Development of novel mass spectrometry guided techniques for the discovery, structural elucidation and biological assessment of different peptide toxins

vorgelegt von

Diplom-Ingenieur (FH)

Daniel Petras

aus Worms

Von der Fakultät II – Mathematik und Naturwissenschaften

der Technischen Universität Berlin

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften

Dr. rer. nat.

genehmigte Dissertation

Promotionsausschuss:

Vorsitzende: Prof. Dr. Maria Andrea Mroginski

Gutachter: Prof. Dr. Roderich D. Süßmuth

Gutachter: Prof. Dr. Juan J. Calvete

Tag der wissenschaftlichen Aussprache: 05. Februar 2016

Berlin 2016

D83

Dedicated to all my teachers

Zusammenfassung

Natürliche Toxine sind von großer Bedeutung für den Menschen. Ihre biochemischen Eigenschaften resultieren in einem „*Ying und Yang*" an tödlichem und zugleich lebensrettenden Potenzial für uns. Etliche pharmazeutische Wirkstoffe basieren auf natürlichen Giften, insbesondere auf peptidischen Toxinen, welche jedes Jahr hunderttausende Menschenleben retten. Im gleichen Moment sind einige dieser Toxine verantwortlich für hunderttausende Todesfälle und fatale gesundheitliche Einschränkungen.

Durch die konstante Verbesserung bio-analytischer und speziell massenspektrometrischer Methoden sind wir auf dem Weg zu einem „kompletten Bild" an Informationen bezüglich Identität, Konzentration und Konformation bzw. räumliche Interaktionstelle von Naturstoffen oder Toxinen in verschiedenen biologischen Systemen. Diese neuen technologischen Möglichkeiten werden uns zwangsläufig zu einem umfassenden Verständnis von Toxinen und anderen Naturstoffen und ihrer Interaktion mit anderen biologisch relevanten Molekülen führen. Hierdurch werden wir das enorme Reservoir an solchen Verbindungen besser aufdecken und ihr Potential als Modellverbindungen, diagnostische Werkzeuge und schlussendlich als neue therapeutische Wirkstoffe nutzen.

In dieser Arbeit sind die Ergebnisse bezüglich der Suche, Strukturaufklärung, der biochemischen und pharmazeutischen Untersuchung zweier verschiedener Toxin-Klassen, unter Zuhilfenahme neuer massenspektrometrischer Analysemethoden, gezeigt und diskutiert.

Das erste Beispiel handelt von dem nicht-ribosomal synthetisierten Peptid-Polyketid-Hybrid Albicidin, welches von dem Zuckerrohr-Pathogen *Xanthomonas albilineans* synthetisiert wird. Durch die Sequenzierung des Genoms von *Xanthomonas albilineans* war es uns möglich die einzigartige biologische Nische und den besonderen Genotyp, innerhalb der Xanthomonaden, welcher etliche ungewöhnliche Biosynthese Gencluster enthält, zu beschreiben. Einer dieser Gencluster ist für die Biosynthese des Phytotoxins Albicidin, einem starken DNA-Gyrase Inhibitor, welcher auch hochwirksam gegen Gram-positive und Gram-negative Bakterien ist, verantwortlich. Die Strukturaufklärung von Albicidin deckte ein bisher unbekanntes chemisches Gerüst auf, welche bis auf Cyano-Alanin und Zimtsäure ausschließlich aus para-Aminobenzoesäure-Derivaten besteht. Die heterologe Produktion und darauf aufbauende *in vitro* Experimente verschiedener Module der Nicht-Ribosomalen-Peptidsynthetase gab uns eine detaillierte Einsicht in die Biosynthesemaschinerie, was uns half ein umfassendes Biosynthesemodell für Albicidin aufzustellen.

Durch *non-targeted* Tandem-Massenspektrometrie (MS/MS) konnten wir mehrere Albicidin Derivate, wie z.B. N-terminal carbamoyliertes Albicidin identifizieren. Die Struktur von Carbamoyl-Albicidin konnte durch hochauflösende MS/MS aufgeklärt und die Biosynthese ebenfalls durch heterologe Produktion der Carbamoyltransferase und in *vitro Assays* sowie einer Geninaktivierungsmutante verifiziert werden.

Die Totalsynthese von Albicidin sowie Carbamoyl-Albicidin ermöglichte uns schließlich die detaillierte Testierung der biologischen Aktivitäten durch *in vitro* Gyrase-inhibierung und Testierung der antibakteriellen Eigenschaften. Carbamoyl-Albicidin zeigte hierbei stärkere inhibitorische Effekte gegenüber Gyrase als Albicidin, was dafür spricht, dass die Carbamoylierung ein Teil der biosynthetischen Reifung von Albicidin ist.

Der zweite Teil dieser Arbeit handelt von der Einführung und ersten Anwendung von *Top-down* Massenspektrometrie für die Analyse von Schlangengiften. In einer Machbarkeitsstudie haben wir

Zusammenfassung

Gift der Königskobra, *Ophiophagus hannah,* mit Hilfe einer *Top-down* Methode analysiert und etliche Toxine verschiedener Proteinfamilien nachgewiesen. Unter diesen Toxinen waren hauptsächlich Drei-Finger-Toxine, ein Kunitz-artiger Proteaseinhibitor, Ohanin, eine bisher unbekannte A2-Phospholipase sowie ein Cystein-reiches-Sekretionsprotein (Ophanin). Eine komplementäre *bottom-up* Analyse führte zu einem kompletten, Lokus-aufgelösten Venom-Repertoire der Königskobra - der größten bekannten Giftschlange.

Als erstes Beispiel für eine Otter-Spezies haben wir das Venom der Anatolischen Wiesenotter, *Vipera anatolica*, durch eine Kombination aus *Top-down* und *Bottom-up Venomics* analysiert. Durch diese Analyse konnten wir etliche Metalloproteasen, Cystein-reiche-Sekretionsproteine, einen Metalloproteaseinhibitor, A2-Phospholipasen, Disintegrine, eine Serinprotease, Typ-C Lektine sowie einen Kunitz-artigen Proteaseinhibitor identifizieren. Des Weiteren konnten wir etliche Isoformen der oben beschriebenen sowie unbekannten Toxine detektieren, was für eine große, bisher nicht beschriebene Komplexität von Schlangengiften spricht, die mit herkömmlichen *bottom-up* Techniken unentdeckt geblieben wäre. Schlussendlich konnten wir hierdurch eine bioaktive Komponente des Venoms als ein heterodimeres Disintegrin identifizieren, welches vielversprechende Aktivität gegen Krebszellen zeigte.

Abstract

Natural toxins play an important role for human health. Their biochemical properties result in a "*ying and yang*" of deadly and lifesaving potential. Several drugs are based on natural and especially peptidic toxins, saving hundred thousands of human lives every year. At the same time, some of these toxins are responsible for hundreds of thousands of deaths and permanent health issues.

With the constant improvement of bioanalytical and especially mass spectrometric techniques, progress has been made towards the accomplishment of "complete pictures" of identities, concentration and spatial location of biomolecules, and in particular toxins, in different biological systems. These technological improvements will enable us to gain a more complete understanding of toxins and all other natural products and their biological interactions. Furthermore these improvements will enable us to better uncover and utilize their enormous potential as biomedical models, diagnostic tools and ultimately novel therapeutics.

In this work, the discovery, structural elucidation, biochemical investigation and finally assessment as potential drugs of two classes of natural toxins are shown and discussed.

The first part is dedicated to the non-ribosomally synthesized peptide-polyketide-hybrid albicidin, which is produced by a sugar cane pathogenic bacterium, *Xanthomonas albilineans.* By means of the sequencing of the genome of *Xanthomonas albilineans* we were able to describe its unique biological niche and special genotype among other xanthomonads, containing a number of unusual biosynthetic gene clusters. One of these gene clusters is responsible for the biosynthesis of the phytotoxin albicidin, a potent DNA gyrase inhibitor, which is also highly active against Gram positive and Gram negative bacteria. The structure elucidation revealed the hitherto unknown structure of albicidin, revealing a unique polyaromatic oligopeptide mainly composed of *p*-amino benzoic acids and cyano-alanine. Heterologous production and *in-vitro* studies of several domains of the non-ribosomal peptide synthetase rendered detailed insights into the biosynthetic machinery, revealing a comprehensive biosynthetic model for albicidin. Besides the main metabolite, we found several derivatives by non-targeted tandem mass spectrometry (MS/MS) experiments, for example an N-terminally carbamoylated version of albicidin. Carbamoyl-albicidin was characterized by high resolution MS/MS from cultures of *X. albilineans* and the putative biosynthesis confirmed by gene inactivation of the carbamoyltransferase, heterologous production of the carbamoyltransferase and *in vitro* reconstitution of the carbamoylation reaction. The chemical synthesis of both albicidin and carbamoyl-albicidin finally enabled us to assess their bioactivities by means of *in vitro* gyrase inhibition and antibacterial assays. Carbamoyl-albicidin showed stronger inhibitory effects against the gyrase compared to albicidin, which identifies the carbamoylation as an important biosynthetic step of albicidin maturation.

In the second part of this work, the first application of top-down mass spectrometry in snake venomics is shown and discussed. In a case study with venom from king cobra, the application of top-down mass spectrometry enabled us to identify several proteins from different toxin families. Among those toxins are eleven three-finger toxins (7-7.9 kDa), a Kunitz-type inhibitor (6.3 kDa), ohanin (11.9 kDa), a novel phospholipase A_2 molecule (13.8 kDa) and a cysteine-rich secretory protein (CRISP) ophanin (25 kDa). Complementary bottom-up MS/MS analyses contributed to complete a locus-resolved venom phenotypic map for *Ophiophagus hannah*, the world's largest venomous snake.

Abstract

As a first example for a viperid species we characterized the venom proteome *Vipera anatolica*, the Anatolian Meadow Viper by a combination of top-down and bottom-up venomics. From this analysis we identified snake venom metalloproteases, cysteine-rich secretory protein isoforms, a metalloprotease inhibitor, several type A2 phospholipases, disintegrins, a snake venom serine protease, a C-type lectin and a Kunitz-type protease inhibitor. Furthermore, we detected several isoforms of above mentioned proteins as well as previously unknown proteins, indicating an extensive complexity of the venom which would have remained undetected with conventional venomic approaches. With these results we could identify one of the bioactive fractions of the venom to be a hetero-dimeric disintegrin, which showed remarkable activity against cancer cells.

Publications

All parts of chapter 2-6 of this thesis have been published in the articles listed below. Detailed author contributions are stated on the next page.

"What makes *Xanthomonas albilineans* unique amongst xanthomonads?"
Pieretti I, Pesic A, Petras D, Royer M, Süssmuth RD and Cociancich S
Frontiers in Plant Science, 2015, published online (doi:10.3389/fpls.2015.00289)

"The gyrase inhibitor albicidin consists of para-aminobenzoic acids and cyanoalanine"
Cociancich S*, Pesic A*, Petras D*, Uhlmann S, Kretz J, Schubert V, Vieweg L, Duplan S, Marguerettaz M, Noëll J, Pieretti I, Hügelland M, Kemper S, Mainz A, Rott P, Royer M and Süssmuth RD
Nature Chemical Biology, 2015, 11(3):195-197 (doi:10.1038/nchembio.1734)

"The *O*-carbamoyltransferase Alb15 is responsible for the modification of albicidin"
Petras D, Kerwat D, Pesic A, Hempel BF, Von Eckardstein L, Semsari S, Arasté J, Marguerettaz M, Cociancich S, Royer M and Süssmuth RD
ACS Chemical Biology, 2016, submitted

"Venom proteomics of Indonesian king cobra, *Ophiophagus hannah*: integrating top-down and bottom-up approaches"
Petras D, Heiss P, Süssmuth RD and Calvete JJ,
Journal of Proteome Research, 2015, 14(6): 2539–2556 (DOI: 10.1021/acs.jproteome.5b00305)

"Mass spectrometry guided venom profiling and bioactivity screening of the Anatolian Meadow Viper, *Vipera anatolica*"
Gocmen B*, Heiss P*, Petras D, Nalbantsoy A and Süssmuth RD
Toxicon, 2015, 1(107):163-174. (doi:10.1016/j.toxicon.2015.09.013)

(*these authors contributed equally)

Author contributions

"What makes *Xanthomonas albilineans* unique amongst xanthomonads?"
Pieretti I, Pesic A, Petras D, Royer M, Süssmuth RD and Cociancich S
I.P., A.P., **D.P.**, M.R., R.D.S. and S.C. reviewed the literature and wrote the manuscript.

"The gyrase inhibitor albicidin consists of para-aminobenzoic acids and cyanoalanine"
Cociancich S, Pesic A, Petras D, Uhlmann S, Kretz J, Schubert V, Vieweg L, Duplan S, Marguerettaz M, Noëll J, Pieretti I, Hügelland M, Kemper S, Mainz A, Rott P, Royer M and Süssmuth RD
S.C., A.P., S.D., M.M., J.N., I.P., M.H., P.R. and M.R. performed the cultivation and isolation of albicidin. A.P., **D.P.**, V.S., S.K. and A.M. performed the structural elucidation of albicidin. **D.P.**, S.U., A.M. and M.R. performed the bioinformatic analysis of the biosynthesis genes. **D.P.**, S.U. and M.H. produced the proteins and performed the in vitro assays. J.K. synthetized the substrates for the in vitro assay. S.C., A.P., **D.P.**, A.M., M.R. and R.D.S. designed the study and analyzed the data. S.C., A.P., **D.P.**, L.V., A.M., M.R. and R.D.S. wrote the manuscript.

"**The *O*-carbamoyltransferase Alb15 is responsible for the maturation of albicidin**"
Petras D, Kerwat D, Pesic A, Hempel BF, Von Eckardstein L, Semsari S, Arasté J, Marguerettaz M, Cociancich S, Royer M and Süssmuth RD
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.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.

"Venom proteomics of Indonesian king cobra, *Ophiophagus hannah*: integrating top-down and bottom-up approaches"
Petras D, Heiss P, Süssmuth RD and Calvete JJ,
D.P. performed the Top-down measurements. **D.P.** and P.H. performed the bottom-up experiments. **D.P.** and J.J.C. designed the study and analyzed the data. **D.P.**, R.D.S. and J.J.C. wrote the manuscript.

"Mass spectrometry guided venom profiling and bioactivity screening of the Anatolian Meadow Viper, *Vipera anatolica*"
Gocmen B, Heiss P, Petras D, Nalbantsoy A and Süssmuth RD
B.G. collected the snakes and extracted the venom. P.H. and **D.P.** performed the top-down and bottom-up venomics experiments. P.H. fractionated the venom. A.N. performed the bioactivity assays. **D.P.** and A.N. designed the study and analyzed the data. **D.P.** and R.D.S wrote the manuscript.

Publications which were prepared during my doctoral studies that are not part of this thesis

"Top-down venomics of the east African green mamba, Dendroaspis angusticeps, and the black mamba, Dendroaspis polylepis, highlights the complexity of their toxin arsenals"
Petras D, Heiss P, Harrison RA, Suessmut RD, Calvete JJ
Molecular Cellular Proteomics, 2016, submitted

"Deuterium-labeled precursor feeding reveals a new *p*ABA-containing meroterpenoid from the mango pathogen *Xanthomonas citri pv. mangiferaeindicae*"
Saleh H, Petras D, Mainz A, Kerwat D, Nalbantsoy A, Erzurumlu Y and Süssmuth RD
Journal of Natural Products, 2016, submitted

"The Minimum Information about a Biosynthetic Gene cluster (MIBiG) specification"
Medema MH, Kottmann R, Yilmaz P, Petras D, Süssmuth RD, Glöckner FO *et al.*
Nature Chemical Biology, 2015; 11(9):625-631. (doi:10.1038/nchembio.1890)

Content

Content

1   Motivation

The term toxin refers to any poisonous compound produced by a living cell or organism and was first used by the German Professor Ludwig Brieder (1849-1919) [1]. There is no kingdom of life which has not evolved toxins at a certain point of evolution for the purpose of hunting, defense or competitor dissuasion [2-5]. Every known ecosystem on earth contains organisms capable of producing deadly toxins including over 170,000 venomous species throughout all major phyla of the evolutionary tree of animals [2]. Due to their high degree of target specificity, evolved over eons between a predator and its prey, toxins represent also a vast and largely untapped source of pharmacological tools and lead compounds for drug development, a fact that attracts increasing interest in academic, industrial and medical arenas [6-10].

The medical potential of toxins was already known in the ancient world, where it was used in small doses in traditional medicine [6]. The first usage of a toxin in modern medicine dates back to the 1940s when tubocurarine, one of the active ingredients of curare, was used for anesthetic practice [11]. Tubocurarine is a nicotinic acetylcholine receptor inhibitor which blocks the transmission of signals from motor nerves to skeletal muscles causing muscle paralysis [12, 13]. Afterwards other curare-like alkaloids of the South American arrow poisonous plants led to the development of new muscle relaxants with less side effects [12]. The most successful muscle relaxants, atracurium besilate is built by the linkage of two curare-like alkaloids and was first synthesized 1974 by George H. Dewar [14, 15]. Today atracurium besilate is considered as one of the most important drugs employed in basic health system [16]. Another prominent example of the medical potential of a plant-derived toxin is ricin. Ricin is a highly toxic lectin found in the seeds of castor oil plants [17]. Although there is no FDA-approved drug yet, ricin has the potential to be used as a terminator agent against tumor cells by linking it to a monoclonal antibody to guide it towards tumor specific receptors [18]. Very similar approaches are currently investigated in clinical trials with recombinant toxin-antibody conjugates of pokeweed antiviral protein from *Phytolacca americana* [19], saporin from

Saponaria officinalis [20] and gelonin from *Gelonium multiflorum* [21]; which are all targeting the ribosome and protein biosynthesis [22].

The first recombinant toxin therapeutic interacting with interleukin-2 receptor that got an FDA approval was the diphtheria toxin. Diphtheria toxin or diftitox is a secreted protein from the bacterium *Corynebacterium diphtheriae*, the causative agent of diphteritis [23], and is sold as Ontak [24]. Ontak is used in the treatment of cutaneous T-cell lymphoma, a type of cancer of the immune system. Another very prominent example of bacterial toxins with medical usage are botulinum toxins A and B, produced from *Clostridium botulinum*. Botulinum toxin A was introduced the late 1980s for the treatment of cervical dystonia and axillary hyperhidrosis [25]. More public attention is nevertheless paid for the usage of botulinumtoxin A for cosmetic purposes, where it reached questionable glory as botox which is used to prevent the development of wrinkles via a relaxant effect on facial muscles [26]. Other examples for the medical potential of bacterial toxins, mainly in the context searching new cancer treatments, are toxin A and toxin B from *Clostridium difficile* [22, 27] and pseudomonas exotoxin from *Pseudomonas aeruginosa* [22, 28]. Besides these proteinaceous toxins, bacteria, or more general microorganisms offer an enormous source of small molecules with therapeutic use or potential which revolutionized the pharmaceutical industry [29]. The grain-contaminating genus of fungi, Fusarium spp., produce a whole set of mycotoxins [30] containing the antibiotic non-ribosomal synthesized (depsi)peptide (NRP) enniatin, which is used for the treatment of oral bacterial infections [31, 32]. However, as most microbial natural products do not show toxic properties against higher eukaryotes, they are not considered as toxins, although they luckily can act quite poisonous against other microorganisms - as for example the antibiotics penicillin [33], erythromycin [34] and vancomycin [35].

But also with these great drugs available, we are facing a growing problem of multiresistant bacterial strains forcing us to constantly search for novel antibacterial compounds. In particular, the so called ESKAPE pathogens (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, and Enterobacter*) are the cause of serious health problems in a global scale as for example increasing number of untreatable infections in hospitals [36]. Despite the huge efforts in the search for new anti-infective agents by applying new technologies, such as high-throughput screening and rational drug design, most new drug candidates failed to work effectively against gram-negative bacteria [37]. A major obstacle in the treatment of gram-negative bacteria is the outer cell membrane which represents an insuperable barrier to many newly discovered compounds [38]. A current trend in pharmaceutical research of antibacterial compounds is the comeback of natural products. Through the latest developments in DNA

sequencing technologies and bioinformatics approaches we stand at a tipping point in the discovery of untapped natural products of microbial ecosystems [39-42], which contain an enormous potential of antibacterial agents evolved over eons of years in the constant fight between competing organisms [43].

A promising new compound found in "uncultivable" soil bacteria is the peptide antibiotic teixobactin [44] which received huge attention in the scientific community [45-47] and public media [48-51]. Teixobactin, a non-ribosomally synthesized peptide, shows high antibacterial activity against several resistant strains. It acts as a lipid II binding antibiotic [44] but is structurally distinct from other lipid II binders such as glycopeptides and lantibiotics [52, 53]. First experiments regarding the development of bacterial resistance in a 27 day trial, where different bacterial strains were incubated with sub-lethal doses of teixobactin, showed promising results [44]. These results led to the assumption in some public news organs that teixobactin would not face bacterial resistance. Taking into account the lessons learned about similar claims made for former newly discovered antibiotics, for example vancomycin, this seems rather unlikely to be the case [54].

A novel antibiotic, also active against gram negative strains, is the non-ribosomally synthesize peptide-polyketide hybrid (NRP-PK) albicidin [55-57]. Albicidin, primarily known as a phytotoxin produced by the xylem invading sugar cane pathogen *Xanthomonas albilineans* [58, 59], induces chlorosis through the inhibition of the plant DNA replication [60]. The blockage of DNA replication is caused by the inhibition of the plant DNA-topoisomerase or gyrase which also affects a wide range of gram-positive and gram-negative bacteria with half maximal inhibitory concentrations (IC$_{50}$) around 40 nM [61]. Besides the promising IC$_{50}$ value, an interesting feature of albicidin is its mode of action which appears different from that of other gyrase inhibitors, such as coumarins and quinolones [60]. While coumarins act at the ATP binding side of the GyrB subunit [62], albicidin interferes in the covalent DNA binding of the GyrA subunit by trapping a conformational state [60]. The structure elucidation of albicidin, which was recently solved in our group, consists only of unusual non-canonical amino acids and α-methyl-4-hydroxycinnamic acid [57]. These features make albicidin a very interesting novel lead structure in the search for new anti-infective drugs. Through the establishment of the total synthesis of albicidin in our group [61] we are currently evaluating different synthetic derivatives in order to assess their structure-activity relationships.

From a different aspect, albicidin is a very interesting novel structure in the context of academic natural product research. The biosynthesis of albicidin is the first reported biochemical pathway of a non-ribosomal peptide synthetase (NRPS) assembling *para*-aminobenzoic acid derivatives and beta-cyanoalanine [57]. Interestingly, shortly after our findings, a number of *p*ABA derived NRPs, termed

cystobactamides were isolated from a *Myxobacteria* strain which shows conspicuous structural similarity to albicidin [63]. The structure of cystobactamides differs mainly at the *N*-terminal hydroxycinnamic acid and L-cyanoalanine as the central amino acid of albicidin, which is most likely produced at an insertional NRPS domain based on L-asparagine as a precursor [57] that interestingly is also present in the cystobactamide NRPS machinery. The gene cluster of albicidin and cystobactamide show in general a high homology, not only in amino acid similarity but also in the organization of the genes [57, 63], indicating a common evolutionary origin. The search and structural elucidation of albicidin and its carbamoyl derivative and the aim to understand their biosynthesis are part of this work and are discussed in detail in chapter 2, 3 and 4.

Moving to a different topic of this thesis, venomous secretions are found in many animals, from invertebrates to vertebrates thousands of toxic peptides and proteins are known which represent the most abundant and functionally relevant components of those deadly mixtures [64]. The medical potential of venoms is known since ancient times, where for example snake venom was used in small doses as a traditional medicine [6]. Until today six FDA-approved drugs based on animal venoms are on the market of which snake toxin derived compounds form the majority. The most prominent example might be captopril, a lifesaving anti-hypertensive drug which got FDA approval in the early 1980s. Captopril targets the angiotensin-converting enzyme and is derived from the structure of a bradykinin potentiating peptide from *Bothrops jararaca*, a pit viper found in the tropical and subtropical forests in South America [65, 66]. For the treatment of thrombosis in cardiovascular vessels, the glycoprotein IIb/IIIa targeting antiplatelet drugs, eptifibatide (derived from a toxin from *Sistrurus miliarius barbouri*) and tirofiban (derived by computer aided drug design from a toxin from *Echis carinatus*) were both introduced in the late 1990s [12, 67]. Hirudin is a naturally occurring thrombin inhibitor from medical leeches (*Hirudo medicinalis*) [68] and was used for the treatment of heparin-induced thrombocytopenia from the early 1990s until 2012. The very interesting glucagon-like peptide-1 (GLP-1) analogue, exenatide, was found in the saliva from the gila monster (*Heloderma suspectum*) a lizard found in North America [7]. Exenatide got an FDA approval in 2005 for the treatment of Type-II diabetes mellitus [69]. Besides these well-established drugs, several animal toxins are currently tested for their pharmaceutical potential in preclinical trials. A highly interesting source might be the high diversity of conotoxins acting on different neuro-receptors [7, 70-72]. Especially for the treatment of neuropathic pain, neurotoxins seem to be very interesting lead structures. A good example therefor is the success of ziconotide, a neurotoxin from a cone snail which got FDA approval for the treatment if neuropathic pain in 2004 [73, 74]. Other interesting analgesic compounds were found in the venom of black mamba (*Dendroaspis polyepsis*) [75]. The venom of black mamba consists mainly of three finger neurotoxins (3FTX) and Kunitz-type protease

inhibitors, also called dendrotoxins [76]. A novel class of mamba 3FTX termed mambalgins showed low toxicity in mice models but high anti-pain activity by blocking acid-sensing ion channels in peripheral and central neurons [12, 75, 77]. Other interesting analgesic compounds have been found in the venom of king cobra [78] which consists mainly of neurotoxic venom components [79, 80]. Several patents have been filed on analgesic peptide derivatives based on the alpha-neurotoxin hannalgensin [12, 81]. Other interesting potential anti-cancer drug leads come from viper venoms. Disintegrins, as for example echistatin and contortrostatin, inhibit angiogenesis and are currently being tested as anti-cancer therapeutics in mouse models [82, 83]. In chapter 6 of this work, the results of the venom characterization and bioactivity screening of *Vipera anatolica*, the Anatolian meadow viper are shown. We found thereby a novel hetero-dimeric disintegrin which showed remarkable selective activity against several cancer cell lines [84].

Paracelsus, a Swiss-German physician, botanist and alchemist, already stated in the 16th century that everything is toxic and only the doses decides if something might be a poison or not [85]: „Alle Dinge sind Gift, und nichts ist ohne Gift; allein die Dosis macht es, dass ein Ding kein Gift ist." [86] Thus, besides the above described lifesaving potential of natural toxins, accidental encounters with venomous creatures, toxic plants or pathogenic microorganisms can have of course fatal consequences.

Through the better understanding of natural toxins and pathogenic organisms on one hand and better hygiene on the other, health issues caused natural toxins however have been dramatically reduced within the last century [87]. For example, nowadays botulism, a mainly foodborne and potentially fatal disease, caused by the botulinum toxin A producing *Clostridium botulinum,* occurs only rarely [88]. From 1990-2000 only 263 cases were reported in the US [89], whereas in medieval times food poisoning through *Clostridium botulinum* had much more serious consequences [90]. Fatalities through intoxication with the plant toxin ricin for example are also fairly seldom. The probably best known event is the assassination of Georgi Markov, a Bulgarian dissident who was killed by Bulgarian secret service by firing a tiny pellet containing ricin into his leg [91]. More public attention for poisoning with natural toxins was achieved through the TV series "Breaking Bad", where the protagonist "Walter White" intended to poison a drug dealer named "Tuco" with ricin contaminated methamphetamine [92]. In another incident "Walter White" poisoned the 6 years old "Brock Cantillo" with berries from Lily of the Valley (*Convallaria majalis*) [93] containing cardiac glycoside as toxic ingredients [94].

Much more serious but mainly neglected health problems are envenomations caused by snake bites [95-98]. A rough estimate of global annual snake bites lies between 1 and 5.5 million cases, with up

to approximately 125,000 fatalities and around 400,000 victims with permanent health effects such as amputation caused by proceeding tissue necrosis [98-102]. As most of the accidents occur in tropical and subtropical underdeveloped countries with poor medical care systems, the envenomations from snakes can be called a "disease of poverty" [99]. The best treatment of snake envenomation is the neutralization of the toxins by antivenom. It is typically produced in horse by injecting the corresponding venom to induce an immune response of the animal to the toxins. The antibodies are then separated from the blood and intravenously applied to the snake bite victim with the goal to prevent mortality and reducing tissue damage [103]. The importance of a good distribution and affordability of antivenom can be seen for example in Africa, where a severe shortage has become a major problem [104], leading also to unethical distribution of inappropriate antivenoms with catastrophic consequences for the victims [105]. Here the development of new polyspecific antivenoms has greatest social-medical importance. Efforts were made for example by Instituto Clodomiro Picado in the development of EchiTAb-Plus-ICP, a polyspecific antivenom to handle the most common snake envenomations in Africa [106].

For the analysis of medically important peptides, mass spectrometry represents a sensitive and selective method [79, 107]. The bioanalytical capabilities have hereby been continuously enhanced by technological advances, as for example through the introduction of electrospray ionization (ESI) [108] or matrix-assisted laser desorption ionization (MALDI) [109]. The latest "game changer" has been the introduction of the orbitrap mass analyzer [110-112]. In the orbitrap the mass-dependent signal is recorded from a induced current produced by ion packets which oscillate around and along the spindle-shaped inner electrode of the trap [110]. In parallel also the improvement of time of flight mass spectrometers (TOF), mainly in combination with a quadrupole mass filter (QTOF) has improved the high throughput analysis of metabolities [113] and proteins [114].

For the identification of venom toxins, several "venomic" workflows have been established based on proteomic *bottom-up* strategies [2]. Typically, the characterization starts with an HPLC separation of the intact toxins, followed by N-terminal sequencing, SDS-PAGE, MALDI-TOF- and ESI-MS-analysis [115]. The different bands of the SDS-PAGE are then submitted to in-gel trypsin digestion and subsequently analyzed by LCMS [112, 115-121].

To assess antivenom *in vitro*, several workflows termed "antivenomics" have been introduced by Calvete and co-workers [115, 119, 122]. These assays are based on the immunodepletion of the toxins while incubation with antivenom, followed by subsequent HPLC analysis of the non-precipitated fraction [116]. For the *in vitro* assessment of Echi-Tab-Plus-ICP, we performed the antivenomic analysis [115, 122] which showed important neutralization capabilities and drawbacks

of the antivenom that ultimately aided to improve the production process and saved already first lives in clinical trials [106]. However, the current venomic techniques have several drawbacks, as for example toxin co-elution in the chromatographic separation or low mass resolution and poor quantification capability in the electrophoretic analysis. Most importantly, the proteolytic digestion of venom toxins results most often in difficulties to differentiate closely related toxin isoforms, reporting thus only a partial picture of venom complexity [79, 123]. In the last decade, high-resolution mass spectrometry has become increasingly available in many laboratories, making it possible to resolve high molecular mass proteins at isotopic levels [124]. The identification of intact proteins by mass spectrometry, termed top-down proteomics, has the potential to eliminate the above mentioned shortcomings of the standard venomic workflow [125, 126]. Top-down workflows, typically apply electrospray ionization and different gas-phase dissociation methods and high-resolution mass analysis, followed by bioinformatic comparison of precursor and product ion masses with species-specific protein sequence databases [124-126]. In Chapter 5 and 6 of this dissertation our efforts for the introduction of a top-down venomic workflow are shown and discussed.

A huge advantage in the analysis of ribosomally synthesized peptidic and proteinogenic toxins are the limited number of potential building blocks, mainly the 20 canonical amino acids including some potential post-translational modifications (PTM), such as phosphorylation, acetylation, disulfide formation [127] and building of thioether bridges [128]. As peptide fragmentation in gas phase ion reactions occurs mainly at the peptide bond [129], mass differences of differential fragmented peptides (n compared to n-1) are thus represented by one of the 20 amino acids masses. This predictable fragmentation makes it very easy to sequence a peptide or even a protein by tandem MS *de novo* [130], or to calculate theoretical spectra based on protein or nucleotide databases and match them to experimental mass spectra, so called peptide spectra matching [131].

The analysis of NRPs on the other hand is unfortunately not as straight forward, as these compounds are not limited to canonical amino acids and can thus have a huge amount of potential building blocks with partly unusual chemical moieties [132]. This higher number of potential mass differences with many possible isobaric moieties makes it challenging to elucidate NRPs structures *de novo.* Even though mass spectrometry is several orders of magnitudes more sensitive [133], most often NMR-Spectroscopy closes the missing link to ultimately unravel novel molecules, as for example in the structural elucidation of albicidin [57]. A main drawback of NMR-Spectroscopy however is the relatively high concentration and purity of sample required, because of which NMR structure elucidation is not well suited for high-throughput analysis of complex mixtures as for example whole metabolomes.

Aid for MS based structure elucidation comes from genomics and the bioinformatic analysis of biosynthesis gene clusters (BGC) [134]. Several tools have been developed to facilitate MS/MS based

structure elucidation by information derived from the biosynthetic logic of the corresponding BGCs [135-138].

Besides the *de novo* identification of metabolites or in particular peptides, a more powerful approach is the dereplication of compounds by spectra-spectra comparison in order to identify "*known knowns*" or "*known unknowns*" [133]. The terms "*known knowns*" and "*known unknowns*" are based on a quote of Donald Rumsfeld (US Secretary of Defense, 1975-1977 and 2001-2006) on "finding known unknowns" regarding weapons of mass destruction in Iraq [133, 139]. A widely applied method of mass spectral dereplication is the spectra-library matching as it is applied typically for electron impact fragment spectra (EI-MS) in gas-chromatography mass spectrometry (GCMS) [140]. More recently efforts were made to generate LC-MS/MS based libraries as for example massbank [141], metlin [142] or NIST [143]. A second way of dereplicating mass spectrometric data is the comparison of all MS/MS patterns within one dataset and the visualization of "spectral proximity" in a molecular network [133, 144]. Mass spectrometric networking has a great potential to better make use of the enormous amounts of non-targeted MS/MS spectra acquired in today's metabolic screening experiments. But not only for the dereplication of known and unknown metabolites [145-147], also for biosynthetic investigation [148, 149] and targeted search of peptides or metabolites containing certain building blocks [150], spectral networking has an enormous potential, which will be discussed in chapter 7.

In summary, mass spectrometry and in particular mass spectrometry coupled to liquid chromatography has become the fundamental tool for the search, structure elucidation and further assessment of novel bioactive peptides. In this dissertation our articles on the genomic investigation of *Xanthomonas albilineans* and the corresponding structural elucidation of albicidin and carbamoyl albicidin by mass spectrometry and NMR-Spectroscopy as well as the further biosynthetic and bioactivity investigations are shown in chapter 2, 3 and 4 and discussed in chapter 7. In the second part, our articles of the introduction of a top-down mass spectrometric approach for the characterization of snake venoms on the examples of king cobra and the Anatolian meadow viper in chapters 5 and 6 are shown and discussed in chapter 7.

2   What makes *Xanthomonas albilineans* unique amongst xanthomonads?

Introduction

Xanthomonas albilineans (Ashby) Dowson is known to invade the xylem of sugarcane and to cause leaf scald disease [151, 152]. Symptoms of this disease vary from a single, white, narrow, sharply defined stripe to complete wilting and necrosis of infected leaves, leading to plant death. Dissemination of *X. albilineans* occurs mainly mechanically through use of contaminated harvesting tools and by distribution and planting of infected cuttings. However, aerial transmission and potential for epiphytic survival have also been reported for this pathogen [153-155].

Xanthomonas albilineans is a representative of the genus *Xanthomonas*, members of which are exclusively Gram-negative plant-associated bacteria that collectively cause dramatic damage to hundreds of plant species of ornamental or agronomical interest. Indeed, both monocotyledonous (e.g. rice, sugarcane or banana) and dicotyledonous (e.g. citrus, cauliflower, bean, pepper, cabbage and tomato) plants are targeted worldwide by various *Xanthomonas* species. While sharing numerous phenotypic characteristics, at least 27 species and over 120 pathovars (variants of pathogeny) of the genus *Xanthomonas* are currently recognized. Each pathovar individually exhibits a very restricted host range and/or tissue-specificity and this leads to clustering of bacterial strains causing similar symptoms on the same host.

Multilocus sequence analysis (MLSA) with 5 housekeeping genes resulted in the distribution of *Xanthomonas* species in two clades. The main one contains the majority of species whereas the secondary clade contains *X. albilineans*, *Xanthomonas sacchari*, *Xanthomonas theicola*, *Xanthomonas hyacinthi* and *Xanthomonas translucen* [156]. Phylogenetic analyses with the *gyrB* sequence indicate that this secondary group also contains several uncharacterized species of *Xanthomonas* isolated mainly on rice, banana or sugarcane [157, 158]. Intriguingly, two multiMLSA studies with 28 genes and 228 genes, respectively, in which *X. albilineans* is the only representative of this secondary clade,

resulted in the branching of *Xylella fastidiosa* between *X. albilineans* and the main clade [159, 160]. *X. fastidiosa* is a xylem-limited bacterium, which is insect-vectored to a variety of diverse hosts, has a reduced genome and lacks the Hrp-T3SS (hypersensitive response and pathogenicity–type III secretion system) [161]. Analysis of the *X. albilineans* genome has revealed unusual features compared to other xanthomonads, the most prominent being the absence of the Hrp-T3SS gene cluster and the occurrence of genome erosion. Furthermore, to our knowledge, *X. albilineans* is the only xanthomonad that produces the phytotoxin albicidin. This mini-review aims to summarize the characteristics that, taken together, make *X. albilineans* so unique.

Genome erosion

The genome of *X. albilineans* strain GPE PC73 has been fully sequenced and annotated. It consists of a 3,768,695-bp circular chromosome with a G+C content of 63%, and three plasmids of 31,555-bp, 27,212-bp and 24,837-bp, respectively [162]. This genome size is much smaller than that of any other xanthomonad sequenced to date (commonly ~5 Mb). Examination of the genome of strain GPE PC73 together with OrthoMCL comparative analyses performed with other sequenced xanthomonads highlights several genomic features that distinguish *X. albilineans* from its near relatives [162-165].

Orthologous analyses show that *X. albilineans* and *X. fastidiosa* have experienced a convergent genome reduction during their respective speciation, with a more extensive genome reduction for *X. fastidiosa* [162]. Based on these analyses, *X. albilineans* has lost at least 592 genes that were present in the last common ancestor of the xanthomonads. Interestingly, most of these ancestral genes are conserved in the genome of *X. sacchari* strains NCPPB4393 and LMG 476 and the *Xanthomonas* spp. strains NCPPB1131 and NCPPB1132, which are the sequenced strains phylogenetically closest to *X. albilineans* [157, 158, 166]. This indicates that genome erosion is specific to *X. albilineans*. Convergent genome erosion of *X. albilineans* and *X. fastidiosa* could be linked to a similar adaptation to a xylem-invading lifestyle in which interactions with living plant tissues are minimal [162]. More recently, a study of the somewhat reduced genome of *Xanthomonas fragariae* (4.2 Mb), led to the hypothesis that the convergent genome reduction observed in some xanthomonads could be linked to their endophytic lifestyle and typically to their commitment to a single host [167].

Specific genes linked to a xylem-invading lifestyle

Although determinants for host- or tissue-specificity of *X. albilineans* remain unclear, the presence in its genome of genes encoding cell-wall-degrading enzymes (CWDEs) with specific features is probably important for its ability to spread in xylem and for pathogenicity. Indeed, all CWDEs from *X. albilineans* harbor a cellulose-binding domain (CBD) and a long linker region both adapted to the utilization of cell-wall breakdown products as carbon source and to the ability to spread in sugarcane xylem vessels [163]. These enzymes may also be required to disrupt pit membranes in sugarcane, thereby promoting propagation of the bacteria in the plant. Interestingly, *X. fastidiosa* also encodes two CWDEs containing a long linker and a CBD. It has been shown that one of these two CWDEs is involved in the spread of *X. fastidiosa* in the xylem by increasing the pore size of pit membranes. CWDE are therefore considered as virulence factors [168-170]. TonB-dependent transporters (TBDTs) may be used by *X. albilineans* to transport cell-wall-degrading products resulting from the activity of CWDEs, and thus may facilitate spread of the organism in the nutrient-poor conditions prevailing in the xylem of sugarcane. In the genome of *X. albilineans*, 35 TBDT genes have been identified, including one specific to this species and two others that are functionally associated to pathogenicity of the bacterium [163, 171].

