-agar (4 mL) containing 20 µL of an overnight culture of *E. coli* DH5α. Albicidin and carbamoyl-albicidin were applied, dissolved in DMSO to yield total amounts of 0.1, 1, 2 and 10 ng/spot. Inhibition zones were measured after overnight incubation. The assay was performed in duplicates.

The minimal inhibitory concentration was determined according to the Clinical and Laboratory Standards Institute, M31-A2[253]: In short, MHB Medium was inoculated with overnight cultures of *Bacillus subtilis*, *Mycobacterium phlei*, *E. coli* K-12 *BW25113* and *Salmonella typhimurium* and aliquoted with 200 µL/well into sterile flat bottom microtiter plates. After the addition of the albicidin, carbamoyl-albicidin or apramycin dilution series, the plates were incubated overnight at 37 °C without shaking. Finally, the optical density (OD) at 625 nm of each well was analyzed using an Infinite 200 plate reader (Tecan).

Gyrase activity

Gyrase supercoiling experiments were performed in a total volume of 20 µL gyrase buffer (protocol by NEB, Frankfurt, Germany). The incubations contained 60 ng relaxed pUC19 plasmid DNA (NEB, Frankfurt, Germany), 1 unit DNA-gyrase (2.2 nM) (NEB) and various concentrations of carbamoyl-albicidin (5-100 nM). Samples were incubated at 37 °C for 45 min and subsequently heated at 65 °C for 15 min in order to inactivate the gyrase. Electrophoretic analysis was performed on a 1% agarose gel. Staining of bands was performed with ethidium bromide. For the determination of the IC_{50}, the gels were photographed and densitometrically analyzed with ImageJ (National Institutes of Health). The peak area of the densitometric analysis of the control without inhibitor was set as 100% enzyme activity. IC_{50} values were calculated through numeric regression using solver (Microsoft) based on the logarithmic equation: $f(x)=100/(1+10\char94((x-LOG(IC_{50}))*slope))$.

3 The Biosynthesis of Albicidin Involves On-line Enzymatic Methylation During Nonribosomal Peptide Assembly

Introduction

Nonribosomal peptide synthetases (NRPSs) are complex and ribosomally independent machineries, responsible for the biosynthesis of many important, peptide-derived natural secondary metabolites. NRPS assembly lines are subject to a repetitive set of domains, which are organised into modules for specific amino acid incorporation. The multimodular protein architecture of NRPS assembly lines combined with the access of a versatile pool of homologous amino acid monomers allows for great variation in stereochemistry, peptide length, cyclisation state and other post-NRPS modifications.[254–256]

A central element of NRPS-mediated biosynthesis is the peptidyl carrier protein (PCP) or thiolation (T) domain, to which all intermediates are bound after initial monomer activation. The peptidyl carrier domains shuttle activated substrates both, upstream and downstream, to catalytic centers within modules along the NRPS assembly line.[255,256] The inactive *apo*-PCP domain (> 100 aa, ~10 kDa) adopts a four-helix bundle, harboring a highly conserved serine residue (GxxS core motif) located at the N-terminus of helix α2, which is transformed into the active *holo*-PCP form by posttranslational attachment of a phosphopantetheinyl (Ppant) arm by a phosphopantetheine transferase (PPTase).[256,257] As a functional prerequisite, the extended arm plays an essential role for conformational flexibility and allows for a temporary storage of versatile substrate monomers and peptidyl intermediates on the PCPs (*holo**-PCP) in the form of a reactive aminoacyl-/peptidyl thioester.[255,256] This thiotemplate mechanism prevents loss of activated substrates and ensures spatial and temporal control during the assembly process, defining the choice of building blocks, order of incorporation, degree of processing and the size of the chains.[254,255,258]

Furthermore, *holo**-PCPs serve as a platform to install further structural diversity of chached substrate intermediates on the assembly line, acronymically named 'on-line', and thus mediate for recruitment of diverse *cis*- and *trans*-acting enzymes by transient interaction interfaces.[254,256,258] With regard to the adjacent *cis*-acting domains in the NRPS assembly line, it is clearly proven that affiliated PCPs do not undergoing major conformational changes and retain the orientation of the helix bundles during substrate shuttling from one partner domain to another.[256] The slight structural reorientations of the rigid-folded PCP domains, irrespective of the loading state, are rather caused due to the different catalytic states along with altered reconfigurations of the

neighbouring modification domains.[255,256] In a typical elongation module, central PCP domains carry out a fine-tuned crosstalk in interplay with three major interaction partners, the intramodular adenylation (A) domain as well as the acceptor site of an upstream and an acceptor-donor site of a downstream condensation (C) domain (**Figure 3.1A**).[255,259] Taking into account the considerable contribution of the prosthetic group and its decisive interaction in the respective catalytic funnel, specific PCP interfaces are located in helix α2 and α3 via hydrophobic as well as ionic interactions (**Figure 3.1B**). In combination with the additional interface for the preceding loop α1' that forms a network of charged interactions, these interaction sites are all in close proximity to the highly conserved Ppant attachment site (**Figure 3.1B**).[260,261,262] In addition, linker sequences between the C-terminale A_{sub} domains and downstream cognate PCPs demonstrated conserved motifs that being important for the relative orientation of both domains and the catalytic efficiency of the respective A domain.[263] Further optional *cis*-acting domains in more advanced NRPS modules, including cyclization (Cy)[264], epimerization (E)[265], and thioesterase (Te)[266,267] domains or modification domains, such as formylation (F)[261] domains, all in close proximity to the intramodular PCP domain, highlight identical structural interfaces.[255] Interestingly, a hydrophobic cleft between helices α2 and α3 also permits a transient resting state of the substrate-loaded Ppant arm for cognate substrate storage.[255,256]

While *cis*-mediated interactions are an integral element within an assembly line and different structural biology studies allow deep mechanistic insights, little is known about the recruitment and interaction between NRPS modules and *trans*-active modification enzymes so far. Few examples are reported for *trans*-acting enzymes in interplay with PCP-bounded substrate monomers, such as transferases[268], cyclases[269], monooxygenases[270] or halogenases[271] enzymes, allowing the introduction of versatile chemical traits. However, limited structural evidence is given to explain fundamental mechanisms that ensure selective recruitment of the t*rans*-modifying enzymes solely with the correct carrier protein domains. Exceptions to this are the structural investigations of biosynthetic machineries both, producing glycopeptide antibiotics (GPAs) as well as the cyclic depsipeptide skyllamycin, whose interaction mechanism between *holo**-PCPs and *trans*-acting cytochrome P450 monooxygenase enzymes (P450s) have been extensively characterized.[256] Interestingly, interfaces in the structural complex of P450$_{sky}$ and the respective *holo**-PCP domain are located in helix α3 as well as the loop α1' between helices α1 and α2 by hydrophobic interactions (**Figure 3.1B**). In addition, the complex is further stabilized by a network of hydrogen bonds within the P450$_{sky}$ and the Ppant arm of the corresponding PCP.[272] Another *in trans* modification, reliant on the recruitment of several P450 enzymes has been identified in the GPA biosynthesis. The system performs sequential oxidative cyclisation to generate rigid, bioactive aglycones from the linear heptapeptide, which is bound to the terminal PCP domain.[273,274] The sequential cyclisation cascade requires a separate NRPS domain, known as X domain, that ensure of separate recruitment for consecutive trans-modifying P450 enzymes

(also known as Oxy enzymes). Further, crystal structures indicate that the X domain act as a binding platform for the Oxy enzymes and that the PCP-bound heptapeptide alone is not always sufficient to generate a competent substrate. In addition, mutation experiments and model structures of the three-domain complex revealed that the loop region after helix α_4 of the final PCP domain is essential for the interaction.[275]
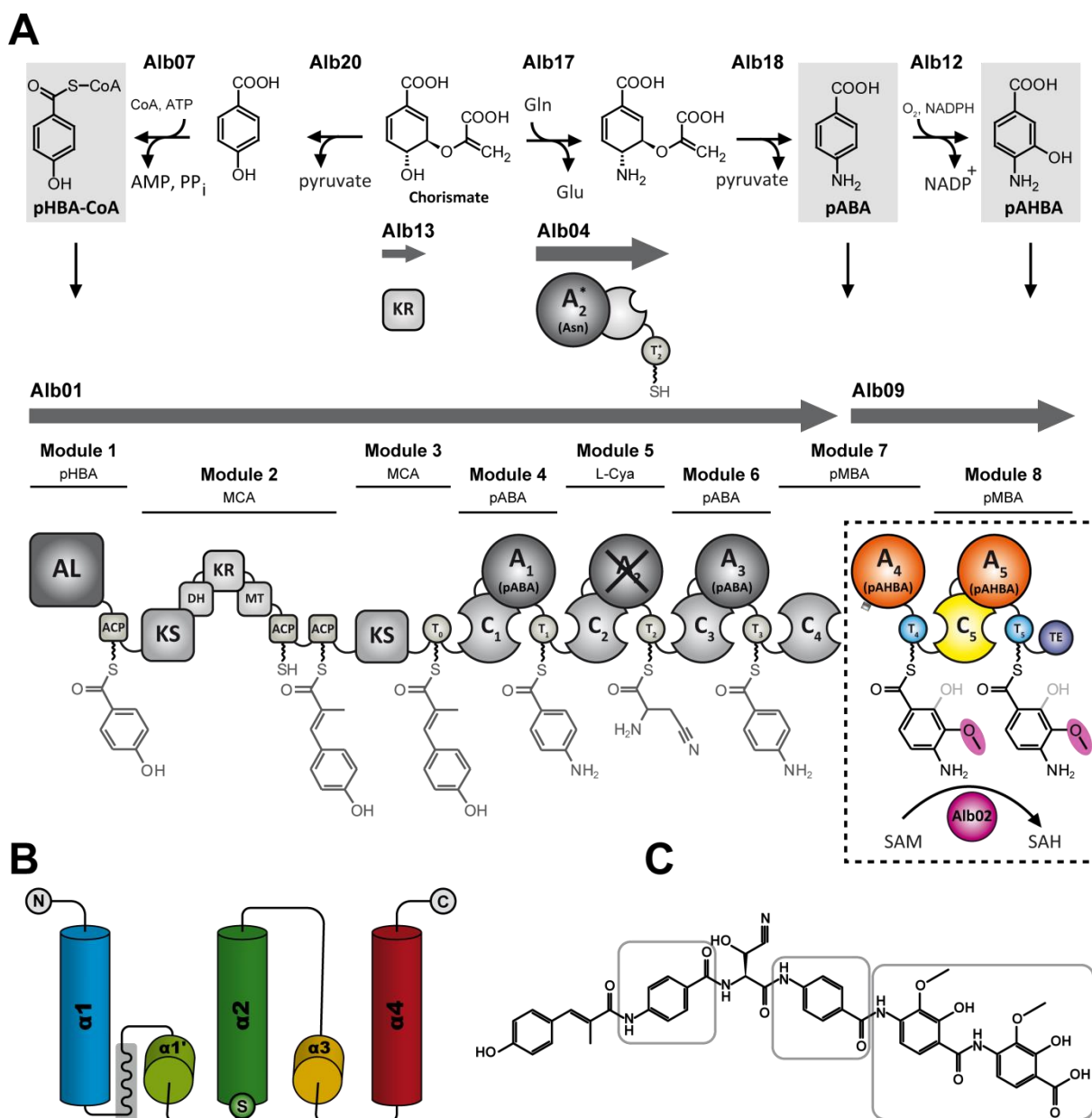
Figure 3.1. Albicidin biosynthesis assembly line, topology of PCPs and molecular structure of albicidin. (**A**) Schematic polyketide synthetase-nonribosomal peptide synthetase (PKS-NRPS) assembly line including substrate synthesis. In dashed box highlighted is our proposed on-line modification hypothesis. (**B**) Topology model of PCP domains showing rigid four-helix bundle and conserved serine residue for phosphopantetheinyl attachment. Highlighted in the grey box is the position of the inserted loop for PCP-4 and PCP-5 of the albicidin NRPS assembly line. (**C**) Molecular structure of albicidin highlighting different pABA analoga.

However, other in-*trans* modification systems may be governed by additional interaction features, and the current lack of study of these interactions hampers future efforts for combinatorial biosynthesis.

The biosynthesis of albicidin, the phytotoxin of the xylem-invading agent *Xanthomonas albilineans*, is assembled by a polyketide synthase-nonribosomal peptide synthetase (PKS-NRPS) hybrid (**Figure 3.1A**).[149,239,243] The exceptional building blocks for the unique structure of albicidin have been recently reported based on the previous structure elucidation and substrate specificity of the adenylation domains (A-domain) (**Figure 3.1A**).[149] It was shown that final NRPS modules, NRPS-4 and NRPS-5 (Alb09), activate *para*-amino-3-hydroxybenzoic acid (pABA-3OH), although the structure of albicidin comprises two *para*-amino-3-hydroxy-4-methoxybenzoic acid (*p*MBA) residues (**Figure 3.1C**). Phylogenetic analysis of the albicidin biosynthesis gene cluster (*alb* cluster) revealed gene *alb02* as a putative methyltransferase belonging to the family of *S*-adenosylmethionine (SAM)-dependent *O*-methyltransferases, which we postulate is responsible for an in-*trans* modification to the PCP-bound substrate.[149] In addition, with respect to known structural investigations on *trans*-acting modification systems and the involvement of distinct PCP interfaces for respective interactions, we performed an *in silico* analysis to search for structural features of PCP domains in modules NRPS-4 and NRPS-5. Interestingly, in homology models we found a highly conserved insertion between helices α1 and α1' that form an extended loop region (**Figure 3.1B**). In large-scale sequence alignments, we solely found PCPs including a similar insertion in the biosynthesis gene cluster that belong both, cystobactamide[150] and coralmycin[151], to the same novel molecular class of gyrase inhibitors. Therefore, we propose that the insertion for a PCP-mediated in-*trans* modification of the pABA building blocks is of crucial importance.

Herein, we report on the *in vitro* characterisation of the SAM-dependent methyltransferase (MTase) Alb02 and its interaction with artificially loaded partner PCP domains (*crypto*-PCP) using mass spectrometry (MS) as well as nuclear magnetic resonance (NMR) spectroscopy. Initial MS analysis allowed us to show substrate recognition of the MTase Alb02 and track *in trans* modification of substrates bound to the cognate PCP. Cross-linking experiments allowed us to demonstrate interactions between MTase Alb02 and different PCP domains. Moreover, employing *apo*-/*holo*- and *crypto*-PCP forms we were able to monitor their interactions with Alb02 in NMR titration experiments. The results highlight that whilst both, cognate and non-cognate, PCP domains are subject to a basal level of transient interactions with Alb02, the tethered substrate on the *crypto*-PCP domain efficiently triggers Alb02 recruitment. Our findings thus give insights how the substrate flow in NRPS assembly is regulated and support the notion that the substrate itself plays an important role in this orchestration.

Results and Discussion

In vitro characterization of putative *O*-methyltransferase Alb02

Previous experimental substrate specificity data implied that NRPS-4 and NRPS-5 (Alb09) preferentially activate the substrate pABA-3OH rather than pMBA. In order to confirm the molecular function of Alb02 as an *O*-methyltransferase, we performed a functional assay using several related pABA analogs (**Figure 3.2A**). The substrates were incubated with Alb02 and isotope-labeled ^{14}C-SAM, such that we expected a transfer of the ^{14}CH$_3$ group of the cofactor only to pABA analogs tolerated by Alb02. None of the free-acid analogs was converted to the corresponding methylated product (**Figure 3.2A**), which was in agreement with previous substrate-selectivity experiments of respective A domains.[149] Based on these results, we excluded the possibility of pre-NRPS modification of building blocks, but instead considered on-line enzymatic tailoring through Alb02 acting in *trans* on the reactive thioester tethered to PCP-4 and PCP-5, respectively. To test our hypothesis, we chemically synthesized the corresponding *N*-acetylcysteamine thioester (SNAC) derivatives, which imitate part of the Ppant of *crypto*-PCP (**Figure 3.2A**).
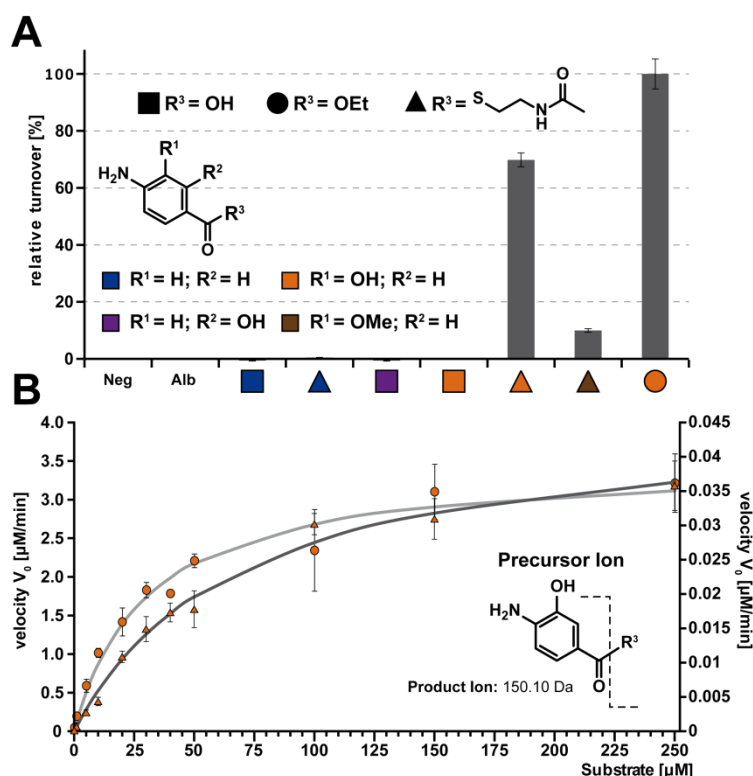
Figure 3.2. *In vitro* characterization of methyltransferase (MTase) Alb02. (**A**) Substrate specificity screening of MTase Alb02 for highly related pABA derivatives both, as free amino acids and *N*-acetylcysteamine (SNAC) amino acid thioesters. Negative control (Neg) was performed without substrate addition and control experiment with albicidin (Alb) to exclude post-NRPS modification. (**B**) Substrate-dependent kinetics of MTase is carried out by multiple reaction monitoring (MRM) in presence of rising concentrations of pABA-3OH-SNAC (light grey line, orange triangle, left y axis) or pABA-3OH-OEt (dark grey line, orange circle, right y axis). All experiments were performed in triplicates.

The subsequent analysis of the pantetheinyl thioester showed corresponding specificity and confirmed a significant turnover (70%) for pABA-3OH-SNAC, in contrast to the corresponding free acid. Notably, the pABA-3OH ethyl ester derivative was also well-accepted by Alb02 with the highest relative conversion (set to 100%). The substrate screening demonstrated the importance of a masked and nonpolar C-terminal extension group for substrate recognition and modification of pABA-3OH in the active site of Alb02. The observed substrate specificity of Alb02 with respect to the ethyl ester and SNAC derivatives of pABA-3OH was further studied through Michaelis-Menten kinetic analysis for a single-substrate reaction.[276,277] We performed a concentration-dependent tandem MS/MS turnover experiment, by screening the isolated product ion (150.10 Da) intensity from the corresponding precursor ion (ethyl ester: 195.09 Da, SNAC: 296.10 Da) of the enzyme-catalyzed methylation of pABA-3OH derivatives, to compare steady-state kinetic parameters, like the maximal velocity (v_{max}) and the Michaelis-Menten constant (K_M), but also fundamental and enzyme-specific constants like the turnover number (k_{cat}) and the specificity constant (k_{cat}/K_M). A direct comparison of both derivatives showed significantly faster maximal velocity for the pABA-3OH-SNAC (v_{max} = 4.1 μM/min) in contrast to the pABA-3OH ethyl ester (v_{max} = 0.04 μM/min) (**Figure 3.2B**). On the other hand, the Michaelis-Menten constant for the SNAC derivative (K_M = 68 μM) was more than two times higher than that of the ethyl ester derivative (K_M = 31 μM). In order to show an enzyme-specific comparability between both substrates, we calculated the apparent unimolecular rate, that denotes the maximum number of enzymatic reactions catalysed per minute, and the catalytic efficiency, that describe the conversion efficiency of an enzyme.[278] The turnover number of MTase Alb02 for pABA-3OH-SNAC (k_{cat} = 0.205 min^{-1}) showed a significant higher catalytic rate of MTase Alb02 in contrast to the pABA-3OH ethyl ester (k_{cat} = 0.002 min^{-1}). In addition, the catalytic efficiency (SNAC: k_{cat}/K_M = 3.01×10^3 min^{-1}M^{-1}; OEt: k_{cat}/K_M = 4.51 min^{-1}M^{-1}) supported our hypothesis of a PCP-anchored substrate and thus PCP-directed recruitment of Alb02 to the NRPS assembly line.

On-line enzymatic modification

In silico analysis of PCP domains in the BGC of albicidin revealed an exclusive sequence insertion (TPAQAAPL) between helices α1 and α1' in PCP-4 and PCP-5, which could not be found in PCP-1 and PCP-3, nor in other PCP domains with the exception of a similar insertion found in PCP domains of the closely related cystobactamide and coralmycin cluster (**Appendix Figure 3.1**).[150,151] Previously, the unique region was already considered as a possible mediator for enzyme recruitment and thus *in trans* modification.[149] To prove the potential function of this insertion, we performed control experiments with a structural highly related peptidyl carrier domain (PCP-1) in the albicidin NRPS biosynthesis line lack of the exclusive insertion.

The intact MS analysis was performed in a two-step experiment and was carried out to support our previous functional analysis as well as proof our hypothesis of an on-line enzymatic

modification. Therefore, we employed the PPTase Sfp and respective PCP domains were converted in a first step from *apo*-PCPs to *crypto*-PCPs by loading with synthetic *p*ABA-*S*-CoA (**Appendix Figure 3.2**) and *p*ABA-3OH-*S*-CoA (**Figure 3.3**).
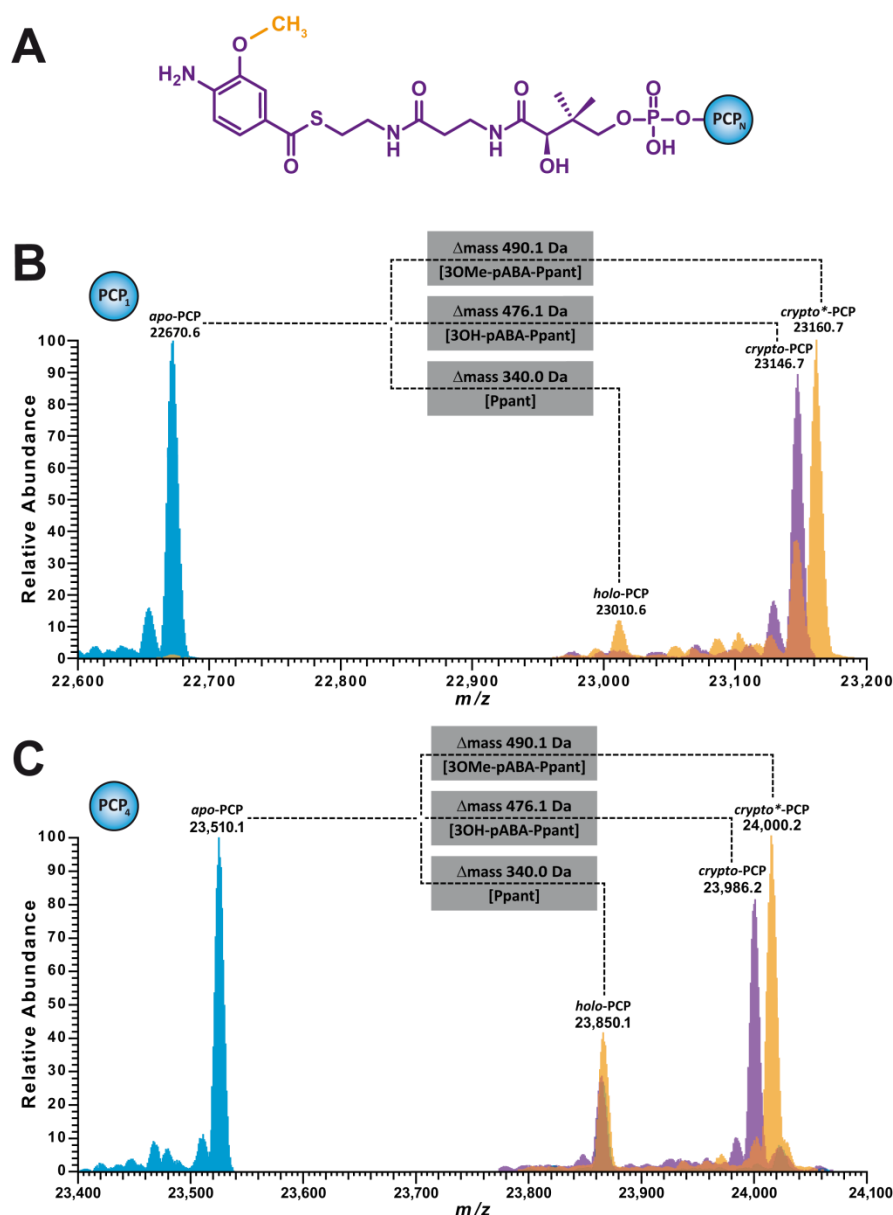
Figure 3.3. On-line modification of *crypto*-PCPs monitored by intact mass spectrometry. (**A**) Schematic overview of different loading/modification steps of respective PCP domains with pABA-3OH. Loading and on-line modification of (**B**) PCP-1 and (**C**) PCP-4. Please note that above presented spectra are overlaid by three different spectra. *Blue*: inactive *apo*-PCP domain state. *Purple*: active *crypto*-PCP domain state including the Ppant arm loaded with pABA-3OH. *Orange*: active *crypto*-PCP domain incubated with MTase Alb02 and SAM and subsequent end point *crypto**-PCP detection. Single spectra show presence of pre-activated *holo*-PCP state caused by PPTase during recombinant expression.

Subsequently, *crypto*-PCP domains were incubated in presence of Alb02 and cofactor SAM to verify on-line methylation of the incorporated derivatives. The two steps were measured separately via top-down LC-MS for both, PCP-1 and PCP-4 domain. The first deconvoluted mass spectra show the native *apo* form of the respective PCP domains (PCP-1: 22,670.6 Da; PCP-4:

23,510.1 Da). The second, layered mass spectra (purple) of the respective PCP domains (PCP-1: 23146.7 Da; PCP-4: 23986.2 Da) show a mass shift of +476.1 Da, which indicates a conversion from the *apo*- to the *crypto*-PCP domain including the thioester-bound pABA-3OH. After incubation with cofactor SAM and Alb02, the third set of mass spectra (orange) show a mass shift of +14.00 Da for both, PCP-1 and PCP-4 domain (PCP-1: 23,160.7 Da; PCP-4: 24,000.2 Da), which in turn revealed methylation of the 3-hydroxy group to pABA-3-methoxy benzoic acid (pABA-3OMe). Additionally we observed deconvoluted mass peaks corresponding to *holo*-PCPs (PCP-1: 23,010.6 Da; PCP-4: 23,850.1 Da) in all spectra caused by the natural PPTase in *Escherichia coli* during recombinant expression.

Furthermore, in our control experiment with the structurally related building block pABA, we did not observed any modifications at the *crypto*-PCP domains (**Appendix Figure 3.2**). Importantly, conversion analysis of our tested *crypto*-PCP domains demonstrated an almost quantitative turnover for both, cognate PCP-4 as well as non-cognate PCP-1. Therefore, we concluded that the inserted region played no significant role for the recruitment of Alb02 and subsequent substrate conversion. Nevertheless, our results support an *in situ* enzymatic tailoring probably controlled by a substrate-driven screening. In our model, the highly-specific *trans*-acting enzyme Alb02 rapidly scans tethered substrates bound to peptidyl carrier domains and modifies only correctly recognized substrate. In the following, we confirm this statement of a substrate-controlled screening and investigated potential interactions by both crosslinking experiments and chemical shift NMR assays.

Investigation of Alb02-PCP interactions

Initial gel filtration experiments injecting methyltransferase Alb02 in presence of different concentrations of *crypto*-PCP domains did not show any co-elution (**Appendix Figure 3.3**). Gel filtration is a basic technique to monitor functional and stable protein PCP complexes with dissociation constants up to the low micromolar range. This observation is in support of very transient interaction events between MTase Alb02 and cognate PCP domains.

To confirm our hypothesis, we covalently trapped the interaction between the methyltransferase Alb02 and *apo/holo*-PCP domains by chemical fixation and evaluated protein complexes by SDS-PAGE (**Appendix Figure 3.4**). Therefore, we performed crosslinking experiments using disuccinimidyl sulfoxide (DSSO), an amine-reactive *N*-hydroxysuccinimide (NHS) ester that chemically reacts with amine and hydroxy side chains. Starting with separate control experiments, we observed for single Alb02 crosslinking experiments a clear appearance of a methyltransferase dimer complex under native conditions that was additionally confirmed by gel filtration experiments and bioinformatic analysis of the N-terminal dimerization domain (**Appendix Figure 3.3** & **3.5**). In the following *apo*-PCPs were incubated in presence of various equivalent concentrations of Alb02. It could be observed that already in equimolar amounts an

additional protein complex in the mass range of an *apo*-PCP/Alb02 complex (∼50 kDa) appears. The individual protein constituents of this protein association were confirmed by in-gel digestion and subsequent peptide mass fingerprinting. It showed the expected protein complex formed by *apo*-PCP and Alb02. No distinction in the complex formation between both *apo*-PCPs for none of the equimolar ratios could be observed, which is in favour of a substrate-controlled Alb02 acquisition and transient interaction.

To further probe the potential transient nature of interactions between Alb02 and PCPs, we decided to employ NMR spectroscopy as a powerful technique for characterizing weak interactions at atomic resolution. To this end, we first established an almost complete assignment of backbone resonances of ^{13}C- and ^{15}N-labeled PCP-4 (**Appendix Table 3.2**). We could not detect or resolve the amide signals of Glu10, Ser46 and Leu47, with the latter two representing the Ppant attachment site (**Figure 3.4**).
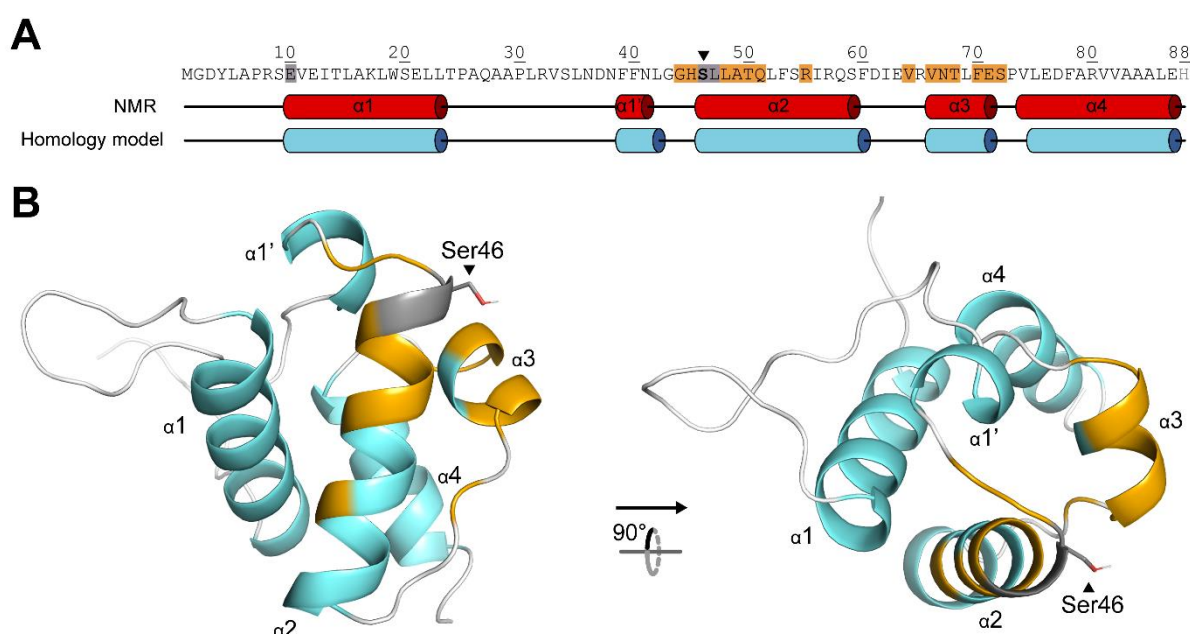
Figure 3.4. NMR analysis of PCP-4. (**A**) The sequence of PCP-4 is shown with unassigned residues and residues with doubled resonances being highlighted in grey and orange, respectively. Except His88 (light grey), all residues from the C-terminal His-tag as well as two N-terminal residues after TEV-cleavage have been omitted for clarity. Cylinders below the sequence represent the position of α-helices obtained either experimentally from NMR chemical shift analysis (red) or through in-*silico* structural homology modelling (blue). (**B**) Structure homology model of PCP-4 with α-helices in cyan and unstructured regions in white. Residues highlighted in panel A are mapped onto the model using the same color code. The Ppant attachment site Ser46 is indicated.

The finger-print ^{1}H-^{15}N HSQC spectrum of PCP-4 indicated a well-folded protein (**Appendix Figure 3.6**) and the chemical shift analysis allowed for the unambiguous identification of an predominantly α-helical structure, which was in very good agreement with previous structures of PCP domains and thus our initial structure homology model (**Figure 3.4**). Importantly, the aforementioned sequence insertion in PCP-4 was classified as dynamic with no tendency to adopt

a specific secondary structure under the measured conditions. Moreover, we observed a second resonance set for several amide signals arising from residues located in helix $\alpha 2$ and $\alpha 3$ in the vicinity of the Ppant attachment site Ser46 (**Figure 3.4**). We ascribed these doubled resonance sets to the presence of both *apo-* and *holo-*PCP-4 in the NMR sample, which we also confirmed by MS analysis (**Figure 3.3** & **Appendix Figure 3.2**). Given the high structural conservation of PCP domains and the very good agreement with our chemical shift data, we used the homology model as an eligible basis for further interaction studies.

We next monitored changes in ^{1}H-^{15}N HSQC spectra of *apo-/holo-*PCP-4 caused by the presence of increasing concentrations of unlabeled Alb02. In agreement with the molecular size of Alb02 (39.1 kDa) and a rather weak binding affinity we observed a gradual attenuation of cross-peak intensities. However, these effects were not homogenous for the entire protein, but strongly residue-dependent. Mapping the strongest effects at ten-fold molar excess of Alb02 onto the structural model of PCP-4 demonstrated that mostly residues of helix $\alpha 2$ and $\alpha 3$, e.g. Leu48, Ala49, Thr50, Leu52, Ser54, Ile56, Arg57, Val66 and Leu69, are involved in binding the partner enzyme Alb02 (**Figure 3.5**). Such binding interface is consistent with the vicinal position of the Ppant anchor as well as with other structural studies that identified helices $\alpha 2$ and $\alpha 3$ as the main contact surface to various interaction partners during NRP biosynthesis.[256] Notably, we observed the strongest attenuation for one of the amide signals of the Ppant moiety itself (NH41, denoted as X), which is expected to penetrate the substrate tunnel of Alb02 (**Figure 3.5A**).

By contrast, we could not detect major binding effects for the region of PCP-4 comprising its unique sequence insertion, which is located at the opposite side from the $\alpha 2/\alpha 3$ surface (**Figure 3.5B**). Non-linear fits of titration curves yielded dissociation constants K_d in the range of 100-900 μM, with the lowest K_d values found for residues of the identified $\alpha 2/\alpha 3$ interface (**Figure 3.5C**).
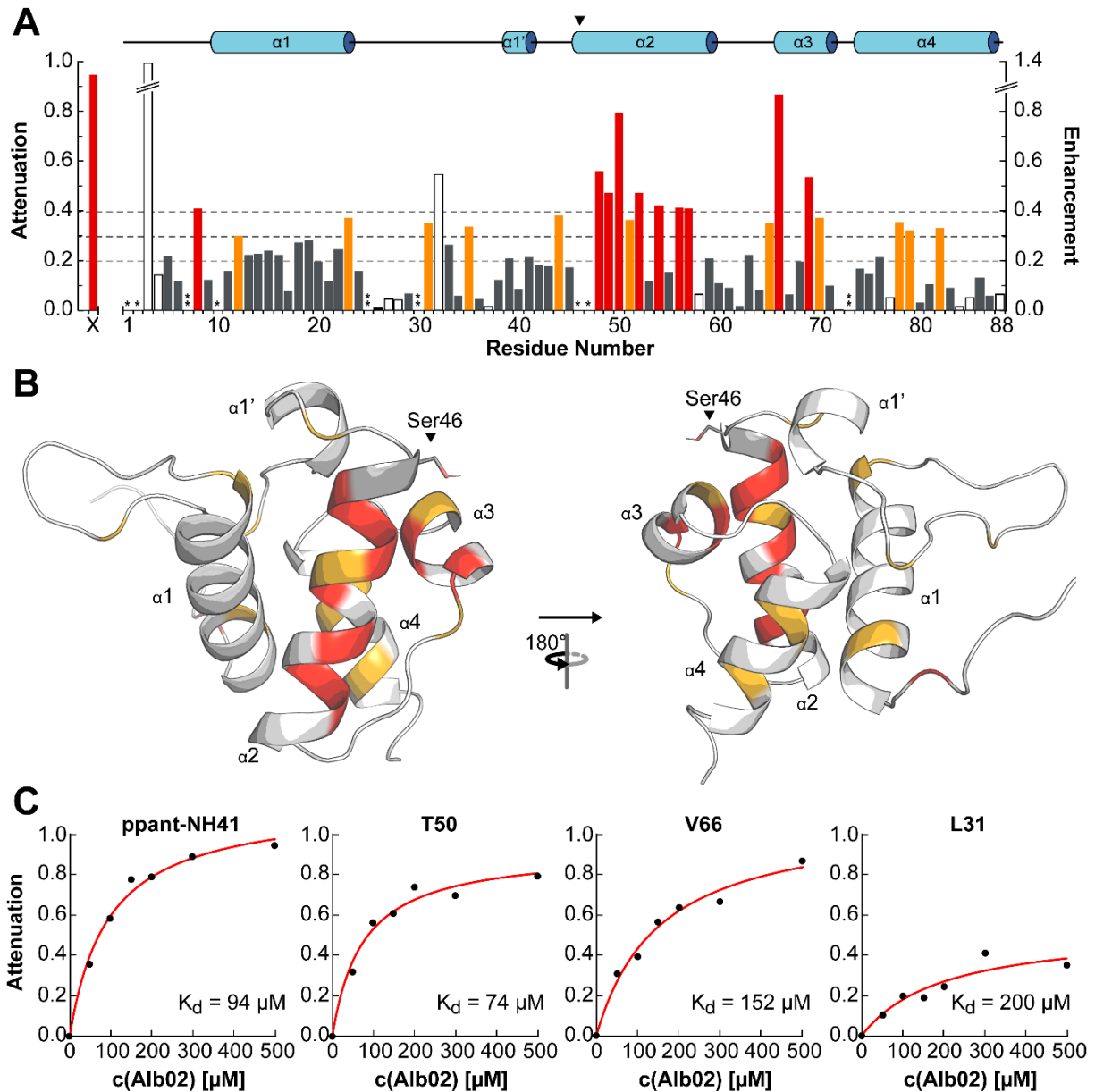
Figure 3.5. Interaction studies of *apo-/holo*-PCP-4 and MTase Alb02. (**A**) Attenuation of cross-peaks in ^{1}H-^{15}N HSQC spectra of ^{15}N-*apo-/holo*-PCP-4 (50 μM) in the presence of unlabeled Alb02 (500 μM). Dashed lines indicate the average attenuation (0.2) as well as arbitrarily chosen limits for medium (orange bars, attenuation $\geq$ 0.3) and strong effects (red bars, attenuation $\geq$ 0.4). Some cross-peaks (white bars) showed negative attenuations (probably due to dynamic effects), which are given as signal enhancements on the secondary axis (*right*). The signal attenuation of NH41 of Ppant in *holo*-PCP-4 (denoted as X) is shown on the left. Helical elements are illustrated on top. Asterisks denote unassigned residues (*) and Pro residues (**), respectively. (**B**) Structural model of PCP-4 (white) with Alb02-binding effects mapped using the same color code as in (A). (**C**) Non-linear fits (red lines) of titration data points (black circles) and corresponding K_d values for representative residues.

Hence, our NMR experiments confirmed a weak binding affinity between *apo-/holo*-PCP-4 and Alb02, however, we cannot clearly differentiate binding affinities for the *apo-* and *holo*-states. We hypothesized that loading of the cognate substrate of Alb02 onto the Ppant of PCP-4 would strengthen their interaction. To this end, we employed the surfactin phosphopantetheinyl

transferase (Sfp) from *Bacillus subtilis* and synthetic pABA-3OH-*S*-CoA for priming of ^{15}N-PCP-4 to yield the corresponding [15]N-*crypto*-PCP-4. It should be noted, that since our recombinant ^{15}N-PCP-4 samples contained both, *apo-* and *holo*-PCP-4, only the *apo* fraction was loaded with pABA-3OH-*S*-Ppant. We recorded ^{1}H-^{15}N HSQC spectra of [15]N-*holo-/crypto*-PCP-4 in the absence and presence of ten-fold molar excess of Alb02 and observed severe signal attenuation of about 70%. Importantly, this time the decrease in signal intensities was fairly homogenous when plotted against the protein sequence (**Figure 3.6**), which is indicative of a more stable Alb02-PCP-4 complex that is too large to be detected. We can thus deduce that the actual substrate of Alb02 increases the binding affinity between the two proteins.
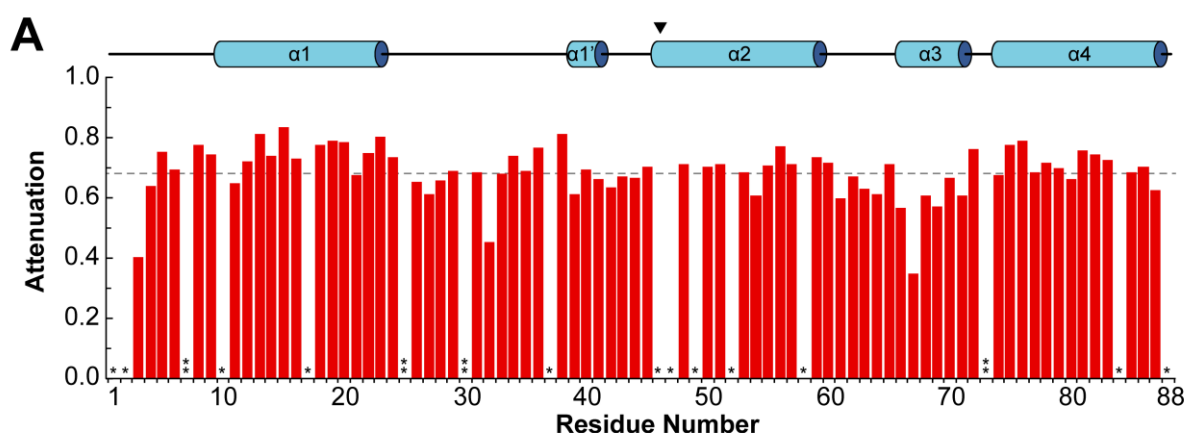
Figure 3.6. Interaction studies of *holo-/crypto*-PCP-4 and Alb02. A) Attenuation of cross-peaks in ^{1}H-^{15}N HSQC spectra of [15]N-*holo-/crypto*-PCP-4 (50 µM) in the presence of unlabeled Alb02 (500 µM). The dashed line indicates the average attenuation (0.68). Helical elements are illustrated on top. Asterisks denote unassigned residues as well as signal overlap (*) and Pro residues (**), respectively.

Significance

In conclusion, our results show that there is basal affinity between a PCP domain and its *trans*-acting partner enzyme, which appears to scan for its cognate binding partner. The loading of the cognate substrate onto the *holo*-PCP, i.e. activation and loading of pABA-3OH by the corresponding A domain, triggers the interaction with Alb02 for efficient enzymatic conversion. Previous investigations on the interplay of *trans*-standing enzymes and PCP-bound substrate monomers are rarely and structural evidence is limited to explain fundamental mechanisms for selective recruitment.[256] The structural investigation between *holo**-PCPs and *trans*-acting P450 in the biosynthesis of the depsipeptide skyllamycin is an exception to that and have been extensively characterized. Nevertheless, interaction studies are based on crystal structures and the complex is artificially trapped by an azole inhibitor bound to the PCP, which not allow the monitoring of transient substrate-dependent interaction events. It was highlighted that important interfaces are located in helix α3 as well as the loop α1' mediated by hydrophobic interactions, further stabilized by a network of hydrogen bonds within the P450$_{sky}$ and the Ppant arm.[272] Another *in trans*

modification, reliant on the recruitment of several P450 enzymes by an additional *cis*-standing interaction platform named X domain, has been structural investigated by crystallization showing similar interfaces.[275] Our interaction experiments confirmed residual interfaces in helix α3 but also revealed additional residual contacts located to helix α2. In case of the for the highly conserved insertion loop, downstream of helix α1, it could not be confirmed to be important for the recruitment of the MTase Alb02. Our study shows the structural characterization of a dynamic and transient complex formation between a *trans*-modifying enzyme and a *holo**-PCP domain, highlighting the important role of a substrate-controlled binding equilibria during NRP biosynthesis. The on-line enzymatic tailoring by a substrate-controlled *trans*-methyltransferase in combination to a basal level of transient *holo**-PCP scanning has so far not been described and impacts our understanding of the assembly line logic, module function, biosynthetic pathway prediction, and will open new opportunities for structural diversity of natural products.

Experimental Section

Cloning, recombinant expression and purification of Alb02, PCP-1 and PCP-4 domains

A detailed protocol is included in the Appendix. The selected genes from *Alb01* and *Alb09* containing PCP-1 and PCP-4 domains of NRPS-1 and NRPS-4 were amplified from the cosmid pALB540 and *Alb02* from pALB571 (**Appendix Table 3.1**).[243] The templates were cloned in the pETtrx_1c vector and subsequently transformed into *E. coli* BL21-Gold (DE3) strain (Merck Millipore, Darmstadt, Germany) by heat shock (42 °C, 2 min). Transformants were selected on Luria Bertani (LB)-agar plates supplemented with 50 μg/mL kanamycin and confirmed by DNA sequencing. The *E. coli* expression strains including the vectors pETtrx_1c-PCP1, pETtrx_1c-PCP4 and pETtrx_1c-Alb02 were each incubated in Terrific-Broth (TB) medium containing auto induction solution (37 °C, 180 rpm. for 2 h followed by 10 h at 18 °C) to induce *lac*-controlled gene expression. All recombinant proteins were expressed as C-terminally His_6-tagged fusion proteins with an N-terminal thioredoxin-solubility (trx) tag. After cell lysis, supernatants were purified by using Ni-affinity chromatography with a stepwise gradient of increasing imidazole concentration. Afterwards, solubility tags for all fusion proteins were cleaved by tobacco etch virus (TEV) overnight (16 h) at 4 °C. Size exclusion chromatography was performed on a Superdex 200 HiLoad 16/60 prep grade column linked to an ÄKTA purifier system (GE Healthcare).

Expression and purification of isotope-labeled PCP-1 and PCP-4 domain

Respective vectors were transformed in *E. coli* BL21-Gold (DE3) strain (Merck Millipore, Darmstadt, Germany), grown in M9 minimal medium containing ^{15}N-labeled ammonium chloride (1 g/L) without or with ^{13}C-labeled D-glucose (3 g/L). The subsequent protein expression and purification was performed as aforementioned or in the appendix.

Analytical size-exclusion chromatography

Experiments were performed with a Superdex 200 increase 10/300 column (GE Healthcare) connected to an ÄKTA purifier system (GE Healthcare) and optically monitored with Monitor UV-900, a multi-wavelength UV-Vis detector Monitor (GE Healthcare). All samples were prepared in 120 µL 50 mM Tris/HCl (pH 7.8), 100 mM NaCl with the Alb02 concentrations adjusted to 10 µM and those of the *crypto*-PCP domains to different concentration ratios. After an incubation of 15 min at room temperature, 100 µL of the samples were injected into the HPLC system and analysed with UNICORN7 (GE Healthcare) evaluation software.

Synthesis of *N*-acetylcysteamine (SNAC) and coenzyme A (CoA) derivatives

For a detailed procedure, see the Appendix. Briefly summarized, SNAC or CoA trilithium salt respectively was added to a solution of *para*-amino benzoic acid (pABA) or *para*-amino-3-hydroxy benzoic acid (pABA-3OH), PyBOP and DIPEA in DMF. The reaction mixture was stirred overnight (14 h) at room temperature and purified via RP-HPLC chromatography.

Substrate activation assay of MTase Alb02

In vitro substrate activation samples were prepared in 100 µL 100 mM Tris/HCl (pH 7.8), 1 mM $MgCl_2$, and 2 mM of respective derivatives. The samples were adjusted with addition of 10 µM Alb02 that was and started with addition of ^{14}C-labeled *S*-Adenosylmethionine (1 µL; 0.5 µCi; PerkinElmer Inc.). Thereafter the reaction mixtures were incubated (45 min, 30 °C) and stopped with hydrochloric acid (1 M). Thereafter, samples were diluted with 500 µL water and extraction was carried out two times by adding ethyl acetate. Supernatants were removed, subsequently mixed with scintillation fluid (4 mL) and measured via liquid scintillation counter (60 sec ^{14}C, Hidex 300 SL). All experiments were performed in triplicates.

Substrate-dependent kinetic of Mtase Alb02

The methyltransferase Alb02 (10 µL, 20 µM) was added to the reaction mixture (5 µL 10 mM $MgCl_2$, 5 µL 2 mM SAM) including different concentrations of pABA-3OH-SNAC or pABA-3OEt respectively (250 µM, 150 µM, 100 µM, 50 µM, 40 µM, 30 µM, 20 µM, 10 µM, 5 µM, 1 µM, 0.1 µM) for obtaining Michaelis-Menten parameters. Reaction mixtures were incubated 5 min at 30 °C and stopped by quenching with hydrochloric acid (5 µL, 1 M). Samples were extracted by adding ethyl acetate. The quenched samples were then centrifuged (5 min, 16.000×g) before measuring by LC-MS. All experiments were performed in triplicates.[276–278]

Multiple reaction monitoring (MRM) MS

Tandem mass spectrometry measurements were performed on a TripleQuad LC/MS 6460 (Agilent Technologies) coupled to an analytical UHPLC 1290 system (Agilent Technologies) using

a Discovery BIO Wide Pore C18-3 (3 μm, 150 x 2.1 mm) column with a linear gradient at 0.3 mL/min from 5% B (A = water + 0.1% formic acid (HFO), B = acetonitrile + 0.1% HFO) to 95% B over 25 min. Detection was performed through multiple reaction monitoring (MRM) using the precursor ions m/z 269.10 (pABA-3OMe-SNAC) and m/z 195.09 (pABA-3OMe-OEt) for the product ion scans. The selected product ion m/z 151.10 (pABA-3OMe b-ion) was selected for both precursor ions.

Carrier protein *crypto* labeling

The *in vitro* modification of unlabeled *apo*-PCP domains (50 μM) was incubated with the respective CoA-constructs in a reaction mixture (100 mM Tris/HCl pH 7.8, 10 mM $MgCl_2$, 1 μM Sfp, 5 mM CoA-derivatives and 10 mM DTT) for 20 min at 30 °C. The reaction mixture was subsequently concentrated using an Amicon Ultra 3k centrifugal filter (Millipore) with a molecular weight cut off (MWCO) of 3 kDa.

Crypto-PCP on-line modification and intact mass analysis

Previously labeled *crypto*-PCP domains (15 μL, 50 μM) were implemented with methyltransferase Alb02 (3 μL, 20 μM) and SAM (2 μL, 200 μM). Top-down MS measurements were performed on an Exactive ESI-Orbitrap (Thermo Fisher Scientific GmbH) coupled to an analytical HPLC 1200 system (Agilent Technologies) using a Discovery BIO Wide Pore C18-3 (3 μm, 150 x 2.1 mm) column with a linear gradient at 0.3 mL/min from 5% B (A = water + 0.1% formic acid (HFO), B = acetonitrile + 0.1% HFO) to 95% B over 25 minutes in a CID full scan positive mode. The following settings were adjusted for Orbitrap measurements: mass range (m/z 400 – 4000), spray voltage (4.5 kV), tube lens voltage (185 V), capillary voltage (35 V) and collision energy (35 V).

In vitro crosslinking assay

A detailed protocol is described in the Appendix. The proteins were concentrated in HEPES buffer (20 mM, pH 7.8) to a final concentration of 100 μM. The crosslinking reagent, disuccinimidyl sulfoxide (DSSO), was prepared in a DMSO stock solution (2 mM). The peptidyl-carrier protein samples (PCP-1 and PCP-4) were mixed with DSSO from the stock solution in a ratio 1:20. The methyltransferase Alb02 (10 μM) was added to all samples. Samples were incubated for 1 h at room temperature and terminated by adding NH_4HCO_3 (1 M) to a final concentration of 50 mM. Afterwards samples were submitted to SDS-PAGE gel electrophoresis (12% polyacrylamide) to evaluate crosslinking events.

In-gel digestion and protein identification

The Coomassie-stained bands of individual or crosslinked proteins were excised from the gel and subjected to in-gel digestion (14 h at 37 °C with 66 ng sequencing-grade trypsin/µL in 25 mM ammonium bicarbonate, 10% ACN; 0.25 µg/sample) after disulfide reduction (10 mM dithiothreitol) and thiol protection (50 mM iodoacetamide). Tryptic peptides were submitted to tandem MS analysis on an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific) coupled to an Agilent 1200 HPLC system (Agilent Technologies). The LC-MS data files (.raw) obtained from the in-gel digestion were converted to mascot generic format (.mgf) files via MSConvert GUI of the ProteoWizard package (http://proteowizard.sourceforge.net; version 3.0.10328). LC–MS/MS data files (.raw) were converted to mascot generic format (.mgf) files via MSConvert GUI of the ProteoWizard package (http://proteowizard.sourceforge.net; v3.0.10577) and annotated by peptide spectrum matching. The SearchGUI[279] (v3.3.11) software tool was used with X!Tandem as the search engine. The free available MeroX[208] software was used for automated data analysis based on characteristic fragment ion signatures created by the MS-cleavable linker.

NMR data acquisition and assignment

NMR experiments were performed on an Avance III 700 MHz spectrometer equipped with a room-temperature TXI probe (Bruker, Karlsruhe, Germany). TopSpin 3.5 (Bruker, Karlsruhe, Germany) was used for data acquisition and processing.

Triple-resonance experiments were performed on ^{13}C,^{15}N-*apo/holo*-PCP-4 (0.7 mM) in PBS buffer, pH 7.8 (10% D$_2$O) at 298 K. In addition to 2D ^{1}H-^{15}N HSQC and constant-time ^{1}H-^{13}C HSQC spectra, we acquired the following 3D spectra for resonance assignment: HNCO, HN(CA)CO, HNCA, HN(CO)CA, CBCANH, CBCA(CO)NH, HCC(CO)NH (CC-TOCSY mixing time of 12 ms), HNHA, as well as ^{1}H-^{15}N HSQC-NOESY (mixing time τ_m of 140 ms). 3D spectra were recorded with acquisition times of 120 ms in the direct ^{1}H dimension, 19 ms in the indirect ^{15}N dimension, 9 ms in the indirect ^{13}CA dimension, 6 ms in the indirect ^{13}CACB and ^{13}CC dimension, 23 ms in the indirect ^{13}C' dimension and 10 ms in the indirect ^{1}H dimensions.

We used rapid data accumulation (^{1}H-^{15}N SOFAST-HMQC[280]) for titration experiments in order to achieve high molar ratios of Alb02:PCP employing a PCP concentration of only 0.05 mM. ^{1}H-^{15}N SOFAST-HMQC spectra were acquired with acquisition times of 90 ms and 30 ms in the direct ^{1}H and indirect ^{15}N dimension, respectively. The interscan delay was set to 100 ms. Band-selective excitation and inversion of amide protons was achieved by applying shaped pulses with durations of 2.4 ms (Pc9[281]) and 1.6 ms (REBURP[282]), respectively, with an offset of 2.3 kHz.

Apodization of time domain data was performed using a squared sine-bell function shifted by 90°. The 2D and 3D data was processed by applying linear forward prediction in the indirect dimensions and zero filling prior to Fourier transformation.

[1]H chemical shifts were referenced externally using a sample of trimethylsilylpropanoic acid (TMSP-d_4, Deutero GmbH, Kastellaun, Germany) in PBS buffer, pH 7.8 at 298 K. [13]C and ^{15}N chemical shifts were referenced indirectly using correction factors of $f_{13C/1H}$ = 0.251449530 and $f_{15N/1H}$ = 0.101329118, respectively.[272,283]

NMR titrations and complex structure calculation

NMR spectra analysis and resonance assignments were performed using NMRFAM-SPARKY.[284] Chemical shifts were analyzed using TALOS+ to assess secondary structural elements.[285] The experimental secondary structure was compared with that of various structural homology models generated via the webtools PHYRE2[286], ITASSER[287] and MODELLER[288]. All homology models were consistent with regard to the boundaries of α-helices as well as their tertiary packing (independent of the structural template) underlining the strong conservation of the PCP fold. However, PHYRE2 and ITASSER predicted partial secondary structure for the elongated loop between helices α1 and α1' of PCP-4. However, chemical shift analysis clearly indicated this region to be highly dynamic. We thus used the model generated by MODELLER as a valid representative for the NMR data. PyMOL 2.0.0 was used for visualization of protein structures.[289]

Titration curves were generated by extracting peak intensities from [1]H-^{15}N SOFAST-HMQC spectra and calculating signal attenuation Q according to the formula $Q = (I_0 - I_{\text{titration}}) I_0^{-1}$, with I_0 and $I_{\text{titration}}$ representing the cross-peak intensities in the absence and presence of Alb02, respectively. For nonlinear least-square fitting of titration data and determination of dissociation constants K_d, we used the following equation:

$$Q = \frac{Q_{\text{max}}([\text{PCP}]_t + [\text{Alb02}]_t + K_d) - \sqrt{([\text{PCP}]_t + [\text{Alb02}]_t + K_d)^2 - 4[\text{PCP}]_t[\text{Alb02}]_t}}{2[\text{PCP}]_t}$$

where $[\text{x}]_t$ refers to the total concentrations of the corresponding protein x and Q_{max} is the maximal attenuation.

4   Comparative Venomics of the *Vipera ammodytes transcaucasiana* and *Vipera ammodytes montandoni* from Turkey Provides Insights into Kinship

Introduction

Venom research has an ongoing significance for various disciplines and applications ranging from drug development, pharmacology for rational antivenom production even to the cosmetics industry.[62,290,291] After the discovery of the first venom-derived therapeutic, Captopril, in 1975, which was developed from the Brazilian pit viper (*Bothrops jararaca*), in the following years, investigations of several other venomous snakes revealed further venom-based drugs with different medical applications.[292,293] Nevertheless, there is still an uncountable number of venomous animals in diverse habitats to be suspected, of which only a small part of venom proteomes have been characterized, and as a consequence the potential for new applications of these venoms and individual components thereof still have to be explored.[291,294,295]

The advances in "-omics" technologies allowed for the characterization of an increasing number of animal venoms and plays a pivotal role for the development of new potential drugs against several human diseases. The increase in sensitivity and the development of soft ionization methods, e.g., electrospray ionization (ESI) or matrix-assisted laser desorption/ionization (MALDI), for mass spectrometry and next-generation high-throughput sequencing dramatically enhanced the analysis of venoms.[291,296] Nowadays, there exist different well-established protocols to characterize the venom in its entirety. The so-called bottom-up strategies can be divided in gel-based approaches and liquid chromatographic (LC)-based approaches. A combination of both strategies, termed "snake venomics", uses both separation techniques successively followed by an in-gel digestion of excised protein bands and a mass spectrometric measurement by tandem mass analysis.[225,297] Recently described methods for the proteomic analysis of snake venoms include the "intact mass profiling", which directly separate the components out of the crude venom without any previous fractionation. The power of this tool has already been demonstrated at the example of whole snake venom proteomes.[233,234] The intact mass analysis of native proteins compared to chemically reduced proteins allows for a classification, based on the number of existing intra- and intermolecular disulfide bonds, and represents an important characteristic of different viper-venom protein families.[224,226,298] However, the method is critical for proteins of higher molecular masses, e.g., snake venom metalloproteases, because the high resolution of accurate isotopic masses becomes

challenging.[234] The subsequent step to the intact mass profiling is the MS/MS analysis by collision-induced dissociation (CID) of intact molecular masses, termed "top-down venomics". The major advantage of this approach is a high-throughput analysis of venoms without the necessity of further pre-MS separation steps. A decisive drawback is the application mainly limited to peptides and small-sized proteins (>15 kDa), which are the main constituents of *Elapidae* (e.g., three-finger toxins, Kunitz-type inhibitor, etc.).[227,231,299] In contrast, *Viperidae* venoms contain higher molecular mass components that makes them less suitable for the top-down approach as *Elapidae* venoms.[233,234] Finally, the mass spectrometry-based absolute intact mass quantification by isotope dilution is a further cutting-edge approach, which could replace the semi-quantitative densitometric determination.[226,228,300]

The combination of several workflows allows for an encompassing characterization of different kinds of venoms. In particular, the venom of vipers is a promising source of new substances and therapeutics, due to their different venom compositions.[62,292] They are distributed in a wide range all over the world, and are especially located around the Mediterranean Sea.[301] A great variety of habitats and zones of subtropical climate along the north coast side of Turkey provides suitable places to shelter for many species that belong to the *Viperidae* family.[302] Important major protein families found in analyzed viperid venoms are snake venom metalloproteases (svMP), snake venom serine proteases (svSP), hyaluronidases, 5′-nucleotidase, phospholipases A$_2$ (PLA$_2$), disintegrins, C-type lectin like proteins (CTL), cysteine-rich secretory proteins (CRISP), natriuretic peptides, bradykinin-potentiating peptides (BPP), nerve growth factors (NGF), snake venom vascular endothelial growth factors (VEGF-F) and Kunitz-type protease inhibitors.[84,303]

Our ongoing studies on snake venoms focus on the venom characterization of unrecorded *Vipera* in the Turkish area and initial cytotoxicity screenings against cancerous as well as non-cancerous cell lines of potent bioactive peptides and proteins. From this point of view, we aimed to screen viper venoms from different regions of Turkey. For this purpose, the regional endemic Transcaucasian Nose-horned Viper (*Vipera ammodytes transcaucasiana*) from central Anatolia and the *V. a. montandoni* from Northwest of Turkey (Turkish Thrace) were chosen for a comparative venom investigation. The Nose-horned Viper (*Vipera ammodytes*) is one of the most venomous snakes in Europe and common pathophysiological conditions range from local tissue damage, hemorrhage, pain, paralysis up to necrosis and in some cases even death.[304,305] After the description of *Vipera ammodytes ammodytes* (Linnaeus, 1758), five further subspecies have been described: *V. a. meridionalis*[306], *V. a. montandoni*[307], *V. a. transcaucasiana*[308], *V. a. ruffoi*[309] and *V. a. gregorwallneri*[310]. *V. a. transcaucasiana* is considered a separate species by some authors.[311] Heckes et al.[312] (2005) and Tomovic[313] (2006) accepted only four valid taxa for *V. ammodytes* (*V. a. ammodytes*, *V. a. meridionalis*, *V. a. montandoni* and *V. a. transcaucasiana*) with an extensive investigation on the species. Phylogenetic and phylogeographic studies, using mtDNA gene sequences obtained from cytochrome *b* (cyt *b*), 16S rRNA and the noncoding control region,

supported the validation of subspecies status of *V. a. ammodytes*, *V. a. meridionalis* and *V. a. montandoni*, but in turn *V. a. ruffoi* and *V. a. gregorwallneri* were only accepted as synonyms to the nominotypic subspecies, *V. a. ammodytes.* In addition, the taxonomic status of *V. a. transcaucasiana* was tentatively classified as subspecies due to a low sample size.[314]

Figure 4.1. Geographical distribution of subspecies from *Vipera ammodytes*. The distribution areas of the four *Vipera ammodytes* subspecies are highlighted in color: *V. a. ammodytes* (yellow), *V. a. montandoni* (blue), *V. a. meridionalis* (green) and *V. a. transcaucasiana* (red). Overlapping distribution areas are highlighted by shaded colors. The locations for catches of *V. a. montandoni* (star, blue) and *V. a. transcaucasiana* (star, red) are marked and exemplary snake habitats are shown.

The occurrence of *Vipera ammodytes* distributes around the Mediterranean Sea and reaches from the Alps over to Turkey, Georgia, Azerbaijan and Iran. The Transcaucasian Nose-horned Viper (*Vipera ammodytes transcaucasiana* (Vat)) shows a distribution in the Northeast of Turkey and sections of Georgia along the Black Sea coast and some inland provinces in Turkey (see **Figure 4.1**, red).[313,315] The Transdanubian Sand Viper (*Vipera ammodytes montandoni* (Vam)) is spread from Turkish Thrace, Bulgaria to Romania and shares its distribution area in parts with all three other subspecies (**Figure 4.1**, blue).[313] Beside those previously mentioned, there exist two further subspecies whose venoms were already characterized: The Western Sand Viper (*Vipera ammodytes ammodytes*) can be found from the Alps over Croatia to the borders of Macedonia (**Figure 4.1**, yellow) and the Eastern Sand Viper (*Vipera ammodytes meridionalis*), only endemic in Greece and several Hellenic islands (**Figure 4.1**, green).[316] All subspecies of *Vipera ammodytes* can be found from sea level up to 2000 m a.s.l. in many kinds of suitable habitats (forests,

meadows, arid regions, rocky areas, and even sandy coastal parts), thus there is no special habitat selectivity. The Nose-horned viper (*Vipera ammodytes*) is one of the most and venomous species in Europe and therefore of significance for public health.[316,317]

Previous investigations on the neutralization of lethality by several antisera against *Vipera ammodytes* subspecies revealed low paraspecific neutralization potency.[318,319] Therefore, the elucidation of the undescribed venom proteome is significant for public health and could help to bypass the lack of sufficient venom neutralization. Here, we give deeper insight into the composition of the venom proteome and peptidome of the two Nose-Horned vipers by bottom-up venomics and an intact mass profiling of the crude venoms.

The detailed characterization and comparison of the venom proteomes with other subspecies showed a remarkable matching of the venom components, which could be an additional helpful tool to overcome the controversial question of the taxonomic status of *Vipera ammodytes transcaucasiana* in connection with the phylogenetic analysis.[314]

Results and Discussion

The venom proteome

Intact mass profiling

First, we applied an intact venom molecular mass profiling to obtain an overview of molecular masses from all venom components, including low abundant and low molecular mass compounds. Therefore, the crude venom as well as the RP-HPLC separated peptide fractions were used. The initial profiling of *Vipera ammodytes transcaucasiana* (Vat) revealed 117 molecular masses for different venom components (**Figure 4.2A**, **Tables 4.1** and **Appendix Figure 4.1**): 55 (<1 kDa), 11 (1–3 kDa) and 28 (3–9 kDa), which represents the peptide part of the venom in total with 13.49%. Higher molecular masses from 10 to 28 kDa were detected 23 times in the following composition: 1 (10 kDa), 14 (12–16 kDa), 2 (21 kDa) and 6 (25–28 kDa). The venom of *Vipera ammodytes montandoni* (Vam) showed a comparable distribution pattern with 115 different venom components (**Figure 4.2B**, **Tables 4.1** and **Appendix Figure 4.2**): 47 (<1 kDa), 26 (1–3 kDa) and 19 (3–8 kDa), which corresponds to a slightly higher peptide content of 17.49%. We also found 25 components with molecular masses between 13 and 34 kDa: 10 (13–16 kDa), 2 (21 kDa), 7 (24–27 kDa) and 6 (32 kDa).
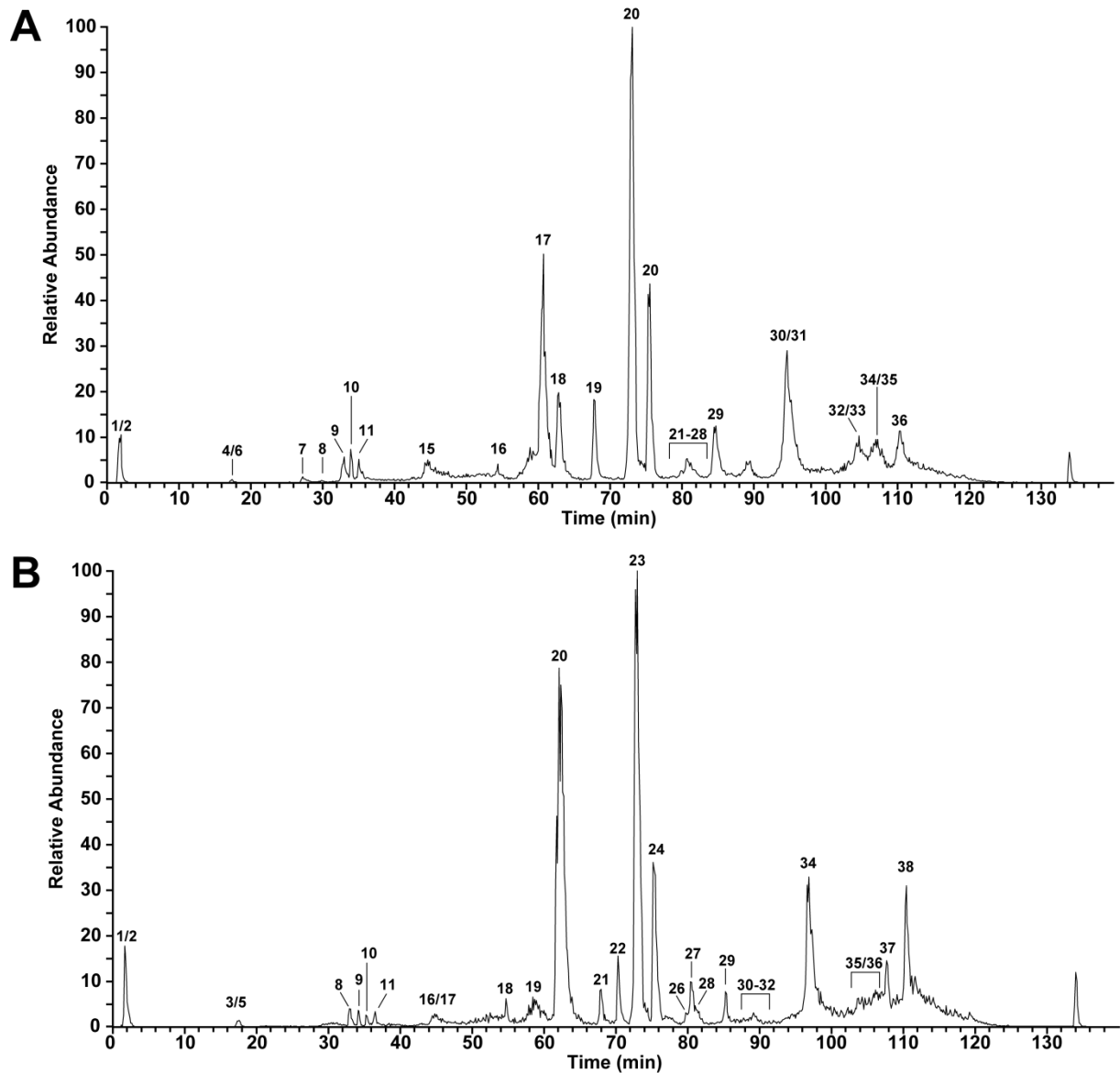
Figure 4.2. Intact molecular mass profiles of *V. a. transcaucasiana* (Vat) and *V. a. montandoni* (Vam). The total ion counts (TIC) of crude venoms from: (**A**) Vat; and (**B**) Vam were measured by HPLC-ESI-MS. The peak nomenclature is based on the chromatogram fractions and is shown in figure 4.3. The identified molecular masses of intact proteins and peptides are listed for Vat in appendix table 4.1 and Vam in appendix table 4.2.

Furthermore, the initial mass profiling enabled us to identify two peptides as members of the tripeptide metalloprotease inhibitor (svMP-i) family. The protease inhibitors serve to avoiding damages during storage of the venom in the gland tissue as well as to prevent auto-proteolysis of the venom.
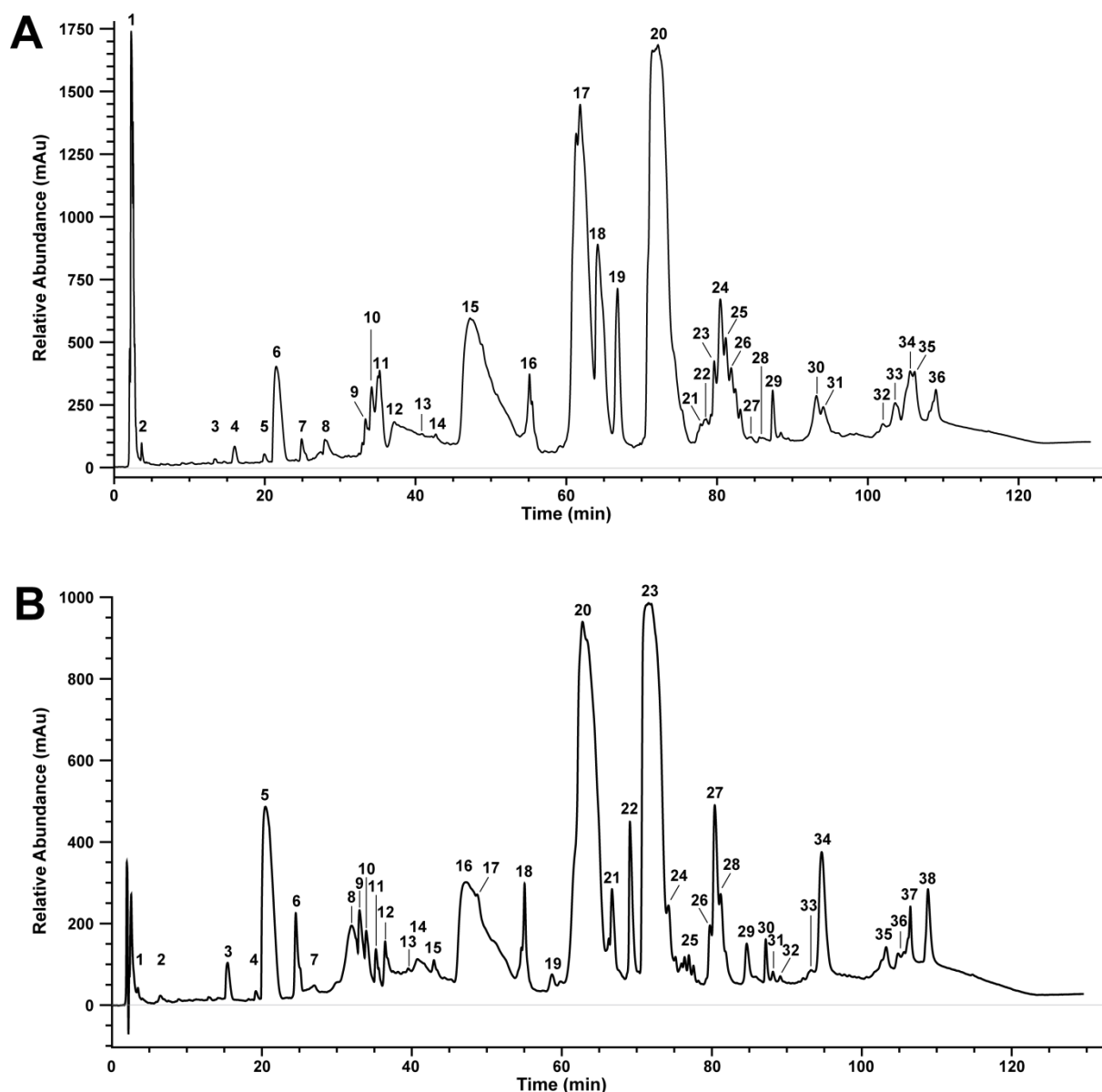
Figure 4.3. Chromatograms of *V. a. transcaucasiana* (Vat) and *V. a. montandoni* (Vam) venoms separated by semi-preparative reversed-phase HPLC. Venom separation of: (**A**) Vat; and (**B**) Vam was performed by a Supelco Discovery BIO wide Pore C18-3 RP-HPLC column and UV absorbance measured at 214 nm.

Protective effects have been described for the well-known endogenous pyroglutamic tripeptide metalloprotease inhibitors, e.g., the small tripeptides pEQW and pENW.[116] Their strong inhibitory effect against svMP from different vipers was intensively studied.[117,320] In our studies, the venoms of *V. a. transcaucasiana* and *V. a. montandoni* exhibit two molecular masses with *m/z* 444.23 and *m/z* 430.17 (**Figure 4.4** and **Appendix Tables 4.1** & **4.2**). While the molecular mass of *m/z* 444.23 corresponds to a well separated peptide signal found in the venoms of Vat (**Figure 4.2**; Peak 4/6) and Vam (**Figure 4.2**; Peak 3, 4/5), the intensity of the signal at *m/z* 430.17 for Vat (**Figure 4.2**; Peak 7) and Vam (**Figure 4.2**; Peak 6) is less prominent.
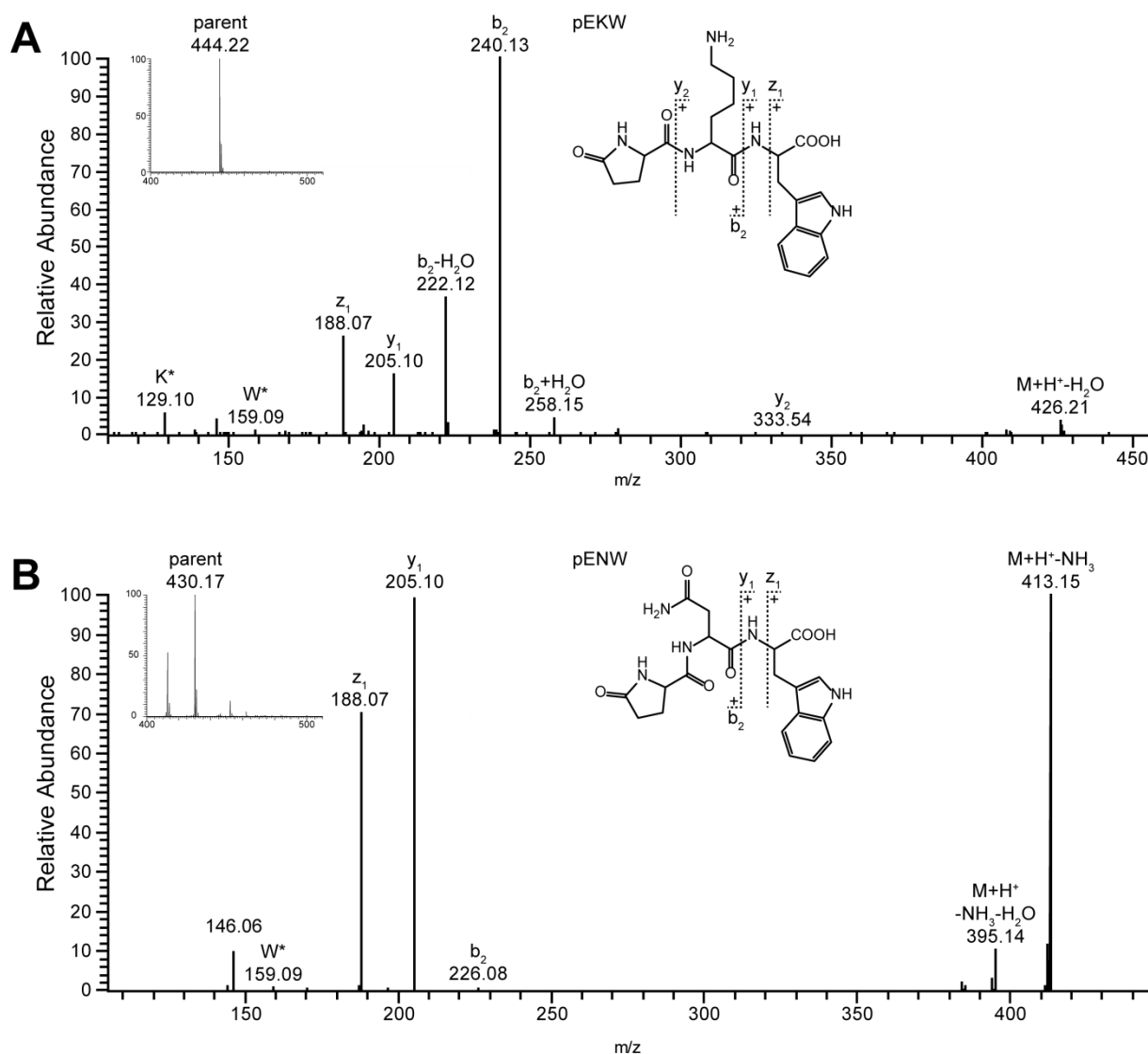
Figure 4.4. CID-MS/MS spectra of two snake venom metalloproteinase (svMP) inhibitor tripeptides. The two snake venom metalloproteinase inhibitor tripeptides: (**A**) pEKW; and (**B**) pENW were identified by intact mass profiling in the venoms of Vat and Vam (here Vat). The HPLC-ESI-MS1 spectra of selected parent ions are shown at the left corner. Standard fragmentation ions were indicated at the ion mass peaks and amino acid related ions by asterisked single letter code.

Accordingly, the measured mass of m/z 430.17 matches to the predicted monoisotopic mass of the pyroglutamic tripeptide pENW, while the measured m/z 444.23 matches to the predicted monoisotopic mass of the pyroglutamic tripeptide pEKW. The identities of the pyroglutamic tripeptides pEKW and pENW were ultimately confirmed by ESI-MS/MS experiments (**Figure 4.4**). In addition, comparison with spectra for the pEKW inhibitor from the literature coincides with our experimentally observed fragmentation pattern.[321] As an additional evidence for tripeptide metalloprotease inhibitors in the venom of both vipers, the presence of several Gly- and Pro-rich protein fragments of the endogenous precursors were detected in the venoms of *V. a. transcaucasiana* as well as *V. a. montandoni*.[322]

Table 4.1. Venom proteins and peptides identified from *Vipera ammodytes transcaucasiana* and *Vipera ammodytes montandoni*. Venomic components were assigned by crude venom intact mass profiling and bottom-up approach. Peak numbers and retention time (RT) based on RP-HPLC (**Figure 4.3**) and TIC (**Figure 4.2**) annotation. SDS-PAGE and intact mass profile analysis provided the molecular weight. Most abundant masses are asterisked (*). For peptide fractions, only the most abundant masses were noted. Detailed lists of all exhibit masses are mentioned in appendix tables 4.1 and 4.2.

		V. a. transcaucasiana					*V. a. montandoni*		
RT (min)	Fraction No.	Protein Species	SDS PAGE M_{av} (kDa)	Most Abundant IMP (m/z)	RT (min)	Fraction No.	Protein Species	SDS PAGE M_{av} (kDa)	Most Abundant IMP (m/z)
2.14	1	unknown	-	571.25 *	-	-	-	-	-
3.45	2	unknown	-	484.20 *	3.46	1	unknown	-	484.21 *
-	-	-	-	-	6.45	2	svMP-i, unknown	-	1234.65 *
12.86	3	unknown	-	600.32 *	-	-	-	-	-
15.48	4	svMP-i, svMP unknown	-	547.22 *	15.43	3	svMP-i, unknown	-	444.22 *
19.47	5	svMP, unknown	-	814.35 *	19.19	4	svMP-i, unknown	-	444.23 *
21.08	6	svMP-i, unknown	-	444.22 *	20.48	5	svMP-i, unknown	-	444.22 *
24.49	7	svMP-i, unknown	-	430.17 *	24.52	6	svMP-i, unknown	-	430.17 *
27.85	8	unknown	-	569.28 *	26.99	7	unknown	-	1072.60 *
-	-	-	-	-	31.95	8	svMP-i, unknown	-	3930.96 *
33.07	9	svMP-i, unknown	-	5775.64 *	33.02	9	unknown	-	3761.72 *
33.86	10	unknown	-	4176.85 *	33.91	10	unknown	-	3796.72 *
34.95	11	BPP, unknown	-	3769.75 *	35.20	11	unknown	-	3796.73 *
36.97	12	unknown	-	1143.64 *	36.44	12	unknown	-	1144.62 *
-	-	-	-	-	39.50	13	unknown	-	1314.73 *
40.43	13	unknown	-	1143.64 *	40.75	14	unknown	-	1144.62 *
42.44	14	svMP-i, unknown	-	1143.64 *	42.94	15	unknown	-	-
46.90	15	VEGF-F	15*	-	47.16	16	VEGF-F	14 *	1159.59 *
-	-	-	-	-	48.69	17	VEGF-F	14 *	1159.59 *
54.79	16	VEGF-F, svMP, unknown	15, 27 *, 50	10,676.97, 21,311.88 *	55.03	18	unknown	14, 25 *	21,199.58 *, 21,298.85
-	-	-	-	-	58.69	19	unknown	-	13,553.82 *, 13,590.75
61.54	17	PLA$_2$, unknown	13 *, 25, 40	13,553.83 *, 13,590.76, 13,814.21, 13,842.19, 13,911.15	62.75	20	PLA$_2$	13 *	13,553.82 *, 13,890.28, 13,988.22
63.85	18	PLA$_2$	13 *, 25, 40	13,918.28 *, 14,016.22	-	-	-	-	-
66.42	19	PLA$_2$, CRISP	13, 25 *, 50	12,346.55, 24,653.41 *, 24,752.38, 24,848.30	66.68	21	PLA$_2$, CRISP	13.23 *	24,654.40 *, 24,750.41
-	-	-	-	-	69.08	22	CRISP	23 *	24,547.04 *

Table 4.1. Venom proteins and peptides identified from *Vipera ammodytes transcaucasiana* and *Vipera ammodytes montandoni.* *continued*

		V. a. transcaucasiana					*V. a. montandoni*		
RT (min)	Fraction No.	Protein Species	SDS PAGE M_{av} (kDa)	Most Abundant IMP (m/z)	RT (min)	Fraction No.	Protein Species	SDS PAGE M_{av} (kDa)	Most Abundant IMP (m/z)
71.28	20	PLA$_2$, svMP, CRISP	13, 22.5 *, 27.5, 60	13,624.69, 13,625.73, 13,676.78, 24,515.96 *	71.38	23	PLA$_2$	13, 21 *	13,624.69
-	-	-	-	-	74.21	24	PLA$_2$	13 *, 21	13,676.81 *
-	-	-	-	-	76.93	25	PLA$_2$, svSP, unknown	15, 23, 37 *	-
78.24	21	svSP	35 *	-	-	-	-	-	-
78.90	22	unknown	35 *	-	-	-	-	-	-
79.43	23	svSP	35 *, 85	-	79.74	26	svSP, unknown	15, 37 *, 60, >200	32,026.88 *, 32,899.08
80.23	24	svSP	35 *, 85	-	80.39	27	svSP	37 *	32,686.16 *, 35,124.93
80.95	25	svSP	35 *, 85	-	81.17	28	CRISP, svSP, unknown	15, 25, 37 *	32,686.34 *, 33,342.02
81.59	26	svSP	40 *, 100	-	-	-	-	-	-
84.01	27	svMP	70 *	-	-	-	-	-	-
85.34	28	svSP	35 *	-	84.65	29	svMP, trypsin-like, unknown	15, 25, 37, 60 *	24,547.96, 27,654.98
87.17	29	CTL, svSP, PDE	13 *, 35, 65	16,108.34 *, 16,208.30	87.19	30	CTL, svSP, svMP	13 *, 30, 37, 55	13,890.25 *
-	-	-	-	-	88.14	31	svSP, unknown	30, 37 *, 55	-
-	-	-	-	-	89.12	32	svSP, unknown	30 *, 37	-
92.85	30	CTL, LAAO	13, 20, 40, 60 *	-	93.23	33	CTL, svSP, LAAO	11, 20, 30, 37, 50, 55 *	-
93.88	31	CTL, LAAO	20, 60 *	-	94.64	34	LAAO	30, 37, 55 *	-
101.76	32	CTL, DI, svMP, unknown	20, 35, 60 *	-	-	-	-	-	-
103.42	33	svSP, svMP, unknown	20, 35, 60 *	-	103.23	35	CTL, svSP, LAAO	11, 20, 30, 37, 50 *	-
104.82	34	svMP, LAAO, svSP	20, 35, 65 *	-	104.83	36	aminopeptidase, svMP, LAAO, unknown	30, 50 *, 70	-
105.70	35	svMP	65 *	-	106.47	37	svMP, LAAO	30, 50 *	-
108.84	36	svSP, LAAO, svMP	35, 60 *	-	108.82	38	unknown	30, 37, 50 *	-

The peptide PEGPPLMEPHE, a previously described tripeptide precursor fragment[323] from the related snake *V. a. ammodytes*, was detected in both investigated venoms (**Appendix Figures 4.1 & 4.2** and **Appendix Tables 4.1** and **4.2**, Peak 9 and 8). Furthermore, several other precursor peptide fragments of this type were detected in the venoms of Vat (GGGGGGW, PPQMPGPKVPP) and Vam (DNEPPKKVPPN) (**Appendix Figures 4.3–4.6** and **Appendix Tables 4.1 & 4.2**). Taken together, the metalloprotease inhibitors and their precursor peptides form the major share of the peptidome of Vat (28.80%) as well as of Vam (45.93%), with the tripeptide pEKW as the main component. Even if details on secretion and processing of the tripeptides are still unknown, a formation of the pyroglutamic inhibitors is assumed to happen during the exocytosis. A processing in the gland lumen itself is rather unlikely based on missing glutaminyl cyclases.[321]

A drawback of the intact mass profiling is that high molecular masses become difficult to properly detect. A reason is that the higher charged states of molecular ions cause a high peak density which increasingly challenges the limits in resolution of the orbitrap mass analyzer. Hence, the incomplete characterization of the venoms discriminating higher masses requires the complementary bottom-up approach.