Lack of Hrp-T3SS

Most phytopathogenic bacteria rely on the type III secretion system (T3SS) of the hypersensitive response and pathogenicity family (Hrp1 and Hrp2, respectively). This syringe-like apparatus allows pathogens to deliver into their host cells proteins (type III effectors) that modulate plant physiology and immunity for the benefit of the pathogen. Interestingly, genes encoding the injectisome and associated effectors of the Hrp-T3SS are missing in the genome of *X. albilineans*, as is also the case in the genomes of *X. sacchari* strains NCPPB4393 and LMG 476 and *Xanthomonas* spp. strains NCPPB1131 and NCPPB1132 [157, 158, 166]. Yet, an Hrp system is present in other close neighbour species of *X. albilineans*, such as *Xanthomonas translucens* pv. *graminis* strain 29, *X. translucens* pv. *translucens* strain DSM18974 and *X. translucens* strain DAR 61454 [172, 173]. Although the Hrp-T3SS is described as a crucial key component in plant–host interactions for most *Xanthomonas* spp, it seems not to be essential in *X. translucens* pv. *graminis* strain 29 for xylem colonization, although it is involved in symptom development [172]. Similarly, despite being devoid of any Hrp T3SS, *X. albilineans* displays pathogenicity and is able to cause serious damage to sugarcane.

Acquisition of a SPI-1 T3SS

The annotated sequence of the genome of *X. albilineans* strain GPE PC73 reveals the presence of a T3SS belonging to the *Salmonella* pathogenicity island-1 (SPI-1) injectisome family. Genes encoding this system are located near the terminus of the replication site of the chromosome and were probably acquired by lateral gene transfer. This secretion system, found mainly in mammals and insect bacterial pathogens or symbionts, exhibits high similarity to that described in *Burkholderia pseudomallei*—a human pathogen causing melioidosis [174]. The SPI-1 needle-like assemblies of *X. albilineans* strain GPE PC73 and *B. pseudomallei* strain K96243 are homologous. Both species share all but two genes—*orgA* and *orgB*, encoding putative oxygen-regulated invasion proteins involved in type three secretion that are not conserved in *B. pseudomallei*. The genome composition of the SPI-1 T3SS in *X. albilineans* additionally includes genes encoding translocon components (*xipB, xipC* and *xipD*), injectisome components (*xsaJ* to *xsaS* and *xsaV* to *xsaZ*) and a chaperone (*xicA*). Furthermore, the locus contains 15 additional genes referred to as *xapA–xapO*, encoding hypothetical proteins. These genes, which show homology neither to sequences from *B. pseudomallei* nor to sequences available from protein sequence databases, are specific to *X. albilineans* and their products represent good candidates to be considered as effectors for this SPI-1 T3SS [164]. Interestingly, this SPI-1 T3SS is conserved in *X. axonopodis* pv. *phaseoli* strains CFBP 2534, CFBP 6164 and CFBP 6982, which moreover possess a second T3SS belonging to the Hrp2 family [164, 175]. Pathogenicity of *X. albilineans* strains seems not to be linked to the presence of the SPI-1 T3SS in their genome; besides, no SPI-1 T3SS locus has been identified in strain PNG130 of *X. albilineans* even though it is able to spread in sugarcane. Functional analyses showed that, *in planta*, multiplication of a SPI-1 T3SS knockout mutant of *X. albilineans* was not impaired when compared to the wild-type, indicating that the SPI-1 T3SS is not required for spread in sugarcane vessels or for development of leaf scald symptoms. The role of the SPI-1 T3SS of *X. albilineans* remains unclear, although it has been conserved during its evolution in *X. albilineans* without frame-shifting indels or nonsense mutations [164]. It remains possible, in conditions other than those tested with our knockout mutant, that the SPI-1 T3SS system may be required for interaction with sugarcane, as in the case of SPI-1 of *Salmonella*, which is involved in interactions with *Arabidopsis thaliana* [176]. The SPI-1 T3SS system may also be associated with other aspects of the *X. albilineans* lifestyle, e.g. an involvement in adherence as reported for *Erwinia tasmaniensis* [177] or in formation of pellicle or biofilm-like structures [178] which could be related to epiphytic survival on sugarcane leaves. Although no insect vector has been identified for *X. albilineans* to date, we cannot rule out that SPI-1 T3SS could be involved in insect association or might mediate persistence of the bacterium in an insect vector as was shown for *Pantoea stewartii* [179].

Lack of t6ss and the xanthan gum gene cluster

X. albilineans lacks two other major pathogenicity factors that are common features of most xanthomonads. First, it lacks the gum gene cluster for extracellular polysaccharide (EPS) synthesis. This gene cluster is responsible for biofilm and xanthan gum formation, and is associated with pathogenesis in xanthomonads [180-182]. Exceptions are *X. fragariae*, which lacks the *gumN*, *gumO* and *gumP* genes, and *X. albilineans*, which lacks the complete set of gum genes, indicating those are not essential for virulence of both these pathogens [163, 167].

X. albilineans is also devoid of any type VI secretion system (T6SS) described in other xanthomonads, as for example in *X. fuscans* pv. *fuscans* strain 4834-R and *X. citri* subsp. *citri* strain 306, which each contain a single T6SS [183, 184] or *X. translucens* strain DAR61454, which encodes two distinct T6SS [173]. Structurally, the T6SS looks like an inverted bacteriophage. Functionally, this system is able to interact with both eukaryotic and prokaryotic cells by delivering effectors or toxins into host cells to subvert the signaling process to its own advantage, but also into other bacteria from the same habitat to outcompete them during infection [185, 186]. Despite its multifunctional roles during host–pathogen interactions, the lack of T6SS in *X. campestris* pv. *campestris* strain 8004, *X. gardneri* strain 101 and *X. albilineans* seems to have no effect on pathogenesis of these xanthomonads.

Albicidin and other non-ribosomally synthesized peptides

A unique feature of *X. albilineans* is the production of albicidin—a phytotoxin causing the white foliar stripe symptoms characteristic of leaf scald disease of sugarcane [55]. Albicidin is a potent DNA gyrase inhibitor that blocks the differentiation of chloroplasts (Figure 2.1). It also targets bacterial gyrase by a mechanism different from that of other DNA gyrase inhibitors like coumarins and quinolones [60]. This mode of action accounts for the potent antibacterial activity of albicidin, which inhibits the growth of Gram-positive and Gram-negative pathogenic bacteria at nanomolar concentrations [55]. Albicidin gives a competitive advantage to *X. albilineans* against other bacteria within the xylem vessels of sugarcane [187].
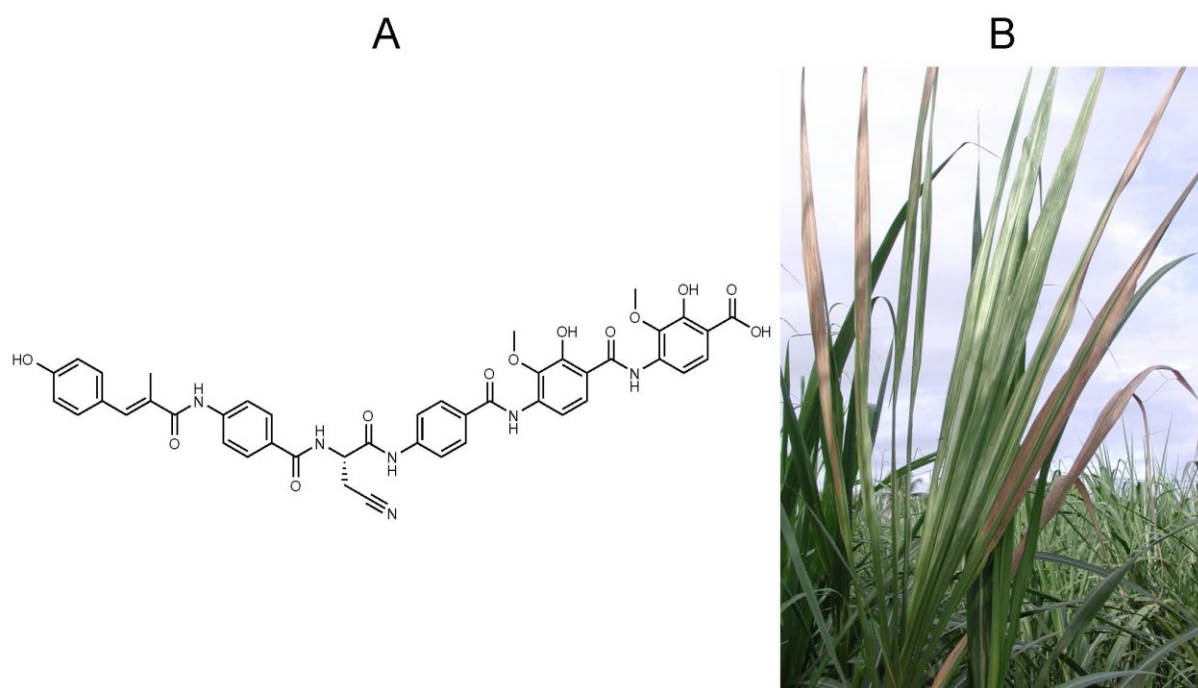
Figure 2.1. *Xanthomonas albilineans* produces the phytotoxin albicidin—a potent gyrase inhibitor that blocks chloroplast differentiation, resulting in sugarcane leaf scald disease symptoms. (A) Structure of albicidin, a hybrid PKS/NRPS compound with unique composition including *p*-aminobenzoic acid and cyanoalanine. (B) Diseased sugarcane plant with characteristic leaf scald symptoms: white foliar bleaching and necrosis of infected leaves (© J. H. Daugrois/Cirad).

Interestingly, two sugarcane-living bacteria harbor an albicidin resistance gene: *Leifsonia xyli* [188] and *Pantoea dispersa* [189]. Albicidin is produced by a hybrid polyketide synthase (PKS)/non-ribosomal peptide synthetase (NRPS) enzyme complex. PKS and NRPS genes are often clustered together with a large set of regulatory, transport or modification (tailoring) genes, as well as genes involved in the biosynthesis of non-proteinogenic amino acids. In addition to a phosphopantetheinyl transferase required for activation of the PKS/NRPS system and a HtpG chaperone, the role of which remains unclear, a locus (*alb* cluster) containing 20 genes is required for albicidin biosynthesis. Among these 20 genes, 3 encode the PKS/NRPS system; 15 others act as transport, regulatory, modification or resistance genes [56]. NRPSs are multimodular megasynthetases used by bacteria and fungi to produce peptides in a ribosome-independent manner [190]. Each module governs the specific incorporation of an amino acid substrate based on signature sequences in the adenylation (A) domains [191], which are loaded onto peptidyl carrier protein (PCP) domains. Elongation of the peptide is mediated by condensation (C) domains present within each module. PKSs function according to the principles of fatty acid biosynthesis [192].

For decades, the structure elucidation of albicidin was impeded by its extremely low production yield by *X. albilineans*. A first step to overcome this bottleneck was achieved by transferring the biosynthetic genes into a heterologous host, namely *X. axonopodis* pv. *vesicatoria*, resulting in a significant increase in albicidin production [193]. Extensive HPLC purification of albicidin and thorough analysis of the purified compound by means of mass spectrometry and nuclear magnetic resonance spectroscopy then allowed us to unravel its unique structure (Figure 1). Albicidin proved to be a linear pentapeptide composed of cyanoalanine and *p*-amino benzoic acids N-terminally linked to a *p*-coumaric acid derivative [57]. Although over 500 different monomers (amino acid substrates) have been identified to date as being incorporated by NRPS systems, elucidation of the structure of albicidin revealed for the first time the incorporation by NRPS of cyanoalanine and *p*-amino benzoic acids. Moreover, the incorporation of *p*-amino benzoic acids is the first example of incorporation of a δ-aminoacid by NRPS, since all NRPS described to date incorporate only α or β aminoacids. The use of unusual amino acid substrates is linked to unique features that were identified *in silico* ten years ago within the albicidin NRPS modules sequence [56]. The formation and incorporation of cyanoalanine most likely occurs *in situ* through an additional module present in the PKS-NRPS assembly line that was investigated in one of our present studies [57]. Chemical synthesis of albicidin is now available, allowing either production of high quantities of the compound for further study of its mode of action and activity spectrum, or the synthesis of analogues [61]. The uniqueness of its structure and the specific mode of action of this compound make albicidin a strong lead structure for antibiotic development.

Data mining of the genome of *X. albilineans* strain GPE PC73 has led to the identification, in addition to the albicidin biosynthesis locus, of five other NRPS loci [163, 165]. The first, named Meta-B, encodes megasynthases performing peptidic elongation of a 16-amino acid lipopeptide. This locus also encodes a transcription regulator belonging to the AraC family, a cyclic peptide transporter, and enzymes involved in biosynthesis of the non-proteinogenic amino acids di-amino butyric acid and dihydroxyphenylglycine. Interestingly, the NRPS gene locus Meta-B has been identified in the genome of strains of three other *Xanthomonas* species, namely *X. oryzae* pv. *oryzae* strains BAI3 and X11-5A, *X. translucens* strain DAR61454 and *Xanthomonas* spp. strain XaS3 [165]. Despite a similar organization of the genes within these loci, the *in silico* prediction of the sequences of the peptides produced indicates that each strain produces a different lipopeptide.

Two other NRPS gene clusters, Meta-A and Meta-C, have been identified in the genome of *X. albilineans* strain GPE PC73. These encode megasynthases that perform the biosynthesis of peptides of 12 and 7 amino acids, respectively. A partial sequence has been predicted for each of these

peptides [165]. Finally, two short NRPS genes have also been identified on the chromosome of *X. albilineans*: they both encode only one NRPS module. Interestingly, there is an overlap between both these genes and a gene encoding a glycosyltransferase. It has been hypothesised that these genes encode glycosylated amino acids, to which, however, no precise function could yet be attributed [165].

Conclusions

Although most xanthomonads require pathogenicity factors such as *gum* genes, T3SS Hrp and T6SS for survival, growth and spread within host plants, *X. albilineans* lacks these pathogenicity factors, *de facto* reducing its artillery to circumvent sugarcane defense mechanisms and innate immunity. While being "disarmed" could be disadvantageous for a vascular plant pathogen, *X. albilineans* remains able to invade and spread in sugarcane, suggesting that it uses other strategies, such as stealth, i.e. being unobtrusive *in planta*, to minimise inducible host defense responses. On the other hand, the reduced genome of *X. albilineans* has specific features that may be involved in the adaptation of the bacterium to live and spread in sugarcane xylem vessels. For example, specific CWDEs and TBDTs appear to be optimized for life in the nutrient-poor sugarcane xylem environment. The uniqueness of *X. albilineans* resides also in the production of the phytotoxin and antibiotic albicidin. The recently unraveled structure and concomitant development of a chemical synthesis protocol for this compound leads to additional prospects for its use in the antibiotherapy field. According to the specificities deriving from the biological, biochemical, phylogenetic and genomic analyses described in this review, one can truly say that *X. albilineans* is quite unique amongst the genus *Xanthomonas* or, in other words, a xanthomonad on the edge.

3 The gyrase inhibitor albicidin consists of para-aminobenzoic acids and cyanoalanine

Introduction

Bacteria of the genus *Xanthomonas* belong to a large group of Gram-negative plant pathogens causing enormous economic losses in various crops, *e.g.* sugarcane, rice, soybean, cotton, wheat and citrus species [156]. *Xanthomonas albilineans* is a xylem invading plant pathogen, which causes leaf scald disease in sugarcane [194]. Sugarcane is cultivated on 26 million hectares at more than 60 locations around the world and is used for the production of sugar, bioethanol and rum [195]. Albicidin has been already described in 1983 as a phytotoxic compound of *X. albilineans* [58], which blocks the DNA gyrase in sugarcane chloroplasts [196, 197]. Apart from its significance for the pathogenicity of *X. albilineans*, albicidin is bactericidal at low nanomolar concentrations to a wide range of Gram-positive and Gram-negative bacteria, e.g. *Enterobacter aerogenes*, *Escherichia coli*, *Haemophilus influenza*, *Klebsiella pneumonia*, *Shigella sonnei* and *Staphylococcus aureus* [198]. It has been shown that albicidin targets the bacterial DNA gyrase (topoisomerase II) [60], which is an essential enzyme for DNA replication, transcription and gene regulation due to its ability to modulate the extent of DNA supercoiling [199]. Albicidin is a potent inhibitor of this supercoiling activity with a 50% inhibitory concentration of ~ 40 nM, comparable to those of other important DNA gyrase inhibitors, e.g. the coumarins and quinolones [60]. Whilst the coumarins affect the ATPase activity of the GyrB subunit [62], albicidin interferes in the catalytic DNA cleavage-religation cycle of the GyrA subunit and traps a different conformational state than that targeted by quinolones [60]. These unique characteristics qualify albicidin as a potential lead compound for antibacterial drug development. However, structural characterization of albicidin has been impeded for decades owing to the extremely low production rate of albicidin in its original host organism *X. albilineans* and the unknown substrates of its enigmatic hybrid polyketide synthetase/non-ribosomal peptide synthetase (PKS/NRPS) machinery [200]. The development of a viable heterologous expression system in

X. axonopodis pv. *vesicatoria* resulted in a significant increase in albicidin yields [200] and enabled us to gain an in-depth understanding of its chemical structure and its biosynthesis.

Results and Discussion

The obtained structure of albicidin is depicted in Figure 3.1. As can be seen, the phytotoxin is composed of the non-proteinogenic α-amino acid β-cyano-*L*-alanine (Cya-3) as well as the aromatic δ-amino acids *p*-aminobenzoic acid (pABA-2, pABA-4) and 4-amino-2-hydroxy-3-methoxybenzoic acid (pMBA-5, pMBA-6), respectively. The pentapeptide of these building blocks is N-terminally attached to 3-(4-hydroxyphenyl)-2-methyl acrylic acid, which can be regarded as a methyl derivative of *p*-coumaric acid (MCA-1).

Structural analysis by means of high resolution mass spectrometry (HR-MS) and nuclear magnetic resonance (NMR) spectroscopy required sufficient amounts of albicidin. We therefore optimized the heterologous production of albicidin in *X. axonopodis* pv. *vesicatoria* and established an extended purification protocol using preparative high-pressure liquid chromatography (HPLC). The purification procedure was guided by antibacterial tests to ensure isolation of bioactive albicidin. Final yields of purified albicidin were in the range of 1 mg per 100 liter cell culture. Comparison of albicidin samples produced by the original host *X. albilineans* and the heterologous host *X. axonopodis* pv. *vesicatoria* verified identical molecules (Supplementary Fig. 1 in appendix A).
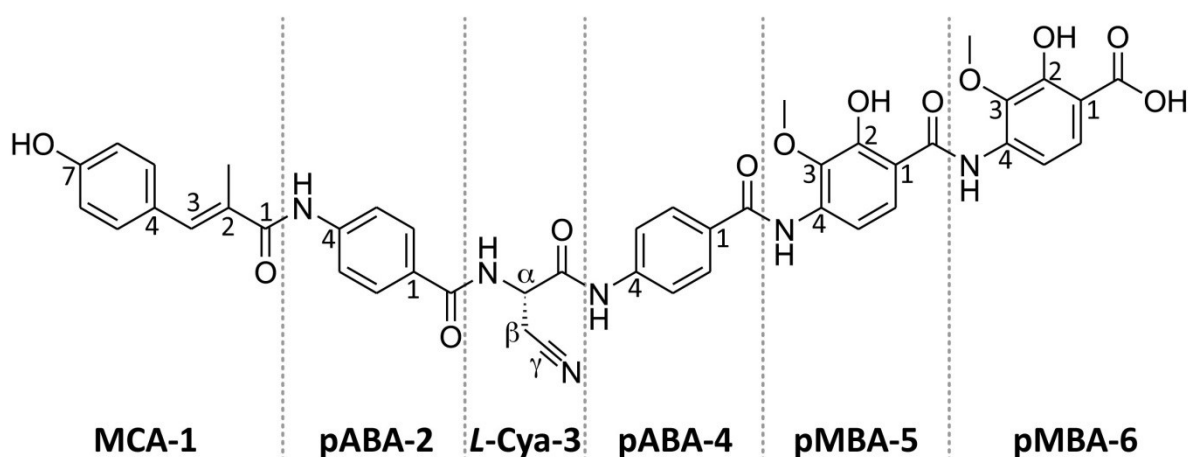
Figure 3.1. The structure of the phytotoxin albicidin. Albicidin is composed of a methylated derivative of *p*-coumaric acid (MCA-1), the non-proteinogenic α-amino acid β-cyano-*L*-alanine (*L*-Cya-3) as well as the aromatic δ-amino acids *p*-aminobenzoic acid (pABA-2, pABA-4) and 4-amino-2-hydroxy-3-methoxybenzoic acid (pMBA-5, pMBA-6).
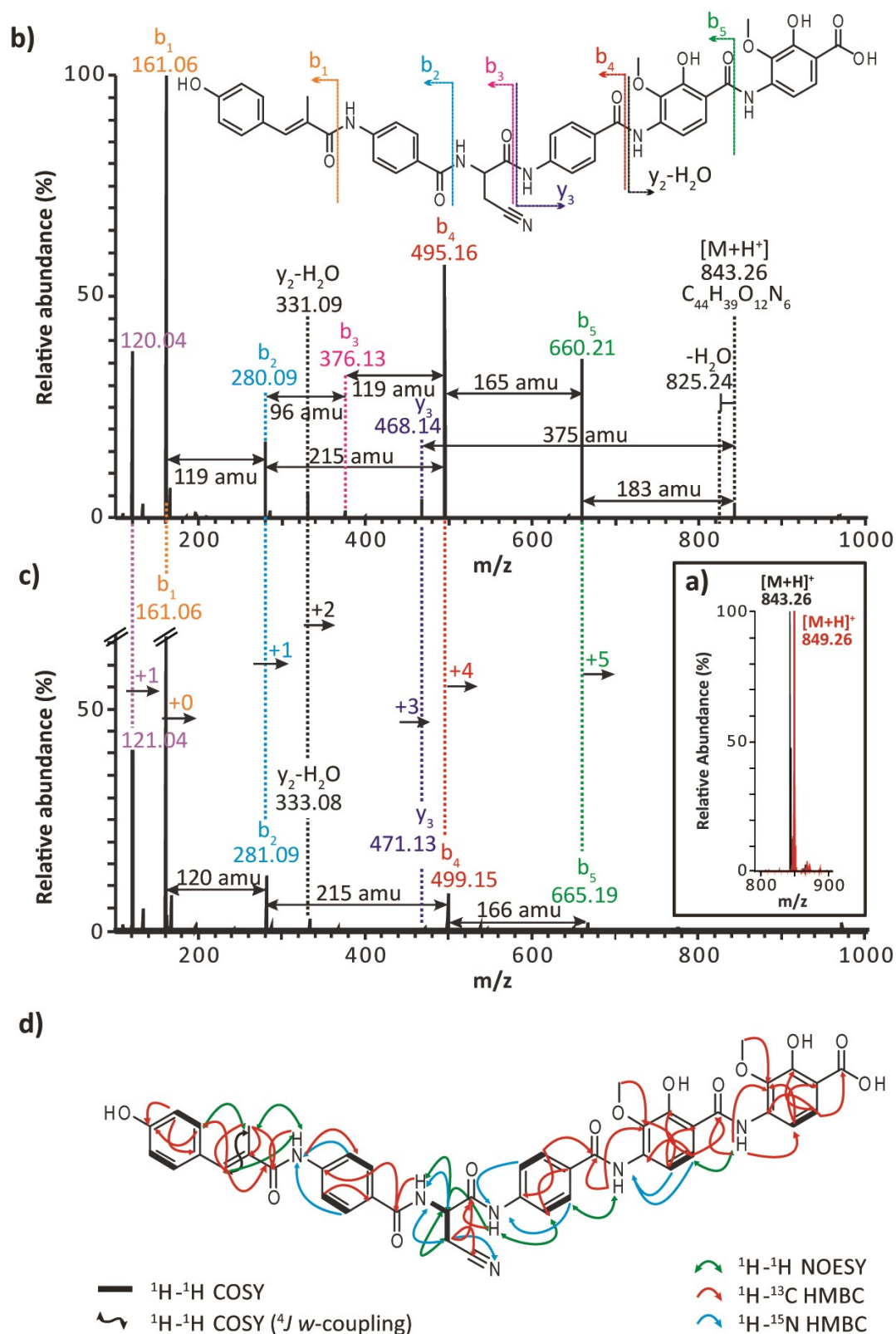
Figure 3.2. Structural characterization of albicidin by means of MS and NMR spectroscopy. a) HR-ESI-(+)-Orbitrap-MS spectra of unlabeled albicidin (black) and ^{15}N-labeled albicidin (red) illustrating the mass shift of 6 amu. The HR-ESI-(+)-Orbitrap-MS[2] fragment spectrum is shown for **b)** unlabeled albicidin and **c)** ^{15}N-labeled albicidin. The observed b-ion and y-ion fragments together with their mass shifts are color-coded and indicated in the structure. **d)** NMR spectroscopic through-bond (COSY, HMBC) and through-space correlations (NOESY) are illustrated in the albicidin structure (see legend).

HPLC-HR-Orbitrap electrospray ionization (ESI)-MS of purified albicidin revealed for the first time the tentative molecular formula $C_{44}H_{39}O_{12}N_6$ $[M+H]^+$ (m/z_{exp} 843.26305 $[M+H]^+$, m/z_{calc} 843.26205 $[M+H]^+$, Δm 1.189 ppm). In order to facilitate heteronuclear NMR experiments, we also prepared ^{15}N-labeled albicidin, which renders a molecular mass of m/z_{exp} 849.24781 $[M+H]^+$ (m/z_{calc} 849.24426 ($C_{44}H_{39}O_{12}{}^{15}N_6$) $[M+H]^+$, Δm 4.185 ppm). The observed difference of 6 amu thus confirms the presence of six nitrogen atoms in the molecule (Fig. 3.2a). The sequential walk by Tandem-MS experiments reveals a peptidic backbone with b-ion and y-ion fragment series (Fig. 3.2b). Consequently, the accurate number of nitrogen atoms and molecular formulae are derived for each fragment by comparison of unlabeled and ^{15}N-labeled albicidin (Fig. 3.2b and 3.2c).

Various two-dimensional NMR experiments were performed to ultimately determine the constitution of albicidin (Supplementary Table 2, appendix A). NMR chemical shifts and scalar coupling constants are listed in Supplementary Table 3 in appendix A. Figure 2d summarizes the observed through-bond and through-space NMR connectivities. In brief, the ^{1}H NMR spectrum of albicidin contains various aromatic signals (δ_H 6-8 ppm) with coupling constants characteristic of para-substituted aromatic spin systems (Supplementary Table 3). The [1]H-^{15}N HMQC spectrum with ^{15}N chemical shifts in the range of 110-130 ppm confirms the presence of five amide moieties (Supplementary Fig. 2 in appendix A). The sixth nitrogen atom is visible in the ^{1}H-^{15}N HMBC spectrum at δ_N 250 ppm (Supplementary Fig. 2 in appendix A) and is part of the nitrile group of the α-amino acid β-cyanoalanine (Cya-3). As shown in Supplementary Figure 3 in appendix A, chiral gas chromatography-MS analysis indicates an *L*-configuration for Cya-3. The remaining four amide groups belong to the aromatic spin systems pABA-2/4 and pMBA-5/6, respectively (Fig. 3.2d). Finally, the sequential arrangement of the identified building blocks including the polyketide-derived MCA-1 at the N-terminus of albicidin is unambiguously established by analysis of ^{1}H-^{13}C HMBC and [1]H-^{1}H NOESY data (Supplementary Fig. 5).

Based on the structure of albicidin and the arrangement of genes in the *alb* gene cluster (Supplementary Table 4 and Supplementary Fig. 6 in appendix A), we propose the following pathway for the formation of albicidin (Fig. 3.3). The N-terminal building block MCA-1 of albicidin is established by the concerted action of the PKS module of Alb01 (Fig. 3.3b), whereas NRPS-1 of Alb01 couples the thioester-bound MCA-1 to the next residue, namely pABA-2. The pABA molecule can be delivered either by the cluster intrinsic gene set of *alb17* and *alb18*, or by their homologs in the primary metabolism (*pabAB* and *pabC,* respectively) [201].
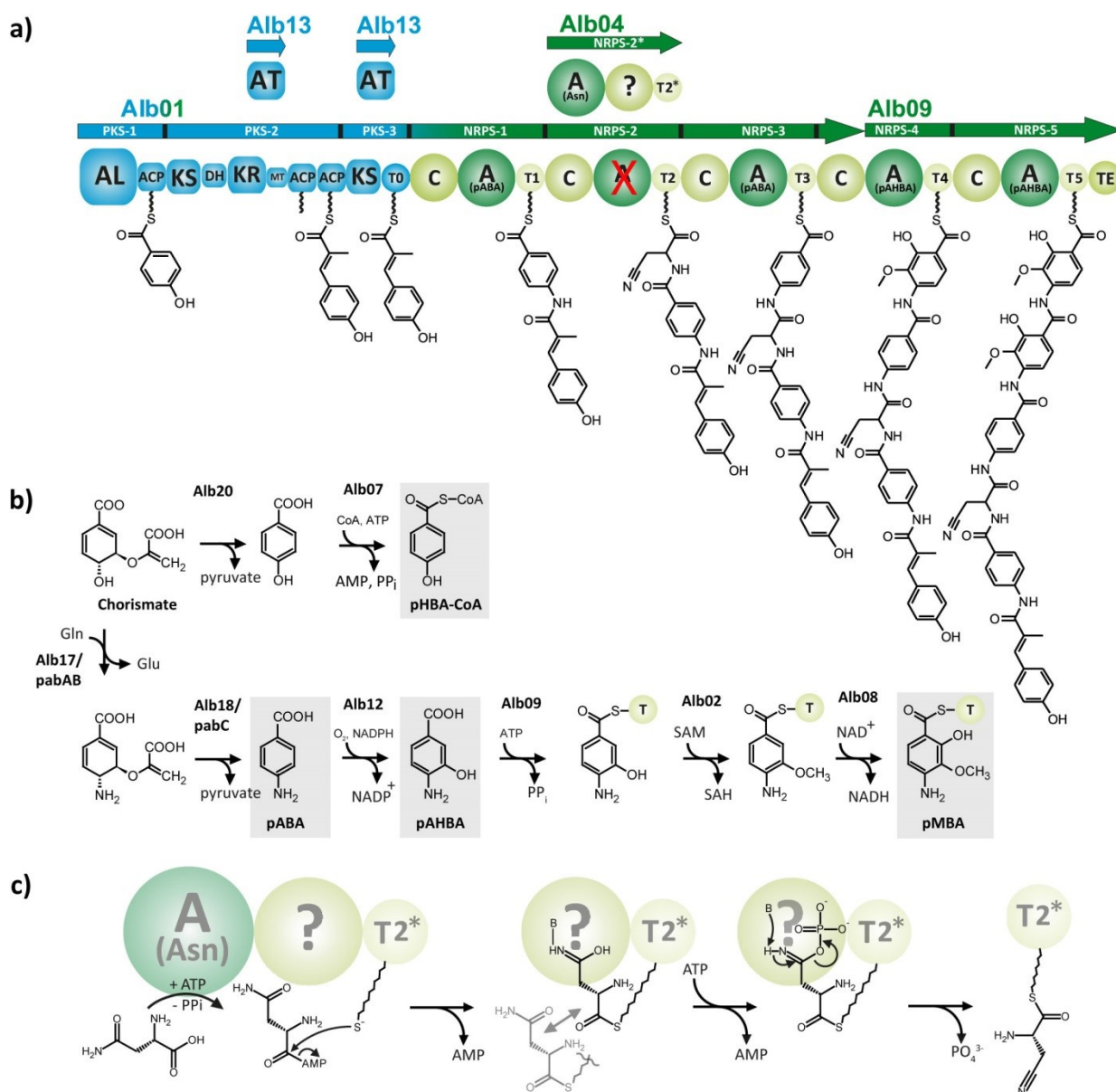
Figure 3.3. Model of albicidin biosynthesis. **a)** Proposed biosynthetic assembly line for albicidin. Substrates of the NRPS are indicated at the A domains. PKS and NRPS modules are color-coded in blue and green, respectively. The PKS-NRPS hybrid is build up by an acyl-CoA ligase (AL) domain, acyl carrier protein (ACP) domains, *trans*-acting acyltransferase (AT), a ketoreductase (KR), a dehydrogenase (DH), methyltransferase (MT), peptide carrier domains (T), condensation domains (C), adenylation domains (A) as well as a thioesterase domain (TE). **b)** Suggested pathways for the biosynthesis of the MCA-1 precursor pHBA-CoA as well as the δ-amino acids pABA and pMBA. **c)** Suggested mechanism for the transformation of activated Asn into Cya by Alb04. Conventional activation of Asn by adenylation is mediated by the A domain of Alb04. The substrate is subsequently stored as a thioester at the 4'-phosphopantetheine arm (T2* domain). A second activation occurs by phosphorylation of the Asn side-chain, which finally leads to the formation of Cya.

The multiple sequence alignment of all adenylation (A) domains of the albicidin gene cluster (Supplementary Fig. 7 in appendix A) demonstrates that NRPS-1 and NRPS-3 as well as NRPS-4 and NRPS-5 share virtually 100% sequence identity in functional core motifs, that have been previously

described to confer substrate specificity [191]. This observation is consistent with similar chemical structures of the building blocks pABA and pMBA. Intriguingly, the activation of aromatic δ-type amino acid rather than α-amino acid substrates appears to arise from an altered localization of a key aspartic acid residue interacting with the corresponding amino moiety of the substrates (Supplementary Fig. 8 in appendix A) [202]. Given these bioinformatic insights, the substrate specificities of the involved NRPS modules were determined experimentally using a radioactivity-based ATP/PP$_i$ exchange assay. As illustrated in Supplementary Figure 9 in appendix A, the A domains of NRPS-1 (pABA-2) as well as NRPS-3 (pABA-4) indeed show the highest activity for the substrate pABA.

The biosynthesis of albicidin proceeds with the incorporation of L-Cya-3. The collinearity rule [191] assigns the A domain of NRPS-2 (Alb01) to be responsible for this biosynthetic step. However, neither Cya nor Asn were activated by NRPS-2 in the ATP/PP$_i$ exchange assay (Supplementary Fig. 9 in appendix A). Bioinformatic analysis supports that NRPS-2 is most likely not functional in terms of a conventional A domain, owing to the low sequence conservation of essential core motifs (Supplementary Fig. 7 in appendix A). Instead, we suggest that incorporation of L-Cya-3 is achieved by the single standing and probably *trans*-acting NRPS-2* (Alb04), which is essential for albicidin biosynthesis [194]. Such a *trans*-complementing mechanism has been already described, e.g. for bleomycin biosynthesis [203]. In agreement with *in-silico* predictions (Supplementary Fig. 8 in appendix A), NRPS-2* reveals specificity for L-Cya and even more pronounced for L-Asn in the *in-vitro* assay (Supplementary Fig. 9 in appendix A). It should be noted that the ATP/PP$_i$ exchange assay generally yields end-point data only and may thus imply activation of less preferred substrates. The A domain of module NRPS-2* contains an unusual 342-amino acid insertion harboring an ATP-binding motif (SGGKD), which is strongly conserved among members of the adenosine nucleotide alpha hydrolase (alpha ANH-like III) superfamily (Supplementary Fig. 7 in appendix A) [204]. Hence, we postulate the activation and *in-situ* processing of L-Asn to L-Cya by NRPS-2* in a two-step scenario: i) adenylation of the α-carboxylic acid moiety and storage as a thioester as well as ii) phosphorylation of the side-chain amide oxygen (Fig. 3c). Subsequent dephosphorylation would cause a formal elimination of water and thus the formation of Cya, which is then attached to the growing peptide chain.

The following biosynthetic step includes the activation and coupling of pABA-4 by NRPS-3 (Alb01) consistent with the *in-vitro* substrate specificity of its A domain (Supplementary Fig. 9 in appendix A). Finally, residues pMBA-5 and pMBA-6 are incorporated by modules NRPS-4 and NRPS-5 (Alb09), respectively, with the thioesterase domain of NRPS-5 mediating the release of synthesized albicidin.

While pABA is utilized as a direct substrate by NRPS-1 and NRPS-3 (Alb01), our experimental data implies that NRPS-4 and NRPS-5 (Alb09) preferentially activate the substrates 2- and 3-hydroxy-pABA rather than pMBA (Supplementary Fig. 9 in appendix A). We therefore propose prior formation of 3-hydroxy-pABA (pAHBA) from pABA through hydroxylation by the benzoyl-CoA oxygenase-like enzyme Alb12 (Fig. 3.3b). Subsequently, the activated pAHBA, being attached to the T domains of NRPS-4 or NRPS-5, is most likely processed to pMBA by β-hydroxylation and methylation through the putative β-hydroxylase Alb08 and the essential methyltransferase Alb02 [205], respectively (Fig. 3.3b). Interestingly, the multiple sequence alignment of several T domains reveals a non-conserved insertion of 13 amino acids for both T domains of NRPS-4 as well as NRPS-5 (Appendix A, upplementary Fig. 7). Based on published T domain structures [206, 207], the sequence insertion apparently extends the loop region between helix I and helix II, and might reflect additional protein-protein interaction modes in the course of *in-situ* pAHBA processing.

In search of new potent antibiotics, albicidin represents a very promising lead structure in anti-infective research. This work uncovers the unique chemical structure and the fascinating biosynthesis apparatus of albicidin, and thus paves the way for structure-activity relationship studies.

Experimental

Fermentation of heterologous albicidin producer

Liquid cultures of heterologous albicidin producing strain (strain Xves-alb of *Xanthomonas axonopodis* pv. *vesicatoria*) [200] were prepared in plastic Falcon tubes of 500 mL to avoid adsorption of albicidin to the glass surface. Cultures were grown under agitation for 5 days at 28 °C in 72 tubes each containing 200 mL (14.4 L) of an optimized medium (XVM3B medium, Appendix A supplementary Table 5). Several of these batches had to be prepared for purification of unlabeled albicidin corresponding to approx. 1500 tubes of 200 mL medium each (approx. 300 L medium in total). For preparation of ^{15}N-labeled albicidin, the same medium was used for feeding of isotopically labeled (^{15}NH$_4$)$_2$SO$_4$ (^{15}N-XVM3B medium, Appendix A, supplementary Table 5).

Isolation and purification of albicidin

Isolation of albicidin was adapted from a previous procedure [55, 198]. Briefly, XAD-7 Amberlite was added to the fermentation broth of strain Xves-alb, in order to adsorb albicidin from the supernatant. Albicidin was eluted from XAD-7 during a MeOH/H_2O-step-gradient (20 – 100% MeOH, Δc = 20%) at a final concentration of 100% MeOH. Bioactivity guided fractions were concentrated in a Genevac Speedvac (Ipswich, Great Britain) and re-dissolved in MeOH. After centrifugation, the supernatant was purified by preparative HPLC on an Agilent 1100 system (Agilent, Waldbronn, Germany) at a detection wavelength of 310 nm on a C18 reversed phase column (GromSil 120 ODS 5 ST, 10 μm; 250 x 20 mm Grace, Rottenburg-Hailfingen, Germany) using a linear MeOH-gradient starting from 35% MeOH + 0.1% HCOOH to 80% MeOH + 0.1% HCOOH for 40 min with a flow rate of 15 mL/min. Albicidin eluted at a retention time R_t of 33 min. The fractions were tested for antibacterial activity, freeze-dried and re-dissolved in aqueous tetrahydrofuran (44%) with acetic acid (1%) and purified by semi-preparative Agilent 1100 HPLC system (Agilent, Waldbronn, Germany) at a detection wavelength of 310 nm, using isocratic conditions on a polymeric reversed phase (PRP-1, 5 μm; 305 x 7 mm, Hamilton Bonaduz, Switzerland). Albicidin eluted at a retention time R_t of 68 min and was obtained as a white solid (3 mg unlabeled albicidin and <1 mg ^{15}N-labeled albicidin).