Bottom-up venomics

The bottom-up analysis by the combined approach, termed snake venomics, was performed with lyophilized snake venoms of Vat and Vam. Subsequent to fractionation by reversed phase-HPLC (see Figure 3), the protein containing fractions were size-separated by SDS-PAGE (**Figure 4.5**). The prominent bands were excised followed by tryptic in-gel digestion and *de novo* sequencing via MS/MS. The quantitative venom composition was calculated based on the RP-HPLC peak integration and in case of co-eluting components, the ratio of optical intensities and densities from SDS-PAGE was deduced.[228,294,300]

A concluding analysis of the Vat and Vam venoms rendered the following results (**Figure 4.6**): the most abundant toxin family of the Vat venom is represented by snake venom phospholipases A_2 (svPLA$_2$, 44.96%) followed by vascular endothelial growth factors (VEGF-F, 9.81%), snake venom serine proteases (svSP, 9.47%), snake venom metalloproteases (svMP, 8.76%), L-amino acid oxidases (LAAO, 6.41%), cysteine-rich secretory proteins (CRISP, 3.41%) and C-type lectin like proteins (CTL, 2.99%). The remaining constituents, such as a disintegrin and a phosphodiesterase, were summarized as other proteins (0.07%) and unannotated proteins (0.60%) (**Figure 4.6A**), respectively. Similarly, the venom of Vam is composed of snake venom phospholipases A_2 (svPLA$_2$, 52.44%) as main part followed by vascular endothelial growth factors (VEGF-F, 10.69%), snake venom serine proteases (svSP, 5.48%), L-amino acid oxidases (LAAO, 4.83%), cysteine-rich secretory proteins (CRISP, 3.81%), snake venom metalloproteases (svMP, 1.79%), C-type lectin like proteins (CTL, 0.26%), several trypsin-like proteins and one aminopeptidase (other proteins, 0.11%) and non-characterized proteins (3.11%) (**Figure 4.6B**).
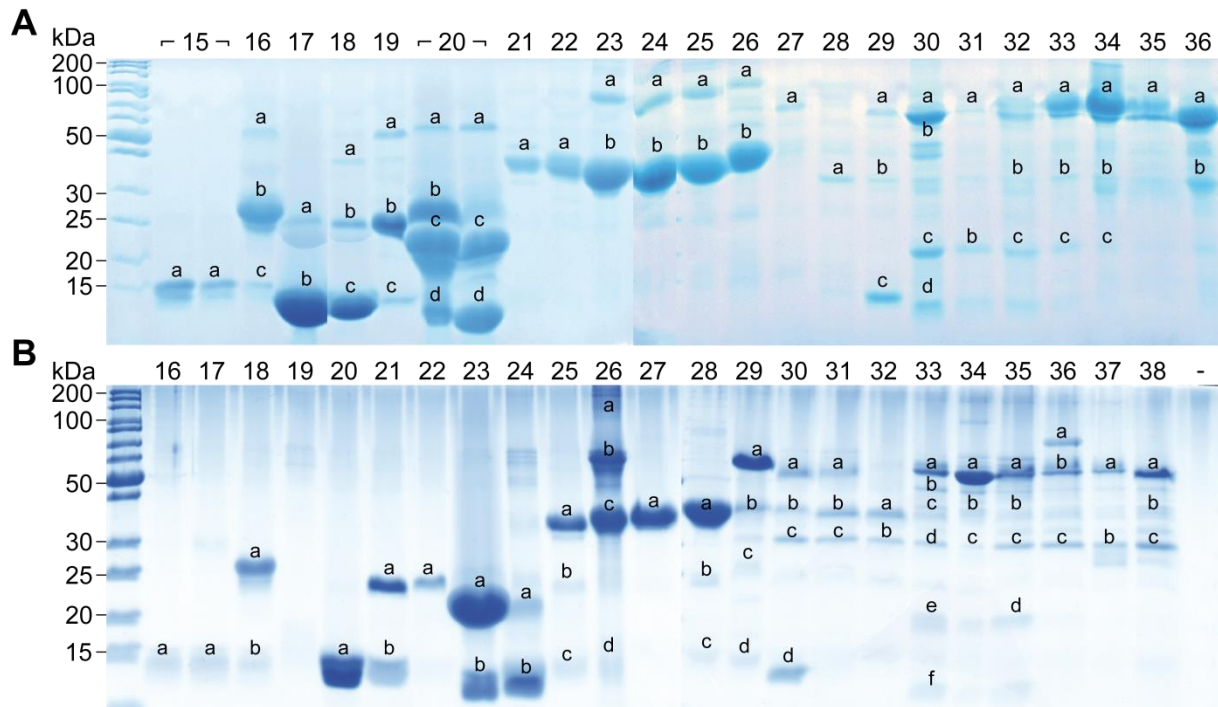
Figure 4.5. Venom fraction analysis of *V. a. transcaucasiana* and *V. a. montandoni* by SDS-PAGE. The RP-HPLC fractions (indicated above the lane are based on **Figure 4.3**) of the (**A**) Vat and (**B**) Vam venoms were analyzed by SDS-PAGE under reducing conditions. Alphabetically marked bands per line were excised for subsequent tryptic in-gel digestion.

In total, 118 fragments with 84 different sequences for Vat (**Table 4.1** and **Appendix Table 4.1**) and 87 fragments to 66 sequences for Vam (**Table 4.1** and **Appendix Table 4.2**) could be assigned by bottom-up analysis with a subsequent *de novo* sequencing. The assignment revealed, e.g., for the phospholipase A$_2$ protein family, homologs of both chains from the heterodimeric Vaspin (acidic, Uniprot-ID: CAE47105.1; basic, Uniprot-ID: CAE47300.1), one of Vipoxin (B chain, Uniprot-ID: 1AOK_B) and the Ammodytin. This monomeric PLA$_2$ was identified in the case of Vat as the neutral Ammodytin I2(A) variant (Uniprot-ID: CAE47197.1) and in the venom of Vam by two isoforms of the acidic Ammodytin I1(A) (Uniprot-ID: CAE47141.1) and I1(C) (Uniprot-ID: CAE47172.1). The main snake venom serine proteases (svSP) could be identified as homologs of Nikobin (Uniprot-ID: E5AJX2.1) and the Cadam10-svSP11 (Uniprot-ID: JAV48393.1). The identity of several bottom-up determined proteins, e.g., Vipoxin or Ammodytin, could be additionally assigned by intact mass profiling through comparison to their average protein family masses.
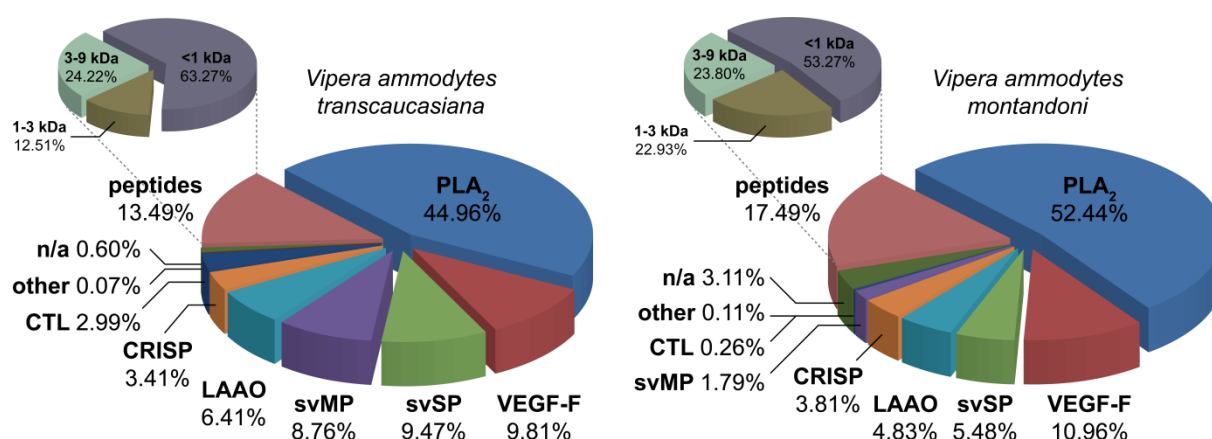
Figure 4.6. Semi-quantitative venom composition of *V. a. transcaucasiana* and *V. a. montandoni*. The relative occurrence of different toxin families of: (**A**) Vat and (**B**) Vam are represented by pie charts. Identification of phospholipases A_2 (PLA2, blue), vascular endothelial growth factors (VEGF-F, red), snake venom serine proteases (svSP, green), snake venom metalloproteases (svMP, violet), L-amino acid oxidases (LAAO, light blue), cysteine rich secretory proteins (CRISP, orange), C-type lectin like proteins (CTL, dark blue), other proteins (other, dark red), unknown proteins (n/a, dark green) and peptides (light red). Groups of different peptide sizes are summarized in an additional pie chart as percentages of the total peptide content and clustered to <1 kDa (dull purple), 1–3 kDa (dull brown) and 3–9 kDa (dull green).

Comparative venomics of *Vipera ammodytes*

Correlations between venom composition and relationship from snakes belonging to the same genera have been shown in the literature.[4,324] Nevertheless, many examples for interspecies variations in the composition of snake venoms are also described.[84,325,326] Variations in the venom composition can be associated to different diets, regional separation of populations, sex or age.[327–331] Likewise, Tashima et al.[332] reported significant variations in the venom composition of close-related pitvipers, but could identify taxonomy markers, which can be employed for an unambiguous differentiation. To date, the comparison of venom compositions from related species in the context of taxonomic classification is still a controversial debate.[326] In the following, we compare the venom composition of the two related Nose-Horned vipers, Vat and Vam, at the level of subspecies that may aid in recognizing of a close relationship, which could help to a potential taxonomic assessment.

The close phylogenetic relationship of the two studied snakes was previously implied by Ursenbacher et al.[314] using molecular phylogeography. Our study attempts to underpin the implied taxonomic status of Vat by comparative analysis at the venomic level, underscored by the comparison of the UV chromatograms (**Figure 4.3**) as well as the TICs (**Figure 4.2**) of the venoms. The curve progression and peak distribution in the chromatograms show a superimposable arrangement containing the same toxin families, which was confirmed by SDS-PAGE followed by similar *de novo* sequencing as described above (**Table 4.1** and **Appendix Tables 4.1** & **4.2**).

A closer look at the intact mass profiling exhibits the svMP-i pEKW and pENW as most abundant peptide part in both venoms, but, with ca. 7.0% and 1.0%, they are more prominent in Vam then

in Vat (3.1% and 0.4%). On the other hand, the major peptide of fraction 11 (Vat *m/z* 3796.75 and Vam *m/z* 3796.73) with 0.9% to 0.4% is twice as abundant in Vat, as *m/z* 1143.64 (Vat) and *m/z* 1144.43 (Vam) as major Vat peptide in fractions 12, 13 and 14. In total, the snakes have eight close related peptide masses. This shows that the lower molecular masses are in the main contents similar, but differ strongly between the studied venoms in the lower abundant peptides and mostly in the abundance of 3–9 kDa masses. Additionally, the intact mass profiling exemplarily revealed six proteins from the venoms of Vat and Vam that are either in part or fully identical between these two vipers, as well as matched database entries for other species members (**Figure 4.7**).
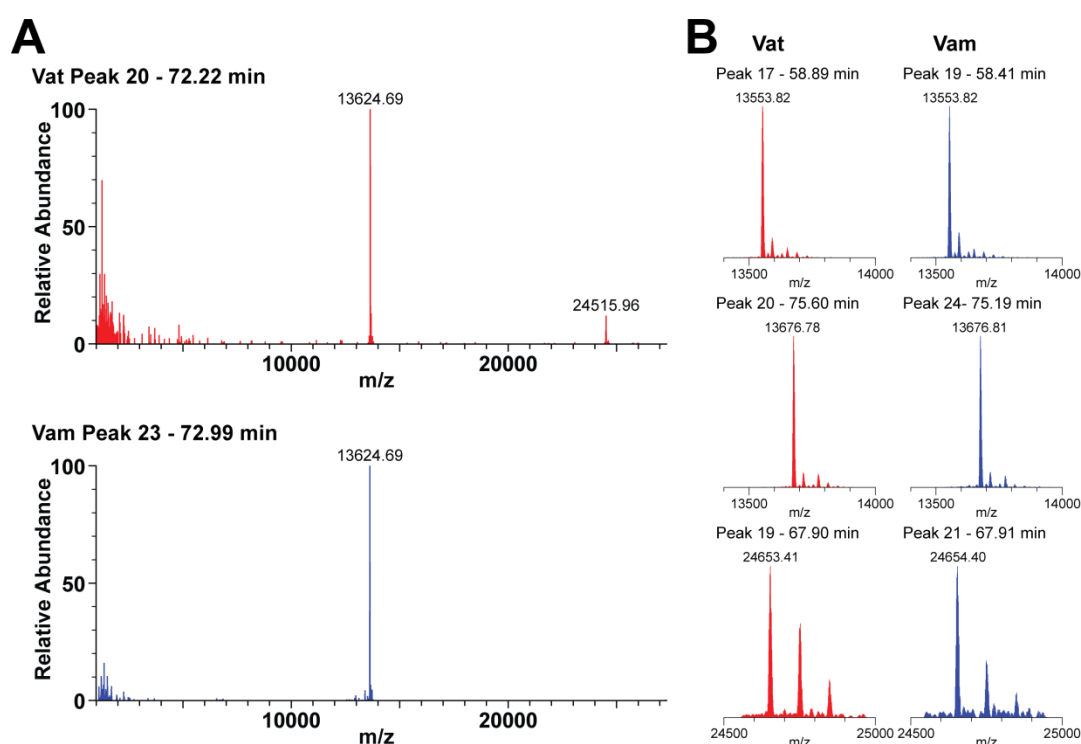
Figure 4.7. Comparative venom analysis of *V. a. transcaucasiana* and *V. a. montandoni*. Intact mass profiling of crude venoms from Vat (red) and Vam (blue) shows several identical masses [M + H]$^+$ of different toxin families: (**A**) Vipoxin A chain in peak 20 (Vat) and peak 23 (Vam); and (**B**) from top to bottom: two phospholipase A$_2$ (PLA$_2$, ~13–14 kDa) in peaks 17 and 20 (Vat) and in peaks 19 and 24 (Vam), as well as one CRISP (~25 kDa) in peak 19 (Vat) and peak 21 (Vat).

Two molecular masses ~24 kDa were observed each in peak 19 of Vat (24,652.41 Da and 24,751.38 Da) and peak 21 of Vam (24,653.40 Da and 24,749.41 Da), which were both determined by *de novo* sequencing as CRISPs. The remaining four molecular masses ~13 kDa belong to the PLA$_2$ family and were found in the strong peaks 17 and 20 of Vat (13,552.83 Da, 13,589.76 Da, 13,623.69 Da and 13,675.78 Da), and in peaks 19, 23 and 24 (13,552.82 Da, 13,589.75 Da, 13,623.69 Da and 13,675.81 Da) of Vam (**Appendix Figures S4.6–S4.13**), respectively.

The previously mentioned molecular mass of 13,623.69 Da identified from Vat (peak 20) and Vam (peak 23) was assigned to the closely related acidic phospholipase A$_2$ homolog Vipoxin A chain

(Uniprot-ID: P04084, including oxidized cysteines) M_{av} = 13,625.04 Da of the *V. a. meridionalis*.[333] Until now, Vipoxin was known as one of the most abundant components from the two other *Vipera ammodytes* subspecies.[316] In both tested venoms, these fractions were also the most abundant peaks, which underscore the importance of this toxin in the general venom composition. Additionally, by *de novo* sequencing, we identified fragments of the Vipoxin B chain in venoms of Vat (peak 18) and Vam (peak 20). The intact mass profiling further shows only for Vat (peak 18) a molecular mass of 13,813.21 Da that correlates to the average mass of the basic phospholipase A_2 Vipoxin B chain (Uniprot-ID: P14420, including oxidized cysteines) M_{av} = 13,813.77 Da of the *V. a. meridionalis*. Furthermore, the measured molecular mass of 13,552.8 Da in both venoms mass-correlates with the neutral phospholipase A_2 Ammodytin I2 (Uniprot-ID: P34180, including oxidized cysteines) M_{av} = 13,553.30 Da of *V. a. ammodytes*.[334] The presence of the same metalloprotease inhibitors and identical fragments of the precursor in both snake venoms, mentioned before, are further indicators of a close kinship.

Considering the study by Georgieva et al.[316] on the venom proteomes of *V. a. ammodytes* and *V. a. meridionalis*, we could compare our datasets in a wider context of the venoms of these subspecies. Georgieva et al. showed that several PLA_2s constitute an important part of the venoms, which we also determined in high quantities, e.g., Vammin A, several Ammodytin variants and Vipoxin B.[316] In addition, from venoms of Vat and Vam, we could also identify several venom families shared between the other *V. ammodytes* subspecies members, such as PLA_2, LAAO, growth factors as well as serine- and metalloproteinases. All these findings, which are solely based on the comparative venom analysis of four snakes of the same species, could be an indicator for a closer relationship and help to make a classification of the four snakes of *Vipera ammodytes* as subspecies comprehensible.

On the other hand, a view to other related *Vipera* species that are sharing the geographical habitat with the *Vipera ammodytes* species in parts show remarkable differences in the venom composition. For example, the wide distributed *Vipera berus berus*, which was recently compared to the *V. a. ammodytes*, exhibit svSP (31%) as the main toxin family followed by svMP (19%). In contrast, the most abundant families of the Vat and Vam play a subordinate role in the *V. b. berus* venom with 10% for PLA_2 or were not even detected in the case of the VEGF-F.[323] Furthermore, the venom composition of the close related *Vipera* species *Vipera anatolica* is focused on the presence of the toxin families svMP (42%) and CRISP (16%) with a similar occurrence of PLA_2 and VEGF-F to *V. b. berus*.[233]

Even if there are many similarities in the comparative analysis of the Vat and Vam venoms, some small differences remain, especially in the peptide content and in the protease pattern as well as in the LAAOs, which could be considered as parameters to distinguish between subspecies. However, venom compositions are susceptible to variation due to the influence of various factors (e.g., age, diet, sex and geographic origin) and could be more likely attributed to the occurrence of

intraspecific variations.[327–331] The venom variations influenced by diet or habitat are not suspected in the cases of *V. a. transcaucasiana* and *V. a. montandoni*, whose analyzed specimens share the same geographical origin, but due to the small pooled sample size of each subspecific population, complete variation compensation cannot be excluded. Additionally, the Vat mass profile shows two dominant peaks (peak 17 and 18), which includes PLA_2s (13,813.21 Da and 13,917.24 Da) that could not be detected in the Vam venom profile. In contrast, the Vam venom contains two abundant peaks (peaks 20 and 22) containing a PLA_2 (13,889.25 Da) and a CRISP (24,546.04 Da) as a major difference that are missing in the Vat profile (**Figures 4.2** and **4.4**). These small differences, even between subspecies, play a crucial role in the development of effective antidotes and the understanding of reduced effects of polyvalent antivenoms. For example, a study of *V. a. ammodytes* antidotes has shown strong reduced neutralization potency against *V. a. montandoni*.[319]

Cytotoxicity screening

Snake venoms constitute complex mixtures of enzymes, peptides and proteins with a high toxicity potential, which can selectively and specifically act on various cellular targets by modulating the physiological function. This turns snake venoms into an attractive source for potential anticancer agents.[335] As part of our ongoing studies on Turkish snake venoms, the potency against various human cancer cells, and the cytotoxicities of the crude venom for *V. a. transcaucasiana* and *V. a. montandoni* were tested on a panel of cancer cell lines together with non-cancerous cell lines in a MTT assay. For *V. a. transcaucasiana* and *V. a. montandoni* crude venoms the MTT assay resulted as IC_{50} values of 1.34–22.75 µg/mL and 0.06–50.00 µg/mL (**Table 4.2**), respectively.

The crude venoms of both snakes show a similar activity against breast (MDA-MB-231), colon (Caco-2) and bladder (253J-BV) cancer cell lines. *V. a. montandoni* shows a high cytotoxicity against four cell lines (HEK-293, U-87 MG, A549 and HeLa) out of the eight tested (**Figure 4.8**). The determination of the IC_{50} shows strong differences of the closely related vipers against the MCF-7 breast cancer cells and the A549 lung cancer cells. The *V. a. transcaucasiana* venom was found to have an IC_{50} of 21.75 ± 2.45 µg/mL and 18.03 ± 2.09 µg/mL against MCF-7 and A549 cells, respectively. In contrast, the *V. a. montandoni* venom has an $IC_{50} > 50.00$ µg/mL and 4.40 ± 0.03 µg/mL against MCF-7 and A549 cells, respectively (**Table 4.3**). Interestingly, the *V. a. transcaucasiana* crude venom exhibited the highest cytotoxic effect (1.84 ± 0.76 µg/mL) against the receptor triple negative breast cancer cells MDA-MB-231 in comparison to *V. a. montandoni* and the 2.6-fold less active positive control (4.80 ± 1.10 µg/mL), while both venoms are less toxic to the second tested breast cancer cells MCF-7 (ER negative, PR positive, HER2 positive) with IC_{50} values of 21.75 ± 2.45 µg/mL and >50.00 µg/mL to Parthenolide (6.01 ± 1.15 µg/mL). The subtypes of breast cancer have been generally identified based on the presence of three receptors:

estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor-2 (HER-2).[336]

Table 4.2. IC$_{50}$ values of *Vipera ammodytes transcaucasiana* and *Vipera ammodytes montandoni* venoms against various human cell lines. The half maximal inhibitory concentration (IC$_{50}$ in µg/mL) for the venom of Vat and Vam were determined after 48 h exposure. Parthenolide was used as reference compound. Noncancerous human cells: HEK293 (human embryonic kidney). Cancerous human cells: U87MG (epithelial-like glioblastoma-astrocytoma); SHSY5Y (neuroblastoma); MDA-MB-231 (breast epithelial adenocarcinoma); A549 (lung adenocarcinoma); MPanc-96 (pancreas adenocarcinoma); MCF-7 (epithelial breast adenocarcinoma); CaCo-2 (epithelial colorectal adenocarcinoma); 253-JBV (bladder carcinoma); HeLa (epithelial cervical carcinoma); PC-3 (prostate adenocarcinoma). A minus (−) mentioned not tested cell lines and error in mean ± SD.

Cell line	*V. ammodytes transcaucasiana* IC$_{50}$ (µg/mL)	*V. ammodytes montandoni* IC$_{50}$ (µg/mL)	Parthenolide IC$_{50}$ (µg/mL)
HEK293	1.34 ± 0.72	3.55 ± 0.61	1.23 ± 0.24
U87MG	6.02 ± 1.38	1.02 ± 0.20	3.33 ± 0.59
SHSY5Y	-	0.06 ± 0.01	0.15 ± 0.01
MDA-MB-231	1.84 ± 0.76	2.36 ± 0.20	4.80 ± 1.10
MCF-7	21.75 ± 2.45	>50.00	6.01 ± 1.15
A549	18.03 ± 2.09	4.40 ± 0.03	4.42 ± 0.87
MPanc-96	22.75 ± 2.25	-	4.70 ± 0.87
CaCo-2	4.21 ± 0.96	1.82 ± 0.14	4.90 ± 1.10
253J-BV	4.22 ± 1.41	3.00 ± 1.98	5.45 ± 1.16
HeLa	6.14 ± 1.12	1.27 ± 0.20	5.75 ± 1.07
PC3	6.95 ± 1.19	-	3.33 ± 0.96

Based on these markers, a classification of breast cancer tumors as hormone receptor positive, HER-2/Neu amplified tumors, and those which do not express ER, PR and do not have a HER-2/Neu amplification were defined as triple-negative breast cancer (TNBC). This might have a significant role for patient-tailored treatment strategies, as TNBC demonstrates approximately 10–15% of all breast cancers and patients with TNBC have an unsuccessful outcome compared to the other breast cancer subtypes.[337] Unfortunately, the tested crude venoms also exhibit a comparable cytotoxic effect against the tested non-cancerous kidney cells with values of 1.34 ± 0.72 and 3.55 ± 0.61 µg/mL. The comparison of further non-cancerous to attributed cancerous cell lines is part of ongoing due to restricted sample size. Nevertheless, we suspect that the high toxic effect on non-cancerous cell lines can be overcome by developing a targeted drug delivery system for potential treatment followed by detailed studies for each drug candidate.
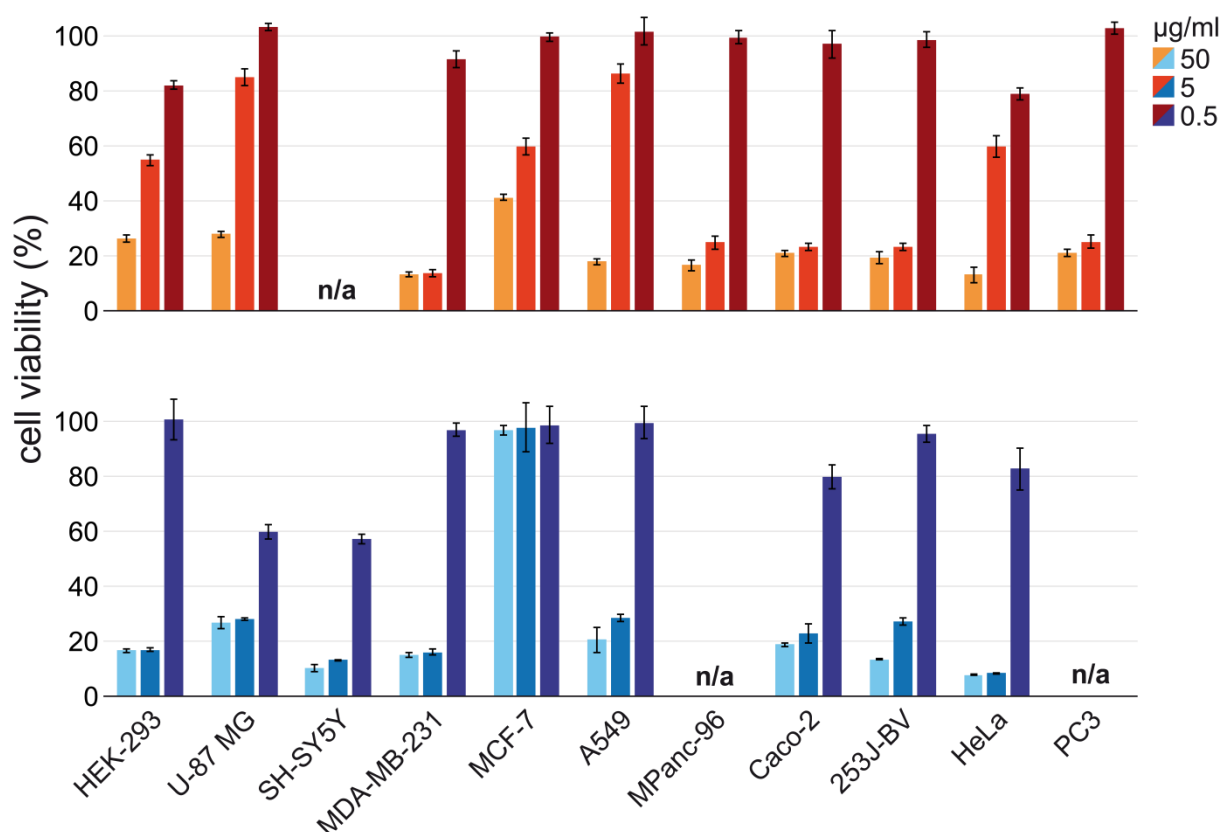
Figure 4.8. MTT human cell viability after 48 h crude venom treatment. Cytotoxicity of: *V. a. transcaucasiana* (red) (**top**); and *V. a. montandoni* (blue) (**bottom**) crude venom in three concentrations (50, 5 and 0.5 µg/mL) against different human cell lines. Cell viability was measured by MTT assay after 48 h at 570 nm. Noncancerous human cells: HEK-293 (human embryonic kidney). Cancerous human cells: U-87 MG (epithelial-like glioblastoma-astrocytoma); SH-SY5Y (neuroblastoma); MDA-MB-231 (breast epithelial adenocarcinoma); A549 (lung adenocarcinoma); MPanc-96 (pancreas adenocarcinoma); MCF-7 (epithelial breast adenocarcinoma); CaCo-2 (epithelial colorectal adenocarcinoma); 253-JBV (bladder carcinoma); HeLa (epithelial cervical carcinoma); PC-3 (prostate adenocarcinoma). Not tested cell lines by n/a and error in mean ± SD.

According to the MTT assay, *V. a. transcaucasiana* venom showed active venom fractions against the human breast adenocarcinoma epithelial cells (MDA-MB-231) (**Appendix Figure 4.14**). Fractions 1, 17, 18 and 19 exhibited the strongest cytotoxic effect on the triple negative MDA-MB-231 breast cancer cells with IC$_{50}$ of 2.96 µg/mL to 9.22 µg/mL (**Figure 4.9** and **Table 4.3**). Mass spectrometric analysis of the bioactive fractions identified fraction 17 as an Ammodytin I2(A) variant, fraction 18 as a Vipoxin chain B or Vaspin basic subunit variant and fraction 19 as a cysteine-rich venom protein. Fraction 1 is a mixture of various small peptides with still unknown sequence. The effectivity and specificity of snake venom phospholipases against cancer cells are known.[338] The cytotoxicity of the CRISP in fraction 19 is due to the blocking ability against several ion channels.[339,340]

Table 4.3. IC$_{50}$ values of selected *V. a. transcaucasiana* HPLC fractions against MDA-MB-231 cells. The half maximal inhibitory concentration (IC$_{50}$ in µg/mL) for the HPLC fractions 1, 17, 18 and 19 of Vat venom against MDA-MB-231 breast adenocarcinoma epithelial cell line. Doxorubicin was used as reference compound and error in mean ± SD.

Samples ID	*V. ammodytes transcaucasiana* Fraction				Doxorubicin
	1	**17**	**18**	**19**	
IC$_{50}$ (µg/mL)	5.92 ± 0.14	3.98 ± 0.85	2.96 ± 0.38	9.22 ± 0.62	>20.00
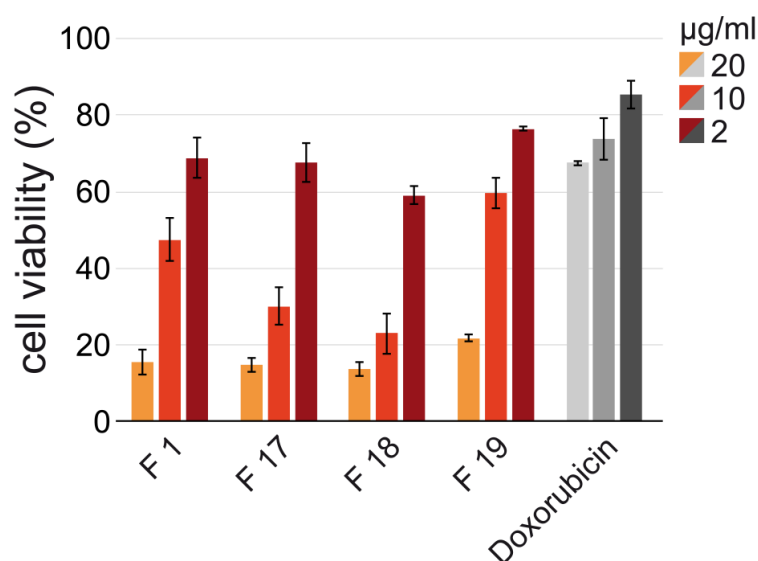
Figure 4.9. MTT human cell viability after 48 h *V. a. transcaucasiana* venom fraction treatment. Cytotoxicity of selected *V. a. transcaucasiana* venom HPLC fractions (F1,17–19) in three concentrations (20 µg/mL, 10 µg/mL and 2 µg/mL) against MDA-MB-231 breast adenocarcinoma epithelial cell line. Cell viability was measured by MTT assay after 48 h at 570 nm. Doxorubicin was used as reference compound and error in mean ± SD.

Conclusion

Here, we report on the first proteomic characterization of the venoms from *Vipera ammodytes transcaucasiana* and *Vipera ammodytes montandoni* by using a combined mass spectrometry-guided approach. The initial intact mass profiling of the venoms facilitated the detection of ~50 venom components for *V. a. transcaucasiana* and ~59 venom components for *V. a. montandoni*. Additionally, the intact mass profiling revealed the presence of two tripeptide metalloprotease inhibitors and their precursors in the venoms, which would not have been detected by the bottom-up approach. However, due to the limited applicability to high molecular mass compounds the analysis was further expanded to the bottom-up approach. The *de novo* sequencing showed for both snake venoms the presence of 11 major *Viperidae* toxin families with the exception of Kunitz type proteinase inhibitors and hyaluronidases, which are not present in either venom.

The comparative analysis by initial mass profiling in combination with the well-established bottom-up protocol revealed strong similarities in the venom composition of the two studied

snake venoms of *V. a. transcaucasiana* and *V. a. montadoni*. While related toxin families or even identical proteins could be identified proving a close relationship of the proteome and under consideration of intraspecific venom variation, small differences in the venom compositions in turn could be parameters for a differentiation into subspecies.

In summary, the mass spectrometry-guided comparative analysis of the venom proteome extended by intact mass profiles provides an excellent method to highlight close relationships of venomous snakes at the proteome level and especially in combination with top-down venomics this would give you an even more detailed picture of venom diversity. Hence, the investigation of venoms for the phylogenetic analysis, using a combination of chromatographic, electrophoretic and different mass spectrometric techniques, is very sensitive and fast. In this context, reliable databases are an important basis for the *de novo* identification and the variation of the venom composition by differences in age and food supply should always be considered as a critical point.[341] The future transcriptomic analysis of the venom glands represents a valuable technique, which could support the proteomic analysis and elucidate the venom proteome in its entirety. The complete analysis of the venom proteome of *Vipera ammodytes transcaucasiana* in connection with mitochondrial DNA could finally clarify the controversial question about the taxonomic status.

Finally, the preliminary *in vitro* cytotoxicity screening against various cancer cell lines for the crude venoms and components for *V. a. transcaucasiana* demonstrated significant cytotoxic effects on the triple negative MDA MB 231 breast cancer cells with IC_{50} of 2.96 µg/mL to 9.22 µg/mL for an Ammodytin I2(A) variant, a Vipoxin chain B or Vaspin basic subunit variant and a cysteine-rich venom protein. Nevertheless, the high toxic effect also against non-cancerous cell lines requires the development of a targeted drug delivery system for potential treatment.

Material and Methods

Collection and preparation of venom samples

Venom samples of the *V. a. transcaucasiana* were collected in June 2015 from four specimens in total, one from Işık Mountain (Çerkeş district, Çankırı Province, Turkey) and three from Çavuş Mountain (Sivas Province, Turkey). *Vipera ammodytes montandoni* venom samples were collected from one individual in the Tekirdağ province and one in the Kırklareli province (Turkish Thrace) in April 2016. Crude venoms were extracted, using a paraffin-covered laboratory beaker without exerting pressure on the venom glands, pooled for each subspecies and lyophilized. Ethical permission (Ege University, Animal Experiments Ethics Committee, 2010#43) and special permission (2011#7110) for field studies from the Republic of Turkey, Ministry of Forestry and Water Affairs were received.

Determination of protein content

Protein concentrations were determined from diluted venom sample (1 or 2 mg/mL) in ultrapure water by Micro-BCA (Bicinchoninic Acid) Protein Assay using a UV/Vis spectrophotometer (Thermo-Scientific, Darmstadt, Germany) at a wavelength of $\lambda = 595$ nm. Bovine serum albumin (BSA) was used as a reference.

Cell culture and *in vitro* cytotoxicity assay

The following human cell lines were used for determination of cytotoxicity: noncancerous cells: HEK293 (human embryonic kidney); cancerous human cells: U87MG (epithelial-like glioblastoma-astrocytoma); SHSY5Y (neuroblastoma); MDA-MB-231 (breast epithelial adenocarcinoma); A549 (lung adenocarcinoma); MPanc-96 (pancreas adenocarcinoma); MCF-7 (epithelial breast adenocarcinoma); CaCo-2 (epithelial colorectal adenocarcinoma); 253-JBV (bladder carcinoma); HeLa (epithelial cervical carcinoma); PC-3 (prostate adenocarcinoma). All cell lines were purchased from the US American Type Culture Collection (ATCC, Manassas, VA, USA) except for 253J-BV cells, which were obtained from Creative Bioarray (Shirley, NY, USA). All cells were cultivated in Dulbecco's modified Eagle's medium F12 (DMEM/F12), supplemented with 10% fetal bovine serum (FBS), 2 mM/L glutamine, 100 U/mL of penicillin and 100 mg/mL of streptomycin (Gibco, Visp, Switzerland). The *in vitro* cytotoxicity testing with crude venoms and fractions was performed according to the protocol of Nalbantsoy and Hempel et al.[234]. The morphological changes of the cells after treatment with crude venom or its fractions were observed with an inverted microscope (Olympus, Tokyo, Japan) compared to the control group following for 48 h.

Determination of half maximal inhibitory concentration (IC$_{50}$)

The IC$_{50}$ values were calculated by fitting the data to a sigmoidal curve and using a four-parameter logistic model and presented as an average of three independent measurements. The IC$_{50}$ values were reported at 95% confidence interval, and calculations were performed using Prism 5 software (GraphPad5, San Diego, CA, USA). The values of the blank wells were subtracted from each well of treated and control cells and half maximal inhibition of growth (IC$_{50}$) were calculated in comparison to the untreated controls.

Preparation of venom samples for intact mass profiling

The crude venoms were dissolved in aqueous 1% (*v/v*) formic acid (HFo) to a final concentration of 10 mg/mL, and centrifuged at 20,000×g for 5 min to spin down insoluble content. Dissolved venoms were then mixed with 30 μL of citrate buffer (0.1 M, pH 4.0). The samples were mixed with an equal volume of 1% aqueous formic acid and centrifuged at 20,000×g for 5 min. Subsequently, samples were submitted to HPLC-high-resolution (HR) ESI-MS/MS measurements.

Intact mass profiling

The intact mass profiling was performed by LC-ESI-HR-MS experiments on an LTQ Orbitrap XL mass spectrometer (Thermo, Bremen, Germany) coupled to an Agilent 1260 HPLC system (Agilent, Waldbronn, Germany) using a Supelco Discovery 300 Å C18 (2 × 150 mm, 3 µm particle size) column. The instrument settings for the HPLC system and ESI-MS were adopted from Nalbantsoy and Hempel et al.[234]. The intact mass profiles were inspected with the Xcalibur Qual Browser (Thermo Xcalibur 2.2 SP1.48, Thermo Fisher Scientific, Waltham, USA) and deconvolution of isotopically resolved spectra was carried out by using the XTRACT algorithm of Xcalibur Qual Browser. The protein assignment was done by comparison to the retention time of the HPLC run and corresponding LC-MS/MS information from SDS-PAGE trypsin digests.

Bottom-up venomics

The lyophilized crude venoms (4 mg) were dissolved to a final concentration of 20 mg/mL in aqueous 3% (*v/v*) ACN with 1% (*v/v*) HFo and centrifuged at 20,000×g for 5 min to spin down insoluble content. The supernatant was loaded onto a semi-preparative reversed-phase HPLC with a Supelco Discovery BIO wide Pore C18-3 column (4.6 × 150 mm, 3 µm particle size) using an Agilent 1260 Low Pressure Gradient System (Agilent, Waldbronn, Germany). The column was operated with a flow rate of 1 mL/min and performed using ultrapure water with 0.1% (*v/v*) HFo (buffer A) and ACN with 0.1% (*v/v*) HFo (buffer B). The technical settings and bottom-up workflow was performed according to the protocol of Nalbantsoy and Hempel et al.[234] with the following amendments. After the chromatographic separation of the crude venoms, the vacuum-dried peak fractions were submitted to a SDS-PAGE with a content of 15% polyacrylamide. Afterwards the coomassie-stained band were excised and reduced, via in-gel trypsin digestion, with freshly prepared dithiothreitol solution (100 mM DTT in 100 mM ammonium hydrogencarbonate, pH 8.3, heated for 30 min at 56 °C) and alkylated with iodoacetamide (55 mM IAC in 100 mM ammonium hydrogencarbonate, pH 8.3, stored for 30 min at 25 °C in the dark). The peptides were extracted with 100 µL aqueous 30% (*v/v*) ACN just as 5% (*v/v*) HFo for 15 min at 37 °C. The supernatant was vacuum dried (Thermo SpeedVac, Bremen, Germany), re-dissolved in 20 µL aqueous 3% (*v/v*) ACN with 1% (*v/v*) HFo and submitted to HPLC-MS/MS analysis.

The bottom-up analysis was performed with an Orbitrap XL mass spectrometer (Thermo, Bremen, Germany) via an Agilent 1260 HPLC system (Agilent Technologies, Waldbronn, Germany) using a reversed-phase Grace Vydac 218MSC18 (2.1 × 150 mm, 5 µm particle size) column. The pre-chromatographic separation was performed with the following settings: After an isocratic equilibration (5% B) for 1 min, the peptides were eluted with a linear gradient of 5–40% B for 10 min, 40–99% B for 3 min, washed with 99% B for 3 min and re-equilibrated in 5% B for 3 min. LC-MS/MS data files (.raw) were converted to mgf-files using MSConvert GUI of the ProteoWizard Software Foundation (ProteoWizard package, version 3.0.10577, open-source platform for

proteomic data analysis, California, USA) and annotated by DeNovo GUI[342] (ProteoWizard package, version 1.15.8, open-source platform for proteomic data analysis, California, USA) with the following settings: fixed modifications: carbamidomethyl Cys (+57.02 Da); variable modifications: acetylation of Lys (+42.01 Da) and phosphorylation of Ser and Thr (+79.97 Da). The deduced amino acid sequences were squared against a non-redundant protein NCBI database of *Viperidae* (taxid:8689) using BLASTP[343] (http://blast.ncbi.nlm.nih.gov).

Data accessibility

Mass spectrometry proteomics data (.mgf, .raw and output files) have been deposited with the ProteomeXchange Consortium[344] (http://proteomecentral.proteomexchange.org) via the MassIVE partner repository under Project Name "Venomics of the *V. a. transcaucasiana* and *V. a. montandoni*" and data set identifier PXD007609.

Relative toxin quantification

The quantification of venom composition is based on the RP-HPLC peak integration (UV_{214nm}) in comparison to the total integral of all analyzed peaks based on the protocol of Juárez et al.[223]. In the case of HPLC co-eluting toxins components, the SDS-PAGE band ratio of optical intensities and densities was, respectively, used for emphasis of peak integral portion.[228,294,300]

5 Intact Protein Mass Spectrometry Reveals Intraspecies Variations in Venom Composition of a Local Population of *Vipera kaznakovi* in Northeastern Turkey

Introduction

Venomics is considered an integrative approach, that can combine proteomics, transcriptomics and/or genomics to study venoms.[345] Although the term was initially used to describe the mass spectrometry-based proteomic characterization of venoms[223,229], genomic[346] or more commonly venom gland transcriptomic sequencing[232,347,348,349–351] have also been used to characterize venom compositions. These molecular approaches provide an overview of venom composition by characterizing the nucleotide sequences of venom toxin-encoding genes (among others) and, in the case of transcriptomics, also provide an estimation of their relative expression in the venom gland. Furthermore, (translated) protein sequence databases are crucial for the robust annotation of tandem mass spectra from proteomic analyses in peptide/protein spectrum matching (P*r*SM). A bibliographic search to the keyword "Snake venomics" in PubMed identified 147 hits between 2004 and 2018, including a rapid expansion in the application of venomics approaches in more recent years.

Initial proteomic analyses of snake venoms included the combination of multidimensional separation techniques (chromatographic and gel electrophoresis), N-terminal Edman degradation, and *de novo* sequencing by tandem mass spectrometry of tryptic peptides generated by in-gel digestion of SDS-PAGE bands.[223,352] Since these initial studies, the proteomic characterization of snake venoms has become more comprehensive due to technical advances in mass spectrometry and next generation nucleotide sequencing. Several complementary strategies were developed to unveil the venom proteomes of more than 100 snake species.[353] Most of these studies applied the so called 'bottom-up' proteomic approaches, whereby intact proteins are typically digested with trypsin before tandem mass spectrometry analysis. Many workflows perform venom decomplexation prior to digestion, either by liquid chromatography (LC) or gel electrophoresis, or a combination of both.[225] The direct, in-solution digestion, or so called 'shotgun proteomics', allows for a fast qualitative overview, but suffers from a less quantitative breakdown of snake venom composition.[225,354] For example, in shotgun experiments, the problem of protein inference often does not permit the differentiation of the numerous toxin isoforms present in venom.[355] Thus, chromatographic or electrophoretic separation of venom samples greatly aids in differentiating between toxin isoforms (encoded by paralogs). In addition,

decomplexing prior to trypsin digestion often does not allow for the clear identification of differential post-translational modified variants, so-called proteoforms.[356]

To circumvent these limitations, a logical solution would be the of the digestion step and the application of direct analysis of intact proteins by tandem mass spectrometry, so called top-down proteomics. Recently top-down protein analysis has been applied alone, or in combination with other venomics approaches, to study the venoms of the King cobra (*Ophiophagus hannah*)[227,299], the entire genus of mambas (*Dendroaspis* spp.)[231,357], the brown tree snake (*Boiga irregularis*)[232], the Okinawa habu pit viper (*Protobothrops flavoviridis*)[358], and several viper species from Turkey[233,234,359]. In the case of viperid species, top-down analysis typically only results in partial characterization of the venom, as a number of the main toxin components, such as high molecular weight snake venom metalloproteinases (svMPs) (>30 kDa), are challenging to efficiently ionize by denaturing electrospray ionization (ESI) and might only provide few observable fragments in tandem MS.[360] A possible way to overcome difficulties in terms of ionization of high molecular weight proteins is the application of native ESI, as described by Melani et al.[299]. However native top-down mass spectrometry typically requires a special type of mass spectrometer with extended mass range and more extensive sample preparation, which makes this type of analysis more technically challenging.

In the majority of the aforementioned studies, the top-down workflows were performed with a front-end LC-based sample decomplexation. This allows for the generation of MS1 mass profiles (XICs) of intact proteoforms. Typically, the MS1 information is accompanied by tandem MS (MS2) information acquired in data-dependent acquisition (DDA) mode. The MS2 fragment spectra are than matched to a translated transcriptome/genome database in order to identify the proteins. In the case that there are not enough MS2 fragment peaks of a particular proteoform, the intact molecular mass can still enable identification, especially if the intact mass can be associated to masses observed in complementary experiments, such as retention time, mass range of SDS-PAGE and/or bottom-up protein IDs of decomplexed bottom-up venomics.[233] The additional information gained through exact intact protein masses can be particularly informative to differentiate between isoforms or proteoforms. As several studies have shown correlations between different ecological, geographical, genetic and/or developmental factors and the venom proteome, e.g. different diets[326,328,329,361], regional separation of populations[341,350,362], sex[331,363] or age[330,364] it would be ideal to characterize venom composition from a representative cohort of individual animals. However, due to the time consuming nature and expensive costs associated with de-complexing bottom-up venom analysis, most venom compositions reported so far[353], were analyzed from a single pool of venom that was sourced from different numbers of individuals. In top-down approaches on the other hand, the simple sample preparation, high sensitivity and fast analysis time allows for a rapid and cheaper comparison of venom composition, and thus seems well suited for large scale quantitative comparisons[365] of snake

venoms. In addition to providing a better biological understanding of the toxins present in venom[366], and the evolutionary processes underpinning population level venom variations[367], population wide venom analyses would provide important information to better understand regional and intraspecific variations in the venom composition of medically important snake species, which has considerable relevance for the development and clinical utility of antivenom treatment for snakebite.[368,369]

In this study, we explored the utility of top-down intact mass profiling to identify intraspecific venom variation by applying it to a local population of the medical relevant Caucasus viper (*Vipera kaznakovi*). The Caucasus viper is a subtropical, medium-sized, viper species with a wide distribution range from the Caucasus Black Sea coastal provinces of Artvin and Rize in northeastern Turkey (**Figure 5.1**), through Georgia to Russia. A distinctive characteristic of this species is the black coloration with elements of an orange to red zigzag-looking on the dorsal side of the body (**Figure 5.1**). This species feeds predominately on small vertebrates (mice, lizards etc.) and insects.[370]

In a previous shotgun proteomics study of this species, Kovalchuk and coworkers described the venom of *V. kaznakovi* (Krasnodar Territory, near Adler, Russia) to be composed of phospholipase A_2 (PLA$_2$, 19.0%), snake venom metalloproteases (svMP, 16.2%), snake venom serine proteases (svSP, 10.8%), Cysteine-rich secretory proteins (CRISP, 9.7%), C-type lectins (CTL, 12.5%), L-amino acid oxidase (LAAO, 4.0%), vascular endothelial growth factor (VEGF, 4.0%), disintegrins (Dis, 0.5%), phospholipase B (PLB, 0.3%), nerve growth factors (NGF, 0.14%), as well as a number of other venom proteins of lower abundance.[371] Here, we used a combination of venom gland transcriptomics, decomplexing bottom-up proteomics and comparative top-down proteomics to provide a more detailed characterization of the venom composition of *V. kaznakovi*, and to gain first insights into intraspecies variation of its venom composition. Furthermore, our findings highlight the potential of intact protein mass profiling for future population level studies of viperid venoms.

Material and Methods

Sampling

Venom samples of *V. kaznakovi* were collected from 6 adult (2 female, 4 male) and 3 juvenile specimens (unknown sex). All specimens were captured in late June 2015 in their natural habitat and released back into their natural environment after venom extraction. The *V. kaznakovi* individuals were collected in Artvin province in Turkey near the Georgian border, with 6 individuals sampled from Hopa district, 2 individuals from Borçka district and 1 specimen in the Arhavi district. An additional female individual found in Borçka district was collected for venom gland dissection for transcriptomic analysis. Ethical permission (Ege University Animal Experiments Ethics Committee, 2013#049) and special permission (2015#124662) for the

sampling of wild-caught *V. kaznakovi* were received from the Republic of Turkey, Ministry of Forestry and Water Affairs. Geographic coordinates for all *V. kaznakovi* individuals are provided in appendix table 5.1.

Sample storage and preparation

Crude *V. kaznakovi* venom was extracted by using a parafilm-covered laboratory beaker without exerting pressure on the venom glands. Venom samples were centrifuged at 2000×g for 10 min at 4 °C to remove cell debris. Supernatants were immediately frozen at −80 °C, lyophilized, and the samples stored at 4 °C until use.

Determination of lethal dose (LD$_{50}$)

The lethal potency (LD$_{50}$) of pooled *V. kaznakovi* venom to mice (mg/kg) was determined by an up-and-down method as recommended by the Organization for Economic Cooperation and Development (OECD) guidelines (Test No. 425).[372] Groups of five Swiss albino mice (n = 15; age, 8 to 10 weeks; female 8 and male 7 individuals) were used per venom dose. Various venom concentrations (5, 2 and 1 mg/kg, milligrams of protein per kg calculated from dry weight venom by Bradford assay) were diluted in ultrapure water to a final volume of 100 μL and injected by the intraperitoneal (IP) route. Control mice (n = 5; female 2 and male 3 individuals) received a single IP injection of sterile saline (0.9%, 100 μL). All assays and procedures involving animals strictly followed the ethical principles in animal research adopted by the Swiss Academy of Medical Sciences.[373] Additionally, they were approved by a local ethics committee (2013#049). Mortality was recorded 24 h after injection, and the median lethal dose was determined by a nonlinear regression fitting procedure in GraphPad Prism 5 (Version 5.01, Inc., San Diego, CA, USA).

RNA isolation and purification

Venom glands were dissected from a wild caught adult female specimen of *V. kaznakovi* in Kanlıdere, Hopa district (Artvin province) and processed as previously described.[349,357] Briefly, immediately following euthanasia, venom glands were dissected and were immediately flash frozen in liquid nitrogen and stored cryogenically prior to RNA extraction. Venom glands were next homogenized under liquid nitrogen and total RNA extracted using a TRIzol Plus RNA purification kit (Invitrogen), DNAse treated with the PureLink DNase set (Invitrogen) and poly(A) selected using the Dynabeads mRNA DIRECT purification kit (Life Technologies), as previously detailed.[349,357]

RNA sequencing, assembly and annotation

RNA-Seq was performed as described in earlier studies.[349,357] The RNA-Seq library was prepared from 50 ng of enriched RNA material using the ScriptSeq v2 RNA-Seq Library

Preparation Kit (epicenter, Madison, WI, USA), following 12 cycles of amplification. The resulting sequencing library was purified using AMPure XP beads (Agencourt, Brea, CA, USA), quantified using the Qubit dsDNA HS Assay Kit (Life Technologies), before the size distribution was assessed using a Bioanalyser (Agilent). The library was then multiplexed and sequenced (alongside other sequencing libraries not reported in this study) on a single lane of an Illumina MiSeq (250 bp paired end sequencing), housed at the Centre for Genomic Research, Liverpool, UK. The *V. kaznakovi* library amounted to 1/6th of the total sequencing lane. The ensuing read data was quality processed by (i) removing the presence of any adapter sequences using Cutadapt (https://code.google.com/p/cutadapt/) and (ii) trimming low quality bases using Sickle (https://github.com/najoshi/sickle). Reads were trimmed if bases at the 3' end matched the adapter sequence for 3 bp or more, and further trimmed with a minimum window quality score of 20. After trimming, reads shorter than 10 bp were removed.

For sequence assembly we used VTBuilder, a *de novo* transcriptome assembly program previously designed for discriminating between multiple related toxin isoforms during the construction of snake venom gland transcriptomes[374], which has previously been utilized for integrating venom toxin gene data with venom proteomic data. [349,357] Paired-end read data was entered into VTBuilder and executed with the following parameters: min. input read length 150 bp; min. output transcript length 300 bp; min. isoform similarity 96%. Assembled contigs were annotated with BLAST2GO Pro v3[375] using the blastx-fast algorithm with a significance threshold of 1e-5, to provide BLAST annotations (max 20 hits) against NCBI's non redundant (NR) protein database (41 volumes; Nov 2015) followed by mapping to gene ontology terms, and Interpro domain annotation using default parameters. Following generic annotation, venom toxins were initially identified based on their BLAST similarity to sequences previously identified in the literature or in molecular databases as snake venom toxins, and then manually curated for validation.

Venom proteomics (bottom-up)

The crude venom (1 mg) was dissolved to a final concentration of 10 mg/mL in aqueous 3% (v/v) acetonitrile (ACN) with 1% (v/v) formic acid (FA) and centrifuged at 16,000×g for 5 min to spin down insoluble content. The supernatant was loaded onto reversed-phase HPLC with a Supelco Discovery BIO wide Pore C18-3 column (4.6 x 150 mm, 3 μm particle size) using an Agilent 1260 Low Pressure Gradient System (Agilent, Waldbronn, Germany). The column was operated with a flow rate of 1 mL/min and run with ultrapure water (solution A) and ACN (solution B), both including 0.1% (v/v) FA. A standard separation gradient was used with solution A and solution B, starting isocratically (5% B) for 5 min, followed by linear gradients of 5-40% B for 95 min and 40-70% for 20 min, then 70% B for 10 min, and finally re-equilibration at 5% B for 10 min. Peak detection was performed at λ = 214 nm using a diode array detector (DAD). After the chromatographic separation of the crude venom, the collected and vacuum-dried peak fractions

were submitted to SDS-PAGE gel electrophoresis (12% polyacrylamide). Subsequently, the coomassie-stained bands were excised, and submitted to in-gel trypsin digestion, reduced with fresh dithiothreitol (100 mM DTT in 100 mM ammonium hydrogencarbonate, pH 8.3, for 30 min at 56 °C) and alkylated with iodoacetamide (55 mM IAC in 100 mM ammonium hydrogencarbonate, pH 8.3, for 30 min at 25 °C in the dark). The resulting peptides were then extracted with 100 µL aqueous 30% (v/v) ACN with 5% (v/v) FA for 15 min at 37 °C. The supernatant was vacuum-dried (Thermo SpeedVac, Bremen, Germany), redissolved in 20 µL aqueous 3% (v/v) ACN with 1% (v/v) FA and submitted to LC-MS/MS analysis.

The bottom-up analyses were performed with an Orbitrap XL mass spectrometer (Thermo, Bremen, Germany) via an Agilent 1260 HPLC system (Agilent Technologies, Waldbronn, Germany) using a reversed-phase Grace Vydac 218MSC18 (2.1 x 150 mm, 5 µm particle size) column. The pre-chromatographic separation was performed with the following settings: after an isocratic equilibration (5% B) for 1 min, the peptides were eluted with a linear gradient of 5-40% B for 10 min, 40-99% B in 3 min, held at 99% B for 3 min and re-equilibrated in 5% B for 3 min.

Population level venom profiling (top-down)

The top-down MS analysis was performed by dissolving the crude venoms in ultrapure water containing formic acid (FA, 1%) to a final concentration of 10 mg/mL, and centrifuged at 20,000×g for 5 min. Aliquots of 10 µL dissolved venom samples were submitted to reverse-phase (RP) HPLC-high-resolution (HR)-MS analyses. RP-HPLC-HR-MS experiments were performed on an Agilent 1260 HPLC system (Agilent, Waldbronn, Germany) coupled to an Orbitrap LTQ XL mass spectrometer (Thermo, Bremen, Germany). RP-HPLC separation was performed on a Supelco Discovery Biowide C18 column (300 Å pore size, 2 × 150 mm column size, 3 µm particle size). The flow rate was set to 0.3 mL/min and the column was eluted with a gradient of 0.1% FA in water (solution A) and 0.1% FA in ACN (solution B): 5% B for 5 min, followed by 5–40% B for 95 min, and 40–70% for 20 min. Finally, the gradient was held isocratic with 70% B for 10 min and re-equilibrated at 5% B for 10 min. ESI settings were: 11 L/min sheath gas; 35 L/min auxiliary gas; spray voltage, 4.8 kV; capillary voltage, 63 V; tube lens voltage, 135 V; and capillary temperature, 330 °C. MS/MS spectra were obtained in data-dependent acquisition (DDA) mode. FTMS measurements were performed with 1 µ scans and 1000 ms maximal fill time. AGC targets were set to 10^6 for full scans and to 3×10^5 for MS/MS scans, and the survey scan as well as both data dependent MS/MS scans were performed with a mass resolution (R) of 100,000 (at m/z 400). For MS/MS the two most abundant ions of the survey scan with known charge were selected. Normalized CID energy was set to 30% for the first, and 35% for the second, MS/MS event of each duty cycle. The default charge state was set to z = 6, and the activation time to 30 ms. Additional HCD experiments were performed with 35% normalized collision energy, 30 ms activation time and z = 5 default charge state. The mass window for precursor ion selection was set to m/z 2 or 6.

A window of m/z 3 was set for dynamic exclusion of up to 50 precursor ions with a repeat of 1 within 10 s for the next 20 s.

Bioinformatic analysis

The LC-MS/MS data files (.raw) obtained from the in-gel digestion were converted to mascot generic format (.mgf) files via MSConvert GUI of the ProteoWizard package (http://proteowizard.sourceforge.net; version 3.0.10328) and annotated by DeNovo GUI[342] (version 1.14.5) with a mass accuracy of 10 ppm for precursor mass and m/z 0.2 for fragment peaks. A fixed modification carbamidomethyl cysteine (C +57.02 Da) was selected. Resulting sequence tags were examined manually and searched against the non-redundant *Viperidae* protein database (taxid: 8689) using the basic local alignment search tool (BLAST).[343]

For peptide spectrum matching, the SearchGUI software tool was used with X!Tandem as the search engine.[279] The MS2 spectra were searched against the non-redundant *Viperidae* protein NCBI (taxid: 8689, 3rd Nov 2017, 1727 sequences), our in-house *Vipera kaznakovi* toxin sequence database (translated from our venom gland transcriptomic analyses; 46 toxin sequences) and a set of proteins found as common contaminants (CRAP, 116 sequences), containing in total 1,889 sequences. Mass accuracy was set to 10 ppm for the precursor mass and m/z 0.2 for the MS2 level. The alkylation of Cys was set as a fixed modification and acetylation of the N-terminus, of Lys, as well as oxidation of Met, were allowed as variable modifications. A false discovery rate (FDR) was estimated through a target-decoy approach and a cut-off of 1% was applied. All PSMs were validated manually and at least two PSMs were required for a protein ID to be considered.

For the top-down data analysis, the .raw data were converted to .mzXML files using MSconvert of the ProteoWizard package (http://proteowizard.sourceforge.net; version 3.0.10328), and multiple charged spectra were deconvoluted using the XTRACT algorithm of the Xcalibur Qual Browser version 2.2 (Thermo, Bremen, Germany). For isotopically unresolved spectra, charge distribution deconvolution was performed using the software tool magic transformer (MagTran).

Multivariable statistics

Principal component analysis (PCoA), using the relative percentages of the major toxin families as well as different proteoforms as variables, was applied to investigate determinants of compositional variation among venoms. PCoA was performed in R (R Foundation for Statistical Computing, 2016) with the extension Graphic Package rgl, available from https://www.R-Project.org.

Data sharing

Mass spectrometry proteomics data (.mgf, .raw and results files and search database) have been deposited to ProteomeXchange[344] with the ID PXD010857 via the MassIVE partner

repository under project name "Venom proteomics of *Vipera kaznakovi*" and massive ID MSV000082845. Raw sequencing reads and the assembled contigs generated for the venom gland transcriptome (.fastq and .fasta, respectively) have been deposited in the NCBI sequence read archive (SRA) under accession SRR8198764 and linked to the BioProject identifier PRJNA505487.
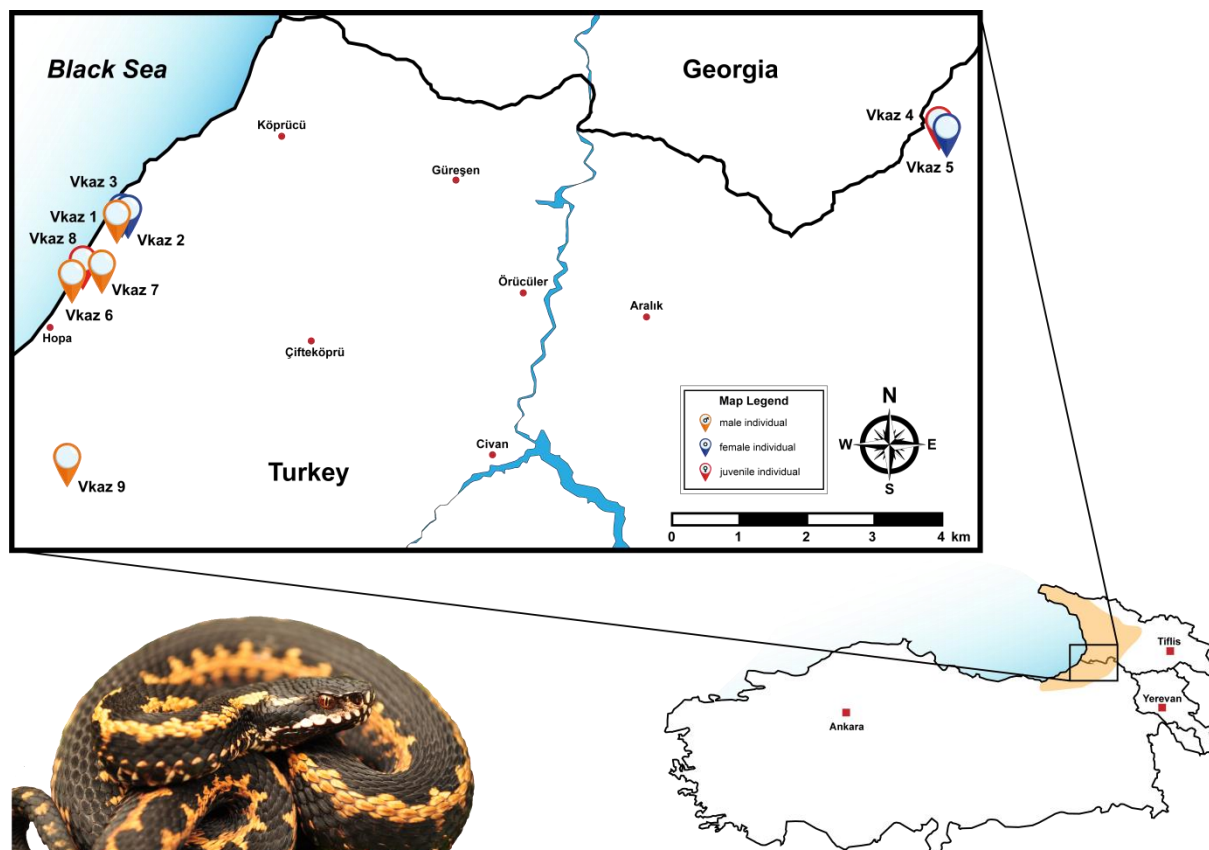
Figure 5.1. The geographical distribution and sampling localities of *Vipera kaznakovi*. The distribution area of the Caucasus viper (*Vipera kaznakovi*, genus *Viperidae*) is highlighted on the map in the lower right corner and adapted from Geniez et al.[376]. The locations and sex/age of the collected individuals are marked on the map (orange – adult male, red –adult female, blue - juvenile).

Results and Discussion

Field work and venom toxicity

During our fieldwork in June 2015 we collected nine *V. kaznakovi* individuals (6 adults and 3 juveniles) in their natural habitat, and extracted their venom before releasing them back into their natural environment. The different *V. kaznakovi* individuals were found in the Hopa (6 spec.), Borçka (2 spec.) and Arhavi (1 spec.) districts of Artvin province (**Figure 5.1**). The LD_{50} mean values of venom pooled from all collected *V. kaznakovi* individuals was assessed by the intraperitoneal (IP) route using a random sample survey of five Swiss albino mice for three venom dose (5, 2 and 1 mg/kg), which is summarized in appendix table 5.2. The LD_{50} value obtained for the pooled *V. kaznakovi* venom was calculated as ~2.6 mg/kg (2.1-3.4 mg/kg) and can be

categorized to have slightly weaker toxicity in this model, when compared to other related viper species (0.9-1.99 mg/kg).[377]

Venom gland transcriptomics

The *V. kaznakovi* venom gland transcriptome resulted in 1,742 assembled contigs, of which 46 exhibited gene annotations relating to 15 venom toxin families previously described in the literature (**Figure 5.2**). The majority of these contigs (33) encode genes expressing toxin isoforms relating to four multi-locus gene families, namely the svMPs, CTLs, svSPs and PLA$_2$s (**Figure 5.2**). Moreover, these four toxin families also exhibited the highest expression levels of the toxin families identified; in combination accounting for >78% of all toxin expressions (**Figure 5.2**). These findings are consistent with many prior studies of viperid venom gland transcriptomes.[350,351,368,378,379]

The svMPs were the most abundantly expressed of the toxin families detected, accounting for 33.4% of the total toxin expression, and were encoded by 17 contigs (**Figure 5.2**). However, these contig numbers are likely to be an overestimation of the total number of expressed svMP genes found in the *V. kaznakovi* venom gland, as six of these contigs were incomplete and non-overlapping in terms of their nucleotide sequence, and therefore likely reflect a degree of low transcriptome coverage and/or under-assembly. Of those contigs that we were able to identify to svMP class level (e.g. P-I, P-II or P-III[380]), ten exhibited structural domains unique to P-III svMPs, one to P-II svMPs and one to a short coding disintegrin. The svMP contig that exhibited the highest expression level encoded for the sole P-II svMP (5.1% of all venom toxins), whereas the short coding disintegrin, which exhibited 98% identity to the platelet aggregation inhibitor lebein-1-alpha from *Macrovipera lebetina*[381], was more moderately expressed (2.1%). Interestingly, we found no evidence for the representation of the P-I class of svMPs in the *V. kaznakovi* venom gland transcriptome.
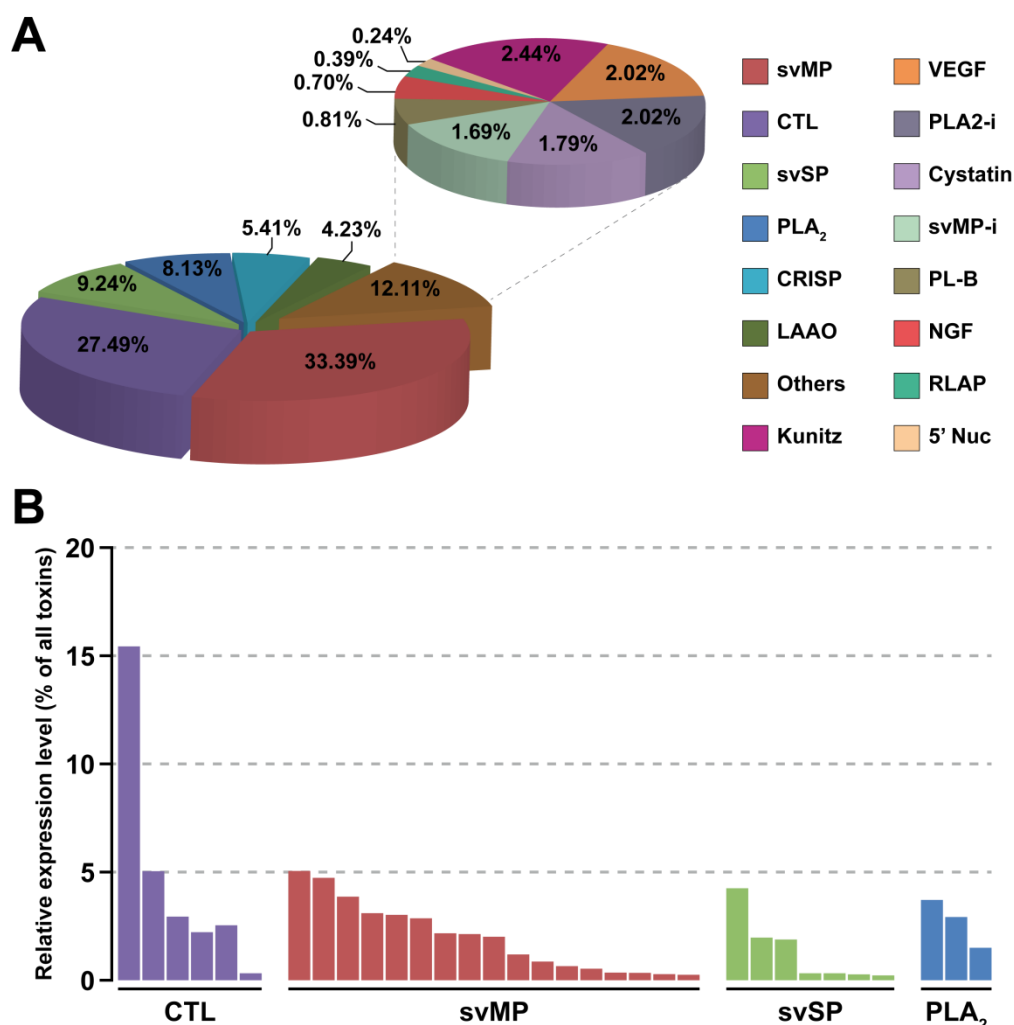
Figure 5.2. The relative expression levels of toxin families identified in the *Vipera kaznakovi* venom gland transcriptome. (**A**) The left pie chart shows the relative expression levels of the major toxin families, each of which accounts for greater than 4% of all toxins encoded in the venom gland. The right pie chart shows the relative expression levels of the remaining toxin families, which in combination account for 12.11% of all toxins encoded in the venom gland ("others"). Percentage values on both charts reflect the expression level of each toxin family as a percentage of the total expression of all identified toxin families. (**B**) The relative expression levels of individual contigs encoded by the most abundantly expressed toxin families (CTL, svMP, svSP and PLA2). Key: svMP – snake venom metalloproteinase; CTL – C-type lectin; svSP – snake venom serine protease; PLA2 – phospholipase A2; CRISP – cysteine-rich secretory protein; LAAO – L-amino acid oxidase; kunitz – Kunitz-type inhibitors; VEGF – vascular endothelial growth factor; PLA2-i – PLA2 inhibitors, svMP-i – SVMP inhibitors PLB – phospholipase B; NGF – nerve growth factor; RLAP – renin-like aspartic proteases; 5' Nuc – 5' nucleotidase.

The CTLs were the next most abundant toxin family, with six contigs representing 27.5% of all toxin gene expression (**Figure 5.2**). One of these CTLs, which exhibits closest similarity to snaclec-7 from *Vipera ammodtyes* venom (GenBank: APB93444.1), was by far the most abundantly expressed toxin identified in the venom gland transcriptome (15.4% of all toxins) (**Figure 5.2**). We identified lower expression levels for the multi-locus svSP and PLA2 toxin families, which accounted for 9.2% and 8.1% of the total toxins, expressed in the venom gland transcriptome respectively, and were encoded by seven and three contigs (**Figure 5.2**). Of the remaining toxin

families identified, only two exhibited expression levels >3% of the total toxin expression; CRISPs were encoded by two contigs amounting to 5.4% of total toxin expression, and LAAO by a single contig representing 4.2% (**Figure 5.2**). The remaining nine, lowly expressed, toxin families identified in the venom gland transcriptome are displayed in figure 5.2, and combined amounted to 12.1% of total toxin expression.

Venom proteomics of pooled venom

To broadly characterize the venom composition of *V. kaznakovi*, we performed bottom-up analysis of pooled venom by reversed phase-HPLC separation (**Figure 5.3A**) and direct online intact mass analysis by ESI-HR-MS (**Figure** 5.**3B**). The prominent bands of the subsequent separation by SDS-PAGE (**Figure 5.3C**) were excised followed by trypsin in-gel digestion and LC-MS/MS analysis. During the first analysis we did not have a species-specific transcriptome database available, hence the spectra were analyzed by *de novo* sequencing. The resulting sequence tags were searched against the NCBI non-redundant viperid protein database using BLAST.[343] The 57 sequence tags resulted in the identification of 25 proteins covering seven toxin families (**Table 5.1**), namely svMP, PLA$_2$, svSP, CTL, CRISP, VEGF and LAAO.

De novo sequencing of MS/MS spectra of native small peptides (peaks 1-9) resulted in four additional sequence tags and the identification of a svMP inhibitor (svMP-i) and two bradykinin potentiating peptides (BPP). When we obtained the assembled transcriptome data, we re-analyzed the MS/MS data from the tryptic peptides by peptide spectrum matching (PSM) using the translated protein sequences of the transcriptome as well as the NCBI Viperidae protein database. PSM resulted in 114 peptide matches in total, which doubled the number of annotated spectra in comparison to the *de novo* annotation. The analysis revealed the same seven major toxin families as identified by the tryptic *de novo* tags, but showed 29 identified proteins (compared to 25 by the prior approach) and thus a modest improvement. Not surprisingly, most of the peptide matches were from the transcriptome derived sequences, with only six protein IDs sourced from other viperid sequences in the NCBI database. Relative quantification through integration of the UV-HPLC peaks and densitometric analysis of the SDS-PAGE gels revealed that the most abundant toxin families were svMP (37.7%), followed by PLA$_2$ (19.0%), svSP (9.6%), LAAO (7.1%), CTL (6.9%), CRISP (5.0%), and VEGF (0.3%). In the small molecular mass range (< 2kDa), SVMP-i represented 12.6%, BPP 2.0%, and unknown peptides 4.0% of the overall venom composition (**Figure 5.3D**).

Table 5.1. Venom protein identifications from *Vipera kaznakovi*. The table shows all protein identification of HPLC fractions (**Figure 5.3**) by LC-MS and LC-MS/MS analysis from pooled venom. Peak numbering corresponds to the UV and MS chromatograms. Sequence tags were obtained by analysis of tryptic peptides by MS/MS *de novo* sequencing and/or peptide spectrum matching. Molecular weights of intact proteins were determined by SDS-PAGE and intact mass profiling (LC-MS). *For the sake of clarity, a simplified version is given here.*

Peak	RT	SDS-PAGE Band	Mass [Da] (ESI-MS)	Mass [kDa] (SDS-PAGE)	Sequence [PSM]	Prot-ID	Sequence (*de novo*)	Identity (blast)	e-value	Accession Number	Protein Family
1	20.1										
			443.2				pEQW	tripeptide svMPi	-	-	svMP-i
2	27.4										
			432.2					unknown	-	-	Peptide
3	28.2										
			3390.6					unknown			Peptide
4	30.3										
			3942.8				-	unknown	-	-	-
			860.3				EPGEEDW	Bradykinin-potentiating peptide	-	BAN04688.1	BPP
			822.4				pEKWPGPK	Bradykinin-potentiating peptide	-	BAN04688.1	BPP
5	32.6										
			3118.5				-	unknown	-	-	Peptide
			680.3				-	unknown	-	-	Peptide
6	35.7										
			3665.7				-	unknown	-	-	Peptide
7	36.8										
			13848.7				-	unknown	-	-	unknown
8	41.2										
			1101.6				K/QPGPVSV	unkown	-	-	Peptide
9	49.7										
			7228.2				-	unknown	-	-	unkown
			6680.9					unknown			unknown

Table 5.1. Venom protein identifications from *Vipera kaznakovi*. *continued*

Peak	RT	SDS-PAGE Band	Mass [Da] (ESI-MS)	Mass [kDa] (SDS-PAGE)	Sequence [PSM]	Prot-ID	Sequence (*de novo*)	Identity (blast)	e-value	Accession Number	Protein Family
10	50.4										
		10A		14	PFXEVYQR	T0203_R_0.0314_L_1049	PFXEVYQR	Nerve groth factor	2.2E-02	P83942.1	VEGF
					HTVDXQXM*R		HTVDXQXMR	Nerve groth factor	5.1E-02		
					ETXVPXXQEYPDEXSDXFRPSCVAVXR		-	-	-	-	
		10B		15	AAAXCFGENVNTYDKK	F8QN51.1	AAAXCAFGENVNTYDKK	acidic phospholipase A2	3.0E-08	F8QN51.1	PLA2
					pCCFVHDCCYGR		NXFQFGK	acidic phospholipase A2	1.1E+00		
					M*DTYSYSFXNGDXVCGDDPCXR	T1290_R_0.0575_L_419	MFCAGYXEGGK	cationic trypsin-3-like	4.0E-05	XP_015670852.1	
					SAYGCYCGWGGQGRPQDPTDR		-	-	-	-	
11, 12	61.2 61.6	11,12									
				15	AAAXCFGENVN*TYDKK	F8QN51.1	AAAXCAFGENVNTYDKK	acidic phospholipase A2	3.0E-08	F8QN51.1	PLA2
			13557.7; 13540.8; 13523.7		M*DTYSYSFXN*GDXVCDGDDDPCXR	T1290_R_0.0575_L_419	-	-	-	-	
					SAYGCYCGWGGQGRPQDPTDR		-	-	-	-	
					SAXXSYSAYGCYCGWGGQGRPQDPTDR		-	-	-	-	
13	64.4										
			13541.8								unknown
14	69.7										
			24671.3								unknown
15	70.5										
		15A	24655.5	25	Q*GCNNNYXK	B7FDI1.1	QGCNNNYXK	cyteine-rich venom protein	3.0E-03	B7FDI1.1	CRISP
					KPEXQN*EXXDXHNSXRR		KPEXQNEXXDXHNSXR	cyteine-rich venom protein	3.0E-05	XP_015678374.1	
					NVDFDSESPR		WTAXXHEWHGEEK	cyteine-rich venom protein	4.0E-07	B7FDI1.1	
					M*EWYPEAAANAER		SVDFDSESPR	cyteine-rich venom protein	2.0E-05	P86537.1	
					SVNPTASNM*XK		-	-	-	-	
					VDFDSESPR		-	-	-	-	
					DFVYGQGASPANAVVGHYTQXVWYK		-	-	-	-	

Table 5.1. Venom protein identifications from *Vipera kaznakovi*. *continued*

Peak	RT	SDS-PAGE Band	Mass [Da] (ESI-MS)	Mass [kDa] (SDS-PAGE)	Sequence [PSM]	Prot-ID	Sequence (*de novo*)	Identity (blast)	e-value	Accession Number	Protein Family
		15B	13691.8	13	HXSQFGDMXNK		HXSQFGDMXNK	Ammodytin I1(A) variant	6.0E-04	CAE47141.1	PLA$_2$
					pCCFVHDCCYGR	Q910A1	-	-	-	-	
16, 17	73.3 74.4				VAAXCFGENM*NTYDQKK		VAAXCAFGENMNTYDQK	Ammodytin I1(A) variant	3.0E-09	CAE47176.1	
		17A	51761	50	FXTNFKPDCTXXRPSR	T0053_R_0.0734_L_1810	VPXVGVEFWXNR	snake venom metalloproteinase III	5.0E-04	ADW54336.1	svMP
					SECDXPEYCTGK		XVXVVDHSMVEK	snake venom metalloproteinase	9.0E-05	ADI47673.1	
					XGQDXYYCR		-	-	-	-	
					KEN*DVPXPCAPEDVK		-	-	-	-	
		16, 17B	13675.9	14	HXSQFGDMXNK		HXSQFGDMXNK	Ammodytin I1(A) variant	6.0E-04	CAE47141.1	PLA$_2$
					pCCFVHDCCYGR	Q910A1	-	-	-	-	
					VAAXCFGEN*M*NTYDQKK		VAAXCAFGENMNTYDQK	Ammodytin I1(A) variant	3.0E-09	CAE47176.1	
					YMLYSIFDCK		-	-	-	-	
18	80.6										
		18A	N.D.	65	XVXVVDHSM*VTK		XVXVVDHSMVTK	snake venom metalloproteinase group III	1.0E-04	CAJ01689.1	svMP
					YN*SDXTVXR		YNSDXTVXR	snake venom metalloproteinase	2.9E-01	ADI47687.1	
					VPXVGVEXWDHR		VPXVGVEXWDHR	snake venom metalloproteinase	6.5E-02	ADI47590.1	
					pQXVATSEQQR		-	-	-	-	
					VNXXNEM*YXPXNXR	T0033_R_0.0599_L_2024	-	-	-	-	
					KRHDNAQXXTTXDFDGSVXGK		-	-	-	-	
					HSVAXVEDYSPXDR		-	-	-	-	
					FXTNDKPDCTXXRPSR		-	-	-	-	
					KGESYFYCR		KGESYFYCR	snake venom metalloproteinase	9.0E-05	ADI47619.1	
					KENDVPXPCAPEDXK		-	-	-	-	

Table 5.1. Venom protein identifications from *Vipera kaznakovi*. *continued*

Peak	RT	SDS-PAGE Band	Mass [Da] (ESI-MS)	Mass [kDa] (SDS-PAGE)	Sequence [PSM]	Prot-ID	Sequence (*de novo*)	Identity (blast)	e-value	Accession Number	Protein Family
18		18B	51600	50	FXTNFKPDCTXXRPSR		VPXVGVEFWXNR	snake venom metalloproteinase group III	5.0E-04	ADW54336.1	svMP
					SECDXPEYCTGK	T0053_R_0.0734_L_1810	XVXVVDHSMVEK	snake venom metalloproteinase	9.0E-05	ADI47673.1	
					XGQDXYYCR		-	-	-	-	
					KEN*DVPXPCAPEDVK		-	-	-	-	
		18C	30133	35	VXGGDECNXNEHPFXVAXHTAR	T1355_R_0.005_L_400	VXGGDECANXNEHPFXAFVTSDR	snake venom serine proteinase nikobin	2.0E-12	E5AJX2.1	svSP
					FYCAGTLXNQEWVXTAAR		XMGWGTXSSTK	snake venom serine proteinase	3.0E-05	ART88740.1	
					VVCAGXWQGGK		VVCAGXWQGGK	snake venom serine proteinase nikobin	1.0E-04	E5AJX2.1	
					C*AGTXXNQEWVXTAAHCNGK	E5AJX2	XMGWGTXTTTK	snake venom serine proteinase	6.0E-04	ADE45141.1	
					XXPDVPHCANXEXXK		-	-	-	-	
					VHPEXPAK		-	-	-	-	
		18D	17249	14	KTWEDAEKFCTEQAR		WTEDAENFCQK	C-type lectin snaclec-1	1.0E-04	AMB36338.1	CTL
					SPEEVDFM*XK	T0841_R_0.0782_L_536	SPEEVDFMXK	C-type lectin-like protein 2B	1.1E-02	AJO70722.1	
					ADXVWXGXR		HXATXEWXGK	C-type lectin snaclec A16	1.8E-01	B4XSZ1.1	
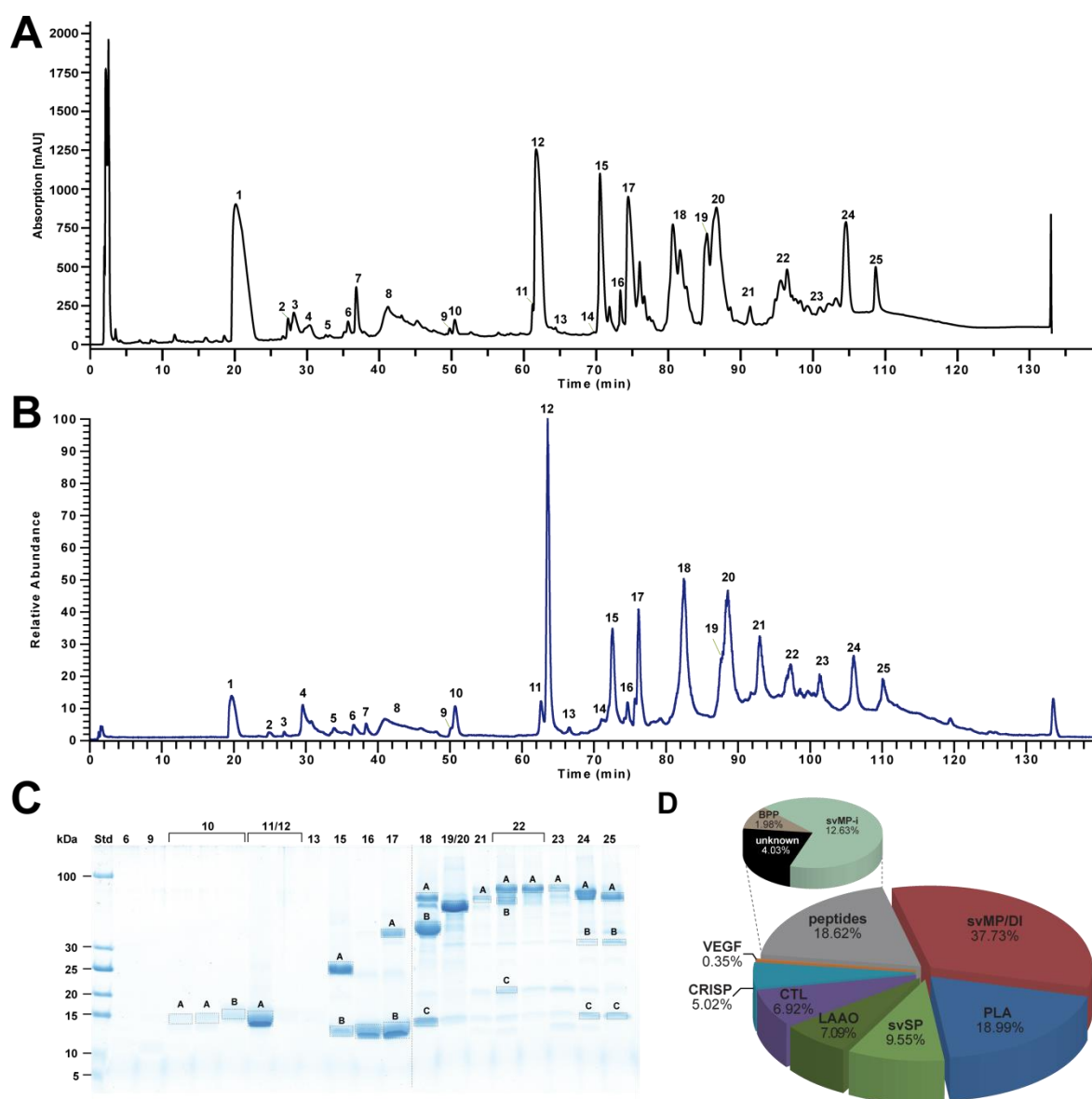
Figure 5.3. Bottom-up snake venomics of *Vipera kaznakovi*. (**A**) Venom separation of *V. kaznakovi* was performed by a Supelco Discovery BIO wide Pore C18-3 RP-HPLC column and UV absorbance measured at λ = 214 nm. (**B**) Total ion current (TIC) profile of crude *V. kaznakovi* venom. The peak nomenclature is based on the chromatogram fractions. (**C**) The RP-HPLC fractions (indicated above the lane) of the *V. kaznakovi* venom was analysed by SDS-PAGE under reducing conditions (Coomassie staining). Alphabetically marked bands per line were excised for subsequent tryptic in-gel digestion. (**D**) The relative occurrence of different toxin families of *V. kaznakovi* are represented by the pie chart. Identification of snake venom metalloproteinase (svMP, red), phospholipases A$_2$ (PLA$_2$, blue), snake venom serine proteinase (svSP, green), C-type lectin like proteins (CTL, purple), cysteine rich secretory proteins (CRISP, light blue), bradykinin-potentiating peptides (BPP, light brown), vascular endothelial growth factors (VEGF-F, red), unknown proteins (n/a, black) and peptides (grey). The *de novo* identified peptides are listed in table 5.1.