Antibacterial assay

Antibacterial activity of albicidin was assessed according to the microbiological bioassay described by Zhang et al. [208]. Briefly, Petri dishes containing 20 mL of LB agar were overlayed with a mixture of 2 ml of exponential phase bacterial culture (2×10^7 cells/mL in LB medium) with 2 mL of molten 1.5% noble agar. Twenty μL of various concentrations of albicidin were deposited in 5-mm diameter wells cut in the overlaid plates. After overnight incubation at 37°C, the width of the crown of growth inhibition surrounding each well was recorded. Bacterial strains used in the assay are described in Supplementary Table 1 in the Appendix A section.

Total hydrolysis and chiral GC-MS

Total hydrolyses of albicidin and L-Cya was performed at 110°C in aqueous 6 M hydrochloric acid solution (200 μL) for 6 h in glass vials. After hydrolysis, the hydrochloric acid was removed in a gentle stream of nitrogen.

For chiral GC/MS analysis of hydrolysates as well as of D- and L-Asp, the amino acids were transformed into N-trifluoroacetyl/ethyl ester derivatives. In short, a total volume of 150 μL of 2 M ethanolic HCl-solution was added. Then the samples were heated for 30 min at 110°C, and reagents

were removed at 110°C in a gentle stream of nitrogen. Acetylation was performed by adding 100 μL dichloromethane and 50 μL trifluoroacetic anhydride. The mixtures were heated again for 10 min at 110°C. Excess of reagent was removed at ambient temperature in a stream of nitrogen. Resulting residues were dissolved in 50 μL anhydrous toluene, and 1 μL was subjected to GC/MS (5975C, Agilent Technologies, Waldbron, Germany) using an Agilent J&W CP-Chirasil L-Val GC column, Antipode D, 0.25 mm, 25 m, 0.08 μm (Agilent Technologies, Waldbron, Germany) with an split 1:50.

The temperature gradient started with an initial hold for 5 min at 70°C followed by an increase of 2°C/min up to 100°C and then 3.5°C/min up to 190°C with a 10 min hold. The column was then heated to 200°C within 1 min with a 1 min hold at 200°C. Selected ion monitoring (SIM) was set to 140 and 212 m/z in positive electron impact (EI) mode (5 min solvent delay, MS source: 300°C, MS quad: 150°C, EMV mode, gain factor 1, EM voltage 1388 V) with a flow of 1.2 mL/min with helium as carrier gas.

HPLC-HR-MS

Full-scan measurements were routinely performed on an Exactive ESI-Orbitrap-MS (Thermo Fisher Scientific GmbH, Bremen, Germany) coupled to an analytical Agilent HPLC 1200 system (Agilent, Waldbronn, Germany) equipped with a Hypersil-Gold (5 μm, 50 x 2.1 mm, Thermo Fisher Scientific GmbH, Bremen, Germany) using MeOH and H_2O + 0.1% HCOOH as mobile phase. Albicidin was analyzed during a linear gradient from 5% to 100% MeOH for 6 min followed by a washout phase of 100% MeOH for 4 min and a re-equilibration phase at 5% MeOH for 2 min. The flow rate was set to 0.25 mL/min. MS conditions were R = 100,000 (@ 200 m/z) with a maximal C trap fill time of 500 ms. The mass range was set to 150-1500 m/z.

HPLC-HR-MS/MS

MS/MS measurements were performed on an LTQ-Orbitrap XL hybrid instrument (Thermo Fisher Scientific GmbH, Bremen, Germany) coupled to an analytical HPLC 1200 system (Agilent, Waldbronn, Germany) with a Hypersil-Gold (5 μm, 50 x 2.1 mm, Thermo Fisher Scientific GmbH, Bremen, Germany) using the same gradient as described above for full scan measurements. R was set to 30,000 (@ 400 m/z) with a maximal C trap fill time of 200 ms. HCD collision energy was set to 18% with 30 ms activation time and an isolation window of 2 m/z for the precursor selection (843.3 and 849.3 m/z). The mass range was set to 100-1000 m/z.

HPLC-QqQ-MRM

Multi reaction monitoring (MRM) of XAD extracts was carried out on an Agilent 6460 triple quadrupole instrument (Agilent Technologies, Waldbron, Germany). A C18 Core-shell column (Supleco Ascentis Express, 100 x 2.1 mm, 5 µm, Sigma Aldrich, San Louis, USA) with a linear gradient from 20-95 % ACN over 5 min, followed by an 2 min washout and 2 min re-equilibration, with an flow rate of 0.3 mL/min was used. ESI negative with -3500 V capillary voltage and -1200 V nozzle voltage was used. For the SRM transient a unit precursor selection window at 841.2 m/z and a MS2 m/z value of 362.1 for a typical albicidin product ion were used. Normalized collision energy was set to 38. Dwell time was set to 200 ms, Fragmentator voltage was set to 135 V and the cell accelerating voltage was set to 7 V.

NMR spectroscopy

NMR spectra of unlabeled and ^{15}N-labeled albicidin were acquired on an Avance III 700 MHz NMR spectrometer (Bruker, Karlsruhe, Germany) equipped with a TXI 1.7 mm room temperature probe head. Samples of 50 µL in d_8-tetrahydrofuran (THF) were filled into 1.7 mm NMR tubes. All experiments were performed at 298 K. Chemical shifts were referenced relative to tetramethylsilane (TMS) (^{1}H, ^{13}C) and liquid NH_3 (^{15}N), respectively. Supplementary Table 2 and Supplementary Table 3 summarize relevant NMR acquisition parameters and the NMR data for albicidin, respectively. All NMR experiments used standard parameter sets employed by the manufacturer (Bruker, Karlsruhe, Germany).

Bioinformatic analysis

DNA sequence information was obtained from Royer *et al.* [194]. Nucleotide sequences were translated using the software Clonemanager (Sci-Ed Software, Cary, NC, USA). Putative protein functions were determined using pBLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) [209]. Multiple alignments of NRPS were performed with clustaW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) [210]. Final adjustment to the sequence of GrsA was performed manually. NRPS substrate prediction was drawn out using the PKS-NRPS predictor of the University of Maryland (http://nrps.igs.umaryland.edu/nrps/) [211]. Structural prediction of A-domain binding pockets was carried out using I-TASSER [212].

Construction of expression plasmids

The *alb04* gene (A/T didomain of NRPS-2*) and selected parts of genes *alb01* (A/T didomain of NRPS-1, NRPS-2 and NRPS-3) and *alb09* (A domains of NRPS-4 and NRPS-5), respectively, were amplified from the cosmid pALB571[200] by polymerase chain reaction (PCR) with Herculase polymerase

(*Fermentas, Thermo Scientific*) using the primers described in Supplementary Table 6. PCR-products were digested with restriction enzymes *Not*I and *Nco*I, cloned into the vector pETtrx_1c [213] or pQE80 and subsequently transformed into *E. coli* BL21 gold by either electroporation or heat shock. Transformed clones were selected on kanamycin or ampicillin LB-agar plates (50 μg/mL) and verified by plasmid purification following restriction analysis and DNA sequencing of the inserted genes.

Gene expression and protein purification

The *E. coli* expression strains containing the vectors pETtrx_1c-NRPS-1, -NRPS-2*, -NRPS-3,-NRPS-4, -NRPS-5 and pQE80_NRPS-2 were incubated in 500 mL TB medium at 37 °C, 200 rpm for 2 h following 10 h at 18 °C, containing 0.2% lactose to induce the *lac* controlled gene expression. Subsequently, cells were harvested by centrifugation at 7,000 g for 10 min. Cell pellets were resuspended in wash buffer (50 mM Tris-HCl, pH 7.8, 300 mM NaCl) containing 5% (v/v) glycerol. Cell lysis was achieved by the addition of lysozyme and sonication (3 x 2 min; 66% intensity) and the resulting cell debris was separated by centrifugation at 50,000 g for 20 min. The supernatant was purified by Ni-affinity chromatography using an Äkta purifier (GE Healthcare, Little Chalfont, UK). Elution was performed on 1 mL or 5 mL His-Trap columns (Quiagen, Venlo, Netherlands or GE Healthcare, Little Chalfont, UK). After loading the sample onto the column at 1 or 3 mL/min flow rate, the proteins were washed (10 column volumes (CV) 30 mM imidazole) and eluted with an imidazole step gradient (10 CV 200 mM imidazole and 10 CV 500 mM imidazole). Elution fractions were combined and submitted to SDS-PAGE and Anti-His6 Westernblot (Supplementary Fig. S10 in Appendix A). The calculated protein masses for the fusion proteins are: 81 kDa for NRPS-1, 66 kDa for NRPS-2, 119 kDa for NRPS-2*, 81 kDa for NRPS-3, 74 kDa for NRPS-4 and 73 kDa for NRPS-5, and fit to bands observed during SDS-PAGE (Supplementary Fig. S10). Furthermore, the identity of the fusion proteins was confirmed by western blot against the poly-His tag (Supplementary Fig. S10 in Appendix A). Minor impurities of the *E. coli* proteome were observed for NRPS-4 and NRPS-5. Subsequent to Ni-affinity purification, the fractions were pooled and dialyzed with an ultrafiltration device (10 kDa molecular weight cut-off, Merck-Millipore, Billerica, MA, USA) using wash buffer. Protein concentrations were determined via intrinsic absorption at 280 nm by using a NanoPhotometer (Implen, München, Germany). The protein yields of NRPS-1, NRPS-2, NRPS-2* NRPS-3, NRPS-4 and NRPS-5 were 15.0, 3.4, 19.0, 1.0, 3.3 and 9.4 mg/L expression culture, respectively. After addition of 20% glycerol, aliquots were shock-frozen in liquid nitrogen and stored at -80°C for further experiments.

ATP/PPi exchange assay

ATP/PP$_i$ exchange assays were performed using a modified protocol of Calendar and Berg [214]. ATP mix (100 µL containing 0.1 mM PP$_i$, 10 mM ATP, 10 mM MgCl$_2$ and ^{32}P-labeled PP$_i$ set to 200000 counts/100 µL) were mixed with 20 µL of 100 mM amino acid substrates. After the addition of 100 µL ~ 50 µM enzyme solution, samples were incubated for 15 min at 30°C. The reaction was stopped by adding 1 mL stop solution (1.4% activated charcoal, 3.8% perchloric acid and 12.2 g/L Na$_4$PP$_i$). ATP bound to activated charcoal was collected by centrifugation, washed three times with 1 mL H$_2$O and subsequently measured using a scintillation counter (Perkin Elmer, Waltham, MA, USA) following the manufacturers instructions.

4　The O-carbamoyltransferase Alb15 is responsible for the maturation of Albicidin

Introduction

Sugarcane leaf scald, a lethal disease of sugarcane, is caused by the xylem-invading bacterium *Xanthomonas albilineans[151]* which belongs to the class of γ-*Proteobacteria*. Transmission of the pathogen usually occurs through human practices with the reproduction of the sugarcane plants during the preparation of cuttings of already infected plants or by using contaminated cutting tools [215]. However, potential aerial transmission has been reported as well [216, 217]. The invasion occurs primarily 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 [218]. 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 [56, 57, 219] and is one of the pathogenic factors of sugarcane leaf scald [55]. 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 since its discovery. However, the high antibacterial activity against a wide range of gram-positive and gram-negative strains [220] combined with a new inhibition mechanism of gyrase makes albicidin a potential candidate for the development of new anti-infectives [60]. Unlike fluoroquinolones and coumarins, which target the ATP-binding pocket [62], albicidin stabilizes the covalent DNA-gyrase complex [60].

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 [193] which finally enabled us to solve the structure of albicidin by a combination of extensive MS/MS and NMR experiments [57]. Having the structure of albicidin in hands, we established its total synthesis, aiming at structure activity relation (SAR) studies to assess

albicidin derivatives for drug development purposes [61]. 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) [57]. Next to pABA as a characteristic structural feature, albicidin (Figure 1.1A) 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 [57]. 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 [57]. 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.

In the biosynthesis gene cluster of albicidin, a putative *O*-carbamoyl transferase gene is present [56, 57] which finds it precedence in other natural products, e.g. cephalomycin [221], novobiocin [222] and tobramycin [223]. Through a combination of tandem mass spectrometry, *in vitro*- and gene knockout experiments, we could identify a putative carbamoylated albicidin derivative, of which we ultimately confirmed the structure by the total synthesis of carbamoyl-albicidin. Its synthetic access enabled us to further study the impact on the bioactivity and pharmacological properties which forms part of our ongoing synthetic optimization of the albicidin structure as a new antibacterial drug.

Results and Discussion

A main obstacle for the structure elucidation of albicidin was the low production yield in the *X. albilineans* wild type. The heterologous host *Xanthomonas axonopodis* pv. *vesicatoria* delivered higher amounts, which were finally sufficient to unravel its structure by mass spectrometry and NMR spectroscopy [57]. During the purification of albicidin by preparative HPLC and MS, we identified another minor abundant compound with a delta mass of +43.0 Da compared to albicidin. The exact mass of [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) that differs from the molecular formula of albicidin by a fragment of CONH which could be interpreted as a carbamoyl group. This assumption was in line with the presence of a putative carbamoyltransferase gene (*alb15*) in the albicidin gene cluster. Alb15 shows high similarity (~30%) to TobZ, an ATP-dependent *O*-carbamoyltransferase involved in the carbamoylation of tobramycin of which structural data is available [223]. 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 supplemental figure 11). This indicates that Alb15 is most likely an ATP depended carbamoyl-transferase.

In order to identify the carbamoylation site and because the little amounts of carbamoyl-albicidin which were insufficient for NMR experiments, we performed mass spectrometric product ion scans of both albicidin and carbamoyl-albicidin (Figure 4.1), 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_2O) and 468.0 Da (y_3). For the b-ion series of 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 an 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, according to the recently published total synthesis of albicidin by our group [61]. 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.* [224].

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 Δm = -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 (supplemental figure 12).

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*-Δ*alb15* 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 4.2).
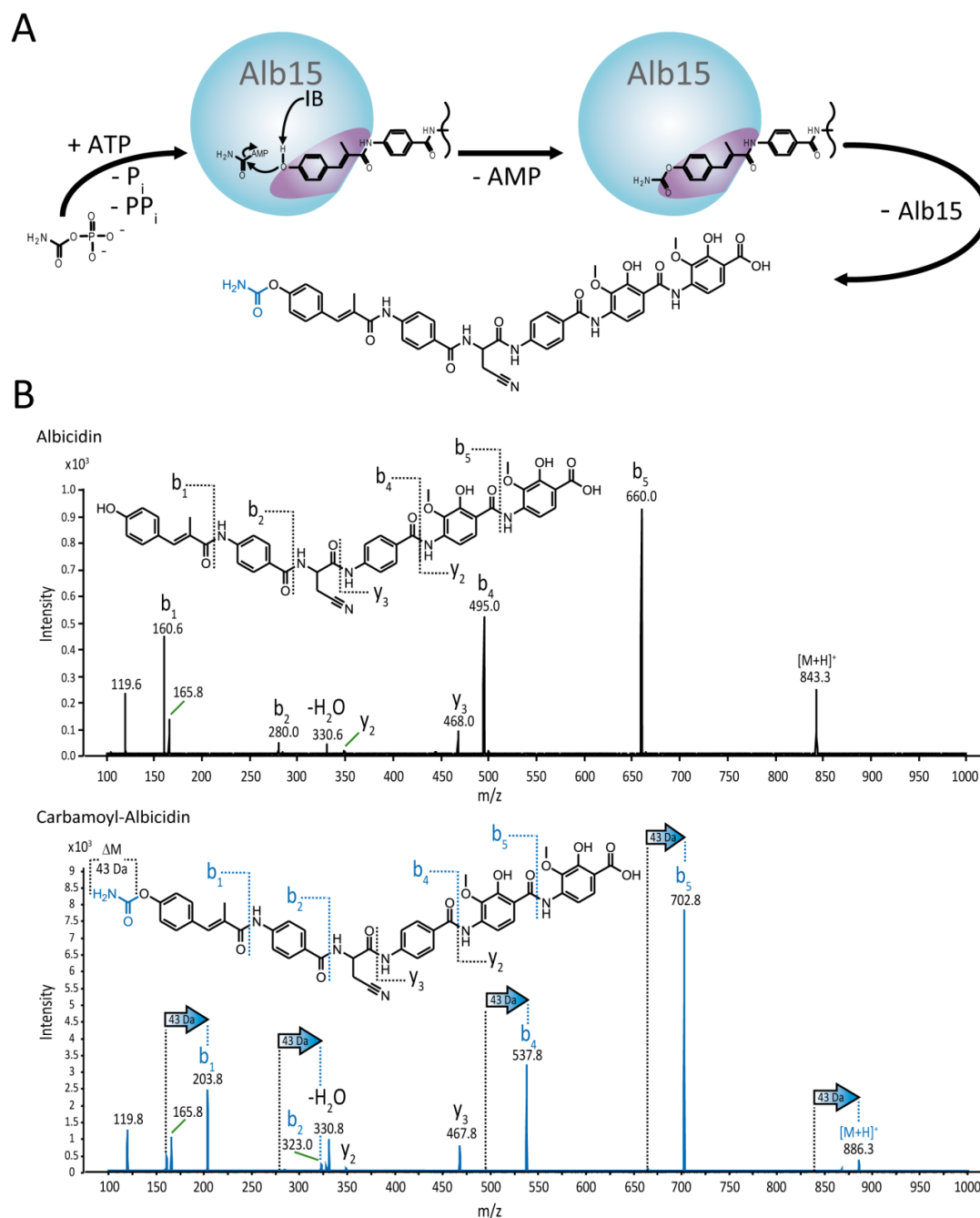
Figure 4.1: A: Reaction scheme of post-NRPS carbamoylation of albicidin. Alb15 catalyzes first the reaction from carbamoyl-phosphate and ATP to carbamoyl-AMP followed by the carbamoylation of albicidin. **B: MS/MS structural elucidation und comparison of albicidin and carbamoyl-albicidin.** The observed b-ion and y-ion fragments and the mass shift through carbamoylation are indicated in the spectra and structure.

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 were ~5 mg/L$_{culture}$ (supplemental figure 13). The purity and identity of the fusion protein was verified by SDS-PAGE, in-

gel trypsin digestion and LC-MS/MS analysis of the tryptic peptides, both shown in the Appendix B, supplemental figures 13 and 14.
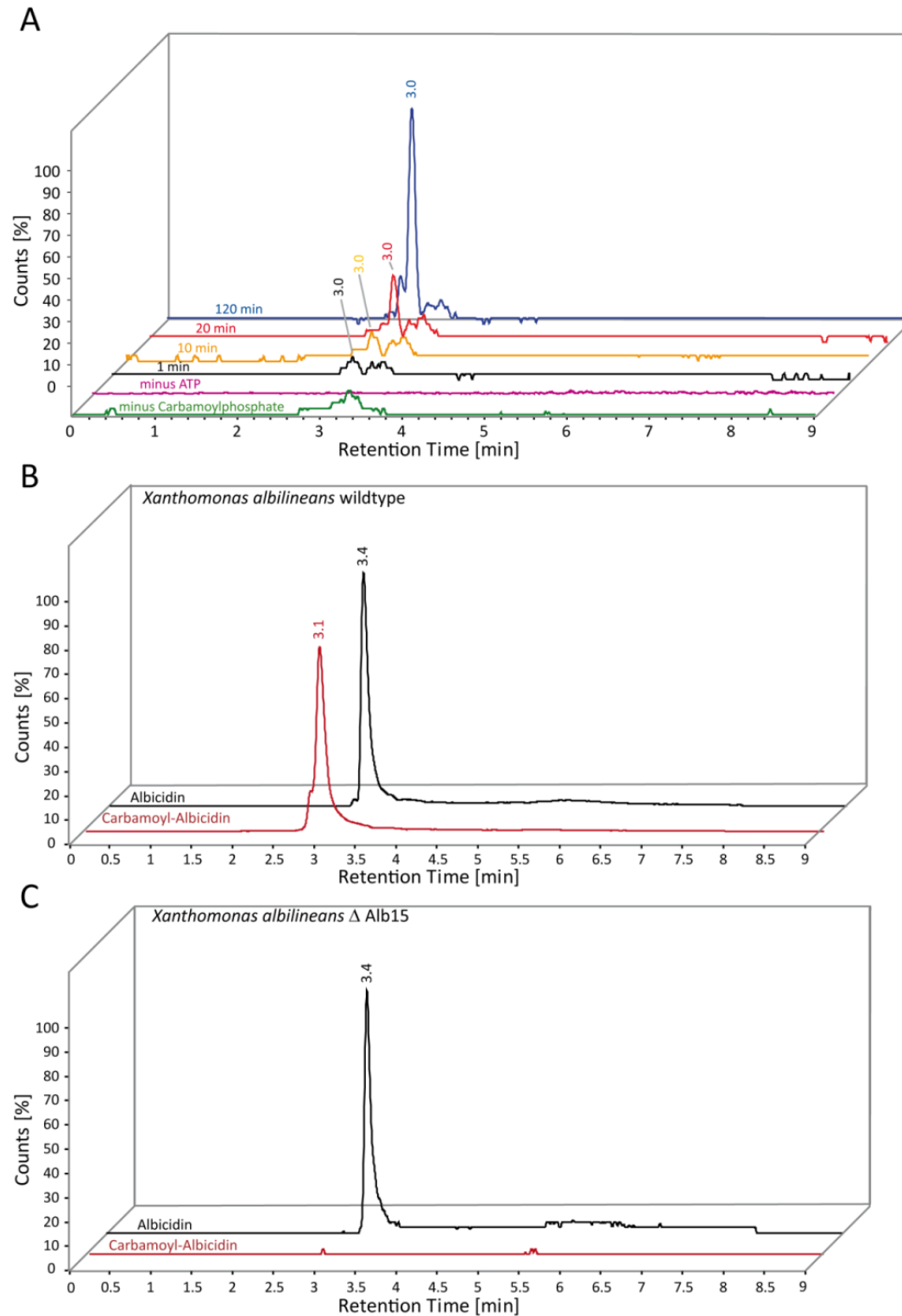
Figure 4.2. A: *In vitro* carbamoylation of albicidin. Multi-reaction monitoring (MRM) chromatograms of different time points and negative controls (- Carbamoyl-phosphate and - ATP) of the *in vitro* conversion assay are shown. **B: MRM chromatogram of XAD extract of *X. albilineans* wildtype. C: MRM chromatogram of XAD extract of *X. albilineans* Δ*alb15* gene inactivation mutant.** Presence of carbamoyl-albicidin (R_t = 3.1 min) can only be seen in the wildtype whereas albicidin (R_t = 3.4 min) is present in both extracts.

The O-carbamoyltransferase Alb15 is responsible for the maturation of Albicidin

We reconstituted the carbamoylation by incubating purified Alb15 with ATP, albicidin and carbamoyl-phosphate. The detection of carbamoylated albicidin was performed by ESI-MRM. The MRM chromatograms in figures 4.2A 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 3.0 min from approximately 20 min after the start of incubation. 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 [225, 226].

To investigate the biological effects of the carbamoylation of albicidin and the influence on its physicochemical properties which may be of significance for further medicinal chemistry development, we tested the antibacterial properties in a simple agar halo assay against *E. coli* DH5alpha (Appendix B, supplemental figure 15). Here, 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 determine the antibacterial activity more in detail, we assessed the minimal inhibitory concentration (MIC) against two gram-positive (*Bacillus subtilis* and *Mycobacterium phlei*) and two gram-negative strains (*E.coli K12* and *Salmonella typhimurium*) (Appendix B, supplemental table 8). We observed MICs around 0.2 ng/µl for albicidin and carbamoyl-albicidin for both gram-positive strains. The reference antibiotic, apramycin, instead showed MIC values ~12.5 ng/µL against *Mycobacterium phlei* and 25 ng/µL against *Bacillus subtilis.* The difference of the MIC against *Salmonella typhimurium* of carbamoyl-albicidin compared to albicidin and apramycin was 3.1 (carbamoyl-albicidin) *vs* 6.3 ng/µL (both albicidin + apramycin) which corresponds to an approximate 50% better inhibition. For *E. coli K12* we observed MIC values ~0.1 ng/µL for carbamoyl-albcidin and an approximately 50% better value for albicidin. In comparison to the apramycin this MICs were more than one order of magnitude lower.

To investigate further the influence of the carbamoyl-group on the inhibitory potential on the gyrase, we performed *in vitro* gyrase DNA supercoiling assays. Therefore, we incubated relaxed DNA with bacterial gyrase and different concentrations of carbamoyl-albicidin. In the absence of the inhibitor gyrase generates supercoiled DNA which runs faster during agarose gel electrophoresis. As shown in a representative gel in figure 4.3A and the densiometric plot in figures 4.3B, supercoiling occurs in the reaction without gyrase (-gyrase) and in a dose-dependent inhibition (0- 100 nM) which reaches from almost no supercoiling activity (100 nM) to almost complete conversion (0 nM). We observed an IC_{50} of carbamoyl-albicidin between 5 and 10 nM, which is around four times stronger compared to albicidin [61] (run as control in the utmost band).
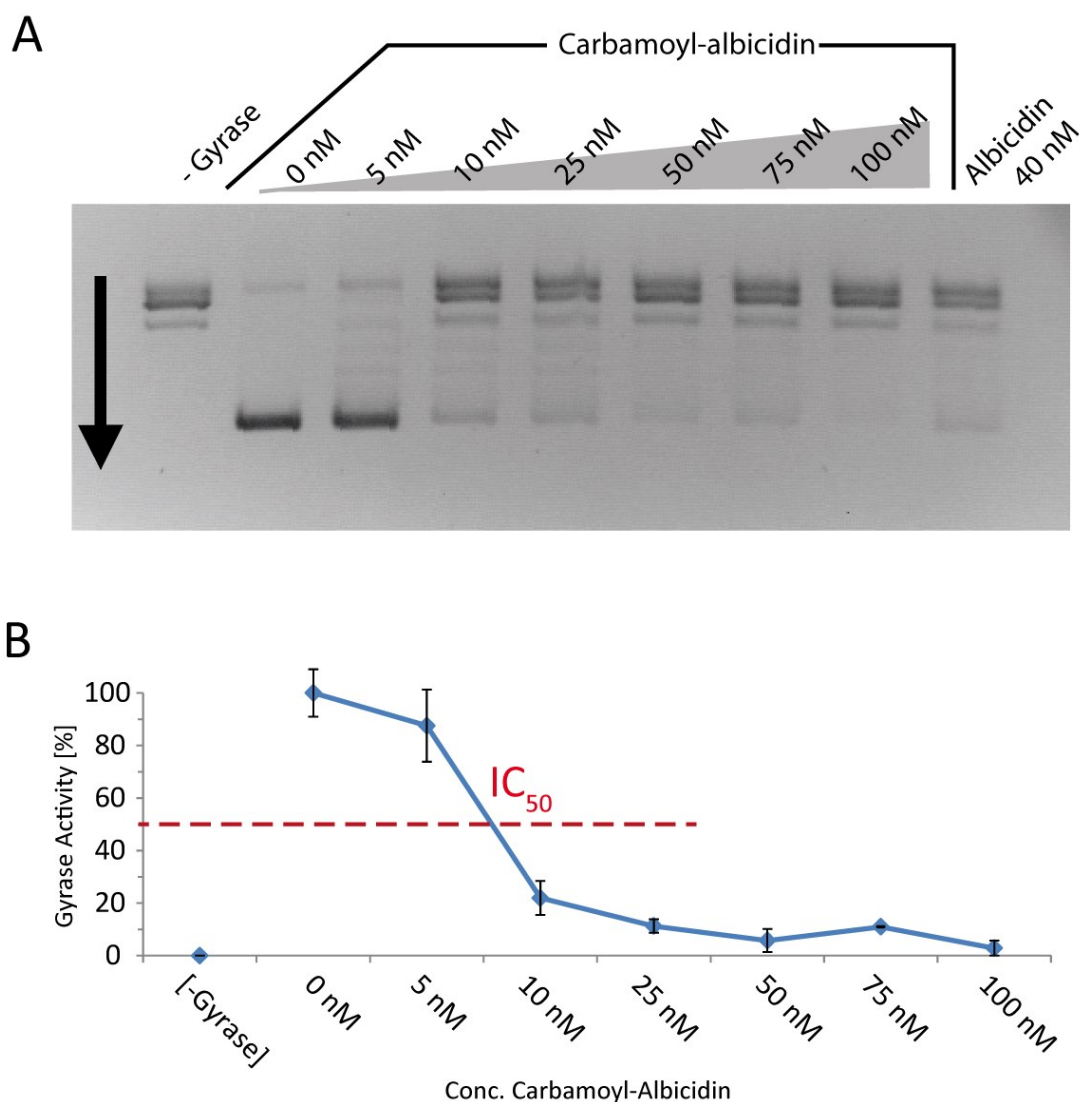
Figure 4.3: A: *In vitro* determination of the half maximal inhibitory concentration (IC$_{50}$) of carbamoyl-albicidin and an albicidin control against bacterial gyrase. B: Densitomentric analysis of biological triplicates of the gyrase assay shown in A. The intensity of the lower supercoiled lane is blotted 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 was set as 100% e.g. no inhibition.

In summary, we can state that Alb15 carbamoylates albicidin at the hydroxy-group at the N-terminus, most likely as a post-NRPS reaction. Through the carbamoylation, albicidin shows higher antibacterial potency compared to non-carbamoylated albicidin, which indicates that albicidin matures in terms of bioactivity. Besides the biology relevance, these finding are important for the optimization of albicidin as a potential drug. The carbamoylation improves solubility of albicidin in water which remains as one of the drawbacks of the drug properties of albicidin. Furthermore, the higher *in vitro* activity indicates that the N-terminal region of albicidin is important for gyrase interaction, giving important insights in the structure-activity relationship. With this finding, we will further investigate

the influence of N-terminal modifications in our ongoing synthetic development of albicidin as a 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*) were prepared in plastic tubes (Corning Inc., Corning, New York, USA). Cultures were grown under agitation for five days 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 [57]. 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, Waldbronn, Germany) at a detection wavelength of 308 nm on a C18 reversed-phase column (GromSil 120 ODS 5 ST, 10 μm; 250 × 20 mm Grace, Rottenburg-Hailfingen, Germany) 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, Waldbronn, Germany) 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, Bonaduz, Switzerland). An additional step of purification was performed with reversed-phase HPLC on an analytical Agilent 1200 HPLC system (Agilent, Waldbronn, Germany) at a detection wavelength of 308 nm using an Agilent Zorbax RX-C_{18} column (250 x 4.6 mm; 5 μm; Agilent, Waldbronn, Germany) 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.* [171]. All primers used can be found in the supplemental information (Appendix B, supplemental

50

table 7). 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* [162]) 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 [227], 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^{-1} kanamycin. Plates were incubated at 28°C for five to seven days until apparition of isolated colonies 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, Bremen, Germany) coupled to an analytical HPLC 1200 system (Agilent, Waldbronn, Germany) 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 % 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 triple-quadrupole mass spectrometer coupled to an analytical UHPLC 1290 system (Agilent, Waldbronn, Germany) using a Grace C18 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, 843 and 846 m/z 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 multi reaction monitoring (MRM) using 843 and 846 m/z as precursor masses and 660 and 703 m/z 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 supplemental information. In short, the *alb15* gene was amplified by PCR using the cosmid pALB540 [56] 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, München, Germany). The identity of the Alb15 fusion protein was verified by SDS-PAGE and in-gel trypsin digestion (supplemental information). 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 to -80°C.

In vitro **carbamoylation assay**

A volume of 25 µL of reaction mix containing 4 mM DTT, 400 µM $MnCl_2$, 4 mM $MgCl_2$, 40 mM Carbamoyl-phosphate and 20 mM ATP) were 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 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, Waltham, MS, USA), 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.*[61] In short, we modified (detailed description can be found in the supplemental information) 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.*[61] 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 using a ruler. The assay was performed in duplicates.

The minimal inhibitory concentration was determined according to the Clinical and Laboratory Standards Institute, M31-A2[228]: In short, MHB Medium was inoculated with overnight cultures of *Bacillus subtilis*, *Mycobacterium phlei*, *E. coli K12 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, Männedorf, Switzerland).

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, Frankfurt, Germany) 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, Maryland, USA). The peak area of the densitomentric analysis of the control without inhibitor was set as 100% enzyme activity.

5   Venom proteomics of Indonesian king cobra, *Ophiophagus hannah*: integrating top-down and bottom-up approaches

Introduction

Venoms represent an adaptive trait and an example of both divergent and convergent evolution [229]. The ecological advantages conferred by the possession of a venom system are evident from the extraordinarily diverse range of animals that have evolved venoms for hunting, defense efficiency or competitor dissuasion. Every ecosystem on Earth supporting life contains venomous organisms and the extant suite of venomous animals includes over 170,000 species throughout all major phyla of the evolutionary tree of the animal kingdom, from ancient cnidarians through annelids, nemertines, echinoderms, mollusks, arthropods, and chordates. Understanding the molecular mechanisms and natural selective forces that underlie venom evolution is an important yet largely unrealized goal in evolutionary biology studies of venomous systems [80, 230-236]. Due to their high degree of target specificity, refined during eons of arms race between a predator and its prey, venom toxins represent also a vast and largely untapped source of pharmacological tools and lead compounds for drug development, a fact that is attracting increasing interest in academic, industrial and medical arenas [6-10]. On the other hand, unwanted encounters with venomous creatures can have fatal consequences for the envenomed victim. In this regard, accidental snake-bite envenomings are a significant but largely neglected, public health problem worldwide [95-98]. Conservative data suggest that this "disease of poverty" [99] affects yearly between 1.2 and 5.5 million people of predominantly poor agrarian and pastoralist communities in Africa, Asia, Latin America and Oceania, causing up to ~125,000 fatalities, and leaving around 400,000 people with permanent sequelae that greatly jeopardize the quality of life of both the victims and their dependents [98-102]. A comprehensive understanding of the composition and natural history of venoms can provide fundamental insights into the evolutionary forces shaping their biodiversity, which may be of applied importance for drug-discovery research and antivenom development [237, 238].

Venom research has been continuously enhanced by technological advances. However, relatively few venoms from the over 600 species of medically relevant front-fanged snakes of the Viperidae and Elapidae families have been fully examined using cutting-edge proteomic and transcriptomic approaches [238, 239]. An important challenge in the characterization of venoms is the inability of any single method to analyze all unique proteins in the venom proteome. The development of sample preparation protocols (nano-UHPLC, multidimensional-HPLC, 2D-electrophoresis) coupled to soft-ionization (MALDI and ESI) mass spectrometry have been pivotal to get an accurate picture of the startling complexity of venoms [240, 241]. Different bottom-up proteomics approaches have been applied to unlock the peptide and protein composition of a number of snake venoms [238], including shotgun proteomics [242, 243] and hyphenated separation technologies [2, 244-246]. In general, these low-throughput venomics platforms provide incomplete sequence coverage, yielding toxin-family resolution. To a certain extent, locus-specific assignments can be achieved using an homologous snake venom gland transcriptome as a database for the assignment of mass spectra [247-251]. However, often the low resolution of the pre-MS separation steps does not allow to distinguish between different proteoforms or closely related (eg. polymorphic) isoforms of toxin family members due to extensive sequence similarity and/or overlapping protein masses. Top-down mass spectrometry has the potential to eliminate this shortcoming [125, 126]. It involves analysis of intact proteins, typically using electrospray ionization and high-resolution mass analysis, as isotope-resolution is needed for accurate determination of charge state of multiply-charged product ions, and thus mandatory for protein identification by BLASTP analysis of *de novo* generated top-down tag sequences, or by comparison of precursor and product ion masses with a species-specific protein sequence database. The potential for achieving full protein characterization has made the top-down mass spectrometry approach extremely useful for analysis of single proteins or simple protein mixtures [125, 126]. Here, we evaluate the possibility of including top-down mass spectrometry as part of our venomics protocol [2] to push the limits of whole venom analysis.

Results and Discussion

Proof-of-concept

The relative abundance and exact mass of the isotopic distribution peaks are distinctive features that identify any polyisotopic molecule having a unique elemental isotopic composition. Proteins of different qualitative or quantitative composition of amino acids are characterized by different molecular formulae ($C_vH_wN_xO_yS_z$) and thus by different molecular masses. With the improvement of instrument resolution, accurate determination of the mass and isotopic fine structure represents an important practical tool for identifying nonisobaric protein species. On the other hand, proteins having the same amino acid composition but arranged into different sequences are isobaric and thus additional information is needed to differentiate one protein from the other. Generally, this extra information is provided as an internal sequence tag generated through a bottom-up or a top-down MS/MS approach [252-254]. In the case of most venom proteins, the number of intra- and inter-molecular disulfide linkages represents an additional class-specific feature that allows the classification of venom toxins within the ~15 known venom protein families [240]. The aim of the present work was to assess the proof-of-concept that, given the appropriate expertise, the combination of high resolution mass determination and sequence tags generated by top-down MS CID and HCD scans represents a fast and accurate tool for locus-specific assignment of venom proteins. To this end, we chose the venom of king cobra, *Ophiophagus hannah* [255]. *Ophiophagus* is derived from the Greek, meaning "snake-eating" and *hannah* is derived from the name of tree-dwelling nymphs in the Greek mythology, referring to the arboreal lifestyle of this snake. King cobra represents the world's longest venomous snake and a species of medical concern across its distribution range in the western Ghats and north-east India, Southeast Asia, and southern areas of East Asia [256, 257]. To date, Indonesian king cobra (a Balinese specimen) is the only venomous snake species whose genome and venom gland transcriptome have been sequenced [80]. Its venom contains, among other toxins, 2-3 natriuretic peptides (3-4 kDa), 47 short and long neurotoxins of the 3FTx family (41.1% of the total venom proteome; 7-8 kDa), 4 PLA_2s (5.5%, 13-14 kDa), 2 Kunitz-type inhibitors (4.1%, 6-7 kDa), a cystatin inhibitor (9 kDa), 1 ohanin molecule (11.9 kDa), 1 cysteine-rich secretory protein (CRISP, 24 kDa), 1-2 PIII-snake venom metalloproteinases, and an L-amino acid oxidase molecule (Appendix C, Supplementary Table 9) [80, 258]. Owing to their molecular masses and lack of posttranslational modifications other than disulfide bonds, these venom proteins are particularly amenable to analysis by top-down MS.