When comparing the abundance of venom toxins (**Figure 5.3D**) with transcriptomic predictions of expression levels (**Figure 5.2A**), we observed an overall positive correlation, but also noted some major differences, particularly relating to the CTLs: transcriptomic expression levels showed CTLs to be the second most abundant toxin family (27.5% of all toxins) while proteomic

analysis showed a much lower abundance (6.9%). Interestingly, some of the molecular masses observed for CTLs (~20 kDa) during SDS-PAGE did not correspond to the expected molecular mass derived from the transcriptome sequences. As reported in other studies, we assume that some of the observed CTLs are hetero-dimers.[382] SvMPs showed highly consistent profiles, as both the most abundantly expressed (33.4%) and translated (32.7%) toxin family. Similarly, the svSPs (9.2%) and CRISPs (5.4%) exhibited transcription levels highly comparable to their relative protein abundance in venom (9.6% and 5.02%). A lower transcription level was shown for PLA_2 (8.1%) in contrast to the two times higher protein level (19.0%). As anticipated, with the exception of VEGF (2.0% T; 0.4% P) and svMP-i (1.7%; 12.6%) as part of the peptidic content, other lowly expressed 'toxin' families could not be assigned on the proteomic level.

The observed discrepancies in proteomic abundance and transcriptomic expression (e.g. CTLs and PLA_2s) could be influenced by many factors, e.g. post-genomic factors acting on toxin genes [368], such as the regulation of expression patterns by MicroRNAs (miRNA)[347,383], degradation processes[384], systematic or stochastic variations[385] or technical limitations in our experimental approaches, including the lower sensitivity of the proteomics workflow. Perhaps most importantly, we compared the toxin transcription level of a single individual (adult female) to a pooled venom protein sample (n=9), and thus, while it is possible that these differences are predominately due to the above mentioned regulatory processes, it seems likely that intra-specific venom variations may also influence our findings. Due to sampling/ethical restrictions relating to the sacrifice of multiple individuals, we were unable to sequence venom gland transcriptomes of additional specimens of *V. kaznakovi* to investigate this further.

The previous proteomic characterization of *V. kaznakovi* venom by Kovalchuk and coworkers was performed by in-solution trypsin proteolysis followed by nanoLC-MS/MS.[371] The PSM against a full NCBI Serpentes database identified 116 proteins from 14 typical snake venom protein families. The semi-quantitative venom composition showed PLA_2 (41.0%) as the most abundant component, followed by svMPs (16.2%), CTL (12.5%), svSP (10.8%), CRISP (9.7%), LAAO (4.0%), VEGF (4.0%) and other lowly abundant proteins (< 1%).[371] The main difference between these findings and those described herein, are the considerably higher levels of PLA_2 and the lower abundance of svMPs (~ 4 fold difference for both protein families). The reasons for the additional detection of lowly abundant proteins could be of technical nature, as the nanoLC-MS/MS and mass spectrometer used in the previous study is typically more sensitive than the LC-MS/MS setup we applied. Differences in protein abundance could also be the result of the different quantification methods applied (UV abundance vs. summed peptide abundance[371]), but the observed variations could also be biological in nature, i.e. the result of intra-specific venom variation, as the animals were collected in distinct geographic regions (Krasnodar Territory, Russia[370], with a distance of ~ 400 km to our collection site in Turkey). However, as in most other venom proteomics studies, Kovalchuk et al. determined toxin composition using a pooled venom sample (15 individuals[371]),

which has the potential to offset variation among individuals. Therefore, in order to robustly assess the extent of intra-specific (e.g. population level) venom variation in *V. kaznakovi*, analysis of a representative group of individuals is necessary.

Population venom profiling

It is understandable that many venom proteomic studies were undertaken using pooled samples, due to the associated costs and analysis time of decomplexing bottom-up venomics. For example, herein we fractionated pooled venom from *V. kaznakovi* into 25 fractions and further separated the protein containing fractions (MW >5kDa) by SDS-PAGE. This multidimensional separation resulted in 25 digested peptide samples which were analyzed by LC-MS/MS, requiring ~10 h MS run time (25 min/sample), and an estimated ~US$1250 costs (US$50/sample). Multiplying this effort and cost by numerous venom samples from individuals would of course make such a study comparatively expensive. Hence, many previous studies investigating venom variability within a species have used pooled venom for in-depth proteomic analysis, and then illuminated individual variability by the comparison of HPLC chromatograms and/or SDS-PAGE images.[369,386] However, such an approach allows at best a comparison at the protein family level (if protein families are clearly separated by HPLC or SDS-PAGE). As an alternative, top-down or shotgun proteomics would facilitate differential comparisons on the protein, or potentially proteoform, level, by performing a single LC-MS/MS run per individual. Taking into account the longer LC-MS run-times (120 min/run) for shotgun or top-down analysis and costs than can be estimated to be ~US$200 per venom sample – the analysis would be significantly cheaper and faster. However, shotgun approaches are likely to suffer from the aforementioned issues with protein inference, while top-down approaches have the drawback of not resolving high molecular mass proteins. This is particularly the case if the identification and comparison of proteins are based on protein spectrum matching (P*r*SM), as high molecular weight toxins may not result in isotope resolved peaks and sufficient precursor signal, and thus are unlikely to provide sufficient fragment ions. However, a comparison by MS1 mass profiling only[387] would eliminate the problem of insufficient MS/MS fragments and isotope resolution, as spectra can be easily deconvoluted based on their charge state distribution. Such an approach could be particularly interesting for laboratories that are equipped with low resolution mass spectrometers.
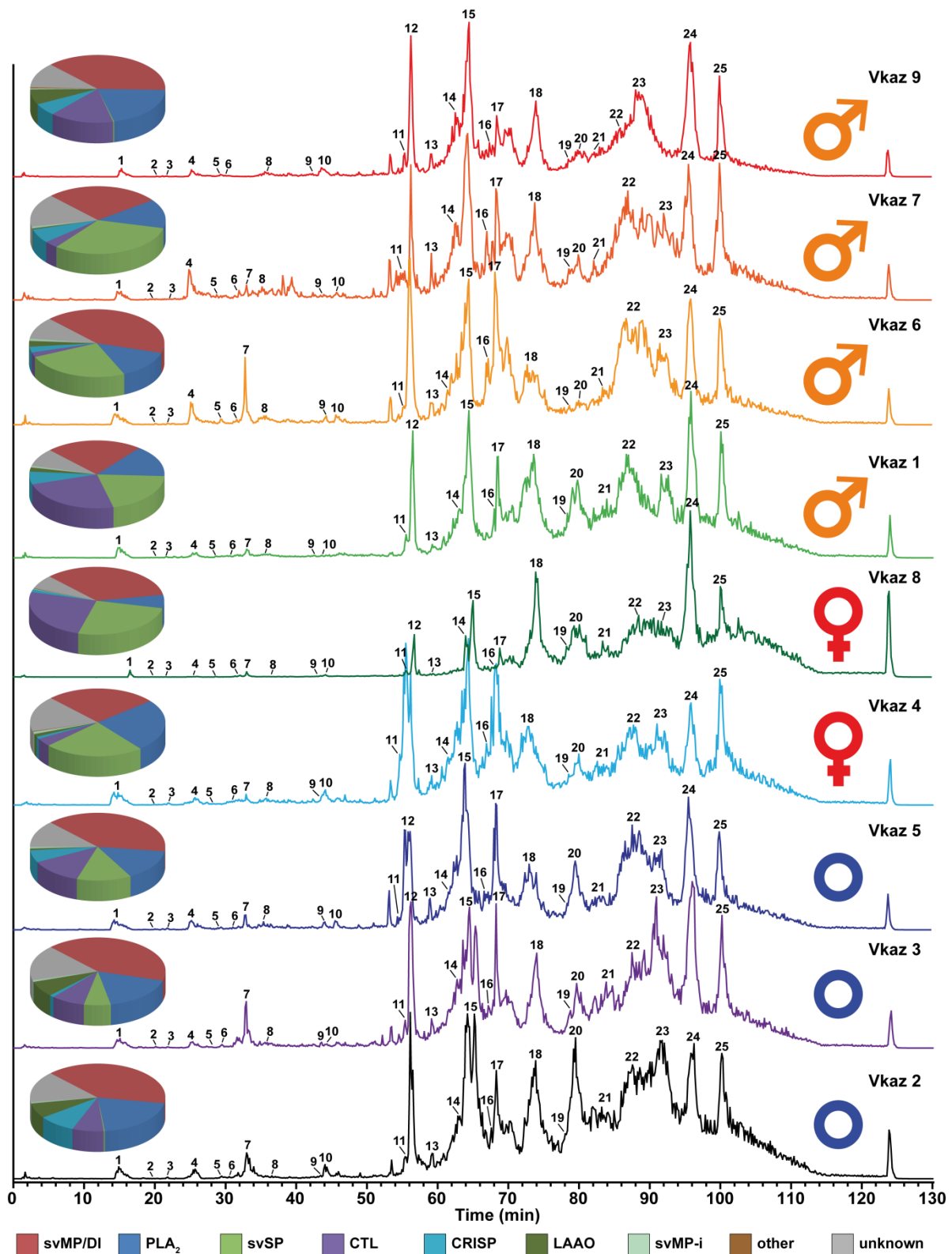
Figure 5.4. Intact molecular mass profiles of venom from several individuals of *V. kaznakovi*. The total ion counts (TIC) of native, crude venoms from several *V. kaznakovi* individuals were measured by HPLC-ESI-MS. The relative abundance was set to 100% for the highest peak. The peak nomenclature is based on the chromatogram fractions and is shown in figure 5.3A. The identified molecular masses of intact proteins and peptides are listed in table 5.1 and shown in appendix figure 5.1. The intact molecular mass profiling includes three juveniles of unknown sex (blue circle), and two female (red Venus symbol) and four male (orange Mars symbol) adult individuals.

In order to explore the potential of venom comparison by top-down mass profiling, we analyzed the venoms of nine *V. kaznakovi* individuals by LC-MS using the same chromatographic method as for our initial HPLC separation of our decomplexing bottom-up venom analysis. Chromatographic peak extraction of all individuals resulted in 119 consensus extracted ion chromatograms (XIC) or so-called ion features. The alignment of XICs by retention time and mass enabled the comparison of samples between individuals, and a comparison with the mass profile of the pooled venom sample for a protein level annotation. The binary distribution of ion features showed that individual venoms contained between 62 and 107 features, with a slightly higher average feature number in juveniles vs. adults. Comparing the total ion currents (TIC) of the LC-MS runs, the individual with the lowest feature number also had the lowest overall signal. Hence it is likely that the lower number of features in this individual was due to lower overall signal intensity and therefore might not be biologically representative. For further statistical evaluation we normalized feature abundance to TIC. Matching the features to the pooled bottom-up venomics results yielded an annotation rate of between 83.4% and 93.5% of the features (based on XIC peak area). As anticipated, the annotation rate is slightly lower than the relative annotation of the pooled sample (96.0%; based on the UV_{214} peak area).

A comparison of the resulting protein family venom compositions is shown in figure 5.4. The highest variance was observed for svSP, CTL and LAAO toxin families (**Figure 5.5A**). Taking the age of the individuals into account, the abundance of svSPs was generally higher in the adult individuals than in the juveniles (average of 21.7% vs. 5.5%), but no significant difference between male and female individuals, or between different geographic regions was observed. The svSPs play a significant role in mammalian envenomation by affecting the hemostatic system through perturbing blood coagulation, typically via the inducement of fibrinogenolytic effects.[83,388] Taking this into account, a possible explanation could be that the lower svSP concentration observed in juveniles could be the result of differences in diet, as young animals typically prey on insects, before switching to feed upon small mammals and lizards as they become adults.[389] Despite observed variations in abundance, no significant differences were found between the individual groups for the CTL and LAAO toxin families (**Figure 5.5A**). However, there was evidence that svSP concentration is correlated to levels of LAAO, as the three individuals with the lowest svSP abundance showed the highest content of LAAO (**Figure 5.5A**). Whether this is a true biological effect or perhaps is the result of differences in ion suppression of the co-eluting compounds requires further investigation with large sample sizes. While we also observed variation between the PLA_2 levels identified in the different venoms, which ranged from 6.5-25.1%, in all cases these abundances remained considerably lower than those previously reported by Kovalchuk et al. (41%).[371]
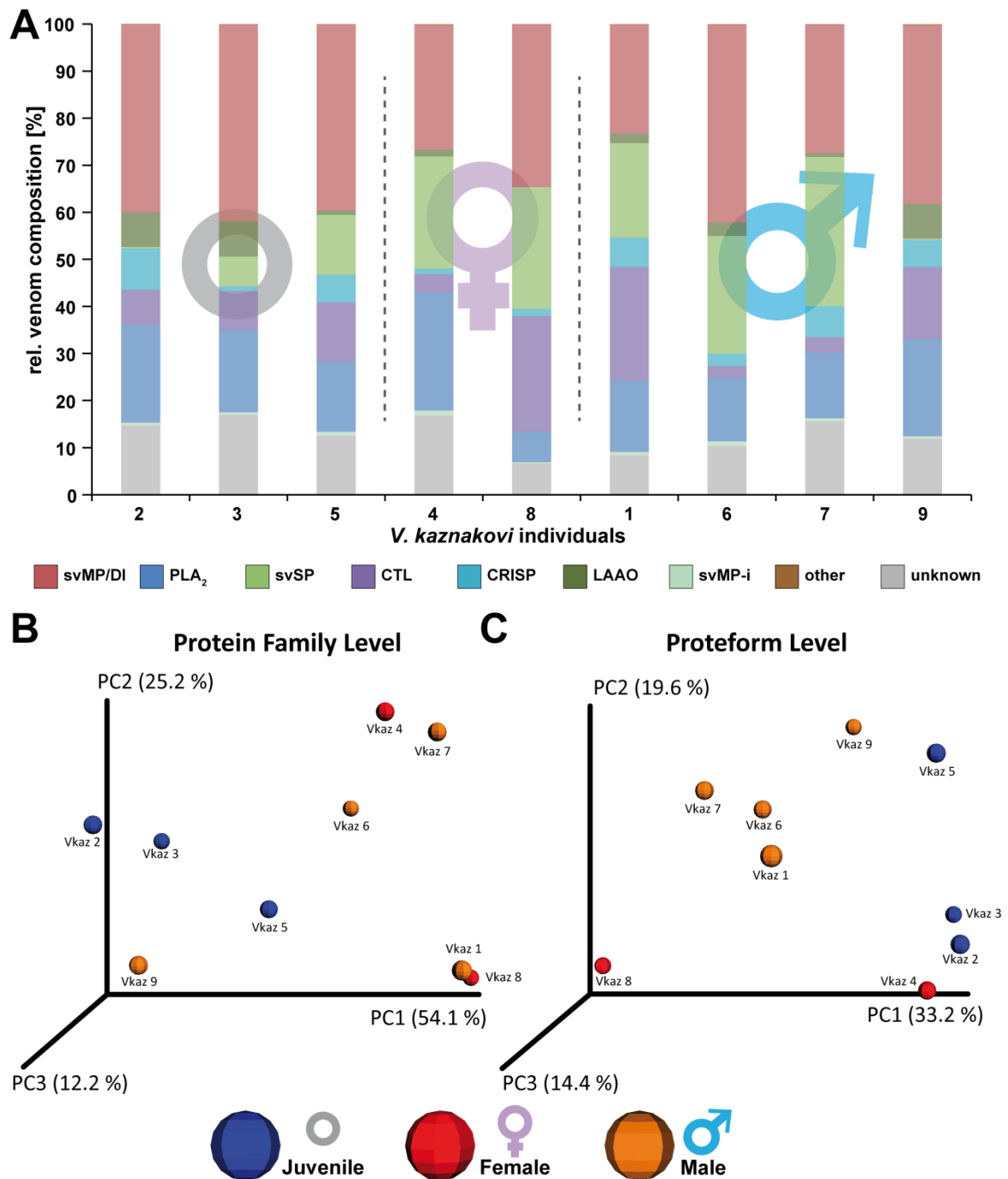
Figure 5.5. Principal Component Analysis (PCoA) and relative compositions of the individual *V. kaznakovi* venoms. (**A**) The proteome overview includes three juveniles of unknown sex (blue circle), and two female (red Venus symbol) and four male (orange Mars symbol) adult individuals. (**B**) and (**C**) The compositional similarity of venom is displayed through Bray-Curtis-Faith distance in PCoA space. Toxin similarity is visualized at the protein family level (**B**) and proteoform level (**C**).

In order to investigate the intra-species differences by multivariate statistics we performed a principal coordinate analysis (PCoA) using the Bray-Curtis dissimilarity metric. The PCoA plots of protein-level and proteoform-level data are shown in figure 5. Clustering between adult and juvenile individuals in protein-family level PCoA space was not found to be significant (**Figure 5.5B**, ADONIS: $r^2 = 0.32$, $p = 0.27$). In addition, and as anticipated from the univariate

statistics, no separation based on gender or geographical region could be observed. Since a possible explanation for not resolving phenotype differences could be the reduction of variables through the binning of proteoforms, we next used proteoform abundance as the input matrix for PCoA. The outcomes of this second analysis revealed a weak grouping between both juvenile and adults, as well as between male and female snakes (**Figure 5.5C**, ADONIS: r^2 = 0.33, p = 0.07). To investigating the toxin variants underpinning these separations, we used univariate comparisons of the two groups and plotted the fold change of toxin abundance (log2) vs. the statistical significance (-log10 p-value, t-test) shown in appendix figure 5.2. Aside from the aforementioned differences in svSP toxins, the most significant (p-value < 0.05, log2 fold change >2 or <-2) differences observed between juvenile and adult individuals was the higher abundance of small proteins (with masses 7707.26 Da, 5565.02 Da and 5693.10 Da) in the juvenile group, all of which were unidentified in our proteomic analyses. In addition, we observed several smaller peptides (with masses 589.27 Da, 1244.56 Da, and 575.26 Da), as well as a putative PLA$_2$ (13667.91 Da), that were more abundant in venom of the juveniles. Contrastingly, a putative PLA$_2$ (13683.86 Da) was of lower abundance in the juvenile group. While we observed fewer significant changes between the venom toxins of the male and female individuals, the observed masses of the differential features indicated that some of those differential toxins belong to different protein families than those involved in differentiating between juvenile and adult snakes. Two toxins with the masses 22829.66 Da and 24641.23 Da were of higher abundant in male individuals, and could be putatively annotated as hetero-dimeric CTLs, while a third toxin (13549.87 Da, putative PLA$_2$) was also of higher abundance in the males.

Concluding remarks

Herein we described comprehensive analyses of the venom composition of *Vipera kaznakovi* by a combination of venom gland transcriptomics and decomplexing bottom-up and top-down venom proteomics. Our findings revealed the presence of 15 toxin families, of which the most abundant toxins were svMPs (37.7%), followed by PLA$_2$s (19.0%), svSPs (9.5%), CTLs (6.9%) and CRISP (5.0%). Intact mass profiling enabled the rapid comparison of venom sourced from multiple individuals. This population venomics approach enabled higher sensitivity of direct intact protein analysis by LC-MS, in comparison to decomplexing bottom-up venomics, and thus enabled us to work with multiple venom samples of low quantity (<0.5 mg venom). This approach also permitted us to capture the snakes, perform venom extractions, and then immediately release the animals back into the field. Our findings revealed intraspecies venom variation in *Vipera kaznakovi*, including both ontogenetic differences between juvenile and adult snakes, and to a lesser extent, sexual differences between adult males and females. The highest significant difference in venom proteome composition was observed between the adult and juvenile group, with svSP toxins found to exhibit the greatest variance. However, individuals in all groups showed

a generally high relative variance of CTL and LAAO concentrations. svMPs on the other hand seemed to be constantly the most abundant venom component in all *V. kaznakovi* individuals analyzed in our study. However, as the statistical power with a relatively small subject size (n=9) is limited, we acknowledge that extending this study to a larger sample cohort, ideally covering all geographical regions (from Northeastern Turkey to Georgia and Russia) of the *V. kaznakovi* distribution zone, would be revealing. In conclusion, the top-down venom profiling approach that we applied herein appears to be well suited for extensive venom analysis at the population level, and will hopefully enable venom researchers to expand their experimental toolbox towards robust comparisons of intraspecies venom variation, and move beyond the characterization of pooled venoms.

6 Extended Snake Venomics by Top-Down In-Source Decay: Investigating the Newly Discovered Anatolian Meadow Viper Subspecies, *Vipera anatolica senliki*

Introduction

Venoms are fascinating biological cocktails mainly composed of peptides as well as proteins and are produced by a phylogenetically broad range of organisms in terms of predation, competition, or defense.[37] The global analyses of venom for diverse members of *Serpentes* are termed as "snake venomics", which usually employs an integrative approach combining proteomics, transcriptomics, and occasionally genomics.[291,294,345] The term was coined in 2004 by Juárez et al.[223], who described for the first time the decomplexation of snake venom by a combination of liquid chromatography (LC) separation in the first dimension, followed by a one-dimensional electrophoresis (sodium dodecyl sulfate-polyacrylamide gel electrophoresis, SDS-PAGE) in the second dimension.[352] Since then, the venomics community experienced a rapid expansion into different venomics applications, and a multitude of proteome studies for a wide variety of snake families have been published.[225,353] Today, decomplexation of snake venom proteomes can be achieved by several bottom-up protocols, combining multidimensional separation methods.[224,225] However, bottom-up proteomics suffers from classical proteolytic drawbacks, a step that often obscures the differentiation of toxin proteoforms and prevents the identification of post-translational modifications (PTMs).[294,356,390] In this context, the recently introduced top-down protocol for snake venoms provides remedy by directly analyzing intact proteins and peptides out of crude venoms using high-resolution tandem mass spectrometry.[227,231,299,357] However, in the case of higher-molecular-weight compounds (>30 kDa), which are typically strongly represented in the genus of *Viperidae*, the top-down analysis only provides a partial characterization and is still challenging.[233,234,358,359]

Here, we applied a *de novo* in-source decay-driven (ISD) venomics workflow using matrix-assisted laser desorption/ionization (MALDI) as an alternative top-down approach to characterize high molecular mass venom constituents. MALDI-ISD refers, in contrast to the commonly observed post-source decay (PSD), to fragmentation directly in the MALDI plume during desorption/ionization prior ion extraction.[391,392] The pseudo-MS/MS technique uses the hydrogen radical transfer from the 1,5-diaminonaphthalene (1,5-DAN) matrix to the analyte, thus providing predominantly c- and z-type fragmentation.[393,394–396] The radical-mediated pathway allows a soft dissociation of proteins in contrast to higher-energy fragmentation methods (e.g.,

collision-induced dissociation, CID) and therefore showed the applicability to identify high molecular mass proteoforms as well as to assign PTMs.[397] The MALDI-ISD top-down sequencing (TDS) is used to characterize protein terminal sequences from the N- or C-termini of intact proteins with the exception of roughly 10 terminally located amino acids due to matrix adducts and/or high background noise. Further downstream MS experiments can access to these terminal amino acids provided by pseudo-MS3 fragmentation, termed T^3-sequencing. In addition, it was reported that the matrix 1,5-DAN can act as a reducing agent and facilitates disulfide bond counting, which is an accessory technique to identify snake venom families by their characteristic number of disulfide bridges.[224,398] A proof-of-concept study by Quinton et al.[399] showed the applicability of ISD for small individual and highly purified snake peptides (<10 kDa) such as cardiotoxins and neurotoxins from the monocled cobra (*Naja kaouthia*). Subsequently, MALDI-ISD was applied in various studies to analyze several small peptides from different venomous species.[400,401] Here, we applied MALDI-ISD for analyzing different toxin families and proteoforms of higher molecular mass for the recently discovered Anatolian meadow viper subspecies, *Vipera anatolica senliki* (*V. a. senliki*), in combination with established snake venomics protocols for a comprehensive venom analysis.

The Anatolian meadow viper *Vipera anatolica* (*V. anatolica*) is a small, mainly insectivorous species from high altitude, stony grasslands (alpine to subalpine meadows). Until 2017, *V. anatolica*, was classified as a monotypic species with a narrow distribution area in Kohu Mountain, Çığlıkara Cedar Reserve, Elmali district, western Antalya Province (Turkey).[402] The venom proteome of *V. anatolica* was previously described and is composed of snake venom metalloproteinases (svMPs), cysteine-rich secretory proteins (CRISPs), phospholipases A_2 (PLA$_2$s), disintegrins (DIs), snake venom serine proteases (svSPs), C-type lectin-like proteins (CTLs), Kunitz-type protease inhibitors (KUNs), and various small peptides.[233] In a recent field study by Göçmen et al.[403] a new population of the rare Anatolian meadow viper was discovered at the Senir, Serinyaka and Gelesandra Plateaus of Gündoğmuş district in eastern Antalya Province, around 200 km distant from the known habitats. Subsequently, morphological and genetic analyses against known *V. anatolica* (now *V. a. anatolica*) populations led to a revised taxonomic classification, designating the Anatolian population as new subspecies *V. a. senliki*.[403] The combination of venom gland transcriptomics, bottom-up proteomics of decomplexed venom, intact mass profiling, and ISD top-down proteomics of *V. a. senliki* venom shed further light on the detailed venom composition of *V. a. senliki* and significant variations on an interspecies scale, which are of clinical relevance for snakebite treatment.

Material and Methods

Sampling of crude venom and venom gland dissection

All crude venoms from five female and five male specimens of *V. a. senliki* were pooled to one venom stock. Venom samples were collected from the Mühür Mountain, Senir and Gelesandra Plateaus, Gündoğmuş district (Antalya Province, Turkey) between May and June 2016. Ethical permissions (Ege University Animal Experiments Ethics Committee, 2010#43 and 2019#070) and special permission (2014#51946) for field studies from the Republic of Turkey, Ministry of Forestry and Water Affairs were obtained.

One male specimen was used for venom gland collection. Dissection was performed on a sterile surface under general anesthesia by intramuscular ketamine injection. The venom gland lies just behind the eye and is covered by the compressor glandulae muscle. Primarily, this muscle was dissected to uncover the venom gland. The venom gland was gently dissected and preserved in RNAlater (Thermo, Bremen, Germany) stabilization solution.

Venom gland transcriptomics

RNA extraction and transcriptome sequencing

The venom gland was shipped to the National University of Singapore on dry ice and transferred immediately to a -80 °C freezer for storage until RNA extraction. The sample was thawed on ice and homogenized in 400 µL of RNAzol RT using a homogenizer and polypropylene pestle. An additional 600 µL of RNAzol RT was added and mixed in with the homogenate. Total RNA was extracted as per manufacturer's protocol (in brief, DNA, proteins, and polysaccharides were eliminated, and ethanol washes were performed to purify RNA). The RNA pellet was resuspended in 22 µL of molecular biology grade water. Preliminary RNA quality checks and quantification were performed on NanoVue Plus. Transcriptome cDNA library preparation and sequencing were performed at NovogeneAIT. Total RNA quality and concentration were assessed using a bioanalyzer, and 2 µg was used for mRNA enrichment. cDNA library was constructed using the NEBNext Ultra Directional Library Prep Kit followed by library quality check by a Bioanalyzer RNA kit. Paired End (PE) sequencing (250 bp) was performed on 1/14th of a lane on an Illumina HiSeq 2500 platform.

Transcriptome assembly

The 250 bp PE raw reads were quality checked using FastQC[404] (v0.11.4) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and then processed in Trimmomatic-0.32[405] to trim adapters and primers. Quality trimming was performed using a sliding window of 4 bp with a Phred score cutoff of Q30 for a high confidence read set. Quality filtered reads of length ≥100 bp were retained for assembly. A *de novo* approach was used to

assemble transcriptomes, employing two dedicated assemblers VT-builder[374] (v7.02) and Extender 3[378], as well as the generic assembler Trinity[406] (v2.6.6). All assemblies were then combined and dereplicated at 100% identity using CD-HIT[407,408] (v4.6.7).

Venom transcriptome database and venom gene expression

All dereplicated contigs were searched against a curated venom database downloaded from UniProt. BLASTX was performed with an *e*-value cutoff of 1×10^{-5} in DIAMOND[409] (v0.8.38.100). All significant hits were aligned to venom toxin sequences in the database and then manually curated for accuracy. The curated nucleotide database was translated to proteins and dereplicated at the 100% identity at the protein level using CD-HIT[407,408] (v4.6.7) to prepare a nonredundant protein database for proteomics mapping. Venom gene expression was quantified by mapping the short reads back to the *de novo* transcriptome assembly using Bowtie[410] (v1.2.2). Gene expression was quantified using Transcript Per Million (TPM), which normalizes read counts by the length of the assembled transcript and then normalizes read depth by a per-million scaling factor. For each toxin family, gene expression was represented as the TPM percentage relative to the total expression in the venom gland.

Top-down venomics

The crude venom (0.1 mg) was dissolved in aqueous 1% (v/v) formic acid (FA) to a final concentration of 10 mg/mL and centrifuged at 20,000×g for 5 min to spin down insoluble content. Next, dissolved venom was mixed with 30 μL of citrate buffer (0.1 M, pH 3.0) and halved for native and reduced sample analysis. To chemically reduce disulfide bonds, 10 μL of tris(2-carboxyethyl)-phosphine (TCEP, 0.5 M) was added to one half and the reaction mixture was incubated for 30 min at 65 °C. For the non-reduced sample, 10 μL of ultrapure water was added to the other half. In the following, the samples were centrifuged at 20,000×g for 5 min. Then, 10 μL of both reduced and non-reduced samples were analyzed by high-performance liquid chromatography (HPLC) high-resolution (HR) electrospray ionization (ESI)-MS/MS measurements.

The intact mass profile and top-down analysis were performed by LC-ESI-HR-MS experiments on an LTQ Orbitrap XL mass spectrometer (Thermo, Bremen, Germany) coupled to an Agilent 1200 HPLC system (Agilent, Waldbronn, Germany), using a Supelco Discovery 300 Å C18 (2.1 × 150 mm, particle size, 3 mm) column. The flow rate was set to 0.3 mL/min, and the column was eluted with a gradient of 0.1% FA in water (solution A) and 0.1% FA in acetonitrile (ACN) (solution B): 5% B for 1 min, followed by 5–40% B for 89 min, and 40–70% for 20 min. Finally, the column was washed with 70% B for 10 min and re-equilibrated at 5% B for 10 min. ESI settings were 45 L/min sheath gas; 5 L/min auxiliary gas; spray voltage, 4.5 kV; capillary voltage, 46 V; tube lens voltage, 155 V; and capillary temperature, 275 °C. MS/MS spectra were obtained in the data-dependent acquisition (DDA) mode. Fourier transform mass spectrometry

measurements were performed with 1 μ scans and 500 ms maximal fill time. Automatic gain control targets were set to 10^6 for full scans and to 1×10^5 for MS/MS scans, and the survey scan as well as both data-dependent MS/MS scans were performed with a mass resolution (R) of 100,000 (at m/z 400). For MS/MS the top three most abundant ions of the survey scan with a known charge were selected. Normalized CID energy was set to 30% for the first and 35% for the second MS/MS event of each duty cycle. The default charge state was set to $z = 9$, and the activation time was set to 30 ms. Additional higher-energy collisional dissociation experiments were performed with 35% normalized collision energy, 30 ms activation time, and $z = 9$ default charge state. The mass window for precursor ion selection was set to m/z 2 or 4. A window of m/z 3 was set for dynamic exclusion of up to 50 precursor ions with a repeat of 1 within 10 s for the next 20 s. The intact mass profile was inspected with the Xcalibur Qual Browser (Thermo Xcalibur 2.2 SP1.48), and deconvolution of isotopically resolved spectra was carried out by using the Xtract algorithm of Xcalibur Qual Browser or charge distribution deconvolution by magic transformer (MagTran) software. Protein assignment was done by comparison to the retention time of the HPLC run.

Chromatographic separation of venom sample

The lyophilized crude venom (1 mg) was dissolved to a final concentration of 10 mg/mL in aqueous 3% (v/v) ACN with 1% (v/v) FA and centrifuged at 16,000 × g for 5 min to spin down insoluble content. For further analysis, two identical HPLC runs with 1 mg each were performed and fractionated and equivalent fractions were pooled. The supernatant was loaded onto a semi-preparative reversed-phase HPLC (RP-HPLC) with a Supelco Discovery BIO wide pore C18-3 column (4.6 × 150 mm; particle size, 3 μm) using an Agilent 1260 high-pressure gradient system (Agilent, Waldbronn, Germany). The column was operated with a flow rate of 1 mL/min using ultrapure water with 0.1% (v/v) FA (buffer A) and ACN with 0.1% (v/v) FA (buffer B). We used a standard separation gradient with solution A and solution B, starting isocratically (5% B) for 5 min, followed by linear gradients of 5–40% B for 95 min and 40–70% for 20 min, then 70% B for 10 min, and finally re-equilibrated at 5% B for 10 min. Detection was performed at λ = 214 nm using a diode array detector (DAD). After the chromatographic separation, fractions were vacuum dried and used for bottom-up analysis and in-source decay analysis.

Bottom-up characterization and quantification of the venom proteome

The dried fractions from the aforementioned chromatographic separation were subsequently analyzed by SDS-PAGE (12% polyacrylamide). Afterward, the Coomassie-stained band was excised via in-gel trypsin digestion reduced with fresh dithiothreitol (100 mM DTT in 100 mM ammonium hydrogen carbonate, pH 8.3, for 30 min at 56 °C) and alkylated with iodoacetamide (55 mM IAC in 100 mM ammonium hydrogen carbonate, pH 8.3, for 30 min at 25 °C in the dark).

An in-gel trypsin (Thermo, Rockfeld, IL, USA) digestion was performed (6.7 ng/µL in 10 mM ammonium hydrogencarbonate with 10% (v/v) ACN, pH 8.3, for 18 h at 37 °C, 0.27 µg/band). The peptides were extracted with 100 µL of aqueous 30% (v/v) ACN just as 5% (v/v) FA for 15 min at 37 °C. The supernatant was vacuum dried (Thermo SpeedVac, Bremen, Germany), redissolved in 20 µL of aqueous 3% (v/v) ACN with 1% (v/v) FA, and analyzed by LC–MS/MS analysis.

The bottom-up analysis was performed with an Orbitrap XL mass spectrometer (Thermo, Bremen, Germany) via an Agilent 1260 HPLC system (Agilent Technologies, Waldbronn, Germany) using a reversed-phase Grace Vydac 218MS C18 (2.1 × 150 mm; particle size, 5 µm) column. The prechromatographic separation was performed with the following settings: after an isocratic equilibration (5% B) for 1 min, the peptides were eluted with a linear gradient of 5–40% B for 10 min and 40–99% B for 3 min, washed with 99% B for 3 min, and re-equilibrated in 5% B for 3 min.

LC–MS/MS data files (.raw) were converted to mascot generic format (.mgf) files via MSConvert GUI of the ProteoWizard package (http://proteowizard.sourceforge.net; v3.0.10577) and annotated by DeNovoGUI[342] (v1.15.11) with the following settings: fixed modifications, carbamidomethyl cysteine (+57.02 Da); and variable modifications, acetylation of lysine (+42.01 Da) and phosphorylation of serine and threonine (+79.97 Da). The deduced peptide sequences were matched against a nonredundant protein NCBI database of *Viperidae* (taxid: 8689) using the basic local alignment tool (BLAST)[343] (http://blast.ncbi.nlm.nih.gov).

For peptide spectrum matching, the SearchGUI[279] (v3.3.11) software tool was used with X!Tandem as the search engine. The MS/MS spectra were searched against (i) the nonredundant *Viperidae* protein NCBI (taxid: 8689, 17th Dec 2018, 1310 sequences); (ii) our custom nonredundant protein database derived from the translation and protein level dereplication of *V. a. senliki* venom gland transcriptome assembly including a total of 141 unique protein sequences, of which 89 were toxin sequences; and (iii) a set of proteins found as common contaminants (common Repository of Adventitious Proteins, cRAP, 116 sequences), containing a total of 1468 sequences. Precursor mass accuracy was set to 10 ppm and 0.5 Da for the MS2 level. The alkylation of Cys was set as a fixed modification, and acetylation of the N-terminus as well as Lys, deamidation of asparagine and glutamine, oxidation of methionine, and phosphorylation as well as glycosylation of serine and threonine were allowed as variable modifications. A false discovery rate (FDR) was estimated through a target-decoy approach, and a cutoff of 1% was applied.

In-source decay

The top-down MS analysis by in-source decay was performed by redissolving vacuum-dried HPLC fractions (two HPLC runs with 1 mg per column load) in ultrapure water. The redissolved fractions (2 µL of ultrapure water) were gently mixed 1:1 (v/v) with the reducing matrix 1,5-diaminonaphthalene (1,5-DAN, 20 mg/mL in 50:50 (v/v) acetonitrile (TA50) and 0.1% FA in

ultrapure water) or 2,5-dihydroxybenzoic acid (2,5-DHB, 20 mg/mL in TA30) and subsequently spotted on a polished ground steel plate (Bruker, Bremen, Germany). The MALDI-ISD measurements were performed on a Bruker ultraflex TOF/TOF device (Bruker, Bremen, Germany). Instrument calibration was performed using a peptide mix (peptide calibration standard from Bruker Daltonics; *m/z* range, 700–3200). The laser power was set to ~50% for experiments with 2,5-DHB and increased to ~70% for experiments using 1,5-DAN. The resulting raw files were analyzed using the Bruker BioTools software (version 2.3). Peak picking was performed with the following settings: a minimal mass difference between peaks of 0.3 Da, threshold (*S/N*) of 1, quality factor threshold of 20, and maximum number of peaks of 1000. For the *de novo* annotation option, the ISD-TOF extended ion list was selected, including a, c, y, and z + 2 ions. A characteristic of MALDI-ISD *de novo* sequencing are short gaps in the c- or z-ion series caused by proline (Pro) residues (secondary amine), named "proline gap", and have to be considered for interpretation of ISD spectra. The following fixed modification was set for top-down analysis: reduced disulfide bridges (+2 Da). Resulting peptide sequences were searched against the nonredundant *Viperidae* protein NCBI (taxid: 8689, 17th Dec 2018, 1310 sequences), our custom nonredundant protein database derived *V. a. senliki* venom gland transcriptome assembly, and a set of proteins found as common contaminants (cRAP, 116 sequences), containing a total of 1567 sequences.

Data accessibility

All transcriptome assembled sequences have been deposited in GenBank using a BankIt sequence submission tool and have been assigned with accession numbers MN831207–MN831363. Transcriptome data, including sample information and RNAseq short reads, have been deposited in the NCBI database (http://www.ncbi.nlm.nih.gov/sra) with the BioProject identifier PRJNA560445. Mass spectrometry proteomics data (.mgf, .raw, .fid, and output files) have been deposited with the ProteomeXchange Consortium[344] (http://proteomecentral.proteomexchange.org) via the MassIVE partner repository under project name "Extended snake venomics on the newly discovered subspecies *Vipera anatolica senliki*" with the data set identifier PXD014805.

Relative toxin quantification

The quantification of venom composition is based on the RP-HPLC fraction integration (UV_{214nm}) in comparison to the total integral of all analyzed fractions based on the protocol of Juárez et al.[223]. In the case of co-eluting toxin components, observable by SDS Coomassie staining, the ratio of optical intensities and densities was respectively used for emphasis of fraction integral portion.

Results and Discussion

Venom of *V. a. senliki*

Herein, we describe the detailed venom composition of the recently discovered subspecies of the Anatolian meadow viper, *V. a. senliki,* by a combination of decomplexing bottom-up and top-down proteomics approaches as well as venom gland transcriptomics. mRNA sequencing of the venom gland was followed by a *de novo* approach to assemble the venom transcriptome for *V. a. senliki*. Further, we applied established protocols for a comprehensive characterization and relative quantification of the venom subsequently extended by MALDI-ISD experiments. Finally, we analyzed and evaluated venom variations on the (sub)species level of closely related European vipers in context to their clinical relevance to human envenoming. The initial analysis of *V. a. senliki* venom was performed by a one-dimensional RP-HPLC separation, followed by a subsequent SDS-PAGE separation and in-gel trypsin digestion as well as a direct online intact mass analysis by ESI-HR-MS. All *de novo* and peptide sequence matches from our transcriptomics and proteomics approaches are listed in more detail in table 6.1 and appendix tables 6.1 and 6.2.

Venom gland transcriptome

Our transcriptome assembly of cDNA from the venom gland of *V. a. senliki* resulted in a total of 158 nucleotide sequences, of which 96 were toxin-related transcripts and 62 were other transcripts commonly known to be expressed in snake venom glands. The 96 toxin-related transcripts were found to belong to 18 venom toxin families. Of these, 32 were full length toxin sequences, whereas 64 sequences were partially assembled toxins. Full length toxins were represented by phospholipase A_2 (PLA_2), snake venom serine protease (svSP), serine protease inhibitor (SP-I) including Kunitz-type SP-I (KUN), nerve growth factor (NGF), vascular endothelial growth factor (VEGF), venom factor (VF), WAP four-disulfide core domain protein (WAP), cystatin (CYS), veficolin (VEF), disintegrin metalloproteinases (DIS-MPs), hyaluronidase (HYAL), glutaminyl-peptide cyclotransferase (GPC), and phosphodiesterase (PDE).

Table 6.1. **Venom proteins and peptides identified from *Vipera anatolica senliki*.** Assignment of venomic components was performed by crude venom intact mass profiling (IMP, method A), IMP of a single RP-HPLC fraction with low molecular mass (method B), bottom-up (BU, method C) and in-source decay annotation (ISD, method D). Fraction numbers are based on the RP-HPLC chromatogram (**Figure 6.1**). Annotation was performed *de novo* and by peptide spectrum matching from in-gel digested protein bands (**Appendix Figure 6.1**). Identification was carried out against a non-redundant Viperidae protein database (taxid: 8689), our custom transcriptome database and a set of proteins found as common contaminants (cRAP). SDS-PAGE and intact mass profile analysis provided the average molecular weight. For IMP only most abundant mass is listed (all masses in **Appendix Table 6.2**). IMP performed by charge-state deconvolution was carried out with MagicTransformer (MagTran) and is marked by #.

HPLC Fraction	Method	RT [min]	SDS-PAGE Band	Mass [Da] (ESI-MS)#	Mass [kDa] (SDS-PAGE)	Transcriptome Hit	No. PSM Hit	No. *de novo* Hit	Protein Family
0	A	1.91	-	347.07	-	-	-	-	n/a peptide
1	B	5.65	-	187.07	-	-	-	-	n/a peptide
2	A; B	7.23	-	717.40	-	-	-	-	n/a peptide
3	A; B	24.11	-	443.22	-	-	-	1	svMP-i
4	A; B	26.75	-	1,128.61	-	-	-	1	BPP
5	A; B	31.09	-	1,128.61	-	-	-	1	BPP
6	A; B	33.01	-	808.39	-	-	-	-	n/a peptide
7	A; B	36.50	-	6,737.95	-	-	-	1	KUN
8	A; B	37.69	-	6,737.95	-	-	-	1	KUN
9	A; C	39.49	9a	13,982.77	14	1	2	1	DI
10	A; C	40.68	10a	13,999.75	14	2	3	1	DI
11	A; C; D	43.90	-	6,636.95	-	1	1	-	KUN
12	A; C; D	45.82	12a	6,627.86	11	3	3	1	KUN
13	A; C	48.69-50.73	13a	22,467.34	27	-	-	1	n/a
14	A; C; D	54.06	14a	-	31	-	-	-	n/a
			14b	27,372.70	25	-	2	2	PLA$_2$
			14c	13,639.90	14	-	5	4	PLA$_2$
15	A; C; D	56.53	15a	22,942.84	30	-	-	-	n/a
			15b	13,639.90	15	-	4	2	PLA$_2$
16	A; C; D	69.62	16a	-	55	13	13	4	CRISP
			16b	-	40	7	7	2	CRISP
			16c	24,660.15	25	13	13	5	CRISP
			16d	18,632.82	20	8	8	2	CRISP
			16e	13,639.95	14	-	3	1	PLA$_2$
			16f	12,330.07	11	4	4	2	CRISP
			16g	8,192.04	10	-	-	2	svMP/DI
17	C; D	71.58	17a	-	55	-	6	5	svMP
			17b	-	40	-	1	1	svMP
			17c	24,660.14	25	8	8	2	CRISP
			17d	13,639.91	15	-	2	1	PLA$_2$
			17e	12,308.06	10	-	-	2	DI

Table 6.1. **Venom proteins and peptides identified from *Vipera anatolica senliki*.** *continued*

HPLC Fraction	Method	RT [min]	SDS-PAGE Band	Mass [Da] (ESI-MS)#	Mass [kDa] (SDS-PAGE)	Transcriptome Hit	No. PSM Hit	No. *de novo* Hit	Protein Family
			18a	-	60	-	1	-	LAAO
			18b	-	55	-	1	2	svMP
			18c	-	50	-	1	2	svMP
18	C; D	73.20	18d	-	40	-	-	4	svMP
			18e	27,493.67	35	5	5	2	svSP
			18f	24,660.13	25	3	3	2	CRISP
			18g	13,638.91	15	-	2	1	PLA$_2$
			19a	-	60	3	3	-	svSP
			19b	-	50	-	2	3	svMP
19	A; C; D	74.73	19c	37,748.44	40	3	4	3	svMP/svSP
			19d	31,578.06	35	4	6	4	svSP
			19e	-	11	-	1	2	svMP/DI
			19f	-	10	-	1	1	svMP/DI
			20a	-	>200	3	3	2	svSP
			20b	-	>200	5	5	2	svSP
20	A; C; D	75.70	20c	-	65	5	5	3	svMP/svSP
			20d	-	50	3	4	3	svMP/svSP
			20e	32,234.29	35	9	9	4	svSP
			21a	-	65	4	4	2	svMP/svSP
21	A; C; D	76.80	21b	-	50	3	3	2	svMP/svSP
			21c	32,889.50	40	9	10	3	svSP
			21d	10,962.85	10	-	2	-	PLA$_2$
			22a	-	65	2	2	1	svMP/svSP
			22b	-	50	3	3	3	svMP/svSP
22	A; C; D	77.48	22c	33,549.74#	40	6	10	4	svSP
			22d	24,659.12	20	-	2	1	CTL
			22e	15,070.74	15	-	5	5	CTL
			22f	-	-	-	-	-	n/a
			23a	-	200	-	-	1	svMP
			23b	-	150	-	4	3	svMP
23	A; C; D	79.04	23c	-	50	-	9	4	svMP
			23d	24,451.00#	20	-	3	3	CTL
			23e	13,665.80#	14	-	7	6	CTL

While alternate isoforms from some of these families were also represented as partially assembled contigs, additional toxin families such as cysteine-rich secretory protein (CRISP), snake venom metalloproteinase (svMP), 5′-nucleotidase (5NUC), and dipeptidyl peptidase (PEPT) were wholly represented by partial contigs. The full set of venom toxins retrieved from the transcriptome and the breakdown of number of full and partial isoforms in each toxin family are shown in figure 6.1C and appendix table 6.1.
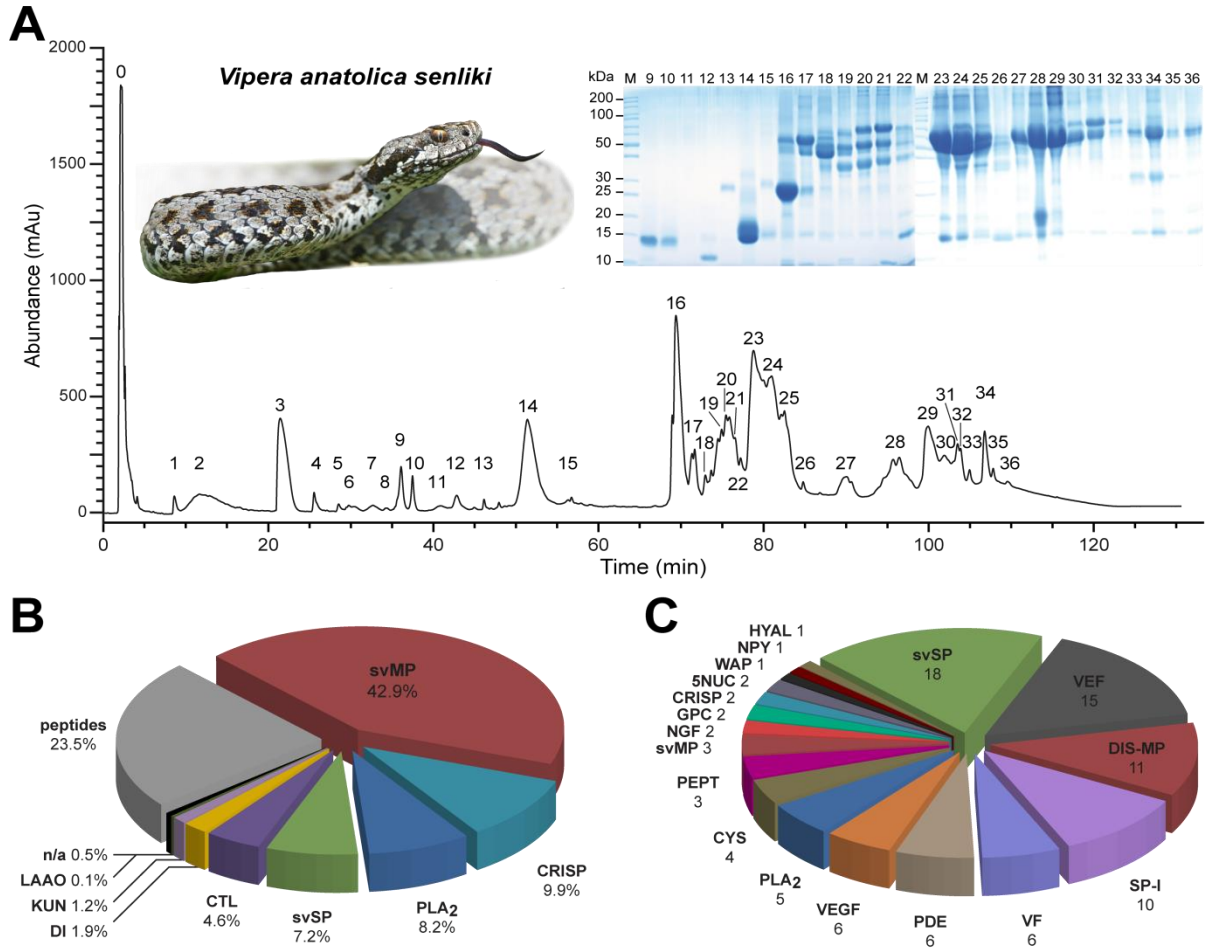
Figure 6.1. Snake venomics analysis of *Vipera anatolica senliki* using venom proteomics and venom gland transcriptomics. (**A**) *V. a. senliki* venom fractions of the RP-HPLC separation, with UV$_{214nm}$ detection signal, were subsequently subjected to 1D-SDS-PAGE, followed by tryptic digests of the most abundant bands. PAGE line nomenclature based on HPLC fractions. A detailed band nomenclature is shown in appendix figure 6.1. (**B**) The relative quantitative occurrence of the toxin families from *V. a. senliki*. Peptides (grey) are summarized of different known (e.g. svMP-i and BPP) and unknown molecular mass represented by pie chart (**Appendix Table 6.2**). (**C**) Numbers of different transcripts of the toxin families identified from transcriptome assembly are represented by pie chart (**Appendix Table 6.1**). All toxin families identified by integrative venomics: 5'-nucleotidase (5NUC), C-type lectin-like protein (CTL), cysteine-rich secretory protein (CRISP), cystatin (CYS), disintegrin (DI), dipeptidyl peptidase (PEPT), disintegrin-metalloproteinase (DIS-MP), glutaminyl peptide cyclotransferase (GPC), hyaluronidase (HYAL), Kunitz-type protease inhibitors (KUN), L-amino acid oxidase (LAAO), nerve growth factor (NGF), proneuropeptide-Y (NPY), not annotated (n/a), phosphodiesterase (PDE), phospholipase A$_2$ (PLA$_2$), serine protease inhibitors (SP-i), snake venom metalloproteinases (svMP), snake venom serine proteases (svSP), veficolin (VEF), vascular endothelial growth factor (VEGF), venom factor (VF), and WAP four-disulfide core domain protein (WAP).

We performed an analysis of toxin gene expression in the venom gland by mapping quality filtered RNAseq reads back to the assembled venom gland transcriptome. We use TPM values that are normalized for both transcript length and read depth to represent gene expression. We found that the most highly expressed toxins were SP-I (38%), CYS (27%), VEF (18%), and svSP (6%), followed by DIS-MP (3%), VF and VEGF both at 2%, and PLA$_2$ and WAP both at 1%. Other families such as svMP, CRISP, 5NUC, and NGF showed negligible expression (<1%), or as in the case of GPC,

no quantifiable expression was found. Prior to venom gland collection for transcriptomics studies, venom is usually extracted a few days in advance to start the gene expression cascade. Venom gene expression and protein resynthesis are known to peak between 3 and 7 days after venom extraction.[411] In the case of *V. a. senliki*, however, venom extraction was performed 10–14 days in advance of gland dissection under anesthesia. Our results show that high expression of inhibitors such as SP-I (132,326 TPM) and CYS (92,421 TPM) likely followed the peak expression of major venom components such as svSP, svMP, or PLA$_2$. This is further supported by results from venom proteomics analysis where svMP, PLA$_2$, and CRISP were readily found to be the most abundantly represented toxin families. While SP-I inhibits the activity of serine proteases, the CYS family is thought to play a crucial role in preserving venom potency by acting as strong inhibitors of cysteine proteases of the prey that can inactivate snake venom proteins.[412] CYS being the second highest expressed family in our study provides an interesting snapshot of the CYS expression pattern in the venom gland assigned to post peak venom synthesis.

Bottom-up venomics

The decomplexation of the *V. a. senliki* venom by RP-HPLC shows 36 characteristic fractions (F), which were subsequently fractionated in the second dimension by SDS-PAGE and revealed the presence of 95 separate protein bands (**Figure 6.1A**). The resulting sequence tags, after in-gel trypsin digestion, were searched against the NCBI nonredundant *Viperidae* protein database using BLASTP. The 266 *de novo* annotated sequence tags resulted in the identification of several proteins, covering seven toxin families (**Appendix Table 6.2**). The most abundant protein family is represented by svMP, followed by CRISP, PLA$_2$, CTL, svSP, and DI, whereas a small percentage (~1%) could not be annotated (n/a). Another abundant part of the venom is formed by peptides, which we further investigated by the direct measurement of the nine first fractions (F 0–8) and IMP analysis described below. *De novo* sequence assignment of the MS/MS spectra of small peptides (F 0–8) resulted in the identification of a bradykinin potentiating peptide (BPP) and a snake venom metalloproteinase inhibitor (svMP-i) pEKW that coincides with MS/MS spectra in the literature (**Appendix Figure 6.3** and **Appendix Table 6.2**).[321,358]

The reanalysis of the MS/MS data from the tryptic peptides was performed by peptide spectrum matching (PSM), using the assembled custom toxin transcriptome database, the NCBI *Viperidae* protein database, and a list of commonly found contaminants (cRAP). This resulted in 346 peptide matches and nine major toxin families in total. This output is a slight improvement of annotated spectra in comparison to the *de novo* annotation (**Table 6.1**), which can be attributed to the venom gland being replete rather than in a venom regeneration state.[413] The vast majority of previously annotated *de novo* peptides could be confirmed by PSM analysis but also revealed, in contrast to the *de novo* sequencing, the presence of a KUN and an L-amino acid oxidase (LAAO). Furthermore, peptide matching analysis corrected a previous false annotation from *de novo* sequencing of two

serine/threonine protein kinases and a predicted zinc finger protein. Indeed, *de novo* sequencing is an error-prone process complicated by uneven fragmentation patterns due to missing fragmentation, limited algorithm accuracy for various interferences, homeometric peptides, low resolution, and the lack of a species-specific protein database.[414] Therefore, the corresponding spectral peptide matches showed the presence of svSP and CTL instead. Overall, our PSM showed a modest improvement, with about half of all obtained peptide matches were from viperid sequences in the NCBI database and 172 protein IDs sourced from transcriptome derived sequences. The peptide matches mainly represent hits against the transcriptome database for the toxin families svMP, svSP, and CRISP. In some cases, where no hits were obtained by peptide spectrum matching, *de novo* sequence tags were resorted to classify the protein. The relative quantification showed svMP (42.9%) as the most abundant toxin family, followed by CRISP (9.9%), PLA_2 (8.2%), svSP (7.2%), CTL (4.6%), DI (1.9%), KUN (1.2%), and LAAO (0.1%). In the small molecular mass range (<2 kDa), the following peptides were found: svMP-i (5.9%), bradykinin potentiating peptides (BPPs, 0.6%), and unknown peptides (17.0%) of the overall venom composition (**Figure 6.1B**).

Furthermore, we were able to identify several biological PTMs attached to the tryptic digested peptides (e.g., pyroglutamylation, glycosylation, or acetylation) as well as other modifications such as experimental artifacts (e.g., oxidation and deamidation) for various venom protein families (**Appendix Figures 6.3–6.5**). The pyroglutamylation by glutaminyl cyclase is a common N-terminal modification of snake venom peptides.[322,415] The formation of cyclic pyroglutamate protects venom peptides from enzymatic degradation by exopeptidases or induces conformational changes to improve receptor binding and therefore can indirectly contributing to venom toxicity.[416] In our study, we identified the tripeptidic metalloprotease inhibitor pEKW (F 3) by tandem MS (**Appendix Figure 6.3**). The peptide mass spectra of several other fractions (F 19–25, except F 23) showed mass shifts, allowing the identification of various glycosylations assembled by hexoses (Hex, 162.05 Da), *N*-acetyl-hexoseamines (HexNAc, 203.08 Da), and *N*-acetyl-neuraminic (NeuAc, 291.09 Da) (**Appendix Figure 6.4A,B**). Highly complex glycans are large and hydrophilic modifications that can play an important functional role in acting as a sterically protection group against proteolysis, solubility tag, or antigenic recognition of venom proteins.[417] A complete reconstruction of the glycosylation patterns were prevented due to insufficient tandem mass spectra. Another well-known biological PTM, the acetylation of lysine, was identified at a svMP disintegrin domain (F 29b) by peptide annotation (**Appendix Figure 6.4C**). The acetylation can act as a crucial regulator and assume several functions, like spatial localization of the protein, regulation of the enzymatic activity, or stability against degradation, which can directly affect the toxicity.[418]

The bottom-up analysis of dynamically regulated and underrepresented alterations is challenging because they are attached to specific amino acid residues and delocalized in a large environment

such as high-molecular-weight venom proteins. Furthermore, sample preparation is complicated due to increased instability during MS experiments and hydrophobicity of several modifications, which may affect the cleavage efficiency of proteases or reduce ionization and detection efficiency in MS.[419] Moreover, several experimental artifacts, which occur *in situ* during sample preparation or analysis, can even affect the identification accuracy of biological relevant PTMs (**Appendix Figure 6.4D**).

To bypass the problem of false or non-annotated peptide matches, a combination of accurate transcriptome assembly or even an alternative bottom-up method, like in-solution shotgun proteomics, could be useful for a targeted identification. However, shotgun proteomics suffers from a lowered mass spectral sensitivity and quantitative breakdown of snake venom composition due to an inefficient differentiation of the numerous toxin isoforms present in crude venoms. Ultimately, employing top-down proteomics combined with a front-end chromatographic separation is a suitable alternative to uncover existing proteoforms and PTMs and to identify toxin components by tandem MS or disulfide bond counting.[294,420]

Top-down venomics

The direct online intact mass profiling by ESI-HR-MS of *V. a. senliki* venom coupled to a front-end LC-based decomplexation resulted in a total ion chromatogram profile (**Appendix Figure 6.2**) comparable to the previous chromatographic separation (fraction nomenclature is based on Figure 6.1). The initial measurement of native venom generated an overview of 67 compounds by intact mass, including low abundant compounds and small peptides. The initial fractions (F 0–8) exclusively contain smaller peptide components (>7 kDa), like svMP-i and BPP as well as a KUN proteoform (F 7 and 8). The aforementioned svMP-i (pEKW) in F 3 and BPP in F 4 and F 5 are detectable but less prominent compared to the semi-preparative separation, which can be explained by low ionization yields of the peptides by electrospray ionization (**Appendix Figure 6.2**). The following fractions (F 9–15) are toxin families in the molecular mass range of 10–30 kDa composed of two heterodimeric disintegrin proteoforms, three KUN proteoforms, and one PLA$_2$. Subsequent fractions (F 16–36) show intact masses across the entire molecular mass range from different venom protein families (10–70 kDa), like svMP, CRISP, LAAO, svSP, or CTL with various proteoforms (**Appendix Table 6.2**).

In addition to the native mass profiling, crude venom was further investigated by chemical reduction using the reagent TCEP. The comparison of native and chemically reduced venom components allows for the identification by inter- and intramolecular disulfide bonds as an important characteristic for several toxin families.[294] A closer look to the native and reduced venom components showed the identification of three different venom protein families (**Figure 6.2**). The intact mass for fraction 7 in the native venom showed two distinct molecular masses (6738.0 and 7280.2 Da) in the typical range for Kunitz-type inhibitors (**Figure 6.2A,B**,

blue). Both observed monoisotopic masses (M_{mono} = 6733.94 and 7276.20 Da) are identical to the theoretically calculated masses of two proteoforms from a transcriptomics database serine protease inhibitor entry (GenBank: MN831324). Both KUN lack an N-terminal signal peptide (MSSGGLLLLLGLLTLWAELTPVSG) and start with an N-terminal pyroglutamic acid. The difference between these proteoforms is a loss of the C-terminal MGRPT, a known propeptide, for example, in the highly similar Kunitz-type serine protease inhibitor Vur-KIn (NCBI: P0DKL8.1) from *Vipera renardi*.[421] The Kunitz-type inhibitor in fraction 12 is a third proteoform of our GenBank entry (GenBank: MN831324), which mass correlates with an N-terminal truncation and pyroglutamylation in addition to the C-terminal propeptide loss. However, to fully confirm the suspected protein family, we compared it to the reduced venom profile and found the corresponding molecular masses that were shifted by Δ6.0 Da (**Figure 6.2A, B**, red). The individual mass shifts indicate the presence of three disulfide bridges, which is a characteristic for Kunitz-type inhibitors in snake venoms.[422]

In combination to the previous *de novo* annotation and PSM, we were able to identify fraction 14 as a single basic PLA$_2$ by disulfide bond counting (**Table 6.1**). The comparison of native (13,639.9 Da) and reduced (13,654.0 Da) mass spectra showed a mass shift of Δ14.1 Da, which correlates to seven disulfide bridges and is a clear characteristic for snake venom PLA$_2$ (**Figure 6.2C**)[423]. We were able to further characterize four different DI subunits (α, 6979.0 Da; β, 6998.0 Da; γ, 7022.0 Da; and δ, 7023.9 Da), forming the three dimeric proteoforms 13,982.8 Da (αδ), 13,999.8 Da (βγ), and 14,001.8 Da (βδ), present in fractions 9 and 10. Both proteoforms contain two intersubunit (Δ2.0 Da) and twice four intra-subunit (Δ8.1 Da) disulfide bridges (**Figure 6.2D–I**).[424] Moreover, we were able to identify again small peptides, like the svMP-i pEKW by tandem MS (**Appendix Figure 6.3**).

Beside the aforementioned N-terminal pyroglutamic acid in KUN inhibitors and the disulfide bond counting, we were able to assign several glycosylation fragments (F 19–22) by IMP (**Appendix Figure 6.5**). The corresponding masses (~31 kDa) were previously identified by BU as serine proteases (GenBank: MN831305–MN831307). The transcriptional sequences showed two potential *N*-glycosylation sites N-X-S/T as well as several serines and threonines for *O*-glycosylation, whereby svSPs are the most frequently *N*-glycosylated venom proteins.[425] The three observed glycan masses, like the previously described tryptic digested peptides of these fractions, show a multiple building block structure (HexNAc-Hex-NeuAc), which is a typical pattern for glycosylation branches. In F 21, this pattern was observed three times directly in a row (**Appendix Figure 6.5C**). The HexNAc-Hex motif was often detected in all four fractions and more abundant as a combined mass difference of 365.13 Da, instead of single Hex or HexNAc losses.
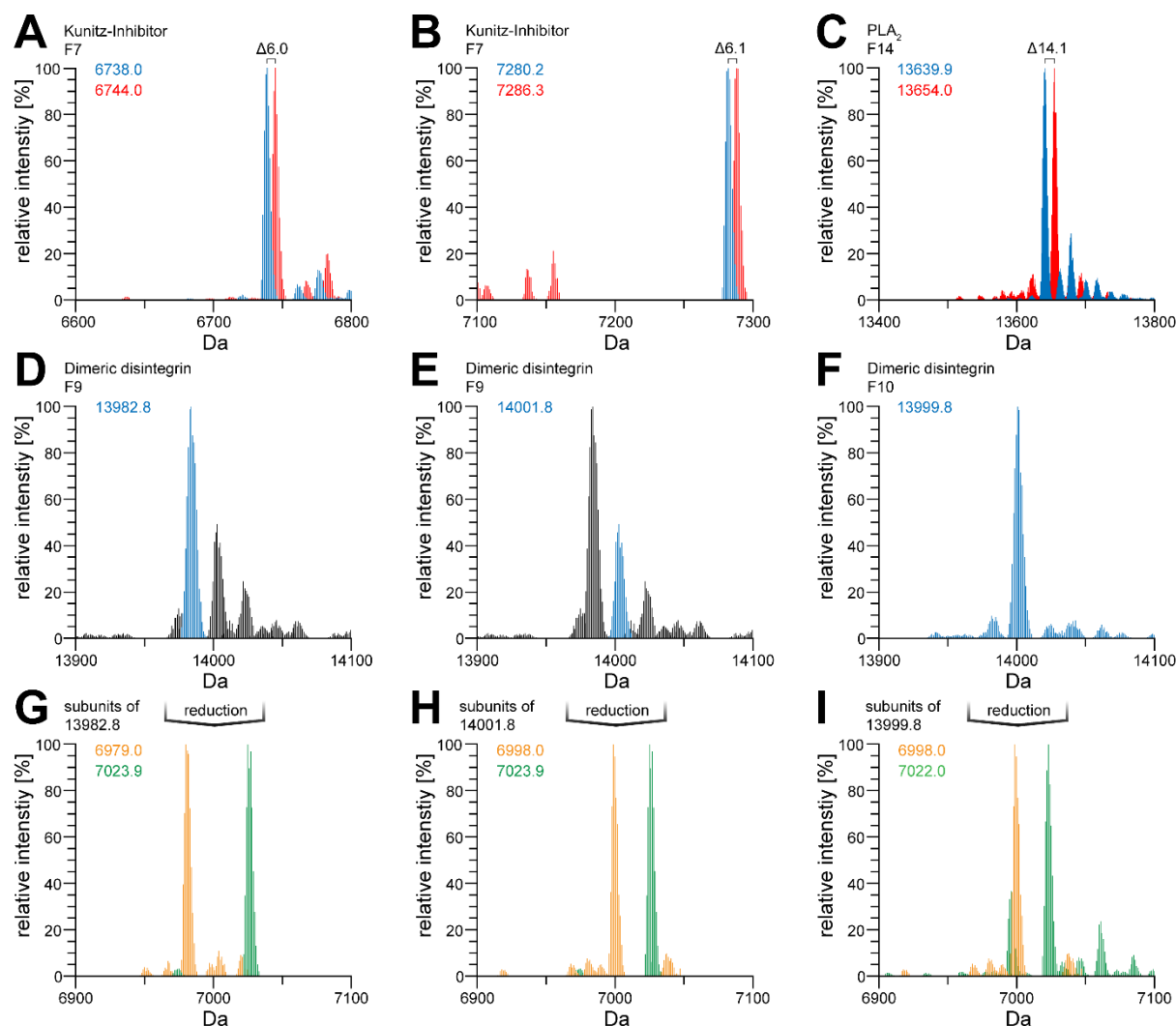
Figure 6.2. Intact mass profiling and disulfide bond mapping of selected toxin components from *Vipera anatolica senliki*. The deconvoluted native and reduced venom profiles show typical mass shifts in several fractions for three different venom protein families. (**A, B**) Venom fraction F 7 includes two Kunitz-type serine protease (KUN) inhibitors with three (Δ6 Da) and (**C**) Venom fraction F 14 shows a single phospholipase A_2 (PLA_2) with seven disulfide (S-S) bonds (Δ14 Da). (**D, E, F**) Venom fractions 9 and 10 include in total the three main dimeric disintegrins (DI) with two inter-subunit S-S and four intra-subunit S-S bonds each (in total Δ20 Da), shown by reduced venom (**G, H, F**).

However, the top-down venomics approach applied here suffers from distinct drawbacks of insufficient ionization by denaturing electrospray ionization and poor isotope resolution for a number of main toxin components, such as high-molecular-weight svMPs or LAAOs, that form a substantial part of viperid venoms. For this reason, we established an alternative top-down method, which offers the same advantages without the limitation of molecular weight that is ideal for venom protein families of viperid species.

Top-down by in-source decay

A possible way to overcome the abovementioned difficulties is the application of MALDI coupled to a time-of-flight (TOF) mass analyzer. MALDI-TOF is a valuable tool for intact mass

analysis of proteins in a higher molecular mass range. In combination with 1,5-DAN as a matrix compound, which allows for a partial reduction of disulfide bonds and significantly enhanced yields of ISD fragmentation, MALDI-TOF becomes a fast and efficient method for top-down venom sequencing. Here, we go beyond a previous proof-of-concept study by Quinton et al.[399] and apply MALDI-ISD to analyze a snake venom from the *Viperidae* family to particularly identify high molecular venom components.

First, crude venom of *V. a. senliki* was front-end fractionated before MALDI-ISD analysis, similar to the initial HPLC separation of the aforementioned bottom-up analysis. The prefractionation is an indispensable requirement for top-down annotation by MALDI-ISD due to the lack of precursor ion selection during the ionization process in the MALDI source. The venom fractions were either prepared with the reductive 1,5-DAN matrix or nonreductive DHB in a ratio 1:1 (v/v) and subsequently spotted on the MALDI target to allow for crystallization. The direct comparison of reduced and non-reduced spectra facilitates a classification of venom protein families by disulfide bond counting, as described for the aforementioned top-down venomics approach. The non-reduced intact mass signal of fraction 11 shows a mass of 6630.1 Da, which is an indicator for a Kunitz-type serine protease inhibitor (**Figure 6.3A**).[426] The comparison with the reduced intact mass (6636.2 Da) reveals a mass shift of Δ6.1 Da (corresponding to three disulfide bridges), which confirms the Kunitz-type architecture. Furthermore, we were able to identify several PLA$_2$ proteoforms in fractions 14 and 15 by disulfide bond counting (**Figure 6.3B**). The most abundant native intact masses of fractions 14 and 15 were 13,637.4 or 13,645.6 Da, respectively, and thus in the molecular mass range of PLA$_2$.[427] The corresponding reduced masses of fractions 14 and 15 (13,652.5 and 13,660.7 Da) show a mass shift of ~Δ14 Da, which corresponds to seven disulfide bridges, a typical feature of PLA$_2$s. Further assignments of venom components by disulfide bond counting were not possible due to the low signal intensity of intact protein masses. The low signal intensity of the reduced intact masses (signal loss, $\sim10^3$ in abundance) makes disulfide bond counting more difficult particularly for proteins present in low amounts.

The effect contributing to much lower signal intensity is caused by hydrogen transfer and the following radical-induced cleavage of the peptide backbone. However, this is a desired process for venom proteome annotation by sequence alignment.[394–396] The peptide backbone cleavage caused by transfer of hydrogen radicals allows for MALDI-TOF top-down sequencing predominantly from the N-terminal part of several toxin components (**Figures 6.3** and **6.4**, **Appendix Figure 6.6**, and **Table 6.1**). The mass range below 1000 Da was only partially considered for *de novo* sequencing due to the intense matrix background.[392] Nevertheless, MALDI top-down sequencing (TDS) provides a straightforward protein sequence analysis approach for small but also high molecular mass toxin components.
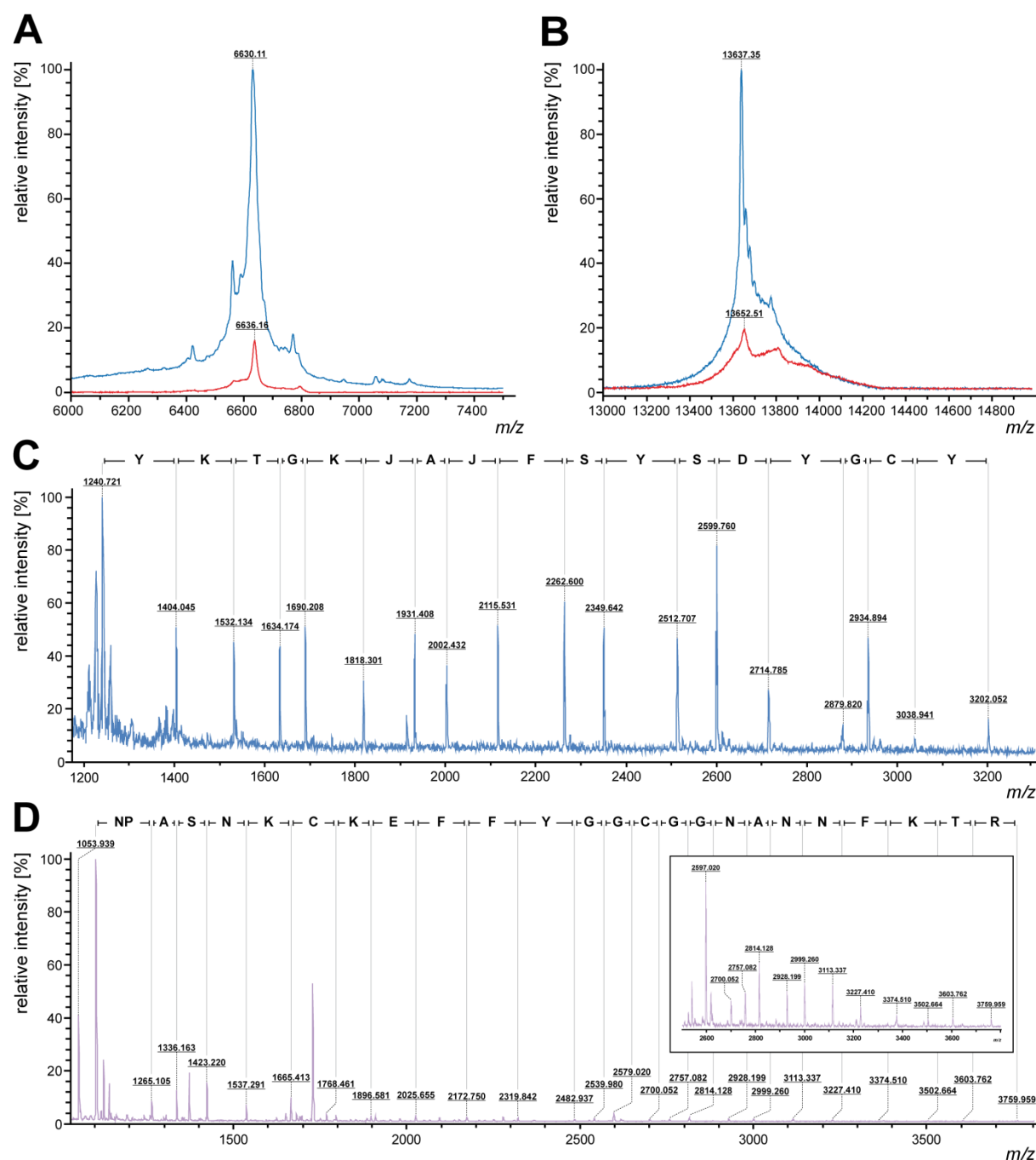
Figure 6.3. MALDI top-down sequencing by in-source decay and disulfide bond mapping of venom components from *Vipera a. senliki*. The transfer of hydrogen radicals induced by 1,5-diaminonaphthalene (1,5-DAN) matrix and the following peptide backbone cleavage allow MALDI-TOF top-down sequencing of high molecular mass toxin components by N-terminal c and z+2 fragment ions. (**A**) Comparison of native (DHB) and reduced (DAN) intact mass spectra (F 11) allow identification of 3 disulfide bridges, characteristic for Kunitz-type serine protease (KUN) inhibitor. (**B**) Comparison of native (DHB) and reduced (1,5-DAN) intact mass spectra (F 14/15) allow identification of 7 disulfide bridges, characteristic for phospholipase A$_2$ (PLA$_2$). (**C**) Identification of a PLA$_2$ proteoform (F 14/15). (**D**) Identification of a KUN inhibitor (F 12). No distinction can be made between leucine and isoleucine (J = Leu or Ile).

In our analysis, we obtained various specific c_n-fragment ion series in the reflectron mode by ISD fragmentation (reISD). The aforementioned protein constituents, like KUN and PLA$_2$ (F 11 and 14/15), identified by the number of disulfide bonds as well as peptide spectrum mapping, could be additionally confirmed by ISD sequence tags. The identity of the PLA$_2$ (F 14/15) was underpinned by a 17-mer (YKTGKJAJFSYSDYGCY), which showed a high correlation with the PLA$_2$ Ammodytin I2(D) isoform (NCBI: CAE47222.1) (**Figure 6.3C**). Additionally, also the Kunitz-type protease inhibitor (F 12) could be manually assigned by an internal 25-mer peptide sequence (NPASNKCKEFFYGGCGGNANNFKTR), which resulted in a specific hit to our transcriptome database (GenBank: MN831324) (**Figure 6.3D**). In this context, it should be mentioned that the matching transcriptome sequence is a precursor protein composed of an N-terminal signal peptide and a protein core sequence, which is post-translationally processed to the three mature KUN proteoforms, as already described and discussed for the intact mass profiling section. This example shows the necessity of top-down proteomics examinations in integrative venomics, which display sequences of mature proteins, even if the bottom-up approach is not able to provide any terminal sequence confirmation.

In addition to previous *de novo* bottom-up sequencing and its subsequent peptide matching to our in-house database, we confirmed the CRISP proteoform (F 16) by a manually assigned 23-mer peptide fragment (KPEJQNEJJDJHNSJRRSVNPTA) that also matches to the entry of our species-specific transcriptome (GenBank: MN831241) in the N-terminal region (**Figure 6.4A**). Furthermore, we verified a svMP proteoform (F 17) by detection of a 22-mer peptide sequence (VEJWRKKDJJNVVSSSDNTJNS) with a high homology to a metalloproteinase (NCBI: ADI47725.1) from *Echis carinatus sochureki* (**Figure 6.4B**). In addition, we found a short N-terminal sequence of the CRISP proteoform, identified in fraction 16 (**Appendix Figure 6.6D**). In fraction 29, we annotated several svMP proteoforms, which were confirmed by three 15/16-mer sequences (EJVJVVDNVMFX$_1$KYX$_2$ with X$_1$ = K/R and X$_2$ = K/(NG)) and are existing proteoforms in the public NCBI *Viperidae* proteome database. In addition, we identified the first annotation sequence as a modified proteoform to one of our database entries (GenBank: MN831329) (**Figure 6.4C**). The top-down ISD sequencing of svMP (F 31/32) matches, in part, our transcriptome data (GenBank: MN831329) and led to the identification of two different proteoforms with a 12-mer (YVEJVJTVDHRM) and 14-mer (JVJVVDNVMFKKYK) peptide in the N-terminal region (**Figure 6.4D**). These examples show the advantage of ISD top-down sequencing for the identification of related proteoforms. In addition to these examples, we confirmed several other proteoforms from various snake venom constituents by peptide fragments of different lengths (**Appendix Figure 6.6** and **Table 6.1**).
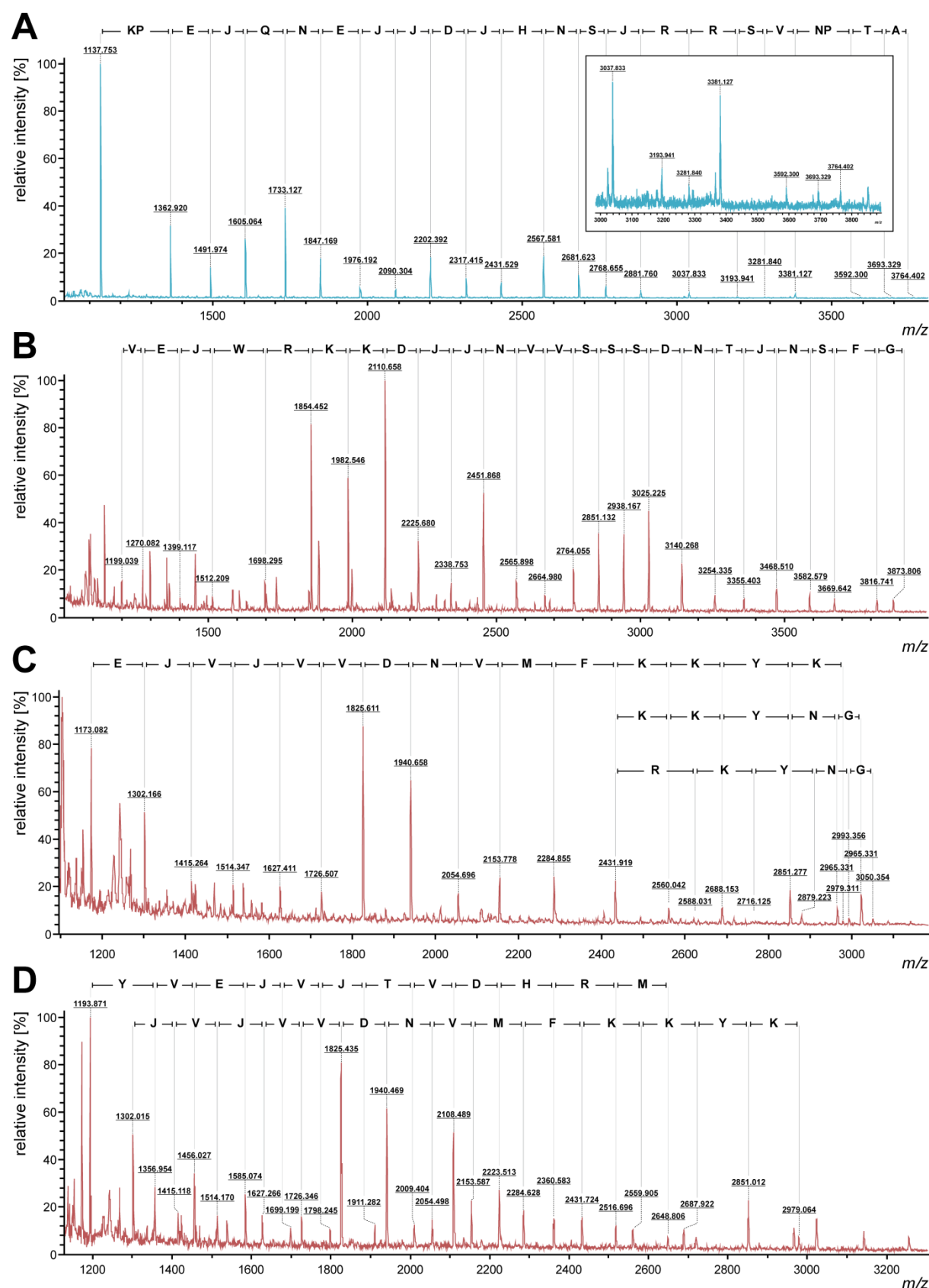
Figure 6.4. MALDI top-down sequencing by N-terminal in-source decay of venom components from *Vipera a. senliki.* (A) Identification of a cysteine-rich venom protein (CRISP) (F 16). **(B)** Identification of a snake venom metalloproteinase (svMP) (F 17). **(C)** Identification of different svMP proteoforms (F 29). **(D)** Identification of different svMP proteoforms (F 31/32). No distinction can be made between leucine and isoleucine (J = Leu or Ile).

Interestingly, the majority of top-down ISD sequences are venom proteins of higher molecular weight, for example, CRISP, svSP, and svMP. However, top-down sequencing by ISD-MALDI suffers from low sensitivity and the restriction to highly concentrated, pure protein samples. The presence of highly blended peptide or protein samples (**Appendix Figure 6.6**) in tandem MS can lead to an overlap of several c_n-fragment ion series of different protein families, which, in turn, could be problematic for peptide and protein identification especially in combination with low abundant signals.[428] In addition, aforementioned T^3-sequencing was performed by MALDI LIFT-TOF/TOF for different toxin families, but based on low signal intensities or overlapping masses for a parent ion with an acceptable small selection window, a sufficient fragmentation was not possible.

Intra- and interspecific venom variation in closely related Eurasian vipers

After the description of the newly discovered subspecies *V. a. senliki*, the populations in Kohu Mountain, Çığlıkara Cedar Reserve became nominotypic (*Vipera anatolica anatolica*).[403] Therefore, a comparative analysis between these two subspecies in context with other closely related and geographically proximal *Vipera* species could provide insights into interspecific and regional venom variation, which has considerable relevance, both for acquiring a better understanding of the pathology of envenomation, as well as for the development of improved treatment of snakebite envenomation.

The venoms of the two *V. anatolica* subspecies are highly similar, both in peak shape as well as fraction intensities in RP-HPLC chromatograms (**Appendix Figure 6.8**). While the *V. a. anatolica* venom profile is sourced from a previous study, both subspecies were analyzed on the same device with identical quantities of crude venom. The retention time-based lineup of the venom peaks shows this similarity in observed toxin families as well as identical intact masses, like in the prominent peaks 3 (svMP-i), 4 (BPP), 9 (DI), 14 (PLA$_2$), and 16 (CRISP) of both subspecies (**Appendix Table 6.3**).[233]

An overall comparative analysis of venom from the two subspecies shows similarity in toxin compositions with some particular features (**Figure 6.5** and **Appendix Table 6.3**). Several svMP variants form the most abundant toxin family in both *V. a. senliki* and *V. a. anatolica* with 42.9% and 41.5%, respectively. This is followed by CRISP, which was at 9.8% in *V. a. senliki* but slightly more dominant in *V. a. anatolica* (15.9%). Snake venom CRISPs are non-enzymatic proteins in the molecular mass range between 20 and 30 kDa and possess ion channel blocking activity.[339] The aforementioned mass profiling showed identical exact masses (average mass, 24,660 Da) for CRISP proteoforms in both *V. anatolica* subspecies. The corresponding transcriptomics entry in our database (GenBank: MN831241) share a high sequence similarity to CRISP of *Vipera ammodytes ammodytes (V. am. ammodytes)*[323] (NCBI: AMB36337.1) (98%) and *Vipera berus*[429] (97%). We assume that the slight different proportion in both subspecies does not result from

various CRISP variants but originates from different expression levels of the CRISP proteoform attributed to separated geographic locales. In addition, PLA_2 (∼8%), DI (∼2%), and peptides (∼22%) do not show major variations.
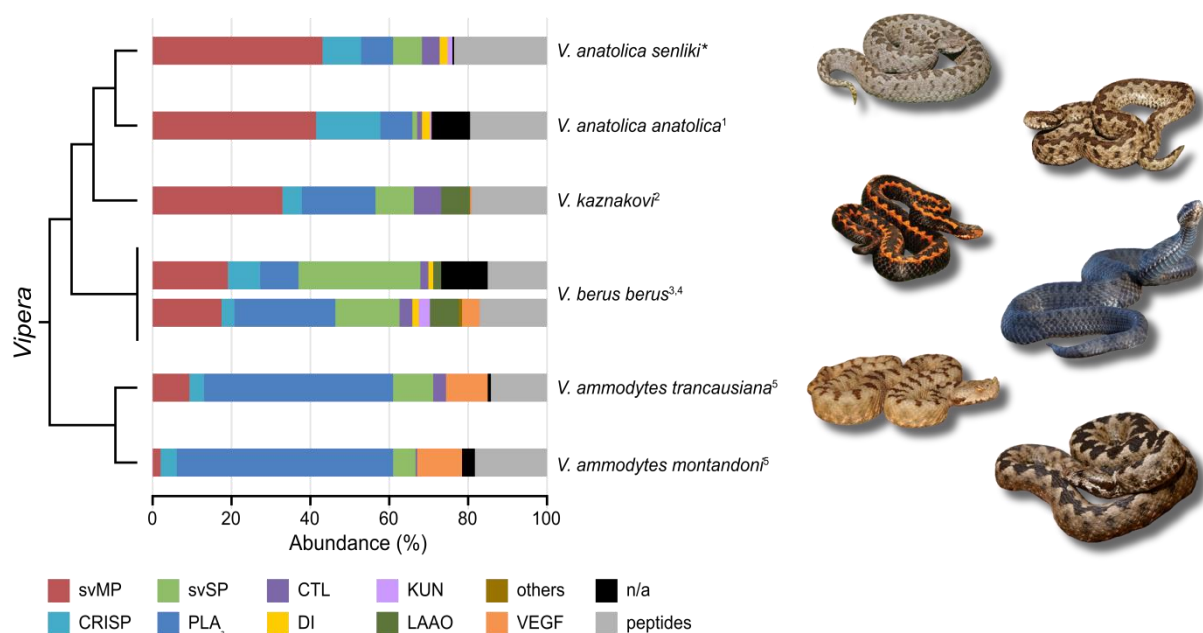

Figure 6.5. Comparative venom proteomic data from six closely related members of the *Vipera* genus in the region of the Black Sea. Seven venoms of six different *Vipera* (*V.*) species and subspecies are compared by their composition. Taxonomic relations based on Alencar et al.[430] and Göçmen et al.[403] are shown without genetic distances. The asterisked data set is a result of this study. The origins of toxin ratios are marked top to bottom numerically: 1 - Göçmen et al.[233], 2 - Petras et al.[431], 3 - Latinović et al.[323], 4 - Al-Shekhadat et al.[432], 5 - Hempel et al.[359]. Snake images correspond in the order to the cladogram species. Image of *V. berus berus* by courtesy of Joshua Baal.

On the other hand, some venom components are more abundant in *V. a. senliki* as opposed to *V. a. anatolica*, including svSP (7.1 to 1.6%), CTL (4.6 to 1.1%), and Kunitz inhibitors (1.2 to 0.3%). Snake venom serine proteases are known to affect various stages of the blood coagulation system, activate platelets, and directly act upon fibrinogen, which disrupts the hemostasis of envenomed prey or victims.[433] It have been previously reported that accelerated evolution[434], exon switching[435], and point mutations[436] can be involved in the generation of novel svSPs due to adaptation to different geographical locations or prey and could therefore be a deciding factor for effective antivenom treatment. The bottom-up venom analysis of *V. a. anatolica* identified one svSP with a high sequence agreement to the serine protease nikobin from Nikolsky adder (*Vipera nikolskii*). In contrast, in our proteotranscriptomics approach, the closely related subspecies *V. a. senliki* revealed several svSP proteoforms, which, however, also have a high sequence similarity (>83%) to the svSP of *V. nikolskii*. It is also interesting to note that the toxin family LAAO was found only in *V. a. senliki* as well as an additional disintegrin (DI) subunit dimerization (**Figure 6.2** and **Appendix Figure 6.7**). While the *V. a. anatolica* venom contains only two dimeric DI (13,982.8 and 14,001.8 Da), the *V. a. senliki* venom includes a third abundant DI mass of

13,999.8 Da. The lack of this dominant fraction 10 and, correspondingly, of the mass in the *V. a. anatolica* venom suggests that the 7022.0 Da subunit is a DI specific to *V. a. senliki*.

In summary, we can conclude that the intraspecies differences of the major toxin families are minor due to the identification of the same toxin classes with partly identical proteoforms or high sequence homologies. Nevertheless, it should be mentioned that a relatively large part of the *V. a. anatolica* venom is still unknown (10%) so that the existing intraspecies differences could be based on that missing information compared to the *V. a. senliki* venom.[233] A reason for these only minor differences between the two *V. anatolica* subspecies seem to be based on similar influences due to likewise habitats, found in doline-rich rocky mountain steppes, and insectivorous diets. The orthopteran-based diet of both meadow vipers, like grasshoppers and crickets, has been described, which, however, is extended by centipedes and scorpions for the *V. a. senliki*.[402,403,437] However, insights into common snake toxin modifications on a subspecies level are very limited because only pyroglutamylation and disulfide bonds were reported for *V. a. anatolica*. Herein, we observed several other PTMs by peptide spectrum matching at the example of several tryptic digested peptides or by intact mass profiling, like different covalent attachments of functional groups or propeptide cleavage (detailed listed in **Appendix Table 6.2**).

Like in this study, some species (e.g., *V. ammodytes*) show just minor intraspecific changes in their venom, while the venom compositions of *Cerastes cerastes* and *Crotalus oreganus* are respectively highly divergent.[359,415,438] In both cases, geographically distant populations were compared, but even between specimen collected from close areas, the venom can be strongly divergent, for example, *Vipera kaznakovi*.[431] This shows that the scope of intraspecific venom variation is strongly diverse and needs further investigation to understand the underlying effects.

Further, we reviewed and compared the previously published venomics data of four closely related and geographically proximal *Vipera* species in relation to *V. anatolica* (**Figure 6.5**). These include the Caucasian *V. kaznakovi*[431]*,* the Eurasian *Vipera berus berus*[323,432]*,* and two subspecies of *V. ammodytes* occurring in Greece and Turkey, *Vipera ammodytes montandoni* and *Vipera ammodytes transcaucasiana*.[359] All of these studies used a consistent snake venomics workflow: quantification with RP-HPLC, 1D-PAGE, and single band bottom-up. While additional quantified venomics studies are available for *V. kaznakovi* (shotgun bottom-up) and *V. b. berus* (2D-PAGE), these were not considered due to variant quantification methods and the direct compositional lineup.[371,439] Generally, svMP/DI, CRISP, PLA$_2$, and svSP are found in venoms of all species and subspecies; however, compositional differences appear to exist between species/subspecies and clades (**Figure 6.5**). The hemorrhagic and tissue destroying svMPs are the most dominant venom components (~40%) in the clade containing *V. kaznakovi* and *V. anatolica* subspecies, whereas myotoxic or neurotoxic PLA$_2$s are the major components (~45%) in *V. ammodytes*. The pathogenesis of venom-induced hemorrhage by svMP involves direct damage of microvessels by degradation of vascular basement membrane components, which profoundly affect the stability

of the endothelium.[440] On the other hand, neurotoxic snake venom PLA_2 causes degeneration of the nerve terminal and skeletal muscle by hydrolysis of membrane phospholipids, whereas myotoxic PLA_2 blocks the post-synaptic acetylcholine receptors (AChR) at the neuromuscular junction.[441] The substantial differences among these viper species predict that a cross-effectiveness of heterologous venoms by a single antivenom is exceedingly problematic. Hence, the neutralization capability may be severely impaired, which should be considered for a proper envenomation treatment. A number of antivenoms for European viper species are commercially available, whereas most of them are monospecific antivenoms and are raised against *V. berus* or *V. ammodytes* venom.[305] Several studies on monospecific antivenoms, such as ViperaTAb, ViperFAV, or Zagreb, showed *in vitro* and *in vivo* immunological cross-reactivity against several venomous European viper species with equipotent cross-reactivity for venoms of related *Vipera* species.[305,442] However, it was also reported that the variability in venom composition for related viper species limits the cross-neutralization of monospecific antivenoms.[305,318,443] A logical bypass is the application of a polyvalent antivenom for related viper species with an effective venom neutralization of variable venom composition, as the recently evaluated polyspecific antivenom Inoserp Europe.[444]

The residual toxin families are only low abundant in some venoms or vary strongly in their presence, for example, VEGF, which is most dominant by 10% in *V. ammodytes*. A consistent aspect among all species, however, is the presence of ~15–20% peptides in the venoms of all *Vipera* venoms compared here. In the case of *V. b. berus*, the two studies show differences in venom compositions and may reflect intraspecific variation at the population level in large geographic areas of species distribution.[323,432] However, this cannot be ascertained as the original localities, from which snakes were collected, are not available in all cases.

Conclusions

Herein, we describe the detailed venom composition of the newly discovered subspecies *V. a. senliki* by an integrative combination of venom gland transcriptomics and decomplexing bottom-up and top-down proteomics.

The transcriptome sequencing revealed 32 full length toxin family transcripts and 64 partially assembled toxins. Relative quantitative snake venomics showed nine venom toxin families with snake venom metalloproteinases (svMPs, 42.9%) as the most abundant protein family, followed by CRISP (9.9%), PLA_2 (8.2%), svSP (7.2%), and CTL (4.6%) as well as DI (1.9%), KUN (1.2%), LAAO (0.1%), and non-annotated components (0.5%). The comparison of native and chemically reduced venom components by established top-down venomics allowed the identification of three venom protein families by inter- and intramolecular disulfide bonds. Furthermore, we identified a high content of peptides (23.5%), including svMP inhibitor (svMP-i, 5.9%) and bradykinin potentiating peptides (BPPs, 0.6%). The comparative venom analysis of the two *V. anatolica*

subspecies showed a high similarity in toxin compositions with some particular features. A broader interspecies analysis of closely related viperid venoms, using consistent snake venomics workflow quantification, gave further insights into the venom proteome distribution. Whether the venom compositions of the reviewed species result from phylogenetic relationships, geographic distribution, or dietary variation could be interesting to investigate in future.

In addition, we extended standard snake venomics strategies by in-source decay top-down sequencing, which enabled disulfide bond counting for KUN and PLA$_2$ by direct comparison of reduced and non-reduced spectra. Furthermore, ISD top-down sequencing renders a sufficient number of terminal mass fingerprints for the characterization of high molecular mass venom components, typically over-represented in viperid venoms, which is an existing limitation to established top-down venomics approaches. Moreover, the method provides sufficient mass fingerprints for several proteoforms even present in one chromatographic fraction, although it was not applicable to sample fractions containing highly mixed toxin components. Top-down sequencing by ISD-MALDI has the potential to overcome persisting limits of current mass spectrometric techniques and allow a deeper insight into the venom composition. Its exploitation as an alternative sequencing method may help to better characterize venom proteomes in combination with established snake venomics approaches. Nevertheless, based on our results, it is still a long way for ISD top-down sequencing to become a high-throughput workflow due to the strongly recommended prerequisite of highly concentrated and pure samples. In this context, a second-dimension chromatographic separation technique (e.g., SEC, IEC, or EBA) is an indispensable step that could help to overcome poor chromatographic resolution. Future investigations including a second prepurification step would enable a more detailed analysis of strongly mixed protein fractions, even in the presence of multiple related proteoforms. However, an orthogonal purification step would go along with expensive costs and a lot of additional preparation time, which, in turn, exclude a high-throughput screening. The software-assisted spectral assignment of venom proteoforms and related PTMs by top-down ISD spectra is a further limitation due to the lack of reliable bioinformatic tools. The manual assignment of such superimposed fragmentation spectra can be a time-consuming process. Therefore, the development of open-source and user-friendly bioinformatic tools would be a first step to turn ISD top-down sequencing into a rapid assignment tool for a broader community. These aspects are currently the main obstacles that prevent ISD top-down sequencing from being an alternative, rapid terminal sequencing method so far. Nevertheless, due to many advantages, ISD top-down sequencing can most likely gain momentum to analyze venom proteomes in its entirety and become a rapid sequencing tool by eliminating existing problems by further experiments and optimizations.

7   Synopsis and Future Perspectives

More than ever, natural toxins play a significant role in human health and the development of novel therapeutic drugs against some of the most serious global diseases.[30,121] The continuous advances of multi-omics technologies and big data analysis are constantly foster novel advances in natural drug discovery.[445] The pathological biology of natural toxins can be fully illuminated by a synergistic use of genomics, transcriptomics, metabolomics, and proteomics as platforms that hold great potential in elucidating high value targets.[446] Advances in next-generation sequencing and viable bioinformatics for analyses of large datasets have led in combination with sustained decreasing costs to an explosion in the number of sequenced genomes and transcriptomes.[447] Particularly, core genomes consisting of essential genes required for primary metabolism are interesting objects that can be used to elucidate potential antibiotic targets.[448] In contrast, genome mutagenesis approaches combined to phenotypic assays can identify genes involved in important aspects of physiology for various toxin producers such as nutrient acquisition, host invasion and infection, or intrinsic resistances to antibiotics.[449] On the other hand, transcriptomics can provide global view of gene expression under different environmental and state physiological conditions. For example, the microbial transcriptome can disclose disease mechanisms for different aspects of host infection or the identification of mode of action for compounds.[450] In addition, comparative proteomics can be consulted to identify gene products of various natural toxins that are expressed under defined environmental conditions, similar to transcriptomic techniques. The application of comparative proteomic approaches give insights on the bacterial response to antibiotics or elucidate resistance mechanisms, which could facilitate the development of inhibitory agents.[451] This is particularly important for the protein interactome and the PTMs of diverse proteomes, because these cannot be recognized by gene sequencing.[452] Metabolomics is used to identify and characterize unknown metabolites within a cell under defined environmental conditions.[453] In contrast to genes and proteins, metabolites provide direct signatures of biochemical activity and can be easier correlated to the phenotype.[454] Furthermore, metabolomic-based clinical applications can contribute for the identification of metabolic causes and biomarkers of chronic diseases and therefore help to understand disease mechanisms from a new perspective.[455] In total, recent advances in 'omics' technologies allow in combination for a rapid analysis, identification and biosynthesis elucidation of natural toxins emerged from biodiverse origins.

In the first part of the present thesis, we have examined the biosynthesis of the secondary metabolite albicidin in more detail and described decisive modifications of the natural phytotoxin. The full structure elucidation of albicidin by a combination of biochemical and analytical experiments in combination with a complete genome from *X. albilineans* paved the way to look for more and minor abundant derivatives in the following. The subsequent analysis of wild type strain cultures by untargeted tandem mass spectrometry revealed a set of different modified albicidin derivatives.[159] In chapter 2, we focused on a promising derivative with a distinct mass shift of 42 Da, which was supposed to be a post-NRPS modification. For this reason, we screened the biosynthesis gene cluster of albicidin for potential modifying enzymes and identified an ATP-dependent carbamoyltransferase (Alb15) by bioinformatic sequence alignments, which is supposed to transfer a carbamoyl moiety to the hydroxy group of the coumaric acid residue at the N-terminus of albicidin. In order to confirm the gene function of *alb15*, we genetically inactivated the corresponding gene and compared crude extracts of the wild type and the mutant strain (Δ*alb15*). The comprehensive characterization was performed by *in vitro* experiments and subsequent mass spectrometric product ion scan analysis with heterologous expressed Alb15. Because of the low abundance in wild type strain cultures, we chemically synthesized carbamoyl-albicidin and tested it in combination with albicidin by *in vitro* inhibition of bacterial DNA gyrase and minimal inhibitory concentration experiments against different multiresistent pathogens. We showed that the N-terminal extended carbamoyl residue exhibits higher inhibition of bacterial gyrase and have, in contrast to the main metabolite, increased solubility properties under native solution conditions. Additional bioactivity experiments for a series of chemically synthesized albicidin analoga allowed a deep understanding of gyrase inhibition and structure–activity relationship (SAR) for the N-terminal region. The considerable structural variants at the terminal acyl residue by cinnamoyl, phenyl-propanoyl, benzoyl, and acyl residues are permitted with retention of activity but did not directly lower the minimal inhibitory concentration. In contrast to our results to carbamoyl-albicidin, it was found that a lipophilic, large acyl moiety is necessary to achieve proper antimicrobial activity that in turn lowers the solubility.[158]

In chapter 3, we focused on an obscure assembly step for the biosynthesis of albicidin that arose in context of the A domain substrate activation experiments and was briefly formulated as a hypothesis before. Substrate screening of respective A domains showed that NRPS-4 and NRPS-5 activate pABA-3OH instead of pMBA, which contradicts to the chemical structure of albicidin. Therefore, we screened the biosynthesis gene cluster for potential modifying enzymes that first introduce an *O*-methylation at the hydroxy group and incorporate afterwards a second hydroxy group or vice versa. A sequence alignment identified a SAM-dependent methyltransferase (Alb02) and β-hydroxylase (Alb08) as potential candidates for an on-line or off-line enzymatic tailoring. To prove our hypothesis, we heterologously expressed Alb02 and characterized the enzyme by a substrate activation experiment *in vitro*. Subsequent substrate-dependent kinetics and top-down

mass spectrometry conversion experiments showed that Alb02 accepts building block pABA-3OH as incorporated ligand to the respective *holo*-PCP domain. Interestingly, sequence similarity networks (SSN) for PCP homologs revealed a non-conserved and highly unique sequence insertion between helix-αI and helix-αII building an extended loop region. Only homologous that cluster to the sequence of respective PCP domains of albicidin including the loop region, belong to the structurally highly similar metabolites cystobactamide[150] and coralmycin[151]. Although secondary and tertiary structures of PCP domains are highly conserved, at the level of primary structures however PCPs show more diverse sequences in the C-terminal half between helix $\alpha 2$ and $\alpha 3$. These variations affect local shape and charge distribution of respective helices and thus PCP surfaces.[259,456,457] Previous structural studies observed that PCPs mainly mediate intramolecular domain interactions by helix-αII/III and the connecting loop.[259,260,266,267] These gave further evidence that the extended stretch is important for recruitment and interaction in-trans that finally support the on-line enzymatic tailoring of pABA-3OH. Additional control experiments using the highly related and artificially loaded *crypto*-PCP-1 domain, which lacks of addressed loop region, showed similar conversion rates in comparison to the native PCP-4 domain. These results did not support our hypothesis of a loop-mediated on-line tailoring. However, structures of *crypto*-PCPs embedded in multidomain environments revealed that Ppant arm loading strengthens domain interactions.[260,267,458] Therefore, we came to the conclusion that *in situ* methylation takes place due to a transient substrate-mediated interaction, which is controlled by correct substrate loading of respective downstream A domain. To fully prove our hypothesis and validate the significance of a substrate-controlled modification process, we examined ^{15}N mono-labeled *apo*-PCP-4 as well as *crypto*-PCP-4 in presence and absence of Alb02 by chemical shift perturbation experiments. In the case of *apo/holo*-PCPs chemical shifts, the prosthetic Ppant arm resulted in no significant alteration of the tertiary structure, which supports the 'swinging arm hypothesis' of a flexible Ppant arm that delivers substrates to adjacent domains whereas PCP domains serves as a largely rigid and chemically inert platforms. In contrast, 2D ^{15}N-HSQC spectra of solely *crypto*-PCP-4 showed for residues in the ultimate vicinity of the post-translational phosphopantheine (Ppant) attachment slight shifted crosspeaks in contrast to the *apo*-PCP-4 and *holo*-PCP-4 sample. Finally, we were able to show an explicit vanishing of distinct backbone amides in the vicinity of the highly-conserved serine residue of *crypto*-PCP-4 after co-incubation with Alb02. Although we gave significant insight into the on-line NRPS assembly of albicidin, knowledge about the non-conserved loop extension of respective PCP domains need further evidence. Nevertheless, a recent structure of a PCP/C di-domain revealed strong interactions for a loaded PCP state, which is a driving force for the product assembly. Taken into account the extended sequence between helix $\alpha 1$ and $\alpha 2$ that introduces subtle conformational changes, these features could therefore provide a decisive sensor signal for intramolecular domain interaction. Given that, dynamic equilibrium changes, which have direct impact on

intermediate on-line tailoring steps (methylation and hydroxylation), since modifying enzymes receive an extended time frame for substrate alteration.[260,275,457]

To date on-line enzymatic tailoring by a substrate-controlled *trans*-acting methyltransferase has not been described and impacts our understanding of the assembly line logic. Previous observed natural products with an *N*-methylated amide peptide backbone are found primarily in fungal *cis*-acting NRPS-machineries.[459] The A domains have been shown to display the ability to contain a downstream catalytic portion of a subdomain, most commonly that of a methyltransferase (MT) enzyme. Alternatively, methylation can be catalysed by separate enzymes within the cluster on the final, often cyclized, peptide.[254,460] Little is known about on-line enzymatic tailoring of PCP-mediated interaction in type 1 bacterial NRPS systems. In 2018, Izoré et al.[256] briefly outlined known in-trans NRPS modifications with PCPs as central interface for enzymatic substrate tailoring, and so far no further findings have been gained. An important class that requires external modification by *in situ* modification can be found for the final maturation of the glycopeptides.[461] These clinically relevant peptide antibiotics undergo an oxidative crosslinking of aromatic side chains catalysed by cytochrome P450 to achieve their final, active conformation.[274,462] It could be shown that a conserved domain present in the final module of all glycopeptide NRPS systems, labeled X domain, is responsible for the recruitment of oxygenases to the PCP-bound peptide to perform the essential side-chain crosslinking.[274,275,463] Nevertheless, *trans*-acting methyltransferases that interact via substrate-controlled pathway without inserted modules can gain further knowledge for transient interactions in NRPS systems that directly affect biosynthetic pathway prediction of natural products and will also open new opportunities for structural diversity in on-going NRPS bioengineering studies.

Future experiments on the biosynthetic assembly of albicidin should shed further light on the subsequent hydroxylation step, as well as on the biosynthesis and incorporation of the central cyanoalanine building block. The decisive question to be answered is whether hydroxylation is installed immediately after the methylation or in a post-NRPS process. A quick proof could be a follow-up experiment to our aforementioned *crypto*-PCP turn over experiment. First attempts have already been made to purify the hydroxylase Alb08 by various solubility tags, but all constructs failed. Here, more effort should be made to obtain a soluble construct and prove our hypothesis of a successive on-line enzymatic tailoring to the final substrate pMBA. Our biosynthetic understanding and the bioinformatic analyses of NRPS-2 and NRPS-2* suggest that the single standing NRPS-2* includes a previously unknown subdomain, which is responsible for the maturation of the bioactivity-based cyanoalanine building block by incorporation of asparagine and subsequent conversion due to an ATP-dependent reduction step. In our group, initial experiments seem to confirm the hypothesis, but full elucidation of the biosynthetic step is still pending. This could be proven by our established protein interaction experiments, as

extensively described in chapter 3, and would represent another groundbreaking and so far unobserved finding.

Talking about the second part of the thesis, we intensively described comprehensive alternative workflows and protocols for the analyses of venom proteomes of the *Viperidae* family, combined with venom gland transcriptomics as the basis for a well-founded database. Hereinafter, we focused on differences of intra- and inter-species analyses by established and introduced methods. Correlations to venom variations can be associated to many external factors, such as diet, regional separation of populations, sex or age, but however can also correlate to species belonging to the same genera.[327–331] The differentiation of venoms down to single individuals is a significant aspect for antivenomics to address the performance and range of clinical use of antivenoms. In previous studies, it was reported that the variability in venom composition for related viper species limits the cross-neutralization of monospecific antivenoms.[218,221,305,318]

In chapter 4, we combined detailed bottom-up venomics with intact mass profiling to comprehensively analyze two separated Nose-Horned vipers of the *Vipera ammodytes* species. Continuous phylogenetic analyses of the *V. ammodytes* complex resulted in a constant reclassification between the subspecies until today. The detailed characterization and comparison of the venom proteomes was performed to shed further light on the kinship of the four distinct *V. ammodytes* subspecies and was used as additive instrumental technique to overcome the controversial question of the taxonomic status of *V. ammodytes transcaucasiana* in connection with the phylogenetic analysis. Intact venom mass profiling was applied to obtain an overview of molecular masses for venom proteomes to highlight distinct differences or similarities as indicator for kinship. Our comparative analysis at the venomic level revealed highly related or in parts identical proteoform masses in both snake subspecies for various major toxin families. In addition, relative venom quantification for respective subspecies underscored close kinship by superimposable arrangement of curve progression and peak distribution in the chromatograms that contain same toxin families. Nevertheless, we observed that established bottom-up venomic protocols are time-consuming and expensive workflows, which prevent high-throughput analysis of large cohorts.

Therefore, we described an alternative approach to be well suited for extensive venom analysis at the population level allowing comprehensive venom quantification in chapter 5. Our robust approach was tested for a defined population of *Vipera kaznakovi* individuals in combination with venom gland transcriptomics, decomplexing bottom-up and top-down venom proteomics. Population venomics by intact mass profiling enabled a rapid venom comparison originated from multiple individuals for higher sensitivity with venom samples of low quantity. It allows us to observe intraspecies venom variation in *V. kaznakovi*, including both ontogenetic differences between juvenile and adult snakes, and to a lesser extent, sexual differences between adult males

and females. The highest difference in the venom proteome composition was observed between groups of adult and juvenile individuals, with svSPs found to exhibit the greatest variance. However, the statistical power was limited due to a relatively small subject size and therefore we infer to extend this study to a larger sample cohort, ideally covering all geographical regions of the *V. kaznakovi* distribution zone. Unfortunately, we observed in the case of higher-molecular-weight compounds (>30 kDa), which are typically strongly represented in the genus of *Viperidae*, the top-down analysis only provides a partial characterization and is still challenging.

Thus, we described a *de novo* in-source decay-driven (ISD) venomics workflow using matrix-assisted laser desorption/ionization (MALDI) as an alternative top-down approach to characterize high molecular mass venom constituents in chapter 6. We combined the approach again with decomplexing bottom-up and top-down venomics as well as venom gland transcriptomics to analyze the venom proteome of the previously unknown venom proteome of the Anatolian meadow viper, *Vipera anatolica senliki*. The transcriptome including 32 full length and 64 partially assembled toxin family transcripts were used as reference database for the identification and relative quantification of the venom proteome by decomplexing bottom-up venomics and intact mass profiling. In addition, MALDI-ISD on the venom proteome of *V. a. senliki* enabled both, identification via disulfide bond counting by direct comparison of native and reduced mass spectra as well as top-down sequencing by terminal mass fingerprints. Moreover, top-down ISD provides sufficient mass fingerprints for several proteoforms even present in a single sample fraction and has the potential to overcome persisting limits of current mass spectrometric techniques allowing a deep insight into the venom composition. Nevertheless, it is still a long way for ISD top-down sequencing to become a high-throughput workflow due to the prerequisite of highly concentrated and pure samples. In this context, an orthogonal chromatographic separation technique (e.g., SEC, IEC, or EBA) is an indispensable step and could help to overcome poor chromatographic resolution. This in turn would result in expensive costs and additional preparation time, which exclude a high-throughput screening.

Beyond venom analysis of secreted venoms, examine toxin profiles in venom glands is a further cutting-edge technology that can answer fundamental question for clinical relevance, like where venoms come from and how it is processed, stored and delivered. The standard technology for protein identification and visualization in tissues sections has been *in situ* hybridization for long time.[464,465] The artificial hybridization is carried out by applying e.g. fluorescence-labeled complementary DNA (FISH) or RNA strands (ISH) to localize specific sequences.[465] Major drawback of *in situ* hybridization techniques is a missing spatial distribution of protein mixtures across tissues as commonly present in highly functionalized and complex venom proteomes. Besides, sequences for proteins of interest must be known and therefore only allow a targeted visualization. The tissue preparation need precise and time-consuming optimization and a crucial drawback, is not allowing the differentiation of proteoforms or identification of PTMs.[466]

In our proof-of-concept study we applied mass spectrometry imaging (MSI) to unravel spatial localization of toxins families and proteoforms in the venom gland of the Egyptian cobra (*Naja haje*). The snake is one of the most medically important snakes implicated in the pathogenesis of snakebite in North Africa.[467] Previous attempts of MSI on venom apparatus showed the localization of discrete toxins peptides in the mass range between 1-6 kDa mainly in poor resolution across the tissue sections of centipedes[401,468] and cnidarians[469] or endogenous components (3-8 kDa) as an example for a snake.[470] All MALDI-imaging experiments were performed by an adapted protocol allowing intact peptide mapping of individual peptides, but have the disadvantage of a poor spatial resolution and allow no identification of toxin families in the higher molecular mass range. Therefore, we performed an enzymatic on-tissue digestion of formalin-fixed and paraffin-embedded (FFPE) tissue sections. The combination of peptide mapping across the tissue sections and a full transcriptome database enabled us to visualize toxin families and toxin proteoforms in a two-dimensional ion intensity map for the venom proteome of the Egyptian cobra. We were able to show that some of the toxin families are spatial separated within the tissue slices, whereas different proteoforms of the same family are highly localized to same tissue regions. Limitation for the protocol is that an identification of several important PTMs attached to toxin families is difficult or even not possible. Nevertheless, MSI will offer new ways in the upcoming years to provide further insights into the biology of venom production and evolution of venomous animals.

In conclusion, we applied or established state-of-the-art mass spectrometry methods for the identification or elucidation of biosynthetic steps of natural toxins from different origins. Mass spectrometry as a high-throughput analytical technique has developed to a broad scientific discipline with a myriad of specialisms and applications.[193] The recent improvements, such as higher sensitivity and resolution, allowed the introduction of mass spectrometry in omics sciences, more precisely metabolomics and proteomics.[471] Here, we identified an exceptional biosynthetic step for a promising antibacterial drug by different MS methods and gained further insights in the structural biology of a substrate-guided interaction between an in-trans enzyme and a PCP domain. Furthermore, we developed MS workflows for the identification and high-throughput screening of venom proteomes. Recently, analytical MS capable of measurements with a high level of accuracy and reproducibility combined with simple preparation and rapid screening of complex mixtures, has received interest in clinical research for its potential of biomarker discovery, development and validation. Future technological advances for enhanced sensitivity and selectivity will help to lift mass spectrometry in personalized clinical research.[472] In combination with other scientific disciplines, such as next-generation sequencing and machine deep learning, mass spectrometry will gain momentum and have a bright future to decipher the complexity of life in its entirety.

„Mit dem Wissen wächst der Zweifel.“

Johann Wolfgang von Goethe (1749-1832)

Appendix

Appendix to Chapter 2

Cloning and heterologous expression of Alb15

The *alb15* gene was amplified by PCR with Q5-Polymerase (New England Biolabs, Ipswich, USA) using the cosmid pALB540 as template. The PCR-products were digested with *Not*I and *Nco*I (Fermentas - Thermo Scientific GmbH, Schwerte, Germany), ligated into pETtrx_1c and subsequently transformed into *E. coli* BL21-Gold by heat shock. Transformants were selected on Luria Broth-agar plates containing 50 µg/mL kanamycin and reviewed by restriction analysis and DNA sequencing. Protein expression was carried out in Terrific Broth medium at 37 °C and 200 rpm for 2 h followed by 16 h at 18 °C and 180 rpm. The medium was supplemented with 50 µg/mL kanamycin, 10 mM $MgCl_2$, 0.5% glycerol, 0.05% glucose and 0.2% lactose to induce the *lac*-controlled gene expression. Afterwards, the cells were harvested by centrifugation at 4000×g for 10 min. The cell pellet was re-suspended in 50 mM Tris-HCl buffer (pH 7.8 + 300 mM NaCl, 5% (v/v) glycerol), lysed by adding lysozyme and DNAse and disrupted by an automated homogenisation system (Avestin, Mannheim, Canada). The cell debris were then separated through centrifugation at 50,000×g for 20 min. Clarified supernatant was finally purified by Ni-affinity chromatography and size exclusion chromatography using an Äkta purification system (GE Healthcare, München, Germany). Ni-affinity chromatography was carried out using a 1-mL HisTrap FF crude® column (GE Healthcare, München, Germany). The sample was loaded onto the column, washed (binding buffer: 50 mM Tris-HCl/300 Mm NaCl/20 mM imidazole, pH 7.8) and eluted by a stepwise gradient of increasing imidazole concentration (elution buffer: 50 mM Tris-HCl/300 Mm NaCl/300 mM imidazole, pH 7.8). The elution fractions were analyzed by SDS-PAGE. To remove the imidazole and the remaining impurities, the fractions were pooled, subsequently concentrated to a final volume of 2 mL and further purified by size exclusion chromatography (gel filtration buffer: 50 mM Tris-HCl/300 Mm NaCl/5% glycerol, pH 7.8) using a 120-mL HiLoad 16/600 Superdex 200 prep grade® column (GE Healthcare, München, Germany).

In-gel trypsin digestion

Coomassie Brilliant Blue G250-stained band of Alb15 was excised from the gel and subjected to in-gel reduction with 10 mM dithiothreitol in 25 mM ammonium bicarbonate, pH 8.3, for 45 min at 65 °C. Alkylation was performed with 50 mM iodoacetamide in 50 mM ammonium bicarbonate, pH 8.3 for 30 min at 25 °C, followed by in-gel trypsin digestion (overnight (12 h) at 37 °C with

66 ng sequencing-grade trypsin/µL in 25 mM ammonium bicarbonate, 10% ACN; 0.25 µg/sample. Tryptic peptides were dried in a vacuum centrifuge, redissolved in 15 µL 5% ACN + 0.1% HFO, and submitted to LC-MS/MS analysis on an Orbitrap XL hybrid mass spectrometer (Thermo, Bremen, Germany) coupled to an HPLC system (Agilent, Waldbronn, Germany). Reversed-phase separation was performed on a Grace Vydac 218MSC$_{18}$ column (2.1 x 15 mm, 5 µm) at 0.3 mL/min flow rate and developed with a gradient of 0.1% HFO in water (solution A) and in ACN (solution B), isocratically at 5% B for 2 min, followed by 5-40% B for 10 min, 40-99% B for 15 min, and 99% B for 5 min. MS experiments were performed with R=15,000 at *m/z* 400 and maximum filling time of 200 msec for survey scans. Product ion scans were recorded in the LTQ. MS/MS fragmentation of the three most intense ions was performed in the LTQ using CID (30 msec activation time); the collision energy was set to 30 %. Precursor ion isolation mass window was *m/z* 2. A window of *m/z* 3 was set for dynamic exclusion of up to 50 precursor ions with a repeat of 2 within 30 sec for the next 30 sec. MS/MS fragmentation spectra were searched against a protein database comprising all protein sequences from *E. coli* BL21-Gold including the sequence of TRX-Alb15-His6 using X!Tandem as the search engine. Peptideshaker was employed for graphical display of MS/MS spectrum matches and visual representation of the protein sequence coverage. For X!Tandem, the precursor mass tolerance was set to 10 ppm and MS/MS mass tolerance was set to ± 0.1 Da. Carbamidomethyl cysteine was selected as fixed modifications and only spectra-database hits with 100% confidence score were considered after manual inspection.

Total synthesis of carbamoyl-albicidin

All chemicals were obtained from commercial suppliers such as ABCR (Karlsruhe, Germany), Acros (Geel, Belgium), Alfar Aesar (Karlsruhe, Germany), Carl Roth GmbH Co. KG (Karlsruhe, Germany), Merck (Darmstadt, Germany), Sigma-Aldrich (Taufkirchen, Germany) and TCI (Zwijndrecht, Belgium) and, if not specified, they were used for the synthesis and analyses without further purification. Deuterated solvents used for NMR-spectroscopy (chloroform-d1 99.8% and dimethylsulfoxide-d6 99.8%) were purchased from Deutero GmbH (Kastellaun, Germany). Thin layer chromatography (TLC) was performed using TLC plates purchased from Merck (Silica gel 60, F254, coating thickness 0.2 mm). The compounds were detected by UV-light with wavelength λ = 254 nm or staining with ninhydrin solution. Flash chromatography was accomplished using silica gel from Merck and Macherey & Nagel (Düren, Germany) (particle size 0.04-0.063 mm). Automatic flash chromatography was performed on a CombiFlash Rf 200 system, with a two-channel UV detector (wavelength range λ = 200-360 nm), combined with RediSep Rf RP C$_{18}$ columns (Teledyne ISCO, Lincoln, NE, USA). ^{1}H-NMR and ^{13}C-NMR spectra were recorded on Bruker Avance 400 and Bruker Avance 500 NMR-spectrometers (Bruker, Rheinstetten, Germany). The signals of the non-deuterated solvent rests were used as standards. Chemical shifts are given in δ-units (ppm) relative to the solvent signal. High-resolution mass-spectrometry (HRMS) using

ESI-technique was performed on a LTQ Orbitrap XL apparatus produced by Thermo Scientific (Waltham, MA, USA). IR spectra were recorded on a Jasco FT-IR 4100 spectrometer (Jasco, Groß Umstadt, Germany).
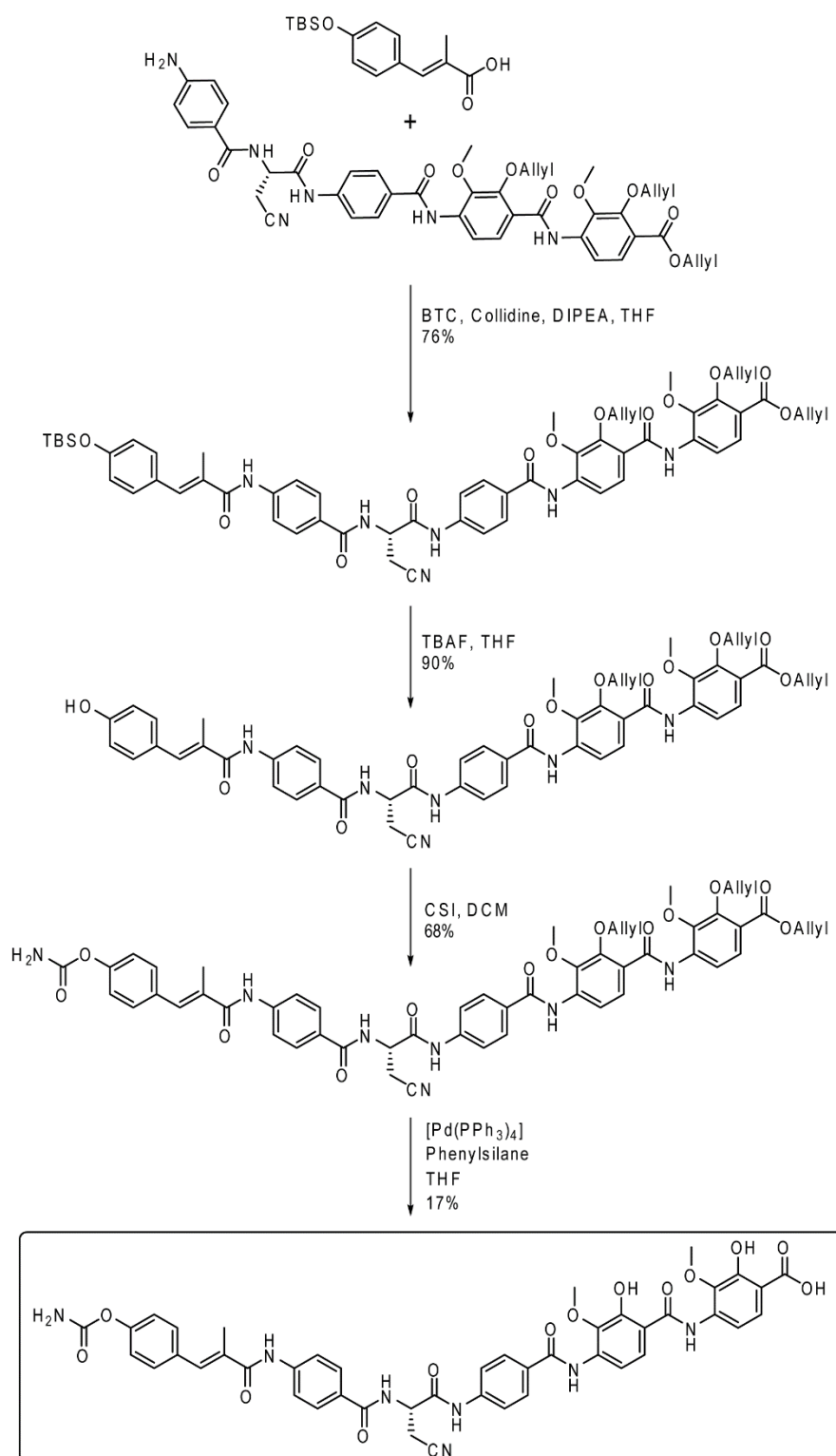
Appx Scheme 2.1. Synthesis strategy of carbamoyl-albicidin.

Structure 1: (*E*)-3-(4-Hydroxyphenyl)-2-methylacrylic acid
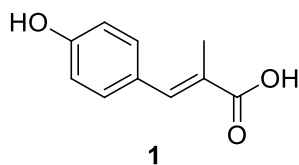
Chemical Formula: $C_{10}H_{10}O_3$
Exact Mass: 178,0630

4-Hydroxybenzaldehyde (1.00 eq, 21.18 mmol, 2.60 g) and 2-Methylmalonic acid (2.00 eq, 42.36 mmol, 5.00 g) were dissolved in piperidine (4.2 mL) and Pyridine (15 mL) and refluxed for 16 h. After cooling down to room temperature, the reaction mixture was decanted into an ice cold HCl-solution (5%, 100 mL). The precipitate was filtered and washed with water. After drying *in vacuo*, the product was obtained as a beige solid (2.6 g, 70%).

^{1}H-NMR (DMSO-d$_6$, 500 MHz): δ [ppm] 2.03 (s, 3H), 6.83 (d, *J* = 8.72 Hz, 2H), 7.35 (d, *J* = 8.52 Hz, 7.51 (s, 1H), 9.82 (s, 1H), 12.28 (bs, 1H)

^{13}C-NMR (DMSO-d$_6$, 125 MHz): δ [ppm] 14.4, 155.9, 125.5, 126.9, 132.1, 138.4, 158.4, 170.2

HRMS (ESI): [M-H]$^-$ calculated: 177.0557

found: 177.0550

Structure 2: (*E*)-*tert*-butyldimethylsilyl-3-(4-(((*tert*-butyldimethylsilyl)oxy)phenyl)-2-methylacrylate
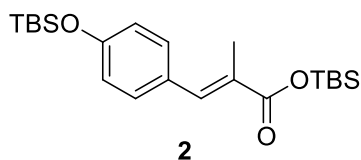
Chemical Formula: $C_{22}H_{38}O_3Si_2$
Exact Mass: 406,2359

Compound **1** (533 mg, 2.99 mmol, 1.00 eq) was dissolved in DMF and imidazole (584 mg, 7.49 mmol, 2.50 eq) was added. TBS-Cl (1128 mg, 7.49 mmol, 2.50 eq) was added at 0 °C and the reaction mixture was let stir for 3 h at room temperature. The reaction mixture was poured onto ice-cold water and the aqueous layer was extracted 3 x with EE. The combined organic layers were washed with brine and dried over Na$_2$SO$_4$. After removing the solvent *in vacuo*, the crude product was chromatographically purified to give the pure product as a yellow oil (650 mg, 53%).

^{1}H-NMR (DMSO-d$_6$, 400 MHz): δ [ppm] 0.21 (s, 6H), 0.28 (s, 6H), 0.95 (s, 9H), 0.96 (s, 9H), 2.03 (s, 3H), 6.92 (d, *J* = 8.60 Hz, 2H), 7.41 (d, *J* = 8.60 Hz, 2H), 7.58 (s, 1H)

^{13}C-NMR (DMSO-d$_6$, 100 MHz): δ [ppm] -3.2, 14.0, 25.5, 25.8, 120.0, 124.6, 126.9, 131.7, 163.2, 170.9

HRMS (ESI): [M+H]$^+$ calculated: 407.2432

found: 407.2429

Structure 3: (*E*)-3-(4-((*tert*-butyldimethylsilyl)oxy)phenyl)-2-methylacrylic acid
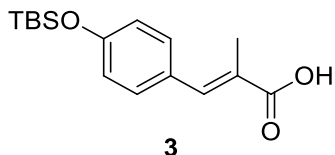
Chemical Formula: $C_{16}H_{24}O_3Si$
Exact Mass: 292,1495

Compound **2** (550 mg, 1.355 mmol) was dissolved in THF/MeOH (3:1) and one part of 10% K_2CO_3-sol. was added. After stirring for 1 h at room temperature, the organic solvents were evaporated. The pH was adjusted to 4 – 5 with citric acid. The aqueous layer was extracted 3x with EE. After drying the combined organic layers over Na_2SO_4, the solvent was evaporated to give the product as a pale yellow solid (309 mg, 78%).

^{1}H-NMR (CDCl$_3$, 500 MHz): δ [ppm] 0.24 (s, 6H), 1.00 (s, 9H), 2.16 (s, 3H), 6.88 (d, *J* = 8.72 Hz, 2H), 7.37 (d, *J* = 8.52 Hz, 2H), 7.77 (s, 1H)

^{13}C-NMR (DMSO-d$_6$, 125 MHz): δ [ppm] 13.6, 18.0, 25.4, 119.9, 124.9, 128.5, 131.4, 140.6, 156.2, 173.2

HRMS (ESI): [M+H]$^+$ calculated: 293.1567

found: 293.1565

Structure 4: (*S,E*)-allyl-2-(allyloxy)-4-(2-(allyloxy)-4-(4-(2-(4-(3-(4-((*tert*-butyldimethylsilyl)oxy)-phenyl)-2-methylacrylamido)benzamido)-3-cyanopropanamido)benzamido)-3-methoxybenzamido)-3-methoxybenzoate
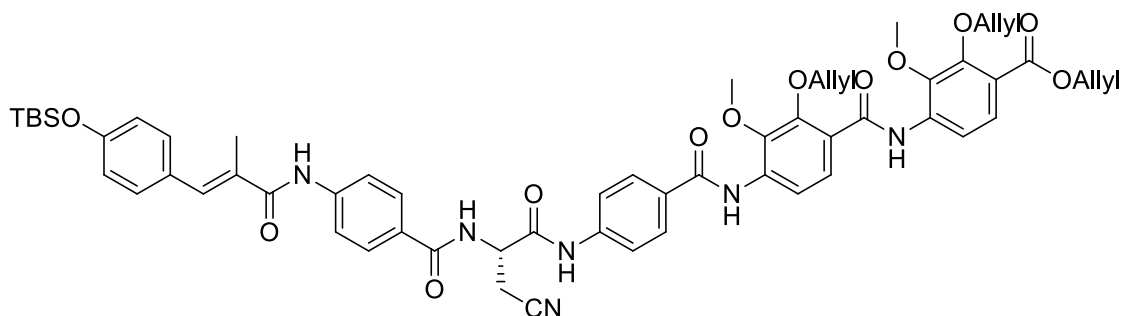
4

Chemical Formula: $C_{59}H_{64}N_6O_{12}Si$
Exact Mass: 1076,4351

Compound **3** (151 mg, 0.517 mmol, 2.50 eq) was dissolved in THF and BTC (51 mg, 0.172 mmol, 0.83 eq) dissolved in THF was added, followed by collidine (167 mg, 182 µL, 1.378 mmol, 8.00 eq). After stirring at room temperature for 30 min, the amine (166 mg, 0.207 mmol, 1.00 eq) and DIPEA (267 mg, 365 µL, 2.068 mmol, 10.00 eq) dissolved in THF were added. After stirring the reaction mixture for 16 h at room temperature, it was diluted with EE and subsequently washed with 1 N HCl, sat. NaHCO₃ solution and brine. After drying over Na₂SO₄, the solvent was evaporated. The crude product was purified chromatographically to obtain the pure product as a pale yellow solid (170 mg, 76%).

^{1}H-NMR (DMSO-d₆, 400 MHz): δ [ppm] 0.22 (s, 6H), 0.96 (s, 9H), 2.12 (s, 3H), 3.12 (m, 2H), 3.92 (s, 3H), 3.93 (s, 3H), 4.53 (m, 2H), 4.79 (m, 4H), 4.99 (m, 1H), 5.27 (m, 3H), 5.40 (s, 3H), 6.07 (m, 3H), 6.93 (d, *J* = 8.60 Hz, 2H), 7.30 (s, 1H), 7.42 (d, *J* = 8.60 Hz, 2H), 7.57 (d, *J* = 8.87 Hz, 1H), 7.80 (m, 3H), 7.85 (d, *J* = 8.87 Hz, 2H), 7.93, (m, 3H), 7.99 (d, *J* = 9.13 Hz, 2H), 8.33 (d, *J* = 9.13 Hz, 1H), 9.03 (d, *J* = 7.79 Hz, 1H), 9.69 (s, 1H), 10.14 (s, 1H), 10.59 (s, 1H), 10.66 (s, 1H)

HRMS (ESI):   [M+H]⁺   calculated: 1077.4424

found: 1077.4407

Structure 5: (*S,E*)-allyl-2-(allyloxy)-4-(2-(allyloxy)-4-(4-(3-cyano-2-(4-(3-(4-hydroxyphenyl)-2-methylacrylamido)benzamido)propanamido)benzamido)-3-methoxybenzamido)-3-methoxybenzoate

5
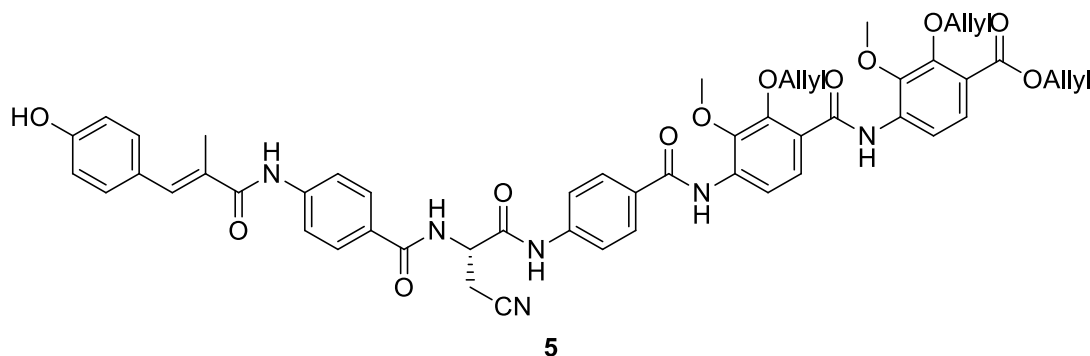
Chemical Formula: $C_{53}H_{50}N_6O_{12}$

Exact Mass: 962,3487

Compound **4** (150 mg, 0.139 mmol, 1.00 eq) was dissolved in THF and a 1 M solution of TBAF (697 µL, 0.697 mmol, 5.00 eq) in THF was slowly added. After stirring for 5 h at room temperature, the reaction mixture was diluted with EE and washed 3 x with 1 N HCl. After drying over Na_2SO_4, the solvent was evaporated to give the pure product as a yellow solid (121 mg, 90%).

^{1}H-NMR (DMSO-d_6, 400 MHz): δ [ppm] 2.12 (s, 3H), 3.13 (m, 2H), 3.92 (s, 3H), 3.93 (s, 3H), 4.54 (m, 2H), 4.79 (m, 4H), 4.98 (m, 1H), 5.27 (m, 3H), 5.40 (s, 3H), 6.07 (m, 3H), 6.84 (d, *J* = 8.60 Hz, 2H), 7.27 (s, 1H), 7.35 (d, *J* = 8.60 Hz, 2H), 7.57 (d, *J* = 8.86 Hz, 1H), 7.80 (m, 3H), 7.85 (d, *J* = 8.87 Hz, 2H), 7.92, (m, 3H), 7.99 (d, *J* = 8.60 Hz, 2H), 8.33 (d, *J* = 8.60 Hz, 1H), 9.02 (d, *J* = 6.98 Hz, 1H), 9.69 (s, 1H), 9.77 (s, 1H), 10.11 (s, 1H), 10.59 (s, 1H), 10.66 (s, 1H)

HRMS (ESI):   [M+H]$^+$ calculated: 963.3559

                                    found: 963.3552

Structure 6: (*S,E*)-allyl 2-(allyloxy)-4-(2-(allyloxy)-4-(4-(2-(4-(3-(4-(carbamoyloxy)phenyl)-2-methylacrylamido)benzamido)-3-cyanopropanamido)benzamido)-3-methoxybenzamido)-3-methoxybenzoate
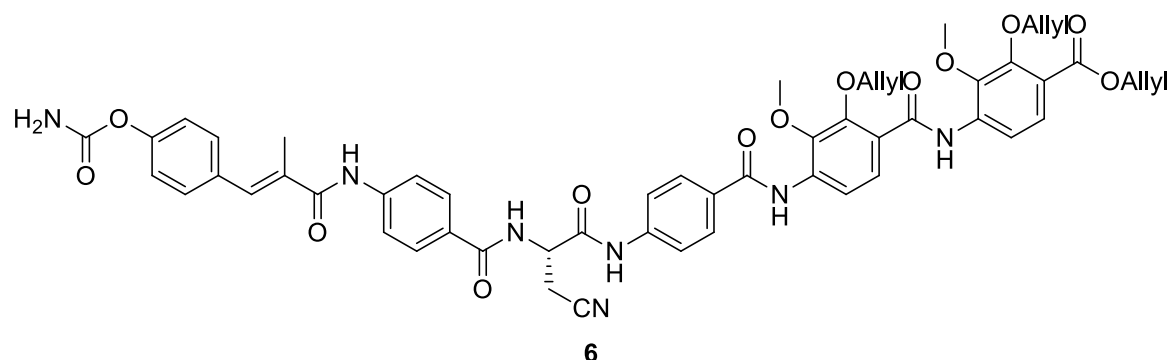
6
Chemical Formula: $C_{54}H_{51}N_7O_{13}$
Exact Mass: 1005,3545

Compound **5** (134 mg, 0.139 mmol, 1.00 eq) was dissolved in DCM and stirred at room temperature. A solution of CSI (20 mg, 12 μL, 0.139 mmol, 1.00 eq) in DCM was added and the reaction mixture was stirred for 2 h. The solvent was removed under reduced pressure and ice-cold water was added to the residue. After stirring overnight at 4 °C, the aqueous layer was extracted 3x with EE. The combined organic layers were washed with brine and subsequently dried over Na_2SO_4. After removing the solvent under reduced pressure, the product was obtained as a yellow solid (94 mg, 68%).

^{1}H-NMR (DMSO-d_6, 400 MHz): δ [ppm] 2.11 (s, 3H), 3.13 (m, 2H), 3.92 (s, 3H), 3.93 (s, 3H), 4.55 (m, 2H), 4.79 (m, 4H), 4.98 (m, 1H), 5.27 (m, 3H), 5.40 (s, 3H), 6.07 (m, 3H), 6.84 (d, *J* = 8.60 Hz, 2H), 7.27 (s, 1H), 7.35 (d, *J* = 8.60 Hz, 2H), 7.57 (d, *J* = 8.86 Hz, 1H), 7.80 (m, 3H), 7.85 (d, *J* = 8.87 Hz, 2H), 7.92, (m, 3H), 7.99 (d, *J* = 8.60 Hz, 2H), 8.33 (d, *J* = 8.60 Hz, 1H), 9.02 (d, *J* = 6.98 Hz, 1H), 9.69 (s, 1H), 9.77 (s, 1H), 10.12 (s, 1H), 10.59 (s, 1H), 10.66 (s, 1H)

HRMS (ESI): [M+H]$^+$ calculated: 1006.3626

found: 1006.3618

Structure 7: (*S*,*E*)-4-(4-(4-(2-(4-(3-(4-(carbamoyloxy)phenyl)-2-methylacrylamido)benzamido)-3-cyanopropanamido)benzamido)-2-hydroxy-3-methoxybenzamido)-2-hydroxy-3-methoxybenzoic acid
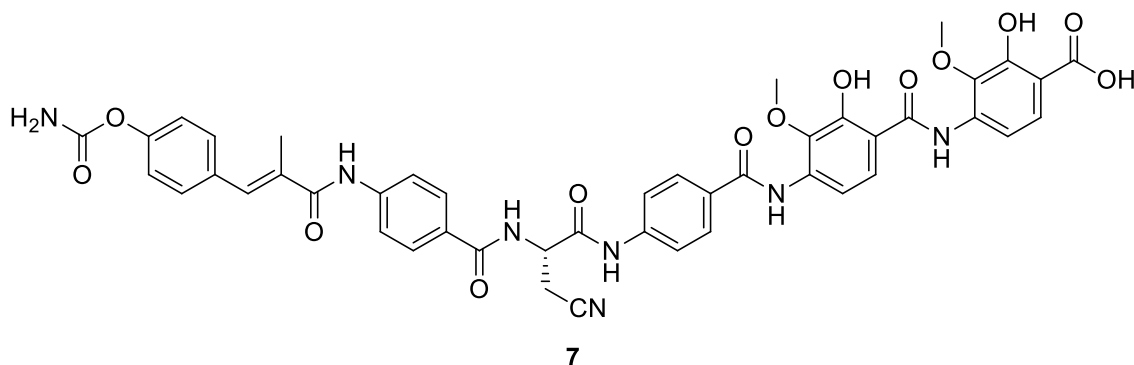
7
Chemical Formula: $C_{45}H_{39}N_7O_{13}$
Exact Mass: 885,2606

Compound **6** (83 mg, 8.259 × 10^{-5} mol, 1.00 eq) was dissolved in THF and phenylsilane (71 mg, 6.607 × 10^{-4} mol, 8.00 eq) was added. Subsequently, Tetrakistriphenylphosphinpalladium (48 mg, 4.129 × 10^{-5} mol, 0.50 eq) was added and the reaction mixture was stirred at room temperature for 16 h. A few drops of acetic acid were added and all volatile components were removed *in vacuo*. The residue was dissolved in MeOH and filtered through a PTFE membrane filter. After removing the MeOH, the crude product was purified via HPLC chromatography. The product was obtained as a pale yellow solid (12 mg, 17%).

^{1}H-NMR (DMSO-d$_6$, 400 MHz): δ [ppm] 2.13 (s, 3H), 3.12 (m, 2H), 3.77 (s, 3H), 3.91 (s, 3H), 4.99 (m, 1H), 6.63 (d, *J* = 8.6 Hz, 1H), 7.12 (d, *J* = 8.3 Hz, 1H), 7.18 (d, *J* = 8.6 Hz, 2H), 7.35 (s, 1H), 7.49 (d, *J* = 8.9 Hz, 1H)), 7.58 (m, 3H), 7.80 (m, 3H), 7.87 (m, 2H), 7.94 (d, *J* = 8.9 Hz, 2H), 7.99 (d, *J* = 8.9 Hz, 2H), 8.05 (d, *J* = 8.9 Hz, 1H), 9.04 (d, *J* = 7.9 Hz, 1H), 9.72 (s, 1H), 10.21 (s, 1H), 10.33 (s, 1H), 10.59 (s, 1H), 11.18 (s, 1H), 11.55 (s, 1H),

HRMS (ESI):　[M-H]$^-$　　calculated:　　884.2522
　　　　　　　　　　　　　　found:　　　　884.2554

Appx Table 2.1. Primers used to prepare the *alb15* deletion knockout mutant.

Name	Sequence	Position on the genome of *X. albilineans* strain GPE PC73
AalbXV	CCTGATCATCCAGACAGTGATGCGTACG	1742951 to 1742932
BalbXV	TACCGACCAAGGTTGACCTTCACTCATGCCGTCCAGTACTGCGC	1742093 to 1742112
CalbXV	TGTGAAGGTCAACCTTGGTCGGTATAACACCAGCTTCAACGTCA	1740989 to 1740970
DalbXV	CCTGATCACTGGAGTTTCTGGCTCATC	1740023 to 1740042
albXV cribIA	GTACAAGCGTTCAATCGGCG	1742513 to 1742494
albXV cribIC	CTGGGCTGTGGCATCACCAT	1742982 to 1742963
albXV cribIB	CACGATCAGCCGCTAGGAAC	1740821 to 1740840
albXV cribID	CTGCTACCACTACCAACAGG	1742072 to 1742053

Appx Table 2.2. Minimal inhibitory concentration (MIC) of carbamoyl-albicidin, albicidin and apramycin against *Bacillus subtilis*, *Mycobacterium phlei, Salmonella typhimurium* and *E. coli* DH5▯.

MICs [ng/µL]:	*Bacillus subtilis*	*Mycobacterium phlei*	*Salmonella typhimurium*	*E. coli* K-12 *BW25113*
Albicidin	< 0.2	< 0.2	~ 6.3	~ 0.06
Carbamoyl-albicidin	< 0.2	< 0.2	~ 3.1	~ 0.1
Apramycin	~ 25	~ 1.6	~ 6.3	~ 4.0

His14 Asp15

His114 His118 Asp137

Asp338

Arg418

His492 Arg498

Thr529 Ser530

TobZ : MRVLGLNCWPRDFHDASAALLVDGRIAAFAEE-ERLTRKKHGYNTAPV--QA------------------AAFC-----LAQAGL--TVDDLDA-----VAFGWDLPA--------MYRERLGGWPHSDSEALDILLPRDVFPRRTDPPLHFVQ : 113
Alb15 : -MKRTYIGLANSFHDSAIAIVGDDGQVRFAEATERYLQYKRSIGVAPDVFQRAIKLVHEYGDPGAELVVATSWSGQTPELMREGLGKTAQAVDQYRSAFGDLPWHVNKQFVAQSFFYRSQLAMVEH-PGHLIEYDLSHMAEPAFKPESYRHYE : 151

TobZ : HHLAHAASAYYFSGEDRGAVLIVDGQGEEECVTLAHAEGGKITVLDT-VPGAW-SLGFFYEHVSEYTGLGGDN--PGKLMGLAAHGTTVDETLSAF--AFDSDGYRLNLIDPQARDPEDWDEYSVTERAWFAHLERIYRLPPNEFVRRYDPAK : 260
Alb15 : HHLTHAVAGCYTSPFEEAVCAVLDGMGEKNALACYHYQQGKLTPTHQSETSSWASLGFFYGMICEVCGFGTLSGEEWKVMGLAAYGQHDRQLYELLRQMLRVDGLTLRFAP----------------AAQFSQLQRTL------Y--------- : 274

TobZ : GRVVRDTRRDPYEYRDLAATAQAALERAVFGLADSVLARTGERTLFVAGGVGLNATMNGKLLTRSTVDKMFVPPVASDIGVSLGAAAAVAVELGDRIAP-MG---DTAAWGPEFSPDQVRAALDRTGLA---YREPANLEREVAALIASGKVVG : 407
Alb15 : -AMRRCKGQPTIELANLAYAGQQVFCDVLFEFLHNLHALGLSDHLVLGGGCALNSSANGRVLAETPFRHLHVFAAPGDDGNAVGAALWAHAEDHPEQTPPAAREQSPYLGSSMSAETLH-NVERFGALSKFTRCLDDAAQRAARLLTEGKIVA : 425

TobZ : WAQGRGEVGPRALGQRSLIGSAHSPTMRDHINLRVKDREWWRPFAPSMLRSVSDQVLEVDADFPYMIMTTKVRAAYAERLPSVVHEDWSTRPQTVTEASNPRYHRMLTELGDLVGDPVCLNTSFNDRGEPIVSSPADALLTFSRLPIDALAVG : 560
Alb15 : WVQGRAEFGPRALGNRSILADPRSPAIKDIINARVKFREEFRPFAPSILHEHGAEYFELYQESPYMERTLKFRAEATRKVPGVVHHDGTGRLQTVKQHWNPRYHALIKEFYRLTGIPLVLNTSFNVMGKPIAHSVEDALSIFFTSGLDAMFID : 578

TobZ : PYLVTKDLRH : 570
Alb15 : DVLIEK---- : 584

Beta Strand Alpha Helix
Turn 3/10-Helix
Bend Beta Bridge

Appx Figure 2.1. Sequence comparison of Alb15 and TobZ. Amino acids involved in the complexation of Fe are marked with black stars; amino acids involved in the binding of carbamoyl-AMP are marked with red diamonds.
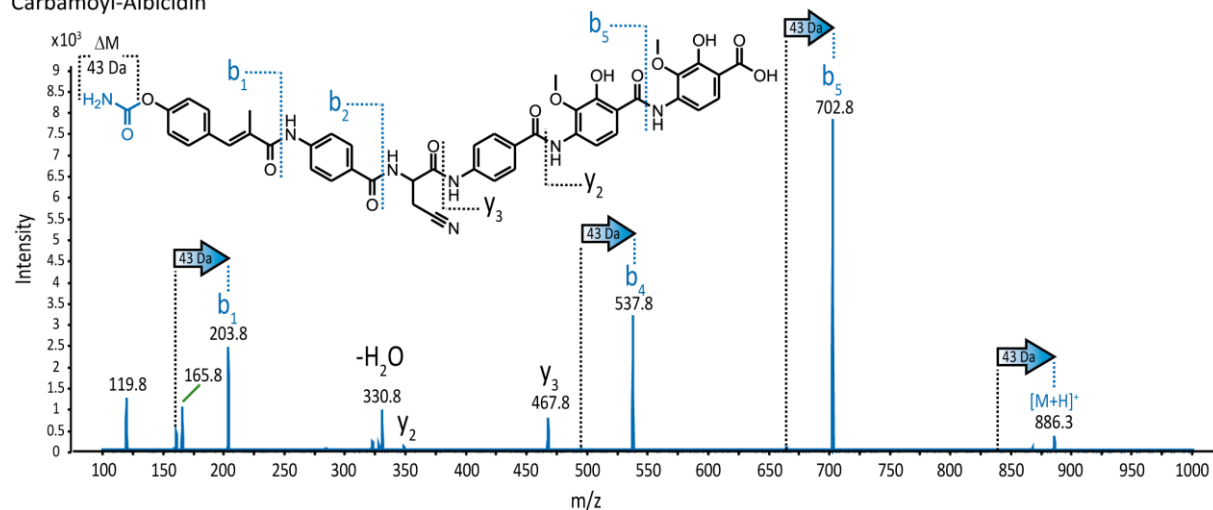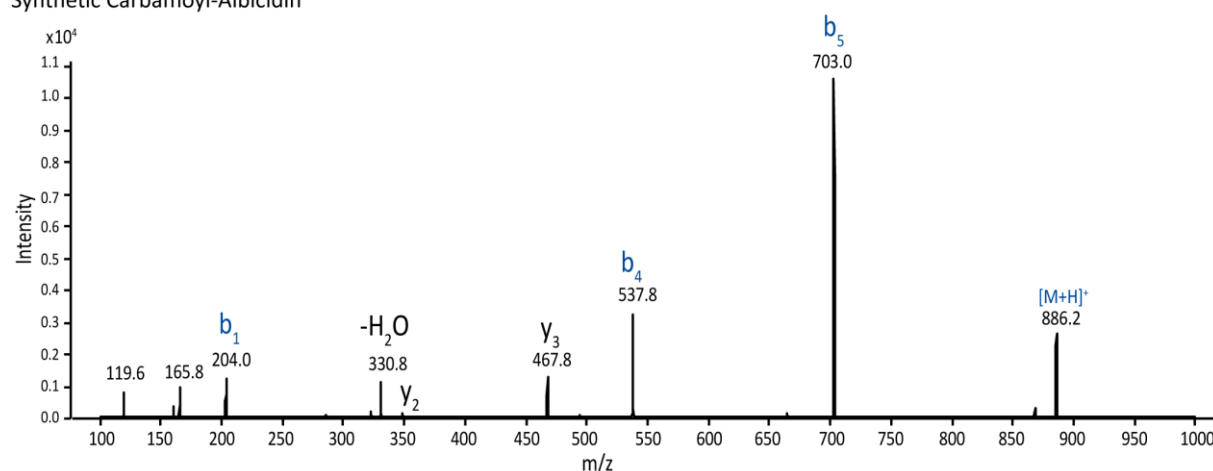
Appx Figure 2.2. MS/MS comparison of natural and synthetic carbamoyl-albicidin. The observed b-ion and y-ion fragments and the mass shift through carbamoylation are indicated in the spectra and structure.
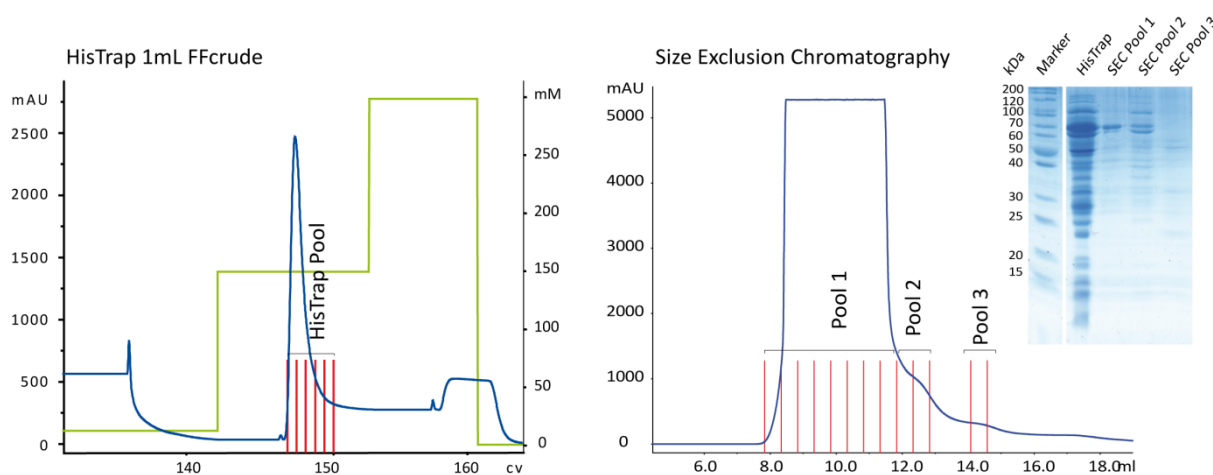
Appx Figure 2.3. Protein purification of Alb15 fusion protein. Ni-affinity chromatogram (HisTrap), size exclusion and SDS-PAGE of the pooled fractions are shown. UV Elution profiles at 280 nm of His-trap and size exclusion chromatography are shown. The red lines mark the pooled fractions. Furthermore SDS-PAGE of the pooled recombinant protein is shown. The expected protein mass is 79.6 kDa.

		PI	Accession	Description	Chr	Coverage	#Peptides	#Spectra	MS2 Quant.	MW	Confidence	
1	★		787773114	Alb15-TRX-His6 [Xanthomonas albilineans]		45.96	47	136	1.29E04	79.58	100	
2	☆		253325999	chaperone protein DnaK [Escherichia coli 'BL21-Gold(DE3)pLysS AG']		19.28	8	16	7.17E03	69.07	100	
3	☆		253325104	ribosomal protein S1 [Escherichia coli 'BL21-Gold(DE3)pLysS AG']		13.11	5	8	3.79E03	61.12	100	
4	☆		253325718	Cytosine deaminase [Escherichia coli 'BL21-Gold(DE3)pLysS AG']		2.34	1	2	3.07E03	47.56	100	
5	☆		253323093	formate acetyltransferase [Escherichia coli 'BL21-Gold(DE3)pLysS AG']		1.44	1	1	1.15E03	85.88	100	
6	☆		253322545	glycoside hydrolase family 31 [Escherichia coli 'BL21-Gold(DE3)pLysS AG']		5.18	2	2	0.00E00	88.04	67	
7	☆		253323429	sulfite reductase (NADPH) flavoprotein, alpha chain [Escherichia coli 'BL21-Gold(...		1.84	1	1	0.00E00	66.22	67	
8	☆		253324234	transcriptional regulator, RpiR family [Escherichia coli 'BL21-Gold(DE3)pLysS AG']		6.23	1	1	0.00E00	31.96	67	
9	☆		253324306	protein of unknown function DUF444 [Escherichia coli 'BL21-Gold(DE3)pLysS AG']		5.39	1	1	0.00E00	49.42	67	
10	☆		253324396	3-dehydroquinate dehydratase, type I [Escherichia coli 'BL21-Gold(DE3)pLysS AG']		3.97	1	1	3.91E03	27.46	67	
11	☆		253325012	Glucose-1-phosphatase [Escherichia coli 'BL21-Gold(DE3)pLysS AG']		6.54	1	1	0.00E00	45.66	67	
12	☆		253326464	aspartate kinase [Escherichia coli 'BL21-Gold(DE3)pLysS AG']		1.48	1	1	0.00E00	88.89	15	

Peptides (44/47 - 37 confident, 7 doubtful)

		PI	Sequence	Start		#Spectra		Confidence	
1	★		NH2-TYIGLANSFHDSAIAIVGDDGQVR-COOH	127		22		100	
2	☆		NH2-EFLDANLAGSGSGSENLYFQGAMER-COOH	102		11		100	
3	☆		NH2-LNIDQNPGTAPK-COOH	59		8		100	
4	☆		NH2-LVHEYGDPGAELVVATSWSGQTPELMR-COOH	178		8		100	
5	☆		NH2-MIAPILDEIADEYQGK-COOH	38		6		100	
6	☆		NH2-GQLKEFLDANLAGSGSGSENLYFQGAMER-COOH	98		5		100	
7	☆		NH2-GLAAYGQHDR-COOH	355		4		100	
8	☆		NH2-SIGVAPDVFQR-COOH	164		3		100	
9	☆		NH2-VMGLAAYGQHDR-COOH	353		3		100	
10	☆		NH2-ALEHHHHHH-COOH	710		2		100	
11	☆		NH2-AIKDIINAR-COOH	574		2		100	
12	☆		NH2-DDGNAVGAALWAHAEDHPEQTPPAAR-COOH	474		2		100	
13	☆		NH2-EQSPYLGSSMSAETLHNVER-COOH	500		2		100	
14	☆		NH2-EAPAAGESQLQR-COOH						

Appx Figure 2.4. Protein verification of in-gel tryptic digest and MS/MS spectra database comparison. MS/MS spectra were searched against an *E. coli* BL21-Gold protein database including the sequence of the TRX-Alb15-His6 construct.

Appx Figure 2.5. Agar inhibition zone assay of *E. coli* DH5α. Carbamoyl-albicidin and albicidin were applied on top-agar plates with *E. coli* DH5α and incubated overnight. Inhibition zones were measured with a ruler and plotted against the amount of compound/spot. The experiment was performed in duplicates.

Appendix to Chapter 3

Supplemental experimental information

Cloning, recombinant expression and purification of Alb02, PCP-1 and PCP-4 domains

The genes were amplified by PCR with Q5-Polymerase (New England Biolabs) using the templates cosmid pALB540 or pALB571 and respective oligonucleotides (**Appendix Table 3.1**). The PCR-products were digested with NotI and NcoI (Fermentas - Thermo Scientific GmbH), ligated into pETtrx_1c and subsequently transformed into *E. coli* BL21-Gold (DE3) strain (Merck Millipore, Darmstadt, Germany) with the T7 promoter expression system by heat shock (42 °C, 2 min). Transformants were selected on Luria Broth (LB) agar plates containing kanamycin (50 μg/mL) and checked by restriction analysis and DNA sequencing. Protein expression was carried out in Terrific Broth (TB) medium at 37 °C and 200 rpm for 2 h followed by 16 h at 18 °C and 180 rpm using autoinduction solution (0.5% glycerol, 0.05% glucose and 0.2% lactose) to induce the lac-controlled gene expression. The medium was supplemented with 10 mM $MgCl_2$ and 50 μg/mL kanamycin as selection marker. Afterwards, the cells were harvested by centrifugation at 8,000×g for 30 min and cell pellets resuspended in 50 mM Tris-HCl buffer (pH 7.8, 500 mM NaCl, 5% (v/v) glycerol, 10 mM imidazole, 20 mM $MgCl_2$). Cell lysis was performed by disruption (15 kPsi) with a high-pressure homogenisation system (Constant Systems Limited). Then, cell debris were separated through centrifugation at 50,000×g for 30 min and clarified supernatants purified by Ni-affinity chromatography. Ni-affinity chromatography was carried out using a 1-mL HisTrap FF crude® column (GE Healthcare). The sample was loaded onto the column, washed (wash buffer: 50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, pH 7.8) and eluted by a stepwise gradient of increasing imidazole concentration (elution buffer: 50 mM Tris-HCl, 300 mM NaCl, 500 mM imidazole, pH 7.8). Afterwards, concentrated fusion proteins were cleaved by an overnight digestion (16 h), with tobacco etch virus (TEV) at 4 °C, to remove the thioredoxin-solubility (trx) tag. Subsequently, a second Ni-affinity purification step with the aforementioned conditions was performed to separate His_6-tagged proteins. Size exclusion chromatography (gel filtration buffer: 50 mM Tris-HCl, 300 mM NaCl, pH 7.8) was performed to remove the remaining imidazole and impurities using a 120-mL HiLoad 16/600 Superdex 200 prep grade® column (GE Healthcare) (**Appendix Figure 3.7**). The identity of all fusion proteins were verified by SDS-PAGE (**Appendix Figure 3.8**) and peptide spectrum sequencing (**Appendix Figure 3.9**).

Synthesis of N-acetylcysteamine (SNAC) derivatives

SNAC (0.587 mmol, 1 eq) was added to a solution of the *p*-aminobenzoic acid (0.617 mmol, 1.05 eq), PyBOP (0.617 mmol, 1.05 eq) and DIPEA (1.23 mmol, 2.10 eq) in DMF (1.2 mL). The reaction was stirred for overnight (14 h) at room temperature and reaction control was carried out by analytical thin layer chromatography (UV light: λ = 254 nm) using aluminium-backed plates coated with Macherey-Nagel silica gel (60, F254). In the following, the reaction mixture was diluted with EtOAc (50 mL) and washed with brine (3 x 50 mL), dried over Na_2SO_4 and concentrated *in vacuo*. The crude material was purified via a RP-Flash column chromatography (20-50 % MeOH in H_2O in 12 min) with a CombiFlash®R_f (Teledyne Isco) system and a C-18 reversed phase cartridge (40 g, Grace). The product was obtained as a colourless solid after lyophilisation (yields: 58-69%). ^{1}H and ^{13}C-spectra were recorded at 298 K using Bruker Avance-II 400 MHz or Bruker Avance-III 500 MHz (Bruker, Karlsruhe, Germany). The chemical shifts are reported in ppm using the residual solvent peak as an internal reference (DMSO-d_6 or CDCl$_3$). Multiplicity (bs = broad singlet, s = singlet, d = doublet, dd = doublet of doublet, t = triplet, q = quartet, m = multiplet) and coupling constants (*J* = Hz) are quoted where possible. HPLC-HRMS spectra were recorded on a QTrap LTQ XL (Thermo Fisher Scientific) with an Agilent 1200 Series HPLC-System (Agilent Technologies) with a C18 column (50 x 2 mm, particle size 3 µm).

<u>*S-(2-acetamidoethyl) 4-aminobenzothioate*</u>

^{1}H-NMR (500 MHz, CDCl$_3$): δ [ppm] = 7.81 (d, *J* = 8.8 Hz, 2H), 6.64 (d, *J* = 8.7 Hz, 2H), 5.98 (bs, 1 H), 4.15 (bs, 2H), 3.52 (q, *J* = 6.3 Hz, 2H), 3.19 (t, *J* = 6.5 Hz, 2H), 1.96 (s, 3H).

^{13}C-NMR (126 MHz, CDCl$_3$): δ [ppm] = 170.3, 151.7, 129.8, 113.8, 40.1, 28.2, 23.2.