On-line venom profiling

High-resolution LC-MS analysis (Fig.5.1) was used to determine the molecular masses of native and reduced *O. hannah* venom proteins. The mass shift between reduced and native components allowed the calculation of the number of disulfide linkages present in these toxins (Table 5.1).
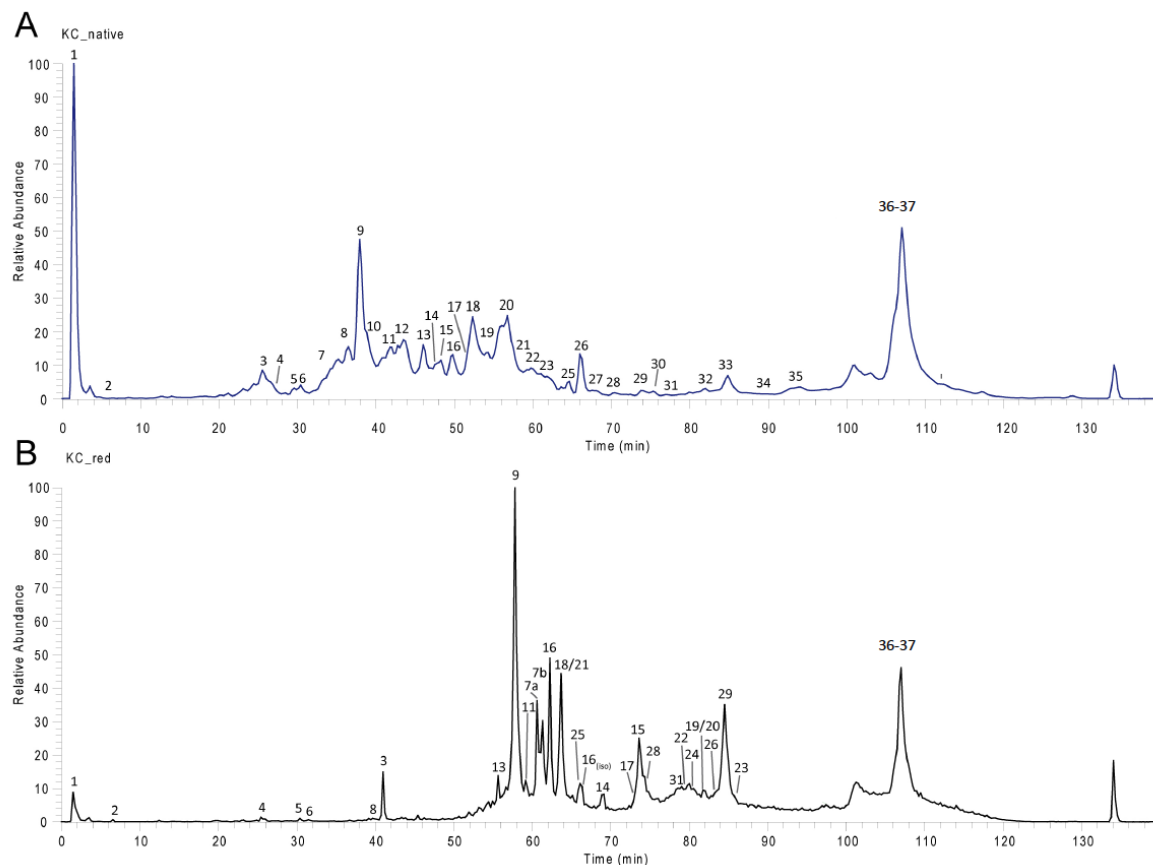
Figure 5.1. Total ion current (TIC) profiles of native (A) and reduced (B) king cobra (KC) venom separated by reverse-phase HPLC.

Using these two parameters, 23 distinct masses could be straightforward assigned to venom proteins encoded among the 18579 genes available from the NCBI database (http://www.ncbi.nlm.nih.gov/genome/genomes/10842?) (Table 5.1 and Appendix C, supplementary Table 9). However, locus-specific assignment for other masses was not possible, and protein classes were only tentatively assigned. Strikingly, the experimentally determined masses of a number of putative 3FTxs (~7kDa, 3-5 disulfide bonds) and the masses of all the four PLA$_2$ molecules did not match those calculated for the reported genome-translated proteins. Intraspecific geographic variation in venom composition has long been documented by herpetologists and toxinologists as a general feature of highly adaptable and widely distributed snake species [259-261]. The geographic

58

distribution of *O. hannah* is very extensive throughout many parts of Asia, and no subspecies are recognized. According to the Reptile Database (http://www.reptile-database.org) the range of king cobra includes Bangladesh, Myanmar, Cambodia, China (Fukien, Kwangtung, Hong Kong, Kwangsi, Hainan, Yunnan, SW Sichuan, Tibet), India (Karnataka (Dandeli); Arunachal Pradesh (Miao - Changlang district, Itanagar – Papum Pare district), Sikkim, West Bengal, Bihar, Orissa, Andhra Pradesh, Andaman Islands), Nepal, Bhutan, Indonesia (Sumatra, Java, Sulawesi, Borneo, Bangka, Bali, Mentawai Islands, Riau Islands), Singapore, Laos, Thailand, Vietnam, W Malaysia, and Philippines (Balabac, Jolo, Luzon, Mindanao, Mindoro, Negros, Palawan, Panay). Genome sequencing was undertaken from a blood sample obtained from an adult, captive, male king cobra that originated from Bali, Indonesia, and the venom provided by Latoxan (http://www.latoxan.com) is also pooled from captive king cobra specimens of Indonesian origin, although no records are available on its precise origin. Our mass profiling data point to large intraspecific genetic variability, suggesting that the currently monospecific genus *Ophiophagus* may actually represent a complex of cryptic locally adapted (sub)species.

Top-down venomics

The identity of the venom components assigned to king cobra loci by their mass and number of disulfide bonds was verified by high resolution top-down mass spectrometry using a first-generation hybrid LTQ Orbitrap XL™ instrument in data-dependent mode. To resolve isotope patterns, particularly those of overlapping differential charged ion patterns, the Orbitrap was run at highest resolution (100.000 at 400 m/z) with a scan speed of 1 Hz. To obtain the highest possible sequence coverage both HCD and CID spectra were acquired back-to-back on the same precursor ions, i.e. the most abundant ion of the survey scan was first selected for CID-fragmentation and then for an HCD-experiment before the next ion was subjected to CID-MS/MS, resulting in a total transient of 3 Hz. HCD fragmentation with Orbitrap detection has no low-mass cutoff of traditional resonant-excitation ion trap-based CID, yielding increased ion fragment coverage resulting in richer and higher quality MS/MS product ion spectra [262]. In addition, dynamic exclusion was selected, thus allowing MS/MS fragmentation of less intense charge states of the same precursor ion, enhancing the sequence coverage. However, interpretation of top-down MS/MS spectra with multiple charged fragment and wider isotope distributions is not supported by the typical database search algorithms used in bottom-up proteomics. Protein identifications from the top-down MS/MS data were accomplished using Prosight Lite. P-scores are shown in Table 5.1.

Table 5.1. Overview of king cobra, *Ophiophagus hannah*, venom proteins assigned by top-down venomics. The table displays the m/z, z values, and the deconvoluted masses of reduced and native venom proteins isolated as in Figs. 5.1 and 5.3; the calculated number of disulfide linkages; MS/MS sequence tags used for protein assignments by BLASTP, and the corresponding E values of the identifications, are listed. P-scores for ProSight Lite[44] assignments are also shown. J, Ile/Leu; B, Gln/Lys. Experimental masses matching venom component identified in the genome/transcriptome of *O. hannah*[7] deposited with the NCBI databank are underlined. Natriuretic peptide sequences assign by monoisotopic mass and disulfide bond content matching are shown between parentheses.
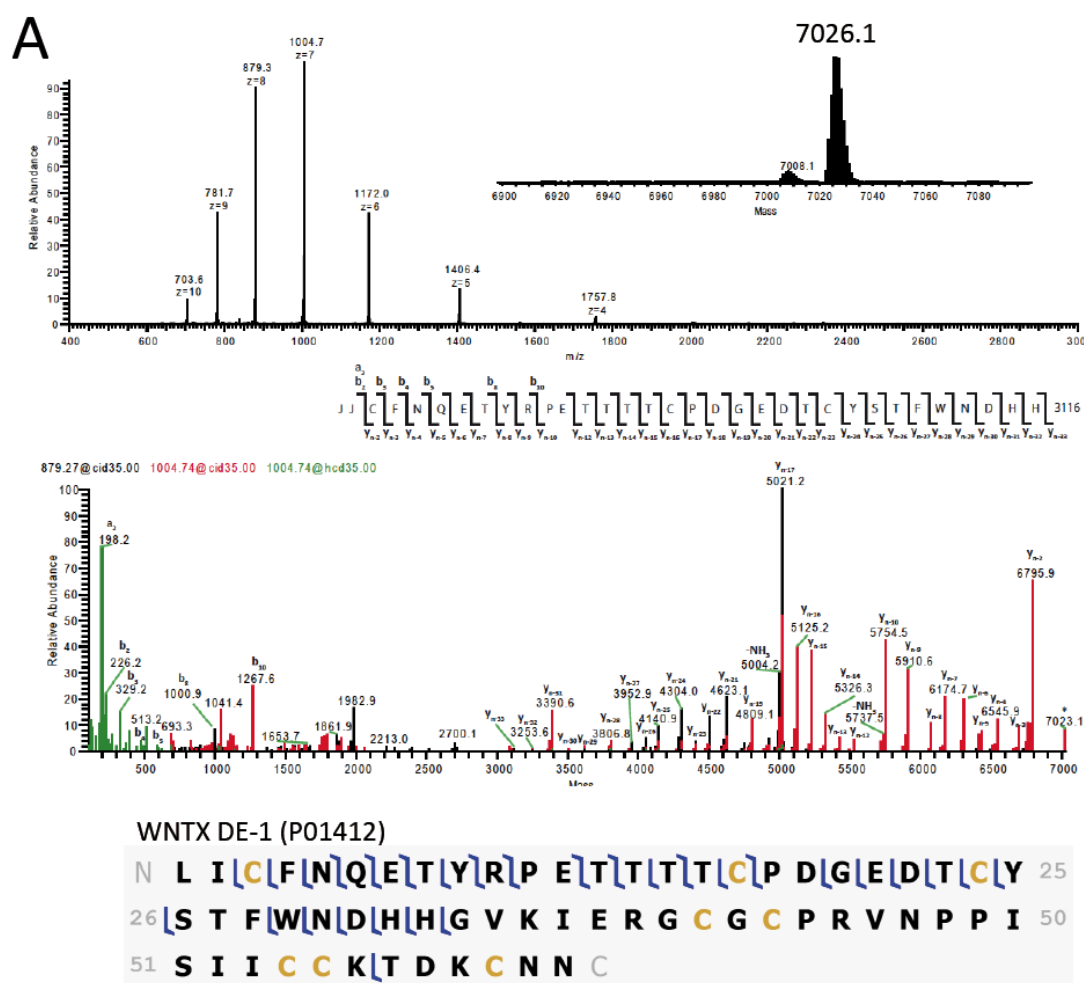
Peak #	m/z native	z	M_{ave} native [Da]	m/z reduced	z	M_{ave} reduced [Da]	nº S-S	MS/MS sequence tag or (assigned sequence by monoisotopic mass and nº of SS)	ID (Protein family)	Accession code	E value	P-score (ProSight Lite)
3	1005.78	7	7033.46	1006.93	7	7041.53	4		(3FTx)			
5	787.98	5	3934.90	788.39	5		1	(MVGPYNPAGGGGGHPSSCFGHKIDRISHSSGMGCGRRPN)	Natriuretic peptide	D9IX98		
6	688.01	6	4122.00	688.34	6	4124.02	1	(LAKKDQHNNCFGRRIDRISHSTDLGCRRRPNPPPAP)	Natriuretic peptide	D9IX98		4.4E-02
7	1135.38	7	7942.66	796.27	10	7952.67	5	(229)CYVTPDATSQTCPD(6238)	LNTX-2 (3FTx)	A8N285	1.00E-04	3.3E-46
	1130.81	7	7908.07	792.88	10	791804	5	(229)CYVTPDVBSETCPAGB(228)CYTETW(2078)TCPJVBP(1330)	LNTX OH-55 (3FTx)	Q53B58	7.00E-11	1.8E-69
8	723.86	6	4337.16	724.19	6	4339.14	1	(QQELAKKDQHNNCFGRRIDRISHSTDLGCRRRPNPPP)	Natriuretic peptide	D9IX98		
9	1003.59	7	7018.13	1004.74	7	7026.18	4	(226)CFNBETYRPETTTTCPDGEDTCYSTFWNDHH(3116)	WNTX DE-1 (3FTx)	P01412	2.00E-15	3.4E-73
10	1140.22	7	7974.54									
11	1001.15	7	7001.05	1002.31	7	7009.17	4	(226)CFNBETYR(225)TT(5209)	~WNTX DE-1 (3FTx)	P01412		2.0E-11
12	997.82	8	7974.56									
13	1077.19	7	7533.33	1078.34	7	7541.37	4	(229)(830)TPETTEJCPD(5390)	MTLP-3 (3FTx)	A8N286	9.00E-05	4.00E-44
14	906.71	7	6339.97	907.57	7	6346.00	3	(214)(604)JPAVSGFCKAYJPSFYYNPDASA(3062)	Kunitz Inhibitor	B6RLX2	3.00E-18	1.2E-101
15	1042.61	7	7291.27	811.93	9	7299.36	4	(226)(722)JFGVTPEJCAD(5201)	SNTX OH-35 (3FTx)	Q53B49	3.00E-04	5.5E-17
16	1075.77	7	7523.39	754.35	10	7533.49	5		LNTXL-6 (3FTx)	P82662		4.4E-39
	1008.23	7	7050.61	785.52	9	7060.66	5		CTX-15 (3FTx)	Q53B46		5.1E-58
17	1080.34	7	7551.38						LNTX-28	Q2VBP4		
	1066.48	6	6392.88	1067.49	6	6398.94	3	(4439)GJANNFNTVEECYJJC(145)	Acetyl-Kunitz Inhibitor	B5L5Q6	1.00E-03	
18	1002.65	7	7011.55	878.45	8	7019.60	4		CTX-27 (3FTx)	Q69CK0		1.1E-69
19/20	1476.08	9	13275.72	1477.42	9	13287.78	6		PLA₂			
21	1002.65	7	7011.55	878.45	8	7019.60	4		CTX-27 (3FTx)	Q69CK0		1.1E-69
22	1470.07	9	13221.63	1471.63	9	13235.67	7		PLA₂			
23	736.17	13	9557.21	736.17	13	9557.21	0		Cystatin	ETE69273		
24	1098.66	7	7683.62	962.21	8	7689.68	3		ILGF-I	ETE61012		
	1109.25	7	7755.77						LNTX OH-57	Q53B56		
25	1328.92	9	11951.28	920.33	13	11951.25	0	(1176)TFDSNTAFESLVV(9353)	Ohanin	P83234	5.00E-06	
	1057.2	7	7393.41						MTLP-38	Q2VBN0		
26	1262.38	6	7568.28	1263.72	6	7576.30	4		SNTX (3FTx)			
	1538.56	9	13838.04	1540.13	9	13852.18	7	(8950)(213)TCN(229)DDECGAFJ(3289)	~acidic PLA2	Q9DF56	3.00E-07	
27	1084.91	7	7583.37						LNTX OH-56 (3FTx)	Q53B57		
28	1266.88	6	7595.28	1268.22	6	7603.35	4		SNTX (3FTx)			
29	1191.07	21	24991.47						CRISP + 2M₀ₓ	Q7ZT98		
29	1387.64	18	24959.52	2082,21	12	24974.55	8		CRISP ophanin	Q7ZT98		1.8E-11
30	~ 50 kDa		~ 50 kDa			~ 50 kDa						
31	1184.04	6	7098.24	1016,19	7	7106.28	4		MTLP-6 (3FTx)	Q2VBN1		6.3E-41
32	1322.5	10	13215.00						PLA₂			
33	1711.35	8	13682.80	1522.77	9	13695.93	7		PLA₂			

When no hit was found, the top-down spectra were manually interpreted and the sequence tags submitted to BLASTP analysis. These sequence tags and the E-values for the corresponding sequence homology search are listed in Table 5.1.

Figure 5.2 displays deconvoluted top down MS/MS spectra generated for different king cobra venom toxin class proteins. Panel A displays a top-down sequence tag generated through combination of two CID spectra of the 8^+ and 7^+ charged state (m/z 879.3 and 1004.7, respectively) and an HCD spectra of the 7^+ charged state of a 3FTx tentatively identified by mass profiling as weak toxin DE-1 (P01412) (Table 5.1). The CID experiments provide mainly amide cleavage in the middle of the peptide, whereas HCD results in low molecular mass b and y ions providing sequential information for C and N termini. All the native isotope-averaged molecular mass (7018.13 Da) and number of intramolecular disulfide linkages (4) (Table 5.1), and the top-down generated sequence tag (Table 5.2), matched 100% to the genomic sequence of weak toxin DE-1 (P01412). This result was fully confirmed by ProSight Lite analysis (Fig. 5.2A, lower panel). Panels B and C show, respectively, the top-down MS/MS analyses of proteins eluted in fractions 14 and 26 of Fig.5.1 that were predicted by the binomial (molecular mass + nº of disulfide bonds) as Kunitz-type inhibitor B5L5Q6 (native M_{ave} 6339.97 Da) and a PLA_2 molecule (native M_{ave} 13838.04 Da) (Table 5.1). BLASTP analysis of the sequence tag generated by combined CID of the 6^+ (m/z 1058.67) and 7^+ (m/z 907.57) charge states and HCD of the 7^+ charge state precursors of the putative Kunitz-type inhibitor unambiguously confirmed the initial assignment. ProSight Lite analysis confirmed this match and extended the identification of fragment ions (Fig. 5.2B, bottom display). The *de novo* sequence tag obtained for the 13.8 kDa protein matched to polypeptide stretch 108TLTCNDDNDECGAF121 of acidic PLA_2 Q9DF56 from *O. hannah*. Discrepancy between the calculated (13176.6 Da) and the ESI-MS (13838.04 Da) isotope-averaged native molecular mass for this PLA_2 molecule indicates that the venom protein may represent a variant not represented in the database or it may contain one or more post-translational modifications that together add up 661.4 Da. Panel D illustrates a unique strength of the top-down proteomic approach using ion trapping mass spectrometry, the ability to sequentially isolate protein species co-eluting in a single chromatographic fraction for accurate mass measurement and selective sequencing. RP-HPLC fraction 7 The m/z 7917.7 component had the same native mass and disulfide bonds than the long neurotoxin LNTX OH-55 (Q53858). On the other hand, m/z 7952.7 could be assigned to the long neurotoxin LNTX-2 homolog (A8N285) (Tables 5.1). *De novo* assignment of CID MS/MS of the 8^+ charged ion at m/z 990.97 yielded two alternative sequence tags (Fig. 5.2, panel D, middle spectrum), one of which, (229)CYVTPDVBSETCPAGB(228)CYTETW(2078)TCPJVBP(1330), hit LNTX OH-55: TKCYVTPDVKSETCPAGQDICYTETWCDAWCTSRGKRVNLGCAATCPIVKPGVEIKCCSTDNC NPFPTRKRP. On the other hand, BLASTP analysis of the sequence tag obtained by CID MS/MS of the 8^+ charged ion at m/z 995.08, (229)CYVTPDATSQTCPD(6238) uniquely identified LNTX-2 (A8N285)

(Table 5.1). These assignments were fully supported by ProSight Lite analysis (Fig. 5.2D, lower displays). Panel E of Fig. 5.2 displays the deconvoluted top-down MS/MS analysis of the 24976.4 Da protein eluted in RP-HPLC fraction 29 (Tables 5.1 and 5.2). Analysis by ProSight Lite of the combined CID patterns of the 23^+, 25^+, and 26^+-charged isotopic peaks identified this protein as the CRISP molecule ophanin [Q7ZT98] predicted by mass and disulfide bond content measurements (Table 5.1).
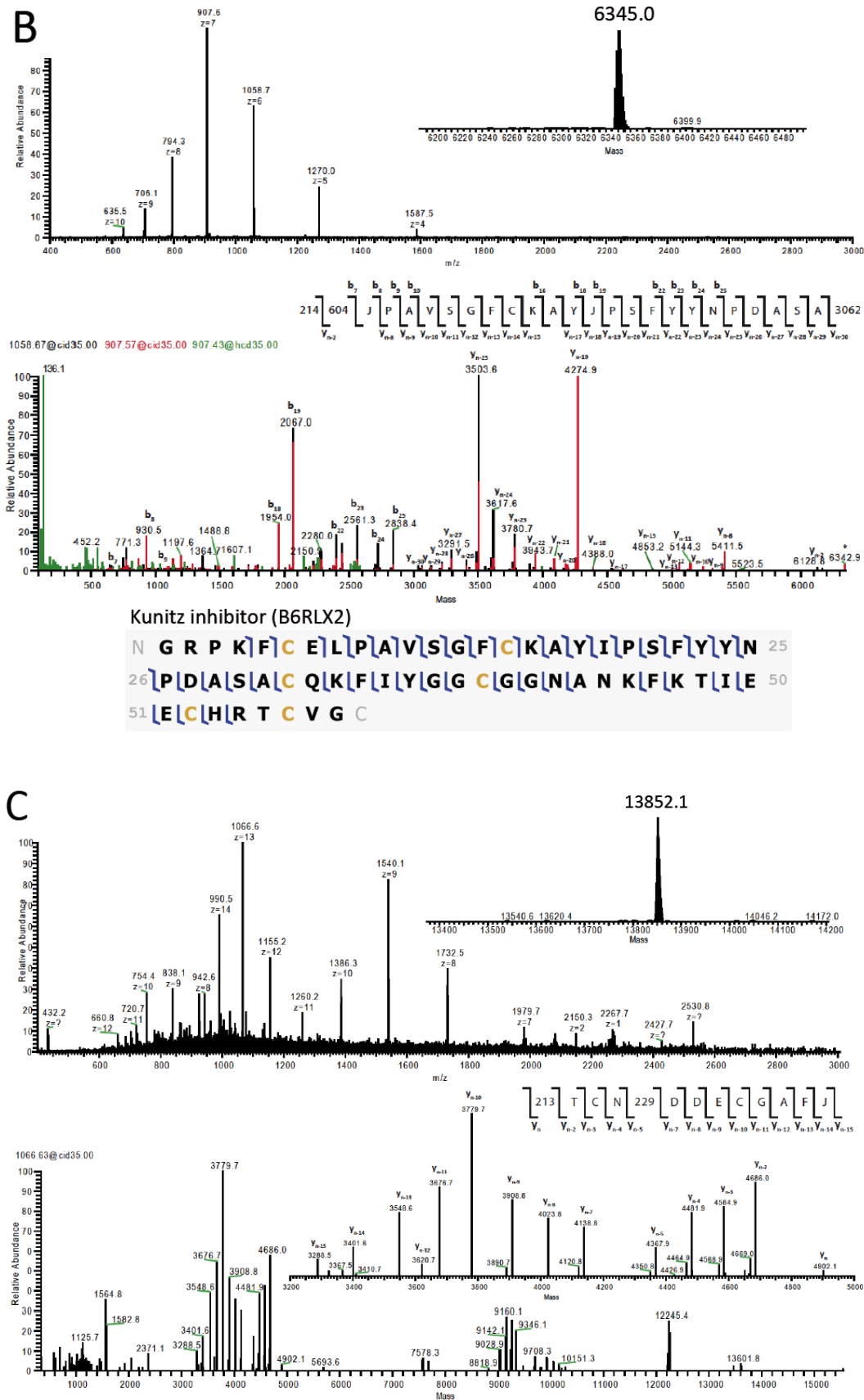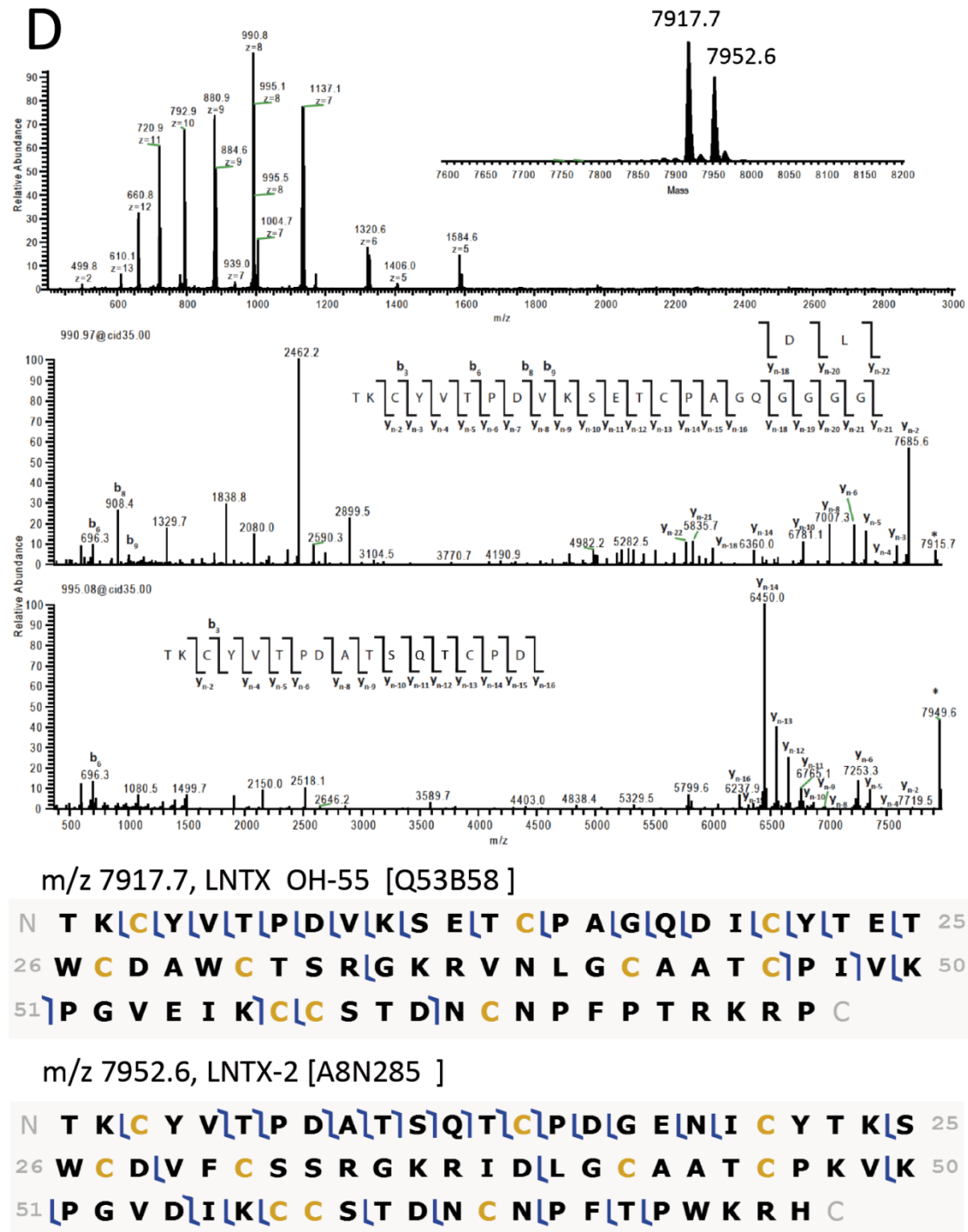

Figure 5.2. **MS and top-down MS/MS spectra**. Panels A, B, C, D, E, and F show multiple charge full scan, deconvoluted full scan, and monoisotopic MS/MS top-down spectrum for WNTX DE-1 (peak 9, Fig. 5.1), Kunitz-type inhibitor (peak 14, Fig. 5.1), a PLA_2 molecule (peak 26, Fig. 5.1), LNTX OH-55 and LNTX-2 (peak 7, Fig. 5.1), CRISP ophanin (peak 29, Fig. 5.1), and a novel putative 3 FTx (peak 3, Fig. 5.1). Identities of these proteins are listed in Table 1. MS/MS spectra comprising merged experiments, displayed in panels A, C, and D, are colour coded: Overlaid CID spectra of differentially charged precursors are displayed in black, red, and blue lines; overlaid HCD spectra are shown in green. Deconvoluted spectra were analyzed with ProSight Lite (matching b- an y- ions are shown in the sequence below the spectra) and/or interpreted manually (*de novo* sequence tags with corresponding b- and y-ion series are shown). Locus-specific fragmentation patterns identified by ProSight Lite are displayed at the bottom of panels A, B, D, and E.

Kunitz inhibitor (B6RLX2)

N G R P K F C E L P A V S G F C K A Y I P S F Y Y N 25
26 P D A S A C Q K F L I Y G G C G G N A N K F K T I E 50
51 E C H R T C V G C

Figure 5.2. **MS and top-down MS/MS spectra (continued).**

m/z 7917.7, LNTX OH-55 [Q53B58]

N T K C Y V T P D V K S E T C P A G Q D I C Y T E T 25
26 W C D A W C T S R G K R V N L G C A A T C P I V K 50
51 P G V E I K C C S T D N C N P F P T R K R P C

m/z 7952.6, LNTX-2 [A8N285]

N T K C Y V T P D A T S Q T C P D G E N I C Y T K S 25
26 W C D V F C S S R G K R I D L G C A A T C P K V K 50
51 P G V D I K C C S T D N C N P F T P W K R H C

Figure 5.2. MS and top-down MS/MS spectra (continued).

CRISP [Q7ZT98]

```
N V D F N S E S T R R Q K K Q K E I V D L H N S L
R R S V S P T A S N M L K M Q W Y P E A A S N A E
R W A S N C N L G H S P D Y S R V L E G I E C G E
N I Y M S S N P R A W T E I I Q L W H D E Y K N F
V Y G V G A N P P G S V T G H Y T Q I V W Y K T Y
R I G C A V N Y C P S S E Y S Y F Y V C Q Y C P S
G N M R G S T A T P Y K S G P T C G D C P S A C D
N G L L C T N P C T L Y N E Y T N C D S L V K Q S S
C Q D E W I K S K C P A S C F C H N K I I
```

Figure 5.2. **MS and top-down MS/MS spectra (continued).**

65

Figure 5.2. **MS and top-down MS/MS spectra (continued).**

The reduced isotope-averaged molecular mass of the venom component eluted in RP-HPLC fraction 3 (7041.53 Da, Table 5.1) and its calculated number of intramolecular disulfide bonds (n=4) suggested that this chromatographic fraction contained a 3FTx. However, this molecular characteristics do not match to any known king cobra mature venom toxin (Appendix C, supplementary Table 9). In addition, BLASTP analysis of the *de novo* sequence tag generated by combining CID and HCD product ion spectra of the m/z 881.3 (z= 8^{+}) shown in panel F of Fig. 5.2 did not hit any known king cobra protein, likely indicating the occurrence of genetic variability between the specimens used for genome sequencing and venom proteome analysis (this work).

Bottom-up venomics

Fragmentation of intact proteins in the gas phase promises a bright future for proteomics by offering access to the complete protein sequence and site-specific PTM analysis in a single experiment [263]. However, current top-down approaches suffer also from several limitations. In particular, the complex fragmentation spectra generated by multiply charged proteins can introduce ambiguity in the interpretation of the *de novo* top-down MS/MS spectra, and the fact that bioinformatic tools for top-down proteomics are still primitive compared to those for bottom-up proteomics, limits database searching. The traditional bottom-up mass spectrometry peptide fragmentation-based approach of inferring the identification of the original protein yields simple product ion spectra easily interpretable through *de novo* sequencing or by high-throughput automated database searching. In addition, MS analyses of peptides are, in general, more sensitive than analyses of intact proteins. Thus, despite the serious "inference" problem associated with peptide-based protein identification, the bottom-up approach still represents a valuable complement to top-down analysis of unknown and large proteins, and was used in this work to confirm top-down assignments and to characterize king cobra venom components uncharacterized in the top-down approach.
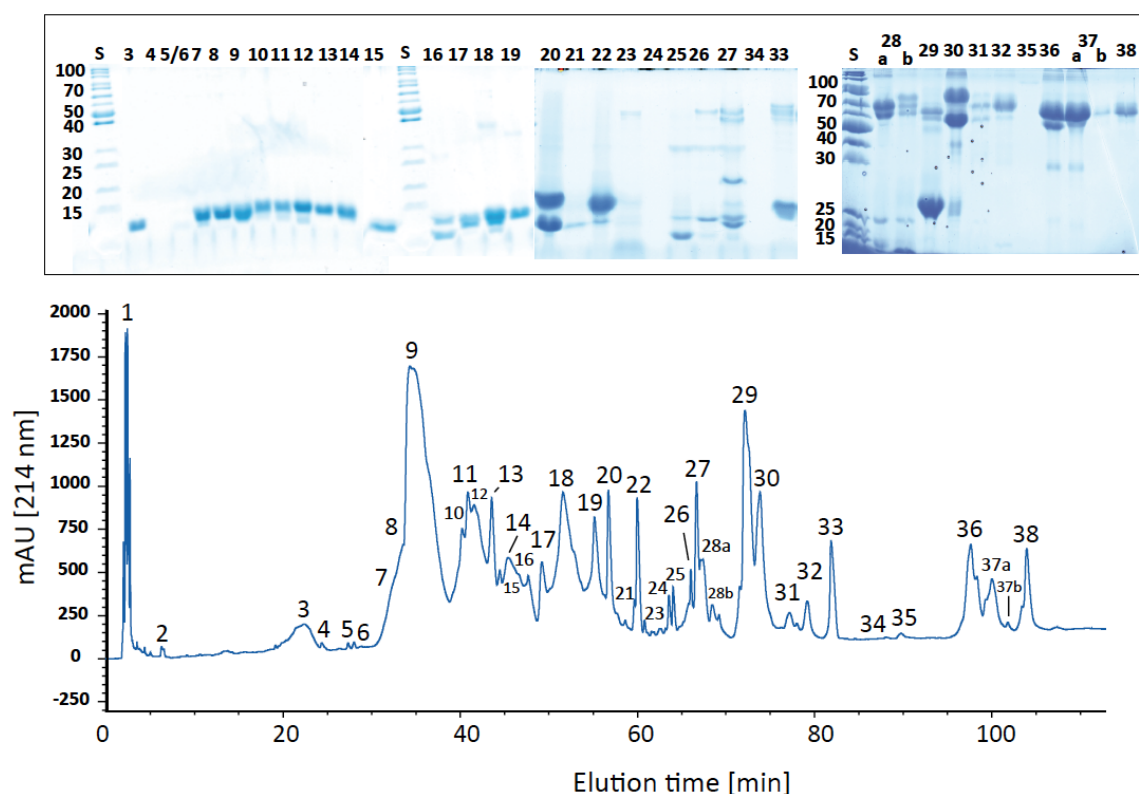
Figure 5.3. Reverse-phase HPLC separation of king cobra venom proteins. Fractions were collected manually and analyzed by SDS-PAGE under reducing conditions (upper panel). Proteins identified by bottom-up MS/MS analysis are listed in Table 5.2.

Reverse-phase separation of Indonesian king cobra venom proteins yielded ~40 chromatographic peaks, many of which contained several components (Fig. 5.3).

The bottom-up results listed in Table 5.2 confirmed the presence of proteins not represented in the NCBI database in the RP-HPLC fractions 3, 7, and 26. These proteins had been assigned by *de novo* top-down sequencing to 3FTx (7041.5 Da, 7952.6 Da) and PLA$_2$ (13852.1 Da) molecules (Fig. 5.2, Table 5.1). The bottom-up approach also confimed the top-down assignment of m/z 11951.3 (RP-HPLC fraction 25) to ohanin. On the other hand, the absence of weak toxin DE-1 and Kunitz-type inhibitor proteins in the bottom-up assignments for RP-HPLC fractions 9 and 14, respectively, may indicate that these proteins do not contain convenient enzymatic cleavage sites and/or the resulting tryptic peptides did not produce significant MS/MS data.

Bottom-up venomics identified the proteins found in the top-down approach and another ~22-25 distinct proteins were characterized in tryptic peptide fragmentation experiments (Table 5.2). However, compared to top-down MS/MS, bottom-up peptide-centered proteomics had the shortcoming that, because of the lack of intact mass information and unique locus-specific peptides, some identifications remained ambiguous (Table 5.2). The combined complementary top-down and bottom-up approaches provide a more comprehensive proteomic view of king cobra venom than each approach separately. The combined data indicate that Indonesian king cobra venom comprises proteins from 10 toxin families (natriuretic peptides, 3FTx, Kunitz-type inhibitor, Group IA PLA$_2$, ohanin, CRISP, cystatin, ILGF-1, PIII-SVMP, and LAAO), whose relative abundances is displayed in Fig. 5.4. Except for the 3FTx and PLA$_2$ protein families, represented respectively by ≥ 20 and 5 members, only one protein of each of the other toxin classes was expressed into the king cobra venom. The high concentration of 3FTxs is in line with the recent demonstration that this toxin family has undergone massive expansion by gene duplication and neofunctionalization under positive selection [80]. This genomic response seemingly occurred in response to a co-evolutionary arms race between ophiophagous king cobra and its neurotoxic snake prey. Complementing the genomic evidence [80], the venom proteome characterization reported here shows that the most expressed proteins of the FTx family were LNTX OH-55 [Q53B58], WNTX DE-1 [P01412], CTX-27 [Q69CK0], and MTLP-3 [A8N286] (Fig. 5.4).

.

Table 5.2. Protein assignments of the peptide ion sequences derived by ESI-MS/MS analysis of tryptic digests of SDS-PAGE-separated protein bands found in the RP-HPLC fractions shown in Fig. 5.3 by *de novo* (*dn*) sequencing and BLASTP against the *Elapidae* (taxid:8602) protein subset of UniProtKB/TrEMBL and NCBI non-redundant databases), and database search against a database comprising all putative venom proteins from *O. hannah* (Appendix C, Supplementary Table 9) and the *O. hannah* dataset of the NCBI database. ~, "similar to". Cysteine residues are carbamidomethylated.