HRMS (ESI): *m/z* calc. for $C_{11}H_{15}N_2O_2S$ (M+H$^+$) 239.0849, detected 239.0850.
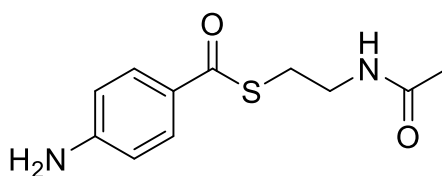
Chemical Formula: $C_{11}H_{14}N_2O_2S$
Exact Mass: 238.07760

S-(2-acetamidoethyl) 4-amino-2-hydroxybenzothioate

^{1}H-NMR (400 MHz, DMSO-d$_6$): δ [ppm] = 11.1 (s, 1 H), 8.10 (t, J = 5.6 Hz, 1H), 7.49 (d, J = 8.8 Hz,1H), 6.40 (s, 2H), 6.16 (dd, J = 2.1 Hz, J = 8.9 Hz, 1H), 5.96 (d, J = 2.2 Hz, 1H), 3.23 (q, J = 6.5 Hz, 2H), 3.01 (t, J = 6.9 Hz, 2H), 1.80 (s, 3H).

^{13}C-NMR (101 MHz, DMSO-d$_6$): δ [ppm] = 191.4, 169.3, 161.3, 156.4, 130.7, 109.0, 107.0, 98.1, 27.4, 22.5.

HRMS (ESI): *m/z* calc. for C$_{11}$H$_{15}$N$_2$O$_3$S (M+H$^+$) 255.0798, found 255.0798.
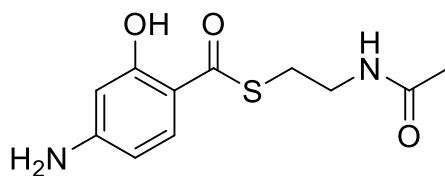
Chemical Formula: C$_{11}$H$_{14}$N$_2$O$_3$S
Exact Mass: 254.07251

S-(2-acetamidoethyl) 4-amino-3-hydroxybenzothioate

^{1}H-NMR (400 MHz, DMSO-d$_6$): δ [ppm] = 9.59 (s, 1H), 8.08 (t, J = 5.5 Hz, 1H), 7.27 (dd, J = 8.32 Hz, J = 2.1 Hz, 1H), 6.60 (d, J = 8.2 Hz, 1H), 5.61 (s, 2H), 3.21 (q, J = 6.4 Hz, 2H), 2.99 (t, J = 6.9 Hz, 2H), 1.80 (s, 3H).

^{13}C-NMR (101 MHz, DMSO-d$_6$): δ [ppm] = 187.6, 169.2, 143.6, 142.9, 123.7, 121.0, 112.2, 112.1, 27.6, 22.5.

HRMS (ESI): *m/z* calc. for C$_{11}$H$_{15}$N$_2$O$_3$S (M+H$^+$) 255.0798, found 255.0791.
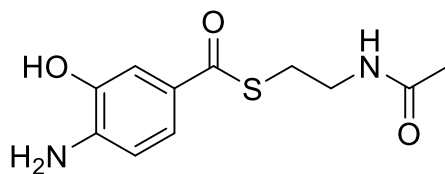
Chemical Formula: C$_{11}$H$_{14}$N$_2$O$_3$S
Exact Mass: 254.07251

Synthesis of coenzyme A (CoA) derivatives

CoA trilithium salt (3.82 μmol, 3 mg, 1.0 eq) was added to a previously activated solution of pABA (11.5 μmol, 1.6 mg, 3.0 eq), PyBOP (11.5 μmol, 6.0 mg, 3.0 eq) and DIPEA (38.2 μmol, 5.0 mg, 10 eq) in DMF (200 μL). The reaction mixture, controlled by analytical TLC (UV light: λ = 254 nm), was stirred for overnight (14 h) at room temperature and neutralized with 1 M HCl (38 μL). After lyophilisation, the crude product was purified via RP-HPLC (15-30 % MeOH in H$_2$O in 15 min) on a 1260 Infinity (Agilent Technologies) system with a C-18 reversed phase column

(21.2 x 250 mm, particle size 10 μm, Agilent Technologies). The product was obtained as a colourless solid after lyophilisation (yields: 38-41%). ^{1}H and ^{13}C-spectra were recorded at 298 K using Bruker Avance-II 400 MHz or Bruker Avance-III 500 MHz (Bruker, Karlsruhe, Germany). The chemical shifts are reported in ppm using the residual solvent peak as an internal reference (DMSO-d$_6$ or CDCl$_3$). Multiplicity (bs = broad singlet, s = singlet, d = doublet, dd = doublet of doublet, t = triplet, q = quartet, m = multiplet) and coupling constants (J = Hz) are quoted where possible. HPLC-HRMS spectra were recorded on a QTrap LTQ XL (Thermo Fisher Scientific) with an Agilent 1200 Series HPLC-System (Agilent Technologies) with a C18 column (50 x 2 mm, particle size 3 μm).

4-aminobenzoic acid-CoA ester

HRMS (ESI): *m/z* calc. for C$_{28}$H$_{42}$N$_8$O$_{17}$P$_3$S (M+H$^+$) 887.1596, detected 887.1616.
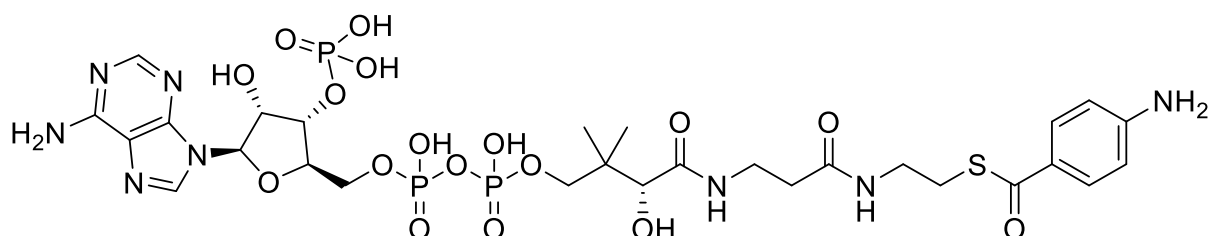
Chemical Formula: C$_{28}$H$_{41}$N$_8$O$_{17}$P$_3$S
Exact Mass: 886.15232

3-hydroxy-4-aminobenzoic acid-CoA ester

HRMS (ESI): *m/z* calc. for C$_{28}$H$_{42}$N$_8$O$_{18}$P$_3$S (M+H$^+$) 903.1545, detected 903.1559.
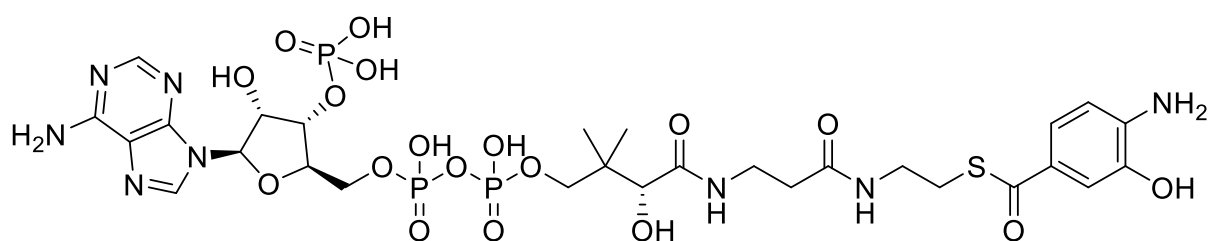
Chemical Formula: C$_{28}$H$_{41}$N$_8$O$_{18}$P$_3$S
Exact Mass: 902.14724

In vitro crosslinking assay

MTase Alb02 as well as PCP-1 and PCP-4 were concentrated in HEPES buffer (20 mM, pH 7.8) to a final concentration of 10 μM or 20 μM respectively. The commercially available and MS-cleavable crosslinker, disuccinimidyl sulfoxide (DSSO), was tested in different concentrations for our protein-interaction system (**Appendix Figure 3.4A**). Finally, DSSO was prepared in a 2 mM stock solution in DMSO. The DSSO crosslinker stock solution was each mixed with the peptidyl-carrier protein samples (PCP-1 & PCP-4) in a 20-fold molar excess (0.1 mM DSSO) (**Appendix Figure 3.4B**). Additionally, control samples containing the peptidyl-carrier proteins without crosslinking reagent (0 mM DSSO) were prepared. Subsequently, the methyltransferase Alb02 (10 μM) was added to all mixtures. Samples were incubated for 1 h at room temperature and terminated by adding NH_4HCO_3 (1 M) to a final concentration of 50 mM. Afterwards samples were submitted to SDS-PAGE gel electrophoresis (12% polyacrylamide) to evaluate the crosslinking.

In-gel digestion

Coomassie Brilliant Blue G250-stained bands were excised from the gel and subjected to in-gel reduction with 10 mM dithiothreitol in 25 mM ammonium bicarbonate, pH 8.3, for 45 min at 65 °C. Alkylation was performed with 50 mM iodoacetamide in 50 mM ammonium bicarbonate, pH 8.3 for 30 min at 25 °C, followed by overnight in-gel trypsin digestion (14 h at 37 °C with 66 ng sequencing-grade trypsin/μL in 25 mM ammonium bicarbonate, 10% ACN; 0.25 μg/sample). Tryptic peptides were dried in a vacuum centrifuge, redissolved in 15 μL 5% ACN + 0.1% HFO, and submitted to LC-MS/MS analysis on an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) coupled to an Agilent 1260 HPLC system (Agilent Technologies, Waldbronn, Germany). Reversed-phase separation was performed on a Grace Vydac 218MS C18 (2.1 × 150 mm; particle size, 5 μm) at 0.5 mL/min flow rate and developed with a gradient of 0.1% HFO in water (solution A) and in ACN (solution B), isocratically at 5% B for 5 min, followed by 5-40% B for 15 min, 40-99% B for 10 min, and 99% B for 5 min. MS experiments were performed with R=15,000 at *m/z* 300 and maximum filling time of 200 msec for survey scans. MS/MS fragmentation of the three most intense ions was performed in the LTQ using CID (30 msec activation time); the collision energy was set to 35% or 30% respectively. A window of *m/z* 3 was set for dynamic exclusion of up to 30 precursor ions with a repeat of 2 within 15 sec for the next 30 sec.

LC–MS/MS data files (.raw) were converted to mascot generic format (.mgf) files via MSConvert GUI of the ProteoWizard package (http://proteowizard.sourceforge.net; v3.0.10577) and annotated by peptide spectrum matching. The SearchGUI[279] (v3.3.11) software tool was used with X!Tandem as the search engine. The MS/MS spectra were searched against our customized

non-redundant protein database consisting of (i) PCPs and MTase Alb02 protein sequences; (ii) *E. coli* UniProt protein database; and (iii) a set of proteins found as common contaminants (common Repository of Adventitious Proteins, cRAP, 116 sequences. Precursor mass accuracy was set to 10 ppm and 0.5 Da for the MS2 level. The alkylation of Cys was set as a fixed modification. A false discovery rate (FDR) was estimated through a target-decoy approach, and a cutoff of 1% was applied (**Appendix Figure 3.9**).

To identify crosslinked peptides, MS fragmentation spectra were searched against a protein database comprising all protein sequences from *E. coli* BL21-Gold including the sequences of Alb02-His$_6$, PCP-1-His$_6$, and PCP-4-His$_6$ using MeroX[208] as the search engine. For MeroX, the precursor mass tolerance was set to 10 ppm and MS/MS mass tolerance was set to ± 0.1 Da.
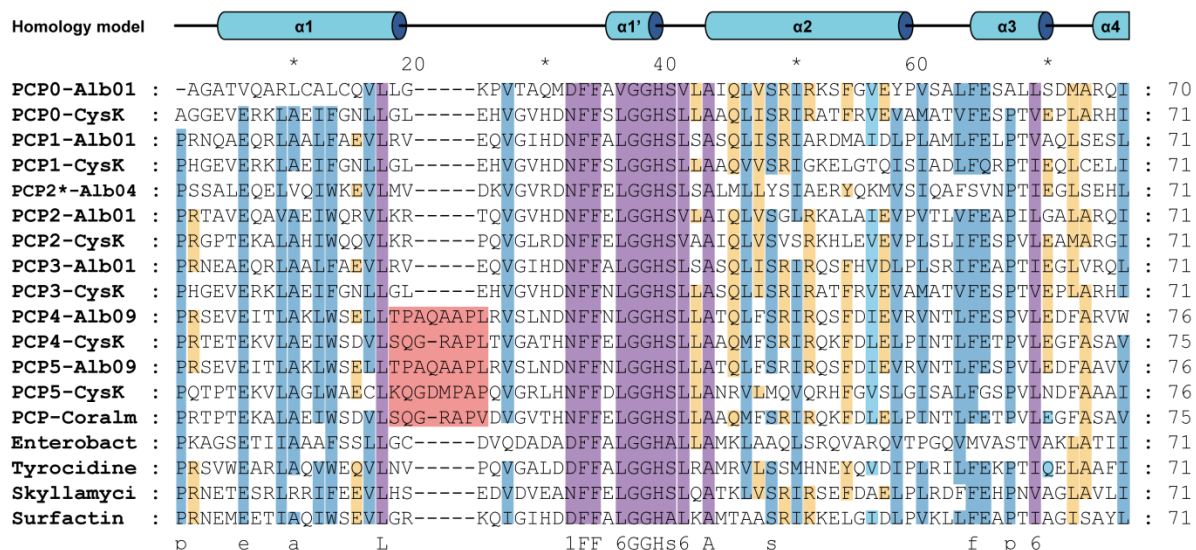
Appx Figure 3.1. Representative PCP homologs highly related to the biosynthesis gene cluster (BGC) of albicidin. *In silico* analysis of PCP domains in the BGC of albicidin revealed an exclusive sequence insertion (TPAQAAPLR) between helices α1 and α1' (light red) in PCP-4 and PCP-5, as well as in closely related cystobactamide and coralmycin cluster.[149–151] Conserved homology between aligned amino acid positions within the different PCP domains are indicated by a color-coded visualization from low sequence (light orange) to high sequence (light purple) homology.
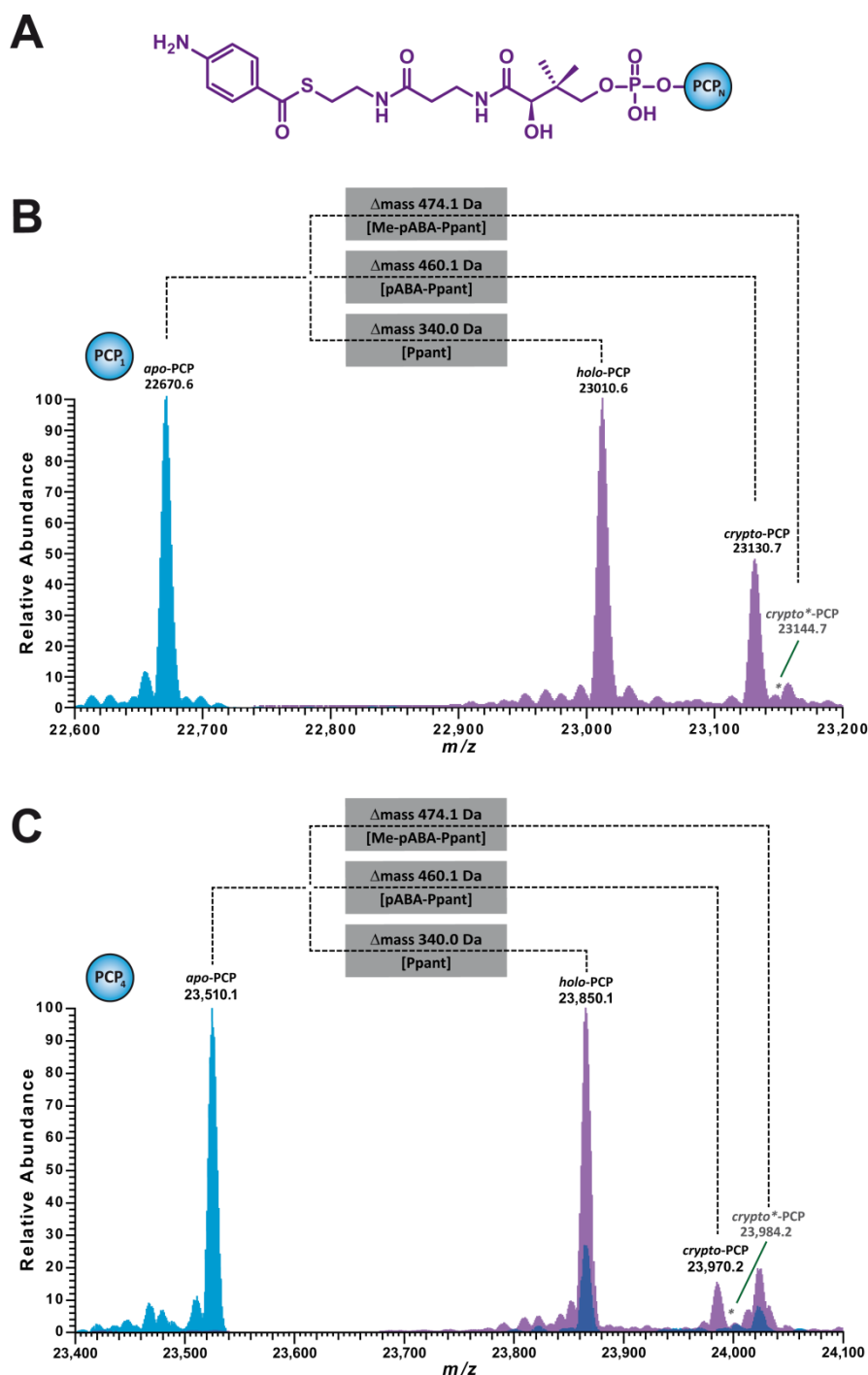
Appx Figure 3.2. On-line modification of *crypto*-PCPs monitored by intact mass spectrometry. (**A**) Schematic overview of different loading/modification steps of respective PCP domains with pABA. Loading and on-line modification of (**B**) PCP-1 and (**C**) PCP-4. Please note that above presented spectra are overlaid by three different spectra. *Blue*: inactive *apo*-PCP domain state. *Purple*: active *crypto*-PCP domain state including the Ppant arm loaded with pABA-3OH. *A*ctived *crypto*-PCP domain incubated with MTase Alb02 and SAM showed no formation of *crypto**-PCP (expected mass shown in grey). Single spectra show presence of pre-activated *holo*-PCP state caused by PPTase during recombinant expression.
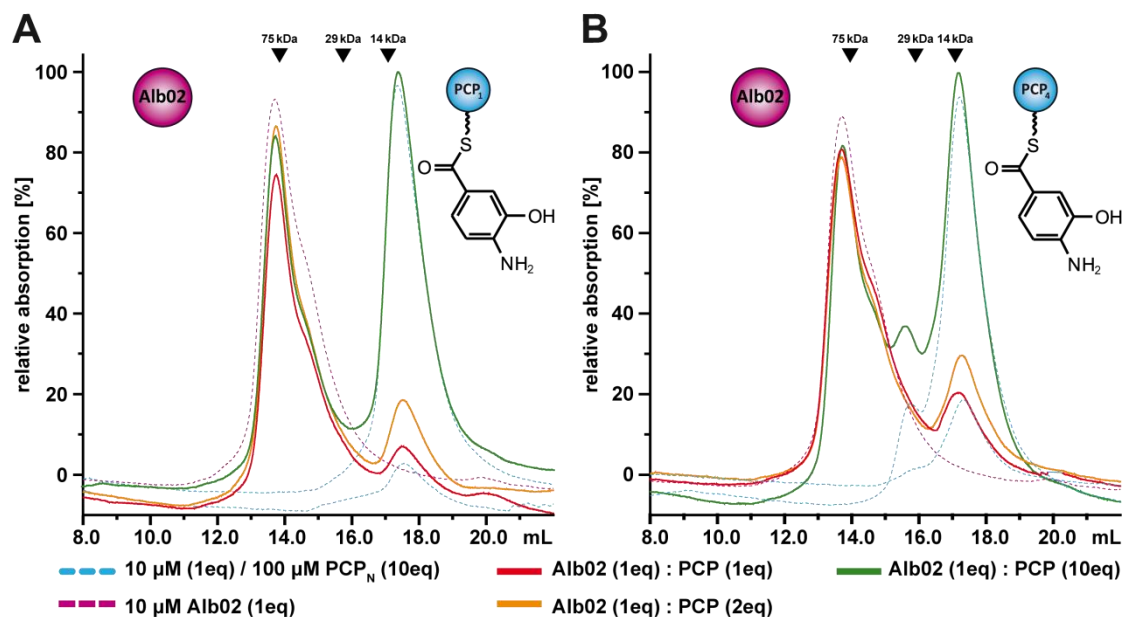
Appx Figure 3.3. Verification of Alb02-PCP complex formation by size exclusion liquid chromatography (SEC). Elution profiles of analytical SEC (λ = 280 nm) and molecular masses of internal standards (black triangles). Dashed lines indicate separate runs of Alb02 (purple) and respective *crypto*-PCP domain (blue). Solid lines indicate mixtures of different ratios for respective PCP domains and Alb02. (**A**) Analytical SEC elution profiles of *crypto*-PCP-1 in presence of MTase Alb02. (**B**) Analytical SEC elution profiles of *crypto*-PCP-4 in presence of MTase Alb02.
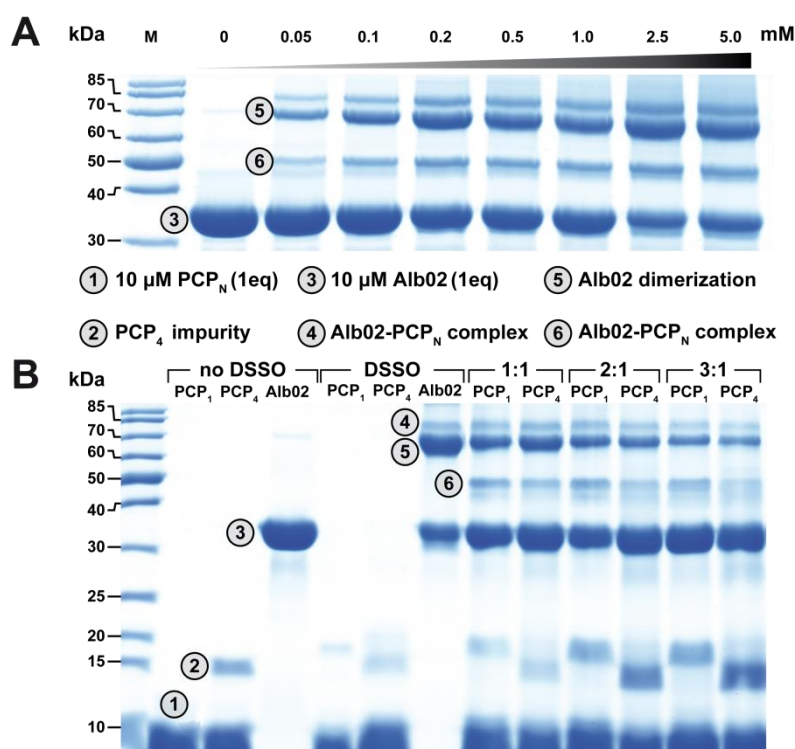
Appx Figure 3.4. *In vitro* crosslink assay and SDS-PAGE interaction complex validation. (**A**) Incubation of MTase Alb02 (10 μM, 1 eq) and PCP-4 (10 μM, 1 eq) in presence of various concentration of disuccinimidyl sulfoxide (DSSO) crosslinker at RT for 1 h and subsequent submitted to SDS-PAGE (12% polyacrylamide). Here, a zoom perspective shows areas of interest. (**B**) *In vitro* crosslinking assay of MTase Alb02 and PCPs. Control experiments performed w/

and w/o crosslinker DSSO (lane 2-7). *In vitro* crosslinking of MTase Alb02 in presence of different ratios PCPs (lane 8-13). (*1*) PCP-1 and PCP-4 (~10 kDa), (*2*) PCP-4 impurity (~15 kDa) (*3*) MTase Alb02 monomer (~39 kDa), (*4*) MTase Alb02 (dimer) -PCP (monomer) complex. (~80 kDa), (*5*) MTase Alb02 dimer (~70 kDa), (*6*) MTase Alb02 (monomer) - PCP (monomer) complex (~ 50 kDa).
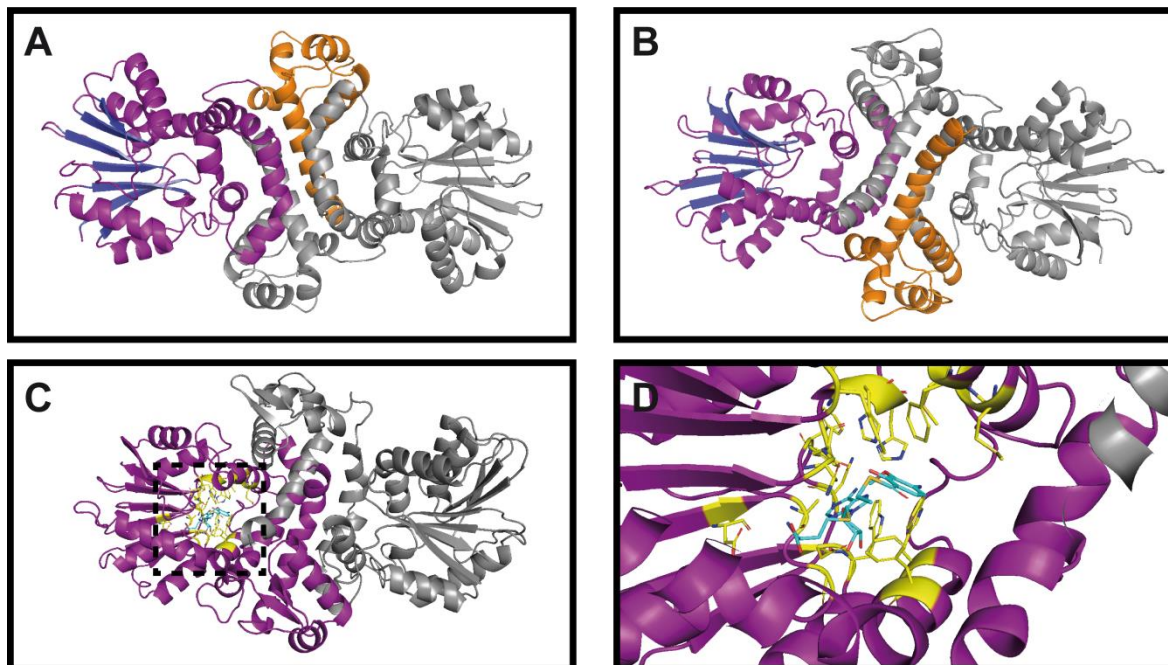
Appx Figure 3.5. Homology model of MTase Alb02. Model structure of a Alb02 dimer based on the related *O*-methyltransferase MmcR (pdb: 3gxo).[473] Regions of interest are color-coded showing the dimerization domain (orange), Rossman fold (blue), catalytic center (yellow), as well as chemical structures of *S*-adenosyl-L-homocysteine (SAH) and pABA-3OMe. (**A**)Full homology model of MTase Alb02 dimer divided into subdomains. For better clarity, the second Alb02 protein structure is shown in grey. (**B**) Full homology model of MTase Alb02 dimer with 180° rotation. (**C**) Full homology model of MTase Alb02 dimer shown the active site and catalytic center for methylation process. (**D**) Zoom perspective of the catalytic center of Alb02 showing pABA-3OMe SNAC and SAH in close proximity.

Appx Figure 3.6. Finger-print ¹H-¹⁵N HSQC spectrum of PCP-4. Chemical shift analysis allows identification of an α-helical structure with good agreement of previous PCP structures and our initial structure homology model (**Figure 3.4**). Sequence insertion in PCP-4 was classified as dynamic with no tendency to adopt a specific secondary structure under the measured conditions. A second resonance set for several amide signals located in helix α2 and α3 in the vicinity of the Ppant attachment site Ser46 showing the presence of both *apo-* and *holo*-PCP-4 (**Figure 3.4**).

Appx Figure 3.7. Protein purification via Ni-affinity and size exclusion chromatography (SEC). Each of the three figures shows two orthogonal chromatograms. The smaller Ni-affinity chromatogram, performed with 1mL HisTrap crude FF crude columns (GE Healthcare), shows an imidazole step gradient (green) and the pooled fractions (red) for the next purification step. The bigger SE chromatogram, performed by a Superdex75 column (GE Healthcare), shows the pooled and stored protein fractions (red). (**A**) Chromatograms of PCP-1 domain purification by two chromatographic steps. (**B**) Chromatograms of PCP-4 domain purification by two chromatographic steps. (**C**) Chromatograms of the Alb02 domain purification by two chromatographic steps.
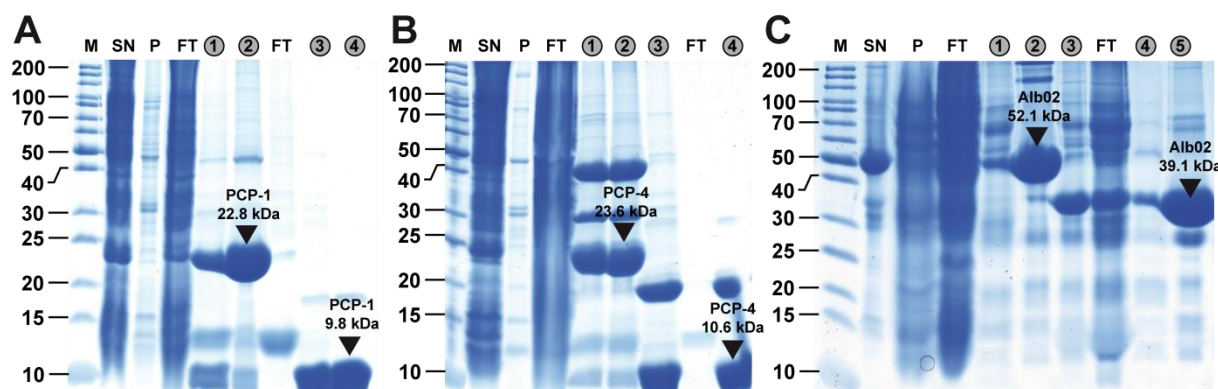
Appx Figure 3.8. Protein identification of PCPs and MTase Alb02 and purity control by SDS-PAGE. Abbreviations are to be interpreted accordingly: M – marker, SN – supernatant of lysed cells, P – cell debris, FT – flow through of Ni affinity HisTrap before or after TEV digestion, Numbers are assigned to the respective purification chromatograms (**Appendix Figure 3.7**) (**A**) Protein verification and purity control of PCP-1 (15% polyacrylamide) (**Appendix Figure 3.7A**). (**B**) Protein verification and purity control of PCP-4 (15% polyacrylamide) (**Appendix Figure 3.7B**). (**C**) Protein verification and purity control of MTase Alb02 (12% polyacrylamide) (**Appendix Figure 3.7C**).

PCP-1 bottom-up sequencing
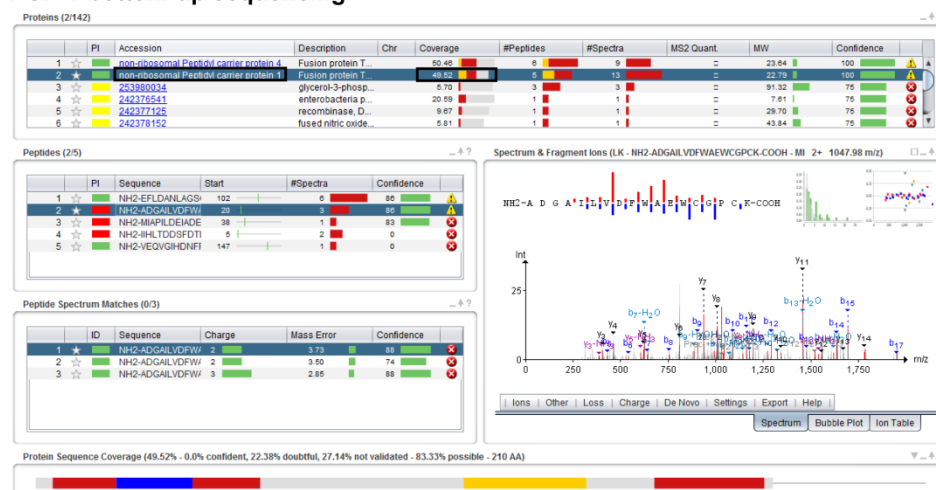
PCP-4 bottom-up sequencing
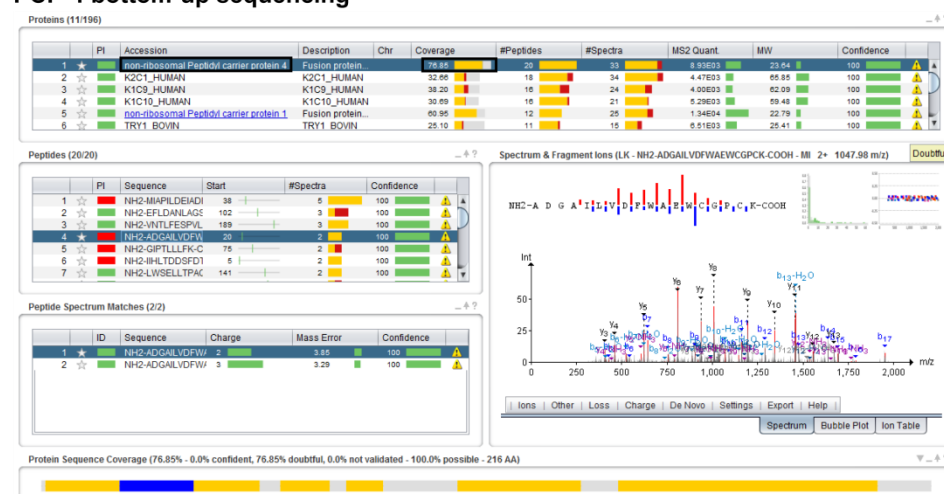
Alb02 bottom-up sequencing
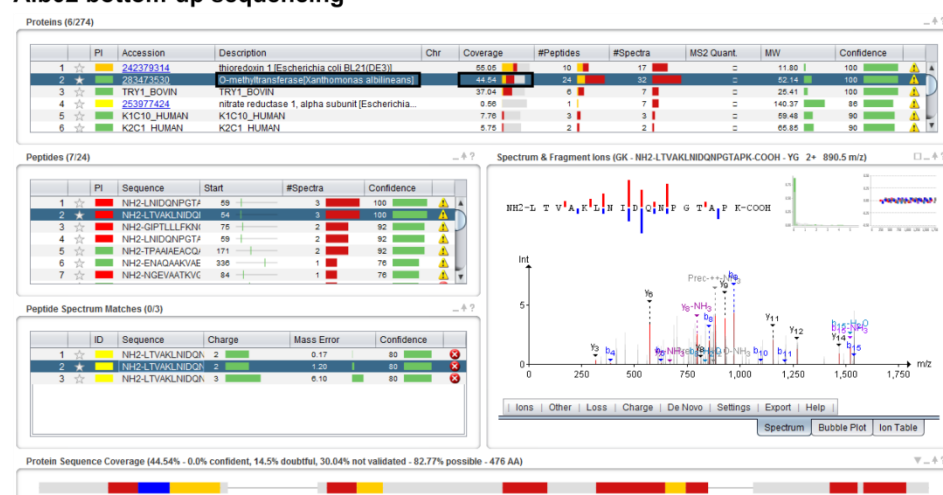
Appx Figure 3.9. Protein identification of in-gel tryptic digest and annotation by peptide spectrum matching. MS/MS spectra were searched against an E. coli BL21-Gold (DE3) protein database including the sequences of PCP-1/4 and MTase Alb02 (program: Peptideshaker 1.16).

Appx Table 3.1. Oligonucleotide sequences. The genes were amplified by PCR using the templates cosmid pALB540 or pALB571. The PCR-products were digested with NotI and NcoI, ligated into pETtrx_1c and subsequently transformed into *E. coli* BL21-Gold (DE3) strain.

Name	Sequence (5'→ 3')	Target	Template
pETtrx1c_PCP-1_for	CACGCCATGGAGTACGTCGCGCCACGCAAC	Alb01, NRPS-1, PCP-1	pALB571
pETtrx1c_PCP-1_rev	GCGTGCTGCGGCCGCGAGGGATTCGCTAAG	Alb01, NRPS-1, PCP-1	pALB571
pETtrx1c_PCP-4_for	CCAGCACCATGGGCGACTACCTCGCTCCGCGCA	Alb09, NRPS-4, PCP-4	pALB571
pETtrx1c_PCP-4_rev	CCGCGCGGCCGCGACCACTCGCGCAAAATCTTCCAG	Alb09, NRPS-4, PCP-4	pALB571
pETtrx1c_Alb02_for	GCAACCATGGATTCAGCGTTACCTACA	Alb02	pALB540
pETtrx1c_Alb02_rev	CCGAGCGGCCGCTTATGGGGCCCTCTTGCGGGA	Alb02	pALB540

Appx Table 3.2. Backbone resonance assignment of ^{13}C- and ^{15}N-labeled PCP-4 in PBS, pH 7.8 at 298 K.

pos.	aa	atom	δ(ppm)
2	G	C	173.834
2	G	CA	45.375
2	G	HA	4.003
3	D	C	175.663
3	D	CA	54.123
3	D	CB	41.318
3	D	HA	4.666
3	D	HN	8.259
3	D	N	120.862
4	Y	C	174.211
4	Y	CA	59.122
4	Y	CB	38.662
4	Y	HA	4.381
4	Y	HN	8.239
4	Y	N	121.779
5	L	C	174.273
5	L	CA	52.852
5	L	CB	43.91
5	L	HA	4.296
5	L	HN	8.133
5	L	N	107.983
6	A	C	175.483
6	A	CA	50.567
6	A	CB	18.276
6	A	HA	4.047
6	A	HN	7.951
6	A	N	127.702
7	P	C	175.382
7	P	CA	63.456
7	P	CB	32.483
7	P	HA	3.975
8	R	C	176.567
8	R	CA	57.254
8	R	CB	32.792
8	R	HA	4.519
8	R	HN	8.844
8	R	N	121.823
9	S	C	174.14
9	S	CA	56.617
9	S	CB	65.57
9	S	HA	4.594
9	S	HB2	4.398
9	S	HB3	3.941
9	S	HN	8.325
9	S	N	116.48
10	E	C	179.742
10	E	CA	60.428
10	E	CB	29.286
10	E	HA	4.063
11	V	C	177.657
11	V	CA	67.217
11	V	CB	31.872
11	V	HA	3.711
11	V	HN	8.13
11	V	N	118.797
12	E	C	178.136
12	E	CA	60.268
12	E	CB	30.308
12	E	HA	3.493
12	E	HN	7.712
12	E	N	119.059
13	I	C	179.008
13	I	CA	65.803
13	I	CB	38.903
13	I	HA	3.4
13	I	HN	8.965
13	I	N	119.298
14	T	C	176.924
14	T	CA	67.227
14	T	CB	68.556
14	T	HA	3.926
14	T	HB	4.385
14	T	HN	8.481
14	T	N	117.969
15	L	C	177.434
15	L	CA	58.202

pos.	aa	atom	δ(ppm)
28	A	C	176.577
28	A	CA	50.986
28	A	CB	21.234
28	A	HA	4.557
28	A	HN	7.745
28	A	N	122.747
29	A	C	174.767
29	A	CA	50.376
29	A	CB	17.227
29	A	HA	4.447
29	A	HN	8.165
29	A	N	123.599
30	P	C	177.99
30	P	CA	63.197
30	P	CB	32
30	P	HA	4.336
31	L	C	177.993
31	L	CA	55.522
31	L	CB	43.755
31	L	HA	4.16
31	L	HN	8.366
31	L	N	125.311
32	R	C	174.902
32	R	CA	55.565
32	R	CB	29.608
32	R	HA	4.466
32	R	HN	8.903
32	R	N	128.176
33	V	C	174.798
33	V	CA	63.543
33	V	CB	32.882
33	V	HA	4.116
33	V	HN	8.367
33	V	N	127.117
34	S	C	176.269
34	S	CA	55.165
34	S	CB	64.99
34	S	HA	4.516
34	S	HB2	4.067
34	S	HB3	3.89
34	S	HN	9.631
34	S	N	123.056
35	L	C	176.273
35	L	CA	57.84
35	L	CB	42.67
35	L	HA	3.882
35	L	HN	8.915
35	L	N	123.81
36	N	C	175.817
36	N	CA	51.924
36	N	CB	38.049
36	N	HA	4.797
36	N	HB2	3.202
36	N	HB3	2.793
36	N	HN	7.401
36	N	N	109.541
37	D	C	174.52
37	D	CA	55.614
37	D	CB	42.729
37	D	HA	4.944
37	D	HN	7.79
37	D	N	120.012
38	N	C	175.232
38	N	CA	51.433
38	N	CB	41.424
38	N	HA	5.496
38	N	HN	8.709
38	N	N	119.752
39	F	C	175.51
39	F	CA	62.152
39	F	CB	40.5
39	F	HA	4.443
39	F	HN	8.577
39	F	N	126.318
40	F	C	179.578
40	F	CA	60.261

pos.	aa	atom	δ(ppm)
40	F	CB	37.696
40	F	HA	3.95
40	F	HN	8.011
40	F	N	114.69
41	N	C	177.43
41	N	CA	55.575
41	N	CB	37.773
41	N	HA	4.56
41	N	HN	8.094
41	N	N	122.538
42	L	C	176.008
42	L	CA	55.102
42	L	CB	43.912
42	L	HA	4.282
42	L	HN	7.572
42	L	N	117.377
43	G	C	174.76
43	G	CA	44.25
43	G	HA	4.357
43	G	HN	7.482
43	G	N	128.629
44	G	C	177.349
44	G	CA	45.613
44	G	HA2	2.947
44	G	HA3	2.157
44	G	HN	7.864
44	G	N	108.683
45	H	C	176.828
45	H	CA	54.586
45	H	CB	32.397
45	H	HA	5.198
45	H	HN	6.136
45	H	N	112.203
47	L	C	179.899
47	L	CA	58.312
47	L	CB	41.798
47	L	HA	4.515
48	L	C	178.69
48	L	CA	56.536
48	L	CB	42.801
48	L	HA	4.255
48	L	HN	7.448
48	L	N	119.833
49	A	C	178.432
49	A	CA	55.291
49	A	CB	18.26
49	A	HA	3.639
49	A	HN	8.347
49	A	N	120.913
50	T	C	177.599
50	T	CA	67.284
50	T	CB	68.651
50	T	HA	4.078
50	T	HB	4.429
50	T	HN	8.051
50	T	N	112.843
51	Q	C	178.638
51	Q	CA	58.968
51	Q	CB	28.489
51	Q	HA	4.261
51	Q	HN	7.654
51	Q	N	121.797
52	L	C	178.466
52	L	CA	58.491
52	L	CB	39.957
52	L	HA	3.921
52	L	HN	8.344
52	L	N	120.915
53	F	C	179.053
53	F	CA	63.178
53	F	CB	37.852
53	F	HA	4.224
53	F	HN	8.408
53	F	N	117.117
54	S	C	177.437
54	S	CA	61.865
54	S	CB	62.614

pos.	aa	atom	δ(ppm)
54	S	HA	4.521
54	S	HB	4.17
54	S	HN	8.057
54	S	N	117.874
55	R	C	180.062
55	R	CA	57.794
55	R	CB	29.123
55	R	HA	4.318
55	R	HN	8.114
55	R	N	121.318
56	I	C	177.649
56	I	CA	66.449
56	I	CB	38.017
56	I	HA	3.705
56	I	HN	8.889
56	I	N	122.709
57	R	C	179.521
57	R	CA	59.893
57	R	CB	29.7
57	R	HA	4.468
57	R	HN	8.159
57	R	N	122.163
58	Q	C	178.001
58	Q	CA	58.671
58	Q	CB	29.172
58	Q	HA	4.156
58	Q	HN	7.881
58	Q	N	117.62
59	S	C	175.089
59	S	CA	62.134
59	S	CB	63.947
59	S	HA	4.211
59	S	HB2	3.463
59	S	HB3	3.137
59	S	HN	8.281
59	S	N	113.265
60	F	C	174.967
60	F	CA	57.468
60	F	CB	41.685
60	F	HA	4.883
60	F	HN	8.572
60	F	N	115.866
61	D	C	174.314
61	D	CA	55.549
61	D	CB	39.949
61	D	HA	4.71
61	D	HN	8.181
61	D	N	118.412
62	I	C	173.828
62	I	CA	58.745
62	I	CB	42.308
62	I	HA	4.557
62	I	HN	6.713
62	I	N	107.62
63	E	C	175.197
63	E	CA	54.937
63	E	CB	30.921
63	E	HA	4.628
63	E	HN	8.379
63	E	N	122.68
64	V	C	174.313
64	V	CA	60.497
64	V	CB	34.003
64	V	HA	4.224
64	V	HN	8.173
64	V	N	125.227
65	R	C	178.212
65	R	CA	56.482
65	R	CB	30.9
65	R	HA	4.153
65	R	HN	8.487
65	R	N	124.65
66	V	C	176.442
66	V	CA	63.976
66	V	CB	31.37
66	V	HA	2.797
66	V	HN	8.308

pos.	aa	atom	δ(ppm)
66	V	N	124.21
67	N	C	176.962
67	N	CA	55.851
67	N	CB	37.62
67	N	HN	8.148
67	N	N	114.345
68	T	C	176.974
68	T	CA	65.988
68	T	CB	69.534
68	T	HA	4.104
68	T	HB	4.276
68	T	HN	7.876
68	T	N	112.214
69	L	C	175.81
69	L	CA	56.646
69	L	CB	42.785
69	L	HA	4.123
69	L	HN	7.45
69	L	N	121.841
70	F	C	178.769
70	F	CA	60.39
70	F	CB	38.642
70	F	HA	4.229
70	F	HN	6.955
70	F	N	113.315
71	E	C	177.727
71	E	CA	58.103
71	E	CB	30.875
71	E	HA	4.216
71	E	HN	7.94
71	E	N	117.277
72	S	CA	55.48
72	S	CB	63.835
72	S	HA	5.17
72	S	HB2	4.196
72	S	HB3	3.863
72	S	HN	7.847
72	S	N	114.935
73	P	C	177.151
73	P	CA	64.196
73	P	CB	32.524
73	P	HA	4.742
74	V	C	178.756
74	V	CA	61.657
74	V	CB	31.944
74	V	HA	5.005
74	V	HN	7.764
74	V	N	122.173
75	L	C	177.822
75	L	CA	59.937
75	L	CB	38.602
75	L	HA	3.742
75	L	HN	8.765
75	L	N	128.732
76	E	C	178.005
76	E	CA	60.864
76	E	CB	30.004
76	E	HA	3.593
76	E	HN	8.842
76	E	N	116.378
77	D	C	178.225
77	D	CA	56.816
77	D	CB	40.991
77	D	HA	4.521
77	D	HN	7.016
77	D	N	116.642
78	F	C	177.84
78	F	CA	62.011
78	F	CB	38.718
78	F	HA	4.26
78	F	HN	9.204
78	F	N	124.259
79	A	C	178.705
79	A	CA	54.54
79	A	CB	18.758
79	A	HA	3.51
79	A	HN	9.356
79	A	N	118.084

pos.	aa	atom	δ(ppm)
80	R	C	179.514
80	R	CA	59.903
80	R	CB	29.986
80	R	HA	4.022
80	R	HN	7.298
80	R	N	118.061
81	V	C	179.278
81	V	CA	66.207
81	V	CB	31.345
81	V	HA	3.688
81	V	HN	7.454
81	V	N	121.022
82	V	C	176.851
82	V	CA	66.892
82	V	CB	30.944
82	V	HA	3.158
82	V	HN	7.645
82	V	N	121.54
83	A	C	180.374
83	A	CA	55.345
83	A	CB	18.045
83	A	HA	3.875
83	A	HN	8.749
83	A	N	120.446
84	A	C	180.131
84	A	CA	54.844
84	A	CB	17.866
84	A	HA	4.199
84	A	HN	7.792
84	A	N	120.017
85	A	C	180.336
85	A	CA	54.548
85	A	CB	18.194
85	A	HA	4.246
85	A	HN	7.688
85	A	N	122.09
86	L	C	178.451
86	L	CA	57.384
86	L	CB	42.355
86	L	HA	4.003
86	L	HN	8.064
86	L	N	119.693
87	E	C	177.95
87	E	CA	58.488
87	E	CB	29.656
87	E	HA	4.074
87	E	HN	7.769
87	E	N	118.898
88	H	C	176.505
88	H	CA	58.124
88	H	CB	30.774
88	H	HA	4.461
88	H	HN	7.889
88	H	N	117.587
X	PNS	HN41	8.130
X	PNS	N41	124.602
X	PNS	HN36	7.880
X	PNS	N36	119.713

Appendix to Chapter 4

Appx Table 4.1. Venom proteins and peptides identified from *Vipera ammodytes transcaucasiana*. Assignment of venomic components by crude venom intact mass profiling (method A) and bottom-up (method C). RP-HPLC fractions with low molecular masses were additionally analyzed by intact mass profiling (method B). Peak numbers based on RP-HPLC (**Figure 4.4A**) and TIC (**Figure 4.2A**) annotation. Sequence tags were obtained *de novo* from MS/MS spectra and identified against a non-redundant *Viperidae* protein database (taxid: 8689) by BLASTP. SDS-PAGE and intact mass profile analysis provided the molecular weight.

Peak number	Method	SDS PAGE M_{av} [kDa]	Identification sequence tag	Protein species	Protein ID	BLAST E-value	NCBI accession number	intact mass *m/z* native average
1	A; B	-	-	-	unknown (peptide)	-	-	409.05, 442.12, 519.19, 549.21, 571.25, 611.20, 689.34, 1005.65, 1607.08
2	A; B	-	DEPLKK	-	unknown (peptide)	-	-	413.26, 484.20, 612.26, 967.40
3	A; B	-	-	-	unknown (peptide)	-	-	413.19, 485.70, 600.32, 617.80, 823.40, 1106.51, 1607.03
4	A; B	-	EDETPKM	svMP	metalloproteinase type III 6a	6.0E-04	JAS05414.1	413.36, 433.18, 490.20, 547.22, 607.27, 685.32, 773.49, 837.50, 849.36, 6424.04, 6693.16, 6793.24
		-	pEKW	svMP-i	tripeptide metalloproteinase inhibitor	-	-	
		-	GGGGGGW	svMP-i	endogenous tripeptide metalloproteinase inhibitor precursor	4.0E-01	AMB36336.1	
5	A; B	-	SEDYSETHYSPDWR	svMP	H3 metalloproteinase precursor 1	3.0E-07	AGL45259.1	
		-	PVSGNEL	svMP	metalloproteinase of class P-II MPII-3	2.8E-01	AMB36351.1	
		-	-	-	unknown (peptide)	-	-	457.11, 715.36, 814.35, 846.35
6	A; B	-	pEKW	svMP-i	svMP-i tripeptide	-	-	444.22, 448.22, 456.22, 460.22, 476.21, 486.23
7	A; B	-	pENW	svMP-i	svMP-i tripeptide	-	-	413.14, 430.17, 452.15, 706.38, 859.33, 1143.79, 3835.92, 4014.86, 4052.80, 4081.80
8	A; B	-	-	-	unknown (peptide)	-	-	569.28, 851.46, 907.44, 3035.36, 3943.82, 4014.88, 7486.42

Appx Table 4.1. Venom proteins and peptides identified from *Vipera ammodytes transcaucasiana*. *continued*

Peak number	Method	SDS PAGE M$_{av}$ [kDa]	Identification sequence tag	Protein species	Protein ID	BLAST E-value	NCBI accession number	intact mass *m/z* native average
9	A; B	-	PEGPPLMEPHE	svMP-i	endogenous tripeptide metalloproteinase inhibitor precursor	9.0E-07	AMB36336.1	569.28, 1232.56, 3859.92, 3931.00, 4296.04, 5775.64, 7470.42
10	A; B	-	-	-	unknown (peptide)	-	-	416.14, 723.35, 869.40, 3761.75, 3931.00, 4176.85, 8352.70
11	A; B	-	PPRCPGPKVPP	BPP	Bradykinin-potentiating peptide 12e	1.1E-01	P0DL03.1	723.36, 1143.64, 3380.75, 3796.75, 4861.17, 7470.46
12	A; B	-	-	-	unknown (peptide)	-	-	904.51, 932.51, 1143.64, 6970.90, 7488.48
13	A; B	-	-	-	unknown (peptide)	-	-	904.51, 932.51, 1143.64, 1472.76, 6785.75, 8351.72
14	A; B	-	QMPPPGPKVPPLK	svMP-i	endogenous tripeptide metalloproteinase inhibitor precursor	2.1E-02	AMB36336.1	904.51, 932.51, 1143.64, 7332.38, 14427.10
			PPQMPGPKVPP	svMP-i	endogenous tripeptide metalloproteinase inhibitor precursor	8.3E-02	AMB36335.1	
15	C	15	CSGCCTDESLK	VEGF-F	Vammin	3.0E-06	P67863.2	
			PFLEVHER			2.0E-03		
			366.21-PFLEVHER			3.0E-03		
			803.34-TVDLQIMR			3.0E-03		
16a	C	50	-	-	unknown (protein)	-	-	
16b	A; C	27	DASYFYCR	svMP	metalloproteinase	1.1E-01	ADI47645.1	21311.88, 21213.90
			GESYFYCR			1.0E-03		
16c	C	15	LLPNTLQFLLK	-	rho GTPase-activating protein 29	9.0E-03	XP_015684754.1	10676.97
			366.21-PFLEVHER	VEGF-F	Vammin	3.0E-03	P67863.2	
17a	C	40	-	-	unknown (protein)	-	-	
17b	C	25	NLYQFGNMIFK	PLA$_2$	Ammodytin I2(A) variant	4.0E-07	CAE47197.1	
			384.24-SYSNYGCYFDS-296.29		Vaspin basic subunit variant	3.0E-03	CAE47300.1	
			LAIYSYSFK			6.0E-04		
			NLFQFAK			6.4E-02		

Appx Table 4.1. Venom proteins and peptides identified from *Vipera ammodytes transcaucasiana*. *continued*

Peak number	Method	SDS PAGE M$_{av}$ [kDa]	Identification sequence tag	Protein species	Protein ID	BLAST E-value	NCBI accession number	intact mass *m/z* native average
17c	A; C	13	GTLLSYSNYGCYCGWGGK	PLA$_2$	Ammodytin I2(A) variant	4.0E-12	CAE47197.1	13553.83, 13590.76, 13814.21, 13842.19, 13911.15
			LGAICFGENLNTYDKK			2.0E-09		
			354.23-CFGENLNTYDKK			1.0E-07		
			NLYQFGNMIFK			4.0E-07		
			227.13-YQFGNFLFK			2.0E-01		
18a	C	40	LAIYYYSFK	PLA$_2$	Vipoxin Chain B	2.0E-04	1AOK_B	
			NLFQFAK		Vaspin basic subunit variant	6.4E-02	CAE47300.1	
18b	C	25	DICDCERVAANCLNGCLP	PLA$_2$	Vaspin basic subunit variant	1.0E-05	CAE47300.1	
			LGANCFHQNK			3.0E-03		
			LAIYYYSFK		Vipoxin Chain B	2.0E-04	1AOK_B	
18c	A; C	13	NLYQFGNMIFK	PLA$_2$	Ammodytin I2(A) variant	4.0E-07	CAE47197.1	13918.28, 14016.22
			VAANCFHQNK		Vaspin basic subunit variant	3.0E-05	CAE47300.1	
			LAIYYYSFK		Vipoxin Chain B	2.0E-04	1AOK_B	
19a	C	50	310.13-YVCQYCPAGNLIGK	CRISP	cysteine-rich venom protein	6.0E-09	XP_015678374.1	
			SVDFDSESPR			2.0E-05	P86537.1	
			LFWYPEAAA-470.22			3.4E-02		
19b	A; C	25	310.13-YVCQYCPAGNLIGK	CRISP	cysteine-rich venom protein	6.0E-09	XP_015678374.1	24653.41, 24752.38, 24848.3
			276.08-WYPEAAANVTR			3.0E-03		
19c	A; C	13	LAIYSYSFK	PLA$_2$	Vaspin basic subunit variant	6.0E-04	CAE47300.1	12346.55
20a	C	60	NLFQFGDMILQK	PLA$_2$	phospholipase A2 I acidic chain	9.0E-08	A60512	
			227.13-FQFGDFILQK		Vaspin acidic subunit (1) variant	2.0E-03	CAE47105.1	
			1110.6-TAIDFNGLITGK	svMP	group III metalloproteinase	2.0E-03	ADI47577.1	
			PCLTLYQCR			9.1E-02	ADW54342.1	
			786.47-PLVGVELWR			3.1E-01		

Appx Table 4.1. Venom proteins and peptides identified from *Vipera ammodytes transcaucasiana*. *continued*

Peak number	Method	SDS PAGE M$_{av}$ [kDa]	Identification sequence tag	Protein species	Protein ID	BLAST E-value	NCBI accession number	intact mass *m/z* native average
20b	C	27.5	ALYGCYCGWGGQGR	PLA$_2$	Vaspin acidic subunit (1) variant	7.0E-09	A60512	
			NLFQFGDMILQK			9.0E-08		
			326.20-SSLGENVNTYDK			1.0E-04	CAE47105.1	
			SVDFDSESPR	CRISP	cysteine-rich venom protein	2.0E-05	P86537.1	
			568.33-SSGENLYMSTSPMK			1.0E-05	B7FDI0.1	
			KPELQQDLLDLH-470.26			7.2E-01		
20c	A; C	22.5	NLFQFGDMILQK	PLA$_2$	Vaspin acidic subunit (1) variant	9.0E-08	A4VBF0.1	24515.96
			326.20-SSLGENVNTYDK			9.0E-05		
20d	A; C	13	HLSQFGDMINKK	PLA$_2$	Ammodytin I2(A) variant	2.0E-07	CAE47141.1	13624.69, 13625.73, 13676.78
			354.24-CFGENMNTYDQK			8.0E-07		
			KLLCFGENMNTYDQK			4.0E-06		
21	C	35	TLCAGLLQGGLDSCK	svSP	serine proteinase-like protein 2	2.0E-06	Q9PT40.1	
22	C	35	-	-	unknown (protein)	-	-	
23a	C	85	LMGWGTITTTK	svSP	enzymatically inactive serine proteinase-like protein SPH-1	3.0E-05	AMB36342.1	
			VVCAGIWQGGK		Nikobin	5.0E-06	E5AJX2.1	
23b	C	35	LMGWGTITTTK	svSP	enzymatically inactive serine proteinase-like protein SPH-1	3.0E-05	AMB36342.1	
			VVCAGIWQGGK		Nikobin	5.0E-06	E5AJX2.1	
24a	C	85	VVCAGIWQGGK	svSP	Nikobin	5.0E-06	E5AJX2.1	
24b	C	35	VVCAGIWQGGK	svSP	Nikobin	5.0E-06	E5AJX2.1	
			LMGWGTISTSK			6.0E-05	ABG26974.1	
25a	C	85	VVCAGIWQGGK	svSP	Nikobin	5.0E-06	E5AJX2.1	
			LMGWGTISTSK		thrombin-like enzyme	6.0E-05	O13069.1	
25b	C	35	VIGGDQCDINEHPFLAFVTDSR	svSP	Nikobin	4.0E-13	E5AJX2.1	
			1142.55-DAVLTAAHCNGK		Nikobin	3.0E-04	E5AJX2.1	
			LMGWGTISSTK		Cadam10_SVSP-11	6.0E-05	JAV48393.1	
26a	C	100	VVCAGIWQGGK	svSP	Nikobin	5.0E-06	E5AJX2.1	
			LMGWGTISSTK		Cadam10_SVSP-11	6.0E-05	JAV48393.1	

Appx Table 4.2. Venom proteins and peptides identified from *Vipera ammodytes montandoni.* Assignment of venomic components by crude venom intact mass profiling (method A) and bottom-up (method C). RP-HPLC fractions with low masses were additionally analyzed by intact mas profiling (method B). Peak numbers based on RP-HPLC (**Figure 4.4B**) and TIC (**Figure 4.2B**) annotation. Sequence tags were obtained *de novo* from MS/MS spectra and identified against a non-redundant *Viperidae* protein database (taxid: 8689) by BLASTP. SDS-PAGE and intact mass profile analysis provided the molecular weight.

Peak Number	Method	SDS PAGE M_{av} [kDa]	Identification sequence tag	Protein species	Protein ID	BLAST E-value	NCBI accession number	intact mass *m/z* native average
1	A; B	-	-	-	unknown (peptide)	-	-	451.00, 484.21, 506.19, 612.27, 686.99, 1801.00, 2229.00, 6842.00, 7384.28
2	A; B	-	DNEPPKKVPPN	svMP-i	endogenous tripeptide metalloproteinase inhibitor precursor	3.0E-06	AMB36336.1	496.24, 688.36, 1234.65
3	A; B	-	pEKW	svMP-i	svMP-i tripeptide	-	-	444.22, 466.21, 887.44, 963.34
4	A; B	-	pEKW	svMP-i	svMP-i tripeptide	-	-	444.23, 460.22, 754.36, 814.36
5	A; B	-	pEKW	svMP-i	svMP-i tripeptide	-	-	444.22, 452.15, 466.21, 468.12, 719.40, 723.36, 747.40, 809.39, 887.44, 963.34, 1072.60, 1110.54, 7384.27, 7531.52
		-	GGGGGGW	svMP-i	endogenous tripeptide metalloproteinase inhibitor precursor	4.0E-01	AMB36336.1	
6	A; B	-	pENW	svMP-i	svMP-i tripeptide	-	-	430.17, 859.34, 1072.6, 1288.51
7	A; B	-	-	-	unknown (peptide)	-	-	470.72, 1072.60
8	A; B	-	PEGPPLMEPHE	svMP-i	endogenous tripeptide metalloproteinase inhibitor precursor	9.0E-07	AMB36336.1	451.31, 2854.40, 2951.45, 3501.81, 3930.96, 5776.60, 6842.05, 7384.32, 7375.43
9	A; B	-	-	-	unknown (peptide)	-	-	416.14, 456.18, 584.28, 699.31, 836.37, 3664.67, 3761.72, 4176.86, 4214.80
10	A; B	-	-	-	unknown (peptide)	-	-	456.22, 722.37, 759.74, 833.44, 904.51, 932.51, 1144.63, 1182.58, 3283.50, 3381.55, 3834.62, 3872.59, 3928.99, 4174.87
11	A; B	-	-	-	unknown (peptide)	-	-	608.30, 681.30, 722.38, 833.44, 932.51, 1006.55, 1144.63, 1182.58, 1460.80, 3796.73, 4730.20, 13971.71
12	A; B	-	-	-	unknown (peptide)	-	-	932.52, 1144.62
13	A; B	-	-	-	unknown (peptide)	-	-	932.52, 1144.62, 1314.73
14	A; B	-	-	-	unknown (peptide)	-	-	573.03, 744.13, 1102.59

Appx Table 4.2. Venom proteins and peptides identified from *Vipera ammodytes montandoni*. *continued*

Peak Number	Method	SDS PAGE M_{av} [kDa]	Identification sequence tag	Protein species	Protein ID	BLAST E-value	NCBI accession number	intact mass *m/z* native average
15	A; B	-	-	-	unknown (peptide)	-	-	-
16	A; C	14	PFLEVHER	VEGF-F	snake venom vascular endothelial growth factor	3.0E-03	P67863.2	834.39, 947.48, 1159.59
17	A; C	14	PFLEVHER	VEGF-F	snake venom vascular endothelial growth factor	3.0E-03	P67863.2	834.39, 947.48, 1159.59
18a	A; C	25	-	-	unknown (protein)	-	-	21199.58, 21298.85
18b	A; C	14	-	-	unknown (protein)	-	-	16307.27
19	A; C	-	-	-	unknown (protein)	-	-	13553.82, 13590.75
20	A; C	13	152.12-FFVHDCCYGR	PLA₂	Ammodytin I1(A) variant	2.0E-05	CAE47140.1	3532.76, 13890.28, 13988.22
			LAIYYYSFK		Vipoxin Chain B	2.0E-04	1AOK_B	
			NLFQFAK		Vaspin basic subunit variant	7.2E-02	CAE47300.1	
21a	A; C	23	310.13-YVCQYCPAGNLIGK	CRISP	cysteine-rich venom protein	6.0E-09	P79845.2	24654.40, 24750.41
			SVDFDSESPR			2.0E-05	P86537.1	
			MEWYPEAAAWQGV			1.0E-04	BAP39957.1	
			CWNLLMSPYPMK			1.8E-02	Q7ZZN9.1	
21b	C	13	NLFQFAK	PLA₂	Vaspin basic subunit variant	7.2E-02	CAE47300.1	-
22	A; C	23	MEWYPEAAANAWV	CRISP	cysteine-rich venom protein	5.0E-07	BAP39957.1	24547.04
			LVRAECGENLYMSTSMPK			7.0E-05	AMB36337.1	
			260.09-WYPEAAANSLR			3.0E-03	XP_015678374.1	
			SSLSNHLVKVDLHNSLR			6.0E-02	CE73575.1	
23a	C	21	YGCYCGWGGQGR	PLA₂	phospholipase A2 I acidic chain	3.0E-08	A60512	-
			RGLCLGENVNTYDK		Vaspin acidic subunit (1)	4.0E-06	CAE47105.1	
			239.13-FQFGDMILQK		Vaspin acidic subunit (1)	1.0E-05		
23b	A; C	13	241.16-LSSFGENMNTYDK	PLA₂	Ammodytin I1(C) variant	1.0E-05	CAE47172.1	13624.69
			262.15-SQFGDMINK			3.0E-04		
			362.17-NGDIVCDNDAKER		Vaspin acidic subunit (1) variant	2.1E-02	CAE47105.1	

Appx Table 4.2. Venom proteins and peptides identified from *Vipera ammodytes montandoni*. *continued*

Peak Number	Method	SDS PAGE M_{av} [kDa]	Identification sequence tag	Protein species	Protein ID	BLAST E-value	NCBI accession number	intact mass *m/z* native average
24a	C	21	GRLCLGENVNTYDK	PLA₂	acidic phospholipase A2 inhibitor chain HPD-1	4.0E-06	A4VBF0.1	
24b	A; C	13	LGAICFGENMNTYSR	PLA₂	Ammodytin I1(A) variant	2.0E-07	CAE47141.1	13676.81
			VAAMRFGENMNTYDK		Ammodytin I1(C) variant	7.0E-07	CAE47172.1	
			262.15-SQFGDMINK			3.0E-04		
			201.99-TFVHDCCYGR		Vaspin basic subunit variant	5.0E-05	CAE47300.1	
			362.15-NGDIVCGGDD-526.27		phospholipase A2	9.0E-05	AHJ09559.1	
25a	C	37	541.21-GTLLNQEWVLTAAR	svSP	venom serine proteinase-like protein 2	1.0E-07	Q9PT40.1	
			TLCAGLLQGGLDSSSK			3.0E-04		
			LMGWGTLTTTK			6.0E-04		
25b	C	23	-	-	unknown (protein)	-	-	
25c	C	15	KAICFGENMNTYSR	PLA₂	Ammodytin I1(A) variant	2.0E-07	CAE47141.1	
26a	C	>200	-	-	unknown (protein)	-	-	
26b	C	60	-	-	unknown (protein)	-	-	
26c	A; C	37	LVNDECNINEHPFLAFVPPHQ	svSP	Nikobin	1.0E-09	E5AJX2.1	32026.88, 32899.08
			409.24-QPGLYTDIFDY-629.33			9.0E-07		
			VVCAGIWQGGK			5.0E-06		
			WKPPVVGSVCR			2.0E-03		
26d	C	15	-	-	unknown (protein)	-	-	
27	A; C	37	409.24-KPGLYTDIFDYS-629.32	svSP	Nikobin	2.0E-06	E5AJX2.1	32686.16, 35124.93
			WKPPVVGSVCR			2.0E-03		
			LMGWGTISSTK		Cadam10_SVSP-11	6.0E-05	JAV48393.1	
28a	A; C	37	PDNHLFAYNEHPFLAFVTSDR	svSP	Nikobin	1.0E-07	E5AJX2.1	32686.34, 33342.02
			VVCAGIWQGGK			5.0E-06		
28b	C	25	MEWYPEAAANPMK	CRISP	cysteine-rich venom protein	7.0E-06	XP_015678374.1	
			838.43-LLDLHNSLR			1.6E-01		
28c	C	15	-	-	unknown (protein)	-	-	

Appx Table 4.2. Venom proteins and peptides identified from *Vipera ammodytes montandoni*. *continued*

Peak Number	Method	SDS PAGE M_{av} [kDa]	Identification sequence tag	Protein species	Protein ID	BLAST E-value	NCBI accession number	intact mass *m/z* native average
29a	A; C	60	LLTAIDFQRTLGK	svMP	metalloproteinase type III	8.0E-05	JAS05425.1	24547.96, 27654.98
			997.58-LTAIDFNGLTKK			3.0E-04	ADI47577.1	
			AYIGTMCQPK			4.0E-04	ADW54356.1	
29b	C	37	-	-	unknown (protein)	-	-	
29c	C	25	1098.52-TLNNDIMLLK	other	trypsin-like	7.0E-04	XP_015672094.1	
29d	C	15	1098.52-TLNNDIMLLK	other	trypsin-like	7.0E-04	XP_015672094.1	
30a	C	55	326.20-LNEMYLPLNLR	svMP	metalloproteinase	8.0E-05	ADI47590.1	
30b	C	37	VVCAGIWQGGK	svSP	Nikobin	5.0E-06	E5AJX2.1	
30c	C	30	AAYPWLLER	svSP	serine proteinase SP-3	1.0E-04	AMB36344.1	
30d	A; C	13	287.13-DDAEMFCR	CTL	C-type lectin	5.0E-03	Q6T7B7.1	13890.25
			204.07-SWEWTDGSSTK			1.3E-02		
			331.12-TEFAEYLADYLK		C-type lectin mannose-binding isoform-like	3.6E-02	XP_015686729.1	
31a	C	55	-	-	unknown (protein)	-	-	
31b	C	37	LYDYSVCR	svSP	serine protease	1.0E-02	ADI47570.1	
31c	C	30	AAYPWLLER	svSP	serine proteinase SP-3	1.0E-04	AMB36344.1	
32a	C	37	-	-	unknown (protein)	-	-	
32b	C	30	AAYPWLLER	svSP	serine proteinase SP-3	1.0E-04	AMB36344.1	
33a	C	55	HDDIFAYEK	LAAO	L-amino-acid oxidase	1.0E-04	Q6WP39.1	
			VTVLEASER			3.0E-03	JAS05318.1	
33b	C	50	HDDIFAYEK	LAAO	L-amino-acid oxidase	1.0E-04	Q6WP39.1	
33c	C	37	-	-	unknown (protein)	-	-	
33d	C	30	AAYPWLLER	svSP	serine proteinase SP-3	1.0E-04	AMB36344.1	
33e	C	20	WTDGSSVIYK	CTL	C-type lectin-like protein 3A	3.0E-05	AJO70726.1	
			537.22-VWLGLWELR			7.4E-02		
33f	C	11	EKNEGINCFVFEIAK	CTL	snaclec-8	2.0E-08	APB93445.1	

Appx Table 4.2. Venom proteins and peptides identified from *Vipera ammodytes montandoni*. *continued*

Peak Number	Method	SDS PAGE M_{av} [kDa]	Identification sequence tag	Protein species	Protein ID	BLAST E-value	NCBI accession number	intact mass *m/z* native average
34a	C	50	EDDYEEFLEIAK	LAAO	L-amino-acid oxidase	1.0E-07	G8XQX1.1	
			213.12-EEGWYANLGNNR			3.0E-04		
			HDDIFAYEK			1.0E-04	Q6WP39.1	
			523.22-FSEALTAPEGR			3.0E-04	P0DI84.1	
			158.07-GQLYEESLK		BATXLAAO1	1.0E-03	JAV01888.1	
34b	C	37	HDDIFAYEK	LAAO	L-amino acid oxidase 1b	1.0E-04	Q6WP39.1	
35a	C	50	VTVLEASER	LAAO	L-amino acid oxidase 1b	3.0E-03	JAS05318.1	
			523.22-FSEALTAPWR			4.7E-02		
35b	C	37	-	-	unknown (protein)	-	-	
35c	C	30	AAYPWLLER	svSP	serine proteinase SP-3	1.0E-04	AMB36344.1	
35d	C	20	537.21-VWLGLWELR	CTL	C-type lectin galatrox	8.1E-02	P0DM53.1	
35e	C	11	371.17-EGINCFVFEIAK	CTL	snaclec-8	2.0E-07	APB93445.1	
36a	C	70	VEDYDQIGASLR	other	xaa-Pro aminopeptidase 2	6.0E-06	XP_015676063.1	
			414.22-FMGSTWQEK			7.0E-05		
			LEDVALVVPAK			2.0E-03		
36b	C	50	SSVGLIQDYCK	svMP	metalloproteinase H4-A	7.0E-06	AHB62069.1	
			HDDIFAYEK	LAAO	L-amino-acid oxidase	1.0E-04	Q6WP39.1	
			VTVLEASER			3.0E-03	JAS04872.1	
			523.22-FSEALTAPEGR			3.0E-04	P0DI84.1	
36c	C	30	-	-	unknown (protein)	-	-	
37a	C	50	SSVGLIQDYCK	svMP	metalloproteinase H4-A	7.0E-06	AHB62069.1	
			HDDIFAYEK	LAAO	L-amino-acid oxidase	1.0E-04	Q6WP39.1	
			VTVLEASER			3.0E-03	JAS04872.1	
37b	C	30	NPQCILNQLPR	svMP	metalloproteinase	4.0E-03	ADI47711.1	
			1199.58-LTLELMLLK	other	ATP synthase F0 subunit 8	7.5E-02	AMW93172.1	

Appx Table 4.2. Venom proteins and peptides identified from *Vipera ammodytes montandoni*. *continued*

Peak Number	Method	SDS PAGE M$_{av}$ [kDa]	Identification sequence tag	Protein species	Protein ID	BLAST E-value	NCBI accession number	intact mass *m/z* native average
38a	C	50	247.16-LMSGTLSCEASIR	svMP	metalloproteinase precursor 1	1.0E-06	AGL45259.1	
			NPCQIYYTPR			3.0E-06		
38b	C	37	-	-	unknown (protein)	-	-	
38c	C	30	-	-	unknown (protein)	-	-	
X	A	-	-	-	unknown (protein)	-	-	13890.25, 27654.95, 27801.99
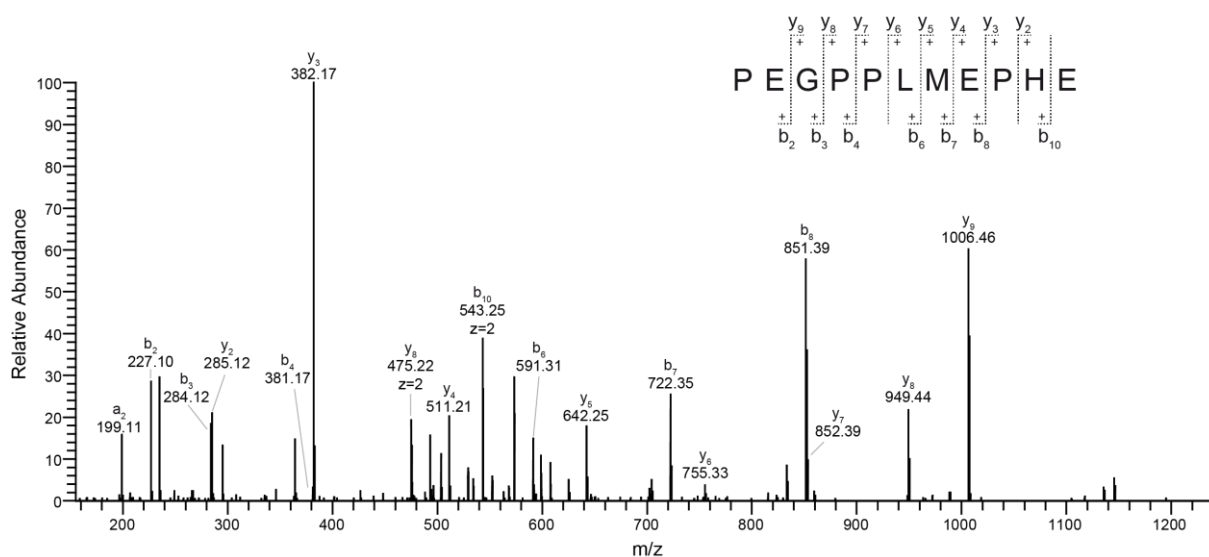
Appx Figure 4.1. Tripeptide precursor fragment of *V. a. transcaucasiana* peak 9. The annotated MS1 spectrum of Vat by intact mass profiling peak 9 at RT 1.74 min shows the mass of the tripeptide precursor fragment PEGPPLMEPHE.
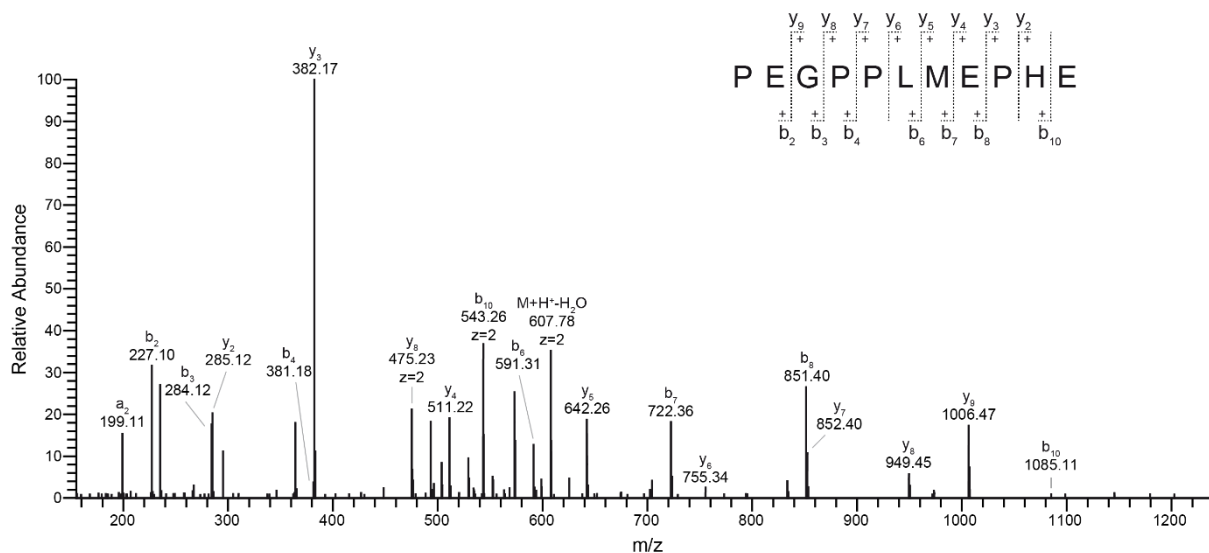
Appx Figure 4.2. Tripeptide precursor fragment of *V. a. montandoni* peak 8. The annotated MS1 spectrum of Vam by intact mass profiling peak 8 at RT 1.90 min shows the mass of the tripeptide precursor fragment PEGPPLMEPHE.
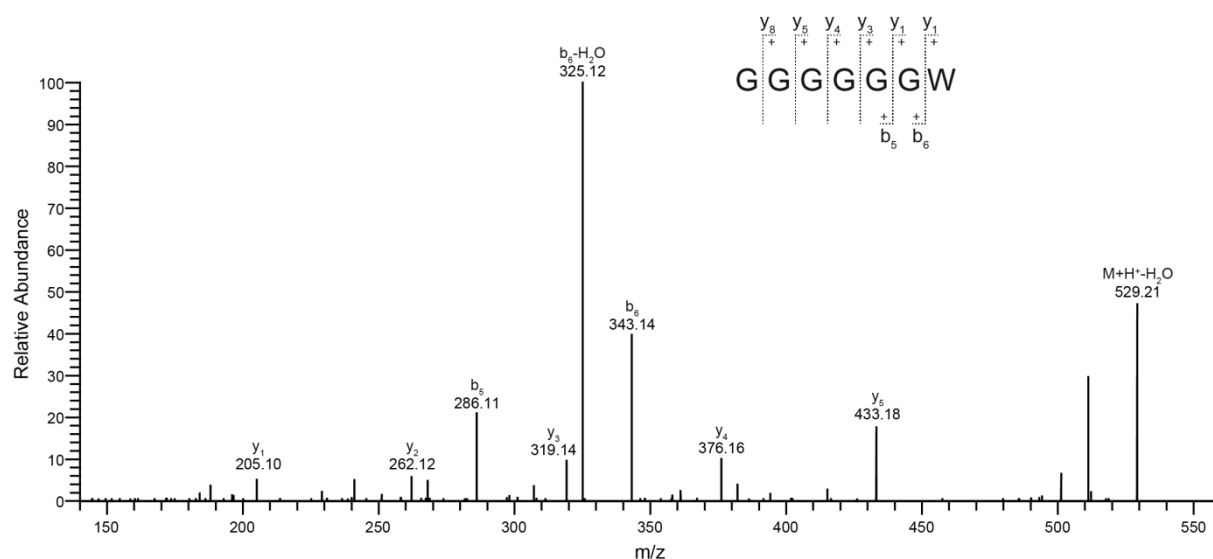
Appx Figure 4.3. Tripeptide precursor fragment of *V. a. transcaucasiana* peak 4. The annotated MS1 spectrum of Vat by intact mass profiling peak 4 at RT 2.28 min shows the mass of the tripeptide precursor fragment GGGGGGW.
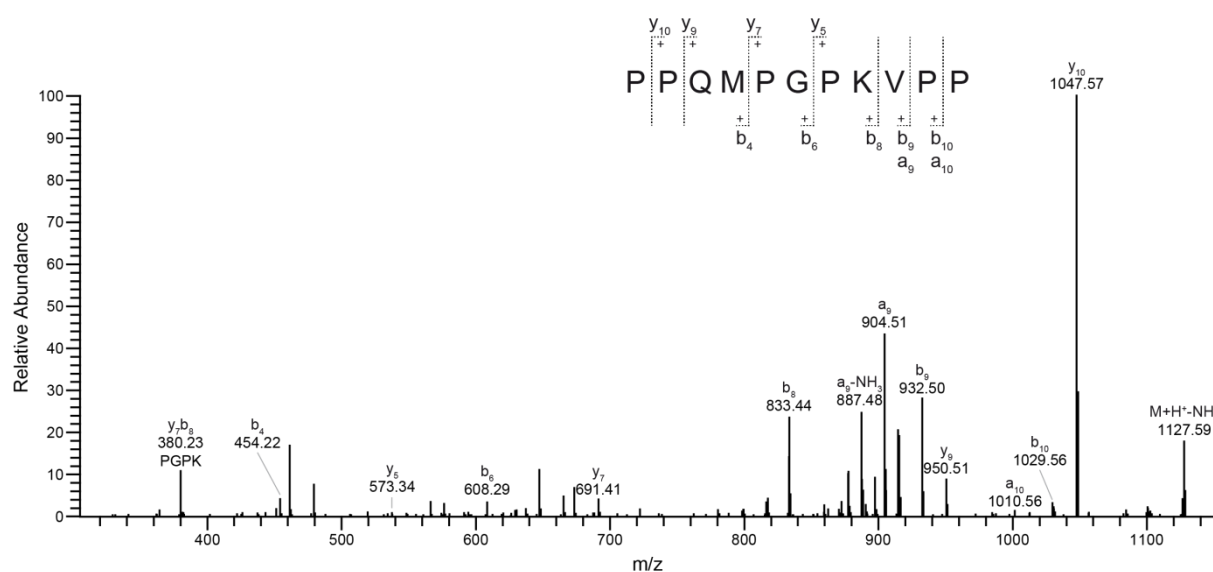
Appx Figure 4.4. Tripeptide precursor fragment of *V. a. transcaucasiana* peak 14. The annotated MS1 spectrum of Vat by intact mass profiling peak 14 at RT 1.96 min shows the mass of the tripeptide precursor fragment PPQMPGPKVPP.

Appx Figure 4.5. Tripeptide precursor fragment of *V. a. montandoni* peak 2. The annotated MS1 spectrum of Vam by intact mass profiling peak 2 at RT 1.57 min shows the mass of the tripeptide precursor fragment DNEPPKKVPPN.

Vat Peak 17 - TIC RT 58.89 min
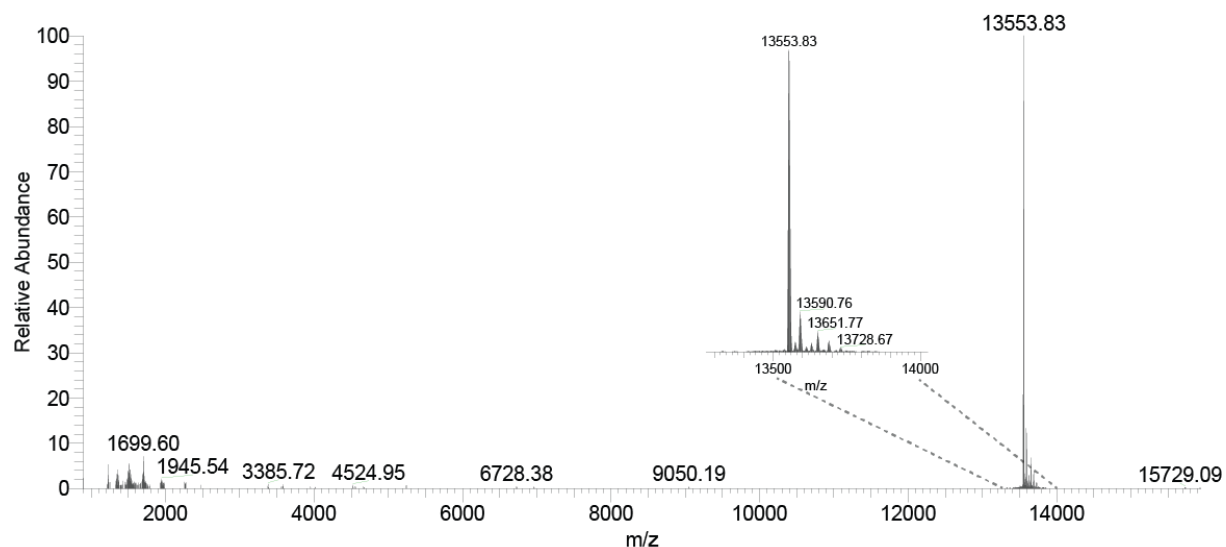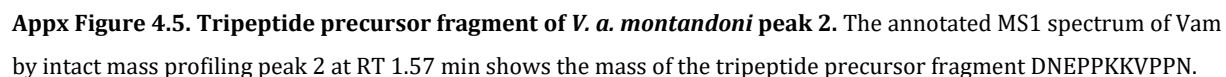
Appx Figure 4.6. Deconvoluted masses of *V. a. transcaucasiana* peak 17. The deconvoluted MS1 spectrum of Vat intact mass profiling peak 17 at RT 58.89 min shows the mass of the neutral phospholipase A_2 Ammodytin I2.

Vat Peak 19 - TIC RT 67.90 min
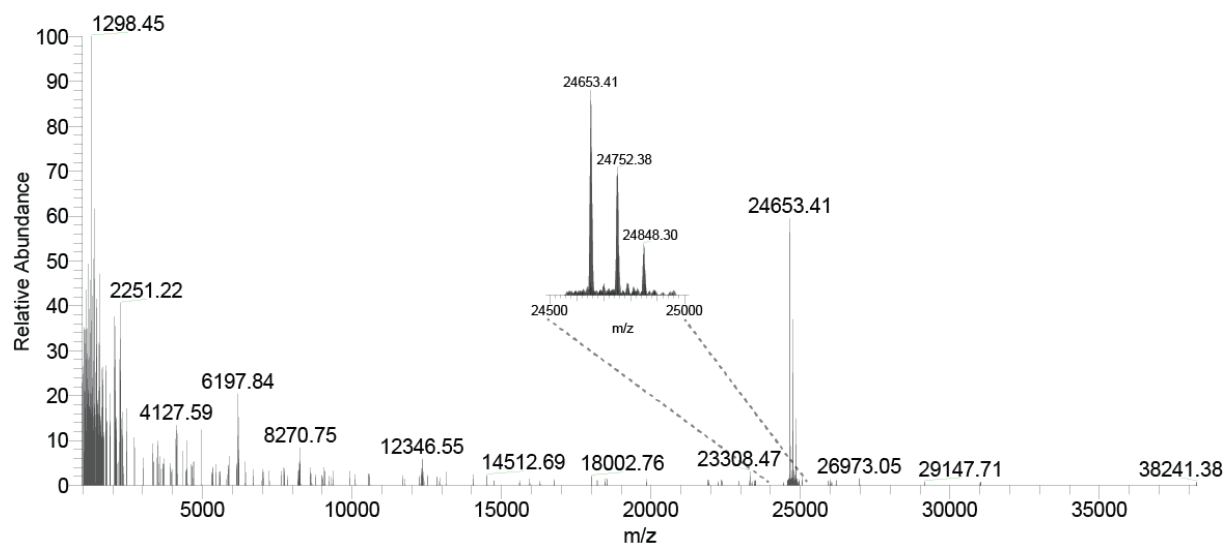
Appx Figure 4.7. Deconvoluted masses of *V. a. transcaucasiana* peak 19. The deconvoluted MS1 spectrum intact mass profiling peak 19 at RT 67.91 min shows the masses of CRISP's.

Vat Peak 20 - TIC RT 72.22 min
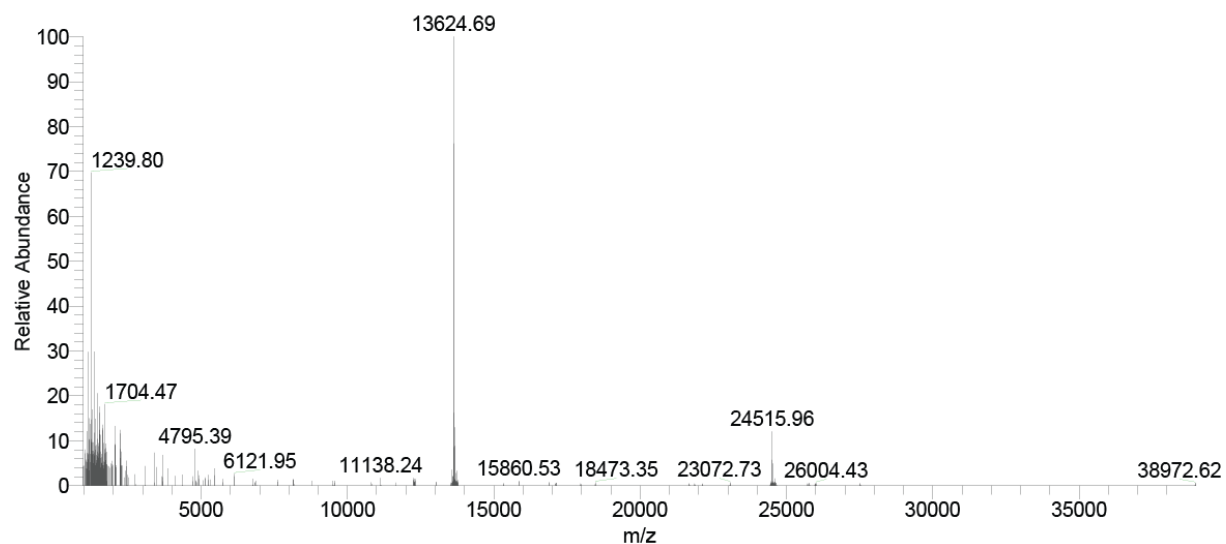
Appx Figure 4.8. Deconvoluted masses of *V. a. transcaucasiana* peak 20. The deconvoluted MS1 spectrum of Vat intact mass profiling peak 20 at RT 72.22 min shows the mass closely related to the acidic phospholipase A_2 homolog Vipoxin A chain.

Vat Peak 20 - TIC RT 75.60 min
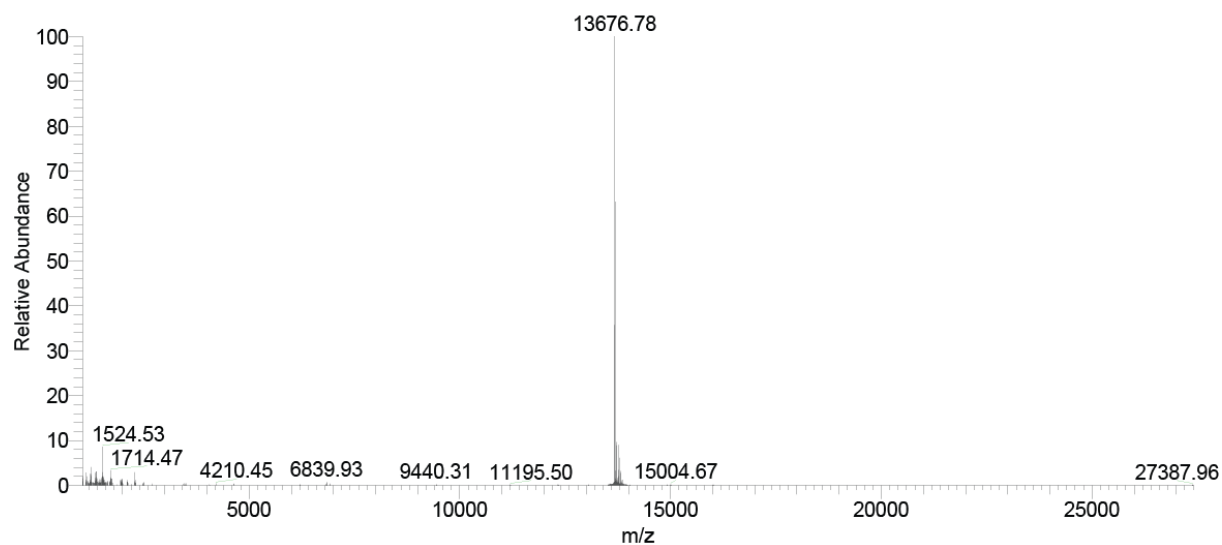
Appx Figure 4.9. Deconvoluted masses of *V. a. transcaucasiana* peak 20. The deconvoluted MS1 spectrum of Vat intact mass profiling peak 20 at RT 75.60 min shows the mass of a PLA_2.

Vam Peak 19 - TIC RT 58.41 min
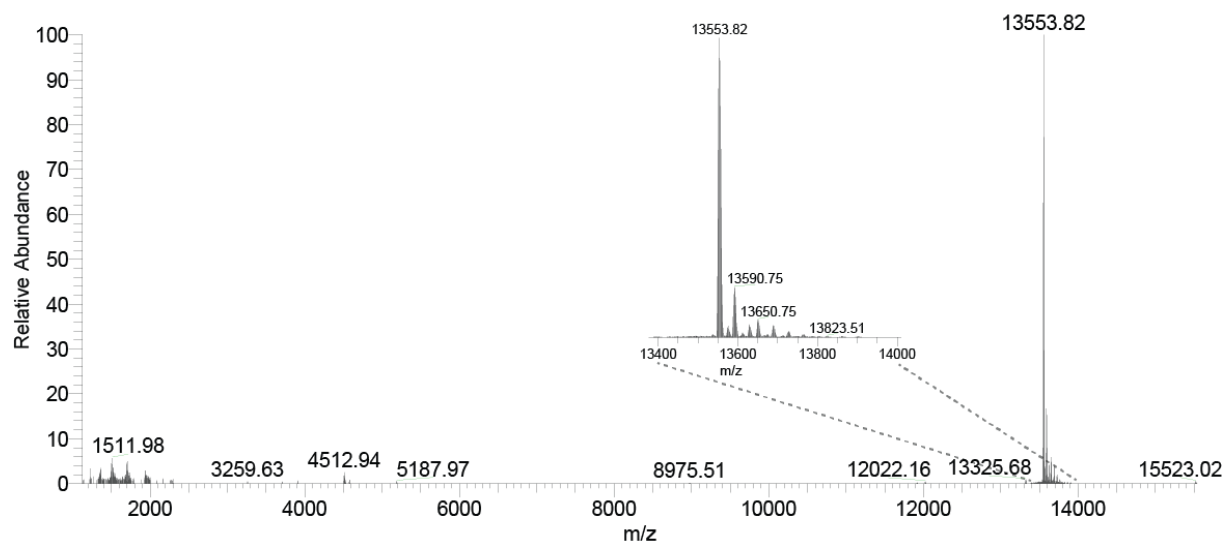
Appx Figure 4.10. Deconvoluted masses of *V. a. montandoni* peak 19. The deconvoluted MS1 spectrum of Vam intact mass profiling peak 19 at RT 58.41 min shows the mass of the neutral phospholipase A_2 Ammodytin I2.

Vam Peak 21 - TIC RT 67.91 min
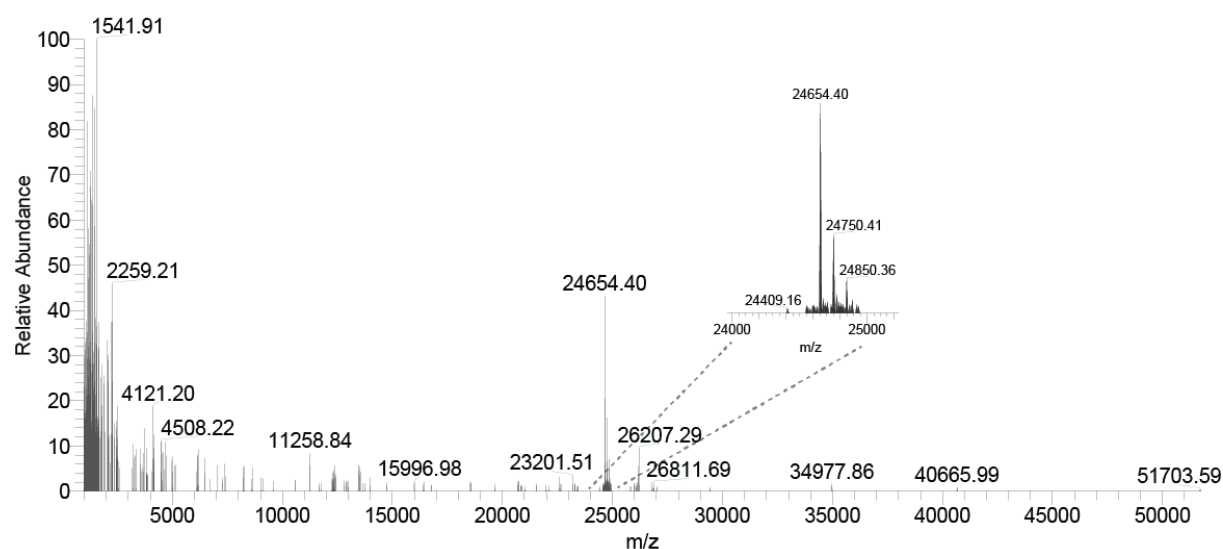
Appx Figure 4.11. Deconvoluted masses of *V. a. montandoni* peak 21. The deconvoluted MS1 spectrum of Vam intact mass profiling peak 21 at RT 67.91 min shows the mass of CRISP's.

Vam Peak 23 - TIC RT 72.99 min
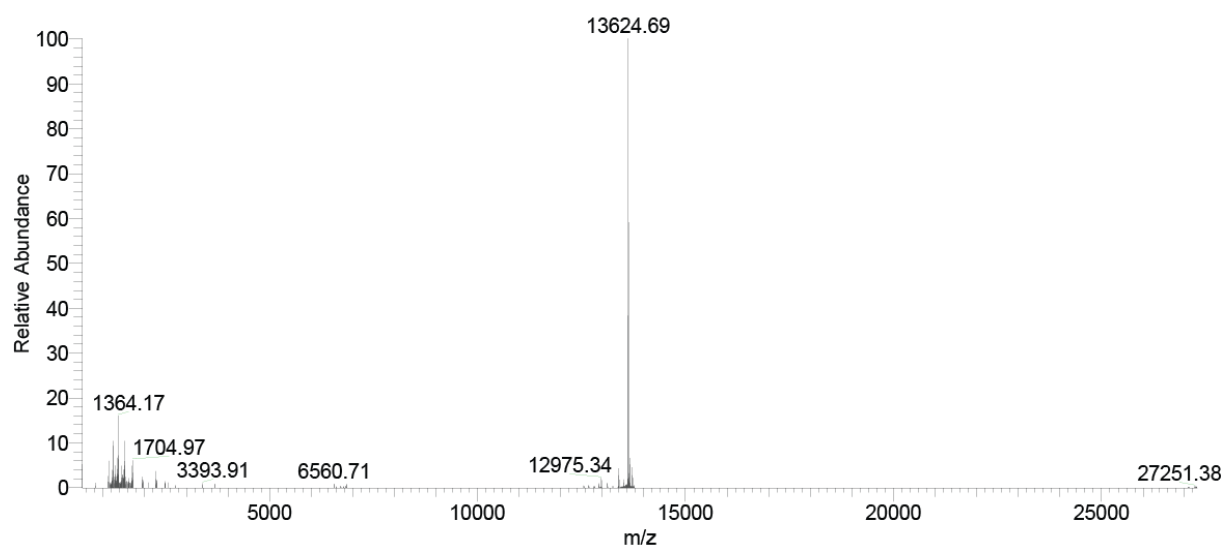
Appx Figure 4.12. Deconvoluted masses of *V. a. montandoni* peak 23. The deconvoluted MS1 spectrum of Vam intact mass profiling peak 23 at RT 72.99 min shows the mass closely related to the acidic phospholipase A_2 homolog Vipoxin A chain.

Vam Peak 24 - TIC RT 75.19 min
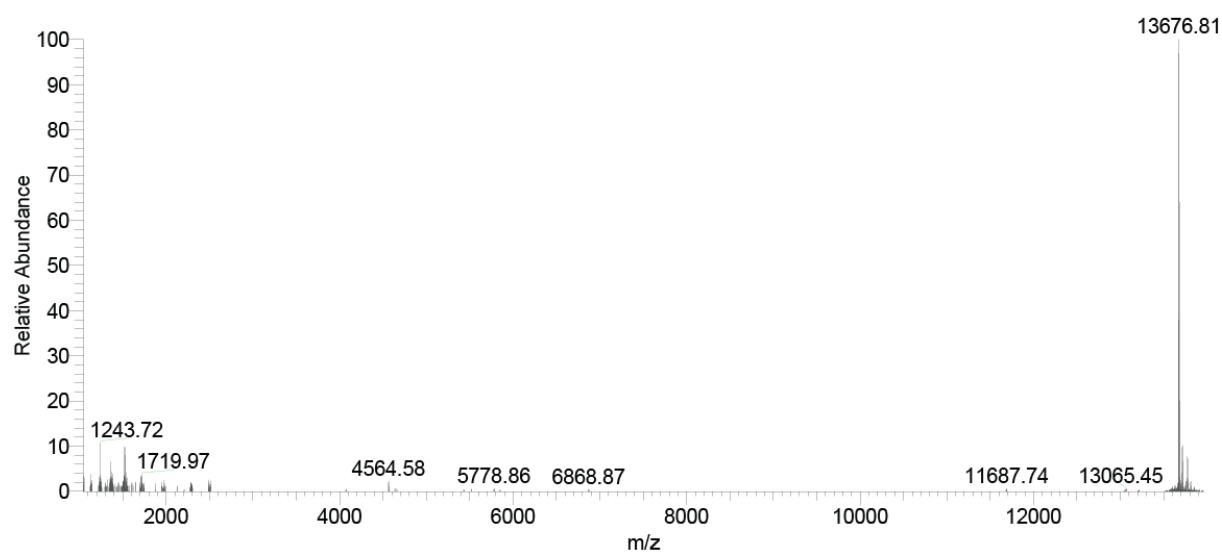
Appx Figure 4.13. Deconvoluted masses of *V. a. montandoni* peak 24. The deconvoluted MS1 spectrum of Vam intact mass profiling peak 24 at RT 75.19 min shows the mass of a PLA$_2$.
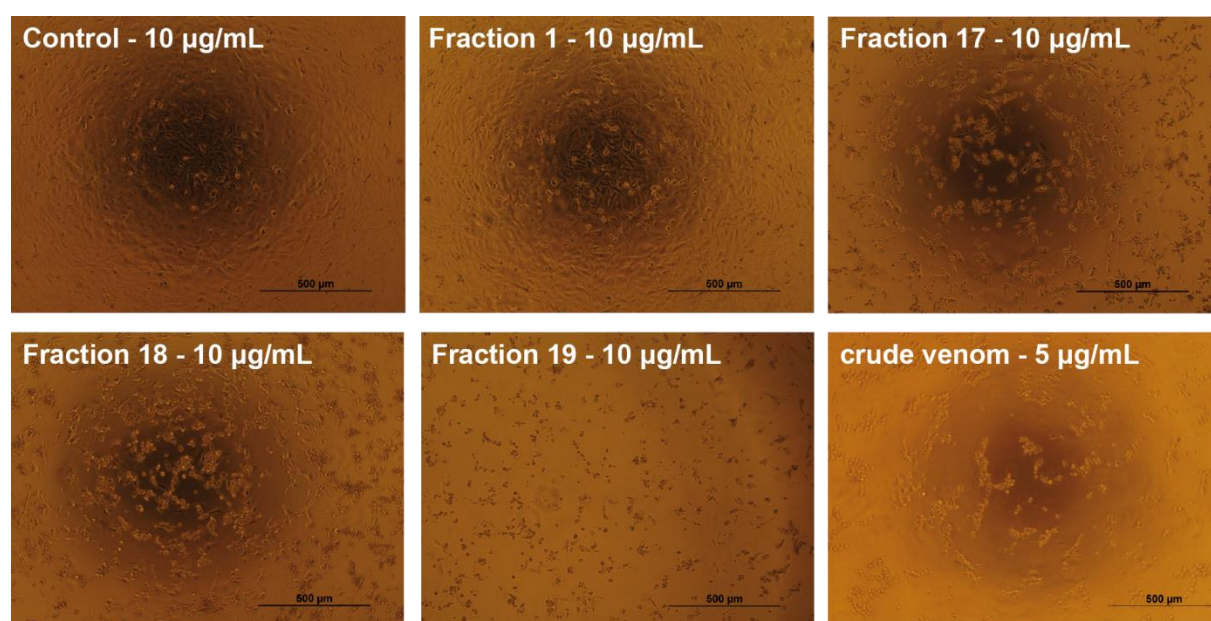
Appx Figure 4.14. Cytotoxicity analysis of fractionated components from *Vipera a. transcaucasiana*. The *in vitro* cytotoxicity testing with crude venoms and fractions was performed according to the protocol of Nalbantsoy and Hempel et al.[234]. Cytotoxicity assay was performed with crude and fractionated venom components for *Vipera a. transcaucasiana*. Cell viability was determined by MTT assay and control was exposed to vehicle only which was taken as 100% viability. We tested all fractions against most sensitive cells MDA-MB-231 according to the screening of crude venom. The fractions with strongest cytotoxic effect on the triple negative MDA-MB-231 breast cancer cells are shown above.

Appendix to Chapter 5

Peak 1
deconvulted by Xtract
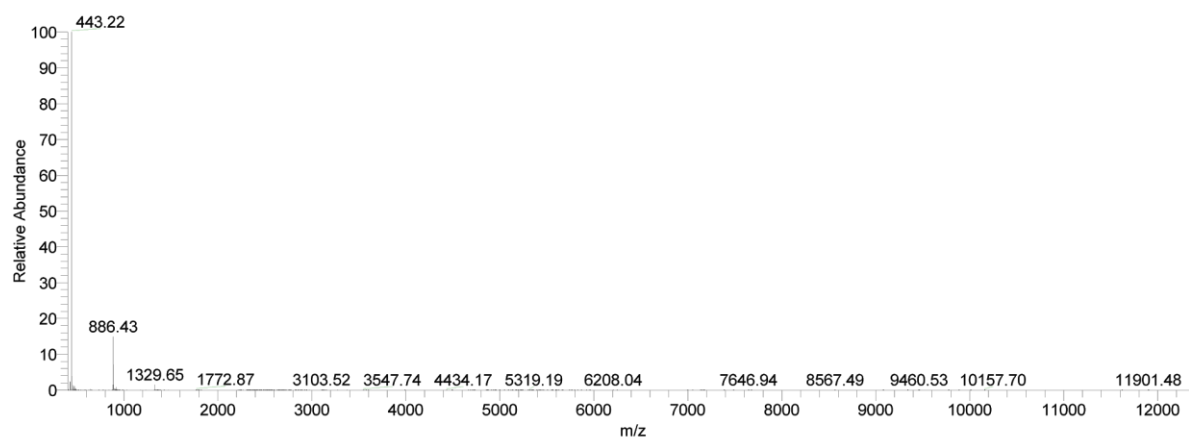
Peak 2
deconvulted by Xtract
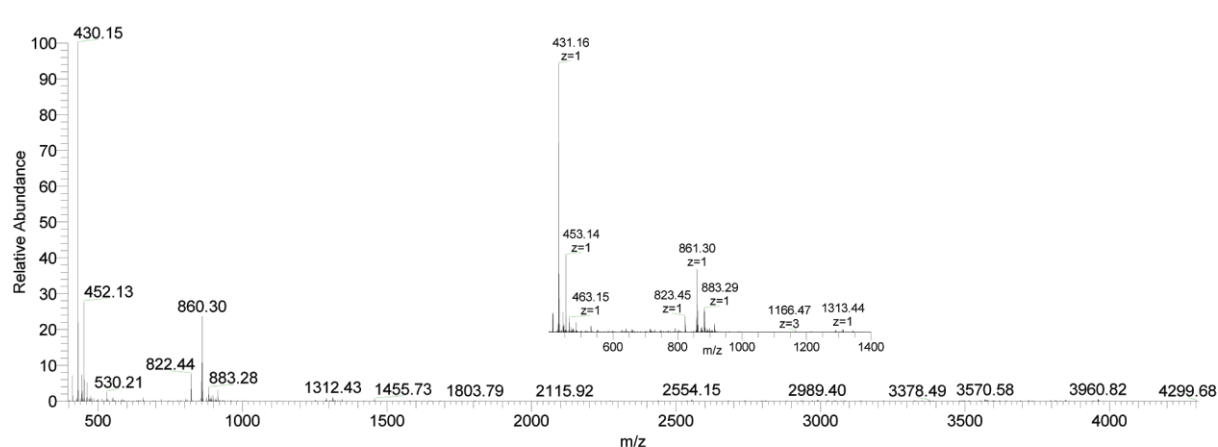
Peak 3
deconvulted by Xtract
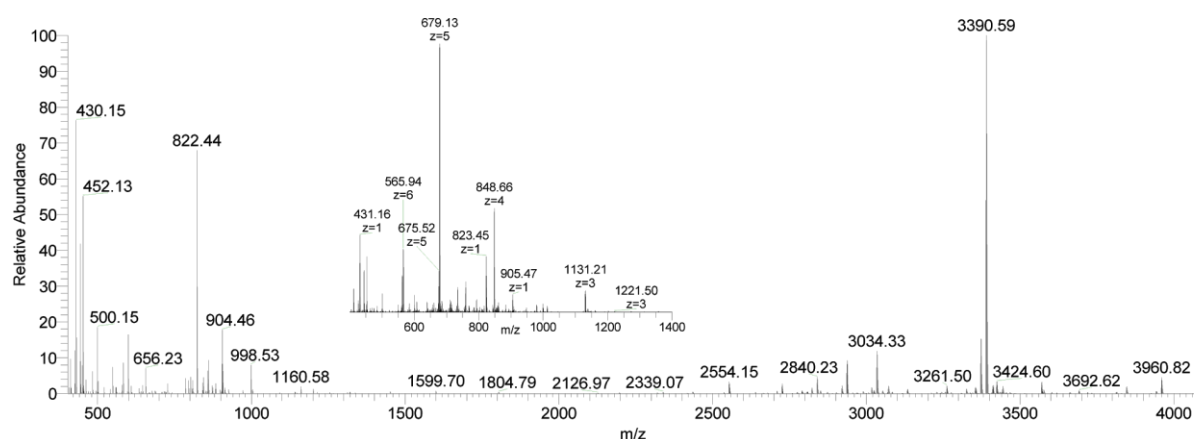
Appx Figure 5.1. Intact mass spectra of pooled *Vipera kaznakovi* venom. Peak nomenclature is based on the chromatogram fractions (**Figure 5.3**). Mass spectra were either isotopically deconvoluted with Xcalibur or charge deconvoluted with magic transformer (MagTran).

Peak 4
deconvulted by Xtract
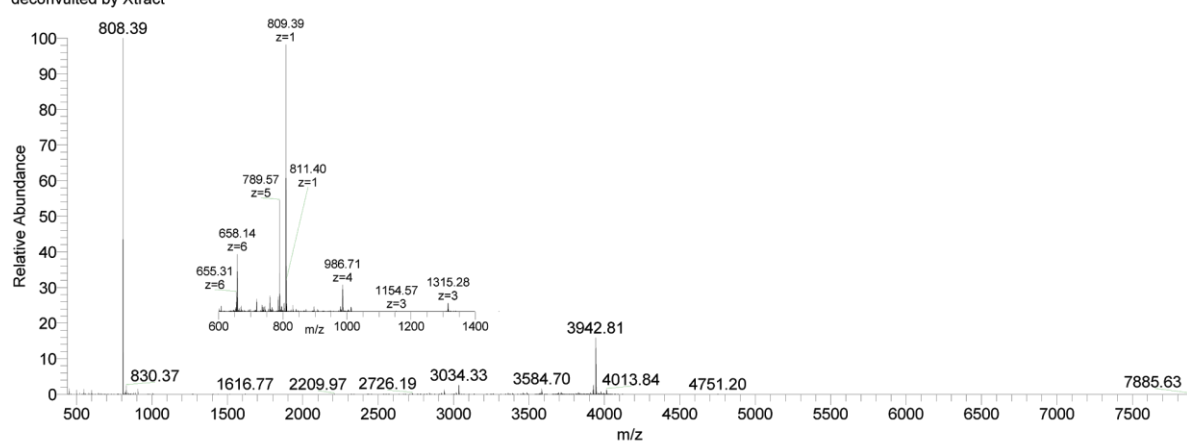
Peak 5
deconvulted by Xtract
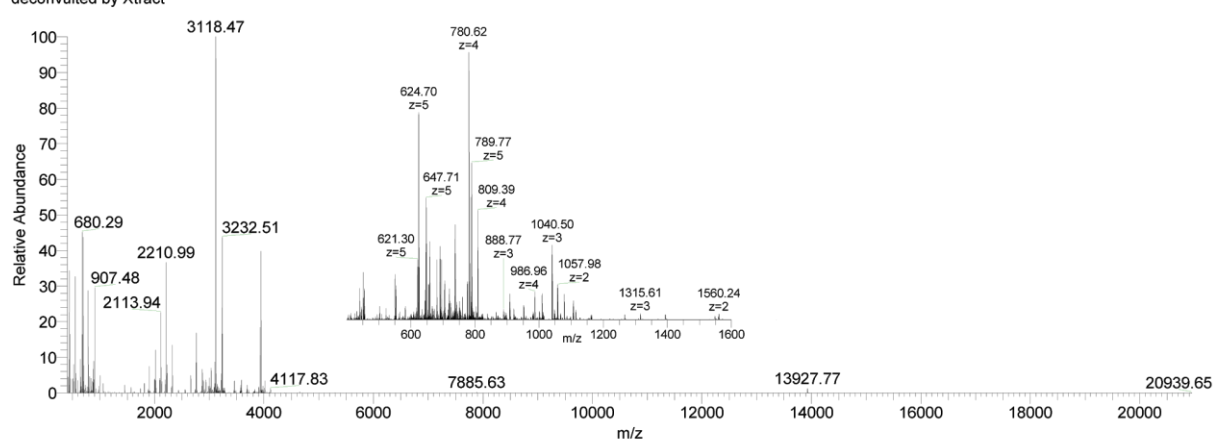
Peak 6
deconvulted by Xtract
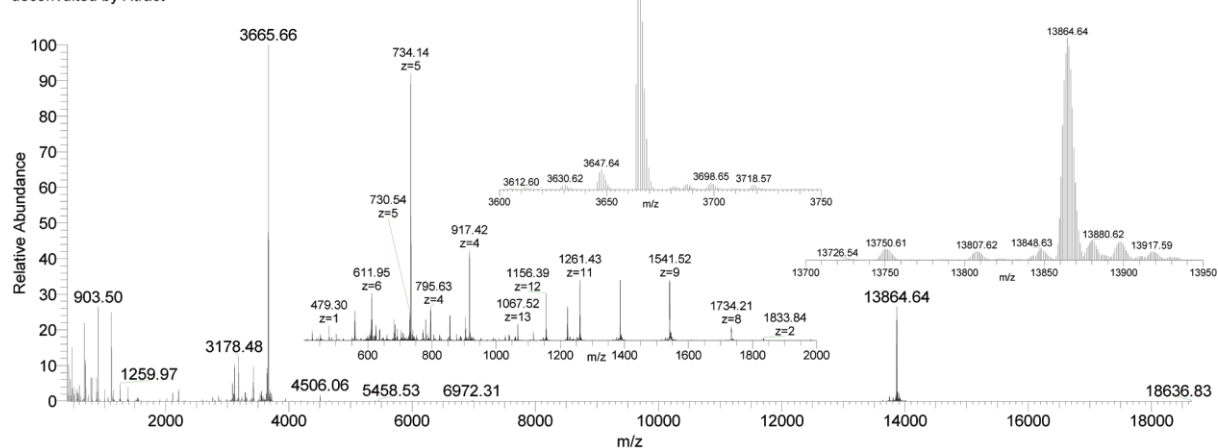
Appx Figure 5.1. Intact mass spectra of pooled *Vipera kaznakovi* venom. *continued*

Peak 7
deconvulted by Xtract
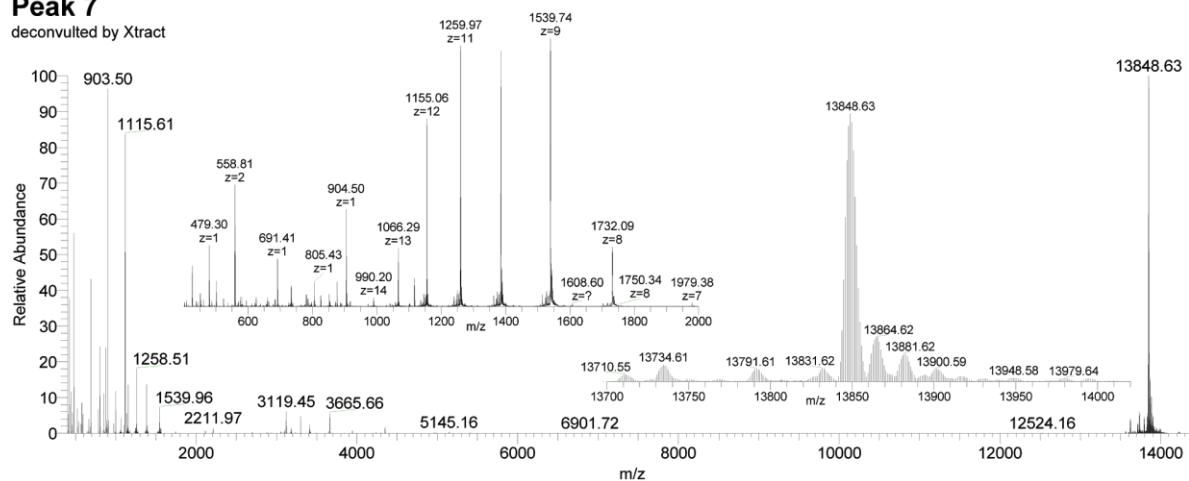
Peak 8
deconvulted by Xtract
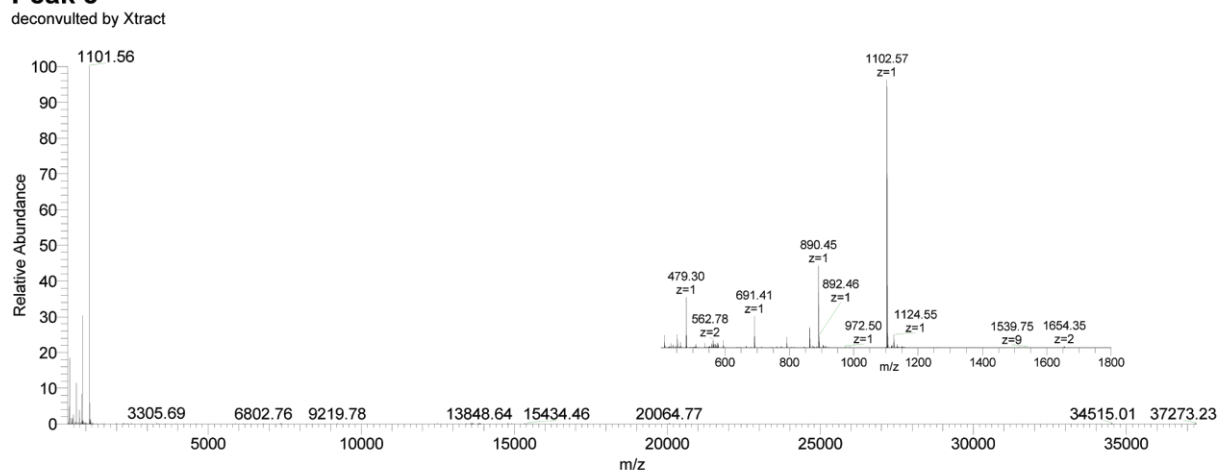
Peak 9
deconvulted by Xtract
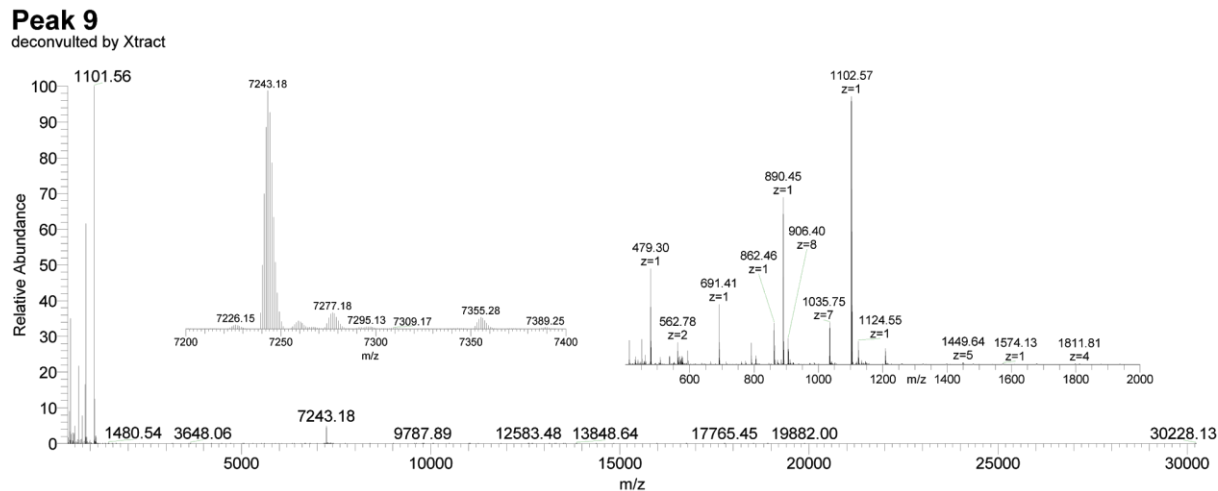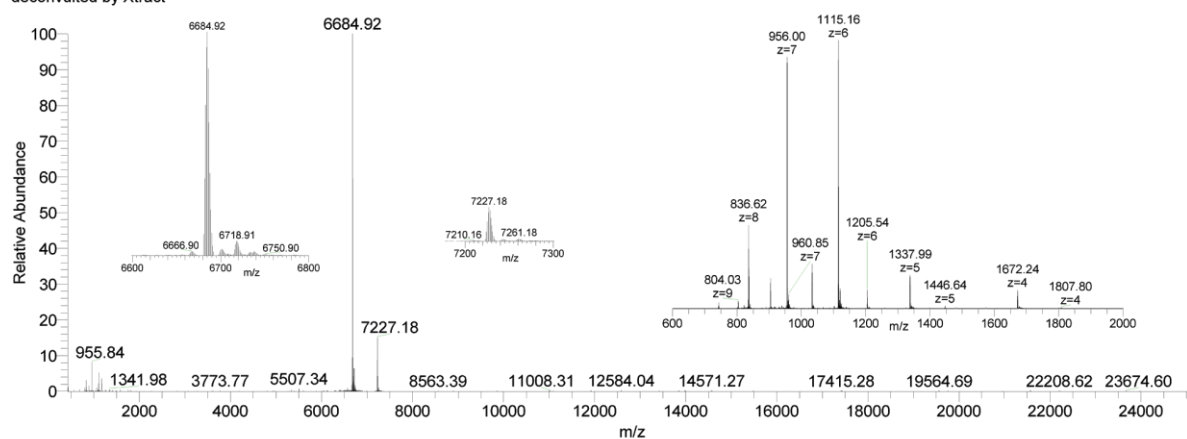
Appx Figure 5.1. Intact mass spectra of pooled *Vipera kaznakovi* venom. *continued*
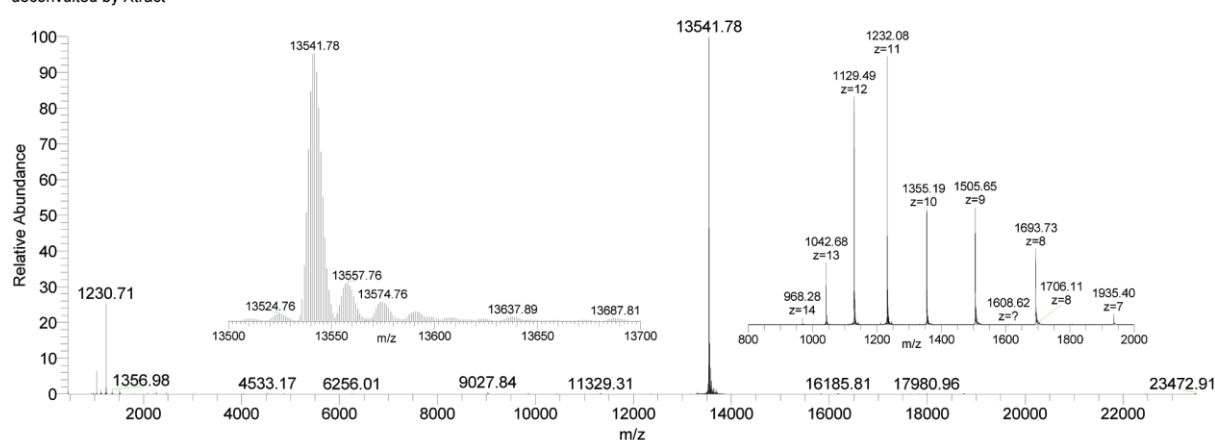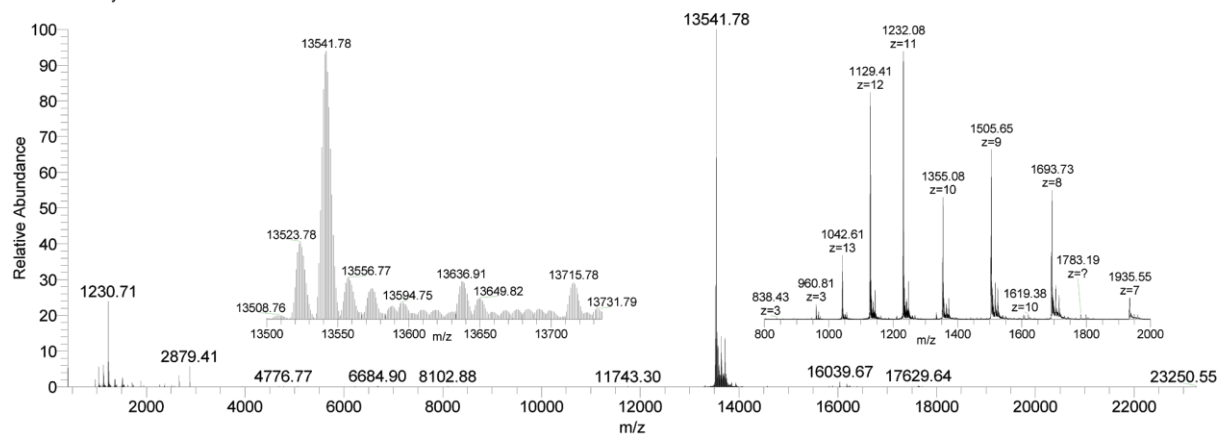
Appx Figure 5.1. Intact mass spectra of pooled *Vipera kaznakovi* venom. *continued*

Peak 14
deconvulted by Xtract
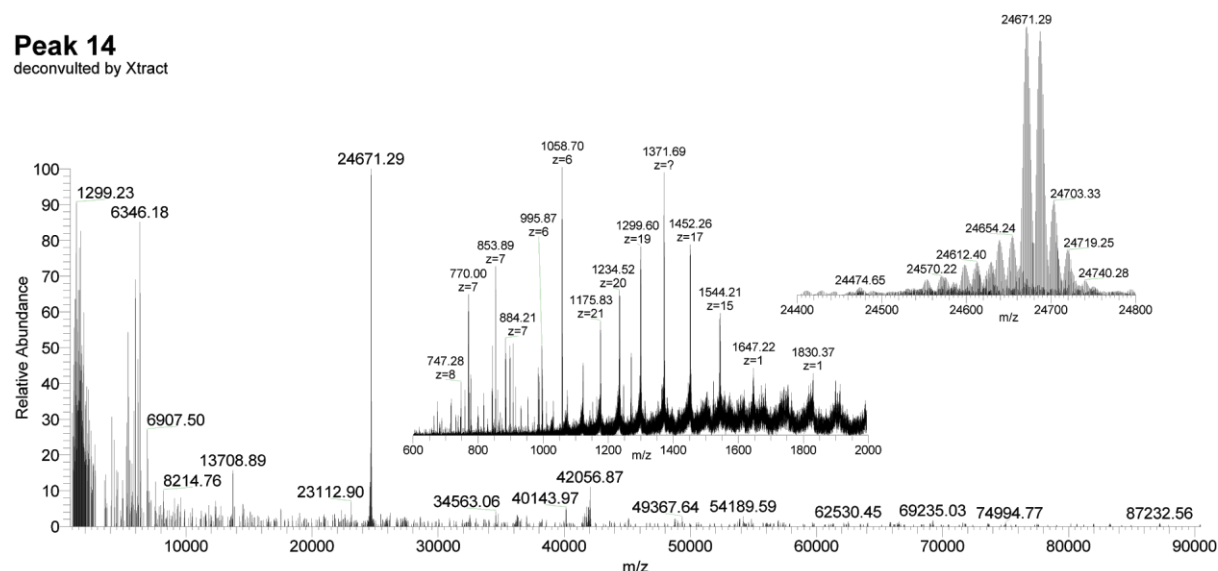
Peak 15
deconvulted by Xtract
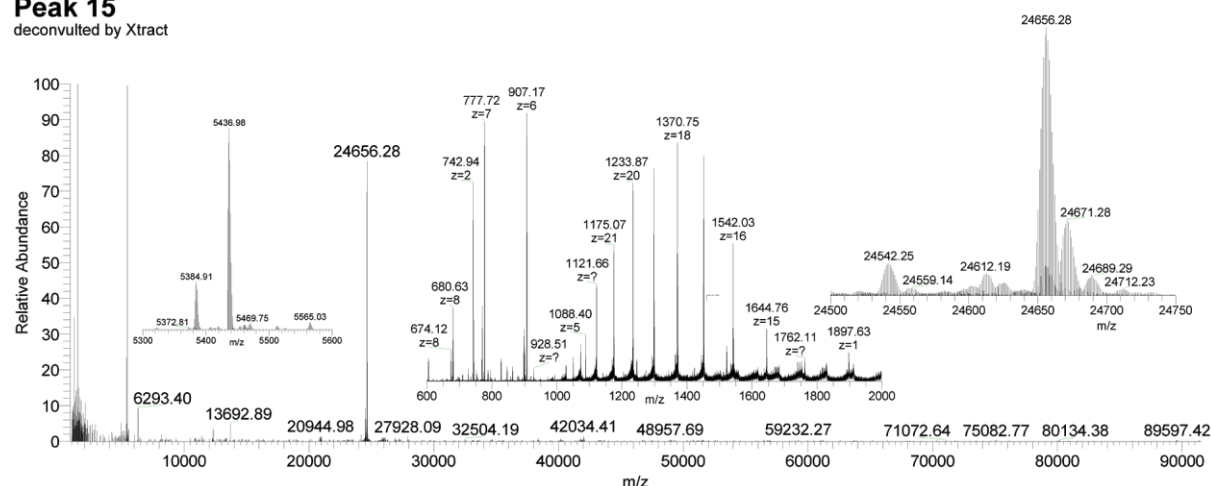
Peak 16
deconvulted by Xtract
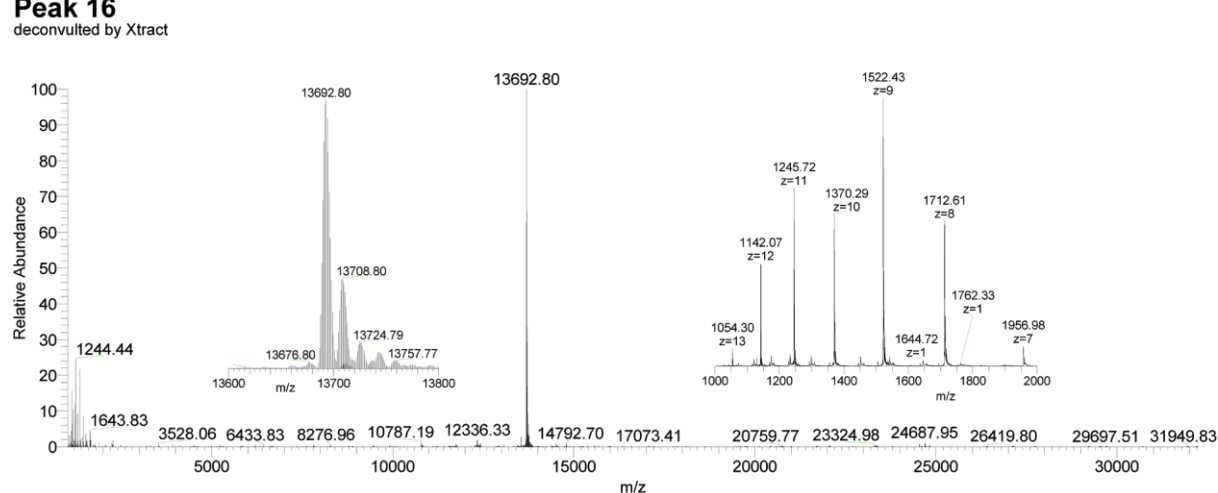
Appx Figure 5.1. Intact mass spectra of pooled *Vipera kaznakovi* venom. *continued*

Peak 17
deconvulted by Xtract
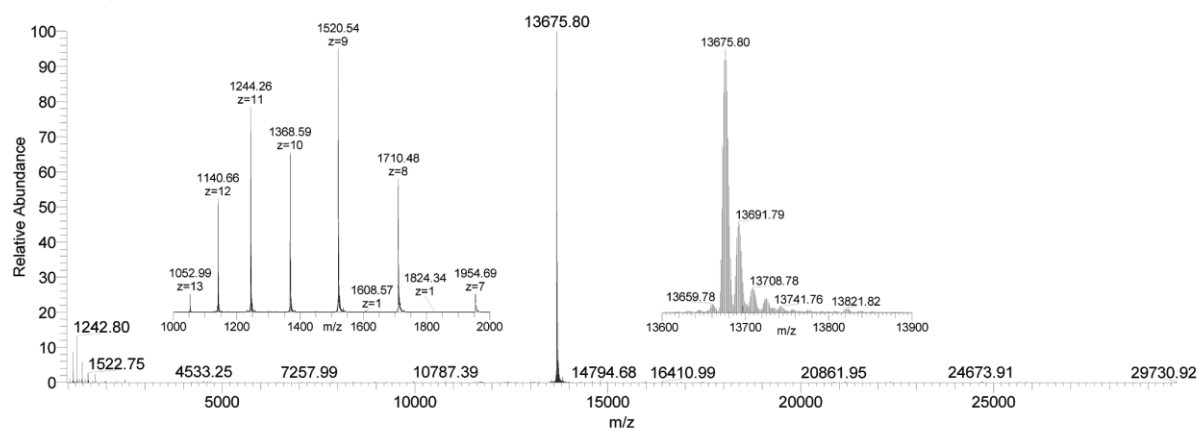
Peak 18
deconvulted by MagTran
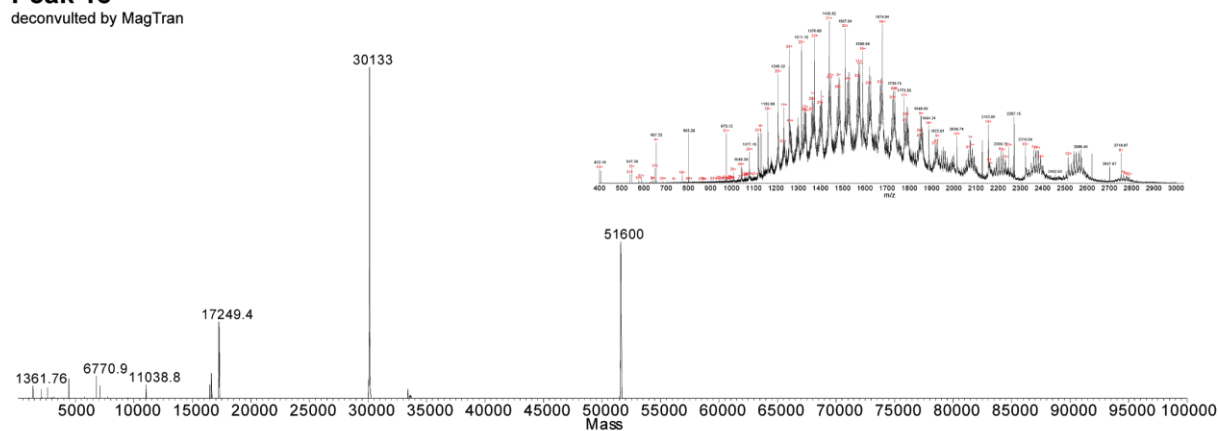
Peak 19
deconvulted by MagTran
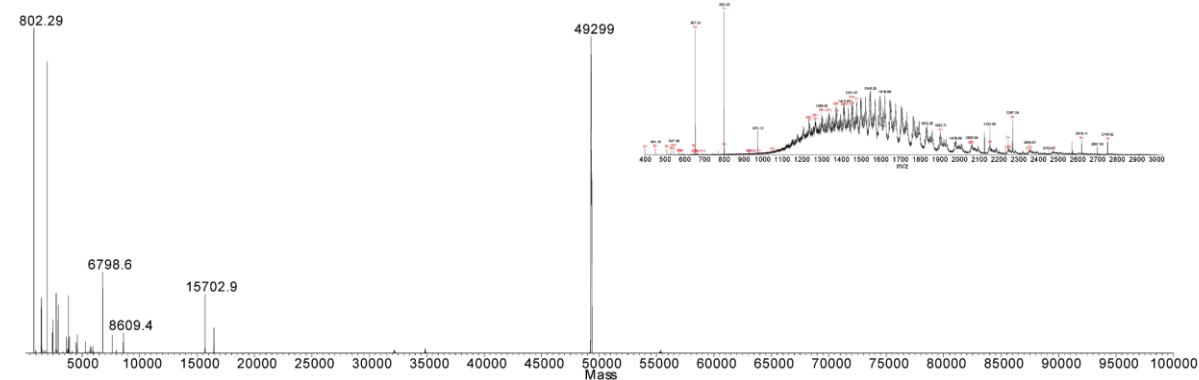
Appx Figure 5.1. Intact mass spectra of pooled *Vipera kaznakovi* venom. *continued*

Peak 20
deconvulted by MagTran
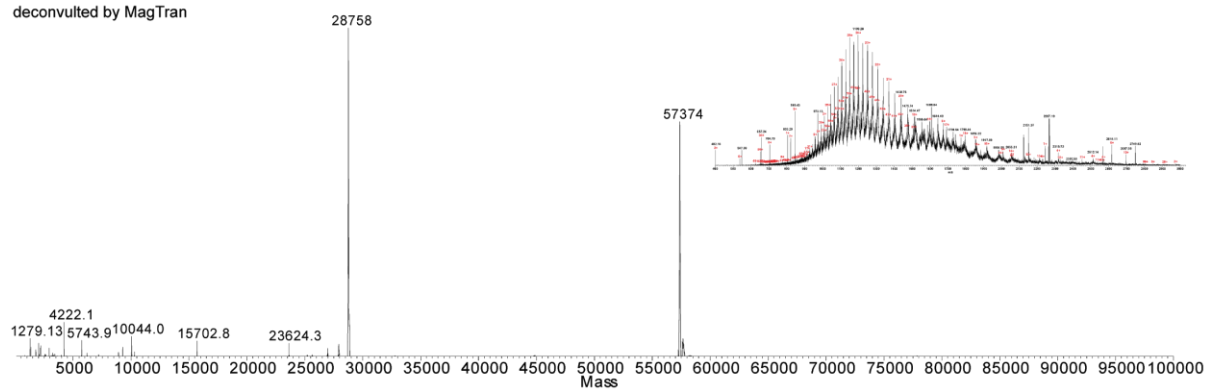
Peak 21
deconvulted by MagTran
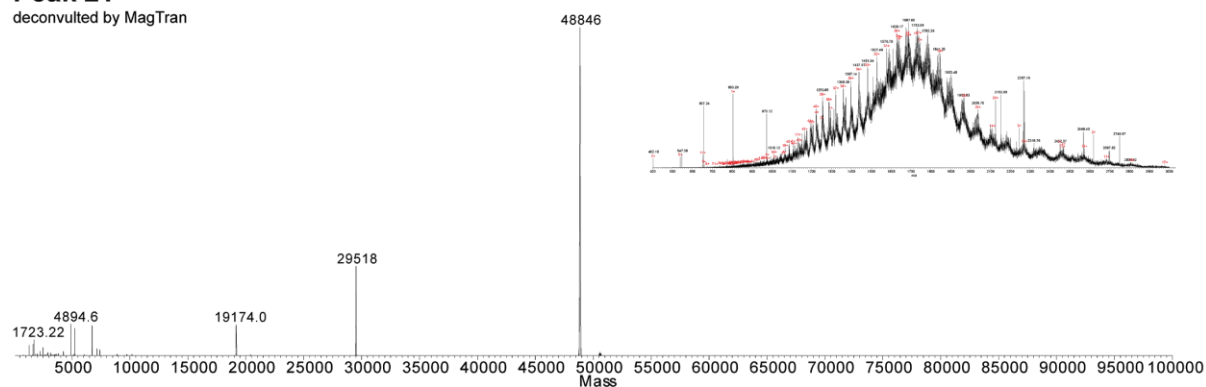
Peak 22
deconvulted by MagTran
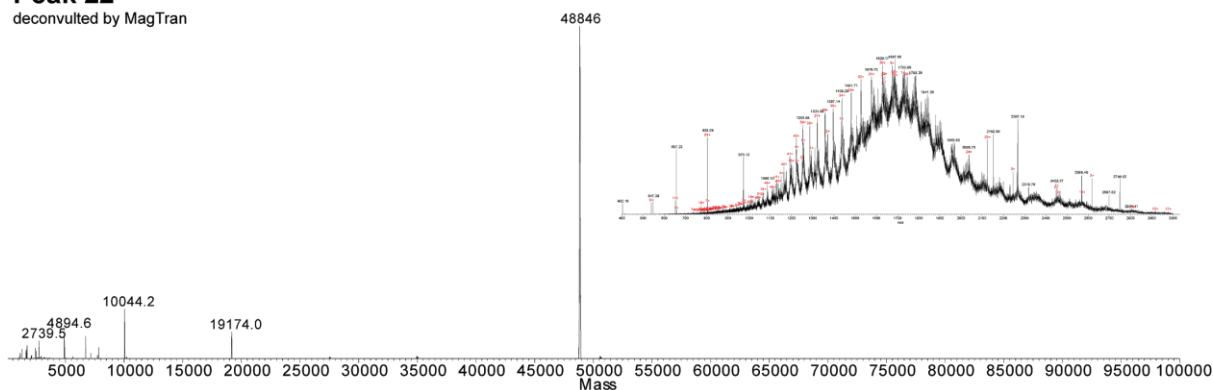
Appx Figure 5.1. Intact mass spectra of pooled *Vipera kaznakovi* venom. *continued*

Peak 23
deconvulted by MagTran
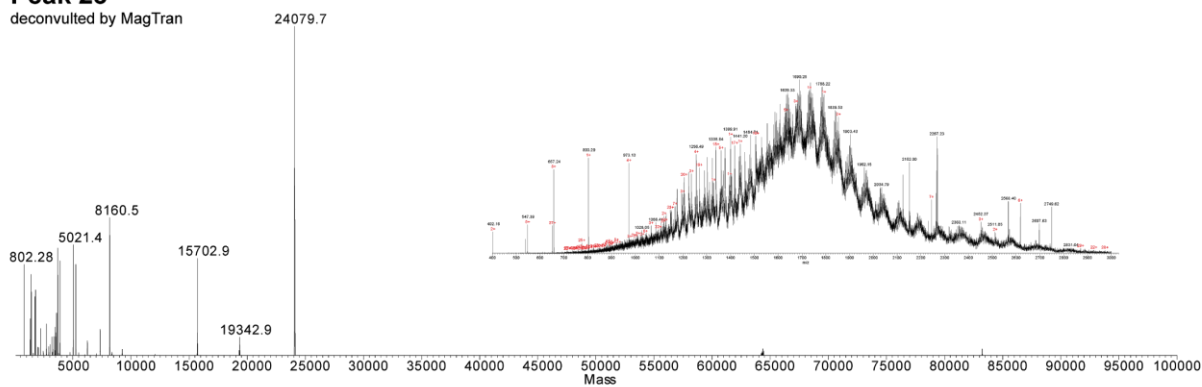
Peak 24
deconvulted by MagTran

Peak 25
deconvulted by MagTran

Appx Figure 5.1. Intact mass spectra of pooled *Vipera kaznakovi* venom. *continued*

Appx Figure 5.2. Vulcano plots of female vs. male individuals and juvenile vs. adult animals. The fold change of proteoform abundance (log2 Ratio) vs. statistical significance (-log10 p-value) is shown. Log2 ratios > 2 or <-2 with -log10 p-values > 1.3 (p-value < 0.05) were considered as significantly differentially expressed proteins.

Appx Table 5.1. Geographic coordinates and sexes for all *V. kaznakovi* individuals. Venom samples of *V. kaznakovi* were collected from 6 adult (2 female, 4 male) and 3 juvenile specimens (unknown sex).

Snake	Number	Family	Subspecies	Sex	Continent	Country	Coordinates
2_Vipera_kaznakovi_juventile _unknown	5	Viperidae	kaznakovi	juvenile	Europe	Turkey	N 41° 25.878' E 41° 27.427'
3_Vipera_kaznakovi_juventile _unknown	6	Viperidae	kaznakovi	juvenile	Europe	Turkey	N 41° 25.938' E 41° 27.413'
5_Vipera_kaznakovi_juventile _unknown	10	Viperidae	kaznakovi	juvenile	Europe	Turkey	N 41° 27.490 E 41° 54.0417'
4_Vipera_kaznakovi_adult_fe male	9	Viperidae	kaznakovi	female	Europe	Turkey	N 41° 27.677' E 41° 54.091'
8_Vipera_kaznakovi_subadult _female	14	Viperidae	kaznakovi	female	Europe	Turkey	N 41° 25.135' E 41° 26.79'
1_Vipera_kaznakovi_adult_ma le	3	Viperidae	kaznakovi	male	Europe	Turkey	N 41° 25.907' E 41° 27.450'
6_Vipera_kaznakovi_adult_ma le	12	Viperidae	kaznakovi	male	Europe	Turkey	N 41° 25.128' E 41° 26.587'
7_Vipera_kaznakovi_subadult _male	13	Viperidae	kaznakovi	male	Europe	Turkey	N 41° 25.300' E 41° 27.50'
9_Vipera_kaznakovi_adult_ma le	15	Viperidae	kaznakovi	male	Europe	Turkey	N 41° 18.76' E 41° 26.381'

Appx Table 5.1. Acute LD$_{50}$ value of *V. kaznakovi* crude venom. Determination of LD$_{50}$s in mice following 24 h exposure by *intraperitoneal* injection.

Crude venom concentration [mg/kg] (n=5)	Dead	Live	Viability rate [%]	Determined LD$_{50}$ value [mg/kg]
5	5	0	0	
2	1	4	80	**2.59**
1	0	5	100	

Appendix to Chapter 6
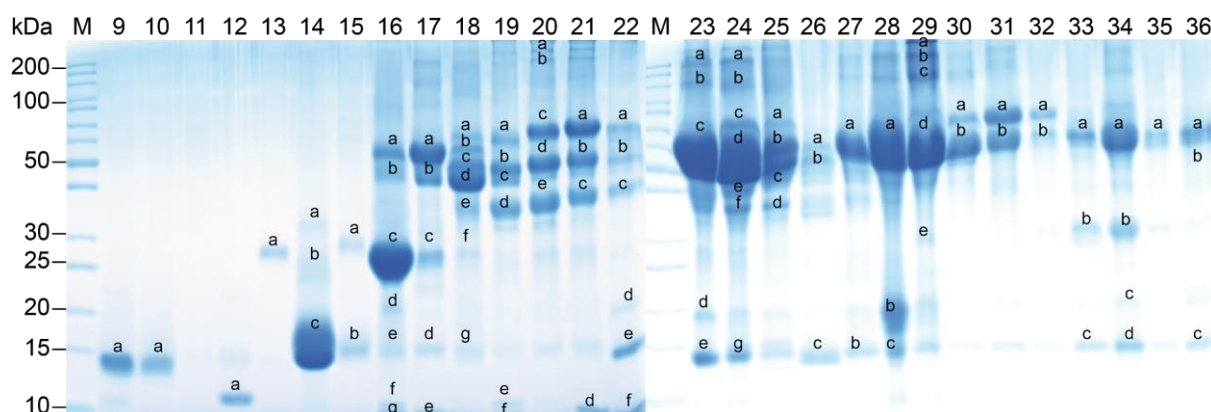
Appx Figure 6.1. SDS-PAGE fractions of *Vipera anatolica senliki* venom under reducing conditions. RP-HPLC venom fractions shown in Figure 1 were further processed by SDS-PAGE analysis. Fraction numbers are indicated above the lanes. Nomenclature shows selected bands for tryptic in-gel digestion and subsequent bottom-up venomics.
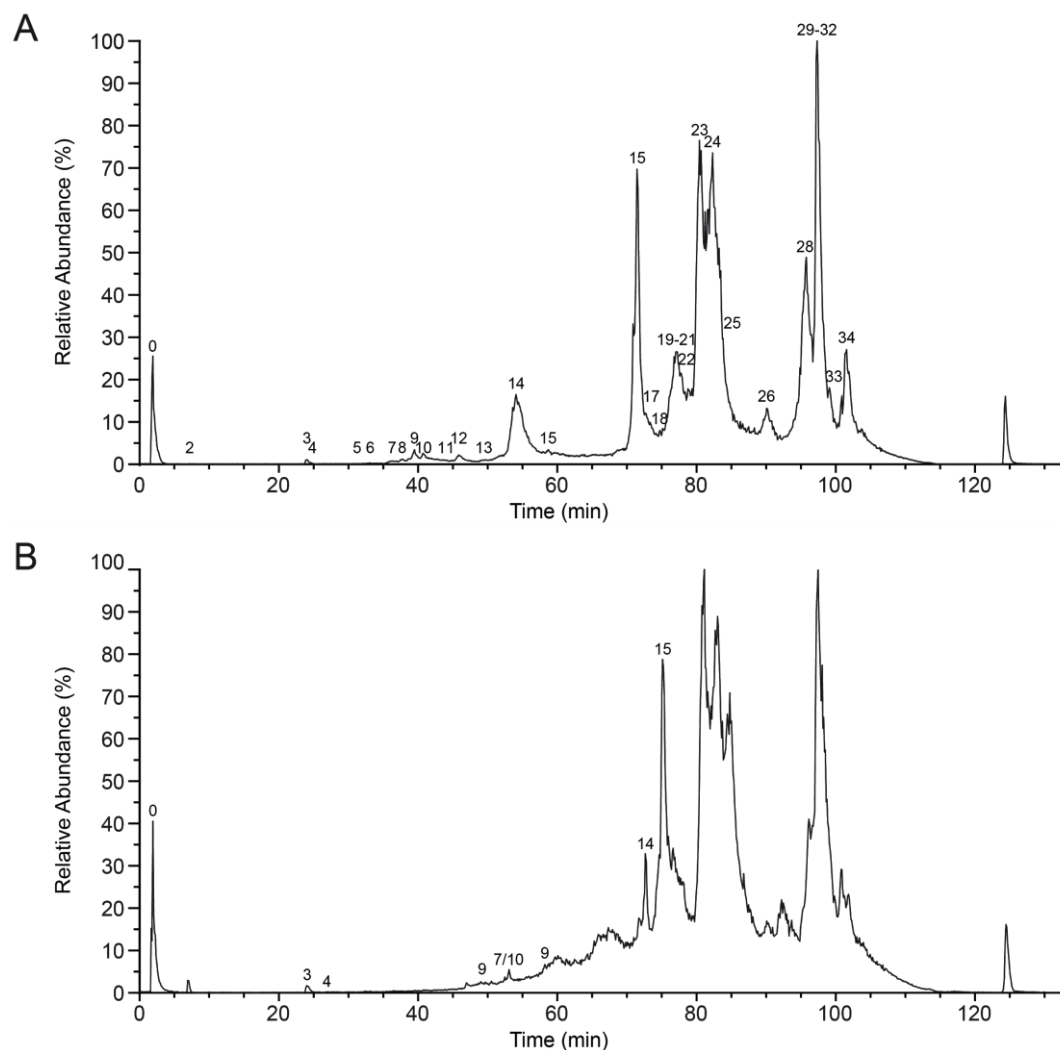
Appx Figure 6.2. Extended snake venomic analysis of native and chemically reduced *Vipera anatolica senliki* crude venom. Total ion chromatogram (TIC) from (**A**) native and (**B**) reduced *V. a. senliki* venom for IMP. The total ion counts were measured by HPLC-ESI-MS and the relative abundance was set to 100% for the highest peak. Fraction nomenclature based on Figure 6.1.
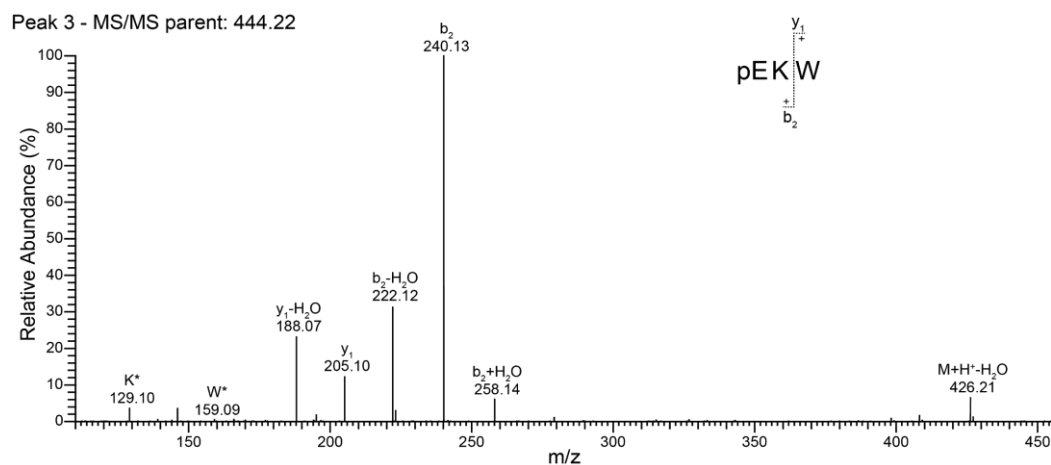
Appx Figure 6.3. Tandem MS spectrum of the tripeptidic metalloprotease inhibitor pEKW. Representative MS/MS spectra of a small tripeptidic svMP inhibitor (svMP-i) with *m/z* 444.22 precursor ion mass for *de novo* annotation in the Vipera anatolica senliki venom.
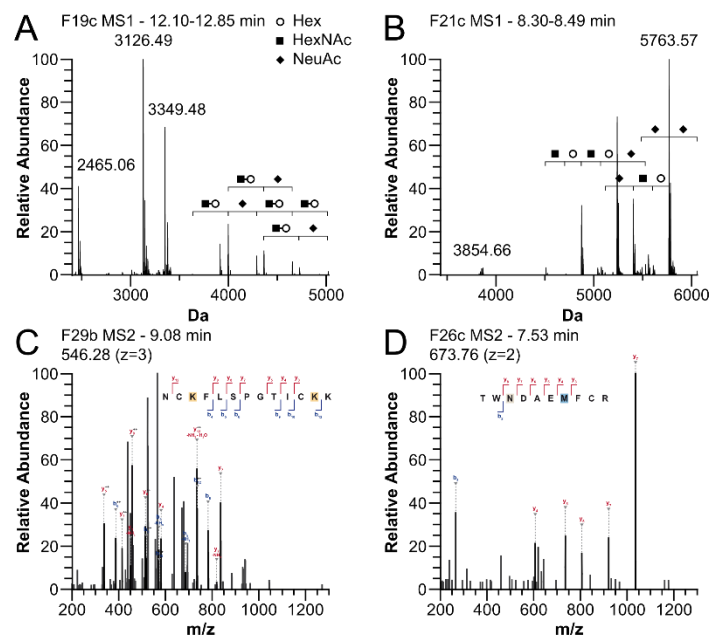
Appx Figure 6.4. Representative PTM MS spectra detected by tryptic digest BU. Deconvoluted MS spectra of two svSP fraction bands (**A**) F19c and (**B**) F21c with the appearance of three different glycosylation building blocks: hexose (Hex), N-acetyl-hexoseamine (HexNAc) and N-acetyl-neuraminic (NeuAc). The MS/MS spectra show (**C**) natural related Lys acetylation of a svMP and (**D**) experimental artificially Met oxidation.
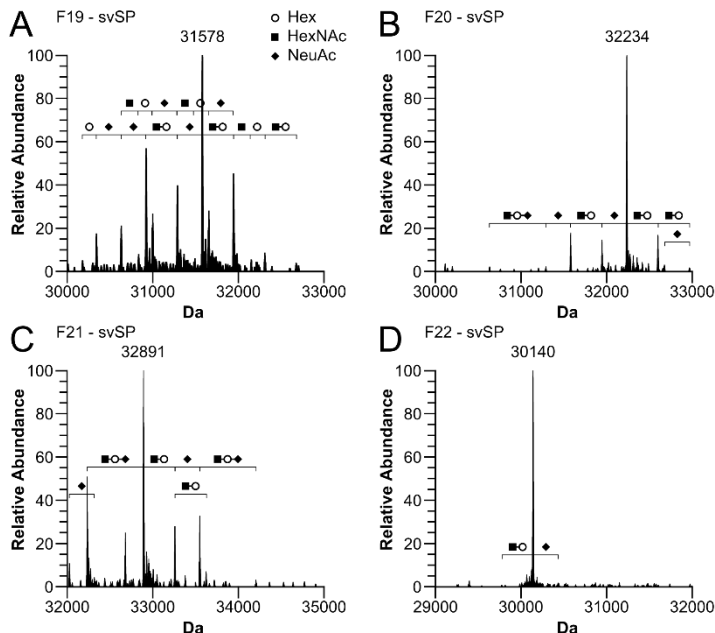
Appx Figure 6.5. Glycosylated svSP observed by IMP. Several glycosylation building blocks (hexose (Hex), N-acetyl-hexoseamine (HexNAc) and N-acetyl-neuraminic (NeuAc)) were observed by MS IMP of snake venom serine proteases (svSP) in (**A**) F19, (**B**) F20, (**C**) F21 and (**D**) F22, with a typical glycosylation branch pattern HexNAc-Hex-NeuAc.
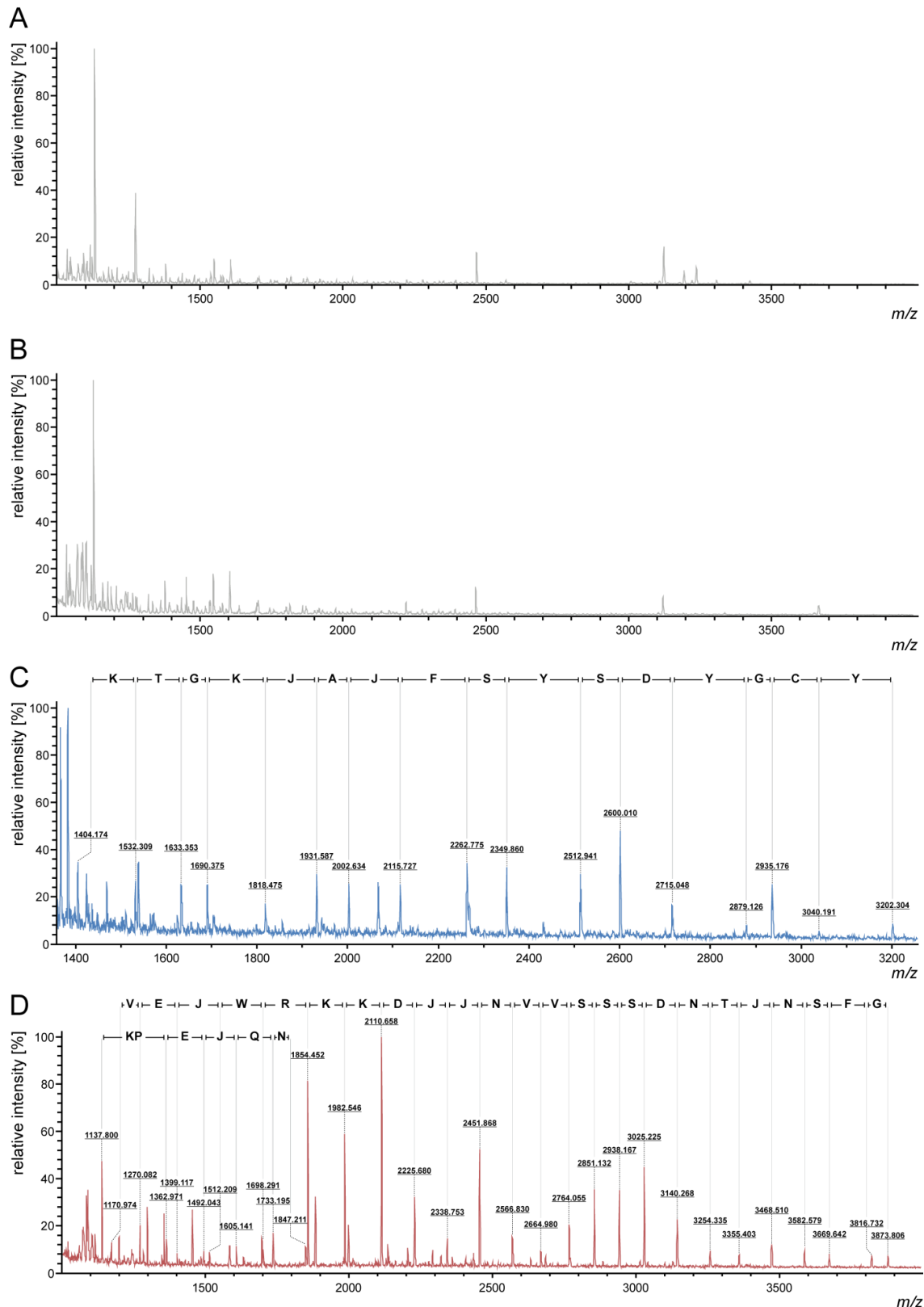
Appx Figure 6.6. MALDI top-down sequencing by in-source decay of different venom components from *Vipera a. senliki.* (**A**) Examples of top-down ISD spectra of peptide fractions (F 8) and (**B**) (F 9/10) showing no distinct sequences. (**C**) Identification of a phospholipase A$_2$ (PLA$_2$) proteoform (Ammodytin I2 (D)) by N-terminal sequence (F 14/15). (**D**) Identification of a short cysteine-rich venom protein (CRISP) peptide fragment by N-terminal sequencing, previously annotated in peak 16, as well as a snake venom metalloproteinase (svMP) proteoform by N-terminal sequencing (F 17). No distinction can be made between leucine and isoleucine (J = Leu or Ile).
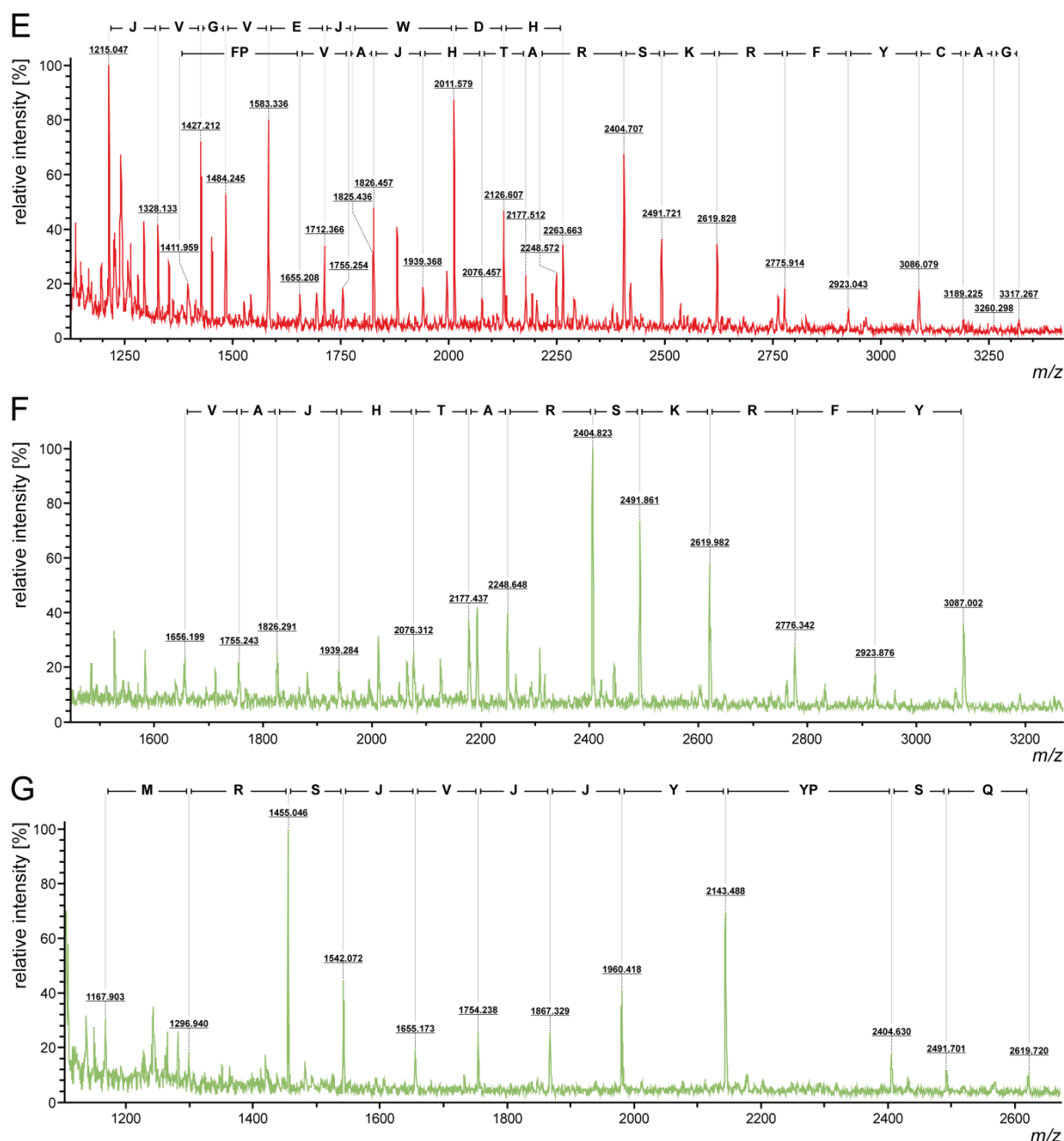
Appx Figure 6.6. *(continued)* **MALDI top-down sequencing by in-source decay of different venom components from *Vipera a. senliki*.** (**E**) Identification of a svMP proteoform by N-terminal sequencing (F 18-22). (**F**) Identification of a snake venom serine protease (svSP) proteoform by N-terminal sequencing (F 23) with a specific transcriptome hit (MN831307). (**G**) Identification of a svSP proteoform by N-terminal sequencing (F 24). No distinction can be made between leucine and isoleucine (J = Leu or Ile).
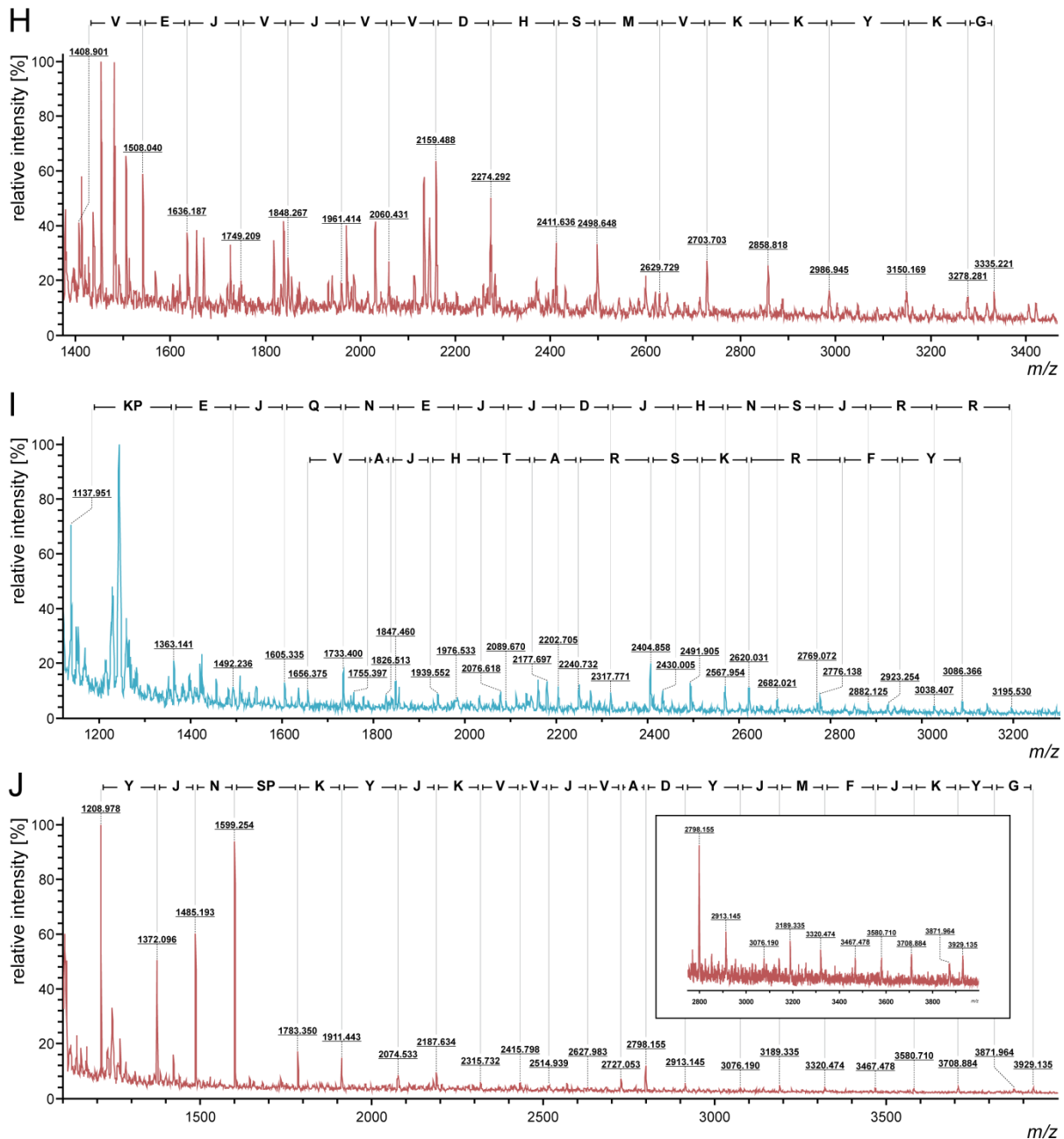
Appx Figure 6.6. *(continued)* **MALDI top-down sequencing by in-source decay of different venom components from *Vipera a. senliki.*** (**H**) Identification of a svMP proteoform by N-terminal sequencing (F 25) with a specific transcriptome hit (MN831329). (**I**) Identification of a CRISP proteoform with a specific transcriptome hit (MN831241) and a svSP proteoform with a specific database hit (MN831307) (F 27). (**J**) Identification of a svMP proteoform by N-terminal sequencing (F 33-35). No distinction can be made between leucine and isoleucine (J = Leu or Ile).
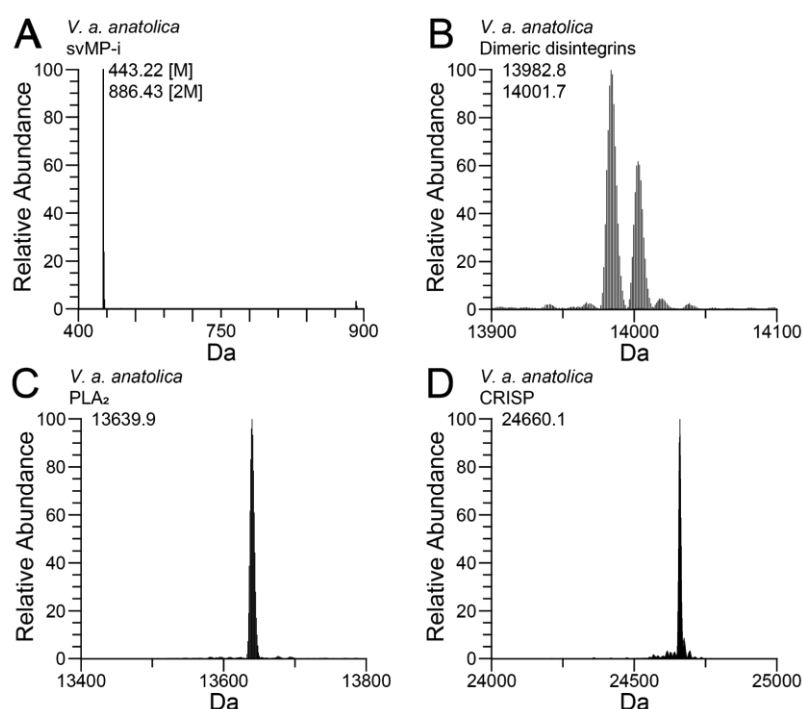
Appx Figure 6.7. Intact mass profiling of exemplary *Vipera anatolica anatolica* venom components. *V. a. anatolica* (Göçmen et al.[233]) shows compared to *V. a. senliki* (**Figure 6.2** and **Appendix Table 6.2**) identical toxin masses, like (**A**) svMP-i, (**B**) dimeric disintegrins, (**C**) PLA$_2$ and (**D**) CRISP.
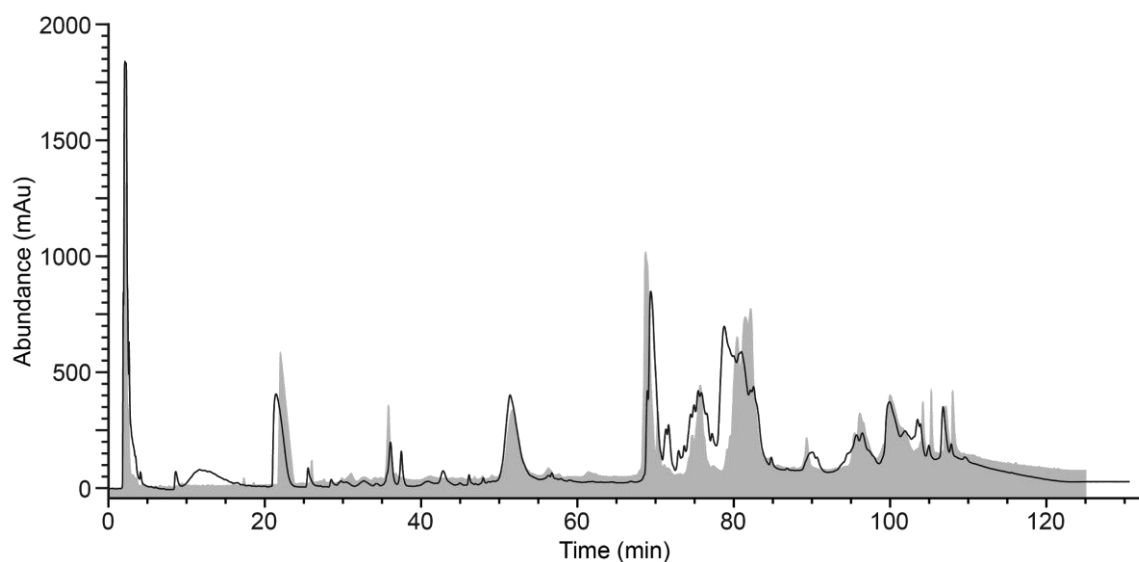
Appx Figure 6.8. Overlay of C18 RP-HPLC venom profile from *Vipera anatolica senliki* and *Vipera anatolica anatolica*. HPLC venom profile of *V. a. senliki* (black line) is shown compared to the *V. a. anatolica* (grey) analysis by Göçmen et al.[233]. Same venom amounts of venom were measured on identical devices and column.