RP-HPLC peak	Apparent mass (SDS-PAGE)	m/z	z	Amino acid sequence	Match	Confidence (%)	Accession code of best match	Snake species
3	7 kDa	814.82	2	CCSTDNCNPFPTR	LNTX OH55	100	Q53B58	*O. hannah*
		625.50	2	NEEVLQCCAK	~NTL4	*dn*	Q9YGI8	*B. multicinctus*
		559.95	2	(216.9)TATCPKPK	3FTX	*dn*	ADN67592	*B. multicinctus*
7	10 kDa	814.82	2	CCSTDNCNPFPTR	LNTX OH55 1-72	100	Q53B58	*O. hannah*
		1013.50	2	VNLGCAATCPIVKPGVEIK		100		*O. hannah*
		741.92	2	AATCPIVKPGVEIK		100		*O. hannah*
8	10 kDa	814.82	2	CCSTDNCNPFPTR	LNTX OH55 1-72	100	Q53B58	*O. hannah*
		1013.50	2	VNLGCAATCPIVKPGVEIK		100		*O. hannah*
9	10 kDa	814.82	2	CCSTDNCNPFPTR	LNTX OH55	100	Q53B58	*O. hannah*
		676.03	3	VNLGCAATCPIVKPGVEIK		100		*O. hannah*
		620.35	2	CPIVKPGVEIK		100		*O. hannah*
11	12 kDa	893.85	2	CCSTDNCNPFTPWK	LNTX2/LNTX OH-17	100	P01386/Q53B54	*O. hannah*
		603.28	2	IDLGCAATCPK		100		*O. hannah*
12	12 kDa	603.28	2	IDLGCAATCPK	~LNTX2 homolog	100	~A8N285	*O. hannah*
		652.26	2	SWCDVFCSSR		100		*O. hannah*
		893.85	2	CCSTDNCNPFTPWK		100		*O. hannah*
13	12 kDa	791.35	2	GCTFTCPELTPNGK	MTLP-3	100	A8N286	*O. hannah*
		986.74	3	CYNHQSTTPETTEICPDSGYFCYK		100		*O. hannah*
14	11 kDa	666.28	2	SWCEVFCTSR	LNTX OH-22	100	Q2VBP5	*O. hannah*
		893.85	2	CCSTDNCNPFTPWK		100		*O. hannah*

		652.26	2	SWCDVFCSSR	~LNTX2 homolog	100	A8N285		*O. hannah*
15	6 kDa	860.38	2	TWHMVYPGGYDHTR	SNTX OH-35	100	Q53B49		*O. hannah*
		735.82	2	EICADGQNLCYK		100			*O. hannah*
		912.91	2	GVTPEICADGQNLCYK		100			*O. hannah*
16	7 kDa	573.93	3	TWHMVYPGGYDHTR	SNTX OH-35	100	Q53B49		*O. hannah*
	6 kDa	912.91	2	GVTPEICADGQNLCYK		100			*O. hannah*
17	7 kDa	705.30	3	PGIDIECCSTDNCNPHPK	LNTX OH-28	100	Q2VBP4		*O. hannah*
		885.42	3	LICFISSHDSVTCAPGENVCFLK		100			*O. hannah*
		1057.94	2	PGIDIECCSTDNCNPHPK	LNTX OH-37	100	Q53B59		*O. hannah*
		506.19	2	CDAWCGSR		100			*O. hannah*
		670.33	2	KLSFGCAATCPK		100			*O. hannah*
18	8 kDa	606.28	2	LSFGCAATCPK	LNTX-6	100	P82662		*O. hannah*
		776.33	3	VNPGIDIECCSTDNCNPHPK		100			*O. hannah*
		624.75	2	SWCDAWCGSR		100			*O. hannah*
19	9 kDa	1102.56	2	CLNTPLPLIYTTCPIGQDK	CTX-27	100	Q69CK0		*O. hannah*
		545.81	2	LPSKYDVIR		100			*O. hannah*
		609.89	2	KLPSKYDVIR		100			*O. hannah*
		799.34	2	SSADVEVLCCDTNK	CTX-9/15	100	Q2VBN8/Q53B46		*O. hannah*
		666.37	2	CLNTPLPLIYK		100			*O. hannah*
		545.81	2	LPSKYDVIR		100			*O. hannah*
		609.86	2	KLPSKYDVIR		100			*O. hannah*
20	15 kDa	909.33	2	ADNDECAAFVCDCDR	PLA$_2$	100	Q9DF33		*O. hannah*
		583.96	3	APYNKENINIDTTTR		100			*O. hannah*
		959.73	3	CCQVHDNCYTQAQQLTECSPYSK		100			*O. hannah*
		753.83	2	CGSGGSGTPVDELDR		100			*O. hannah*
20, 21	6 kDa	1102.56	2	CLNTPLPLIYTTCPIGQDK	CTX-14	100	Q2VBN7		*O. hannah*
		545.81	2	LPSKYDVIR		100			*O. hannah*

		648.32	2	IYTTCPIGQDK		100		*O. hannah*
		778.07	3	KCLNTPLPLIYTTCPIGQDK		100		*O. hannah*
		609.86	2	KLPSKYDVIR		100		*O. hannah*
22	15 kDa	607.83	2	HLVQFNGMIR	PLA$_2$	100	Q9DF33	*O. hannah*
		606.78	2	QLTECSPYSK		100		*O. hannah*
		885.35	4	RYSYDCSEGTLTCKADNDECAAFVCDCDR		100		*O. hannah*
		900.43	3	VAAICFAGAPYNKENINIDTTTRC		100		*O. hannah*
		1128.09	3	YSYDCSEGTLTCKADNDECAAFVCDCDR		100		*O. hannah*
25	12 kDa	757.38	3	ADVTFDSNTAFESLVVSPDKK	Ohanin	100	P83234	*O. hannah*
		749.36	2	FSSSPCVLGSPGFR		100		*O. hannah*
		512.60	3	YGTQREWAVGLAGK		100		*O. hannah*
	5 kDa	716.87	2	LEMYCAPVKPPK	ILGF-I	100	ETE61012	*O. hannah*
		595.81	2	MYCAPVKPPK		100		*O. hannah*
		624.61	3	GIVEECCFQSCDLVR		100		*O. hannah*
26	15 kDa	454.24	2	VAAICFAR	PLA$_2$	*dn*	~Q9PSN5	*N. s. scutatus*
	12 kDa	749.36	2	FSSSPCVLGSPGFR	Ohanin	100	P83234	*O. hannah*
		512.60	3	YGTQREWAVGLAGK		100		*O. hannah*
27	9/10 kDa	642.75	2	SWCDAWCGSR	LNTX-6/28, OH-37	100	P82662, Q2VBP4, Q53B59	*O. hannah*
28 a,b	70 kDa	676.84	2	TSVAVVQDYSKR	PIII-SVMP	*dn*	~ D5LMJ3	*N. atra*
		598.81	2	TSVAVVQDYSK	PIII-SVMP	*dn*	~ Q9PVK7	*N. kaouthia*
		664.31	2	DFEMVNTVNMK	PIII-SVMP	*dn*	~ Q9PVK7	*N. kaouthia*
	56 kDa	797.39	2	NDNAELLTGLDFSGK	PIII-SVMP	*dn*	~ ADD14036	*N. atra*
		614.81	2	SLGVAYMGSLCK	PIII-SVMP	*dn*	~ D6PXE8	*N. atra*
		656.33	2	YVEMVNTLNTK	PIII-SVMP	*dn*	~ D5LMJ3	*N. atra*
29	26 kDa	916.45	2	AWTEIIQLWHDEYK	CRISP	100	Q7ZT98	*O. hannah*
		776.84	2	MQWYPEAASNAER		100		*O. hannah*

		918.79	3	NFVYGVGANPPGSVTGHYTQIVWYK		100		*O. hannah*
		882.38	2	WASNCNLGHSPDYSR		100		*O. hannah*
		640.78	2	QSSCQDEWIK		100		*O. hannah*
		1144.8	3	IGCAVNYCPSSEYSYFYVCQYCPSGNMR		100		*O. hannah*
30	85 kDa	753.39	2	NDNAQLLTGLGFSR	PIII-SVMP	*dn*	~ D6PXE8	*N. atra*
		649.33	2	VYEMVNTVNTK	PIII-SVMP	*dn*	~ D6PXE8	*N. atra*
	58 kDa	663.85	2	TSVAVVQDHSKR	PIII-SVMP	*dn*	~ F8RKV9	*D. coronoides*
32	70 kDa	692.89	2	(128.15)TVGLAYMGTLCK	PIII-SVMP	*dn*	~ F8RKV9	*D. coronoides*
		626.66	3	(626.32)AELLTSLDFSR	PIII-SVMP	*dn*	~ AHZ08819	*M. ikaheka*
36-38	65 kDa	553.62	3	TNNVIIPCKPTDVK	PIII-SVMP	100	A3R0T9	*O. hannah*
		956.96	2	AMCDVLQSVGIVQDYSK		100		*O. hannah*
		574.59	3	ASYSEIEDIGMVDHR		100		*O. hannah*
		795.92	2	VFDMVNYITVVYK		100		*O. hannah*
		1214.48	2	NECDLPEFCIGQSAECPMDR		100		*O. hannah*
		581.56	3	NGHSCQNDQGYCFR		100		*O. hannah*
36, 37	52 kDa	508.92	3	EAGHEVVILEASDR	LAAO	100	P81383	*O. hannah*
		685.42	2	IVIVGAGISGLTAAK		100		*O. hannah*
		804.63	4	HVINLEECFQEPEYENWLATASHGLTK		100		*O. hannah*
		698.98	3	MSANNPENFGYQLNPNER		100		*O. hannah*
		804.63	4	HVINLEECFQEPEYENWLATASHGLTK		100		*O. hannah*
		749.47	2	KIVIVGAGISGLTAAK		100		*O. hannah*
		1047.97	2	MSANNPENFGYQLNPNER		100		*O. hannah*
		784.37	3	QKMSANNPENFGYQLNPNER		100		*O. hannah*
		1072.51	3	HVINLEECFQEPEYENWLATASHGLTK		100		*O. hannah*
		687.01	3	IKTHREDGWYVDVGPMR		100		*O. hannah*
		780.04	3	IYFAGEYTAHPHGWIETSMK		100		*O. hannah*

Composition-toxic activity correlations

The main effect of king cobra bite is its neurotoxicity that rapidly blocks neuro-muscular transmission causing rapid onset of neuromuscle flaccid paralysis in victims, and delayed or inadequate treatment can lead to respiratory failure and death [264-267]. This clinical picture points to a central role for post-synaptically-acting alpha-neurotoxic and cardiotoxic 3FTxs [268] and acidic PLA_2 molecules [269] in the envenomation strategy developed by this species. In concordance with the clinical observations, the venom composition of king cobra is heavily dominated by 3FTxs (Fig. 5.4). Specifically, the venom proteome contained proteins from the 5 major subfamilies into which the 3FTx family can be functionally and structurally subdivided, namely 10-12 long-chain neurotoxins (LNTX) and short-chain neurotoxins (SNTX), 1 weak toxin/neurotoxin (WNTX), 3 muscarinic toxin-like proteins (MTLP) and 3-4 cytotoxins/cardiotoxins (CTX) [268, 269] (Fig. 5.4; Tables 5.1 and 5.2). LNTXs (65-72 residues, 5 disulfide linkages) and SNTXs (60-62 amino acids crosslinked by 4 intramolecular disulfide bonds) are ⍺-neurotoxins (⍺-NTX) that antagonize the function of muscular nicotinic cholinoceptor (nAChR), and LNTXs exhibit also high specificity towards neuronal-type α7, α8 and α9 nAChR. WNTXs also hit muscular-type and neuronal-type nicotinic receptors, but are less potent (LD50 ~5-80 µg/g) than typical α-NTXs (LD50 ~0.1 µg/g) [270, 271]. MTLPs selectively antagonize distinct subtypes of muscarinic receptors (mAChRs) [272]. CTXs have a wide range of pharmacological activities [268, 273], being their major toxic effects cytotoxic and cytolytic leading to tissue destruction [268, 269, 274].

Although containing molecules from the five major 3FTx groups (Tables 5.1 and 5.2), Indonesian *O. hannah* venom is dominated by α-NTXs (37.1% of the total venom proteins; Fig. 5.4), strongly supporting the view that these post-synaptically-acting neurotoxins are the major contributors to the neurotoxic effects of king cobra venom. Acidic PLA2s, interacting synergistically with cytotoxins [275] and extracellular matrix-degrading Zn^{2+}-dependent PIII metalloproteinases [276], may potentiate the extensive tissue destruction and necrosis observed in surviving victims of king cobra envenomings. The occurrence of potent hypotensive natriuretic peptides [277], ohanin, a protein that produces dose-dependent hypolocomotion and hyperalgesia (in mice) [278], and particularly the high concentration of cysteine-rich secretory protein (CRISP), a widely distributed protein in snake venoms whose reported activities include inhibition of smooth muscle contraction and cyclic nucleotide-gated ion channels [279, 280], suggest that these molecules participate in prey incapacitation and perhaps also in human envenoming.
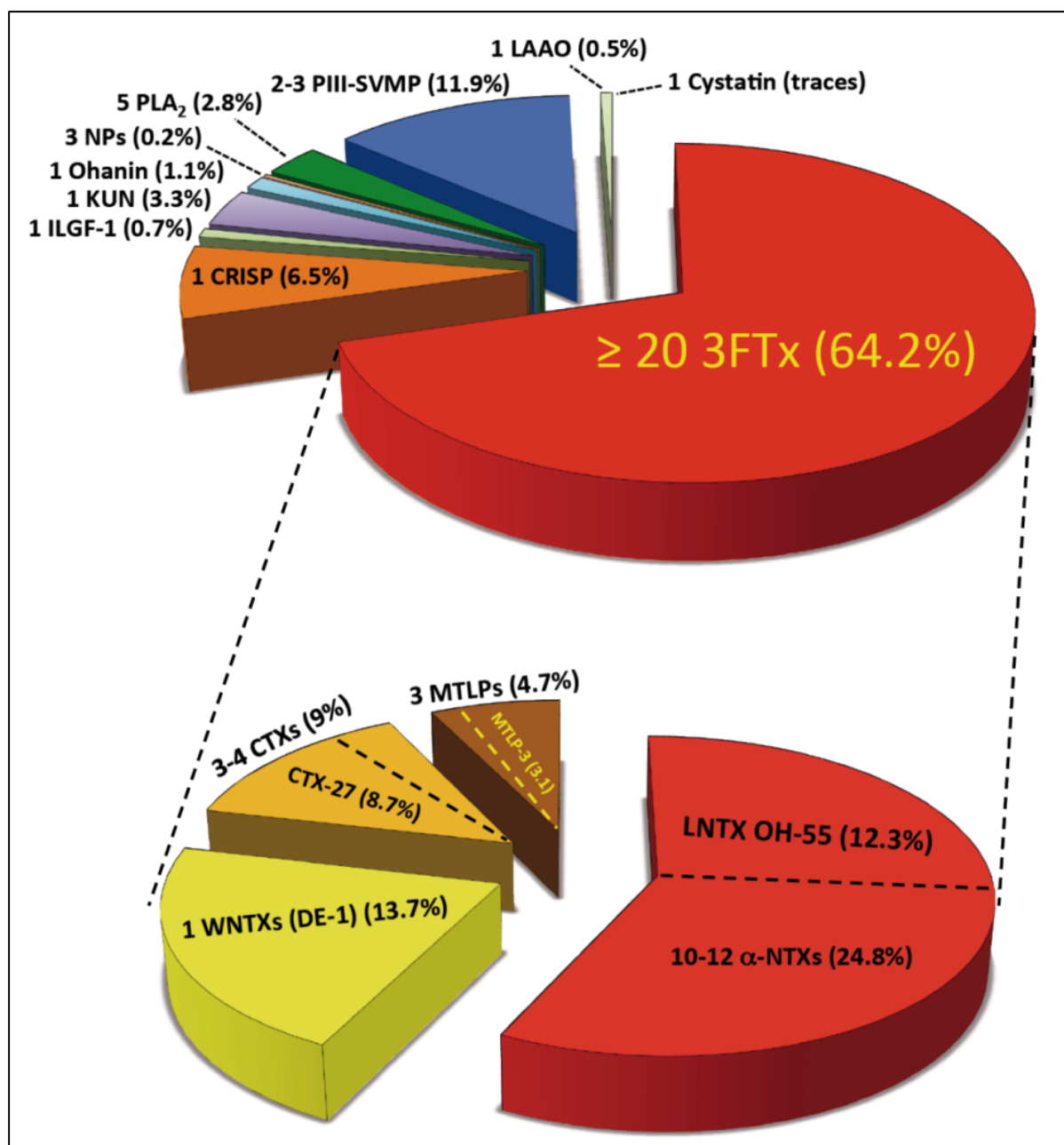
Figure 5.4. **Protein family composition of Indonesian king cobra venoms.** Upper pie chart represents the relative occurrence (in percentage of the total RP-HPLC-separated venom proteins; Fig.5.3) and the estimated number of protein species from the different toxin families as identified in the current work. 3FTX, three-finger toxin; CRISP, cysteine-rich secretory proteins; ILGF-1, insulin-like growth factor 1; KUN, Kunitz-type inhibitor; NP, natriuretic peptide; PLA_2, phospholipase A_2; PIII-SVMP, snake venom metalloproteinase of class PIII; LAAO, L-amino acid oxidase. The lower pie chart shows the relative contributions (in percentage of the total RP-HPLC-separated venom proteins; Fig. 5.3) of different subfamilies of the 3FTx family: α-NTXs, post-synaptic short- and long-neurotoxins; WNTXs, weak neurotoxins; CTX, cytotoxins; MTLP, muscarinic toxin-like proteins. The most expressed protein within each group is specified.

Concluding remarks

Thanks to recent advances in instrumentation, dissociation strategies, and bioinformatic tools, top-down mass spectrometry is increasingly gaining momentum in proteomic analysis [125, 126, 281-287]. In addition of preserving the inherent value of intact protein mass measurement, the top-down approach is especially useful for the characterization of peptides/proteins that do not contain convenient enzymatic cleavage sites that generate sizeable peptides (generally 0.4-3.5 kDa) amenable for peptide-centric bottom-up analysis. This work, which represents the first application of top-down analysis in snake venom research, illustrates the opportunity of integrating this approach to achieve locus-specific resolution in snake venomics. Top-down venomics, alone or combined with bottom-up MS/MS, may represent the method of choice for unravelling the toxin composition of other snake venoms, particularly those of terrestrial and marine elapids [273, 288-290], but also venoms from other venomous creatures, such as arthropod (spider, scorpions, and centipedes), hymenopteran, and cone snail venoms, among others. Complex mixtures of inorganic salts, biogenic amines, peptides and proteins (mainly <10 kDa), that display high specificity and affinity for neuronal ion channels and receptors, account for most of their molecular and functional diversity of these venom proteomes [291-299]. These venom peptides are generally stabilized by multiple intramolecular disulfide bonds that impart high levels of chemical, thermal and biological stability [229, 300], but whose small size and amino acid sequence may also impose serious analytical problems to bottom-up approaches. Efficient separation of the venom components, followed by a top-down sequencing approach and species-specific database search would solve this analytical limitation allowing automating much of the venom analysis.

Experimental

Preparation of venom samples

Lyophilized venom of king cobra, *Ophiophagus hannah*, from Indonesia (no specific locality disclosed) was purchased from Latoxan (Valence, France) and maintained at -20 °C until used. For LC-MS analysis, the crude venom was dissolved in 1% (v/v) formic acid to a final concentration of 10 mg/mL, and centrifuged at 20,000 xg for 5 min. For reduction of disulfide bonds, the procedure described by Zhao *et al*. [301] was followed. Briefly, 10 µL of venom (10 mg/mL) was mixed with 10 µL of tris(2-carboxyethyl)phosphine (TCEP, 0.5 M) and 30 µL of citrate buffer (0.1 M. pH 3), and the reaction

mixture incubated for 30 min at 65 °C. Thereafter, the sample was mixed with an equal volume of 1% formic acid and centrifuged at 20,000 xg for 5 min. 10 µL of each, reduced and non-reduced samples were submitted to top-down measurements.

Top-down mass spectrometry

LC-ESI-HR-MS/MS experiments were carried out on an LTQ Orbitrap XL mass spectrometer (Thermo, Bremen, Germany) equipped with an Agilent 1260 HPLC system (Agilent, Waldbronn, Germany). Reverse-phase HPLC separation was performed using a Suppelco Discovery 300Å C18 (2 x 150 mm, 3 µm particle size) column. The flow rate was set to 0.3 mL/min and the column was developed with a linear gradient of 0.1% formic acid in water (solution A) and 0.1% formic acid in acetonitrile (solution B), isocratically (5% B) for 1 min, followed by 5-50% B over 40 min, 40-70% over 20 min, 70% B for 10 min, and re-equilibration in 5% B. ESI settings were 60 L/min sheath gas; 20 L/min auxiliary gas; spray voltage, 4.8 kV; capillary voltage, 46 V; tube lens voltage, 135 V; and capillary temperature, 330 °C. For information dependent acquisition (IDA), four scan events were set with 2 micro scans and 500 msec maximal fill time. The survey scan was performed with mass resolution (R) of 100.000 (at m/z 400). For MS/MS R was set to 60,000 (at m/z 400). Every cycle contained two CID scans of the two most abundant ions and one HCD scan for the most abundant ion of the survey scan. Normalized collision energy was set to 35% for CID and 35% for HCD. Default charge states were set to 4+ and the activation time to 30 msec. Mass window for precursor ion selection was set to 2 m/z. Dynamic exclusion was set to a 3 m/z exclusion window for precurosor ions with 1 repeat within 10 sec. The exclusion list size was set to 50 ions for duration of 20 sec. The XTRACT algorithm of Xcalibur (Thermo, Bremen, Germany) was used to deconvolute isotopically resolved spectra to produce a monoisotopic mass peak list. The Zscore algorithm [302] implemented in the magic transformer (MagTran) freeware program was used for automated charge distribution deconvolution. For protein identifications, deconvoluted mass spectra were imported into, and analyzed using, ProSight Lite [287] (a free Windows application for matching a single candidate protein sequence and its modifications against a set of mass spectrometric observations, available from http://prosightlite.northwestern.edu/) against the king cobra venom proteins database (Appendix C, Supplementary Table 9) and/or interpreted manually. The resulting *de novo* generated sequence tags were BLASTed against the NCBI non-redundant Elapidae (taxid:8602) database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) using the BLASTP algorithm [209] and/or interpreted manually.

Bottom-up MS/MS

Crude venom was dissolved at 10 mg/mL in aqueous 1% formic acid (HFO) and 5% acetonitrile (ACN), and 100 µL were used for semipreparative reverse-phase (RP) HPLC separation using a Suppelco Discovery 300 Å C18 (4.6 x 150 mm, 3 µm particle size) column. The flow rate was set to 1 mL/min and the column was developed with a linear gradient of 0.1% HFO in water (solution A) and 0.1% HFO in ACN (solution B), isocratically (5% B) for 5 min, followed by linear gradients of 5-40% B for 95 min, 40-70% for 20 min, 70% B for 10 min, and re-equilibration at 5% B for 10 min. Peak detection was performed at λ =214 nm using a DAD detector. Chromatographic fractions were collected manually, dried in a vacuum centrifuge and submitted to SDS-PAGE under reducing conditions. Coomassie Brilliant Blue G250-stained bands were excised from the gel and subjected to in-gel reduction (10 mM dithiothreitol in 25 mM ammonium bicarbonate, pH 8.3, for 45 min at 65 °C) and alkylation (50 mM iodacetamide 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/L in 25 mM ammonium bicarbonate, 10% ACN; 0.25 µg/sample). Tryptic digests were dried in a vacuum centrifuge, redissolved in 15 µL of 5% ACN containing 0.1% HFO, and submitted to an LC-MS/MS on an Orbitrap XL hybrid mass spectrometer (Thermo, Bremen, Germany) coupled to an HPLC system (Agilent, Waldbronn, Germany). Chromatographic separation was performed on a Grace Vydac 218MSC18 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 5%B for 2 min, followed by 10 min from 5-40% B for 10 min, 40-99% B for 15 min, and 99% B for 5 min.

MS experiments were performed in the Orbitrap with R=15,000 at m/z 400 and maximum filling time of 200 msec for both survey and first product ion scans. A second product ion scan was recorded in the LTQ. MS/MS fragmentation of the most intense ion was performed in the LTQ using CID (30 msec activation time); the collision energy was set to 30% for LTQ readouts and 35% for Orbitrap readouts. Precursor ion isolation mass window was 2 m/z. A window of 3 m/z 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 database comprising all putative venom proteins from *O. hannah* (64 entries, Appendix C, Supplementary Table 9). To avoid losing potential assignments to proteins not classified as "venom proteins", MS/MS spectra were also searched against the *O. hannah* dataset of the NCBI database (release of 8th December 2014, comprising 18593 protein sequences) using X!Tandem as the search engine. Peptideshaker [303] was employed for graphical display of MS/MS spectrum matches and visual representation of the protein sequence coverage. For the search engine, 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. Only database hits with 100% confidence score were considered and, in addition, manually inspected. *De novo* annotation of

MSMS spectra of proteins not present in the above databases was performed manually or with the DeNovoGUI tool [304]. Deduced sequence tags were searched against the *Elapidae* (taxid:8602) protein subset of UniProtKB/TrEMBL (release 2014_10) and NCBI non-redundant databases using BLASTP [209].

The relative abundances (expressed as percentage of the total venom proteins) of the different protein families were calculated as the ratio of the sum of the areas of the reverse-phase chromatographic peaks (Fig.5.3) containing proteins from the same family to the total area of venom protein peaks in the reverse-phase chromatogram [2, 246]. When more than one protein band was present in a reverse-phase fraction, their proportions were estimated by densitometry of Coomassie-stained SDS-polyacrylamide gels using ImageJ version 1.47 (http://rsbweb.nih.gov/ij).

DATA ACCESSIBILITY

Mass spectrometry proteomics data (.mzid, .mgf and .raw files) have been deposited with the ProteomeXchange Consortium [305] (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository under Project Name "Venom proteomics of Indonesian king cobra" and dataset identifier PXD001726 and 10.6019/PXD001726.

6   Mass spectrometry guided venom profiling and bioactivity screening of the Anatolian Meadow Viper, *Vipera anatolica*

Introduction

Snake venom, which consists mainly of a mixture of proteins and peptides, has evolved over eons of years to immobilize and kill the prey as fast and as efficient as possible. Snake venoms also cause many human fatalities (> 90,000) each year [235]. Apart from these terrible consequences of snake envenomation, the toxins also represent an important natural source of potential lead structures for the treatment of various human diseases [9, 306]. The medical potential of snake venoms was already known in the ancient world, where it was used in small doses as a traditional medicine [6]. Currently there are six FDA-approved drugs on the market which are based on venom toxins [7] of which the majority originate from snakes. One of the most prominent example is captopril, an lifesaving anti-hypertensive drug derived from a bradykinin potentiating peptide from *Bothrops jararaca* [307, 308], a pit viper from the tropical and subtropical forests in southern Brazil, Paraguay and the North of Argentina [309]. Nevertheless, not only for the treatment of cardiovascular diseases snake toxins may be applicable, as there have been reported effects against neuropathic pain [310] and tumors [311]. Particularly non-toxic doses of snake venom have been shown to reduce the size of solid tumors as well as inhibition of tumor angiogenesis [312, 313].

In general, snake toxins target a wide range of cellular mechanisms, which are typically neural receptors, the cell membrane and the blood coagulation cascade [314]. Accordingly, the toxins have been grouped into neurotoxins, cytotoxins and hemotoxins composed of around 10 protein families (three-finger toxins; protease inhibitors; C-type lectins; vascular-, endothelial- and nerve growth factors; cysteine-rich secretory proteins (CRISPs), metallo-proteases (SVMP); serine-proteases (SVSP); L-amino acid oxidases (LAAO); acetylcholine esterases (AChEs) and nucleotidases) as well as peptides resulting from proteolytic cleavage (bradykinin-potentiating peptides, natriuretic peptides, disintegrins and SVMP-inhibitors) [314, 315].

Over the past decades, the analytical characterization of venom has been continuously enhanced by technological developments. The implementation of mass spectrometric techniques into proteomic workflows as well as next-generation sequencing, have advanced the analysis of snake venoms significantly. Since venom research evolved mainly from hypothesis-driven approaches, focused on single toxins, to a systematic, untargeted analysis, several so-called "bottom-up" approaches have been developed [2]. These techniques typically include a proteolytic digestion of the proteins to facilitate their analysis and identification [238]. The digestion step however multiplies the number of molecules present in a sample and disconnect structural relation within highly homologous isoforms and eventually occurring post-translational modified proteins. To overcome the sample complexation and loss of information by protease digestion we recently introduced a top-down mass spectrometric approach in an exemplary study of King Cobra venom [79]. However, even equipped with a set of toolboxes, including high throughput techniques capable of a rapid and systematic venom analysis, only few venoms from over 600 biomedically relevant snake species have been thoroughly studied [238, 239]. Furthermore, most of these studies only reflect a partial picture of the whole-venom of a certain snake species, since toxin structures and venom composition show considerable intraspecific and geographic variability [80, 235, 236].

The distribution of poisonous snakes in Turkey includes three families: *Colubridae, Viperidae* and *Elapidae*. Particularly, viper species are quite abundant and some species are endemic [316]. The Anatolian Meadow Viper, *Vipera anatolica*, which was first recognized as a differentiated taxon among the Euro-Asiatic vipers in 1983 [317] can only be found in a narrow and limited area near Elmali, in the Antalya province [318]. To the best of *our knowledge*, no studies on the venom composition of *V. anatolica* have been reported in *literature* so far. Thus, *V. anatolica* venom might be a source of untapped bioactive peptides and proteins with interesting pharmacological properties. As a part of our ongoing research on snake venoms from Turkish species, the purpose of this work was to profiling *V. anatolica* venom composition and to assess potential cytotoxic activities against different cancer cell lines.

Results and Discussion

Vipera anatolica

The Anatolian Meadow Viper is an viper species endemic in an area less than 100 km^2 east of Elmali, in Southwestern Anatolia, Turkey (shown in Figure 6.1B). The species is extremely rare and listed as critically endangered on the IUNC Redlist 2015 [319]. It took us several years and expeditions to finally capture three individuals (2 male and 1 female) in lengths between 20 and 40 cm. Figure 6.1 A shows a photograph of a female and a male individual in their natural habitat. Venom of each individual was extracted twice and freeze dried, which yielded a total amount of ~ 1 mg dried venom. The peptide and protein concentration was determined by means of a Bradford assay as ~38 % (wt%).

Figure 6.1 A: Photograph of *Vipera anatolica* – The photo was taken during a fieldtrip in October 2014 at Kohu Mt., Elmali, Antalya, Turkey. The individual on the right is an adult female; the second individual is an adult male. **B: Geographic distribution of *Vipera anatolica*.** The Anatolian Meadow Viper is an endemic viper species in an area less than 100 km^2 east of Elmali, in southwestern Anatolia, Turkey, blue circle.

In order to obtain a detailed picture of the venom composition and to assess the cytotoxicity of isolated toxins we performed a detailed proteomic analysis and fractionation of the venom. It is important to note that the limited access to the snakes, due to their low abundance and the minimal amounts of venom available for mass spectrometric analysis, made a classic venomic protocol (HPLC fractionation, SDS-PAGE, in gel trypsin digestion followed by MSMS *de novo* sequencing) aiming at a thorough proteomic characterization an extremely challenging task [2, 238, 309]. This is exemplified by the fact that the typical amount of venom applied to one semi-preparative HPLC run equals the amount of venom we had available to perform the complete analytical characterization. In order to obtain a general overview of the molecular masses of the venom components, including low

abundant and low molecular mass compounds which might not be gathered during SDS-PAGE, we thus decided to perform an initial top-down mass spectrometric analysis of the crude venom.

Top-down venomics

In top-down approaches the proteins are not digested prior to mass spectrometric analysis [320]. Instead they are directly ionized and thus render the molecular mass of the intact protein. Ideally, if the protein is further fragmented in the gas phase, amino acid sequences and site specific modifications can be determined. Therefore isotope resolution is typically required, in order to deconvolute the complex fragment spectra. This is limited by the resolution provided by the mass analyzer used, typically < 50 kDa for Orbitrap analyzers with R=100,000 at 400 m/z. As the majority of the toxins present in viper venoms exceed this molecular weight (> 50 kDa, SVMPs, SVSPs, LAAOs etc.), one can still make use of deconvolution based on charge distribution of protein MS1 spectra.

An inherent limitation of SDS-PAGE is that small sized peptides and proteins are not efficiently resolved which makes a direct HPLC-supported mass spectrometric analysis the most effective way to gather structural information [79]. In Figure 6.2 the UV chromatogram of the semi-preparative fractionation and the corresponding Coomassie-stained SDS-PAGE analysis of the fractions as well as the total ion current (TIC) of the LCMS analysis of native venom is shown. Due to the difference in column size the retention times are slightly shifted between the LCMS analysis and the semiprep-HPLC run. In the initial top-down mass profiling ~ 100 toxin masses were detected which are listed in Table 6.1 (All deconvoluted mass spectra are shown in the supplemental information). In total we observed 22 molecular masses < 1 kDa and 53 molecular masses between 1 and 9 kDa, which represents ~75% of the peptide and protein species. Furthermore the initial top-down analysis rendered another 27 compounds between 9 and 60 kDa. Interestingly, the majority of venom toxins seem to consist of molecular masses in the low mass range (< 9 kDa) which, at the same time occur in low concentrations. In order to obtain useful fragment spectra of toxins in subsequent MS2 experiments it was required to reduce potentially occurring disulfide bridges of toxins prior to top-down analysis. We thus reduced the venom chemically using TCEP prior to a second LCMS run with data dependent MS2 acquisition. Because a comprehensive analysis of top-down data relies strongly on the availability of transcriptome and/or genome data, which yet does not exist for *V. anatolica* and is still limited for closely related *Viperidae,* we manually analyzed the ESI mass spectra and searched *de novo* generated sequence tags against a *Viperidae* database using the Blast algorithm. Using this method we could identify a PLA$_2$ fragment with the sequence tag (1871)ALFSYSDYGCYCGWG(1931). As the spectrum was obtained from chemically reduced venom the observed fragmentation of the PLA2 could have occurred during the sample preparation before

mass spectrometric analysis. Furthermore we could directly identify a disintegrin like compound with the sequence tag (5085)-DYCTGJS-(221). Figure 6.3 exemplarily shows the deconvoluted top-down MS1 and MS2 spectra of the disintegrin-like venom protein with a molecular weight of 6047.5 Da (peak 11 in Figure 6.2).
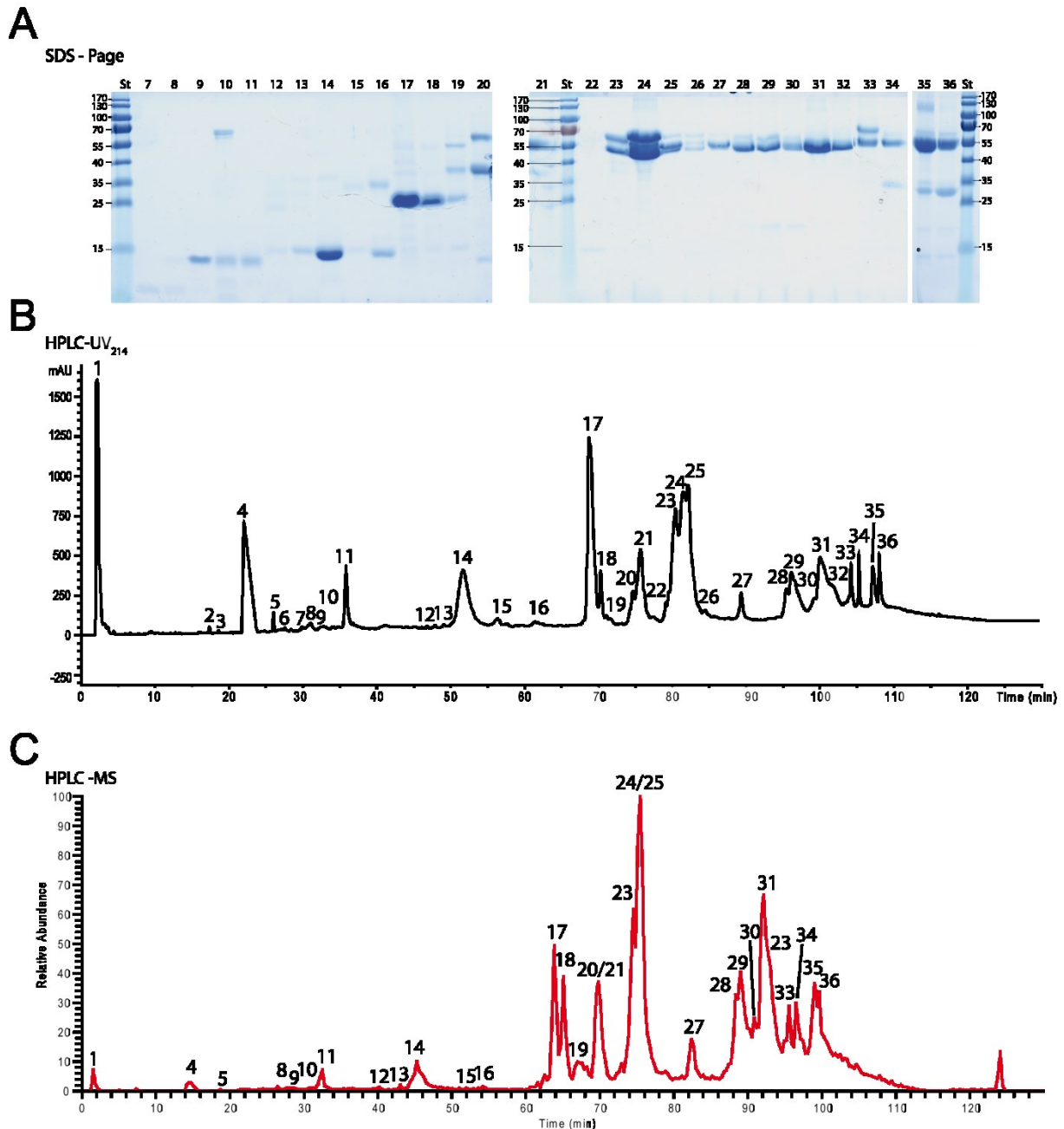
Figure 6.2 Venom profile of *Vipera anatolica*. Reverse-phase HPLC separation of *V. anatolica* venom proteins. A shows the SDS-PAGE analysis of fractions collected from the semi-preparative HPLC run displayed in B. Excised protein bands were identified by bottom-up *de novo* sequencing. Panel C shows the top-down MS1 analysis. [Molecular masses, sequence tags and database hits are listed in Table 2. Extracted and deconvoluted mass spectra are shown in the supplemental material. The deconvoluted MS1 and MS2 spectra of peak 11 are shown in Figure 6.4].

In the top-down MS analysis the non-reduced disintegrin like compounds (peak 11) instead showed average molecular masses of 13982.8, 14001.7 and 14017.7 Da which correspond to the SDS-PAGE analysis of the isolated peak with a band height at ~ 14 kDa and thus indicates that the disintegrin in peak 11 is a hetero-dimer. Furthermore peak 4 (see Figure 6.2) could be identified as a snake venom metalloprotease inhibitor, which is commonly present in *Viperidae* venoms [321, 322]. Nevertheless, through the difficulties in ionizing and deconvoluting high molecular mass components (> 50 kDa) which represent the majority of the *V. anatolica* venom content, the peptide/protein IDs obtained from the initial top-down analysis are limited. A future solution to increase the number of protein IDs obtained by top-down MS/MS would be a pre-fractionation by size exclusion chromatography and thus an enrichment of the low abundant low molecular mass toxins, which are accessible for top-down MS/MS elucidation but remained unselected during the information dependent acquisition (IDA) MS/MS experiments. A further general improvement for the venom proteome analysis would be a transcriptome analysis of *V. anatolica* venom gland tissue, which would enable the use of spectra-database comparisons for data analysis, which is more effective than the generation of *de novo* sequence tags and a subsequent Blast search.

Bottom-up venomics

Besides the initial top-down mass spectrometric characterization rendering intact venom components, which already showed a detailed picture of the structural diversity of the venom, we identified the remaining toxin families through the application of a bottom-up approach. By means of HPLC fractionation, SDS-PAGE separation followed by in-gel digestion and LC-MS/MS analysis we could generate 46 *de novo* sequence tags (Table 6.1), representing seven toxin families. The percentage of the venom protein content is presented in the pie chart in Figure 4. For determination of relative abundance (in %) of venom proteins the intensities of UV peak areas were used. In case of coelution the relative intensity of deconvoluted top-down spectra of proteins was taken as a measure. If an inefficient ionization of proteins was observed the signal strength of SDS-PAGE bands served as a measure. Hence, the most abundant toxin family is represented by snake venom metalloproteases (SVMP, 41.5 %), followed by two cysteine-rich secretory protein isoforms (CRISP, 15.9 %) with molecular masses of 24,645.0 Da and 24,544.1 Da, metalloprotease inhibitor (SVMPi, 9.3 %), A_2 phospholipases (PLA2, 8.1 %), disintegrins (2.0 %), a snake venom serine protease (SVSP, 1.6 %), a C-type lectin (1.1 %) and a Kunitz-type protease inhibitor (0.3 %).

.

Table 6.1 Venom peptides and proteins identified from *Vipera anatolica*. Protein assignment of RP-HPLC fractions (Figure 6.4) by LCMS and MS/MS analysis. Peak numbering corresponds to the UV and MS chromatograms shown in Figure 6.2. Sequence tags are obtained by *de novo* analysis of intact protein and tryptic peptide MS/MS spectra. Protein IDs are obtained by BLASTP analysis of the sequence tags against a viperid non-redundant protein database. Molecular weight was determined by SDS-PAGE (*) and top-down MS analysis (#) and are shown as average masses.