Appx Table 6.1. Transcriptome assembled sequences identified by snake venom gland transcriptomics from *Vipera anatolica senliki.* List of full length and partial length toxin transcripts identified by *de novo* transcriptome assembly of the *Vipera anatolica senliki* venom gland. GenBank accession numbers and functional annotations are provided along with the average and monoisotopic masses of peptide hits from proteomics analysis.

BankIt Accession Number	Annotation	Full/Partial	M average[Da]	M mono [Da]
MN831207	5' nucleotidase 1	Partial		
MN831208	5' nucleotidase 2	Partial		
MN831241	cysteine-rich secretory protein 1	Partial		
MN831242	cysteine-rich secretory protein 2	Partial		
MN831243	cystatin 1	Full	15982.00	15971.38
MN831244	cystatin 2	Full	15982.00	15971.38
MN831245	cystatin 3	Full	15154.72	15144.56
MN831246	cystatin 4	Partial		
MN831247	disintegrin	Partial		
MN831248	disintegrin and metalloproteinase domain-containing protein 1	Full	84993.81	84937.96
MN831249	disintegrin and metalloproteinase domain-containing protein 2	Full	93718.67	93657.22
MN831250	disintegrin and metalloproteinase domain-containing protein 3	Partial		
MN831251	disintegrin and metalloproteinase domain-containing protein 9	Partial		
MN831252	disintegrin and metalloproteinase domain-containing protein 4	Partial		
MN831253	disintegrin and metalloproteinase domain-containing protein 8	Partial		
MN831254	disintegrin and metalloproteinase domain-containing protein 7	Partial		
MN831255	disintegrin and metalloproteinase domain-containing protein 6	Partial		
MN831256	disintegrin and metalloproteinase domain-containing protein 10	Partial		
MN831257	disintegrin and metalloproteinase domain-containing protein 5	Partial		
MN831259	glutaminyl-peptide cyclotransferase 1	Full	42731.51	42704.47
MN831260	glutaminyl-peptide cyclotransferase 2	Partial		
MN831261	hyaluronidase	Full	54710.59	54675.42
MN831278	nerve growth factor 1	Full	27325.12	27307.50
MN831279	nerve growth factor 2	Partial		
MN831280	pro-neuropeptide Y	Full	10949.67	10942.53
MN831281	phosphodiesterase 1	Full	51851.14	51817.37
MN831282	phosphodiesterase 4	Partial		
MN831283	phosphodiesterase 6	Partial		
MN831284	phosphodiesterase 3	Partial		
MN831285	phosphodiesterase 2	Partial		
MN831286	phosphodiesterase 5	Partial		
MN831287	dipeptidylpeptidase 1	Partial		
MN831288	dipeptidylpeptidase 2	Partial		
MN831289	dipeptidylpeptidase 3	Partial		
MN831290	phospholipase A2 2	Partial		
MN831291	phospholipase A2 1	Partial		
MN831292	phospholipase A2 3	Full	17449.36	17437.34
MN831293	phospholipase A2 4	Partial		

Appx Table 6.1. Transcriptome assembled sequences identified by snake venom gland transcriptomics from *Vipera anatolica senliki*. *continued*

BankIt Accession Number	Annotation	Full/Partial	M average [Da]	M mono [Da]
MN831294	phospholipase A2 5	Partial		
MN831301	serine protease 4	Full	36177.59	36153.97
MN831302	serine protease 1	Full	35544.57	35521.07
MN831303	serine protease 3	Full	38020.98	37995.87
MN831304	serine protease 5	Full	46327.53	46297.79
MN831305	serine protease 2	Full	28103.07	28084.40
MN831306	serine protease 6	Full	28913.83	28894.73
MN831307	serine protease 7	Full	28927.86	28908.75
MN831308	serine protease 11	Partial		
MN831309	serine protease 10	Partial		
MN831310	serine protease 12	Partial		
MN831311	serine protease 8	Partial		
MN831312	serine protease 9	Partial		
MN831313	coagulation factor V 4	Partial		
MN831314	coagulation factor V 2	Partial		
MN831315	coagulation factor V 3	Partial		
MN831316	coagulation factor V 5	Partial		
MN831317	coagulation factor V 1	Partial		
MN831318	coagulation factor VIII	Partial		
MN831319	serine protease inhibitor 2	Full		
MN831320	serine protease inhibitor 1	Full	15614.12	15603.13
MN831321	serine protease inhibitor 4	Full	57582.48	57544.36
MN831322	serine protease inhibitor 6	Full	27859.09	27840.57
MN831323	serine protease inhibitor 3	Full	21747.28	21732.23
MN831324	serine protease inhibitor 5	Full	9728.20	9721.63
MN831325	serine protease inhibitor 10	Partial		
MN831326	serine protease inhibitor 9	Partial		
MN831327	serine protease inhibitor 7	Partial		
MN831328	serine protease inhibitor 8	Partial		
MN831329	snake venom metalloprotease 2	Partial		
MN831330	snake venom metalloprotease 1	Partial		
MN831331	snake venom metalloprotease 3	Partial		
MN831338	veficolin 3	Full	36745.98	36722.48
MN831339	veficolin 2	Full	36009.22	35986.12
MN831340	veficolin 1	Full	36618.79	36594.94
MN831341	veficolin 5	Partial		
MN831342	veficolin 7	Partial		
MN831343	veficolin 9	Partial		
MN831344	veficolin 4	Partial		

Appx Table 6.1. Transcriptome assembled sequences identified by snake venom gland transcriptomics from
Vipera anatolica senliki. *continued*

BankIt Accession Number	Annotation	Full/Partial	M average [Da]	M mono [Da]
MN831345	veficolin 15	Partial		
MN831346	veficolin 12	Partial		
MN831347	veficolin 8	Partial		
MN831348	veficolin 10	Partial		
MN831349	veficolin 13	Partial		
MN831350	veficolin 6	Partial		
MN831351	veficolin 11	Partial		
MN831352	vascular endothelial growth factor 4	Full	25442.05	25424.58
MN831353	vascular endothelial growth factor 1	Full	17202.10	17190.37
MN831354	vascular endothelial growth factor 2	Full	22460.19	22444.81
MN831355	vascular endothelial growth factor 3	Full	16446.20	16434.99
MN831356	vascular endothelial growth factor 5	Partial		
MN831357	venom factor 1	Full	111230.94	111160.31
MN831358	venom factor 6	Partial		
MN831359	venom factor 4	Partial		
MN831360	venom factor 2	Partial		
MN831361	venom factor 5	Partial		
MN831362	venom factor 3	Partial		
MN831363	WAP four-disulfide core domain protein	Full	14043.25	14033.26

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki.* Assignments of venomic components by crude venom intact mass profiling (IMP, method A), IMP of a single RP-HPLC fraction with low molecular mass (method B), bottom-up (BU, method C) and in-source decay annotation (ISD, method D; identified sequences marked in green). Fraction numbers are based on the RP-HPLC chromatogram (**Figure 6.1**). Annotation was performed *de novo* and by peptide spectrum matching from in-gel digested protein bands (**Appendix Figure 6.1**). Peak 0 corresponds with injection peak Identification was carried out against a non-redundant *Viperidae* protein database (taxid: 8689), our custom transcriptome database and a set of proteins found as common contaminants (cRAP). SDS-PAGE and intact mass profile analysis provided the average molecular weight. Most abundant mass for IMP analysis s is marked by *. IMP performed by charge-state deconvolution was carried out with MagicTransformer (MagTran; marked by #) and with Thermo Xtract. *For the sake of clarity, a simplified version is given here.*

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
0	A	-	347.07	-	Unknown peptides	-	-	-	-	-
1	B	-	187.07	-	Unknown peptides	-	-	-	-	-
2	A; B	-	420.21; 514.32; 692.38; *717.40	-	Unknown peptides	-	-	-	-	-
3	A; B	-	*443.22; 886.44	-	svMP-i	pEKW	-	-	-	-
4	A; B	-	412.15; 429.17; 451.15; 485.13; 508.15; 531.11; 752.38; *1128.61	-	BPP	267.1-R-380.2-IPP	-	-	-	-
5	A; B	-	808.39; *1128.61; 1618.54; 3391.59; 4014.84	-	BPP	267.1-R-380.2-IPP	-	-	-	-
6	A; B	-	*808.39; 1128.61; 3943.81; 4014.84; 7296.22	-	Unknown peptides	-	-	-	-	-
7	A; B	-	*6737.95; 7280.21	-	KUN	-	-	-	-	-
8	A; B	-	3119.46; 3233.50; 3421.59; *6737.95; 7280.21; 13998.77	-	KUN	-	-	-	-	-
9	A; C	-	6736.95; 13973.77; *13982.78; 14001.77	14	DI	FLNAGTICQYAR	APB93446.1	2.00E-07	FLNAGTICQYAR	TRINITY_DN37600_c0_g1_i1_len_310_SVMP_partial
						-	-	-	NSANPCCDPITCKPR	P0C6B0.1
10	A; C	-	13999.75	14	DI	FLNAGTICQYAK	APB93446.1	3.00E-05	FLNAGTICQYAR	TRINITY_DN37600_c0_g1_i1_len_310_SVMP_partial
						-	-	-	NSANPCCDPITCKPR	P0C6B0.1
11	A; C; D	-	6630.92	-	KUN	-	-	-	FFYGGCGGNANNFKTR*	TRINITY_DN44715_c0_g1_i1_len_508_KUN

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki.* continued

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
12	A; C; D	-	6627.86	11	KUN	NPASNKCKEFFYGGCGGNANNFKTR	P0DKL8.1	2.00E-21	NPASNKCKEFFYGGCGGNANNFKTR	TRINITY_DN44715_c0_g1_i1_len_508_KUN
						-	-	-	CKEFFYGGCGGNANNFK	TRINITY_DN44715_c0_g1_i1_len_508_KUN
						-	-	-	FYYNPASNK	TRINITY_DN44715_c0_g1_i1_len_508_KUN
13	A; C	-	1101.56; 22.467.34	27	n/a	YDYSDDFDL-838.84	XP_015669739.1	2.00E-01	-	-
14	A; C; D									
		14a	-	31	n/a	-	-	-	-	-
		14b	27372.7	25	PLA2	CCFVHDCCYGR	ABU68548.1	5.00E-08	CCFVHDCCYGR	P31854.2
						VAALCFHEL-783.37	P84397.2	8.30E-01	-	-
		14c	*13639.9	14	PLA2	YKTGKIALFSYSDYGCY	CAE47222.1	7.00E-13	YKTGKIALFSYSDYGCY	CAE47222.1
						-	-	-	KTGKIALFSYSDYGCY*	CAE47222.1
						CCFVHDCCYGR	ABU68548.1	5.00E-08	CCFVHDCCYGR	P31854.2
						VAAICFGENLNTYDKK	CAE47197.1	2.00E-11	VAAICFGENLNTYDKK	F8QN53.1
						AVCECDR	AJA90797.1	2.30E-02	LTIYSYSFENGDIVCGGDDSCKRA	F8QN53.1
15	A; C; D		11143.55; 11394.64		n/a	-	-	-	-	-
		15a	22942.84	30	n/a	-	-	-	-	-
		15b	13639.9	15	PLA2	YKTGKIALFSYSDYGCY	CAE47222.1	7.00E-13	YKTGKIALFSYSDYGCY	CAE47222.1
						-	-	-	KTGKIALFSYSDYGCY*	CAE47222.1
						CCFVHDCCYGR	ABU68548.1	5.00E-08	CCFVHDCCYGR	P31854.2
						-	-	-	VAAICFGENLNTYDKK	F8QN53.1
16	A; C; D	16a	-	55	CRISP	KPEIQNEIIDLHNSLRRSVNPTA	XP_015678374.1	1.00E-18	KPEIQNEIIDLHNSLRRSVNPTA	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	KPEIQNEIIDLHNSLR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						VLQVECGENLYMSTSPMK	AMB36337.1	3.00E-10	KPEIQNEIIDLHNSLRRW*	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						MEWYPEAAANAER	BAP39957.1	3.00E-09	VIGGIECGENIYMSTSPMK	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						SVDFDSESPR	P86537.1	2.00E-05	MEWYPEAAANAER	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki.* *continued*

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
									SVDFDSESPR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	DFVYGQGASPANAVVGHYTQIVWYK	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	AVVGHYTQIVWYK	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	CIYDHSPR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	IIDLHNSLR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	NAVVGHYTQIVWYK	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	NEIIDLHNSLR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	YTQIVWYK	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
		16b	-	40	CRISP	MEWYPEAAANAER	BAP39957.1	3.00E-09	MEWYPEAAANAER	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						SVDFDSESPR	P86537.1	2.00E-05	SVDFDSESPR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	DFVYGQGASPANAVVGHYTQIVWYK	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	KPEIQNEIIDLHNSLR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	AVVGHYTQIVWYK	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
		16c	24660.15	25	CRISP	KPEIQNEIIDLHNVTR	BAP39957.1	2.00E-08	KPEIQNEIIDLHNSLR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	KPEIQNEIIDLHNSLRRW*	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						MEWYPEAAADAER	BAP39957.1	3.00E-08	MEWYPEAAANAER	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						SVDFDSESPR	P86537.1	2.00E-05	SVDFDSESPR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						YTQIVWYK	BAP39957.1	3.00E-04	CIYDHSPR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						186.1-NPTASNMLK	BAP39957.1	8.00E-04	-	-
						-	-	-	NAVVGHYTQIVWYK	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	AVVGHYTQIVWYK	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	GASPANAVVGHYTQIVWYK	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	KPEIQNEIIDLHNSLRR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	DFVYGQGASPANAVVGHYTQIVWYK	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	KDFVYGQGASPANAVVGH	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	VIGGIECGENIYMSTSPMK	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
		16d	18632.82	20	CRISP	MEWYPEAAADAER	BAP39957.1	3.00E-08	MEWYPEAAANAER	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki.* continued

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
						SVDFDSESPR	P86537.1	2.00E-05	SVDFDSESPR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	KPEIQNEIIDLHNSLR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	KPEIQNEIIDLHNSLRRW*	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	IIDLHNSLR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	DFVYGQGASPANAVVGHYTQIVWYK	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
		16e	13639.95	14	PLA2	CCFVHDCCYGR	ABU68548.1	5.00E-08	CCFVHDCCYGR	P31854.2
						-	-	-	VAAICFQKNMNTYNK	P31854.3
		16f	12330.07	11	svMP	LVIVVDHSMVEK	ADI47673.1	4.00E-06	-	-
						IVVDHSMVEK	ADI47673.1	3.00E-05	-	-
					CRISP	-	-	-	KPEIQNEIIDLHNSLR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	KPEIQNEIIDLHNSLRRW*	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	MEWYPEAAANAER	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
		16g	8192.04	10	svMP	LVIVVDHSMVEK	ADI47673.1	4.00E-06	-	-
						IVVDHSMVEK	ADI47673.1	3.00E-05	-	-
17	C; D									
		16g								
		17a	-	55	svMP	VEIWRKKDLINVVSSSDNTLNSFG	Q7T046.1	6.00E-09	VEIWRKKDLINVVSSSDNTLNSFG	Q7T046.1
						SCVMSGVLSDQPSK	Q7LZ61.2	7.00E-07	LHSWVECESGECCDQCR	Q2UXQ5.2
						NPQCLLNKPLR	ADI47757.1	1.00E-06	NPQCILNKPLR	B7U492.1
						CPLTLYQCR	ADW54339.1	7.00E-05	CPLTLYQCR	Q2UXQ5.1
						AYIGTMCQPK	ADI47602.1	3.00E-04	VEIWRKKDLINVVSSSDNTLN*	ADI47725.1
		17b	-	40	svMP	ALFGPADAVAPDR	XP_015683146.1	5.50E-02	YVEFVVVLDHGMYTKYKDNLDK	O42138.1
		17c	24.660.14	25	CRISP	MEWYPEAAANAER	BAP39957.1	3.00E-09	MEWYPEAAANAER	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						SVDFDSESPR	P86537.1	2.00E-05	SVDFDSESPR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	AVVGHYTQIVWYK	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	DFVYGQGASPANAVVGHYTQIVWYK	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	IIDLHNSLR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki.* *continued*

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
						-	-	-	KPEIQNEIIDLHNSLR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	YTQIVWYK	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
		17d	13639.91	15	PLA2	CCFVHDCCYGR	ABU68548.1	5.00E-08	CCFVHDCCYGR	P31854.2
						-	-	-	VAAICFQKNMNTYNK	P31854.3
		17e	12308.06	10	svMP	LVIVVDHSMVEK	ADI47673.1	4.00E-06	-	-
						IVVDHSMVEK	ADI47673.1	3.00E-05	-	-
18	C	18a	-	60	LAAO	-	-	-	DLKRTNCSYILNK	P0DI84.1
		18b	-	55	svMP	VPLVGVEIWDHR	ADI47619.1	2.00E-04	CPLTLYQCR	Q2UXQ5
						AYIGTMCQPK	ADW54356.1	3.00E-04	-	-
		18c	-	50	svMP	SSGDDTLDSFGEWR	ADI47643.1	6.00E-08	-	-
						VPLVGVEIWDHR	ADI47619.1	2.00E-04	-	-
		18d	-	40	svMP	500.24-CESGECCEQCR	XP_015681822.1	3.00E-07	-	-
						VPLVGVEIWDHR	ADI47619.1	2.00E-04	-	-
						ALFGPDAAVAPDR	BAP39943.1	6.00E-04	-	-
						MYLPLNIR	ADI47643.1	9.00E-04	-	-
		18e	27493.67	35	svSP	186.08-DKDLMLLR	JAV01824.1	8.10E-02	WDKDIMLIR	TRINITY_DN15387_c1_g2_i4_len_7249_SP_Partial
						LMGWGTITTTK	AMB36342.1	3.00E-05	IMGWGTITTTK	A0A1I9KNP0.1
						-	-	-	TLCAGILQGGIDSCK	A0A1I9KNP0.1
						-	-	-	HAWCEALYPWVPADSR	Q9PT41.1
						-	-	-	FYCAGTLINQEWVLTAAR	A0A1I9KNP0.1
		18f	24660.13	25	CRISP	MEWYPEAAANAER	BAP39957.1	3.00E-09	MEWYPEAAANAER	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						SVDFDSESPR	P86537.1	2.00E-05	SVDFDSESPR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial
						-	-	-	KPEIQNEIIDLHNSLR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP_Partial

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki.* continued

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
		18g	13638.91	15	PLA2	CCFVHDCCYGR	ABU68548.1	5.00E-08	CCFVHDCCYGR	F8QN53.1
						-	-	-	VAAICFGENLNTYDKK	F8QN53.1
18-22	D				svSP	FPVALHTARSKRFYCAG	Q9PT40.1	3.00E-12	FPVALHTARSKRFYCAG	VT_T0953_R_0_0019_L_1199_SP
					svMP	LVGVELWDH	P0C7B0.2	1.20E+00	-	-
19	A; C; D									
		19a	-	60	svSP	-	-	-	WDKDIMLIR	TRINITY_DN15387_c1_g2_i4_len_7249_SP_Partial
						-	-	-	IMGWGTITTTK	Q9PT40.1
						-	-	-	TLCAGILQGGIDSCK	Q9PT40.1
		19b	-	50	svMP	SSGDDTLDSFGEWR	ADI47643.1	6.00E-08	FFYTPTDENIGMVDTGTKCGDK	Q8JIR2.1
						MYLPLNIR	ADI47643.1	9.00E-04	LHSWVECESGECCDQCR	Q8JIR2.1
						HSVALVEDYSPDLR	ADI47645.1	6.00E-03	-	-
		19c	37748.44	40	svMP	SSGDDTLDSFGEWR	ADI47643.1	6.00E-08	-	-
						VPLVGVEIWDHR	ADI47619.1	2.00E-04	-	-
						HGVTLVEDYSPVER	ADI47638.1	3.00E-03	-	-
					svSP	-	-	-	FYCAGTLINQEWVLTAAR	A0A1I9KNP0.1
						-	-	-	IMGWGTITTTK	A0A1I9KNP0.1
						-	-	-	TLCAGILQGGIDSCK	A0A1I9KNP0.1
		19d		35	svSP	FPDGLDKDIMLIR	P18965.2	7.00E-08	FPNGKDKDIMLIR	Q9PT41.1
						NIQNEDEQIR	XP_015671564.1	1.00E-05	-	-
						LMGWGTITTTK	AMB36342.1	3.00E-05	IMGWGTITTTK	A0A1I9KNP0.1
						WDKDIMLLR	ASX97876.1	3.00E-04	WDKDIMLIR	TRINITY_DN15387_c1_g2_i4_len_7249_SP_Partial
						-	-	-	HAWCEALYPWVPADSR	Q9PT41.1
		19e	12330.07	11	svMP	LVIVVDHSMVEK	ADI47673.1	4.00E-06	LVGVEIW*	P0C7B0.2
						IVVDHSMVEK	ADI47673.1	3.00E-05	-	-
		19f	10478.27	10	svMP	LVIVVDHSMVEK	ADI47673.1	4.00E-06	LVGVEIW*	P0C7B0.2

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki.* continued

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
20	A; C									
		20a	-	>200	svSP	LTCAGILQGGIDTCK	AMB36345.1	1.00E-07	TLCAGILQGGIDSCK	A0A1I9KNP0.1
						615.30-LQGGIDSCK	AMB36342.1	2.00E-03	IMGWGTITTTK	A0A1I9KNP0.1
						-	-	-	WDKDIMLIR	TRINITY_DN15387_c1_g2_i4_len_7249_SP_Partial
		20b	-	>200	svSP	LMGWGTITTTK	AMB36342.1	3.00E-05	IMGWGTITTTK	A0A1I9KNP0.1
						TLCAGILQGKDS	XP_015671555.1	9.00E-05	TLCAGILQGGIDSCK	A0A1I9KNP0.1
						-	-	-	FYCAGTLINQEWVLTAAR	A0A1I9KNP0.1
						-	-	-	VTYPDVPHCADINMFDYSVCQK	A0A1I9KNP0.1
						-	-	-	WDKDIMLIR	TRINITY_DN15387_c1_g2_i4_len_7249_SP_Partial
		20c	-	65	svMP	LVNTVNEMYLPLNIR	ADI47643.1	4.00E-10	-	-
						SSGDDTLDSFGEWR	ADI47643.1	6.00E-08	-	-
						427.24-VNEFYLPLNIR	ADI47635.1	5.00E-05	-	-
					svSP	-	-	-	TLCAGILQGGIDSCK	A0A1I9KNP0.1
						-	-	-	FYCAGTLINQEWVLTAAR	A0A1I9KNP0.1
						-	-	-	IMGWGTITTTK	A0A1I9KNP0.1
						-	-	-	VTYPDVPHCADINMFDYSVCQK	A0A1I9KNP0.1
						-	-	-	WDKDIMLIR	TRINITY_DN15387_c1_g2_i4_len_7249_SP_Partial
		20d	-	50	svMP	SSGDDTLDSFGEWR	ADI47643.1	6.00E-08	-	-
						DHNAQLLTAIDFDDKLGK	ADI47585.1	2.00E-05	-	-
						KGESNFYCR	ADI47577.1	2.00E-03	-	-
					svSP	-	-	-	FYCAGTLINQEWVLTAAR	A0A1I9KNP0.1
						-	-	-	TLCAGILQGGIDSCK	A0A1I9KNP0.1
						-	-	-	IMGWGTITTTK	A0A1I9KNP0.1
						-	-	-	RGRGDSLHDYCTGVTPDCPR	Q6T6T3.1
		20e	32234.29	35	svSP	FPNKDKDIMLIR	Q9PT41.1	2.00E-06	WDKDIMLIR	TRINITY_DN15387_c1_g2_i4_len_7249_SP_Partial
						PNGLDKFDIMLIR	P18965.2	7.00E-06	TLCAGILQGGIDSCK	A0A1I9KNP0.1

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki.* continued

Peak	Meth od	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
						NIQNEDEQIR	XP_015671564.1	1.00E-05	FYCAGTLINQEWVLTAAR	A0A1I9KNP0.1
						LMGWGTITTTK	AMB36342.1	3.00E-05	IMGWGTITTTK	A0A1I9KNP0.1
						-	-	-	MFDYSVCQK	A0A1I9KNP0.1
						-	-	-	VIGGDECNINEHPFLVALHTAR	A0A1I9KNP0.1
						-	-	-	VTYPDVPHCADINMFDYSVCQK	A0A1I9KNP0.1
						-	-	-	EKFFCLSSK	A0A1I9KNP0.1
21	A; C									
		21a	-	65	svSP	LMGWGTITTTK	AMB36342.1	3.00E-05	IMGWGTITTTK	A0A1I9KNP0.1
						WDKDIMLLR	ASX97876.1	3.00E-04	WDKDIMLIR	TRINITY_DN15387_c1_g2_i4_len_7249_SP_Partial
						-	-	-	YSKCQRVHPELPAK	TRINITY_DN15387_c1_g2_i4_len_7249_SP_Partial
						-	-	-	FYCAGTLINQEWVLTAAR	A0A1I9KNP0.1
						-	-	-	TLCAGILQGGIDSCK	A0A1I9KNP0.1
		21b	-	50	svMP	SSGDDTLDSFGEWR	ADI47643.1	6.00E-08	-	-
						MYLPLNIR	ADI47643.1	9.00E-04	-	-
					svSP	-	-	-	FYCAGTLINQEWVLTAAR	A0A1I9KNP0.1
						-	-	-	IMGWGTITTTK	A0A1I9KNP0.1
						-	-	-	TLCAGILQGGIDSCK	A0A1I9KNP0.1
		21c	32889.5	40	svSP	FYCAGTLINQEWVLDKR	AMB36342.1	4.00E-10	FYCAGTLINQEWVLTAAR	A0A1I9KNP0.1
						WDKDIMLLR	ASX97876.1	3.00E-04	WDKDIMLIR	TRINITY_DN15387_c1_g2_i4_len_7249_SP_Partial
						PYGWGTITTTK	AMB36342.1	1.00E-03	IMGWGTITTTK	A0A1I9KNP0.1
						-	-	-	TLCAGILQGGIDSCK	A0A1I9KNP0.1
						-	-	-	EKFFCLSSK	A0A1I9KNP0.1
						-	-	-	NVPNEDEQMR	A0A1I9KNP0.1
						-	-	-	VIGGDECNINEHPFLVALHTAR	A0A1I9KNP0.1
						-	-	-	VTYPDVPHCADINMFDYSVCQK	A0A1I9KNP0.1
						-	-	-	EQEEDEEEQLMIASMLGLR	TRINITY_DN68352_c0_g1_i1_len_1119_SP_Partial
						-	-	-	HAWCEALYPWVPADSR	Q9PT41.1

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki*. *continued*

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
		21d	10962.85	10	PLA2	-	-	-	TLWIMAVLLVGVEGNLVQFETLIMKIAGR	D0UGJ0.1
						-	-	-	DLVQFGQMILKVAGR	P0DM49.1
					(svMP)	*LVIVVDHSMVEK*	*ADI47673.1*	*4.00E-06*		
						IVVDHSMVEK	*ADI47673.1*	*3.00E-05*		
22	A; C: D	22a	-	65	svMP	MYLPLNIR	ADI47673.1	9.00E-04	-	-
					svSP	-	-	-	IMGWGTITTTK	A0A1I9KNP0.1
						-	-	-	TLCAGILQGGIDSCK	A0A1I9KNP0.1
		22b	-	50	svMP	SSGDDTLDSFGEWR	ADI47643.1	6.00E-08	-	-
						HSVALVEDYSPVER	ADI47638.1	3.00E-04	-	-
						APISSMCQSK	AMB36352.1	4.00E-03	-	-
					svSP				WDKDIMLIR	TRINITY_DN15387_c1_g2_i4_len_7249_SP_Partial
									IMGWGTITTTK	A0A1I9KNP0.1
									TLCAGILQGGIDSCK	A0A1I9KNP0.1
		22c	33549.74	40	svSP	LMGWGTITTTK	AMB36342.1	3.00E-05	IMGWGTITTTK	A0A1I9KNP0.1
						CAGILQGGIDSSSK	AMB36342.1	4.00E-05	TLCAGILQGGIDSCK	A0A1I9KNP0.1
						WDKDIMLLR	ASX97876.1	3.00E-04	WDKDIMLIR	TRINITY_DN15387_c1_g2_i4_len_7249_SP_Partial
						NMATSKTTSPYEWVLTAAR	ABU68558.1	9.00E-03	FYCAGTLINQEWVLTAAR	A0A1I9KNP0.1
						-	-	-	VTYPDVPHCADINMFDYSVCQK	A0A1I9KNP0.1
						-	-	-	VIGGDECNINEHPFLVALHTAR	A0A1I9KNP0.1
						-	-	-	HTARSKRFY*	Q9PT40.1
						-	-	-	RSLVLLY*	AMB36343.1
		22d	24659.12	20	CTL	SPEEVDFMIK	AJO70722.1	5.00E-04	ADLVWIGLR	B0VXV0.1
						-	-	-	TTDNQWLR	B5U6Y6.1
		22e	15070.74	15	CTL	ANFVAELVTLTK	AMB36338.1	3.00E-06	TTDNQWLR	Q4PRC6.1
						HLATIEWLGK	B4XSZ1.1	3.00E-05	HLATIEWLGK	B4XSY7.1

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki.* *continued*

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
23	A; C; D					ADLVWIGLR	APB93444.1	4.00E-04	-	-
						SPEEVDFMIK	AJO70722.1	5.00E-04	-	-
						TWEDAENFPDAF	AMB36338.1	2.00E-03	TWEDAEKFCNE	Q4PRC6.1
		22f	-	10	n/a	-	-		-	-
					svSP	VAJHTARSKRFY	Q9PT40.1	7.00E-07	VALHTARSKRFY	VT_T0953_R_0_0019_L_1199_SP
		23a	-	200	svMP	YNSDLTVIR	ADI47687.1	1.20E-02	-	-
		23b	-	150	svMP	IVVDHSMVTK	CAJ01687.1	4.00E-05	DLKRTNCSYILNK	P0DI84.1
						APISSMCQSK	AMB36352.1	4.00E-03	LHSWVECESGECCDQCR	Q0NZX8.1
						HSVALVEDYS-481.27	ADI47645.1	4.30E-02	VNGEPVVLYLEKNKQLFSK	Q2QA02.1
						-	-	-	FFYTPTDENIGMVDTGTKCGDK	Q8JIR2.1
		23c	-	50	svMP	HDNAQLLTAIDFDGSVIGK	AEJ31993.1	5.00E-11	LHSWVECESGECCDQCR	Q8JIR2.1
						SSGDDTLDSFGEWR	ADI47643.1	6.00E-08	DTLDSFGEWR	Q8JIR2.1
						LVIVVDHSMVTK	ADI47638.1	5.00E-07	FFYTPTDENIGMVDTGTKCGDK	Q8JIR2.1
						LNEFYLPLNIR	ADI47635.1	6.00E-04	EAHAVFKYENVEK	Q8JIR2.1
						-	-	-	AGAVCRAARTECDIPENCTDQS	Q3HTN2.1
						-	-	-	LQGEIYLIEPLKLRDSEAHAVFK	J3S830.1
						-	-	-	CPLTLYQCR	Q2UXQ5.1
						-	-	-	VNGEPVVLYLEKNKQLFSK	P0CB14.1
		23d	24451.00#	20	CTL	SPEEVDFMIK	AJO70722.1	5.00E-04	TTDNQWLR	I7ICN3.1
						ADLVWIGLR	APB93444.1	4.00E-04	ADLVWIGLR	B0VXV0.1
						TTDNQWLR	Q4PRC6.1	1.00E-03	HLATIEWLGK	B4XSY7.1
		23e	13665.80#	14	CTL	TWEDAENFCQK	AMB36338.1	3.00E-07	KTWEDAEK	B4XSY7.1
						HLATIEWLGK	B4XSZ1.1	3.00E-05	HLATIEWLGK	B4XSY7.1
						ADLVWIGLR	APB93444.1	4.00E-04	ADLVWIGLR	B0VXV0.1

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki.* continued

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
24	A; C; D					SPEEVDFMIK	AJO70722.1	5.00E-04	GGHLISLK	B4XT05.1
						TTDNQWLR	Q4PRC6.1	1.00E-03	TTDNQWLR	I7ICN3.1
						VFAELVTLTKPK	AMB36338.1	6.00E-03	DQECLPGWSFYEGHCYK	B5U6Y7.1
						-	-	-	DQDCLPGWSFYEGHCYK	B4XSY7.1
		24a	-	200	svMP	CWDTLDSFGEWR	XP_015681822.1	5.00E-07	-	-
						SSGDATLDSFGEWR	ADJ67475.1	2.00E-06	-	-
						LVIVVDHSMVEK	ADI47673.1	4.00E-06	-	-
						ALFGPADAVAPDR	XP_015683146.1	5.50E-02	-	-
		24b	-	150	svMP	SSGDATLDSFGEWR	ADJ67475.1	2.00E-06	TCPTMNNQCIALFGPNAAVSQDACFQFNR	Q90ZI3.1
						LVIVVDHSMVEK	ADI47673.1	4.00E-06	-	-
		24c	-	70	svMP	SSDGDTLDSFGEWR	ADI47654.1	2.00E-07	CPLTLYQCR	Q2UXQ5.1
						LVIVVDHSMVEK	ADI47673.1	4.00E-06	NPQCILNKPLR	B7U492.1
						AYIGTMCQPK	ADW54356.1	3.00E-04	MIQVLLVTICLAVFPYQVSSK	Q7ZZM2.1
						VPLVGVELWDHR	ADI47619.1	6.00E-03	NPQCILNKPLR	B7U492.1
		24d	-	50	svMP	SSDGDTLDSFGEWR	ADI47654.1	2.00E-07	TLDSFGEWR	Q2UXQ5.1
						LVIVVDHSMVEK	ADI47673.1	4.00E-06	CEACIMSAVISDKQSK	P17349.2
						VPLVGVEIWDHR	ADI47619.1	2.00E-04	CPLTLYQCR	Q2UXQ5.1
						IVDHVSMVEK	ADI47673.1	8.70E-02	VHELVNTVNGFFRSKQDLIK	P0C7A9.1
						HSVALVEDYSVEPR	ADI47645.1	9.30E-02	-	-
		24e	-	40	svMP	SSDGDTLDSFGEWR	ADI47654.1	2.00E-07	-	-
						LVIVVDHSMVEK	ADI47673.1	4.00E-06	-	-
						VPLVGVEIWDHR	ADI47619.1	2.00E-04	-	-
						ALFGPDAAVAPDR	XP_015683146.1	6.00E-04	-	-

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki.* continued

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
						MYLPLNIR	ADI47643.1	9.00E-04	-	-
					(svSP)	-	-	-	*IMGWGTITTTK*	*A0A1I9KNP0.1*
						-	-	-	*TLCAGILQGGIDTCK*	*E0Y419.1*
		24f	-	35	svSP	LMGWGTITTTK	AMB36342.1	3.00E-05	IMGWGTITTTK	A0A1I9KNP0.1
						WDKDIMLLR	ASX97876.1	3.00E-04	WDKDIMLIR	TRINITY_DN15387_c1_g2_i4_len_7249_SP_Partial
						MRSLVLLYYPSQ	P81038.1	3.00E-03	MRSLVLLYYPSQ	P81038.1
						-	-	-	EYTMWDKDIMLIR	E5AJX2.1
						MNASKTFQPYEWVLTAAR	BAP39909.1	3.10E-02	FYCAGTLINQEWVLTAAR	A0A1I9KNP0.1
						-	-	-	TLCAGILQGGIDSCK	A0A1I9KNP0.1
						-	-	-	VIGGDECNINEHPFLVALHTAR	A0A1I9KNP0.1
						-	-	-	VTYPDVPHCADINMFDYSVCQK	A0A1I9KNP0.1
						-	-	-	VVCAGIWQGGK	E5AJX2.1
						-	-	-	VILPDVPHCANIEILK	E5AJX2.1
						-	-	-	FPNGLDKDIMLIR	P18964.1
		24g	-	14	CTL	TWEDAENFCQK	AMB36338.1	3.00E-07	DQDCLPGWSFYEGHCYK	B4XSY7.1
						HLATIEWLGK	B4XSZ1.1	3.00E-05	HLATIEWLGK	B4XSY7.1
						CELAYNFICQSR	B4XSZ1.1	4.00E-05	ADLVWIGLR	B0VXV0.1
						SPEEVDFMIK	AJO70722.1	5.00E-04	GGHLISLK	B4XT05.1
						-	-	-	TTDNQWLR	I7ICN3.1
25	C; D	-	-	-	svMP	VEIVIVVDHSMVKKYKG	Q9PWJ0.1	3.00E-10	ELVIVVDHRMVKKYKN	TRINITY_DN2248_c0_g1_i1_len_747_SVMP
		25a	-	70	svMP	VTSSGDDTLDSFGEWR	ADI47643.1	7.00E-10	CPLTLYQCR	Q2UXQ5.1
						-	-	-	FFYTPTDENIGMVDTGTKCGDK	Q8JIR2.1
		25b	60110.00#	60	svMP	SSGDDTLDSFGEWR	ADI47643.1	6.00E-08	YNSNVNTIR	C0HJU2.1
						LVIVVDHSMVTK	ADI47638.1	5.00E-07	VHELVNTVNGFFRSKQDLIK	P0C7A9.1
						326.20-LNEFYLPLNIR	ADI47635.1	6.00E-04	CPLTLYQCR	Q2UXQ5.1

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki.* continued

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
						HSVALVEDYSQLDR	ADI47645.1	2.30E-02	-	-
		25c	-	50	svMP	VTSSGDDTLDSFGEWR	ADI47643.1	7.00E-10	TDIVSPPACGNELLER	B8K1W0.1
						LVIVVDHSMVEK	ADI47673.1	4.00E-06	VHELVNTVNGFFRSKQDLIK	P0C7A9.1
						VPLVGVEIWDHR	ADI47619.1	2.00E-04	CPLTLYQCR	Q2UXQ5.1
		25d	36165.00#	35	svSP	TLCAGILQGGIDSCK	AMB36342.1	1.00E-09	TLCAGILQGGIDSCK	A0A1I9KNP0.1
						LMGWGTITTTK	AMB36342.1	3.00E-05	IMGWGTITTTK	A0A1I9KNP0.1
						-	-	-	FYCAGTLINQEWVLTAAR	A0A1I9KNP0.1
						-	-	-	EYTMWDKDIMLIR	E5AJX2.1
						-	-	-	VILPDVPHCANIEILK	E5AJX2.1
						-	-	-	VVCAGIWQGGK	E5AJX2.1
						-	-	-	WDKDIMLIR	TRINITY_DN15387_c1_g2_i4_len_7249_SP_Partial
26	A; C									
		26a	64698.00#	60	svMP	LVIVVDHSMVTK	ADI47638.1	5.00E-07	LHSWVECESGECCEQCR	Q0NZX8.1
						CWDTLDSFGEWR	XP_015681822.1	2.00E-05	-	-
		26b	-	55	svMP	VPLVGVEMWVYK	ADW54336.1	3.20E-02	CPLTLYQCR	Q2UXQ5.1
						-	-	-	QNGVTIPCARKIK	C5H5D2.1
		26c	48685.00#	50	CTL	HLATIEWLGK	B4XSZ1.1	3.00E-05	HLATIEWLGK	B4XSY8.1
						EFCAELVYFTGYR	P86970.2	3.00E-04	DQDCLPGWSFYEGHCYK	B4XSY7.1
						SPEEVDFMIK	AJO70722.1	5.00E-04	ADLVWIGLR	B0VXV0.1
						WTDDAEMFCR	Q6T7B7.1	3.00E-03	TTDNQWLR	Q4PRC6.1
27	A; C; D									
		-	-	-	CRISP	KPEIQNEIIDLHNSLRR	XP_015678374.1	9.00E-13	KPEIQNEIIDLHNSLRR	TRINITY_DN8323_c0_g1_i1_len_755_CRISP
		-	-	-	svSP	VALHTARSKRFY	Q9PT40.1	7.00E-07	VALHTARSKRFY	VT_T0953_R_0_0019_L_1199_SP
		27a	-	50	svMP	VTSSDGDTLDSFGEWR	ADI47643.1	4.00E-08	LHSWVECESGKCCNQCR	B8K1W0.1
						SSGDDTLDSFGEWR	ADI47643.1	6.00E-08	FLTQRNPKCMINKPLR	Q2UXQ5.1

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki.* continued

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
						LVIVVDHSMVTK	ADI47638.1	5.00E-07	-	-
		27b	13641.80#	15	CTL	ANFVAELVTLTK	AMB36338.1	3.00E-06	-	-
						HLATIEWLGK	B4XSZ1.1	3.00E-05	HLATIEWLGK	B4XSY7.1
						CELAYNFICQSR	B4XSZ1.1	4.00E-05	ADLVWIGLR	B5U6Y7.1
						SPEEVDFMIK	AJO70722.1	5.00E-04	DQECLPGWSFYEGHCYK	C0HKZ7.1
						TWEDAENFQCK	AMB36338.1	2.00E-03	TTDNQWLR	B5U6Y6.1
28	A; C									
		28a	60922.00#	60	svMP	ATVAEDSCFQENLK	ADI47635.1	5.00E-09	HDNAQLLTGMR	TRINITY_DN2248_c0_g1_i1_len_747_SVMP_Partial
						LYEMVNTLNVVFR	AMB36352.1	7.00E-08	SVGLIQYDSTTNLR	TRINITY_DN2248_c0_g1_i1_len_747_SVMP_Partial
						FVIVVDHSMVTK	ADI47655.1	1.00E-06	CILNPPLRK	TRINITY_DN2248_c0_g1_i1_len_747_SVMP_Partial
						VPLVGVEFWFHK	ADW54336.1	3.00E-04	FDLSTLGITFYEGMCQAYR	TRINITY_DN2248_c0_g1_i1_len_747_SVMP_Partial
						-	-	-	TAVIMAHEIGHNLGMEHDK	TRINITY_DN2248_c0_g1_i1_len_747_SVMP_Partial
						-	-	-	TFIELVIVVDHR	TRINITY_DN2248_c0_g1_i1_len_747_SVMP_Partial
						-	-	-	TLDSFAEWR	TRINITY_DN2248_c0_g1_i1_len_747_SVMP_Partial
						-	-	-	YYCTCGGNSCIMSAVLSNQPSK	TRINITY_DN2248_c0_g1_i1_len_747_SVMP_Partial
						-	-	-	LHSWVECESGECCDQCR	Q2UXQ5.1
						-	-	-	CPLTLYQCR	Q2UXQ5.1
						-	-	-	KIPCAPQDIK	O42138.1
						-	-	-	DKIIVQSSADVTLDLFAK	C5H5D3.1
		28b	19383.60#	20	CTL	ALAEESYCLLILTHK	Q7T045.1	5.00E-08	EQECSSEWSDGSSVSYDKLGK	Q4PRD2.1
						TWFNLNCEER	Q696W1.1	3.00E-04	CPPDSSPYR	Q696W1.1
						ADLVWIGLR	APB93444.1	4.00E-04	ADLVWIGLR	B0VXV0.1
						THFWIGLR	Q696W1.1	9.00E-04	FITHFWIGLR	Q696W1.1
						-	-	-	HLATIEWLGK	B4XSY7.1

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki.* continued

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
		28c	16221.50#	15	CTL	TWEDAENFCQK	AMB36338.1	3.00E-07	ADLVWIGLR	B0VXV0.1
						HLATIEWLGK	B4XSZ1.1	3.00E-05	HLATIEWLGK	B4XSY7.1
						CELAYNFICQSR	B4XSZ1.1	4.00E-05	DQDCLPGWSFYEGHCYK	B4XSY7.1
						NAFVAELVTLTK	AMB36338.1	4.00E-04	TTDNQWLR	B5U6Y6.1
						SPEEVDFMIK	AJO70722.1	5.00E-04	-	-
29	C; D									
		-	-	-	svMP	EJVJVVDNVMFRKYNG	AHB62069.2	1.00E+00	ELVIVVDHRMVKKYKN	TRINITY_DN2248_c0_g1_i1_len_747_SVMP
						EJVJVVDNVMFKKYNG	AHB62069.1	3.00E-09	ELVIVVDHRMVKKYKN	TRINITY_DN2248_c0_g1_i1_len_747_SVMP
						EJVJVVDNVMFKKYK	AHB62069.1	1.00E-08	ELVIVVDHRMVKKYKN	TRINITY_DN2248_c0_g1_i1_len_747_SVMP
		29a	-	>200	svMP	LYEMVNTLNVVFR	AMB36352.1	7.00E-08	-	-
						HDNAQLLTGIDNKR	XP_015683144.1	2.00E-05	-	-
						EFEFVNTLNVVFR	AHB62069.1	1.00E-04	-	-
						443.19-TSFTLDLFGDWR	ADI47668.1	2.00E-04	-	-
					(PLB)	-	-	-	*SLEDGTLYIIEQVPK*	*F8S101.1*
						-	-	-	*FTAYAINGPPVEK*	*F8S101.1*
		29b	-	200	svMP	LYEMVNTLNVVFR	AMB36352.1	7.00E-08	-	-
						443.19-TSFTLDLFGDWR	ADI47668.1	7.00E-06	-	-
						HDNAQLLTGIDNKR	XP_015683144.1	2.00E-05	-	-
						TCIMSPVAGSK	AHB62069.1	9.00E-03	-	-
					(PLB)	-	-	-	*SLEDGTLYIIEQVPK*	*F8S101.1*
						-	-	-	*FTAYAINGPPVEK*	*F8S101.1*
		29c	-	150	svMP	LYEMVNTLNVVFR	AMB36352.1	7.00E-08	-	-
						HDNAQLLTGID-440.25	XP_015683144.1	9.00E-06	-	-
						LFVALVGVEIWNK	AHB62069.1	2.00E-06	-	-

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki.* continued

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
		29d	-	50	svMP	SSGDDTLDSFGEWR	ADI47643.1	6.00E-08	SVGLIQYDSTTNLR	TRINITY_DN2248_c0_g1_i1_len_747_SVMP_Partial
						HDNAQLLTGIDLDGR	J3SDW6.1	2.00E-07	IPLNIHVTLTGVEFWCDR	TRINITY_DN2248_c0_g1_i1_len_747_SVMP_Partial
						PDYGMVDLGTK	AAX86634.1	3.00E-06	NCKFLSPGTICKK	P83041.1
						VQSEASFTLGLFGDWR	AMB36350.1	6.00E-05	HAPGDNGMVDPGTKCEDK	C0LZJ5.1
						EFEFVNTLNVVFR	AHB62069.1	1.00E-04	-	-
		29e	-	30	CTL	-	-	-	GQAEVWIGLWDKKK	P0DL30.1
30	C	30a	-	70	svMP	TWVFEMVNTLNEIYR	ABG26979.1	3.00E-09	LHSWVECESGECCEQCR	Q0NZX8.1
						DLFGDWR	ADI47668.1	1.10E-02	-	-
		30b	-	55	svMP	LYEMVNTLNVVFR	AMB36352.1	7.00E-08	NCKFLSPGTICKK	P83041.1
						HDNAQLLTGID-440.25	XP_015683144.1	9.00E-06	GMVLPGTKCADGKVCSNR	A2CJE2.1
						VDSNVSFTLGLFGDWR	P20164.4	7.00E-04	-	-
						565.22-LLTGIDLNGR	ADI47641.1	7.00E-04	-	-
						SCIMSPVAGSK	AHB62069.1	9.00E-03	-	-
31	C	31a	-	70	svMP	LYEMVNTLNVVFR	AMB36352.1	7.00E-08	LHSWVECESGECCDQCR	Q0NZX9.1
						YNQKEMVNTLNEIYR	ABG26979.1	1.00E-05	GMVLPGTK	Q0NZX9.1
						211.09-ECILNQPLR	ADI47711.1	2.00E-03	FGSDATMAQDSCFQVNK	P86092.1
		31b	-	55	svMP	SSGDDTLDSFGEWR	ADI47643.1	6.00E-08	NCKFLSPGTICKK	P83041.1
						HDNAQLLTGIDLDGR	J3SDW6.1	2.00E-07	-	-
						LVIVVDHSMVTK	ADI47638.1	5.00E-07	-	-
						VNLLNEMYLPLNLR	ADI47590.1	1.00E-05	-	-
						EFEFVNTLNVVFR	AHB62069.1	1.00E-04	-	-
						EWYFTTCSQNQYQEFR	ABG26980.1	3.00E-03	-	-

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki*. *continued*

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
31-32	D	-	-	-	svMP	LVIVVDNVMFKKYK	AHB62069.1	1.00E-07	LVIVVDHRMVKKYK	TRINITY_DN2248_c0_g1_i1_len_747_SVMP
						YVELVITVDHRM	B8K1W0.1	2.00E-05	FIELVIVVDHRM	TRINITY_DN2248_c0_g1_i1_len_747_SVMP
32	C									
		32a	-	70	svMP	386.18-FEMVNTLNEIYR	XP_015683679.1	2.00E-05	-	-
						LGIEYAYCR	ADW54334.1	1.70E-02	-	-
		32b	-	55	svMP	SSGDDTLDSFGEWR	ADI47643.1	6.00E-08	-	-
33	C									
		33a	-	55	svMP	VTSSGDDTLDSFGEWR	ADI47643.1	7.00E-10	SVGLIQYDSTTNLR	TRINITY_DN2248_c0_g1_i1_len_747_SVMP_Partial
						LVIVVDHSMVTK	ADI47638.1	5.00E-07	LHSWVECESGECCEQCR	Q0NZX9.1
						326.19-LNEFYLPLNIR	ADI47635.1	6.00E-04	-	-
						SVAIVEDYSWNAQ	Q8QG88.1	3.00E-03	-	-
		33b	-	35	svMP	HSVKVQDHSVINR	AMB36350.1	3.00E-03	YIELVIVADHAMVTK	Q98995.1
		33c	-	15	CTL	HLATIEWLGK	B4XSZ1.1	3.00E-05	HLATIEWLGK	B4XSY7.1
						ADLVWIGLR	APB93444.1	4.00E-04	ADLVWIGLR	B0VXV0.1
						SPEEVDFMIK	AJO70722.1	5.00E-04	DQECLPGWSFYEGHCYK	B5U6Y7.1
33-35	D	-	-	-	svMP	YLNSPKYIKVVIVADYIMFLKYG	Q4VM08.1	2.00E-18	-	-
34	C									
		34a	-	60	svMP	VTSSGDDTLDSFGEWGV	ADI47643.1	1.00E-08	SVGLIQYDSTTNLR	TRINITY_DN2248_c0_g1_i1_len_747_SVMP_Partial
						SCIMSGTLSCEGS	ADI47586.1	7.00E-08	SCIMSGTLSCEASIR	Q4VM08.1
						LYCFDNLPEHK	ADW54334.1	4.00E-07	LYCFDNLPEHK	Q4VM08.1
						SSDGDTLDSFGEWR	ADI47654.1	2.00E-07	IYEIVNILNVIYR	Q4VM08.1
						LGNEYGYCR	AGL45259.1	2.00E-06	LGNEYGYCR	Q4VM08.1
						-	-	-	TRIYEIVNILNVIYR	Q4VM08.1
						-	-	-	VTLDLFGK	Q4VM08.1
						-	-	-	MPQCILNK	Q4VM08.1

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki.* *continued*

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
		34b	-	32	svMP	NPQCILNQPLR	ADI47711.1	1.00E-06	NPQCILNKPLR	Q98995.1
						YIELVIVADHAFV-229.14	ADI47713.1	2.00E-06	YIELVIVADHAMVTK	Q3ZD74.1
						VVLVADYIMFLK	C5H5D6.1	2.00E-05	QLVNNIIVFYR	Q3ZD74.1
						SSEALDLFGEWR	ADI47614.1	1.00E-03	QLNLTPEQQR	Q98995.1
						487.59-VAHELGHN-429.24	XP_015669710.1	1.00E-02	-	-
		34c	-	20	CTL	ADLVWIGLR	APB93444.1	4.00E-04	ADLVWIGLR	B0VXV0.1
						-	-	-	HLATIEWLGK	B4XSY7.1
		34d	-	15	CTL	HLATIEWLGK	B4XSZ1.1	3.00E-05	HLATIEWLGK	B4XSY7.1
						CELAYNFICQSR	B4XSZ1.1	4.00E-05	KTWEDAEK	B4XSY7.1
						ADLVWIGLR	APB93444.1	4.00E-04	ADLVWIGLR	B5U6Y7.1
						SPEEVDFMIK	AJO70722.1	5.00E-04	DQECLPGWSFYEGHCYK	B5U6Y7.1
						TTDNQWLR	Q4PRC6.1	1.00E-03	TTDNQWLR	B5U6Y6.1
35	C	35a	-	60	svMP	SSGDDTLDSFGEWR	ADI47643.1	6.00E-08	SVGLIQYDSTTNLR	TRINITY_DN2248_c0_g1_i1_len_747_SVMP_Partial
						LVIVVDHSMVTK	ADI47638.1	5.00E-07	-	-
						VPLVGVEIWDHR	ADI47619.1	2.00E-04	-	-
						YNSDLTVIR	ADI47687.1	1.20E-02	-	-
						EVKLHEDYSPIDR	ADI47638.1	5.50E-02	-	-
36	C									
		36a	-	70	svMP	SSGDDTLDSFGEWR	ADI47643.1	6.00E-08	SVGLIQYDSTTNLR	TRINITY_DN2248_c0_g1_i1_len_747_SVMP_Partial
						LVIVVDHSMVTK	ADI47638.1	5.00E-07	LHSWVECESGECCEQCR	Q0NZX8.1
						VPLLGVEIWDNR	B8K1W0.1	4.00E-03	-	-
						137.05-SVAIVEDYKLTTR	Q8QG88.1	2.70E-02	-	-

Appx Table 6.2. Venom proteins and peptides identified from *Vipera anatolica senliki. continued*

Peak	Method	Band	Mass [Da] (ESI-MS)	Mass [kDa]	Protein Family	Sequence (de novo)	Prot-ID (blast)	E-value	Sequence (PSM)	Prot-ID (PSM)
		36b	-	60	CTL	VPLVGVEIWDHR	ADI47619.1	2.00E-04	LGNEYGYCR	Q4VM08.1
						518.28-VNLLNVIYR	AGL45259.1	4.00E-04	IYEIVNILNVIYR	Q4VM08.1
						GLNEYGYCR	AGL45259.1	1.20E-02	-	-
						LDHGQQLTGIDLNGR	Q98995.1	1.40E-02	-	-
		36c	-	15	CTL	ANFVAELVTLTK	AMB36338.1	3.00E-06	DQECLPGWSFYEGHCYK	B5U6Y7.1
						HLATIEWLGK	B4XSZ1.1	3.00E-05	HLATIEWLGK	B4XSY7.1
						ADLVWIGLR	APB93444.1	4.00E-04	ADLVWIGLR	B0VXV0.1
						SPEEVDFMIK	AJO70722.1	5.00E-04	EAVFVAELLSENVK	P0DJL2.1
						-	-	-	TWEDAERFCLD	P0DJL2.1
						-	-	-	TTDNQWLR	B5U6Y6.1

Appx Table 6.3. Compositional venom lineup of two *Vipera anatolica* subspecies. The most abundant toxin families in the venoms of *Vipera anatolica senliki* (this study) and *Vipera anatolica anatolica* (Göçmen et al.[233]) are compared by their HPLC retention time. Venoms were measured on identical devices and column. Identical identified masses in the IMP are mentioned in the correspondent row.

	V. a. senliki		*V. a. anatolica*		identical IMP masses
Peak No.	most abundant fraction components	t_R in min	t_R in min	most abundant fraction components	in both venoms in Da
0	peptides	2	2	peptides	
1	peptides	9	-	-	-
2	peptides	12	-	-	-
-	-	-	17	peptides	-
-	-	-	18	peptides	-
3	svMP-i (pEKW)	22	23	svMP-i (pEKW)	443.2; 886.4
4	BPP; peptides	26	26	peptides	429.2; 752.4; 1128.6
-	-	-	28	peptides	-
5	BPP; peptides	29	30	peptides	-
6	unknown peptides	30	31	peptides	808.39; 1128.61; 3943.81
7	Kunitz-inhibitor; peptides	33	33	Kunitz-inhibitor, peptides	6738.0; 7280.2
8	Kunitz-inhibitor; peptides	34	34	peptides	3119.46; 3233.50; 3421.59
9	DI	36	36	DI	13982.8; 14001.8
10	DI	38	-	-	-
11	Kunitz-inhibitor	41	-	-	-
12	Kunitz-inhibitor	43	-	-	-
13	unknown protein	46	47	peptides	1101.5
-	-	-	49	peptides	-
14	PLA$_2$	52	52	PLA$_2$	13639.9
15	PLA$_2$	57	56	unknown protein	11143.6; 11394.6; 13639.9
-	-	-	62	PLA$_2$	-
16	CRISP	70	69	CRISP	24660.1
-	-	-	70	CRISP	-
17	svMP; CRISP	72	71	svMP; CRISP	-
18	svMP	73	-	-	-
19	svMP; svSP	75	75	unknown protein; svSP	-
20	svMP; svSP	76	76	svMP	-
21	svMP; svSP	77	-	-	-
22	svMP; svSP; CTL	78	78	unknown protein	-
23	svMP; CTL	79	-	-	-
-	-	-	80	svMP	-
24	svMP	81	81	svMP	-
25	svMP	83	82	svMP	-
26	svMP; CTL	85	85	svMP	-
27	svMP; CTL	90	89	unknown protein	-
-	-	-	95	unknown protein	-
28	svMP; CTL	97	97	CTL	-
-	-	-	100	CTL	-
29	svMP	101	101	svMP	-
30	svMP	102	103	svMP	-
31	svMP	104	-	-	-
32	svMP	104	104	svMP	-
33	svMP; CTL	105	105	unknown protein	-
34	svMP; CTL	107	108	svMP	-
35	svMP	108	109	unknown protein	-
36	svMP; CTL	110	-	-	-

References

[1] a) A. T. Tu, *Handbook of Natural Toxins*, M. Dekker, New York, Basel, **1984**; b) E. Hodgson, *A Textbook of Modern Toxicology*, Wiley-Interscience, Hoboken, N.J., **2004**.

[2] M. Carolyn Hardegree, *Handbook of Natural Toxins*, Marcel Dekker, New York, **1988**.

[3] A. T. Tu, R. F. Keeler, M. C. Hardegree, J. Moss, *Handbook of Natural Toxins*, Marcel Dekker, New York, **1983-95**.

[4] A. T. Tu (Ed.) *Handbook of Natural Toxins*, Dekker, New York, **1991**.

[5] S. M. Gwaltney-Brant in *Reproductive and Developmental Toxicology*, Elsevier, **2017**, pp. 963–972.

[6] R. Early, A. Y. Tamime in *Microbial Toxins in Dairy Products* (Ed.: A. Y. Tamime), John Wiley & Sons, Ltd, Chichester, UK, **2017**, pp. 1–18.

[7] L. Brieger, C. Fraenkel, *Berl. Klin. Wochenschrift* **1890**, *27*, 268.

[8] E. T. Rietschel, O. Westphal in *Endotoxin in Health and Disease*, Marcel Dekker, New York, **1999**, pp. 1–30.

[9] R. Pfeiffer, *Zeitschr. f. Hygiene* **1892**, *11*, 393.

[10] a) C. R. H. Raetz, C. Whitfield, *Annu. Rev. Biochem.* **2002**, *71*, 635; b) V. Sampath, *Agric. Nat. Resour.* **2018**, *52*, 115.

[11] a) A. J. Greaney, S. H. Leppla, M. Moayeri, *Front. Immunol.* **2015**, *6*, 570; b) I. Sastalla, D. M. Monack, K. F. Kubatzky, *Front. Immunol.* **2016**, *7*, 300.

[12] a) D. Dressler, F. Adib Saberi, *Eur. Neurol.* **2005**, *53*, 3; b) C. Montecucco, J. Molgó, *Curr. Opin. Pharmacol.* **2005**, *5*, 274.

[13] J. W. Bennett, M. Klich, *Clin. Microbiol. Rev.* **2003**, *16*, 497.

[14] C. R. Carlini, R. Ligabue-Braun, *Plant Toxins*, Springer Netherlands, Dordrecht, **2017**.

[15] P. J. T. Morris, *FromCclassical to Modern Chemistry*. *The instrumental revolution*, Royal Society of Chemistry, Cambridge, **2002**.

[16] M. I. Karayannis, C. E. Efstathiou, *Talanta* **2012**, *102*, 7.

[17] a) K. K. Maggon, S. K. Gupta, T. A. Venkitasubramanian, *Bacteriol. Rev.* **1977**, *41*, 822; b) A. F. M. Botelho, F. Pierezan, B. Soto-Blanco, M. M. Melo, *Toxicon* **2019**, *158*, 63.

[18] E. Hodgson in *A Textbook of Modern Toxicology* (Ed.: E. Hodgson), John Wiley & Sons, Inc, Hoboken, NJ, USA, **2004**, pp. 1–12.

[19] a) S. P. Mackessy, *Toxicon* **2010**, *55*, 1463; b) B. Fry, *Venomous Reptiles and Their Toxins*. *Evolution, Pathophysiology, and Biodiscovery*, Oxford University Press, New York, NY, **2015**.

[20] a) D. A. Henderson, *Science* **1999**, *283*, 1279; b) N. J. Beeching, D. A. B. Dance, A. R. O. Miller, R. C. Spencer, *Br. Med. J.* **2002**, *324*, 336.

[21] E. Nepovimova, K. Kuca, *Arch. Toxicol.* **2019**, *93*, 11.

[22] a) T. Reynolds, *Phytochemistry* **2005**, *66*, 1399; b) H. Hotti, H. Rischer, *Molecules* **2017**, *22*, 1962; c) A. Mayor in *Toxicology in Antiquity* (Ed.: Philip Wexler), Elsevier, **2019**, pp. 161–174; d) A. Touwaide in *Toxicology in Antiquity* (Ed.: Philip Wexler), Elsevier, **2019**, pp. 131–139.

[23] J. Emsley, *Molecules of Murder. Criminal Molecules and Classic Cases*, Royal Society of Chemistry, **2016**.

[24] a) K. I. Bos, A. Herbig, J. Sahl, N. Waglechner, M. Fourment, S. A. Forrest, J. Klunk, V. J. Schuenemann, D. Poinar, M. Kuch et al., *eLife* **2016**, *5*, e12994; b) M. A. Spyrou, R. I. Tukhbatova, M. Feldman, J. Drath, S. Kacki, J. Beltrán de Heredia, S. Arnold, A. G. Sitdikov, D. Castex, J. Wahl et al., *Cell Host Microbe* **2016**, *19*, 874; c) S. D. Brown, T. C. Montie, *Infect. Immun.* **1977**, *18*, 85; d) B. J. Hinnebusch, A. E. Rudolph, P. Cherepanov, J. E. Dixon, T. G. Schwan, A. Forsberg, *Science* **2002**, *296*, 733; e) K. Kawahara, H. Tsukano, H. Watanabe, B. Lindner, M. Matsuura, *Infect. Immun.* **2002**, *70*, 4092.

[25] J. I. Pitt, J. D. Miller, *J. Agric. Food Chem.* **2017**, *65*, 7021.

[26] a) C. L. Schardl, D. G. Panaccione, P. Tudzynski, *Alkaloids Chem. Biol.* **2006**, *63*, 45; b) M. R. Lee, *J. R. Coll. Physicians Edinb.* **2009**, *39*, 179.

[27] J. Spencer Smith, W. Paul Williams, G. L. Windham, *Mycotoxin Res.* **2019**, *35*, 111.

[28] a) R. Pita, A. Romero, *Forensic Sci. Rev.* **2014**, *26*, 85; b) V. Pitschmann, Z. Hon, *Molecules* **2016**, *21*, 556.

[29] a) G. B. Brown, Y. H. Kim, H. Küntzel, H. S. Mosher, *Toxicon* **1977**, *15*, 115; b) N. G. Bisset, *J. Ethnopharmacol.* **1989**, *25*, 1.

[30] A. Kasturiratne, A. R. Wickremasinghe, N. de Silva, N. K. Gunawardena, A. Pathmeswaran, R. Premaratna, L. Savioli, D. G. Lalloo, H. J. de Silva, *PLoS Med.* **2008**, *5*, e218.

[31] R. A. Harrison, A. Hargreaves, S. C. Wagstaff, B. Faragher, D. G. Lalloo, *PLoS Negl. Trop. Dis.* **2009**, *3*, e569.

[32] D. A. Groneberg, V. Geier, D. Klingelhöfer, A. Gerber, U. Kuch, B. Kloft, *PLoS Negl. Trop. Dis.* **2016**, *10*, e0005046.

[33] J. P. Chippaux, *Bull. World Health Organ.* **1998**, *76*, 515.

[34] J. White, *Ther. Drug Monit.* **2000**, *22*, 65.

[35] *Lancet* **2019**, *393*, 2175.

[36] J.-P. Chippaux, *J. Venom. Anim. Toxins Incl. Trop. Dis.* **2017**, *23*, 38.

[37] B. G. Fry, K. Roelants, D. E. Champagne, H. Scheib, J. D. A. Tyndall, G. F. King, T. J. Nevalainen, J. A. Norman, R. J. Lewis, R. S. Norton et al., *Annu. Rev. Genomics Hum. Genet.* **2009**, *10*, 483.

References

[38]   a) M. F. Balandrin, A. D. Kinghorn, N. R. Farnsworth in *ACS Symposium Series* (Eds.: A. D. Kinghorn, M. F. Balandrin), American Chemical Society, Washington, DC, **1993**, pp. 2–12; b) J. K. Rudkin, R. M. McLoughlin, A. Preston, R. C. Massey, *PLoS Pathog.* **2017**, *13*, e1006452.

[39]   a) B. P. Zietz, H. Dunkelberg, *Int. J. Hyg. Environ. Health* **2004**, *207*, 165; b) T. M. Daniel, *Respir. Med.* **2006**, *100*, 1862; c) S. Haensch, R. Bianucci, M. Signoli, M. Rajerison, M. Schultz, S. Kacki, M. Vermunt, D. A. Weston, D. Hurst, M. Achtman et al., *PLoS Pathog.* **2010**, *6*, e1001134; d) D. Lippi, E. Gotuzzo, *Clin. Microbiol. Infect.* **2014**, *20*, 191; e) D. Lippi, E. Gotuzzo, S. Caini, *Microbiol. Spectr.* **2016**, *4*; f) I. Barberis, N. L. Bragazzi, L. Galluzzo, M. Martini, *J. Prev. Med. Hyg.* **2017**, *58*, E9-E12.

[40]   M. E. Hibbing, C. Fuqua, M. R. Parsek, S. B. Peterson, *Nat. Rev. Microbiol.* **2010**, *8*, 15.

[41]   S. Duckett, *Lancet* **1999**, *354*, 2068.

[42]   R. S. Schwartz, *N. Engl. J. Med.* **2004**, *350*, 1079.

[43]   a) N. C. Lloyd, H. W. Morgan, B. K. Nicholson, R. S. Ronimus, *Angew. Chem.* **2005**, *44*, 941; b) K. J. Williams, *J. R. Soc. Med.* **2009**, *102*, 343.

[44]   F. W. Diggins, *Br. J. Biomed. Sci.* **1999**, *56*, 83.

[45]   B. L. Ligon, *Semin. Pediatr. Infect. Dis.* **2004**, *15*, 52.

[46]   R. Gaynes, *Emerg. Infect. Dis.* **2017**, *23*, 849.

[47]   A. Fleming, *Rev. Infect. Dis.* **1980**, *2*, 129.

[48]   a) R. I. Aminov, *Front. Microbiol.* **2010**, *1*, 134; b) K. I. Mohr, *Curr. Top. Microbiol. Immunol.* **2016**, *398*, 237.

[49]   a) E. M. Costa-Neto, *An. Acad. Bras. Ciênc.* **2005**, *77*, 33; b) R. R. N. Alves, H. N. Alves, *J. Ethnobiol. Ethnomed.* **2011**, *7*, 9; c) B. B. Petrovska, *Pharmacogn. Rev.* **2012**, *6*, 1.

[50]   a) M. Giles, C. Ulbricht, K. P. S. Khalsa, C. D. Kirkwood, C. Park, E. Basch, *J. Herb. Pharmacother.* **2005**, *5*, 119; b) M. R. Lee, *J. R. Coll. Physicians Edinb.* **2007**, *37*, 77.

[51]   L. L. Brunton, J. S. Lazo, I. L. O. Buxton, D. K. Blumenthal, H. Akil, P. C. Amrein, R. J. Baldessarini, J. R. Balser, W. M. Bennett, J. E. Bennett et al., *Goodman & Gilman's the Pharmacological Basis of Therapeutics*, McGraw-Hill, New York.

[52]   R. E. Schultes, A. Hofmann, *Plants of the Gods. Origins of Hallucinogenic Use*, A. van der Marck Editions, New York, **1987**.

[53]   S. D. Meriney, E. E. Fanselow, *Synaptic Transmission*, Academic Press, London, San Diego, CA, **2019**.

[54]   a) A. J. Clark, *J. Physiol.* **1926**, *61*, 547; b) F. Steigmann, L. Kaminski, *Am. J. Dig. Dis.* **1956**, *1*, 174.

[55]   E. B. Russo, *Chem. Biodivers.* **2007**, *4*, 1614.

[56]   E. J. Brand, Z. Zhao, *Front. Pharmacol.* **2017**, *8*, 108.

[57]   A. W. Zuardi, *Rev. Bras. Psiquiatr.* **2006**, *28*, 153.

[58]   M. Ren, Z. Tang, X. Wu, R. Spengler, H. Jiang, Y. Yang, N. Boivin, *Sci. Adv.* **2019**, *5*, e1391.

[59]   a) H. Wright, *Am. J. Inter. Law* **1912**, *6*, 865; b) A. L. Tennyson, *Am. J. Med.* **1953**, *14*, 578; c) H. Carliner, Q. L. Brown, A. L. Sarvet, D. S. Hasin, *Prev. Med.* **2017**, *104*, 13.

[60]   A. Novotna, J. Mares, S. Ratcliffe, I. Novakova, M. Vachova, O. Zapletalova, C. Gasperini, C. Pozzilli, L. Cefaro, G. Comi et al., *Eur. J. Neurol.* **2011**, *18*, 1122.

[61]   Y. N. Utkin, *World J. Biol. Chem.* **2015**, *6*, 28.

[62]   G. F. King, *Expert Opin. Biol. Ther.* **2011**, *11*, 1469.

[63]   a) Y. S. Bakhle, *Nature* **1968**, *220*, 919; b) D. W. Cushman, M. A. Ondetti, *Hypertension* **1991**, *17*, 589.

[64]   T. Paracelsus (Ed.) *Paragranum*, Darmstadt, **1538**.

[65]   L. J. Casarett, J. Doull, *Casarett and Doull's Toxicology*. *The Basic Science of Poisons*, McGraw-Hill Education, New York, **2013**.

[66]   a) Q.-Y. He, Q.-Z. He, X.-C. Deng, L. Yao, E. Meng, Z.-H. Liu, S.-P. Liang, *Nucleic Acids Res.* **2008**, *36*, D293-7; b) R. Judson, A. Richard, D. Dix, K. Houck, F. Elloumi, M. Martin, T. Cathey, T. R. Transue, R. Spencer, M. Wolf, *Toxicol. Appl. Pharmacol.* **2008**, *233*, 7; c) A. P. Davis, C. G. Murphy, C. A. Saraceni-Richards, M. C. Rosenstein, T. C. Wiegers, C. J. Mattingly, *Nucleic Acids Res.* **2009**, *37*, D786-92; d) U. Schmidt, S. Struck, B. Gruening, J. Hossbach, I. S. Jaeger, R. Parol, U. Lindequist, E. Teuscher, R. Preissner, *Nucleic Acids Res.* **2009**, *37*, D295-9; e) E. Lim, A. Pon, Y. Djoumbou, C. Knox, S. Shrivastava, A. C. Guo, V. Neveu, D. S. Wishart, *Nucleic Acids Res.* **2010**, *38*, D781-6.

[67]   J. Sumner, *The Natural History of Medicinal Plants*, Timber Press, Portland, Or., **2000**.

[68]   H. P. Rang, M. M. Dale, *Pharmacology*, Churchill Livingstone, Elsevier, Phildelphia, PA, **2007**.

[69]   H. P. Rang, *Br. J. Pharmacol.* **2006**, *147 Suppl 1*, S9-16.

[70]   a) K. B. Mullis, F. A. Faloona, *Meth. Enzymol.* **1987**, *155*, 335; b) G. S. Sittampalam, S. D. Kahl, W. P. Janzen, *Curr. Opin. Chem. Biol.* **1997**, *1*, 384; c) L. A. A. de Jong, D. R. A. Uges, J. P. Franke, R. Bischoff, *J. Chromatogr. B* **2005**, *829*, 1.

[71]   B. I. Escher, J. L. M. Hermens, *Environ. Sci. Technol.* **2002**, *36*, 4201.

[72]   B. I. Escher, R. Ashauer, S. Dyer, J. L. M. Hermens, J.-H. Lee, H. A. Leslie, P. Mayer, J. P. Meador, M. S. J. Warne, *Integr. Environ. Assess. Manag.* **2011**, *7*, 28.

[73]   a) D. Perlman, *Hindustan. Antibiot. Bull.* **1966**, *8*, 175; b) T. Oki, *Biotechnol. Bioeng.* **1980**, *22 Suppl 1*, 83.

[74]   a) T. Wieland, *Int. J. Pept. Protein Res.* **1983**, *22*, 257; b) T. Wieland, H. Faulstich, *Crit. Rev. Biochem.* **1978**, *5*, 185; c) J. D. Walton, H. E. Hallen-Adams, H. Luo, *Biopolymers* **2010**, *94*, 659; d) M.-A. J. Siegert, C. H. Knittel, R. D. Süssmuth, *Angew. Chem.* **2020**, *59*, 5500.

[75]   a) K. Alam, A. Crowe, X. Wang, P. Zhang, K. Ding, L. Li, W. Yue, *Int. J. Mol.* **2018**, *19*; b) A. Pahl, C. Lutz, T. Hechler, *Drug Discov. Today Technol.* **2018**, *30*, 85.

[76]   J. M. Lord, L. M. Roberts in *Topics in Current Genetics* (Eds.: M. J. Schmitt, R. Schaffrath), Springer Berlin Heidelberg, Berlin, Heidelberg, **2005**, pp. 215–233.

[77]   J. Audi, M. Belson, M. Patel, J. Schier, J. Osterloh, *JAMA.* **2005**, *294*, 2342.

[78]   A. C. Ferraz, M. E. Angelucci, M. L. Da Costa, I. R. Batista, B. H. de Oliveira, C. Da Cunha, *Pharmacol. Biochem. Behav.* **1999**, *63*, 367.

[79]   a) H. T. Wright, J. D. Robertus, *Arch. Biochem. Biophys.* **1987**, *256*, 280; b) S. A. Weston, A. D. Tucker, D. R. Thatcher, D. J. Derbyshire, R. A. Pauptit, *J. Mol. Biol.* **1994**, *244*, 410; c) R. Wales, P. T. Richardson, L. M. Roberts, H. R. Woodland, J. M. Lord, *J. Biol. Chem.* **1991**, *266*, 19172.

[80]   P. V. Dubovskii, D. M. Lesovoy, M. A. Dubinnyi, A. G. Konshina, Y. N. Utkin, R. G. Efremov, A. S. Arseniev, *Biochem. J.* **2005**, *387*, 807.

[81]   a) P. V. Dubovskii, A. G. Konshina, R. G. Efremov, *Curr. Med. Chem.* **2014**, *21*, 270; b) S. E. Gasanov, I. H. Shrivastava, F. S. Israilov, A. A. Kim, K. A. Rylova, B. Zhang, R. K. Dagda, *PLoS One* **2015**, *10*, e0129248.

[82]   J. M. Gutiérrez, T. Escalante, A. Rucavado, C. Herrera, *Toxins* **2016**, *8*, 93.

[83]   J. Slagboom, J. Kool, R. A. Harrison, N. R. Casewell, *Br. J. Haematol.* **2017**, *177*, 947.

[84]   S. P. Mackessy, *Handbook of Venoms and Toxins of Reptiles*, CRC Press, Boca Raton, Fla., **2010**.

[85]   a) J. M. Gutiérrez, A. Rucavado, *Biochimie* **2000**, *82*, 841; b) A. Cortelazzo, R. Guerranti, L. Bini, N. Hope-Onyekwere, C. Muzzi, R. Leoncini, R. Pagani, *Blood Transfus.* **2010**, *8 Suppl 3*, s120-5.

[86]   N. Oguiura, M. Boni-Mitake, G. Rádis-Baptista, *Toxicon* **2005**, *46*, 363.

[87]   a) M. Raffray, G. M. Cohen, *Pharmacol. Ther.* **1997**, *75*, 153; b) S. Y. Proskuryakov, A. G. Konoplyannikov, V. L. Gabai, *Exp. Cell Res.* **2003**, *283*, 1.

[88]   a) G. Majno, I. Joris, *Am. J. Pathol.* **1995**, *146*, 3; b) A. L. Edinger, C. B. Thompson, *Curr. Opin. Cell Biol.* **2004**, *16*, 663.

[89]   a) R. M. Kini, *Toxicon* **2003**, *42*, 827; b) S. M. T. Serrano, R. C. Maroun, *Toxicon* **2005**, *45*, 1115.

[90]   a) R. J. Green, D. C. Dafoe, T. A. Raffin, *Chest* **1996**, *110*, 219; b) S. Paz Maya, D. Dualde Beltrán, P. Lemercier, C. Leiva-Salinas, *Skeletal Radiol.* **2014**, *43*, 577.

[91]   A. Giuliano, F. Lewis, K. Hadley, F. W. Blaisdell, *Am. J. Surg.* **1977**, *134*, 52.

[92]   T. M. File, J. S. Tan, J. R. DiPersio, *Cleve. Clin. J. Med.* **1998**, *65*, 241.

[93]   G. P. Carter, M. M. Awad, Y. Hao, T. Thelen, I. L. Bergin, P. M. Howarth, T. Seemann, J. I. Rood, D. M. Aronoff, D. Lyras, *Infect. Immun.* **2011**, *79*, 1025.

[94]   J. C. M. Brust, *Int. J. Environ. Res. Public Health* **2010**, *7*, 1540.

[95]   A. Lau, M. Tymianski, *Pflug. Arch. Eur. J. Phy.* **2010**, *460*, 525.

[96]   V. Calabrese, C. Mancuso, M. Calvani, E. Rizzarelli, D. A. Butterfield, A. M. G. Stella, *Nat. Rev. Neurosci.* **2007**, *8*, 766.

[97]   M. Pirazzini, O. Rossetto, R. Eleopra, C. Montecucco, *Pharmacol. Rev.* **2017**, *69*, 200.

[98]   G. Schiavo, E. Papini, G. Genna, C. Montecucco, *Infect. Immun.* **1990**, *58*, 4136.

[99]   J. Lago, L. P. Rodríguez, L. Blanco, J. M. Vieites, A. G. Cabado, *Mar. Drugs* **2015**, *13*, 6384.

[100]   a) E. Katoh, H. Nishio, T. Inui, Y. Nishiuchi, T. Kimura, S. Sakakibara, T. Yamazaki, *Biopolymers* **2000**, *54*, 44; b) W. C. Hodgson, J. C. Wickramaratna, *Clin. Exp. Pharmacol. Physiol.* **2002**, *29*, 807.

[101]   a) S. S. Arnon, R. Schechter, T. V. Inglesby, D. A. Henderson, J. G. Bartlett, M. S. Ascher, E. Eitzen, A. D. Fine, J. Hauer, M. Layton et al., *JAMA.* **2001**, *285*, 1059; b) M. C. Kiernan, G. K. Isbister, C. S.-Y. Lin, D. Burke, H. Bostock, *Ann. Neurol.* **2005**, *57*, 339.

[102]   Y. Wu, H. Ma, F. Zhang, C. Zhang, X. Zou, Z. Cao, *ACS Chem. Neurosci.* **2018**, *9*, 187.

[103]   J. Kalia, M. Milescu, J. Salvatierra, J. Wagner, J. K. Klint, G. F. King, B. M. Olivera, F. Bosmans, *J. Mol. Biol.* **2015**, *427*, 158.

[104]   S. Bajaj, J. Han, *Front. Pharmacol.* **2019**, *10*, 58.

[105]   S. A. Ali, M. Alam, A. Abbasi, E. A. B. Undheim, B. G. Fry, H. Kalbacher, W. Voelter, *Toxins* **2016**, *8*, 36.

[106]   D. P. Figgitt, S. Noble, *Drugs* **2002**, *62*, 705.

[107]   D. Dressler, F. A. Saberi, E. R. Barbosa, *Arq. Neuropsiquiatr.* **2005**, *63*, 180.

[108]   A. M. G. Osman, A. G. Chittiboyina, I. A. Khan in *Foodborne Infections and Intoxications*, Elsevier, **2013**, pp. 435–451.

[109]   a) J. Bakker, F. J. Gommers, I. Nieuwenhuis, H. Wynberg, *J. Biol. Chem.* **1979**, *254*, 1841; b) W. M. Rampone, J. L. McCullough, G. D. Weinstein, G. H. Towers, M. W. Berns, B. Abeysekera, *J. Invest. Dermatol.* **1986**, *87*, 354.

[110]   L. S. McCarty, D. Mackay, *Environ. Sci. Technol.* **1993**, *27*, 1718.

[111]   a) S. Perea, T. F. Patterson, *Clin. Infect. Dis.* **2002**, *35*, 1073; b) L. Svabova, A. Lebeda, *J. Phytopathol.* **2005**, *153*, 52; c) F. C. Tenover, *Am. J. Med.* **2006**, *119*, S3-10; discussion S62-70; d) L. Després, J.-P. David, C. Gallet, *Trends Ecol. Evol.* **2007**, *22*, 298; e) M. L. Holding, D. H. Drabeck, S. A. Jansa, H. L. Gibbs, *Integr. Comp. Biol.* **2016**, *56*, 1032; f) K. Arbuckle, R. C. La Rodríguez de Vega, N. R. Casewell, *Toxicon* **2017**, *140*, 118.