Peak number	Identified sequence tag	Protein ID	BLAST E-value	NCBI Accession Number	M [kDa]*	M [Da]#	Method
1		unknown (~peptide)				429.0; 621.0; 835.1; 1071.0; 1241.1	Top-down MS1
2		unknown (~peptide)				579.2; 816.4; 1269.6; 1600.8; 2049.0	Top-down MS1
3		unknown (~peptide)				969.4; 1026.5	Top-down MS1
4	pQKW	Metalloprotease inhibitor				443.2	De novo from Top-down MS2
5		unknown (~peptide)				474.2; 598.4; 752.4; 858.3; 1128.6; 1747.7	Top-down MS1
6		unknown (~peptide)				454.3; 499.4; 726.8	Top-down MS1
7		unknown (~peptide)				452.3; 606.3	Top-down MS1
8		unknown (~peptide)				808.4; 1128.6; 3943.8	Top-down MS1
9	(423.02)-YGGCGGNANNFK-COOH	~Kunitz-type serine protease inhibitor	2.0E-08	P0DKL8.1	14	6737.9	De novo from Trypsin-digest, Top-down MS1, Mass from SDS-Page
		unknown (~peptide)				452.3; 497.4; 707.4; 781.4; 3137.5; 4014.8; 4848.2; 5123.4	Top-down MS1
10		unknown (~Protein)			14; 56		Mass from SDS-Page
		unknown (~peptide)				1115.6; 3119.5; 3233.5	Top-down MS1
11	NH2-FJNAGTJCQY-(227.02)	~Disintegrin VA6	3.0E-04	P0C6A5.1	14	13982.8; 14001.7; 14017.7	De novo from Trypsin-digest, Top-down MS1, Mass from SDS-Page
	(532.05)-DYCDGJSSDGVDR-COOH	~Disintegrin VB7A	2.0E-02	P0C6A6.1			
	(1001.56)-YQTGJSSDCPR-COOH	~Disintegrin VB7B	6.0E-04	P0C6A7.1			
	(5085)-DYCTGJS-(221)	~Disintegrin VB7A	1.2E-02			6047.5 (reduced)	De novo from Top-down MS2

#	Sequence	Protein	E-value	Accession	SDS (kDa)	Mass	Method
12		unknown (~peptide)			14	1457.7; 6817.2	Top-down MS1; Mass from SDS-Page
13		unknown (~peptide)			14		Mass from SDS-Page
		unknown (~peptide)				1101.5; 1357.7; 6965.2; 7001.2; 7227.1	Top-down MS1
14	(241.03)-JCFGDQJNTYDK-COOH (1871)ALFSYSDYGCYCGWG(1931).	~Neutral phospholipase A2 ammodytin I2 ~ammodytin I2(C) isoform	1.0E-05 7.0E-12	P34180.2 CAE47236.1	14	13639.9 (5461.6)	De novo from Trypsin-digest, Top-down MS1, Top-down MS2,Mass from SDS-Page
		unknown (~Protein)				6948.3; 6982.2	Top-down MS1
15		unknown (~Protein)				7175.4; 7212.4; 7293.5	Top-down MS1
16a		unknown (~Protein)			35		Mass from SDS-Page
16b	(890.35)-VAAJCFGENJNS-(548.34)	~Neutral phospholipase A2 ammodytin I2	5.0E-06	P34180.2	14	13637.9	De novo from Trypsin-digest, Top-down MS1, Mass from SDS-Page
	(354.21)-CFGENJNTYDKK-COOH	~Neutral phospholipase A2 ammodytin I2	5.0E-09	P34180.2	14	7276.4; 7903.7	De novo from Trypsin-digest, Top-down MS1, Mass from SDS-Page
		unknown (~Protein)				3772.2; 5549.1; 5586.1; 5677.2; 5713.1	Top-down MS1
17	NH2-SVNPTASNMJK-COOH	~Cysteine-rich venom protein	5.0E-07	B7FDI0.1	25	24645.0	De novo from Trypsin-digest, Top-down MS1, Mass from SDS-Page
	NH2-SVDFDSESPR-COOH	~Cysteine-rich venom protein	3.0E-05	B7FDI0.1			
	NH2-FJDAYPEAAANAER-COOH	~Cysteine-rich venom protein	7.0E-06	B7FDI0.1			
	(225.09)-EJQNEEPDJHNSJR-COOH	~Cysteine-rich venom protein	2.0E-05	B7FDI0.1			
18	(439.24)-ECGENJYMSTSEVK-COOH	~Cysteine-rich venom protein	1.0E-05	B7FDI0.1	25	24544.1	De novo from Trypsin-digest, Top-down MS1, Mass from SDS-Page
19a	NH2-AYJGTMCQPK-COOH	~H3 metalloproteinase precursor 1	1.2E-02	AGL45259.1	55	46397.0	De novo from Trypsin-digest, Top-down MS1, Mass from SDS-Page
19b		unknown (~Protein)			35		Mass from SDS-Page
19c	(186.09)-DFTTESPR-COOH	~Cysteine-rich venom protein	8.6E-02	B7FDI0.1	25		De novo from Trypsin-digest, Top-down MS1, Mass from SDS-Page
19		unknown (~Protein)				1375.8; 6067.3; 6166.3	Top-down MS1
20a		unknown (~Protein)			60		Mass from SDS-Page

Table 6.1 Venom peptide and protein content of *Vipera anatolica* – *(continued)*

20b	NH2-JQGJVSWGS-(487.26)	~Snake venom serine protease nikobin	5.0E-05	E5AJX2.1	35			De novo from Trypsin-digest, Top-down MS1, Mass from SDS-Page
	NH2-JMGWGTJTTTK-COOH	~Snake venom serine protease nikobin	4.5E-02	E5AJX2.1				
	(699.24)-JJJEEWVJDQR-COOH	~Snake venom serine protease nikobin	8.9E-01	E5AJX2.1				
	NH2-FPNGDJKDLMJJR-COOH	~Snake venom serine protease nikobin	4.4E-01	E5AJX2.1				
20c	(1350.51)-JVCDGGDDPGTR-COOH	~ammodytin I1(A) variant (PLA2)	9.5E-02	CAE47176.1	14			De novo from Trypsin-digest, Mass from SDS-Page
21	NH2-TAJDFDGSVJGK-COOH	~metalloproteinase	5.7E-01	ADI47580.1	55			De novo from Trypsin-digest, Mass from SDS-Page
20/21		unknown (~Protein)				3181.2; 4688.4; 5505.0: 7376.8; 7833.8; 12891.8	Top-down MS1	
22		unknown (~Protein)				8172.7; 16346.4, 16397.1	Top-down MS1	
23a +24a	NH2-SSDJGFEDYSQJDR-COOH	~Zinc metalloproteinase-disintegrin-like ammodytagin	7.6E-01	P0DJE2.3	70			De novo from Trypsin-digest, Mass from SDS-Page
	(895.05)-JJTFDDSFGEWR-COOH	~Zinc metalloproteinase-disintegrin-like ammodytagin	5.3E-01	P0DJE2.3				
	NH2-JPECJJNKPJR-COOH	~Zinc metalloproteinase-disintegrin-like ammodytagin	4.4E-01	P0DJE2.3				
24b + 25	(373.93)-GDDTJDSFGEWR-COOH	~metalloproteinase F1	8.0E-05	AJC52543.1	60			De novo from Trypsin-digest, Mass from SDS-Page
	NH2-YDYSEDPDYGFGD-(449.1)	~Zinc metalloproteinase-disintegrin-like ammodytagin	1.6E-01	P0DJE2.3				
	NH2-VNJJNEMYJPLNJR-COOH	~H3 metalloproteinase	2.6E-01	AGL45259.1				
23b + 24c	NH2-VPJVGVEJWDHGDJJK-COOH	~H3 metalloproteinase	4.0E-03	AGL45259.1	45	48693.0		De novo from Trypsin-digest, Mass from SDS-Page, Top-down MS1
	NH2-VNJJNEFYLPJNJR-COOH	~H3 metalloproteinase	3.7E-01	AGL45259.1				
	(174.31)-NJFGEDYSQJDR-COOH	~H3 metalloproteinase	1.5E-02	AGL45259.1				
23-25		unknown (~Protein)				4452.7; 6299.4; 10102.4; 11831.3; 24411.4	Top-down MS1	
26	NH2-VPJVGVEJWDVPJTPR-COOH	~metalloproteinase H4-A	2.2E-02	AHB62069.1				De novo from Trypsin-digest

Table 6.1 Venom peptide and protein content of *Vipera anatolica* – (continued)

26		unknown (~Protein)				8594.1; 16106.5	Top-down MS1
27		unknown (~Protein)			50		Mass from SDS-Page
		unknown (~Protein)				6433.0; 9356.4; 16586.1	TIC Top-down MS1
29b - 30b	NH2-TWFNJNCEER-COOH	~C-type lectin	**1.0E-04**	**Q696W1.1**	20	21892.7	De novo from Trypsin-digest, Top-down MS1, Mass from SDS-Page
28 -29		unknown (~Protein)			50	49184; 24561.5	TIC Top-down MS1, Mass from SDS-Page
30a		unknown (~Protein)				48281;	TIC Top-down MS1
30-32		unknown (~Protein)				27058.0	TIC Top-down MS1
31	(388.99)-PDDPDYGFVDJGTK-COOH	~Zinc metalloproteinase-disintegrin-like ammodytagin	**1.0E-08**	**P0DJE2.3**	50		De novo from Trypsin-digest, Mass from SDS-Page
	NH2-TLGJAPVSGMCQPK-COOH	~metalloproteinase H4-A	**3.0E-06**	**AHB62069.1**			
	NH2-HDNAQJJTGJDJN-(213.07)	~metalloproteinase H4-A	**7.0E-08**	**AHB62069.1**			
32	NH2-MTJEGFJTGLDJNGR-COOH	~Zinc metalloproteinase-disintegrin-like ammodytagin	**2.1E-01**	**P0DJE2.3**	50		De novo from Trypsin-digest, Mass from SDS-Page
	NH2-DVGJAPVSGMCQPK-COOH	~metalloproteinase H4-A	**2.0E-04**	**AHB62069.1**			
	(323.01)-ATSEQQR-COOH	~metalloproteinase H4-A	**1.1E-01**	**AHB62069.1**			
33a	(260.16)-DHJDJJCJJNQPJR-COOH	~Zinc metalloproteinase-disintegrin jararin	**1.5E-02**	**Q0NZX6.1**	65	32076.0	De novo from Trypsin-digest, Mass from SDS-Page, Top-down MS1
	(626.8)-NPESSJJNQPJR-COOH	~metalloproteinase	**1.6E-01**	**ADI47709.1**			
33b	NH2-HDNAQJJTGJD-(440.24)	~Zinc metalloproteinase-disintegrin BlatH1	**8.0E-05**	**U5PZ28.1**	50	57085.0	De novo from Trypsin-digest, Mass from SDS-Page, Top-down MS1
	NH2-VTSSGDDTJDSFGEGER-COOH	~metalloproteinase	**5.0E-07**	**ADI47715.1**			
	NH2-YYJNEMYJPJNJR-COOH	~metalloproteinase	**2.0E-05**	**ADI47590.1**			
34a		unknown (~Protein)			60	26398.0	TIC Top-down MS1, Mass from SDS-Page
34b		unknown (~Protein)			35	27603.0	TIC Top-down MS1, Mass from SDS-Page

88

Table 6.1 Venom peptide and protein content of *Vipera anatolica* – (continued)

35a	(493.29)-VSWGSCAQK-COOH	~Thrombin-like enzyme gyroxin B2.1	**6.6E-02**	**Q58G94.1**	170		De novo from Trypsin-digest
35b + 36a	NH2-JJEWSER-COOH	~metalloproteinase F1	**1.7E-02**	**AJC52543.1**	50	25225.0	De novo from Trypsin-digest, Mass from SDS-Page, Top-down MS1
	(37.19)-QJYYTPR-COOH	~H3 metalloproteinase	**7.0E-03**	**AGL45259.1**			
	(339.04)-JTPEEKR-COOH	~H3 metalloproteinase	**4.9E-01**	**AGL45259.1**			
36a		unknown (~Protein)			27		TIC Top-down MS1, Mass from SDS-Page
36b		unknown (~Protein)					TIC Top-down MS1, Mass from SDS-Page

As the first line of quantification relies on the UV signal from the HPLC run, the molar extinction coefficient ε of the peptides plays a crucial role for the peak areas of the toxins. Thus, our quantification is based on the simple assumption that these compound-specific values are nearly equal. This indeed might be the case for compounds of the same protein family but is rather unlikely for the comparison of peptides with high molecular mass proteins. Especially the concentration of the SVMPi, containing a Trp residue at the C-terminus, might be overestimated as ε-value is most likely higher than of the other compounds. Thus the venom composition has to be considered as semi-quantitative.

Medical implications

In comparison with other viper species, effects through envenomation caused by *V. anatolica* are generally low and not life threatening. After two accidental bite cases with *V. anatolica* during our field studies, the subjects (healthy, 51 and 28 years old, both male) reported only local pain and itch at the bitten area without any other complications. Similar symptoms were reported by Krecsak *et al.* [323] for the closely related small-sized insectivorous snakes viper *Vipera* (*Acridophaga*) *ursinii*. On the other hand, in a case study about accidents with another viper species to be found in Turkey, the blunt-nosed viper, *Macrovipera lebetina obtusa*, strong clinical symptoms and physiological damage were reported [324]. The difference in the toxicity can be explained by the bigger size and higher amount of venom injected by *Macrovipera lebetina obtusa*, but also the higher abundance of PLA$_2$ (~ 34 % vs. ~ 8 % *V. anatolica*) [324] might be a reason for the stronger necrotic effects as well as swelling-associated symptoms.

Cytotoxicity screening

To assess the potential of the toxins against cancer cells, we assessed the cytotoxicity of crude venom against the following cell lines: CACO-2, human colon carcinoma epithelial cells; MCF-7, human breast adenocarcinoma epithelial cells; U87MG, human glioblastoma-astrocytoma epithelial-like cells; PC3, human prostate epithelial cells; HeLa, human cervical epithelial carcinoma cells; MPanc-96, human pancreatic fibroblast cells; A549, human lung epithelial cells; HEK293, human embryonic epithelial kidney cell; Vero, African green monkey fibroblast-like kidney cells. We therefore performed initial cytotoxicity assays by means of an MTT assay which measures the mitochondrial reductase activity of cells and therefore is a measure for viability of cells. In Figure 6.5 it can be seen that crude venom of the Anatolian Meadow Viper inhibits cell viability in a dose-dependent manner. The IC$_{50}$ values for all affected cell lines are shown in table 6.2. MCF-7, MPanc-96

and A549 were herby the most resistant cell lines against all venom doses tested (IC_{50} > 50 µg/mL), while PC3, HEK293, CACO-2, Vero and HeLa cells were only moderately affected (IC_{50} ~ 4.9; 5.5 and 7.2; 11.7 and 13.2 µg/mL respectively). The highest activity was observed against brain cancer cells (U87MG) with an IC_{50} value of ~ 0.8 µg/mL, which is more than one order of magnitude lower than against the non-cancerogenous cell lines (HEK293 and Vero). Remarkably, the crude venom cytotoxic effects on glioblastoma cells as well as the selectivity to cancer cells are notably higher than of the plant-derived sesquiterpene lactone parthenolide, a promising candidate in anti-cancer research. Microphotography of treated cells (shown in Appendix D) showed similar results in comparison to the MTT assay.
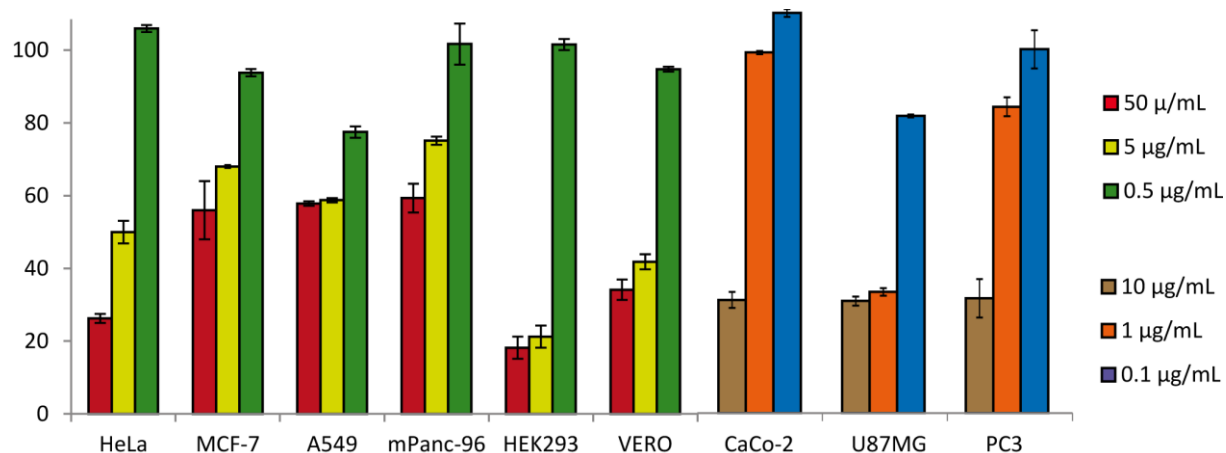
Figure 6.5. Viability of cancer and non-cancerous cell lines after crude venom treatment for 48 h. Cell viability was determined by MTT assay, control was exposed to vehicle only which was taken as 100 % viability. CACO-2, human colon carcinoma epithelial cells; MCF-7, human breast denocarcinoma epithelial cells; U87MG, human glioblastoma-astrocytoma epithelial-like cells; PC3, human prostate epithelial cells; HeLa, human cervical epithelial carcinoma cells; MPanc-96, human pancreatic fibroblast cells; A549, human lung epithelial cells; HEK293, human embryonic epithelial kidney cell; Vero, African green monkey fibroblast-like kidney cells.

Untreated cells were homogeneously distributed in the culture plates showing typical morphological characteristics. Whereas morphological changes were observed after treatment with crude venom for 48 h varied depending on the origin of the cell lines. Increasing venom concentrations resulted in a higher number of rounded, detached cells, elevated mobility and distribution of cells on the surface, multicellular aggregate formation and growth inhibition compared to untreated cells. In addition, cell disorganization and large areas without cells were observed with increasing venom concentrations. Because the amount of fractionated venom generally were very low (0.04-9.03 µg) the isolated toxins were only tested against the most sensitive cell line (U87MG) of the crude venom cytotoxicity results. The results for peak 11 collected by semipreparative HPLC are shown in Figure 6.6.

Table 6.2. IC$_{50}$ values of *V. anatolica* crude venom and parthenolide treated cell lines.
CACO-2, human colon carcinoma epithelial cells; MCF-7, human breast adenocarcinoma epithelial cells; U87MG, human glioblastoma-astrocytoma epithelial-like cells; PC3, human prostate epithelial cells; HeLa, human cervical epithelial carcinoma cells; MPanc-96, human pancreatic fibroblast cells; A549, human lung epithelial cells; HEK293, human embryonic epithelial kidney cell; Vero, African green monkey fibroblast-like kidney cells.

Cell lines	*V. anatolica* venom [µg/mL]	Parthenolide [µg/mL]
CACO-2	7.21±0.12	0.69±0.08
MCF-7	> 50	1.17±0.06
U87MG	0.75±0.010	1.93±0.05
PC3	4.85±0.18	1.28±0.07
HeLa	13.15±0.13	3.33±0.12
MPanc-96	> 50	0.58±0.02
A549	> 50	1.66±0.09
HEK293	5.52±0.21	0.63±0.04
Vero	11.65±0.16	2.58±0.10

Hereby, the only active fraction (Peak 11), containing a dimeric disintegrin, had a significant cytotoxic effect on glioblastoma cells with an IC$_{50}$ value of 0.51±0.04 µg/ml which is slightly better than the IC$_{50}$ of crude venom (0.75 µg/mL). However comparing the fractions activity with the activity of parthenolide we found a twofold higher cytotoxicity. If the molecular mass ratio between parthenolide and disintegrin (248 Da: ~14.0 kDa) are considered, then the molar activity of peak 11 (~ 36 nM) is around 100 times higher than that of parthenolide. In comparison to the cytotoxic effects of recombinant disintegrins, r-viridistatin 2 and r-mojastin 1, peak 11 was significantly more active against BXPC-3 pancreatic cells with IC$_{50}$ value of 10.6 and 8.7 µM [325]. The disintegrin colombistatin, isolated from *Bothrops colombien*sis, on the other hand showed similar IC$_{50}$ against T24 and SK-Mel-28 cells (4.4 µM and 33 nM) [326]. Cytotoxic activity of disintegrins against glioblastoma, has been described to be caused by targeting specific integrins ultimately resulting in decrease of tumor growth, affecting invasion and migration of tumor cells in carcinogenesis [327, 328]. Cell adhesion and migration are important stages in metastasis development in which integrins, a class of receptors that modulate cell attachment, cell-cell and cell-extracellular matrix interactions, are significantly involved [328]. Here, the molecular target could be the RGD (Arg-Gly-Asp)-binding α5β1 and αvβ3 integrins which are known to be important for single cell migration and highly

expressed in U87MG cells [329, 330]. Thus, snake venom and particularly the dimeric disintegrin isolated from *V. anatolica* might have a promising potential as therapeutic anti-cancer agents.
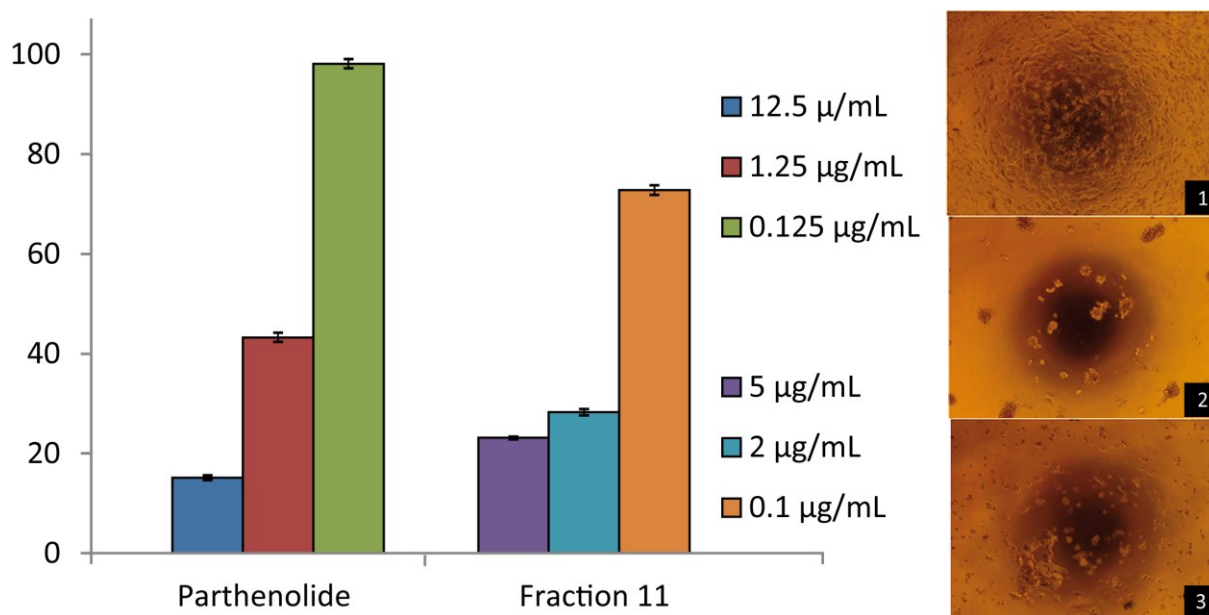
Figure 6.6. A: Viability of U87MG cells after treatment with fraction 11 for 48 h. Cell viability was determined in an MTT assay, normalized to a control (100 % viability). **B: Effect of the isolated disintegrin on U87MG cells**. Cells were treated with fractionated venom (fraction 11) for 48 h at 37°C. 1: untreated, 2: treated with peak 11 2 µg/ml, 3: treated with parthenolide, 1.25 µg/ml.

To further investigate the anti-cancer properties of the herein described disintegrin (peak 11) full protein sequencing is needed in order to establish its primary sequence. Due to the low abundance and difficulties obtaining sufficient amounts of venom, the best approach would be cDNA cloning of the venom gland mRNA, which would enable us to heterologously express and produce the toxin. A transcriptomic analysis of the venom gland tissue would also facilitate the interpretation of mass spectra and thus allow us to obtain deeper insights into the complexity of *Vipera anatolica's* venom.

Conclusions

In conclusion, this study is an important step towards a complete profiling of the venom of *Vipera anatolica*. Based on a minimal amount of venom, we applied a combination of top-down and bottom-up mass spectrometry, which enabled us to detect ~110 venom components of which 100 different molecular masses were detected by intact protein mass spectrometry and ~10 exclusively by SDS-PAGE. Out of 36 isolated fractions we could identify 8 different toxin families. By means of MS/MS *de novo* sequencing of tryptic peptides as well as of intact proteins and peptides we could

identify a snake venom metalloprotease inhibitor; a Kunitz-type protease inhibitor; disintegrins; phospholipases A_2; cysteine-rich secretory proteins; snake venom metalloproteinase; snake venom serine proteases, and a C-type lectin. The application of top-down mass spectrometry facilitated the detection of considerably more venom components as we were able to detect by a bottom-up approach. Even though a full protein identification by top-down MS/MS remains a challenging task for high molecular mass venom compounds, e.g. for metalloproteases, the mass profiling of intact toxins significantly improves the characterization of the complexity of snake venoms.

During our search for pharmacological interesting compounds we identified a dimeric disintegrin which showed remarkable cytotoxic activity against U87MG cells, influencing the cell shape, survival and proliferation and migration, which will be further assessed in our ongoing studies.

Experimental

Collection and preparation of venom samples

The study was approved by the Ege University, Local Ethical Committee of Animal Experiment (Number: 2010/43) and a special permission (2011/7110) for field studies was accepted from the Republic of Turkey, Ministry of Forestry and Water Affairs. All specimens were released back into their natural environment after investigation.

V. anatolica individuals were collected from late April to mid-October 2014 in Cığlıkara forests, North West of Kohu Mt. in Elmalı, Antalya provinces in Turkey, in altitudes between 1650 and 1750 m (MASL). Crude *V. anatolica* venom was extracted from two male and one female adult, using a paraffin-covered laboratory beaker without exerting pressure on the venom glands. Venom samples were pooled in one tube and centrifuged at 2,000 × g for 10 min at + 4°C to remove cell debris. Supernatants were collected, immediately frozen at -80°C, and then lyophilized. Lyophilized samples were stored at 4°C.

Determination of Protein Concentration

Protein concentration was determined from diluted venom sample (4 mg/mL) in deionized water by Bradford assay[331] using a UV/Vis spectrophotometer (VersaMax, Molecular Devices, CA, USA) at a wavelength of λ = 595 nm. Bovine serum albumin was used as a reference.

Cell Culture and in vitro Cytotoxicity Assay

The following cell lines were used for determination of cytotoxicity: HeLa (human cervix adenocarcinoma), A-549 (human alveolar adenocarcinoma), MCF-7 (human breast adenocarcinoma), CACO-2 (human colon colorectal adenocarcinoma), mPANC96 (human pancreas adenocarcinoma), PC-3 (human prostate adenocarcinoma), U87MG (human glioblastoma-astrocytoma) cancer cells and as a non-cancerous cell lines, HEK (human embryonic kidney). Cell lines were purchased from ATCC (Manassas, VA, USA). The PC3 cell line was obtained from Dr. K. Korkmaz (Ege University, Bioengineering Department, Bornova-Izmir, Turkey). 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 µg/mL of streptomycin (Lonza, Visp, Switzerland). The cells were incubated at 37°C in a humidified atmosphere of 5 % CO_2.

Cytotoxicity of crude venom was determined by following the general procedure based on cell viability using a modified colorimetric MTT [3-(4,5- dimethyl-2-thiazolyl)-2,5-diphenyl-2H-tetrazolium bromide)] assay [332, 333]. The optical density (OD) was measured in triplicates at λ = 570 nm (with a reference wavelength λ = 690 nm) by UV/Vis spectrophotometry (Thermo, Bremen, Germany). All cell lines were cultivated for 24 h in 96-well microplates with an initial concentration of 1 x 10^5 cells/mL. Subsequently, the cultured cells were treated with different doses of venom and incubated for 48 h at 37 ºC. The plant-derived compound parthenolide (a sesquiterpene lactone) was used as a positive cytotoxic control agent. Percentages of surviving cells in each culture were determined after incubation with venom. The viability (%) was determined by the following formula:

%Viable cells= [(absorbance of treated cells) – (absorbance of blank)] / [(absorbance of control) – (absorbance of blank)] x 100

Determination of half maximal inhibitory concentration (IC50)

In cell culture studies for untreated cell lines (negative controls) cytotoxicity was set to 0 %. 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 untreated controls.

Morphological Studies

The morphological studies of the cells were performed with an inverted microscope (Olympus, Tokyo, Japan) compared to the control group 48 h after treatment.

Preparation of venom samples for proteomic analysis

For LC-MS analysis and semi-preparative HPLC fractionation, crude venom was dissolved in aqueous 1% formic acid (HFo) (v/v) to a final concentration of 10 mg/mL, and centrifuged at 20,000 g for 5 min. For top-down measurements disulfide bonds were reduced. Therefore 10 µL of venom (10 mg/mL) was mixed with 10 µL of tris(2-carboxyethyl)phosphine (TCEP, 0.5 M) and 30 µL of citrate buffer (0.1 M, pH 3), after which the reaction mixture was incubated for 30 min at 65 °C. Thereafter, the sample was mixed with an equal volume of 1% formic acid and centrifuged at 20,000 x g for 5 min. Then 10 µL of both reduced and non-reduced samples were submitted to top-down measurements. A volume of 100 µL of non-reduced venom was subjected for semi-preparative HPLC fractionation.

Top-down venomics

LC-ESI-HR-MS/MS experiments were performed on an LTQ Orbitrap XL mass spectrometer (Thermo, Bremen, Germany) coupled to an Agilent 1260 HPLC system (Agilent, Waldbronn, Germany). A Suppelco Discovery 300Å C18 (2 x 150 mm, 3 µm particle size) column was used. Typically the flow rate was set to 0.3 mL/min and a gradient of 0.1% HFo in water (solution A) and 0.1% HFo in acetonitrile (ACN) (solution B) was used. The gradient started isocratically (5% B) for 1 min, followed by an increase from 5 to 40% B over 40 min, 40-70% over 20 min, a washout at 70% B for 10 min, and a re-equilibration phase at 5% B. ESI settings were 40 L/min sheath gas; 20 L/min auxiliary gas; spray voltage, 4.8 kV; capillary voltage, 46 V; tube lens voltage, 135 V and capillary temperature, 330 °C. For information dependent acquisition (IDA), four scan events were set with 2 micro scans and 500 ms maximal fill time. The survey scan was performed with mass resolution (R = 100,000 at m/z 400). For MS/MS R was set to 60,000 (at m/z 400). Every cycle contained two CID scans of the two most abundant ions of the survey scan. Normalized collision energy was set to 35% for CID. The default charge state was set to 8+ and the activation time to 30 msec. The precursor selection window was set to 2 m/z. Dynamic exclusion was performed with a 3 m/z exclusion window for precursor ions with 1 repeat within 10 s. The exclusion list contained maximal 50 ions for a duration of 20 s. For the deconvolution of isotopically resolved spectra the XTRACT algorithm of Xcalibur (Thermo, Bremen, Germany) was used. For isotopically not resolved spectra the Zscore algorithm [302] implemented in the <u>mag</u>ic <u>tran</u>sformer (MagTran) tool was used for charge distribution deconvolution. For protein identifications, deconvoluted mass spectra were analyzed manually. The resulting *de novo* generated

sequence tags were BLASTed against the NCBI non-redundant Viperidae database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) using the BLASTP algorithm [209].

Bottom-up venomics

For bottom-up analysis, the venom was dissolved to a concentration of 10 mg/mL in aqueous 1% HFo and 5% acetonitrile (ACN). A volume of 100 μL was subjected to semipreparative reverse-phase (RP) HPLC separation on a Suppelco Discovery 300Å C18 (4.6 x 150 mm, 3 μm particle size) column. The flow rate was set to 1 mL/min. A linear gradient of 0.1% HFo in water (solution A) and 0.1% HFo in ACN (solution B), isocratically (5% B) for 5 min, followed by linear gradients of 5-40% B for 95 min, 40-70% for 20 min, 70% B for 10 min, and finally re-equilibration at 5% B for 10 min was used. Peak detection was performed at λ = 214 nm using a diode array detector (DAD). Chromatographic fractions were collected manually, dried in a vacuum centrifuge, chemically reduced with dithiothreitol (DTT) and submitted to SDS-PAGE. Subsequent Coomassie stained bands were excised from the gel and subjected to in-gel reduction (10 mM DTT in 25 mM $(NH_4)HCO_3$, pH 8.3, for 45 min at 65 ºC) and alkylation (50 mM iodacetamide in 50 mM $(NH_4)HCO_3$, pH 8.3, for 30 min at 25 C), followed by in-gel trypsin digestion (12 h at 37 ºC with 66 ng sequencing-grade trypsin/mL in 25 mM $(NH_4)HCO_3$, 10% ACN; 0.25 mg/sample). Tryptic peptides were dried in a vacuum centrifuge, re-dissolved in 15 mL of 5% ACN containing 0.1% HFo, and submitted to LC-MS/MS analysis using a an Orbitrap XL hybrid mass spectrometer (Thermo, Bremen, Germany) equipped with an HPLC system (Agilent, Waldbronn, Germany). LC separation was performed on a Grace Vydac 218MSC18 column (2.1 x 15 mm, 5μm) with a flow rate of 0.3 mL/min. A gradient was applied using of 0.1% HFo in water (solution A) and in ACN (solution B). The gradient started isocratically with 5% B for 2 min, followed by an increase over 10 min from 5-40% B, 40-99% B over 15 min, and held at 99% B for 5 min with final re-equilibration phase at 5 % B for 5 min. MS experiments were performed in the Orbitrap analyser with R=15,000 at m/z 400 and maximum filling time of 200 ms for both survey and first product ion scans. MS/MS fragmentation of the two most intense ions was performed in the LTQ using CID (30 ms activation time); the collision energy was set to 35%. Precursor-ion isolation was performed within a mass window of 2 m/z. Dynamic exclusion was set up for a 3 m/z windows for up to 50 precursor ions with a repeat of 2 within 30 s. *De novo* annotation of MSMS spectra was performed with the DeNovoGUI tool [304]. Manually proved sequence tags were searched against a Viperidae non-redundant protein database of UniProtKB/TrEMBL using BLASTP [209].

Relative toxin quantification

The relative abundances (percentage of the total venom proteins) of the different protein families were calculated as the ratio of the sum of the areas of the reverse-phase UV_{214}-chromatographic

peaks containing proteins from the same family to the total area of venom protein peaks in the reverse-phase chromatogram according to Calvete *et al.* [2]. If more than one protein was present in a reverse-phase fraction, their proportions were estimated by the relative abundance of deconvoluted top-down spectra. If co-eluting proteins were observed by SDS-PAGE which were not accessible by mass spectrometry, their relative abundance was estimated by optical signal strength of the Coomassie-stained bands [309].

7 Synopsis and Future Perspectives

Natural products as well as the toxins among them play a fundamental role in human health and medicine, since they are extensively used as antibiotics, antifungals, immune-suppressants and anticancer agents [334]. Advances in sequencing technologies and bioinformatics approaches brought us to a tipping point in natural products discovery as they revealed an enormous reservoir of novel compounds in microbial ecosystems [39] and higher organisms such as snakes [80]. The role of microbial natural products as mediators of diseases is discussed in an upcoming article [335].

A prominent example of a disease induced by a natural product is sugarcane leaf-scald caused by *Xanthomonas albilineans;* which invades the xylem of sugarcane. Symptoms of this disease vary from white, sharply defined pencil stripe to complete necrosis of infected leaves leading to plant death [59]. In 2004 the genome of *Xanthomonas albilineans* was sequenced [56] and revealed the biosynthesis genes of several interesting natural products including the phytotoxin albicidin [56, 165]. Besides the interesting biosynthetic gene clusters (BGC), *Xanthomonas albilineans* lacks important genes involved in pathogenicity, for example *gum* genes, T3SS Hrp and T6SS, which are required for survival, growth and spread within host plants [166]. The unusual biosynthetic gene clusters, paired with the genomic erosion, make *X. albilineans* quite unique amongst the genus *Xanthomonas*, which we described as a xanthomonad on the edge [166].

As described in Chapter 2, one of the mediators of sugarcane leaf scald is the phytotoxin albicidin which has been already described in 1983 [58]. Albicidin blocks the DNA gyrase in sugarcane chloroplasts and thus induces chlorosis which ultimately leads to the death of the plants [196, 197]. Besides the dire consequence for the plants and the humans dependent on their cultivation, albicidin is a very promising lead structure in anti-infective research. After more than 30 years since its discovery by Birch *et al*. [55], our group finally unraveled its unusual chemical structure by a combination of chemical derivatizations and stable isotope feeding combined with high resolution mass spectrometry and NMR, shown in Chapter 3 of this thesis. With the knowledge of the structure, we analyzed the NRPS amino acid sequences through multiple alignments and expressed all

adenylation domains (A-domains), responsible for the substrate activation of every NRPS module, as thioredoxin-His6 fusion proteins. With all A-domains we performed *in vitro* substrate screening on which we finally based a comprehensive biosynthetic model shown in figure 3.3. Through the bioinformatic alignment to known A-domains we furthermore defined a new NRPS code for aromatic δ-amino acids in which the negatively charged residue (Asp) is relocated to the end of the binding pocket, shown in Appendix A, supplemental figure 8. These findings also match with the substrate-conferring amino acid code of the cystobactamide NRPS, which was discovered most recently [63].

Ironically, the desperate search for chemically novel, unusual moieties and highly bioactive compounds are the elemental reasons why the structure elucidation of albicidin took so long. First of all, as albicidin consists exclusively of unusual building blocks, the structure elucidation by dereplication approaches like spectral library search was not possible, since spectra of this class of non-ribosomally synthesized peptides did not exist before in the databases. Secondly, through the high bioactivity, the cells need to produce only minor amounts of albicidin to inhibit their target (competing cells in the plant), which in turn makes it difficult to gain enough compound for extensive NMR experiments and the *de novo* structure elucidation. However, the generation of a heterologous producer strain finally enabled us to produce enough albicidin for NMR experiments. An MS based approach would have needed much less amounts and thus less time and especially less resources [133]. With the structure in hand, the MS/MS structure elucidation of further albicidin derivatives, like carbamoyl-albicidin, described in Chapter, was much easier and faster. To find more and especially minor abundant derivatives of albicidin we performed most lately untargeted MS/MS screening of *Xanthomonas albilineans* extracts. To identify related molecules we performed molecular networking of bioactive HPLC fractions using gnps (gnps.ucsd.edu). Besides albicidin, carbamoyl-albicidin and other known derivatives [336] we found additional, till now unknown analogues. The most interesting example is an albicidin derivatives which has only one core substitution on each of the two *p*ABA moieties at the C-terminus [337]. This finding perfectly supports our biosynthesis hypothesis of the incorporation of a mono-hydroxylated *p*ABA building block through the NRPS Alb09 and further *in situ* methylation and a second hydroxylation, shown in figure 3.3. Meanwhile we could archive further evidence for the *in situ* methylation through the heterologous production of the methyltransferase Alb02 as a thioredoxin-His6 fusion protein and *in vitro* methylation assays. We tested several potential substrates as for example 3-hydroxy-para-amino-benzoic acid (3-OH-*p*ABA) and a corresponding *N*-acetylcysteamine (SNAc) linked derivative, which mimics the CoA moiety of the T-domain. The SNAc linked 3-OH-*p*ABA showed herby significant higher turnover rates than the free amino acid, indicating that the methylation most likely occurs while 3-OH-*p*ABA is bound at the NRPS, which is in line with our hypothesis published before [338].

100

Most lately we were able to show the methylation of 3-OH-*p*ABA-CoA bound to a heterologous produced peptide carrier domain (PCP-domain) of the NRPS Alb09 by intact protein mass spectrometry [338].

Interestingly, the structure of an albicidin homologue cystobactamide, found by Baumann *et al.* [63] has only one substitution at the C-terminal *p*ABA moieties as well; which gives further evidence that the second hydroxylation must occur after the incorporation of 3OH-*p*ABA through the NRPS. These findings indicate that an essential part of albicidin biosynthesis occurs either *in situ* at the NRPS or at least post-NRPS as for example the formation of cyano-alanine at an novel module at the *trans* acting NRPS Alb04, or the carbamoylation described in chapter 4. The carbamoylation of albicidin increases thereby the potency of gyrase inhibition, which indicates that the tailoring reaction matures albicidin in terms of bioactivity. Besides the biology relevance, the novel derivatives and their assessment in terms of bioactivity are furthermore important for our ongoing drug development of synthetic albicidin derivatives.

With the knowledge we achieved through the bioinformatic and biochemical investigation of the albicidin biosynthesis and thanks to the rise of genomes sequenced, we are currently searching for other NRPS gene clusters containing A-domains with the unusual substrate conferring code for *p*ABA moieties. We will therefore designed a bioinformatic tool which automatically downloads novel sequence information from the NCBI database and performs genome mining including the annotation of the NRPS codes locally on our server with antismash 3.0 [339]. An interesting gene cluster containing a putative *p*ABA or *p*ABA derivative activating A-domain, we found already in the mango pathogen *Xanthomonas citri. pv. mangiferaendicae* by "manual" genome mining using the antismash online tool. In order to identify *p*ABA containing natural products in *Xanthomonas citri pv. mangiferaendicae*, we performed differential feeding experiments were we supplemented liquid culturing media with *p*ABA in comparison to cultures fed with deuterated *p*ABA (D_4-pABA). Through untargeted LC-MS/MS experiment and dereplication via spectral networking we were able to identify several compounds with 3 Da mass shifts, indicating that these molecules must contain a *mono*-substituted *p*ABA moiety. Through further purification of the most abundant compound and corresponding NMR analysis, we elucidated the structure of a novel *p*ABA containing meroterpenoid, that we named xantomonic acid. Even though the compound is not peptidic and thus most likely does not belong to the NRPS we identified through genome mining, the compound showed a very interesting anti-cancer activity. The results were recently submitted as an article to the *Journal of Natural Products* [150].