[112]   M. Balass, E. Katchalski-Katzir, S. Fuchs, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6054.

[113]   a) D. Barchan, S. Kachalsky, D. Neumann, Z. Vogel, M. Ovadia, E. Kochva, S. Fuchs, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7717; b) O. Asher, M. Lupu-Meiri, B. S. Jensen, T. Paperna, S. Fuchs, Y. Oron, *FEBS Lett.* **1998**, *431*, 411.

[114]   a) D. Neumann, D. Barchan, M. Horowitz, E. Kochva, S. Fuchs, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 7255; b) A. Malhotra (Ed.) *Toxinology*, Springer Netherlands, Dordrecht, **2017**.

[115]   a) J. C. Perez, W. C. Haws, C. H. Hatch, *Toxicon* **1978**, *16*, 198; b) J. C. Perez, S. Pichyangkul, V. E. Garcia, *Toxicon* **1979**, *17*, 601; c) C. A. de Wit, *Toxicon* **1982**, *20*, 709; d) V. E. Garcia, J. C. Perez, *Toxicon* **1984**, *22*, 129.

[116]   H. Kato, S. Iwanaga, T. Suzuki, *Experientia* **1966**, *22*, 49.

[117]   K. F. Huang, C. C. Hung, S. H. Wu, S. H. Chiou, *Biochem. Biophys. Res. Commun.* **1998**, *248*, 562.

[118]   K.-F. Huang, S.-H. Chiou, T.-P. Ko, A. H.-J. Wang, *Eur. J. Biochem.* **2002**, *269*, 3047.

[119]   J. E. Biardi, C. Y. L. Ho, J. Marcinczyk, K. P. Nambiar, *Toxicon* **2011**, *58*, 486.

[120]   J. E. Biardi, R. G. Coss, *Toxicon* **2011**, *57*, 323.

[121]   W. H. Organization, *Antimicrobial Resistance. Global report on surveillance 2014*, World Health Organization, Geneva, **2014**.

[122]   G. M. Rossolini, F. Arena, P. Pecile, S. Pollini, *Curr. Opin. Pharmacol.* **2014**, *18*, 56.

[123]   C. L. Ventola, *Pharm. Ther.* **2015**, *40*, 277.

[124]   a) J. L. Martinez, F. Baquero, *Antimicrob. Agents Chemother.* **2000**, *44*, 1771; b) N. Woodford, M. J. Ellington, *Clin. Microbiol. Infect.* **2007**, *13*, 5; c) C. J. H. von Wintersdorff, J. Penders, J. M. van Niekerk, N. D. Mills, S. Majumder, L. B. van Alphen, P. H. M. Savelkoul, P. F. G. Wolffs, *Front. Microbiol.* **2016**, *7*, 173.

[125]   T. Vellai, G. Vida, *Proc. Biol. Sci.* **1999**, *266*, 1571.

[126]   a) C. A. Michael, D. Dominey-Howes, M. Labbate, *Front. Public Health* **2014**, *2*, 145; b) T. P. van Boeckel, C. Brower, M. Gilbert, B. T. Grenfell, S. A. Levin, T. P. Robinson, A. Teillant, R. Laxminarayan, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5649; c) B. Aslam, W. Wang, M. I. Arshad, M. Khurshid, S. Muzammil, M. H. Rasool, M. A. Nisar, R. F. Alvi, M. A. Aslam, M. U. Qamar et al., *Infect. Drug Resist.* **2018**, *11*, 1645.

[127]   S. Santajit, N. Indrawattana, *Biomed Res. Int.* **2016**, *2016*, 2475067.

[128]   a) W. Yang, I. F. Moore, K. P. Koteva, D. C. Bareich, D. W. Hughes, G. D. Wright, *J. Biol. Chem.* **2004**, *279*, 52346; b) F. Nguyen, A. L. Starosta, S. Arenz, D. Sohmen, A. Dönhöfer, D. N. Wilson, *Biol. Chem.* **2014**, *395*, 559.

[129]   a) K. J. Shaw, P. N. Rather, R. S. Hare, G. H. Miller, *Microbiol. Rev.* **1993**, *57*, 138; b) E. Azucena, S. Mobashery, *Drug Resist. Updat.* **2001**, *4*, 106.

[130]   a) J. R. Knowles, *Acc. Chem. Res.* **1985**, *18*, 97; b) K. Bush, P. A. Bradford, *Cold Spring Harb. Perspect. Med.* **2016**, *6*.

[131]   a) Y. Nakajima, *J. Infect. Chemother.* **1999**, *5*, 61; b) R. Leclercq, *Clin. Infect. Dis.* **2002**, *34*, 482.

[132]   a) M. Arthur, P. Reynolds, P. Courvalin, *Trends. Microbiol.* **1996**, *4*, 401; b) P. Courvalin, *Clin. Infect. Dis.* **2006**, *42 Suppl 1*, S25-34.

[133]   a) J. Vila, A. Fàbrega, I. Roca, A. Hernández, J. L. Martínez, *Adv. Enzymol. Relat. Areas Mol. Biol.* **2011**, *77*, 167; b) H. Nikaido, J.-M. Pagès, *FEMS Microbiol. Rev.* **2012**, *36*, 340.

[134]   a) A. Coates, Y. Hu, R. Bax, C. Page, *Nat. Rev. Drug. Discov.* **2002**, *1*, 895; b) A. R. M. Coates, Y. Hu, *Br. J. Pharmacol.* **2007**, *152*, 1147.

[135]   a) L. B. Rice, *Infect. Control Hosp. Epidemiol.* **2010**, *31 Suppl 1*, S7-10; b) J. N. Pendleton, S. P. Gorman, B. F. Gilmore, *Expert Rev. Anti Infect. Ther.* **2013**, *11*, 297.

[136]   U. Theuretzbacher, K. Bush, S. Harbarth, M. Paul, J. H. Rex, E. Tacconelli, G. E. Thwaites, *Nat. Rev. Microbiol.* **2020**, *18*, 286.

[137]   a) K. Poole, *J. Pharm. Pharmacol.* **2001**, *53*, 283; b) M. S. Mulani, E. E. Kamble, S. N. Kumkar, M. S. Tawre, K. R. Pardesi, *Front. Microbiol.* **2019**, *10*, 539.

[138]   D. J. Tipper, J. L. Strominger, *Proc. Natl. Acad. Sci. USA* **1965**, *54*, 1133.

[139]   D. J. Waxman, J. L. Strominger, *Annu. Rev. Biochem.* **1983**, *52*, 825.

[140]   a) M. P. Jevons, *Br. Med. J.* **1961**, *1*, 124; b) M. C. Enright, D. A. Robinson, G. Randle, E. J. Feil, H. Grundmann, B. G. Spratt, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7687.

[141]   a) C. Hansch, A. R. Steward, *J. Med. Chem.* **1964**, *7*, 691; b) A. Gin, L. Dilay, J. A. Karlowsky, A. Walkty, E. Rubinstein, G. G. Zhanel, *Expert Rev. Anti Infect. Ther.* **2007**, *5*, 365; c) Z. Ashraf, A. Bais, M. M. Manir, U. Niazi, *PLoS One* **2015**, *10*, e0135293.

[142]   a) H. S. Gold, R. C. Moellering, *N. Engl. J. Med.* **1996**, *335*, 1445; b) R. J. Worthington, C. Melander, *J. Org. Chem.* **2013**, *78*, 4207.

[143]   a) M. A. Kohanski, D. J. Dwyer, J. J. Collins, *Nat. Rev. Microbiol.* **2010**, *8*, 423; b) P. Murima, J. D. McKinney, K. Pethe, *Chem. Biol.* **2014**, *21*, 1423.

[144]   S. M. Hashimi, *J. Antibiot.* **2019**, *72*, 785.

[145]   a) R. G. Birch, S. S. Patil, *Physiol. Mol. Plant. P.* **1987**, *30*, 199; b) R. G. Birch, S. S. Patil, *Physiol. Mol. Plant. P.* **1987**, *30*, 207.

[146]   S. M. Hashimi, M. K. Wall, A. B. Smith, A. Maxwell, R. G. Birch, *Antimicrob. Agents Chemother.* **2007**, *51*, 181.

[147]   G. Huang, L. Zhang, R. G. Birch, *Gene* **2000**, *258*, 193.

[148]   G. Huang, L. Zhang, R. G. Birch, *Microbiology* **2001**, *147*, 631.

[149]   S. Cociancich, A. Pesic, D. Petras, S. Uhlmann, J. Kretz, V. Schubert, L. Vieweg, S. Duplan, M. Marguerettaz, J. Noëll et al., *Nat. Chem. Biol.* **2015**, *11*, 195.

[150]   S. Baumann, J. Herrmann, R. Raju, H. Steinmetz, K. I. Mohr, S. Hüttel, K. Harmrolfs, M. Stadler, R. Müller, *Angew. Chem.* **2014**, *53*, 14605.

[151]   Y. J. Kim, H.-J. Kim, G.-W. Kim, K. Cho, S. Takahashi, H. Koshino, W.-G. Kim, *J. Nat. Prod.* **2016**, *79*, 2223.

[152]   M. J. Walker, R. G. Birch, J. M. Pemberton, *Mol. Microbiol.* **1988**, *2*, 443.

[153]   a) L. Zhang, J. Xu, R. G. Birch, *Microbiology* **1998**, *144 (Pt 2)*, 555; b) L.-X. Weng, L.-H. Wang, J.-L. Xu, J.-E. Wu, Q. Li, L.-H. Zhang, *Appl. Environ. Microbiol.* **2005**, *71*, 1445.

[154]   L. Rostock, R. Driller, S. Grätz, D. Kerwat, L. von Eckardstein, D. Petras, M. Kunert, C. Alings, F.-J. Schmitt, T. Friedrich et al., *Nat. Commun.* **2018**, *9*, 3095.

[155]   L. Zhang, R. G. Birch, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 9984.

[156]   L. Vieweg, J. Kretz, A. Pesic, D. Kerwat, S. Grätz, M. Royer, S. Cociancich, A. Mainz, R. D. Süssmuth, *J. Am. Chem. Soc.* **2015**, *137*, 7608.

[157]   a) S. Grätz, D. Kerwat, J. Kretz, L. von Eckardstein, S. Semsary, M. Seidel, M. Kunert, J. B. Weston, R. D. Süssmuth, *ChemMedChem* **2016**, *11*, 1499; b) I. Behroz, P. Durkin, S. Grätz, M. Seidel, L. Rostock, M. Spinczyk, J. B. Weston, R. D. Süssmuth, *Chem. Eur. J.* **2019**.

[158]   D. Kerwat, S. Grätz, J. Kretz, M. Seidel, M. Kunert, J. B. Weston, R. D. Süssmuth, *ChemMedChem* **2016**, *11*, 1899.

[159]   L. von Eckardstein, D. Petras, T. Dang, S. Cociancich, S. Sabri, S. Grätz, D. Kerwat, M. Seidel, A. Pesic, P. C. Dorrestein et al., *Chem. Eur. J.* **2017**, *23*, 15316.

[160]   A. Okada, M. Igarashi, T. Okajima, N. Kinoshita, M. Umekita, R. Sawa, K. Inoue, T. Watanabe, A. Doi, A. Martin et al., *J. Antibiot.* **2010**, *63*, 89.

[161]   T. Watanabe, M. Igarashi, T. Okajima, E. Ishii, H. Kino, M. Hatano, R. Sawa, M. Umekita, T. Kimura, S. Okamoto et al., *Antimicrob. Agents Chemother.* **2012**, *56*, 3657.

[162]   M. Igarashi, T. Watanabe, T. Hashida, M. Umekita, M. Hatano, Y. Yanagida, H. Kino, T. Kimura, N. Kinoshita, K. Inoue et al., *J. Antibiot.* **2013**, *66*, 459.

[163]   Y. Eguchi, T. Okajima, N. Tochio, Y. Inukai, R. Shimizu, S. Ueda, S. Shinya, T. Kigawa, T. Fukamizo, M. Igarashi et al., *J. Antibiot.* **2017**, *70*, 251.

[164]   G. Murakami, T. Sato, T. Takiguchi, *Arch. Histol. Cytol.* **1990**, *53 Suppl*, 219.

[165]   a) S. B. Singh, H. Jayasuriya, J. G. Ondeyka, K. B. Herath, C. Zhang, D. L. Zink, N. N. Tsou, R. G. Ball, A. Basilio, O. Genilloud et al., *J. Am. Chem. Soc.* **2006**, *128*, 11916; b) E. Martens, A. L. Demain, *J. Antibiot.* **2011**, *64*, 705.

[166]   a) S. W. White, J. Zheng, Y.-M. Zhang, Rock, *Annu. Rev. Biochem.* **2005**, *74*, 791; b) Y.-M. Zhang, S. W. White, C. O. Rock, *J. Biol. Chem.* **2006**, *281*, 17541.

[167]   J. Wang, S. Kodali, S. H. Lee, A. Galgoci, R. Painter, K. Dorso, F. Racine, M. Motyl, L. Hernandez, E. Tinney et al., *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 7612.

[168]   J. Wang, S. M. Soisson, K. Young, W. Shoop, S. Kodali, A. Galgoci, R. Painter, G. Parthasarathy, Y. S. Tang, R. Cummings et al., *Nature* **2006**, *441*, 358.

[169]   J. R. Zabrecky, T. Lam, S. J. McKenzie, W. Carney, *J. Biol. Chem.* **1991**, *266*, 1716.

[170]   A. Menetrey, A. Janin, J. Pullman, J. S. Overcash, A. Haouala, F. Leylavergne, L. Turbe, F. Wittke, V. Nicolas-Métral, *Antimicrob. Agents Chemother.* **2019**, *63*.

[171]   N. Chen, S. Xu, Y. Zhang, F. Wang, *Biophys. Rep.* **2018**, *4*, 233.

[172]   M. W. Pennington, A. Czerwinski, R. S. Norton, *Bioorg. Med. Chem.* **2018**, *26*, 2738.

[173]   S. H. Ferreira, D. C. Bartelt, L. J. Greene, *Biochemistry* **1970**, *9*, 2583.

[174]   a) P. Zimmet, D. Magliano, Y. Matsuzawa, G. Alberti, J. Shaw, *J. Atheroscler. Thromb.* **2005**, *12*, 295; b) V. S. Malik, B. M. Popkin, G. A. Bray, J.-P. Després, F. B. Hu, *Circulation* **2010**, *121*, 1356.

[175] a) L. Heilbronn, S. R. Smith, E. Ravussin, *Int. J. Obes. Relat. Metab. Disord.* **2004**, *28 Suppl 4*, S12-21; b) R. Taylor, *Diabetes* **2012**, *61*, 778.

[176] a) J. C. N. Chan, V. Malik, W. Jia, T. Kadowaki, C. S. Yajnik, K.-H. Yoon, F. B. Hu, *JAMA.* **2009**, *301*, 2129; b) F. B. Hu, *Diabetes Care* **2011**, *34*, 1249.

[177] a) J. Eng, W. A. Kleinman, L. Singh, G. Singh, J. P. Raufman, *J. Biol. Chem.* **1992**, *267*, 7402; b) R. Göke, H. C. Fehmann, T. Linn, H. Schmidt, M. Krause, J. Eng, B. Göke, *J. Biol. Chem.* **1993**, *268*, 19650.

[178] a) J. J. Holst, *Diabetes Metab. Res. Rev.* **2002**, *18*, 430; b) L. L. Nielsen, A. A. Young, D. G. Parkes, *Regul. Pept.* **2004**, *117*, 77; c) B. L. Furman, *Toxicon* **2012**, *59*, 464.

[179] J. M. de Souza, B. D. C. Goncalves, M. V. Gomez, L. B. Vieira, F. M. Ribeiro, *Front. Pharmacol.* **2018**, *9*, 145.

[180] a) A. L. Harvey, E. Karlsson, *Naunyn. Schmiedebergs Arch. Pharmacol.* **1980**, *312*, 1; b) W. R. Gray, A. Luque, B. M. Olivera, J. Barrett, L. J. Cruz, *J. Biol. Chem.* **1981**, *256*, 4734; c) R. A. Myers, G. C. Zafaralla, W. R. Gray, J. Abbott, L. J. Cruz, B. M. Olivera, *Biochemistry* **1991**, *30*, 9370; d) M. Harel, G. J. Kleywegt, R. B. Ravelli, I. Silman, J. L. Sussman, *Structure* **1995**, *3*, 1355.

[181] a) N. Satkunanathan, B. Livett, K. Gayler, D. Sandall, J. Down, Z. Khalil, *Brain Res.* **2005**, *1059*, 149; b) B. G. Livett, D. W. Sandall, D. Keays, J. Down, K. R. Gayler, N. Satkunanathan, Z. Khalil, *Toxicon* **2006**, *48*, 810.

[182] a) H. R. Knapp, *World Rev. Nutr. Diet.* **1991**, *66*, 313; b) J. Castro, L. Grundy, A. Deiteren, A. M. Harrington, T. O'Donnell, J. Maddern, J. Moore, S. Garcia-Caraballo, G. Y. Rychkov, R. Yu et al., *Br. J. Pharmacol.* **2018**, *175*, 2384.

[183] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, A. Jemal, *CA Cancer J. Clin.* **2018**, *68*, 394.

[184] J. A. DeBin, G. R. Strichartz, *Toxicon* **1991**, *29*, 1403.

[185] a) L. Soroceanu, Y. Gillespie, M. B. Khazaeli, H. Sontheimer, *Cancer Res.* **1998**, *58*, 4871; b) J. Deshane, C. C. Garner, H. Sontheimer, *J. Biol. Chem.* **2003**, *278*, 4135.

[186] a) C. Sun, C. Fang, Z. Stephen, O. Veiseh, S. Hansen, D. Lee, R. G. Ellenbogen, J. Olson, M. Zhang, *Nanomedicine* **2008**, *3*, 495; b) J. Wan, X. Meng, E. Liu, K. Chen, *Nanotechnology* **2010**, *21*, 235104; c) S. Chen, M. Ahmadiantehrani, N. G. Publicover, K. W. Hunter, X. Zhu, *RSC Adv.* **2015**, *74*, 60612.

[187] a) D. C. Hockaday, S. Shen, J. Fiveash, A. Raubitschek, D. Colcher, A. Liu, V. Alvarez, A. N. Mamelak, *J. Nucl. Med.* **2005**, *46*, 580; b) S. Shen, M. B. Khazaeli, G. Y. Gillespie, V. L. Alvarez, *J. Neuro-Oncol.* **2005**, *71*, 113.

[188] M. Kita, Y. Nakamura, Y. Okumura, S. D. Ohdachi, Y. Oba, M. Yoshikuni, H. Kido, D. Uemura, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7542.

[189] a) C. V. Bowen, D. DeBay, H. S. Ewart, P. Gallant, S. Gormley, T. T. Ilenchuk, U. Iqbal, T. Lutes, M. Martina, G. Mealing et al., *PLoS One* **2013**, *8*, e58866; b) S. Fu, H. Hirte, S. Welch, T. T.

Ilenchuk, T. Lutes, C. Rice, N. Fields, A. Nemet, D. Dugourd, S. Piha-Paul et al., *Invest. New Drugs* **2017**, *35*, 324; c) H. Xue, Y. Wang, T. J. MacCormack, T. Lutes, C. Rice, M. Davey, D. Dugourd, T. T. Ilenchuk, J. M. Stewart, *J. Cancer* **2018**, *9*, 3196.

[190] a) P. W. Latham, *Nat. Biotechnol.* **1999**, *17*, 755; b) A. K. Sato, M. Viswanathan, R. B. Kent, C. R. Wood, *Curr. Opin. Biotechnol.* **2006**, *17*, 638; c) A. Henninot, J. C. Collins, J. M. Nuss, *J. Med. Chem.* **2018**, *61*, 1382.

[191] W. I. Stephen, *Analyst* **1977**, *102*, 793.

[192] a) J. Bonjoch, D. Solé, *Chem. Rev.* **2000**, *100*, 3455; b) R. Chau, J. A. Kalaitzis, B. A. Neilan, *Aquat. Toxicol.* **2011**, *104*, 61; c) K. Brook, J. Bennett, S. P. Desai, *J. Anesth. Hist.* **2017**, *3*, 50.

[193] F. W. McLafferty, *Annu. Rev. Anal. Chem.* **2011**, *4*, 1.

[194] A.-H. M. Emwas, *Methods Mol. Biol.* **2015**, *1277*, 161.

[195] Shukla, Futrell, *J. Mass Spectrom.* **2000**, *35*, 1069.

[196] a) M. Marchetti-Deschmann, G. Allmaier, *J. Proteomics* **2011**, *74*, 915; b) M.-S. Kim, S. M. Pinto, D. Getnet, R. S. Nirujogi, S. S. Manda, R. Chaerkady, A. K. Madugundu, D. S. Kelkar, R. Isserlin, S. Jain et al., *Nature* **2014**, *509*, 575; c) V. U. Weiss, J. Z. Bereszczazk, M. Havlik, P. Kallinger, I. Gösler, M. Kumar, D. Blaas, M. Marchetti-Deschmann, A. J. R. Heck, W. W. Szymanski et al., *Anal. Chem.* **2015**, *87*, 8709; d) D. Sviben, D. Forcic, B. Halassy, G. Allmaier, M. Marchetti-Deschmann, M. Brgles, *Virol. J.* **2018**, *15*, 160; e) M. Holzlechner, K. Strasser, E. Zareva, L. Steinhäuser, H. Birnleitner, A. Beer, M. Bergmann, R. Oehler, M. Marchetti-Deschmann, *J. Proteome Res.* **2017**, *16*, 65; f) J. Chalupová, M. Sedlářová, M. Helmel, P. Rehulka, M. Marchetti-Deschmann, G. Allmaier, M. Sebela, *J. Mass Spectrom.* **2012**, *47*, 978; g) J. Beinhauer, R. Lenobel, D. Loginov, I. Chamrád, P. Řehulka, M. Sedlářová, M. Marchetti-Deschmann, G. Allmaier, M. Šebela, *Electrophoresis* **2016**, *37*, 2940; h) H. Dong, M. Marchetti-Deschmann, W. Winkler, H. Lohninger, G. Allmaier in *NATO Science for Peace and Security Series A: Chemistry and Biology* (Ed.: J. Banoub), Springer Netherlands, Dordrecht, **2011**, pp. 47–63; i) S. J. Cordwell, A. S. Nouwens, B. J. Walsh, *Proteomics* **2001**, *1*, 461.

[197] a) A. J. R. Heck, *Nat. Methods* **2008**, *5*, 927; b) T. Fuhrer, N. Zamboni, *Curr. Opin. Biotechnol.* **2015**, *31*, 73; c) M. H. Medema, M. A. Fischbach, *Nat. Chem. Biol.* **2015**, *11*, 639; d) M. T. Henke, N. L. Kelleher, *Nat. Prod. Rep.* **2016**, *33*, 942.

[198] a) S. Cappadona, P. R. Baker, P. R. Cutillas, A. J. R. Heck, B. van Breukelen, *Amino Acids* **2012**, *43*, 1087; b) Y. Perez-Riverol, R. Wang, H. Hermjakob, M. Müller, V. Vesada, J. A. Vizcaíno, *Biochim. Biophys. Acta* **2014**, *1844*, 63.

[199] M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan et al., *Nat. Biotechnol.* **2016**, *34*, 828.

[200] J. L. Little, C. D. Cleven, S. D. Brown, *J. Am. Soc. Mass. Spectrom.* **2011**, *22*, 348.

[201] J. Watrous, P. Roach, T. Alexandrov, B. S. Heath, J. Y. Yang, R. D. Kersten, M. van der Voort, K. Pogliano, H. Gross, J. M. Raaijmakers et al., *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E1743-52.

[202]   P. Lössl, M. van de Waterbeemd, A. Heck, JR, *EMBO J.* **2016**, *35*, 2634.

[203]   a) G. R. Hilton, J. L. P. Benesch, *J. R. Soc. Interface* **2012**, *9*, 801; b) A. C. Leney, A. J. R. Heck, *J. Am. Soc. Mass. Spectrom.* **2017**, *28*, 5.

[204]   a) M. M. Young, N. Tang, J. C. Hempel, C. M. Oshiro, E. W. Taylor, I. D. Kuntz, B. W. Gibson, G. Dollinger, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5802; b) A. Leitner, M. Faini, F. Stengel, R. Aebersold, *Trends Biochem. Sci.* **2016**, *41*, 20.

[205]   a) A. Sinz, *Mass Spectrom. Rev.* **2006**, *25*, 663; b) J. Rappsilber, *J. Struct. Biol.* **2011**, *173*, 530.

[206]   Z. A. Chen, A. Jawhari, L. Fischer, C. Buchen, S. Tahir, T. Kamenski, M. Rasmussen, L. Lariviere, J.-C. Bukowski-Wills, M. Nilges et al., *EMBO J.* **2010**, *29*, 717.

[207]   a) C. Iacobucci, C. Hage, M. Schäfer, A. Sinz, *J. Am. Soc. Mass. Spectrom.* **2017**, *28*, 2039; b) C. Iacobucci, M. Götze, C. Piotrowski, C. Arlt, A. Rehkamp, C. Ihling, C. Hage, A. Sinz, *Anal. Chem.* **2018**, *90*, 2805.

[208]   C. Iacobucci, M. Götze, C. H. Ihling, C. Piotrowski, C. Arlt, M. Schäfer, C. Hage, R. Schmidt, A. Sinz, *Nat Protoc.* **2018**, *13*, 2864.

[209]   B. Schilling, R. H. Row, B. W. Gibson, X. Guo, M. M. Young, *J. Am. Soc. Mass. Spectrom.* **2003**, *14*, 834.

[210]   a) J. D. Chavez, D. K. Schweppe, J. K. Eng, J. E. Bruce, *Cell Chem. Biol.* **2016**, *23*, 716; b) X. Wu, J. D. Chavez, D. K. Schweppe, C. Zheng, C. R. Weisbrod, J. K. Eng, A. Murali, S. A. Lee, E. Ramage, L. A. Gallagher et al., *Nat. Commun.* **2016**, *7*, 13414; c) X. Zhu, F. Yu, Z. Yang, S. Liu, C. Dai, X. Lu, C. Liu, W. Yu, N. Li, *Proteomics* **2016**, *16*, 1915; d) J. D. Chavez, C. F. Lee, A. Caudal, A. Keller, R. Tian, J. E. Bruce, *Cell Syst.* **2018**, *6*, 136-141.e5.

[211]   Y. Feng, G. de Franceschi, A. Kahraman, M. Soste, A. Melnik, P. J. Boersema, P. P. de Laureto, Y. Nikolaev, A. P. Oliveira, P. Picotti, *Nat. Biotechnol.* **2014**, *32*, 1036.

[212]   S. Schopper, A. Kahraman, P. Leuenberger, Y. Feng, I. Piazza, O. Müller, P. J. Boersema, P. Picotti, *Nat Protoc.* **2017**, *12*, 2391.

[213]   a) D. Sturtevant, Y.-J. Lee, K. D. Chapman, *Curr. Opin. Biotechnol.* **2016**, *37*, 53; b) M. Holzlechner, S. Reitschmidt, S. Gruber, S. Zeilinger, M. Marchetti-Deschmann, *Proteomics* **2016**, *16*, 1742; c) T. Greer, R. Sturm, L. Li, *J. Proteomics* **2011**, *74*, 2617; d) J. Franck, K. Arafah, M. Elayed, D. Bonnel, D. Vergara, A. Jacquet, D. Vinatier, M. Wisztorski, R. Day, I. Fournier et al., *Mol. Cell. Proteomics* **2009**, *8*, 2023.

[214]   E. R. van Amstalden Hove, D. F. Smith, R. M. A. Heeren, *J. Chromatogr. A* **2010**, *1217*, 3946.

[215]   A. Bodzon-Kulakowska, P. Suder, *Mass Spectrom. Rev.* **2016**, *35*, 147.

[216]   a) J. Kriegsmann, M. Kriegsmann, R. Casadonte, *Int. J. Oncol.* **2015**, *46*, 893; b) A. Ly, A. Buck, B. Balluff, N. Sun, K. Gorzolka, A. Feuchtinger, K.-P. Janssen, P. J. K. Kuppen, C. J. H. van de Velde, G. Weirich et al., *Nat Protoc.* **2016**, *11*, 1428; c) R. Casadonte, R. Longuespée, J. Kriegsmann, M. Kriegsmann, *Adv. Cancer Res.* **2017**, *134*, 173.

[217]   a) E. H. Seeley, R. M. Caprioli, *Anal. Chem.* **2012**, *84*, 2105; b) J. Oetjen, M. Aichler, D. Trede, J. Strehlow, J. Berger, S. Heldmann, M. Becker, M. Gottschalk, J. H. Kobarg, S. Wirtz et al., *J. Proteomics* **2013**, *90*, 52; c) H. Thiele, S. Heldmann, D. Trede, J. Strehlow, S. Wirtz, W. Dreher, J. Berger, J. Oetjen, J. H. Kobarg, B. Fischer et al., *Biochim. Biophys. Acta* **2014**, *1844*, 117.

[218]   T. Kurtović, M. Lang Balija, M. Brgles, D. Sviben, M. Tunjić, H. Cajner, M. Marchetti-Deschmann, G. Allmaier, B. Halassy, *PLoS Negl. Trop. Dis.* **2019**, *13*, e0007431.

[219]   a) P. Escoubas, L. Quinton, G. M. Nicholson, *J. Mass Spectrom.* **2008**, *43*, 279; b) B. Lomonte, J. Escolano, J. Fernández, L. Sanz, Y. Angulo, J. M. Gutiérrez, J. J. Calvete, *J. Proteome Res.* **2008**, *7*, 2445.

[220]   J. J. Calvete, L. Sanz, Y. Angulo, B. Lomonte, J. M. Gutiérrez, *FEBS Lett.* **2009**, *583*, 1736.

[221]   B. Halassy, M. Brgles, L. Habjanec, M. L. Balija, T. Kurtović, M. Marchetti-Deschmann, I. Križaj, G. Allmaier, *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* **2011**, *153*, 223.

[222]   a) J. P. Chippaux, V. Williams, J. White, *Toxicon* **1991**, *29*, 1279; b) R. S. Thorpe, *Venomous snakes. Ecology, evolution and snakebite*, Clarendon Press, Oxford, **2002**.

[223]   P. Juárez, L. Sanz, J. J. Calvete, *Proteomics* **2004**, *4*, 327.

[224]   J. J. Calvete, P. Juárez, L. Sanz, *J. Mass Spectrom.* **2007**, *42*, 1405.

[225]   B. Lomonte, J. J. Calvete, *J. Venom. Anim. Toxins Incl. Trop. Dis.* **2017**, *23*, 26.

[226]   J. J. Calvete, D. Petras, F. Calderón-Celis, B. Lomonte, J. R. Encinar, A. Sanz-Medel, *J. Venom. Anim. Toxins Incl. Trop. Dis.* **2017**, *23*, 27.

[227]   D. Petras, P. Heiss, R. D. Süssmuth, J. J. Calvete, *J. Proteome Res.* **2015**, *14*, 2539.

[228]   F. Calderón-Celis, S. Diez-Fernández, J. M. Costa-Fernández, J. R. Encinar, J. J. Calvete, A. Sanz-Medel, *Anal. Chem.* **2016**, *88*, 9699.

[229]   J. J. Calvete, *Expert Rev. Proteomics* **2018**, *15*, 555.

[230]   B. Lomonte, W.-C. Tsai, J. M. Ureña-Diaz, L. Sanz, D. Mora-Obando, E. E. Sánchez, B. G. Fry, J. M. Gutiérrez, H. L. Gibbs, M. G. Sovic et al., *J. Proteomics* **2014**, *96*, 103.

[231]   D. Petras, P. Heiss, R. A. Harrison, R. D. Süssmuth, J. J. Calvete, *J. Proteomics* **2016**, *146*, 148.

[232]   D. Pla, D. Petras, A. J. Saviola, C. M. Modahl, L. Sanz, A. Pérez, E. Juárez, S. Frietze, P. C. Dorrestein, S. P. Mackessy et al., *J. Proteomics* **2018**, *174*, 71.

[233]   B. Göçmen, P. Heiss, D. Petras, A. Nalbantsoy, R. D. Süssmuth, *Toxicon* **2015**, *107*, 163.

[234]   A. Nalbantsoy, B.-F. Hempel, D. Petras, P. Heiss, B. Göçmen, N. Iğci, M. Z. Yildiz, R. D. Süssmuth, *Toxicon* **2017**, *135*, 71.

[235]   R. G. Birch, *Mol. Plant Pathol.* **2001**, *2*, 1.

[236]   P. Rott, *A Guide to Sugarcane Diseases*, CIRAD 2000, Montpellier, **2000**.

[237]   a) J. H. Daugrois, V. Dumont, P. Champoiseau, L. Costet, R. Boisne-Noc, P. Rott, *Eur. J. Plant Pathol.* **2003**, *109*, 445; b) P. Champoiseau, P. Rott, J.-H. Daugrois, *Plant Dis.* **2009**, *93*, 339.

[238]   I. Mensi, M.-S. Vernerey, D. Gargani, M. Nicole, P. Rott, *Open Biol.* **2014**, *4*, 130116.

[239]   M. Royer, L. Costet, E. Vivien, M. Bes, A. Cousin, A. Damais, I. Pieretti, A. Savin, S. Megessier, M. Viard et al., *Mol. Plant-Microbe Interact.* **2004**, *17*, 414.

[240]   R. G. Birch, S. S. Patil, *J. Gen. Microbiol.* **1985**, *131*, 1069.

[241]   Birch, Birch, Robert G., S. S. Patil, U.S. Patent 4,525,354.

[242]   A. Maxwell, D. M. Lawson, *Curr. Top. Med. Chem.* **2003**, *3*, 283.

[243]   E. Vivien, D. Pitorre, S. Cociancich, I. Pieretti, D. W. Gabriel, P. C. Rott, M. Royer, *Antimicrob. Agents Chemother.* **2007**, *51*, 1549.

[244]   J. Kretz, D. Kerwat, V. Schubert, S. Grätz, A. Pesic, S. Semsary, S. Cociancich, M. Royer, R. D. Süssmuth, *Angew. Chem.* **2015**, *54*, 1969.

[245]   S. J. Brewer, P. M. Taylor, M. K. Turner, *Biochem. J.* **1980**, *185*, 555.

[246]   C. L. Freel Meyers, M. Oberthür, H. Xu, L. Heide, D. Kahne, C. T. Walsh, *Angew. Chem.* **2004**, *43*, 67.

[247]   C. Parthier, S. Görlich, F. Jaenecke, C. Breithaupt, U. Bräuer, U. Fandrich, D. Clausnitzer, U. F. Wehmeier, C. Böttcher, D. Scheel et al., *Angew. Chem.* **2012**, *51*, 4046.

[248]   J. Dauvergne, K. Wellington, K. Chibale, *Tetrahedron Lett.* **2004**, *45*, 43.

[249]   a) T.-W. Yu, L. Bai, D. Clade, D. Hoffmann, S. Toelzer, K. Q. Trinh, J. Xu, S. J. Moss, E. Leistner, H. G. Floss, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7968; b) Y. Li, P. Zhao, Q. Kang, J. Ma, L. Bai, Z. Deng, *Chem. Biol.* **2011**, *18*, 1571.

[250]   P. Rott, L. Fleites, G. Marlow, M. Royer, D. W. Gabriel, *Mol. Plant-Microbe Interact.* **2011**, *24*, 594.

[251]   I. Pieretti, M. Royer, V. Barbe, S. Carrere, R. Koebnik, S. Cociancich, A. Couloux, A. Darrasse, J. Gouzy, M.-A. Jacques et al., *BMC Genom.* **2009**, *10*, 616.

[252]   A. Castañeda, J. D. Reddy, B. El-Yacoubi, D. W. Gabriel, *Mol. Plant-Microbe Interact.* **2005**, *18*, 1306.

[253]   T. R. Shryock, *Performance standards for antimicrobial disk and dilution susceptibility tests for bacteria isolated from animals. Approved standard*, NCCLS, Wayne, Pa., **2002**.

[254]   S. Sundaram, C. Hertweck, *Curr. Opin. Chem. Biol.* **2016**, *31*, 82.

[255]   R. D. Süssmuth, A. Mainz, *Angew. Chem.* **2017**, *56*, 3770.

[256]   T. Izoré, M. J. Cryle, *Nat. Prod. Rep.* **2018**, *35*, 1120.

[257]   a) R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla, C. T. Walsh, *Chem. Biol.* **1996**, *3*, 923; b) T. Weber, R. Baumgartner, C. Renner, M. A. Marahiel, T. A. Holak, *Structure* **2000**, *8*, 407; c) A. Koglin, M. R. Mofid, F. Löhr, B. Schäfer, V. V. Rogov, M.-M. Blum, T. Mittag, M. A. Marahiel, F. Bernhard, V. Dötsch, *Science* **2006**, *312*, 273; d) J. Beld, E. C. Sonnenschein, C. R. Vickery, J. P. Noel, M. D. Burkart, *Nat. Prod. Rep.* **2014**, *31*, 61.

[258]   C. T. Walsh, *Acc. Chem. Res.* **2008**, *41*, 4.

[259]   J. R. Lohman, M. Ma, M. E. Cuff, L. Bigelow, J. Bearden, G. Babnigg, A. Joachimiak, G. N. Phillips, B. Shen, *Proteins* **2014**, *82*, 1210.

[260]   E. J. Drake, B. R. Miller, C. Shi, J. T. Tarrasch, J. A. Sundlov, C. L. Allen, G. Skiniotis, C. C. Aldrich, A. M. Gulick, *Nature* **2016**, *529*, 235.

[261]   J. M. Reimer, M. N. Aloise, P. M. Harrison, T. M. Schmeing, *Nature* **2016**, *529*, 239.

[262]   a) A. Tanovic, S. A. Samel, L.-O. Essen, M. A. Marahiel, *Science* **2008**, *321*, 659; b) E. A. Felnagle, J. J. Barkei, H. Park, A. M. Podevels, M. D. McMahon, D. W. Drott, M. G. Thomas, *Biochemistry* **2010**, *49*, 8815.

[263]   B. R. Miller, J. A. Sundlov, E. J. Drake, T. A. Makin, A. M. Gulick, *Proteins* **2014**, *82*, 2691.

[264]   D. P. Dowling, Y. Kung, A. K. Croft, K. Taghizadeh, W. L. Kelly, C. T. Walsh, C. L. Drennan, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12432.

[265]   W.-H. Chen, K. Li, N. S. Guntaka, S. D. Bruner, *ACS Chem. Biol.* **2016**, *11*, 2293.

[266]   D. P. Frueh, H. Arthanari, A. Koglin, D. A. Vosburg, A. E. Bennett, C. T. Walsh, G. Wagner, *Nature* **2008**, *454*, 903.

[267]   Y. Liu, T. Zheng, S. D. Bruner, *Chem. Biol.* **2011**, *18*, 1482.

[268]   a) R. Shi, S. S. Lamb, B. Zakeri, A. Proteau, Q. Cui, T. Sulea, A. Matte, G. D. Wright, M. Cygler, *Chem. Biol.* **2009**, *16*, 401; b) C. Brieke, G. Yim, M. Peschke, G. D. Wright, M. J. Cryle, *Chem. Comm.* **2016**, *52*, 13679.

[269]   D. Thankachan, A. Fazal, D. Francis, L. Song, M. E. Webb, R. F. Seipke, *ACS Chem. Biol.* **2019**, *14*, 845.

[270]   a) D. Bischoff, S. Pelzer, A. Höltzel, G. J. Nicholson, S. Stockert, W. Wohlleben, G. Jung, R. D. Süssmuth, *Angew. Chem.* **2001**, *40*, 1693; b) K. Zerbe, K. Woithe, D. B. Li, F. Vitali, L. Bigler, J. A. Robinson, *Angew. Chem.* **2004**, *43*, 6709; c) K. Woithe, N. Geib, K. Zerbe, D. B. Li, M. Heck, S. Fournier-Rousset, O. Meyer, F. Vitali, N. Matoba, K. Abou-Hadeed et al., *J. Am. Chem. Soc.* **2007**, *129*, 6887; d) K. Woithe, N. Geib, O. Meyer, T. Wörtz, K. Zerbe, J. A. Robinson, *Org. Biomol. Chem.* **2008**, *6*, 2861; e) P. C. Schmartz, K. Wölfel, K. Zerbe, E. Gad, E. S. El Tamany, H. K. Ibrahim, K. Abou-Hadeed, J. A. Robinson, *Angew. Chem.* **2012**, *51*, 11468.

[271]   a) P. C. Dorrestein, E. Yeh, S. Garneau-Tsodikova, N. L. Kelleher, C. T. Walsh, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13843; b) S. Lin, S. G. van Lanen, B. Shen, *J. Am. Chem. Soc.* **2007**, *129*, 12432; c) T. Kittilä, C. Kittel, J. Tailhades, D. Butz, M. Schoppet, A. Büttner, R. J. A. Goode, R. B. Schittenhelm, K.-H. van Pee, R. D. Süssmuth et al., *Chem. Sci.* **2017**, *8*, 5992.

[272]   K. Haslinger, C. Brieke, S. Uhlmann, L. Sieverling, R. D. Süssmuth, M. J. Cryle, *Angew. Chem.* **2014**, *53*, 8518.

[273]   E. Stegmann, H.-J. Frasch, W. Wohlleben, *Curr. Opin. Microbiol.* **2010**, *13*, 595.

[274]   M. Peschke, M. Gonsior, R. D. Süssmuth, M. J. Cryle, *Curr. Opin. Struct. Biol.* **2016**, *41*, 46.

[275]   K. Haslinger, M. Peschke, C. Brieke, E. Maximowitsch, M. J. Cryle, *Nature* **2015**, *521*, 105.

[276]   X. Ge, T. L. Sirich, M. K. Beyer, H. Desaire, J. A. Leary, *Anal. Chem.* **2001**, *73*, 5078.

[277]   A. Liesener, U. Karst, *Anal. Bioanal. Chem.* **2005**, *382*, 1451.

[278]   A. Cornish-Bowden, *Fundamentals of Enzyme Kinetics*, Wiley-VCH, Weinheim, **2013**.

[279]   H. Barsnes, M. Vaudel, *J. Proteome Res.* **2018**, *17*, 2552.

[280]   P. Schanda, B. Brutscher, *J. Am. Chem. Soc.* **2005**, *127*, 8014.

[281]   E. Kupce, R. Freeman, *J. Magn. Reson. Series A* **1994**, *108*, 268.

[282]   H. Geen, R. Freeman, *J. Magn. Reson.* **1991**, *93*, 93.

[283]   a) D. S. Wishart, C. G. Bigam, J. Yao, F. Abildgaard, H. J. Dyson, E. Oldfield, J. L. Markley, B. D. Sykes, *J. Biomol. NMR* **1995**, *6*, 135; b) J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright, K. Wüthrich, *J. Biomol. NMR* **1998**, *12*, 1.

[284]   a) W. Lee, M. Tonelli, J. L. Markley, *Bioinformatics* **2015**, *31*, 1325; b) W. Lee, G. Cornilescu, H. Dashti, H. R. Eghbalnia, M. Tonelli, W. M. Westler, S. E. Butcher, K. A. Henzler-Wildman, J. L. Markley, *J. Biomol. NMR* **2016**, *64*, 307.

[285]   Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, *J. Biomol. NMR* **2009**, *44*, 213.

[286]   L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, *Nat Protoc.* **2015**, *10*, 845.

[287]   J. Yang, Y. Zhang, *Curr. Protoc. Bioinformatics* **2015**, *52*, 5.8.1-5.8.15.

[288]   B. Webb, A. Sali, *Curr. Protoc. Bioinformatics* **2016**, *54*, 5.6.1-5.6.37.

[289]   L. Schrödinger, *The PyMOL Molecular Graphics System*, Schrödinger, LLC.

[290]   V. Oldrati, E. Bianchi, R. Stöcklin in *Spider Ecophysiology* (Ed.: W. Nentwig), Springer Berlin Heidelberg, Berlin, Heidelberg, **2013**, pp. 491–503.

[291]   V. Oldrati, M. Arrell, A. Violette, F. Perret, X. Sprüngli, J.-L. Wolfender, R. Stöcklin, *Mol. Biosyst.* **2016**, *12*, 3530.

[292]   D. C. I. Koh, A. Armugam, K. Jeyaseelan, *Cell. Mol. Life Sci.* **2006**, *63*, 3030.

[293]   I. Vetter, J. L. Davis, L. D. Rash, R. Anangi, M. Mobli, P. F. Alewood, R. J. Lewis, G. F. King, *Amino Acids* **2011**, *40*, 15.

[294]   J. J. Calvete, *Biochem. J.* **2017**, *474*, 611.

[295]   B. Xie, Y. Huang, K. Baumann, B. G. Fry, Q. Shi, *Mar. Drugs* **2017**, *15*.

[296]   a) A. Ménez (Ed.) *Perspectives in Molecular Toxinology*, John Wiley, New York, **04-02**; b) J. R. Prashanth, R. J. Lewis, S. Dutertre, *Toxicon* **2012**, *60*, 470; c) J. M. Heather, B. Chain, *Genomics* **2016**, *107*, 1.

[297]   a) J. W. Fox, L. Ma, K. Nelson, N. E. Sherman, S. M. T. Serrano, *Toxicon* **2006**, *47*, 700; b) D. Georgieva, R. K. Arni, C. Betzel, *Expert Rev. Proteomics* **2008**, *5*, 787.

[298]   J. J. Calvete, *Expert Rev. Proteomics* **2011**, *8*, 739.

[299]   R. D. Melani, O. S. Skinner, L. Fornelli, G. B. Domont, P. D. Compton, N. L. Kelleher, *Mol. Cell. Proteomics* **2016**, *15*, 2423.

[300]   J. J. Calvete, *Expert Rev. Proteomics* **2014**, *11*, 315.

[301]   N. Cox, J. Chanson, S. Stuart, *The Status and Distribution of Reptiles and Amphibians of the Mediterranean Basin*, The World Conservation Union (IUCN), Swizerland, **2006**.

[302]   a) D. Mallow, D. Ludwig, G. Nilson, *True Vipers. Natural History and Toxinology of Old World Vipers*, Krieger Publ, Malabar, Fla., **2003**; b) M. O'Shea, *Venomous Snakes of the World*, New Holland, London, **2008**.

[303]   J. W. Fox, S. M. T. Serrano, *Proteomics* **2008**, *8*, 909.

[304]   V. Stoyanova, R. Aleksandrov, M. Lukarska, D. Duhalov, V. Atanasov, S. Petrova, *Toxicon* **2012**, *60*, 802.

[305]   T. Lamb, L. de Haro, D. Lonati, M. Brvar, M. Eddleston, *Clin. Toxicol.* **2017**, *55*, 557.

[306]   G. A. Boulenger, *Proc. Zool. Soc. Lond.* **1903**, *73*, 185.

[307]   G. A. Boulenger, *Ann. Mag. Nat. Hist.* **1904**, 134.

[308]   G. A. Boulenger, *Ann. Mag. Nat. Hist.* **1913**, 283.

[309]   Bruno S., *Mem. Mus. Civ. Stor. Nat.* **1968**, 289.

[310]   E. Sochurek, *Herpet. Blatt.* **1974**, 1.

[311]   G. Nilson, B. Tuniyev, C. Andrén, N. L. Orlov (Eds.) *Vipers of Caucasus. Taxomonic Considerations*, Joger, U. (Ed.), Kaupia, **1999**.

[312]   Heckes, U., Gruber, H.-J., N. Stumpel in *Handbuch der Reptilien und Amphibien Europas. Schlangen (Serpentes) III, Viperidae* (Eds.: U. Joger, N. Stümpel, W. Böhme), pp. 81–151.

[313]   L. Tomović, *Herpetol. J.* **2006**, 191.

[314]   S. Ursenbacher, S. Schweiger, L. Tomović, J. Crnobrnja-Isailović, L. Fumagalli, W. Mayer, *Mol. Phylogenet. Evol.* **2008**, *46*, 1116.

[315]   B. Göçmen, J. Mulder, M. Karis, K. Mebert, *South-West. J. Hortic. Biol. Environ.* **2015**, *2*, 91.

[316]   D. Georgieva, M. Risch, A. Kardas, F. Buck, M. von Bergen, C. Betzel, *J. Proteome Res.* **2008**, *7*, 866.

[317]   C. Y. Frangides, V. Koulouras, S. N. Kouni, G. V. Tzortzatos, A. Nikolaou, J. Pneumaticos, C. Pierrakeas, C. Niarchos, N. G. Kounis, C. M. Koutsojannis, *Eur. J. Intern. Med.* **2006**, *17*, 24.

[318]   M. Brvar, T. Kurtović, D. Grenc, M. Lang Balija, I. Križaj, B. Halassy, *Clin. Toxicol.* **2017**, *55*, 241.

[319]   I. G. Archundia, A. R. de Roodt, B. Ramos-Cerrillo, J.-P. Chippaux, L. Olguín-Pérez, A. Alagón, R. P. Stock, *Toxicon* **2011**, *57*, 1049.

[320]   a) B. Francis, I. I. Kaiser, *Toxicon* **1993**, *31*, 889; b) S. M. Munekiyo, S. P. Mackessy, *Toxicon* **2005**, *45*, 255.

[321]   K. T. Yee, M. Pitts, P. Tongyoo, P. Rojnuckarin, M. C. Wilkinson, *Toxins* **2016**, *9*, 15.

[322]   S. C. Wagstaff, P. Favreau, O. Cheneval, G. D. Laing, M. C. Wilkinson, R. L. Miller, R. Stöcklin, R. A. Harrison, *Biochem. Biophys. Res. Commun.* **2008**, *365*, 650.

[323]   Z. Latinović, A. Leonardi, J. Šribar, T. Sajevic, M. C. Žužek, R. Frangež, B. Halassy, A. Trampuš-Bakija, J. Pungerčar, I. Križaj, *J. Proteomics* **2016**, *146*, 34.

[324]   A. T. Tu (Ed.) *Venoms. Chemistry and Molecular Biology*, John Wiley & Sons, New York, **1977**.

[325]   a) Y. Angulo, J. Escolano, B. Lomonte, J. M. Gutiérrez, L. Sanz, J. J. Calvete, *J. Proteome Res.* **2008**, *7*, 708; b) J. Fernández, B. Lomonte, L. Sanz, Y. Angulo, J. M. Gutiérrez, J. J. Calvete, *J. Proteome Res.* **2010**, *9*, 4234.

[326]   H. L. Gibbs, L. Sanz, M. G. Sovic, J. J. Calvete, *PLoS One* **2013**, *8*, e67220.

[327]   T. Chijiwa, M. Deshimaru, I. Nobuhisa, M. Nakai, T. Ogawa, N. Oda, K. Nakashima, Y. Fukumaki, Y. Shimohigashi, S. Hattori et al., *Biochem. J.* **2000**, *347*, 491.

[328]   J. C. Daltry, W. Wüster, R. S. Thorpe, *Nature* **1996**, *379*, 537.

[329]   H. L. Gibbs, S. P. Mackessy, *Toxicon* **2009**, *53*, 672.

[330]   S. P. Mackessy, *Copeia* **1988**, *1988*, 92.

[331]   D. C. Pimenta, B. C. Prezoto, K. Konno, R. L. Melo, M. F. Furtado, A. C. M. Camargo, S. M. T. Serrano, *Rapid Commun. Mass Spectrom.* **2007**, *21*, 1034.

[332]   A. K. Tashima, L. Sanz, A. C. M. Camargo, S. M. T. Serrano, J. J. Calvete, *J. Proteomics* **2008**, *71*, 473.

[333]   I. Mancheva, T. Kleinschmidt, B. Aleksiev, G. Braunitzer, *Hoppe-Seyler's Z. Physiol. Chem.* **1984**, *365*, 885.

[334]   I. Krizaj, N. S. Liang, J. Pungercar, B. Strukelj, A. Ritonja, F. Gubensek, *Eur. J. Biochem.* **1992**, *204*, 1057.

[335]   Almeida L. M., J. R. Resende, R. K. Watanabe, V. C. Corassola, S. Huancahuire-Vega, C. A. da S Caldeira, A. Coutinho-Neto, A. M. Soares, N. Vale, P. A. de C Gomes et al., *Curr. Med. Chem.* **2017**, *24*, 3254.

[336]   J. D. Brenton, L. A. Carey, A. A. Ahmed, C. Caldas, *J. Clin. Oncol.* **2005**, *23*, 7350.

[337]   K. Subik, J.-F. Lee, L. Baxter, T. Strzepek, D. Costello, P. Crowley, L. Xing, M.-C. Hung, T. Bonfiglio, D. G. Hicks et al., *Breast Cancer* **2010**, *4*, 35.

[338]   S. Khunsap, O. Khow, S. Buranapraditkun, S. Suntrarachun, S. Puthong, S. Boonchang, *J. Venom. Anim. Toxins Incl. Trop. Dis.* **2016**, *22*, 7.

[339]   Y. Yamazaki, T. Morita, *Toxicon* **2004**, *44*, 227.

[340]   Y.-L. Wang, J.-H. Kuo, S.-C. Lee, J.-S. Liu, Y.-C. Hsieh, Y.-T. Shih, C.-J. Chen, J.-J. Chiu, W.-G. Wu, *J. Biol. Chem.* **2010**, *285*, 37872.

[341]   A. Alape-Girón, L. Sanz, J. Escolano, M. Flores-Díaz, M. Madrigal, M. Sasa, J. J. Calvete, *J. Proteome Res.* **2008**, *7*, 3556.

[342]   T. Muth, L. Weilnböck, E. Rapp, C. G. Huber, L. Martens, M. Vaudel, H. Barsnes, *J. Proteome Res.* **2014**, *13*, 1143.

[343]   S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **1990**, *215*, 403.

[344]   J. A. Vizcaíno, E. W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Ríos, J. A. Dianes, Z. Sun, T. Farrah, N. Bandeira et al., *Nat. Biotechnol.* **2014**, *32*, 223.

[345]   J. J. Calvete, B. Lomonte, *Toxicon* **2015**, *107*, 159.

References

[346] a) F. J. Vonk, N. R. Casewell, C. V. Henkel, A. M. Heimberg, H. J. Jansen, R. J. R. McCleary, H. M. E. Kerkkamp, R. A. Vos, I. Guerreiro, J. J. Calvete et al., *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 20651; b) S. D. Aird, J. Arora, A. Barua, L. Qiu, K. Terada, A. S. Mikheyev, *Genome Biol. Evol.* **2017**, *9*, 2640.

[347] J. Durban, L. Sanz, D. Trevisan-Silva, E. Neri-Castro, A. Alagón, J. J. Calvete, *J. Proteome Res.* **2017**, *16*, 3370.

[348] a) S. D. Aird, N. J. da Silva, L. Qiu, A. Villar-Briones, V. A. Saddi, M. Pires de Campos Telles, M. L. Grau, A. S. Mikheyev, *Toxins* **2017**, *9*, 187; b) B. G. Fry, H. Scheib, I. de L M Junqueira Azevedo, D. A. Silva, N. R. Casewell, *Toxicon* **2012**, *59*, 696; c) D. R. Rokyta, K. P. Wray, M. J. Margres, *BMC Genom.* **2013**, *14*, 394; d) C. H. Tan, K. Y. Tan, S. Y. Fung, N. H. Tan, *BMC Genom.* **2015**, *16*, 687.

[349] D. Pla, L. Sanz, G. Whiteley, S. C. Wagstaff, R. A. Harrison, N. R. Casewell, J. J. Calvete, *Biochim. Biophys. Acta* **2017**, *1861*, 814.

[350] L. Gonçalves-Machado, D. Pla, L. Sanz, R. J. B. Jorge, M. Leitão-De-Araújo, M. L. M. Alves, D. J. Alvares, J. de Miranda, J. Nowatzki, K. de Morais-Zani et al., *J. Proteomics* **2016**, *135*, 73.

[351] N. R. Casewell, R. A. Harrison, W. Wüster, S. C. Wagstaff, *BMC Genom.* **2009**, *10*, 564.

[352] J. Nawarak, S. Sinchaikul, C.-Y. Wu, M.-Y. Liau, S. Phutrakul, S.-T. Chen, *Electrophoresis* **2003**, *24*, 2838.

[353] T. Tasoulis, G. K. Isbister, *Toxins* **2017**, *9*, 290.

[354] R. D. Melani, L. Goto-Silva, F. C. S. Nogueira, M. Junqueira, G. B. Domont in *Venom genomics and proteomics. Venom Genomics and Proteomics* (Eds.: P. Gopalakrishnakone, J. J. Calvete), Springer Netherlands, Dordrecht, **2014**, pp. 1–12.

[355] A. I. Nesvizhskii, R. Aebersold, *Mol. Cell. Proteomics* **2005**, *4*, 1419.

[356] L. M. Smith, N. L. Kelleher, *Nat. Methods* **2013**, *10*, 186.

[357] S. Ainsworth, D. Petras, M. Engmark, R. D. Süssmuth, G. Whiteley, L.-O. Albulescu, T. D. Kazandjian, S. C. Wagstaff, P. Rowley, W. Wüster et al., *J. Proteomics* **2018**, *172*, 173.

[358] M. Damm, B.-F. Hempel, A. Nalbantsoy, R. D. Süssmuth, *Molecules* **2018**, *23*, 1893.

[359] B.-F. Hempel, M. Damm, B. Göçmen, M. Karis, M. A. Oguz, A. Nalbantsoy, R. D. Süssmuth, *Toxins* **2018**, *10*, 23.

[360] P. Ghezellou, V. Garikapati, S. M. Kazemi, K. Strupat, A. Ghassempour, B. Spengler, *Rapid Commun. Mass Spectrom.* **2018**.

[361] A. Barlow, C. E. Pook, R. A. Harrison, W. Wüster, *Proc. Biol. Sci.* **2009**, *276*, 2443.

[362] a) H.-W. Huang, B.-S. Liu, K.-Y. Chien, L.-C. Chiang, S.-Y. Huang, W.-C. Sung, W.-G. Wu, *J. Proteomics* **2015**, *128*, 92; b) R. J. B. Jorge, H. S. A. Monteiro, L. Gonçalves-Machado, M. C. Guarnieri, R. M. Ximenes, D. M. Borges-Nojosa, K. P. d. O. Luna, R. B. Zingali, C. Corrêa-Netto, J. M. Gutiérrez et al., *J. Proteomics* **2015**, *114*, 93.

[363]   a) M. C. Menezes, M. F. Furtado, S. R. Travaglia-Cardoso, A. C. M. Camargo, S. M. T. Serrano, *Toxicon* **2006**, *47*, 304; b) F. G. Amorim, T. R. Costa, D. Baiwir, E. de Pauw, L. Quinton, S. V. Sampaio, *Toxins* **2018**, *10*, 177.

[364]   a) J. M. Gutiérrez, C. Avila, Z. Camacho, B. Lomonte, *Toxicon* **1990**, *28*, 419; b) R. A. P. Guércio, A. Shevchenko, A. Shevchenko, J. L. López-Lozano, J. Paba, M. V. Sousa, C. A. O. Ricart, *Proteome Sci.* **2006**, *4*, 11; c) S. P. Mackessy, N. M. Sixberry, W. H. Heyborne, T. Fritts, *Toxicon* **2006**, *47*, 537.

[365]   a) S. Wu, J. N. Brown, N. Tolić, D. Meng, X. Liu, H. Zhang, R. Zhao, R. J. Moore, P. Pevzner, R. D. Smith et al., *Proteomics* **2014**, *14*, 1211; b) F. Moehring, M. Waas, T. R. Keppel, D. Rathore, A. M. Cowie, C. L. Stucky, R. L. Gundry, *J. Proteome Res.* **2018**, *17*, 2635; c) I. Ntai, T. K. Toby, R. D. LeDuc, N. L. Kelleher, *Methods Mol. Biol.* **2016**, *1410*, 121.

[366]   J. J. Calvete, N. R. Casewell, U. Hernández-Guzmán, S. Quesada-Bernat, L. Sanz, D. R. Rokyta, D. Storey, L.-O. Albulescu, W. Wüster, C. F. Smith et al., *Sci. Rep.* **2018**, *8*, 11539.

[367]   N. R. Casewell, W. Wüster, F. J. Vonk, R. A. Harrison, B. G. Fry, *Trends Ecol. Evol.* **2013**, *28*, 219.

[368]   N. R. Casewell, S. C. Wagstaff, W. Wüster, D. A. N. Cook, F. M. S. Bolton, S. I. King, D. Pla, L. Sanz, J. J. Calvete, R. A. Harrison, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 9205.

[369]   D. J. Massey, J. J. Calvete, E. E. Sánchez, L. Sanz, K. Richards, R. Curtis, K. Boesen, *J. Proteomics* **2012**, *75*, 2576.

[370]   V. G. Starkov, A. V. Osipov, Y. N. Utkin, *Toxicon* **2007**, *49*, 995.

[371]   S. I. Kovalchuk, R. H. Ziganshin, V. G. Starkov, V. I. Tsetlin, Y. N. Utkin, *Toxins* **2016**, *8*, 105.

[372]   a) R. D. Bruce, *Fundam. Appl. Toxicol.* **1985**, *5*, 151; b) C. C. Cates, J. G. McCabe, G. W. Lawson, M. A. Couto, *Comp. Med.* **2014**, *64*, 440.

[373]   Swiss Academy of Medical Sciences, *Experientia* **1996**, *52*, 1.

[374]   J. Archer, G. Whiteley, N. R. Casewell, R. A. Harrison, S. C. Wagstaff, *BMC Bioinform.* **2014**, *15*, 389.

[375]   S. Götz, J. M. García-Gómez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M. Robles, M. Talón, J. Dopazo, A. Conesa, *Nucleic Acids Res.* **2008**, *36*, 3420.

[376]   P. Geniez, *Snakes of Europe, North Africa & the Middle East. A photographic guide*, Princeton University Press, Princeton, **2018**.

[377]   a) D. Sket, F. Gubensek, S. Adamic, D. Lebez, *Toxicon* **1973**, *11*, 47; b) N. Oukkache, R. El Jaoudi, N. Ghalim, F. Chgoury, B. Bouhaouala, N. E. Mdaghri, J.-M. Sabatier, *Toxins* **2014**, *6*, 1873; c) L. Calderón, B. Lomonte, J. M. Gutiérrez, A. Tarkowski, L. A. Hanson, *Toxicon* **1993**, *31*, 743; d) L. de Haro, A. Robbe-Vincent, B. Saliou, M. Valli, C. Bon, V. Choumet, *Hum. Exp. Toxicol.* **2002**, *21*, 137.

[378]   D. R. Rokyta, A. R. Lemmon, M. J. Margres, K. Aronow, *BMC Genom.* **2012**, *13*, 312.

[379]   I. L. M. Junqueira-de-Azevedo, C. M. V. Bastos, P. L. Ho, M. S. Luna, N. Yamanouye, N. R. Casewell, *Mol. Biol. Evol.* **2015**, *32*, 754.

[380]   a) J. W. Fox, S. M. T. Serrano, *FEBS J.* **2008**, *275*, 3016; b) N. R. Casewell, S. C. Wagstaff, R. A. Harrison, C. Renjifo, W. Wüster, *Mol. Biol. Evol.* **2011**, *28*, 2637.

[381]   A. Gasmi, N. Srairi, S. Guermazi, H. Dekhil, H. Dkhil, H. Karoui, M. El Ayeb, *Biochim. Biophys. Acta* **2001**, *1547*, 51.

[382]   A. Navdaev, J. M. Clemetson, J. Polgar, B. E. Kehrel, M. Glauner, E. Magnenat, T. N. Wells, K. J. Clemetson, *J. Biol. Chem.* **2001**, *276*, 20882.

[383]   J. Durban, A. Pérez, L. Sanz, A. Gómez, F. Bonilla, S. Rodríguez, D. Chacón, M. Sasa, Y. Angulo, J. M. Gutiérrez et al., *BMC Genom.* **2013**, *14*, 234.

[384]   C. Vogel, E. M. Marcotte, *Nat. Rev. Genet.* **2012**, *13*, 227.

[385]   K. V. Ruggles, K. Krug, X. Wang, K. R. Clauser, J. Wang, S. H. Payne, D. Fenyö, B. Zhang, D. R. Mani, *Mol. Cell. Proteomics* **2017**, *16*, 959.

[386]   a) J. J. Calvete, L. Sanz, A. Pérez, A. Borges, A. M. Vargas, B. Lomonte, Y. Angulo, J. M. Gutiérrez, H. M. Chalkidis, R. H. V. Mourão et al., *J. Proteomics* **2011**, *74*, 510; b) N. d. C. Galizio, C. Serino-Silva, D. R. Stuginski, P. A. E. Abreu, S. S. Sant'Anna, K. F. Grego, A. K. Tashima, A. M. Tanaka-Azevedo, K. de Morais-Zani, *J. Proteomics* **2018**, *186*, 56; c) M. J. Margres, A. Patton, K. P. Wray, A. T. B. Hassinger, M. J. Ward, E. M. Lemmon, A. R. Lemmon, D. R. Rokyta, *Mol. Biol. Evol.* **2018**.

[387]   B. G. Fry, J. C. Wickramaratna, W. C. Hodgson, P. F. Alewood, R. M. Kini, H. Ho, W. Wüster, *Rapid Commun. Mass Spectrom.* **2002**, *16*, 600.

[388]   S. M. T. Serrano, *Toxicon* **2013**, *62*, 19.

[389]   a) G. S. Dias, E. S. Kitano, A. H. Pagotto, S. S. Sant'anna, M. M. T. Rocha, A. Zelanis, S. M. T. Serrano, *J. Proteome Res.* **2013**, *12*, 4585; b) H. L. Gibbs, L. Sanz, J. E. Chiucchi, T. M. Farrell, J. J. Calvete, *J. Proteomics* **2011**, *74*, 2169; c) A. Zelanis, A. K. Tashima, M. M. T. Rocha, M. F. Furtado, A. C. M. Camargo, P. L. Ho, S. M. T. Serrano, *J. Proteome Res.* **2010**, *9*, 2278.

[390]   a) H. Schlüter, R. Apweiler, H.-G. Holzhütter, P. R. Jungblut, *Chem. Cent. J.* **2009**, *3*, 11; b) P. R. Jungblut, B. Thiede, H. Schlüter, *J. Proteomics* **2016**, *134*, 1.

[391]   R. S. Brown, J. J. Lennon, *Anal. Chem.* **1995**, *67*, 3990.

[392]   K. Demeure, L. Quinton, V. Gabelica, E. de Pauw, *Anal. Chem.* **2007**, *79*, 8678.

[393]   a) B. Spengler, D. Kirsch, R. Kaufmann, R. J. Cotter, *Rapid Commun. Mass Spectrom.* **1991**, *5*, 198; b) J. Hardouin, *Mass Spectrom. Rev.* **2007**, *26*, 672.

[394]   M. Takayama, *J. Am. Soc. Mass. Spectrom.* **2001**, *12*, 420.

[395]   M. Takayama, *J. Am. Soc. Mass. Spectrom.* **2001**, *12*, 1044.

[396]   T. Köcher, A. Engström, R. A. Zubarev, *Anal. Chem.* **2005**, *77*, 172.

[397]   a) J. J. Lennon, K. A. Walsh, *Protein Sci.* **1997**, *6*, 2446; b) J. J. Lennon, K. A. Walsh, *Protein Sci.* **1999**, *8*, 2487; c) M. Macht, *Bioanalysis* **2009**, *1*, 1131; d) D. Calligaris, C. Villard, L. Terras,

D. Braguer, P. Verdier-Pinard, D. Lafitte, *Anal. Chem.* **2010**, *82*, 6176; e) A. Resemann, D. Wunderlich, U. Rothbauer, B. Warscheid, H. Leonhardt, J. Fuchser, K. Kuhlmann, D. Suckau, *Anal. Chem.* **2010**, *82*, 3283; f) K. Demeure, V. Gabelica, E. A. de Pauw, *J. Am. Soc. Mass. Spectrom.* **2010**, *21*, 1906; g) F.-G. Hanisch, *Anal. Chem.* **2011**, *83*, 4829; h) A. W. Jones, H. J. Cooper, *Analyst* **2011**, *136*, 3419.

[398]   Y. Fukuyama, S. Iwamoto, K. Tanaka, *J. Mass Spectrom.* **2006**, *41*, 191.

[399]   L. Quinton, K. Demeure, R. Dobson, N. Gilles, V. Gabelica, E. de Pauw, *J. Proteome Res.* **2007**, *6*, 3216.

[400]   C. Y. Chow, B. Cristofori-Armstrong, E. A. B. Undheim, G. F. King, L. D. Rash, *Toxins* **2015**, *7*, 2494.

[401]   E. A. B. Undheim, K. Sunagar, B. R. Hamilton, A. Jones, D. J. Venter, B. G. Fry, G. F. King, *J. Proteomics* **2014**, *102*, 1.

[402]   B. Göçmen, J. Mulder, M. Kariş, M. Oğuz, *Herpetologica Romanica* **2014**, *8*, 1.

[403]   B. Göçmen, K. Mebert, M. Karış, M. A. Oğuz, S. Ursenbacher, *Amphib. Reptilia* **2017**, *38*, 289.

[404]   S. W. Wingett, S. Andrews, *F1000Research* **2018**, *7*, 1338.

[405]   A. M. Bolger, M. Lohse, B. Usadel, *Bioinformatics* **2014**, *30*, 2114.

[406]   a) M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng et al., *Nat. Biotechnol.* **2011**, *29*, 644; b) B. J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber et al., *Nat Protoc.* **2013**, *8*, 1494.

[407]   L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, *Bioinformatics* **2012**, *28*, 3150.

[408]   W. Li, A. Godzik, *Bioinformatics* **2006**, *22*, 1658.

[409]   B. Buchfink, C. Xie, D. H. Huson, *Nat. Methods* **2015**, *12*, 59.

[410]   B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, *Genome Biol.* **2009**, *10*, R25.

[411]   R. B. Currier, J. J. Calvete, L. Sanz, R. A. Harrison, P. D. Rowley, S. C. Wagstaff, *PLoS One* **2012**, *7*, e41888.

[412]   M. Brillard-Bourdet, V. Nguyên, M. Ferrer-di Martino, F. Gauthier, T. Moreau, *Biochem. J.* **1998**, *331 (Pt 1)*, 239.

[413]   D. Morgenstern, B. H. Rohde, G. F. King, T. Tal, D. Sher, E. Zlotkin, *Toxicon* **2011**, *57*, 695.

[414]   a) B. Lu, T. Chen, *Drug Discov. Today Biosilico* **2004**, *2*, 85; b) V. Gorshkov, S. Y. K. Hotta, T. Verano-Braga, F. Kjeldsen, *Proteomics* **2016**, *16*, 2470; c) T. Muth, B. Y. Renard, *Brief. Bioinformatics* **2018**, *19*, 954.

[415]   C. S. Ozverel, M. Damm, B.-F. Hempel, B. Göçmen, R. Sroka, R. D. Süssmuth, A. Nalbantsoy, *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* **2019**, *220*, 20.

[416]   a) J. Pawlak, R. Manjunatha Kini, *Toxicon* **2006**, *48*, 278; b) A. Chapeaurouge, M. A. Reza, S. P. Mackessy, P. C. Carvalho, R. H. Valente, A. Teixeira-Ferreira, J. Perales, Q. Lin, R. M. Kini, *PLoS One* **2015**, *10*, e0092091.

[417]   a) S. G. Soares, L. L. Oliveira, *Protein Pept. Lett.* **2009**, *16*, 913; b) D. Andrade-Silva, D. Ashline, T. Tran, A. S. Lopes, S. R. Travaglia Cardoso, M. d. S. Reis, A. Zelanis, S. M. T. Serrano, V. Reinhold, *Mol. Cell. Proteomics* **2018**, *17*, 1261.

[418]   E. Verdin, M. Ott, *Nat. Rev. Mol. Cell. Biol.* **2015**, *16*, 258.

[419]   a) M. Mann, O. N. Jensen, *Nat. Biotechnol.* **2003**, *21*, 255; b) M. R. Larsen, M. B. Trelle, T. E. Thingholm, O. N. Jensen, *Biotechniques* **2006**, *40*, 790; c) C. Carapito, C. Klemm, R. Aebersold, B. Domon, *J. Proteome Res.* **2009**, *8*, 2608.

[420]   R. D. Melani, F. C. S. Nogueira, G. B. Domont, *J. Venom. Anim. Toxins Incl. Trop. Dis.* **2017**, *23*, 44.

[421]   I.-H. Tsai, Y.-M. Wang, A. C. Cheng, V. Starkov, A. Osipov, I. Nikitin, Y. Makarova, R. Ziganshin, Y. Utkin, *Toxicon* **2011**, *57*, 332.

[422]   a) P. Ascenzi, A. Bocedi, M. Bolognesi, A. Spallarossa, M. Coletta, R. de Cristofaro, E. Menegatti, *Curr. Protein Pept. Sci.* **2003**, *4*, 231; b) S. E. Banijamali, M. Amininasab, M. M. Elmi, *Arch. Biochem. Biophys.* **2019**, *662*, 1.

[423]   a) N. A. Ghazaryan, L. Ghulikyan, A. Kishmiryan, T. V. Andreeva, Y. N. Utkin, V. I. Tsetlin, B. Lomonte, N. M. Ayvazyan, *Biochim. Biophys. Acta* **2015**, *1848*, 463; b) F. Tonello, M. Rigoni in *Snake Venoms* (Eds.: H. Inagaki, C.-W. Vogel, A. K. Mukherjee, T. R. Rahmy), Springer Netherlands, Dordrecht, **2017**, pp. 49–65.

[424]   a) J. J. Calvete, M. Jürgens, C. Marcinkiewicz, A. Romero, M. Schrader, S. Niewiarowski, *Biochem. J.* **2000**, *345 Pt 3*, 573; b) J. J. Calvete, M. P. Moreno-Murciano, R. D. G. Theakston, D. G. Kisiel, C. Marcinkiewicz, *Biochem. J.* **2003**, *372*, 725.

[425]   a) N. Tanaka, H. Nakada, N. Itoh, Y. Mizuno, M. Takanishi, T. Kawasaki, S. Tate, F. Inagaki, I. Yamashina, *J. Biochem.* **1992**, *112*, 68; b) M. A. Coronado, R. J. Eberle, Arni Raghuvir K., *J. Mol. Cell. Biol. Forecast* **2018**, *1*, 1011.

[426]   a) V. Zupunski, D. Kordis, F. Gubensek, *FEBS Lett.* **2003**, *547*, 131; b) A. Munawar, S. A. Ali, A. Akrem, C. Betzel, *Toxins* **2018**, *10*, 474.

[427]   a) R. C. de Paula, H. C. Castro, C. R. Rodrigues, P. A. Melo, A. L. Fuly, *Protein Pept. Lett.* **2009**, *16*, 899; b) H. Xiao, H. Pan, K. Liao, M. Yang, C. Huang, *Biomed Res. Int.* **2017**, *2017*, 1.

[428]   A. W. Purcell, J. J. Gorman, *J. Immunol. Methods* **2001**, *249*, 17.

[429]   A. S. Ramazanova, V. G. Starkov, A. V. Osipov, R. H. Ziganshin, S. Y. Filkin, V. I. Tsetlin, Y. N. Utkin, *Toxicon* **2009**, *53*, 162.

[430]   L. R. V. Alencar, T. B. Quental, F. G. Grazziotin, M. L. Alfaro, M. Martins, M. Venzon, H. Zaher, *Mol. Phylogenetics Evol.* **2016**, *105*, 50.

[431]   D. Petras, B.-F. Hempel, B. Göçmen, M. Karis, G. Whiteley, S. C. Wagstaff, P. Heiss, N. R. Casewell, A. Nalbantsoy, R. D. Süssmuth, *J. Proteomics* **2019**, *199*, 31.

[432]   R. I. Al-Shekhadat, K. S. Lopushanskaya, Á. Segura, J. M. Gutiérrez, J. J. Calvete, D. Pla, *Toxins* **2019**, *11*, 90.

[433]   a) F. S. Markland, *Toxicon* **1998**, *36*, 1749; b) S. Vaiyapuri, N. Thiyagarajan, E. G. Hutchinson, J. M. Gibbins, *Bioinformation* **2012**, *8*, 763.

[434]   E. Siigur, A. Aaspõllu, J. Siigur, *Gene* **2001**, *263*, 199.

[435]   R. Doley, S. P. Mackessy, R. M. Kini, *BMC Evol. Biol.* **2009**, *9*, 146.

[436]   S. Vaiyapuri, S. C. Wagstaff, R. A. Harrison, J. M. Gibbins, E. G. Hutchinson, *PLoS One* **2011**, *6*, e21532.

[437]   a) O. Zinenko, A. Avci, F. Spitzenberger, A. Tupikov, K. Shiryaev, E. Bozkurt, C. Ilgaz, Stümpel Nikolaus, *Herpetozoa* **2016**, 141; b) K. Mebert, B. Göçmen, M. Karış, *South-west. J. Hortic. Biol. Environ.* **2017**, *8*, 65.

[438]   K. Sunagar, E. A. B. Undheim, H. Scheib, E. C. K. Gren, C. Cochran, C. E. Person, I. Koludarov, W. Kelln, W. K. Hayes, G. F. King et al., *J. Proteomics* **2014**, *99*, 68.

[439]   A. Bocian, M. Urbanik, K. Hus, A. Łyskowski, V. Petrilla, Z. Andrejčáková, M. Petrillová, J. Legath, *Molecules* **2016**, *21*, 1398.

[440]   a) C. Baldo, C. Jamora, N. Yamanouye, T. M. Zorn, A. M. Moura-Da-Silva, *PLoS Negl. Trop. Dis.* **2010**, *4*, e727; b) H. F. Williams, B. A. Mellows, R. Mitchell, P. Sfyri, H. J. Layfield, M. Salamah, R. Vaiyapuri, H. Collins-Hooper, A. B. Bicknell, A. Matsakas et al., *PLoS Negl. Trop. Dis.* **2019**, *13*, e0007041.

[441]   R. M. Kini, *Toxicon* **2003**, *42*, 827.

[442]   a) D. Boels, J. F. Hamel, M. Bretaudeau Deguigne, P. Harry, *Clin. Toxicol.* **2012**, *50*, 189; b) N. R. Casewell, I. Al-Abdulla, D. Smith, R. Coxon, J. Landon, *Toxins* **2014**, *6*, 2471.

[443]   G. Zanetti, E. Duregotti, C. A. Locatelli, A. Giampreti, D. Lonati, O. Rossetto, M. Pirazzini, *Sci. Rep.* **2018**, *8*, 9818.

[444]   A. García-Arredondo, M. Martínez, A. Calderón, A. Saldívar, R. Soria, *Toxins* **2019**, *11*, 149.

[445]   a) M. R. Pulido, M. García-Quintanilla, M. L. Gil-Marqués, M. J. McConnell, *Drug Discov. Today* **2016**, *21*, 465; b) W. M. Aizat, I. Ismail, N. M. Noor, *Adv. Exp. Med. Biol.* **2018**, *1102*, 1; c) H. Dihazi, A. R. Asif, T. Beißbarth, R. Bohrer, K. Feussner, I. Feussner, O. Jahn, C. Lenz, A. Majcherczyk, B. Schmidt et al., *Expert Rev. Proteomics* **2018**, *15*, 463; d) B. B. Misra, C. D. Langefeld, M. Olivier, L. A. Cox, *J. Mol. Endocrinol.* **2018**; e) J. Paananen, V. Fortino, *Brief. Bioinformatics* **2019**.

[446]   a) J. Y. Yang, J. R. Karr, J. D. Watrous, P. C. Dorrestein, *Curr. Opin. Chem. Biol.* **2011**, *15*, 79; b) L. T. Ngo, J. I. Okogun, W. R. Folk, *Nat. Prod. Rep.* **2013**, *30*, 584; c) N. Chandra Mohana, H. C. Yashavantha Rao, D. Rakshith, P. R. Mithun, B. R. Nuthan, S. Satish, *J. Genet. Eng. Biotechnol.* **2018**, *16*, 1.

[447]   a) D. C. Koboldt, K. M. Steinberg, D. E. Larson, R. K. Wilson, E. R. Mardis, *Cell* **2013**, *155*, 27; b) S. Goodwin, J. D. McPherson, W. R. McCombie, *Nat. Rev. Genet.* **2016**, *17*, 333.

[448]   a) J. Rosamond, A. Allsop, *Science* **2000**, *287*, 1973; b) A. G. McArthur, N. Waglechner, F. Nizam, A. Yan, M. A. Azad, A. J. Baylay, K. Bhullar, M. J. Canova, G. de Pascale, L. Ejim et al., *Antimicrob. Agents Chemother.* **2013**, *57*, 3348.

[449]   a) A. M. Gustafson, E. S. Snitkin, S. C. J. Parker, C. DeLisi, S. Kasif, *BMC Genom.* **2006**, *7*, 265; b) P. Mazurkiewicz, C. M. Tang, C. Boone, D. W. Holden, *Nat. Rev. Genet.* **2006**, *7*, 929; c) L.-M. Bobay, H. Ochman, *Front. Genet.* **2017**, *8*, 72.

[450]   a) E. Castro-Nallar, Y. Shen, R. J. Freishtat, M. Pérez-Losada, S. Manimaran, G. Liu, W. E. Johnson, K. A. Crandall, *BMC Med. Genomics* **2015**, *8*, 50; b) E. A. Franzosa, T. Hsu, A. Sirota-Madi, A. Shafquat, G. Abu-Ali, X. C. Morgan, C. Huttenhower, *Nat. Rev. Microbiol.* **2015**, *13*, 360.

[451]   a) J. E. Bandow, H. Brötz, L. I. O. Leichert, H. Labischinski, M. Hecker, *Antimicrob. Agents Chemother.* **2003**, *47*, 948; b) F. J. Pérez-Llarena, G. Bou, *Front. Microbiol.* **2016**, *7*, 410; c) B. Peng, H. Li, X. Peng, *Expert Rev. Proteomics* **2019**, *16*, 829.

[452]   a) J. A. Cain, N. Solis, S. J. Cordwell, *J. Proteomics* **2014**, *97*, 265; b) B. Macek, K. Forchhammer, J. Hardouin, E. Weber-Ban, C. Grangeasse, I. Mijakovic, *Nat. Rev. Microbiol.* **2019**, *17*, 651.

[453]   G. A. N. Gowda, D. Djukovic, *Methods Mol. Biol.* **2014**, *1198*, 3.

[454]   G. J. Patti, O. Yanes, G. Siuzdak, *Nat. Rev. Mol. Cell. Biol.* **2012**, *13*, 263.

[455]   D. S. Wishart, *Nat. Rev. Drug. Discov.* **2016**, *15*, 473.

[456]   A. C. Goodrich, B. J. Harden, D. P. Frueh, *J. Am. Chem. Soc.* **2015**, *137*, 12100.

[457]   T. Kittilä, A. Mollo, L. K. Charkoudian, M. J. Cryle, *Angew. Chem.* **2016**, *55*, 9834.

[458]   C. A. Mitchell, C. Shi, C. C. Aldrich, A. M. Gulick, *Biochemistry* **2012**, *51*, 3252.

[459]   B. R. Miller, A. M. Gulick, *Methods Mol. Biol.* **2016**, *1401*, 3.

[460]   G. H. Hur, C. R. Vickery, M. D. Burkart, *Nat. Prod. Rep.* **2012**, *29*, 1074.

[461]   a) M. S. Butler, K. A. Hansford, M. A. T. Blaskovich, R. Halai, M. A. Cooper, *J. Antibiot.* **2014**, *67*, 631; b) G. Yim, M. N. Thaker, K. Koteva, G. Wright, *J. Antibiot.* **2014**, *67*, 31.

[462]   C. Brieke, V. Kratzig, K. Haslinger, A. Winkler, M. J. Cryle, *Org. Biomol. Chem.* **2015**, *13*, 2012.

[463]   a) M. Peschke, K. Haslinger, C. Brieke, J. Reinstein, M. J. Cryle, *J. Am. Chem. Soc.* **2016**, *138*, 6746; b) C. Brieke, M. Peschke, K. Haslinger, M. J. Cryle, *Angew. Chem.* **2015**, *54*, 15715; c) M. Peschke, C. Brieke, M. J. Cryle, *Sci. Rep.* **2016**, *6*, 35584.

[464]   J. G. Gall, *Methods* **2016**, *98*, 4.

[465]   J. M. Levsky, R. H. Singer, *J. Cell Sci.* **2003**, *116*, 2833.

[466]   a) A. Gozzetti, M. M. Le Beau, *Semin. Hematol.* **2000**, *37*, 320; b) D. Huber, L. Voith von Voithenberg, G. V. Kaigala, *Micro Nano Eng.* **2018**, *1*, 15.

[467]   I. Malih, M. R. Ahmad rusmili, T. Y. Tee, R. Saile, N. Ghalim, I. Othman, *J. Proteomics* **2014**, *96*, 240.

[468]   E. A. B. Undheim, B. R. Hamilton, N. D. Kurniawan, G. Bowlay, B. W. Cribb, D. J. Merritt, B. G. Fry, G. F. King, D. J. Venter, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 4026.

[469]   a) M. L. Mitchell, B. R. Hamilton, B. Madio, R. A. V. Morales, G. Q. Tonkin-Hill, A. T. Papenfuss, A. W. Purcell, G. F. King, E. A. B. Undheim, R. S. Norton, *Aust. J. Chem.* **2017**, *70*, 1235; b) B. Madio, S. Peigneur, Y. K. Y. Chin, B. R. Hamilton, S. T. Henriques, J. J. Smith, B. Cristofori-Armstrong, Z. Dekan, B. A. Boughton, P. F. Alewood et al., *Cell. Mol. Life Sci.* **2018**, *75*, 4511.

[470]   B. R. Hamilton, D. L. Marshall, N. R. Casewell, R. A. Harrison, S. J. Blanksby, E. A. B. Undheim, *Angew. Chem.* **2020**, *59*, 3855.

[471]   G. L. Glish, R. W. Vachet, *Nat. Rev. Drug. Discov.* **2003**, *2*, 140.

[472]   a) P. J. Jannetto, R. L. Fitzgerald, *Clin. Chem.* **2016**, *62*, 92; b) L. M. Heaney, D. J. Jones, T. Suzuki, *Future Sci. OA* **2017**, *3*, FSO213.

[473]   S. Singh, A. Chang, R. D. Goff, C. A. Bingman, S. Grüschow, D. H. Sherman, G. N. Phillips, J. S. Thorson, *Proteins* **2011**, *79*, 2181.