The second part of this thesis (Chapters 5 and 6) are dedicated to the development of top-down mass spectrometric methods for profiling of snake venoms. Thanks to recent developments in mass spectrometry, including the introduction of the orbitrap mass analyzer, novel fragmentation methods such as electron transfer dissociation (ETD) and bioinformatic tools, top-down mass spectrometry is increasingly gaining momentum in the analysis of proteins or whole proteome analysis [125, 126, 281-287]. In Chapter 5, the first application of top-down venom analysis of snake venom, termed top-down venomics, is evaluated on the example of king cobra. We could identify several three finger toxins, Kunitz-type protease inhibitors, A2-phospholipases and cysteine rich secretory proteins including unknown isoforms, straight in the gas phase. The results clearly show the great opportunity of integrating this approach to achieve locus-specific resolution in snake venomics. Top-down venomics, alone or combined with bottom-up MS/MS, represent the method of choice for unravelling the toxin composition of snakes, especially those of terrestrial and marine elapids [273, 288-290]. Besides snakes, top-down venomics will be of value in the analysis of venoms from other creatures, such as arthropod, hymenopteran, and cone snail [79]. Venom peptides are most often stabilized by multiple intramolecular disulfide bonds that impart high levels of chemical, thermal and biological stability [229, 300], but may also cause serious analytical problems to bottom-up approaches. Efficient separation of the venom components, followed by a top-down analysis with species-specific database search would solve this analytical limitation. Furthermore, top-down venomics dramatically facilitates the experimental workflow and would enable a high-throughput venom analysis.

In Chapter 6, the first application of top-down venomics for the analysis of viper venom is presented. This work is also an important step towards a complete profiling of the venom of *Vipera anatolica.* Through the combination of top-down and bottom-up analysis, we were able to detect ~110 venom components, of which we could identify 8 different toxin families. Through *de novo* sequencing of tryptic peptides as well as of intact proteins we could identify a snake venom metalloprotease-inhibitor, a Kunitz-type protease inhibitor, disintegrins, phospholipases A_2, cysteine-rich secretory proteins, snake venom metalloproteinase, snake venom serine proteases, and a C-type lectin. The employment of top-down mass spectrometry pushed hereby the detection of considerably more venom components as we were able to detect only by a *bottom-up* approach. Also, full protein identification by top-down MS/MS remains a challenging task for high molecular mass venom compounds of vipers (metalloproteases, serine proteases, L-amino acid oxidases, etc.) the mass profiling of intact toxins significantly improves the characterization of the complexity of snake venoms. In the bioactivity screening of isolated fractions of *Vipera anatolica*, we identified a dimeric disintegrin which showed remarkable cytotoxic activity against U87MG cancer cells, influencing the

cell shape, survival, proliferation and migration. This is being further assessed in an ongoing study. Currently we are applying top-down venomics to several venoms of other species including cone snails, several Turkish vipers we collected during a two weeks field trip in north-eastern Turkey and four different mamba species. Analyzing the mamba venoms, we could show for the first time that some neuro-toxins may be acetylated or tri-methylated at various lysine residues. This could be caused as a regulatory effect, as the positive charges of the lysine side chains in for example dentrotoxin K are important for target binding in K^+ ion channels [340]. These results have been recently submitted as an article to the journal *Molecular and Cellular Proteomics* [341].

In summary, mass spectrometry techniques improved massively for the search and structural elucidation of natural products, in particular peptidic toxins. Easy sample preparation, rapid measurement and high sensitivity as well as selectivity are crucial for high-throughput profiling of complex mixtures such as microbial and plant metabolomes or venom secretions of animals. With constant improvement of mass spectrometers, nucleotide sequencing techniques and bioinformatic tools, integrating this information, the bright future of mass spectrometry and its central role in the discovery of nature's complexity is more than obvious [64, 79, 133, 134].

A   Appendix to Chapter 3

Supplementary Table 1: Antibacterial activity of purified albicidin against various *E. coli* strains [200]. The table summarizes the crown widths of growth inhibition zones [208] (in mm) caused by varying quantities of isolated albicidin. *E. coli* DH5α is an albicidin-sensitive strain, but can spontaneously acquire resistance (*E. coli* DH5α). The *E. coli* strains DH5α/pBC-AlbXIV and DH5α/pBC-AlbXIX express an albicidin efflux pump and an albicidin *qnr* gene, respectively, both conferring resistance. *E. coli* RYC1000/pMR100 (SbmC) expresses a gyrase inhibitor resistance gene, and *E. coli* DH5α/GST-AlbD achieves resistance through biosynthesis of the albicidin-detoxifying protein AlbD. The observed correlation thus clearly identifies the isolated antibacterial compound as albicidin. The *E. coli* strains DH5α/pBC, RYC1000/pUC19 and DH5α/pGex4T3 represent the corresponding negative controls without the respective genes.

	Isolated albicidin		
Bacterial strain tested	**2.0 ng**	**0.2 ng**	**0.02 ng**
E. coli DH5α	7.5	5.5	2
E. coli DH5α/R spont. PR	0	0	0
E. coli DH5α/pBC	7.5	5	2
E. coli DH5α/pBC-AlbXIV	7.5	5	0.5
E. coli DH5α/pBC-AlbXIX	5.5	3	0
E. coli RYC1000/pUC19	7	5	1
E. coli RYC1000/pMR100 (SbmC)	6	3	0
E. coli DH5α/pGex4T3	8	6	2.5
E. coli DH5α/GST-AlbD	0	0	0

Supplementary Table 2: NMR acquisition parameters for albicidin samples in d_8-THF at 298 K and 16.4 Tesla. The experiments were based on standard pulse programs supplied by Bruker (Karslruhe, Germany).

Experiment	Pulse program	Scans	TD (F1/F2)	Miscellaneous
1D ^{1}H	zg30	128	32768	---
^{1}H-^{1}H COSY	cosygpqf	32	2048/256	---
^{1}H-^{1}H NOESY	noesygpph	64	2048/256	400 ms mixing time
^{1}H-^{13}C HMQC	hmqcgpqf	256	2048/128	---
^{1}H-^{13}C HMBC	hmbcgplpndqf	256	2048/128	---
HSQC-TOCSY	hsqcdietgpsisp.2	256	2048/64	100 ms spin lock
^{1}H-^{15}N HMQC*	hmqcgpqf	128	2048/256	---
^{1}H-^{15}N HMBC*	hmbcgplpndqf	128	2048/256	---

*Recorded with ^{15}N-labeled albicidin.

Supplementary Table 3: ^{1}H, ^{13}C and ^{15}N chemical shifts of albicidin in d_8-THF at 298 K and 16.4 Tesla. Multiplicities, scalar coupling constants J_{HH} and integrals are also given.

Position	$\delta(^1$H) in ppm (multiplicity, integral, J in Hz)	$\delta(^{13}$C, ^{15}N) in ppm
MCA-1		
-/C1	-	168.3 (qC)
-/C2	-	130.0 (qC)
H3/C3	7.30 (s, 0.8)	133.5 (CH)
-/C4	-	127.3 (qC)
H5/C5	7.33 (d, 1.4, 8.8)	131.0 (CH)
H6/C6	6.84 (d, 1.4, 8.5)	115.1 (CH)
-/C7	-	157.9 (qC)
OH @C7	n.d.	-
H8/C8	2.20 (d, 2.6, 1.2)	13.7 (CH$_3$)
pABA-2		
/-C1	-	127.9 (qC)
H2/C2	7.94 (d, 1.6, 8.6)	128.2 (CH)
H3/C3	7.89 (m, 3.9*)	118.8 (CH)
-/C4	-	143.3 (qC)
NH @C4	9.37 (s, 0.7)	126.0
/-C5	-	166.9 (qC)
Cya-3		
NH	8.50 (d, 0.9, 8.1)	109.7
Hα/Cα	5.16 (m, 1.0)	50.8 (CH)
Hβ',Hβ''/Cβ	3.13/3.07 (m,1.0/1.0)	19.5 (CH$_2$)
Cγ/Nδ	-	116.9 (qC)/ 252.5 (tN)
C'	-	167.8 (qC)
pABA-4		
-/C1	-	129.8 (qC)
H2/C2	7.98 (d, 1.8, 8.6)	128.2 (CH)
H3/C3	7.88 (m, 3.9*)	118.9 (CH)
-/C4	-	142.3 (qC)
NH @C4	10.09 (s, 0.9)	131.2
/-C5	-	164.2 (qC)
pMBA-5		
-/C1	-	114.1 (qC)
-/C2	-	151.3 (qC)
OH @C2	n.d.	-
-/C3	-	137.4 (qC)
-/C4	-	136.7 (qC)
NH @C4	9.03 (s, 0.9)	115.6
H5/C5	8.17 (d, 0.8, 8.8)	111.3 (CH)
H6/C6	7.81 (d, 0.9, 8.8)	124.7 (CH)
-/C7	-	165.3 (qC)
H8/C8	3.99 (s, 3.3)	60.0 (CH$_3$)
pMBA-6		
-/C1	-	109.1 (qC)
-/C2	-	155.1 (qC)
OH @C2	n.d.	-
-/C3	-	136.7 (qC)
-/C4	-	137.7 (qC)
NH @C4	10.31/10.09 (s, 0.8)	120.3
H5/C5	8.13 (d, 0.9, 8.8)	110.5 (CH)
H6/C6	7.65 (d, 0.9, 8.5)	125.3 (CH)
-/C7	-	172.1 (qC)
(CO)OH @C7	11.73 (s, 1.1)	-
H8/C8	4.04 (s, 3.0)**	59.6 (CH$_3$)

Supplementary Table 4: Functional assignment of genes from the albicidin gene cluster XALB1 of *Xanthomonas albilineans* strain Xa23R1 (genbank accession no. AJ586576).

Gene	Alternative gene name	Predicted function		Functional assignment
alb01	*albI*	PKS/NRPS		Tn*5* mutants in *alb01* were affected in the production of albicidin (no production) [194]
		PKS modules	PKS domains	
		PKS1	AL ACP1	
		PKS2	KS KR DH MT ACP2 ACP3	
		PKS3	KS T0	
		NRPS modules	NRPS domains	
		NRPS-1	C A T1	
		NRPS-2	C A T2	
		NRPS-3	C A T3	
		NRPS-4	C	
alb02	*albII* *xabC*	O-methyltransferase		One Tn*5* mutant in *alb02* affected in the production of albicidin (no production) [205]
alb03	*albIII*[1]	Type II thioesterase (suggested as required for hydrolysis of aberrant intermediates and optimal NRPS biosynthesis)		
alb04	*albIV*	NRPS		Tn*5* mutants in *alb04* affected in the production of albicidin (no production) [194]
		NRPS module	NRPS domains	
		NRPS-2*	A T2*	
alb05	*albV* *thp*	No function (transposon)		
alb06	albVI *xabD*	O-methyltransferase		
alb07	*albVII* *xabE*	4-hydroxybenzoate CoA ligase (Incorporation of 4-HB CoA as a starter unit by the module PKS1).		Insertion mutant (directed mutagenesis) in *alb07* affected in the production of albicidin (no production) [342].
alb08	*albVIII* *xabF*	SyrP-like (β-hydroxylase, proposed to be an albicidin tailoring enzyme)		
alb09	*albIX* *xabG*	NRPS		Tn*5* mutants in *alb09* affected in the production of albicidin (no production) [194]
		NRPS modules	NRPS domains	
		NRPS-4	A T4	
		NRPS-5	C A T5	
alb10	*albX*	MbtH-like protein		*Alb10* is frame-shifted and not functional in the sequenced *X. albilineans* strain GPE PC73 which produces albicidin (accession no. FP565176). Another MbtH gene is present elsewhere on the genome (YP_003375568)
alb11	*albXI*	SyrC-like protein (suggested to be required for transfer of Asn/Cya from module NRPS-2* to module NRPS-2).		
alb12	*albXII*	Benzoyl-CoA oxygenase (suggested as required for dioxygenation of pABA)		
alb13	*albXIII*	Acyltransferase required for loading of acyl substrates by modules PKS2 and PKS3		
alb14	*albXIV*	Albicidin transporter		The function of *alb14* (resistance to albicidin) was characterized in *E. coli*[343].

CVIII

alb15	*albXV*	Carbamoyl transferase (suggested as an albicidin tailoring enzyme)	
alb16	*albXVI*	No function (transposon)	
alb17	*albXVII* *pabAB*	p-aminobenzoate synthase (PabAB, aminodeoxychorismate synthase)	Insertion mutant (directed mutagenesis), not affected in the production of albicidin [342]. Genes *pabA* and *pabB* are present elsewhere in the genome of *X. albilineans* (YP_003374811 and YP_003375136, respectively).
alb18	*albXVIII* *pabC*	4-Amino-4desoxychorismate lyase (PabC, aminodeoxychorismate lyase) probably non-functional because of frame shifts in the gene.	Gene *pabC* is present elsewhere in the genome of *X. albilineans* (YP_003375456).
alb19	*albXIX* *albG*	McbG-like (immunity against albicidin)	Resistance-conferring gene characterized in *E. coli*[5]
alb20	*albXX* *ubiC*	UbiC-like protein (4-hydroxybenzoate synthetase)	No gene similar to *alb20* present elsewhere in the genome.

Supplementary Table 5: Medium composition for the production of unlabeled and ^{15}N-labeled albicidin in *Xanthomonas axonopodis* pv. *vesicatoria*.

Medium component	XVM3B	^{15}N-XVM3B
Glycerol	0.5 % (v/v) 6 g/L	0.5 % (v/v) 6 g/L
K_2HPO_4	0.24 mM	0.24 mM
KH_2PO_4	0.12 mM	0.12 mM
$(NH_4)_2SO_4$/	10 mM	-
$(^{15}N)\text{-}(NH_4)_2SO_4$	-	10 mM
$MgSO_4*7H_2O$	5 mM	5 mM
Casamino acids	0.015 %	0.015 %
$FeSO_4$	0.01 mM	0.01 mM
$CaCl_2$	1 mM	1 mM
NaCl	20 mM	20 mM

Supplementary Table 6: Oligonucleotides used for PCR amplification of NRPS modules

Construct	Forward primer	Reverse primer
NRPS-1	TTAACCATGGTGCGGGTGGGCAAGTGGGAC	GGTAGCGGCCGCGCTGGCGAGGGATT
NRPS-2	CATGGGTACCATGGCGCCCAACGAGGCGCAACCG	CCGGGCAAGCTTGATCTGCCGCGCCAGCGCCCCCAGT
NRPS-2*	TTAACCCATGGTGAACGAAACTGCAACTG	CCTTGGCGGCCGCTGCGACCAGATGCTC
NRPS-3	TTAACCATGGTGCGGGTGGGCAAGTGGGAC	CGGTAGCGGCCGCCAACGCTAGCTGCCT
NRPS-4	TTAACCATGGTGATGAAACTATCAAGCATGTC	GAGCGGCGGCCGCGCTGTCGCTGCTGGCGGCTT
NRPS-5	CTGGCCATGGTGCGCTGCCGCCTCTCGGATTT	GAGCGGCGGCCGCGCTGTCGCTGCTGGCGGCTT

CX
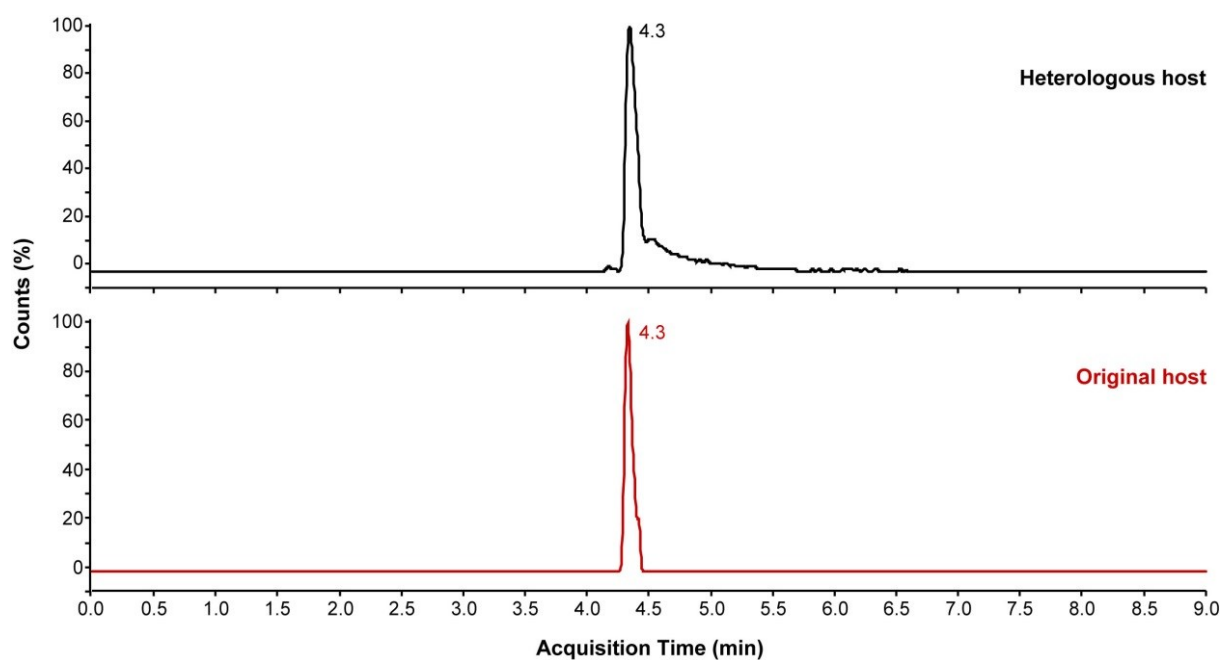
Supplementary Fig. 1: Comparison of albicidins produced by the wild type and heterologous host as observed by ESI negative SRM. XAD extract of *Xanthomonas albilineans* and purified albicidin from the heterologous producer strain are compared by selected reaction monitoring. Precursor ion selection was set to 841.2 m/z and MS2 was set to the typical albicidin product ion of 362.1 m/z.
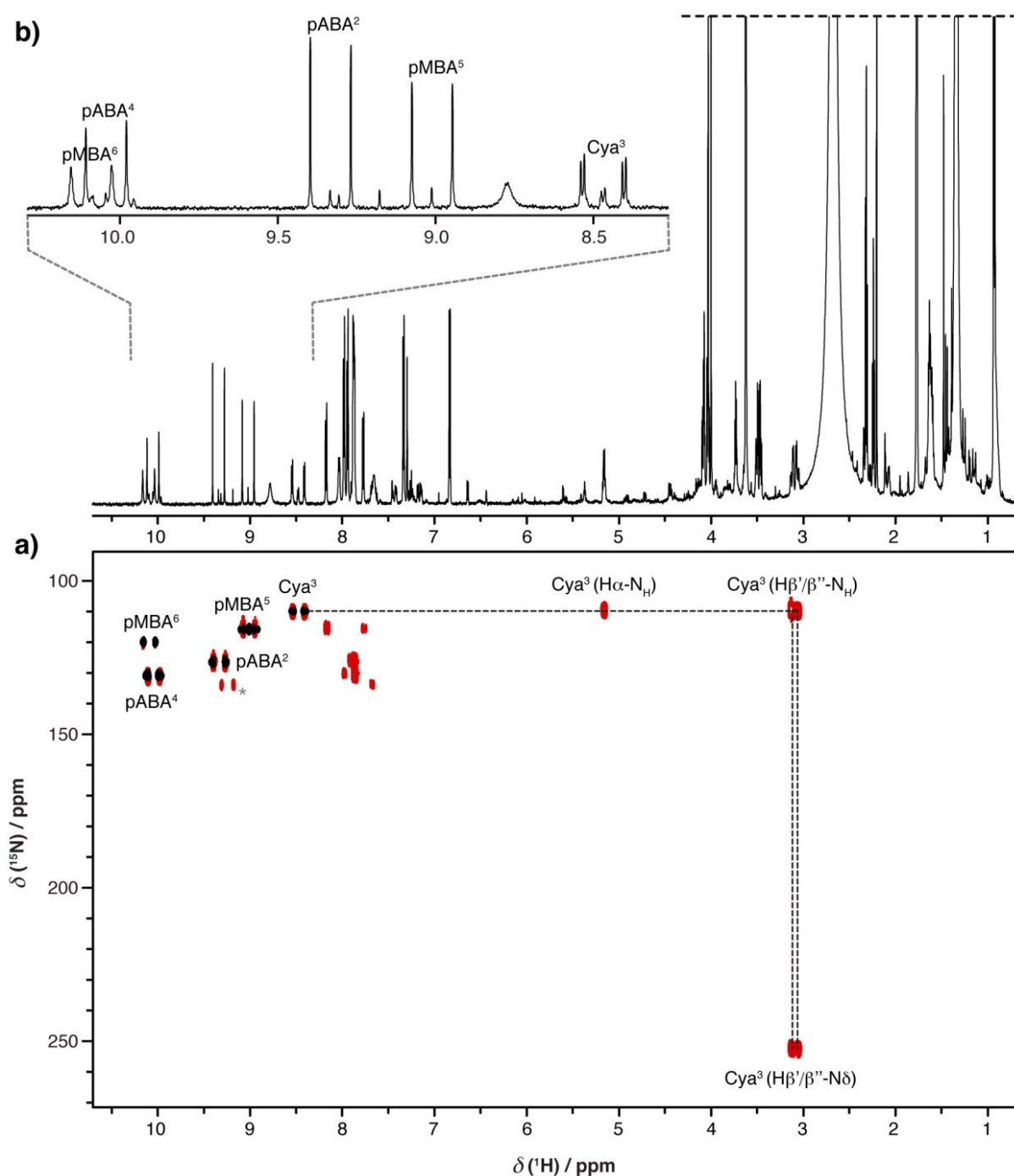
Supplementary Fig. 2: Assignment of ^{1}H and ^{15}N chemical shifts of albicidin. a) Overlay of 2D [1]H-^{15}N HMQC (black) and ^{1}H-^{15}N HMBC (red) spectra of ^{15}N-labeled albicidin in d_8-THF at 298 K and 16.4 Tesla. Resonance assignments are given for the HMQC spectrum. The HMBC peak pattern of Cya[3] is annotated to help the reader's eye. The asterisk denotes impurity. b) 1D ^{1}H spectrum of ^{15}N-labeled albicidin. The amide region of the spectrum is enlarged on top. Please note, that no ^{15}N decoupling was applied during acquisition, therefore giving rise to $^1J_{HN}$ peak splitting (~90 Hz). Solely the amide proton of the α-amino acid Cya[3] shows additional peak splitting due to $^3J_{Hn-Hα}$ coupling (~8 Hz).
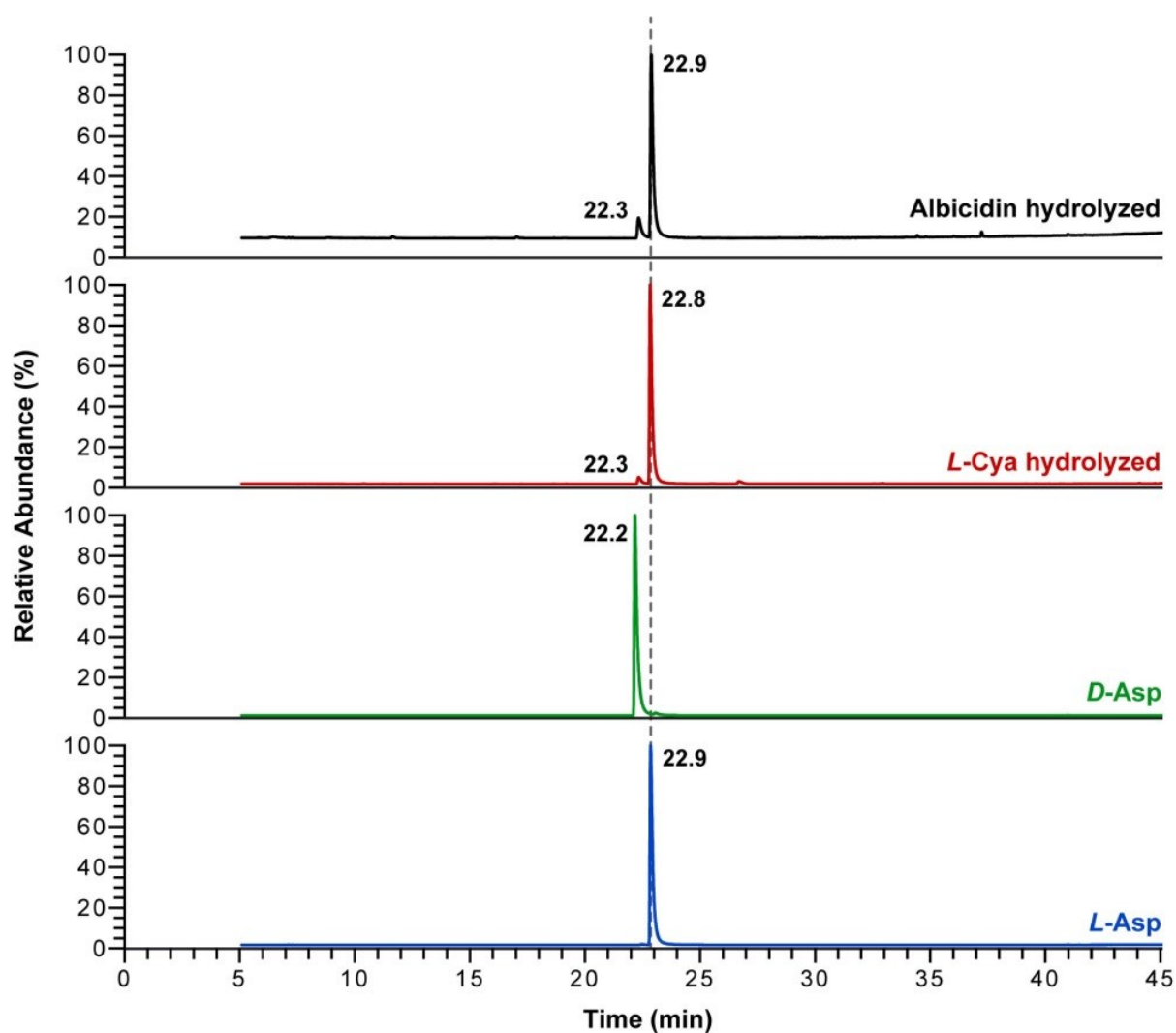
Supplementary Fig. 3: Total hydrolysis and chiral GC-MS for determination of absolute configuration of Cya-3. During the total hydrolysis of albicidin, complete hydrolysis of Cya to Asp is observed. In the selected ion monitoring (SIM) chromatograms on chiral GC we analyzed therefore the derivatized total hydrolysis product of albicidin, L-Cya as well as D- and L-Asp. The results clearly show that Cya-3 is L-configured. The small peak of D-Asp in the hydrolyzed albicidin occurs due to minor racemization during the hydrolysis, as it can be observed for L-Cya.
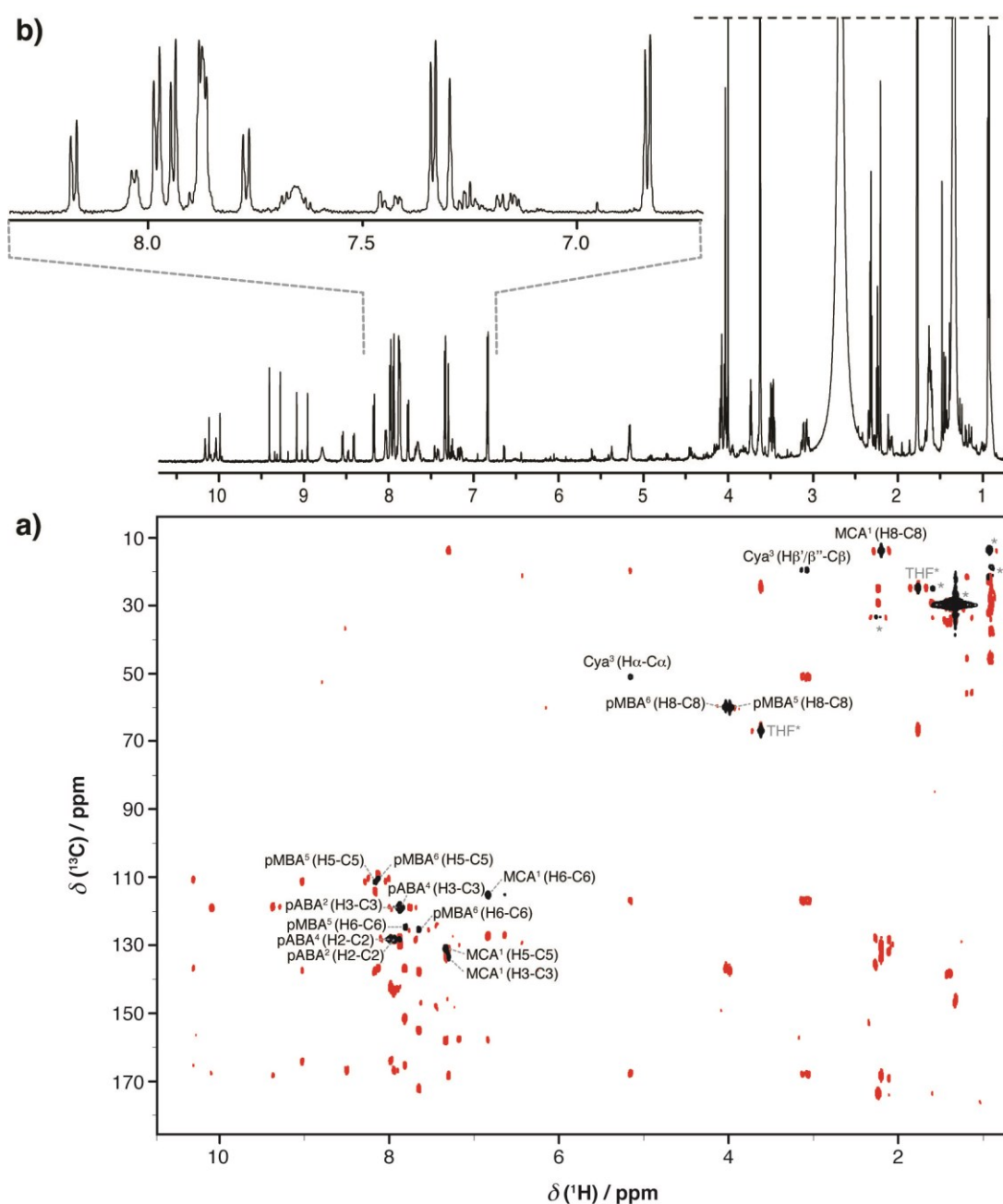
Supplementary Fig. 4: Assignment of ^{1}H and ^{13}C chemical shifts of albicidin. a) Overlay of 2D [1]H-^{13}C HMQC (black) and ^{1}H-^{13}C HMBC (red) spectra of unlabeled albicidin in d_8-THF at 298 K and 16.4 Tesla. Resonance assignments are given for the HMQC spectrum. The asterisks denote signals of impurities and solvent. b) 1D ^{1}H spectrum of ^{15}N-labeled albicidin. The H_N chemical shift of pMBA[6] is slightly shifted compared to the unlabeled sample. The aromatic region of the spectrum is enlarged on top.
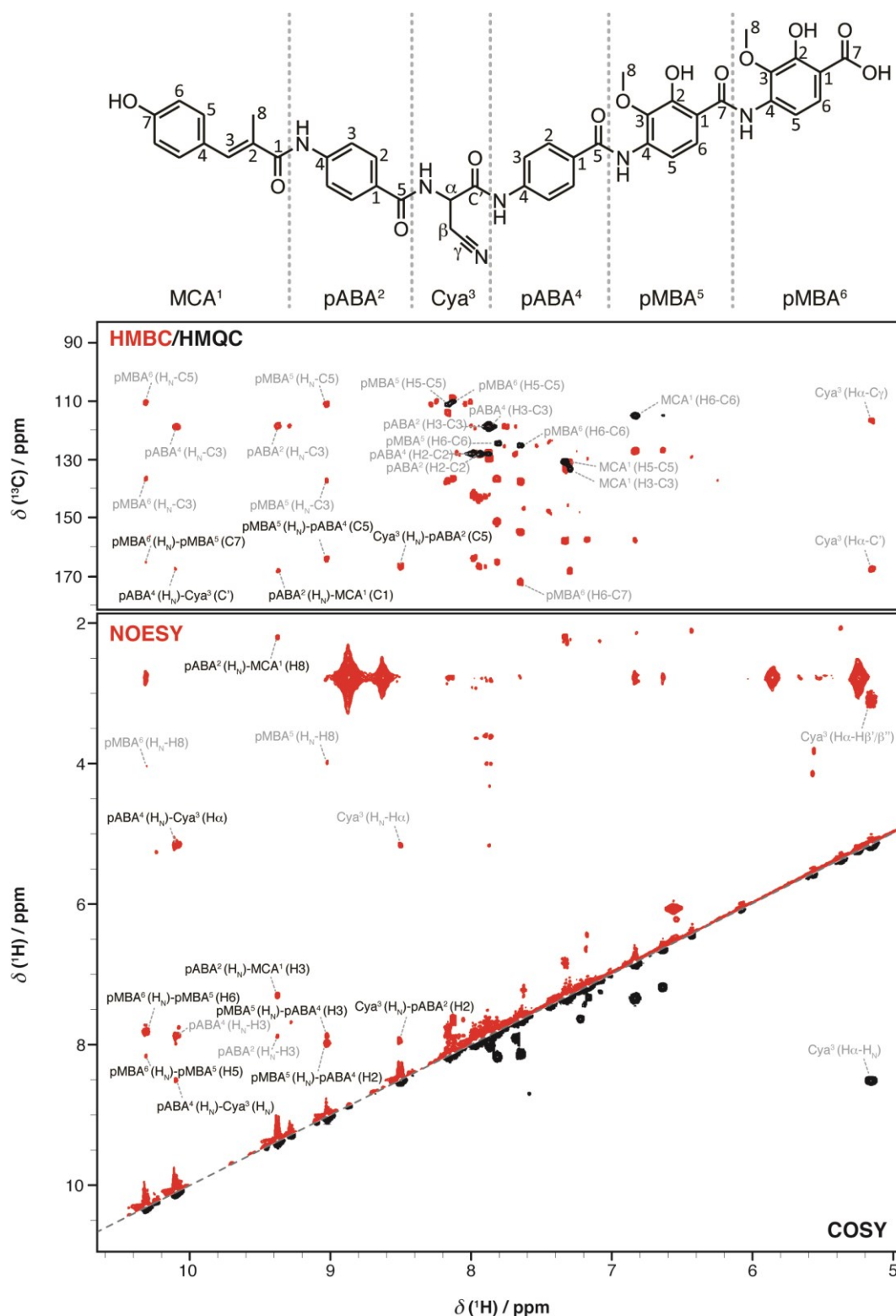
Supplementary Fig. 5: Sequential arrangement of identified building blocks in albicidin as obtained from NMR experiments. 2D NMR spectra are colored as indicated. Inter-residual correlations in the HMBC and NOESY spectra are annotated in black. Selected intra-residual correlations are also labeled (grey) to guide the reader. For visualization purposes, the plot on the bottom shows through-space (red) on one side and through-bond ^{1}H-^{1}H correlations (black) on the other side of the diagonal. The COSY spectrum reveals $^3J_{Hn-H\alpha}$ coupling only for the α-amino acid building block Cya[3]. The structure of albicidin is depicted on top.

Albicidin gene cluster XALB1 (48,131 bps)
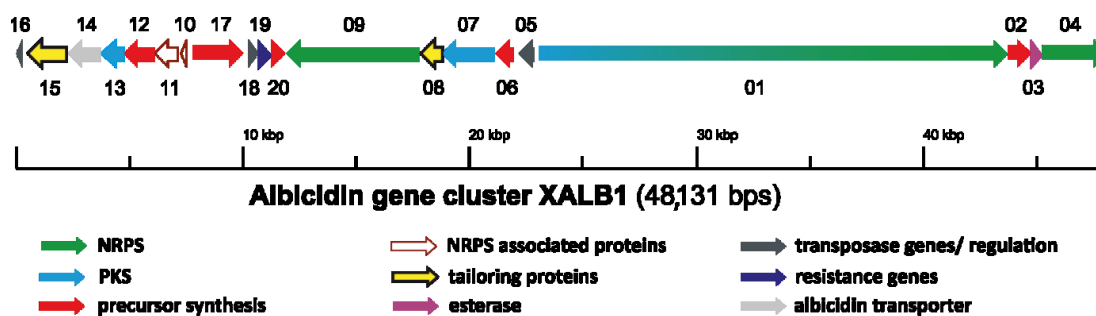
Supplementary Fig. 6: Gene cluster of albicidin biosynthesis. The gene machinery responsible for albicidin biosynthesis is located on three loci (XALB1-3). XALB1 contains the main genes, including the PKS-NRPS assembly line, NRPS and PKS associated proteins, enzymes involved in precursor synthesis of the pABA derivatives, an esterase, albicidin transporter, gene regulation and resistance genes as well as tailoring enzymes (see legend).[194] XALB2 and XALB3 encode each only one ORF, a phosphopantetheinyl transferase and the heat shock protein HtpG.[194, 200]

a)

[Sequence alignment of NRPS A domains showing rows A1, A2, A2*, A3, A4, A5, and GrsA, with core motifs core1–core10 labeled and ATP-side, ATP+Mg²⁺-side binding sites highlighted.]
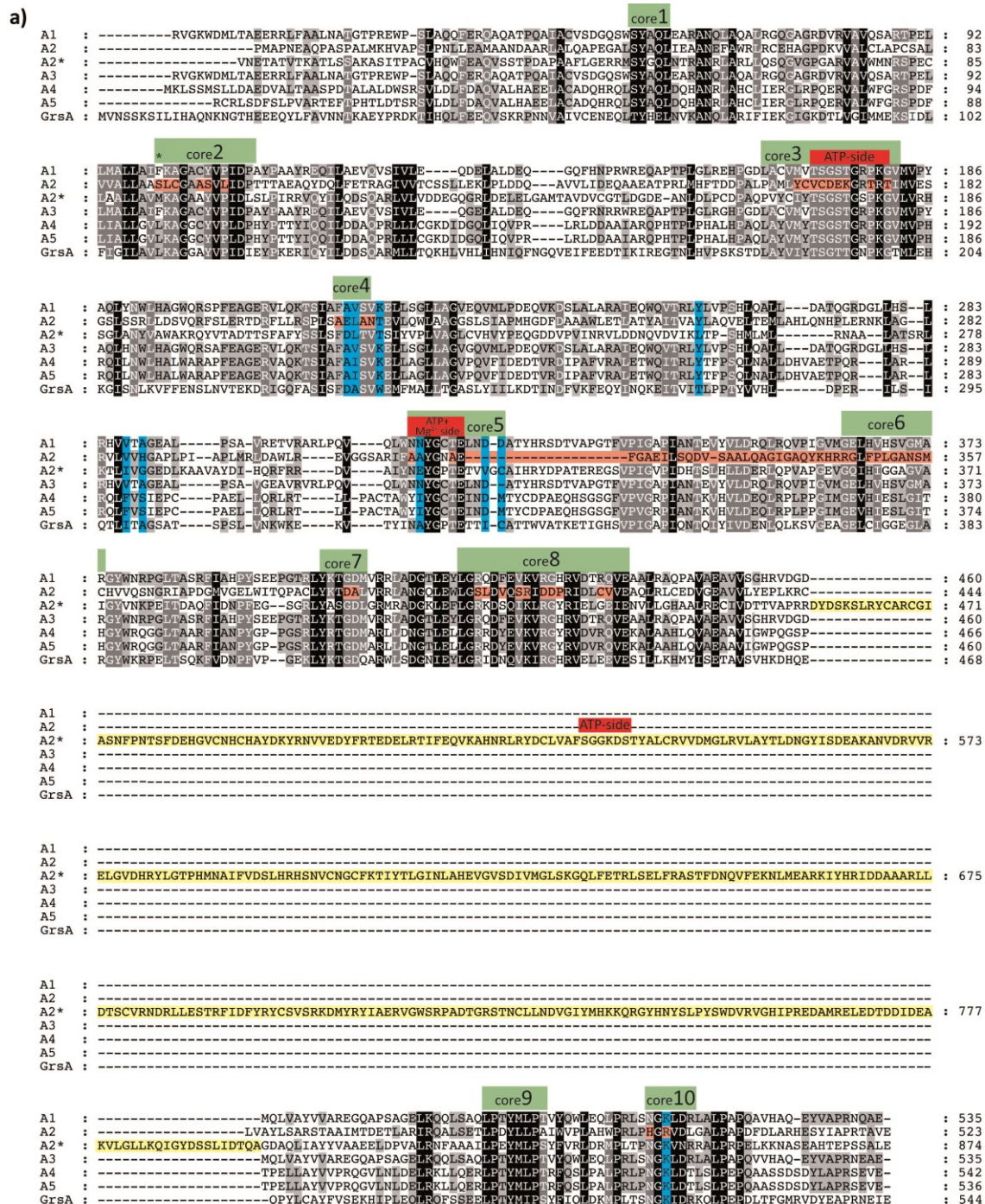
Supplementary Fig. 7: Sequence alignment of albicidin NRPS A domains (a) and T domains (b). The scale corresponds to the number of amino acids. Substrate selectivity-conferring residues are highlighted in blue. The alignment shows severe variations in the A domain of NRPS-2 (ATP, Mg^{2+} and substrate binding sites), indicating that this domain may not be active in terms of an A domain (highlighted in red). Furthermore, the non-conserved insertions in the T domains of NRPS-4 and NRPS-5 suggest further protein-protein interactions (yellow). The insertion sequence in module NRPS-2* (Alb04) is 338 amino acids in length and harbors the SGGKD-motif, which is strongly conserved among members of the adenosine nucleotide alpha hydrolases superfamily (Alpha ANH III like superfamily).[204, 344] Multiple alignment of various ANH III like proteins with the NRPS-2* insertion is shown in (c).

b)


```
                                                                    core T
T0      : ----------------ARICAICQVILGKPVTAQM------DFFAVGGHSVIAIQLVSRIRKSFGVEYPVSALFE---------------- : 53
T1      : ----------------QRIAALFAEVLRVEQVGIH-----DNFFALGGHSLSASQLISRIARDMAIDIPLAMLFELPTVAQISESI----- : 65
T2      : ----------------QAVAEIWQRVLKRTQVGVH-----DNFFELGGHSVLAIQLVSGIRKALAIEVPVTLVPEAPIIGAIARQI----- : 65
T2*     : ----------------QEIVQIWKEVLMVDKVGVR-----DNFFELGGHSLSALMLLYSIAERYQKMVSIQAFSVNPTIEGISEHIVA--- : 67
T3      : ----------------QRIAALFAEVLRVEQVGIH-----DNFFALGGHSLSASQLISRIRQSFHVDIPLSRIFEAPTIEGIVRQI----- : 65
T4      : ---------------ITIAKLWSELLTPAQAAPLRVSLNDNFFNLGGHSLLATQLFSRIRQSFDIEVRVNTLFESPVIEDFARVV----- : 70
T5      : ---------------ITIAKLWSELLTPAQAAPLRVSLNDNFFNLGGHSLLATQLFSRIRQSFDIEVRVNTLFESPVIEDFAAVV----- : 70
Sky2    : ------AGRQPRTSTEQVICGHIAEILQLSSVGID-----DSFFDLGGHSLIASRLISRIQSVQGVAISVRDLFEAPTVALFAKRVERAGT : 80
Sky5    : -------TRQAPRTPQEEIICGLFAEMLNLPSVGVH-----DNFFELGGHSLHATRVVSRIRSVFEVELPIRELWEAPTVAEIVGRIGGAAG : 80
bpsD    : ---------------IAALFAEVLGVSSVGVE-----DSFFEVGGHSLLATRVVSRIRESLRVRIRVQAFIDAPTIAEIAKVIDAALT : 68
sfcB    : ----------------RVQQIWCILLGLEDIGVT-----DDFFQLGGHSILVARMVERIETAFGRRVPIADILYFSPTIARIAAT----- : 64
sfcC    : -----------------IAQVLGLTAVTPD-----TDFFEQGGNSILLTRLAGTLSAKYQVQIPIHEFILTITPAAIAQAI----- : 59
tyrA    : MGVTEAQYVAPTNAVESKIAEIWERVLGVSGIGIL-----DNFFQIGGHSLKAMAVAAQVHREYQVELPLKVLFAQPTIKAIAQYVATRS- : 85
```

c)


```
                        ATP-side
A2* insert : LRYDCLVAFSGGKDSTYALCRVV-DMGLRVLAYTL :  34
gi|4006281 : QLYDCIIPVSGGKDSFFQAHIIK-ELCYNALLVTY :  34
gi|1555918 : -VYDCVIGVSGGKDSTFQVVKML-ELGLNPLCVCF :  33
gi|1579264 : FEYDCAIAVSGGKDSHFQVHIMKEKLGMNPILFSV :  35
gi|3386497 : NSYDVIVPVSGGKDGSYVAYNLKHKFNMNPLCVTV :  35
gi|6688607 : GSYDCVVPGSGGKDSFYAAHILKYKYGMNPLTVTW :  35
gi|3025016 : GPYDVIVCLSGGKDSLYLLYRLKVEYKLNVLAFTT :  35
```

Supplementary Fig. 7 (continued)

a)

Module	A domain signature									Predicted substrate
	235	236	239	278	299	301	322	330	331	517	
GrsA	D	A	W	T	I	A	A	I	C	K	
NRPS-1	A	V	K	Y	V	A	N	D	A	K	NO HIT
NRPS-2	E	L	T	Y	V	H	A	–	–	R	NO HIT
NRPS-2*	D	L	T	K	I	G	E	V	G	K	Asx
NRPS-3	A	V	K	Y	V	A	N	D	A	K	NO HIT
NRPS-4	A	I	K	Y	F	S	I	D	M	K	NO HIT
NRPS-5	A	I	K	Y	F	S	I	D	M	K	NO HIT
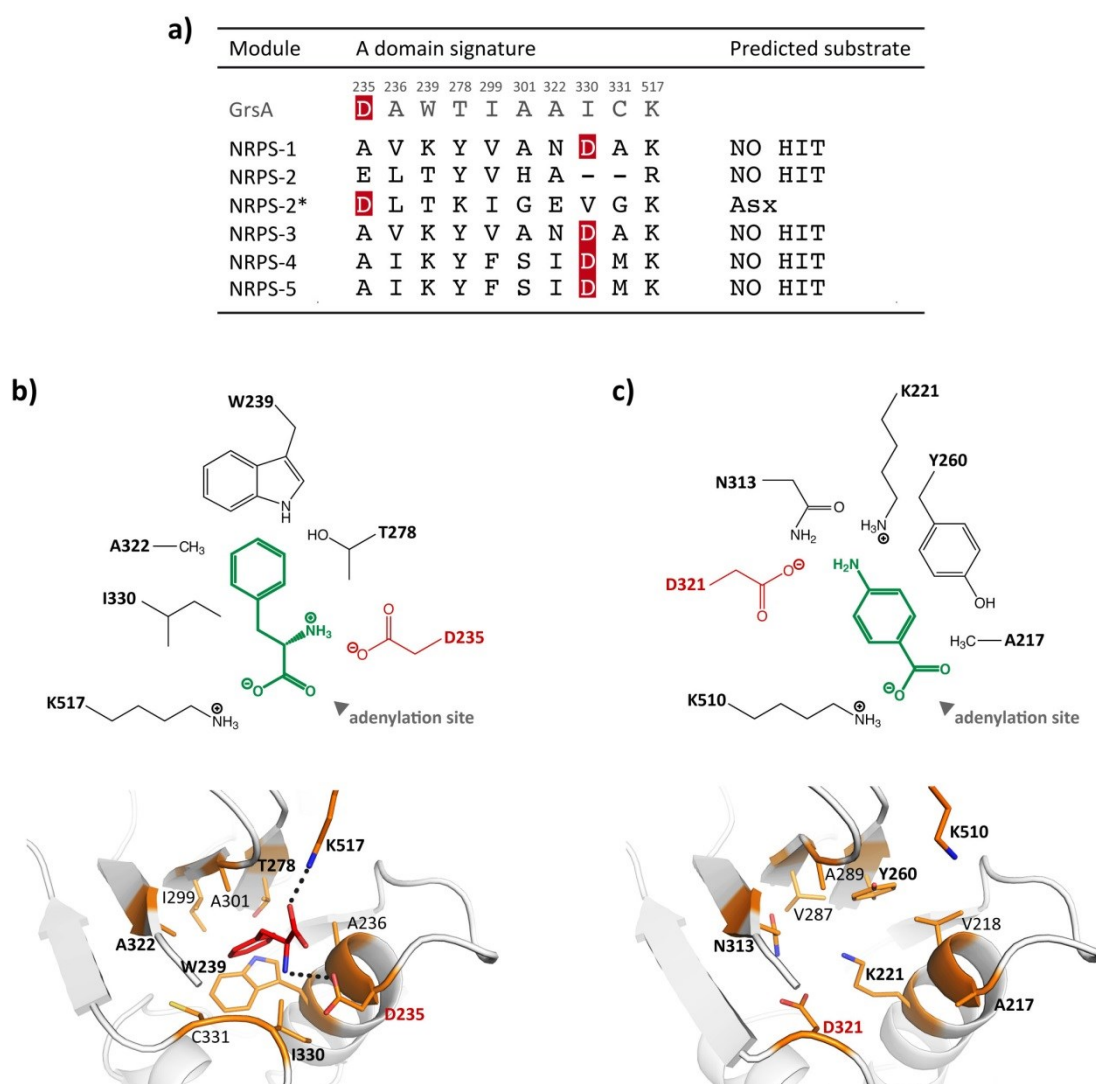
Supplementary Fig. 8: Substrate selectivity of adenylation domains in albicidin NRPS modules. a) Selectivity-conferring residues of A domains[202] of modules NRPS-1, NRPS-2, NRPS-2*, NRPS-3, NRPS-4 and NRPS-5 as derived from sequence alignments (see Supplementary Fig. 7). The corresponding residues in the Phe-activating A domain of gramicidin synthetase are shown as a reference. Note that the highly conserved residue D235 (red), which commonly interacts with the α-amino group of the substrate, is preserved only for the Asn/Cya-activating NRPS-2* A domain. However, a new aspartic acid appears for all A domains that have been shown to activate *p*-aminobenzoic acid derivatives (highlighted in red). NRPS-2 reveals severe variations, in particular the residues corresponding to the conserved D235 as well as K517 (GrsA) are not preserved. Substrate selectivity was predicted using the NRPS predictor tool of the University of Maryland [211]. Relevant interactions in the substrate binding pockets of A domains are shown for (b) the Phe-activating GrsA (PDB 1amu)[345] and (c) the structural model of NRPS-1 (based on PDB 1amu). The substrate Phe is depicted in red. Substrate selectivity-conferring residues are shown as sticks (orange). The re-positioned Asp residues (D235 and D321) are highlighted in red. Relevant residues (shown in bold) are also illustrated in simplistic arrangement patterns on top. The substrates (Phe, pABA) and the Asp residues are colored in green and red, respectively. In the NRPS-1 model, the amino moiety of pABA appears to be stabilized by a hydrogen-bond network established by residues K221, N313 and D321. The structural model of NRPS-1 was obtained from homology modeling using the webtool I-TASSER[212] and PDB entry 1amu [345].
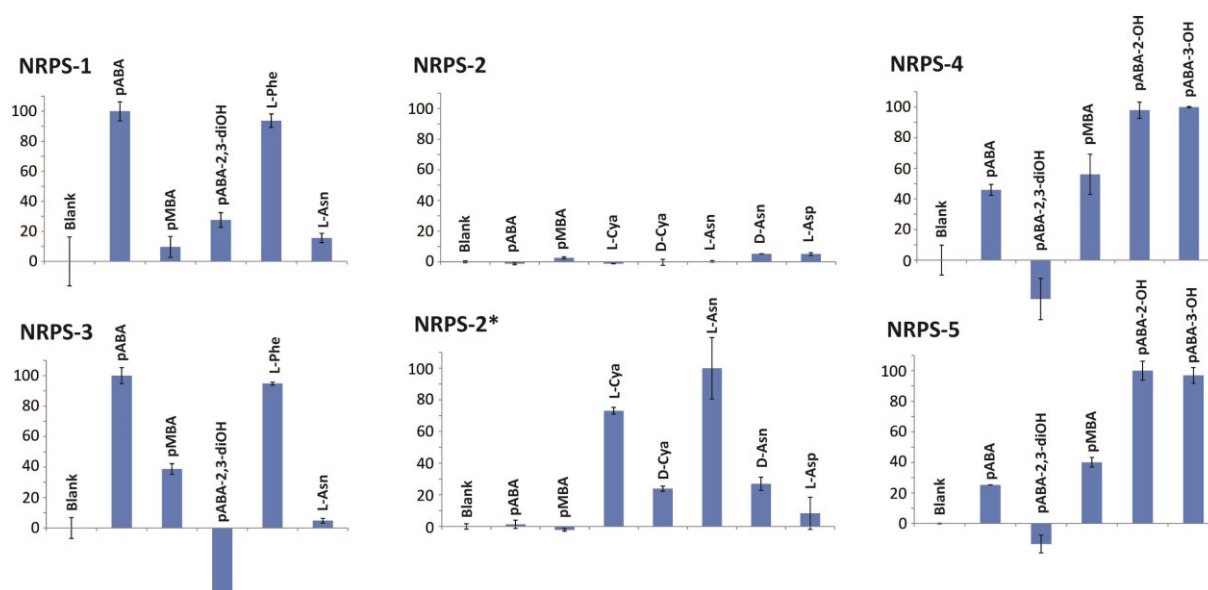
Supplementary Fig. 9: Substrate specificities of NRPS activation domains in albicidin biosynthesis. Relative turnover of activated substrates are shown. Radioactive ATP/PP$_i$ exchange assays were performed for the A domains of NRPS-1, NRPS-2, NRPS-2*, NRPS-3, NRPS-4 and NRPS-5. Normalization to 100% refers to 65.9 µCi/mol (NRPS-1), 1160.8 µCi/mol (NRPS2 and NRPS-2*), 168.3 µCi/mol (NRPS-3), 259.8 µCi/mol (NRPS-4) and 557,2 µCi/mol (NRPS-5), respectively. Experiments were performed in duplicates and error bars of standard deviation are shown.
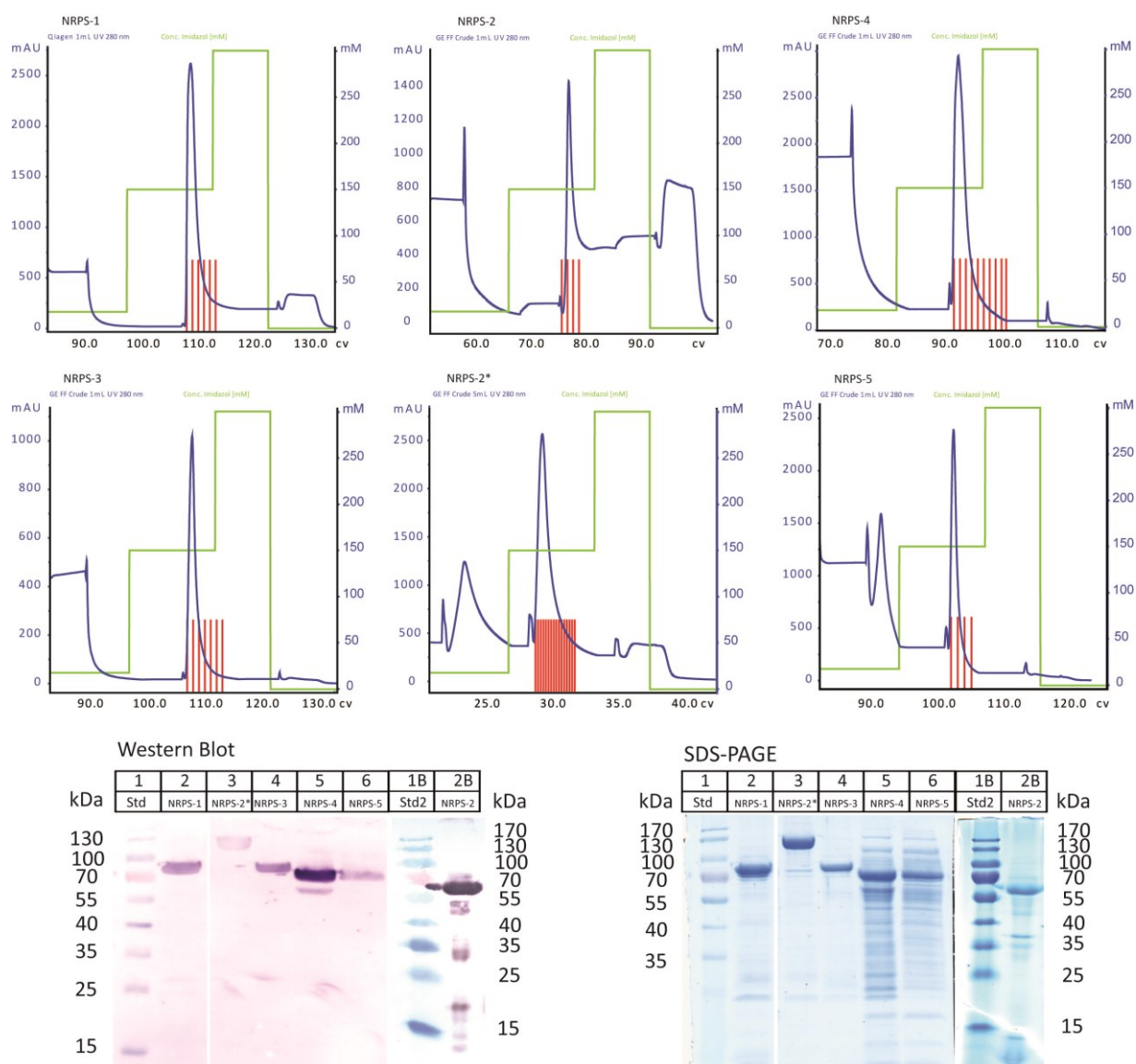
Supplementary Fig. 10: Ni-affinity purification, SDS-PAGE and western blot of NRPS module 1, 2*, 3, 4 and 5. UV Elution profiles at 280 nm of His-trap purifications are shown. The red lines mark the pooled fractions. Furthermore SDS-PAGE and western blot against the His$_6$-tag of the pooled recombinant proteins is shown. The expected protein masses are 81 kDa for NRPS-1, 66 kDa for NRPS-2, 119 kDa for NRPS-2*, 81 kDa for NRPS-3, 74 kDa for NRPS-4 and 73 kDa for NRPS-5.

B   Appendix to Chapter 4

Supplemental experimental information

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 was then separated through centrifugation at 50000 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 TrisHCl/300 Mm NaCl/20 mM imidazole, pH 7.8) and eluted by stepwise gradient of increasing imidazole concentration (elution buffer: 50 mM TrisHCl/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 TrisHcl/300 Mm NaCl/5% glycerine, 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 iodacetamide 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 ºC 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 2 m/z. A window of 3 m/z 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 a Bruker Avance 400 and Bruker Avance 500 NMR-spectrometer (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).

Compound 1: (*E*)-3-(4-Hydroxyphenyl)-2-methylacrylic acid

1
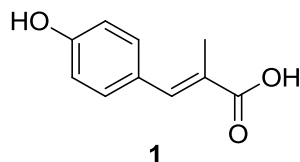
Chemical Formula: $C_{10}H_{10}O_3$
Exact Mass: 178,0630

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

Compound 2: (*E*)-*tert*-butyldimethylsilyl-3-(4-((*tert*-butyldimethylsilyl)oxy)phenyl)-2-methylacrylate

2
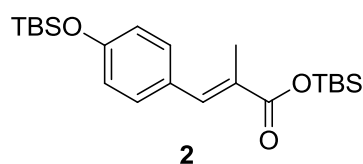
Chemical Formula: $C_{22}H_{38}O_3Si_2$
Exact Mass: 406,2359

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_2SO_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

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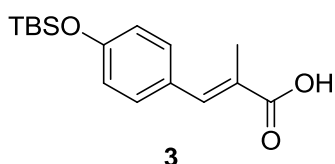
3
Chemical Formula: $C_{16}H_{24}O_3Si$
Exact Mass: 292,1495

Compound **2** (550 mg, 1.355 mmol) was dissolved in THF/MeOH (3:1) and one part of 10 % K_2CO_3-solution. 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 3 x 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

Compound 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
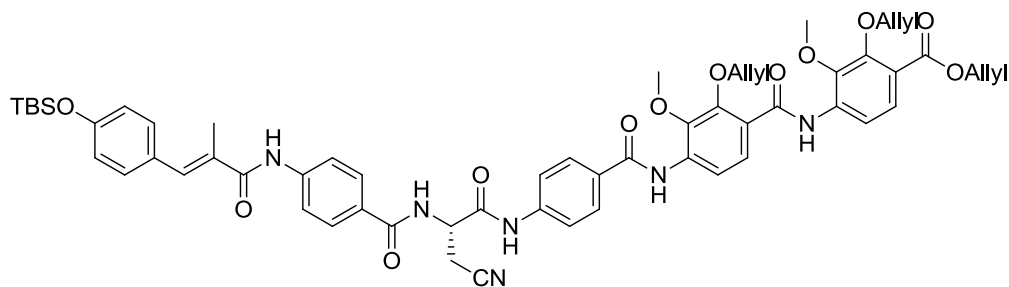
4

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 pentapeptide (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, saturated $NaHCO_3$ solution and brine. After drying over Na_2SO_4, 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-d$_6$, 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

Compound 4: (*S*,*E*)-allyl-2-(allyloxy)-4-(2-(allyloxy)-4-(4-(3-cyano-2-(4-(3-(4-hydroxyphenyl)-2-methyl-acrylamido)benzamido)propanamido)benzamido)-3-methoxybenzamido)-3-methoxybenzoate

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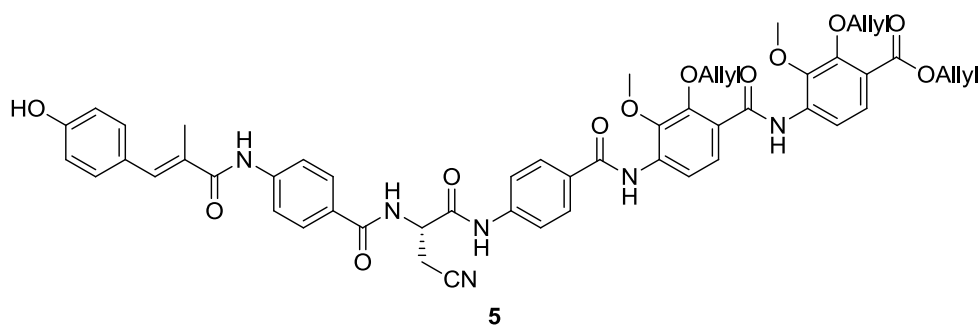
Chemical Formula: $C_{53}H_{50}N_6O_{12}$

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

Compound 5: (*S*,*E*)-allyl 2-(allyloxy)-4-(2-(allyloxy)-4-(4-(2-(4-(3-(4-(carbamoyloxy)phenyl)-2-methyl-acrylamido)benzamido)-3-cyanopropanamido)benzamido)-3-methoxybenzamido)-3-methoxybenzoate

6
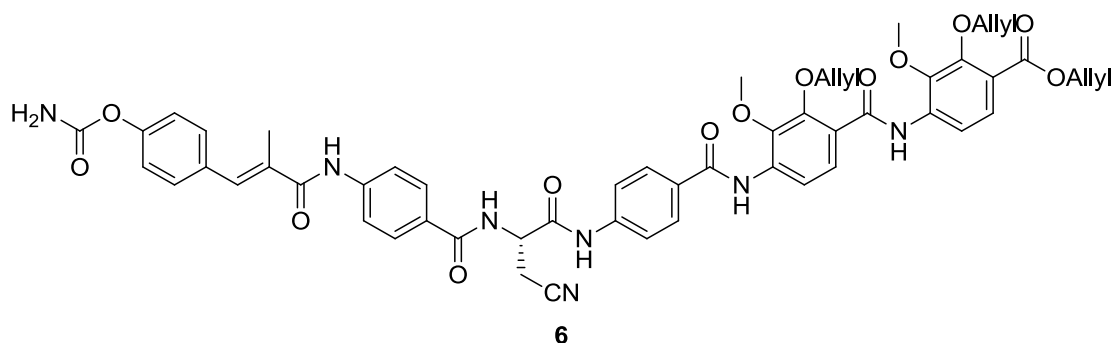
Chemical Formula: $C_{54}H_{51}N_7O_{13}$

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 3 x 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

Compound 6: (*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
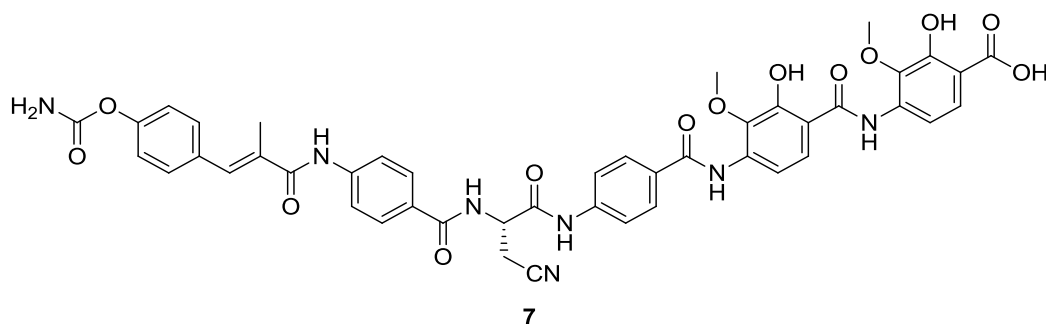
7
Chemical Formula: $C_{45}H_{39}N_7O_{13}$
Exact Mass: 885,2606

Compound **6** (83 mg, $8.259 \cdot 10^{-5}$ mol, 1.00 eq) was dissolved in THF and phenylsilane (71 mg, $6.607 \cdot 10^{-4}$ mol, 8.00 eq) was added. Subsequently, tetrakistriphenylphosphinpalladium (48 mg, $4.129 \cdot 10^{-5}$ 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

Supplemental table 7: Primers used to prepare the *alb15* deletion knockout mutant

Name	Sequence	Position on the genome of *X. albilineans* strain GPE PC73
A AlbXV	CCTGATCATCCAGACAGTGATGCGTACG	1742951 to 1742932
B AlbXV	TACCGACCAAGGTTGACCTTCACTCATGCCGTCCAGTACTGCGC	1742093 to 1742112
C AlbXV	TGTGAAGGTCAACCTTGGTCGGTATAACACCAGCTTCAACGTCA	1740989 to 1740970
D AlbXV	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

Supplemental table 8: Minimal inhibitory concentration (MIC) of carbamoyl-albicidin, albicidin and apramycin against *Bacillus subtilis*, *Mycobacterium phlei*, *Salmonella typhimurium* and *E. coli DH5a*.

MICs [ng/µL]:	*Bacillus subtilis*	*Mycobacterium phlei*	*Salmonella typhimurium*	*E. coli K12 BW25113*
Albicidin	< 0.2	< 0.2	~ 6.3	~ 0.06
Carbamoyl-Albi	< 0.2	< 0.2	~ 3.1	~ 0.1
Apramycin	~ 25	~ 1.6	~ 6.3	~ 4.0
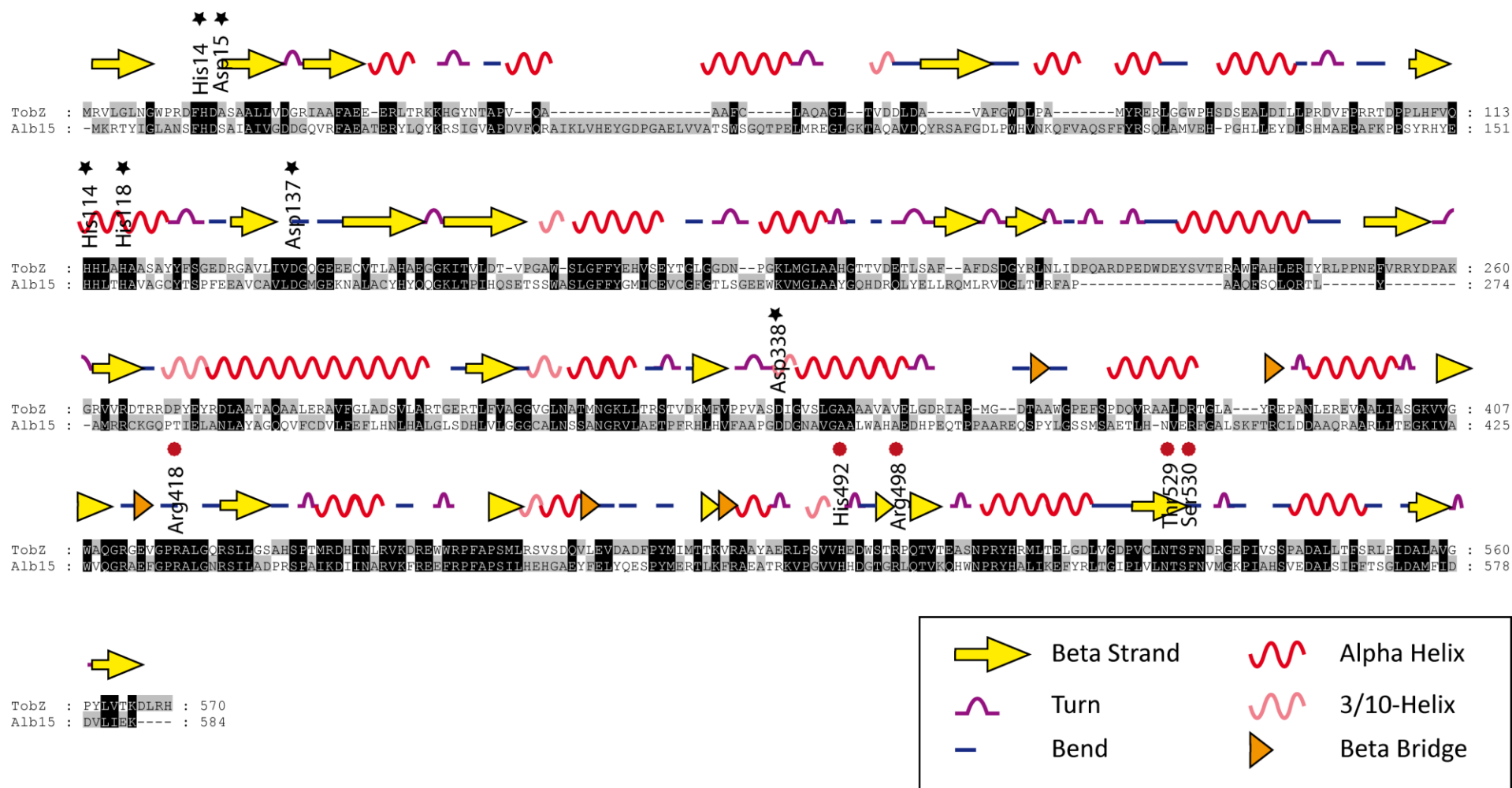
Supplemental Figure 11: Sequence comparison of Alb15 and TobZ. Amino acids involved in the complexation of Fe are marked with black stars; amino acids involved in the binding of carbamoyl-AMP are marked with red diamonds.
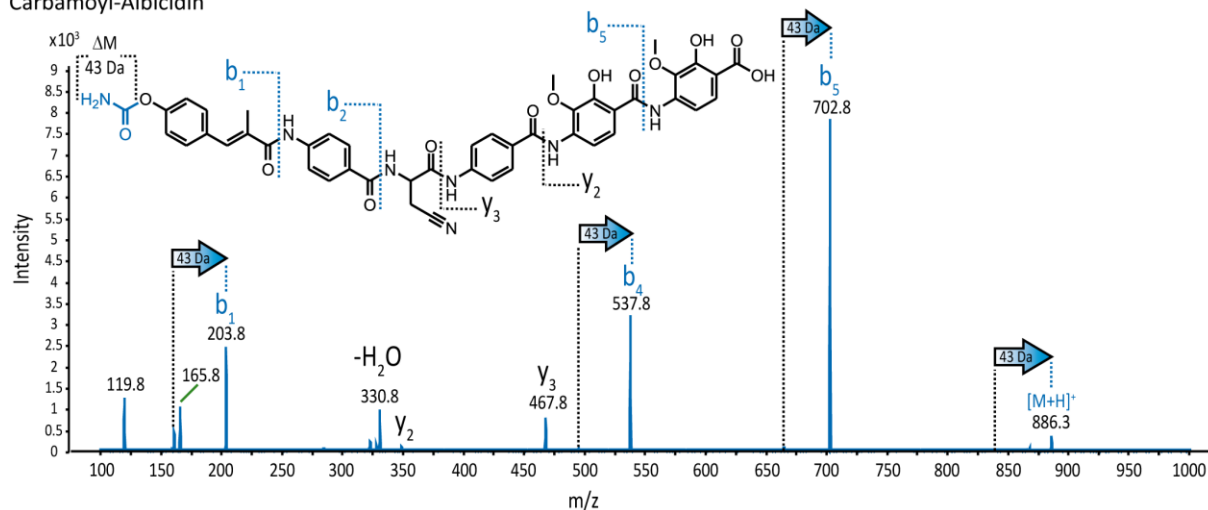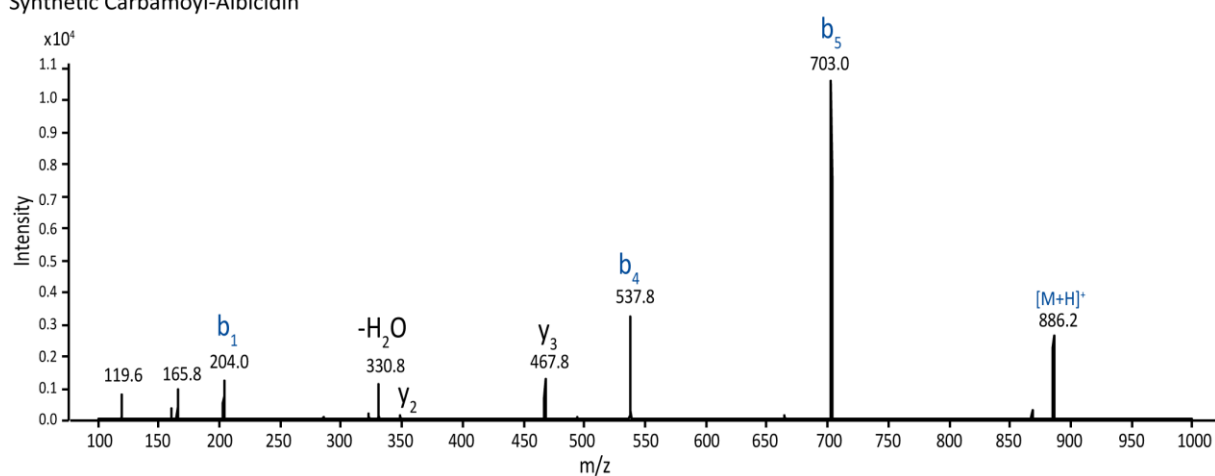
Supplemental Figure 12: MS/MS comparison of natural and synthetic carbamoyl-albicidin. The observed b-ion and y-ion fragments and the mass shift through carbamoylation are indicated in the spectra and structure.
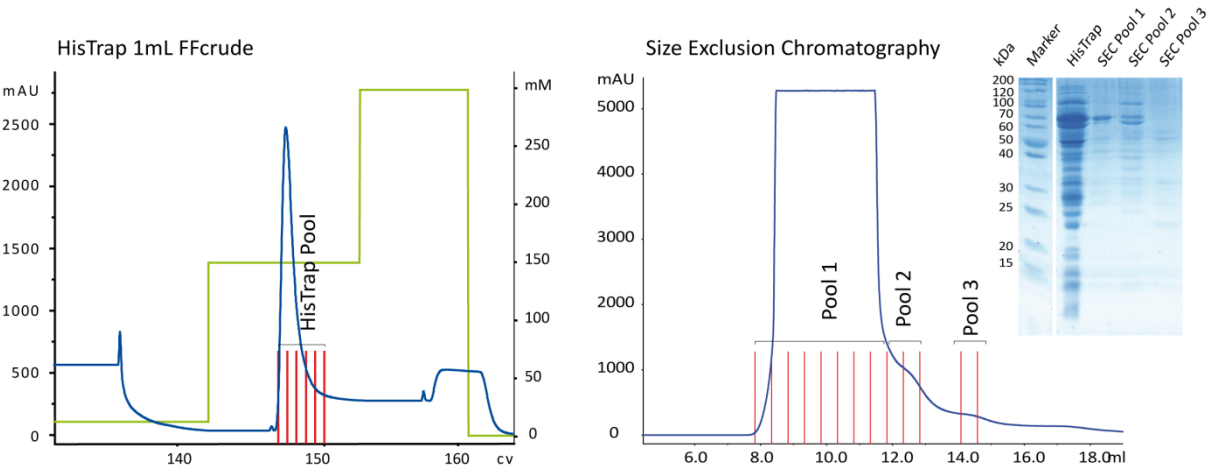
Supplemental Figure 13: Protein purification of Alb15 fusion protein. Ni affinity chromatogram (HisTrap), size exclusion and SDS-PAGE of the pooled fractions are shown. UV Elution profiles at λ = 280 nm of His-trap and size exclusion chromatography are shown. The red lines mark the pooled fractions. Furthermore SDS-PAGE of the pooled recombinant protein is shown. The expected protein mass is 79.6 kDa.
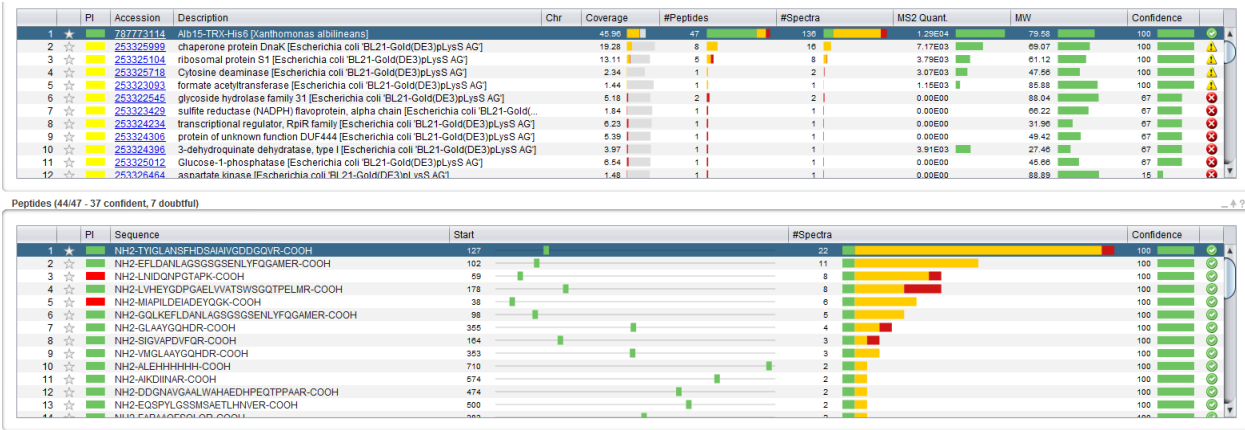
Supplemental Figure 14: Protein verification of in gel tryptic digest and MS/MS spectra database comparison. MS/MS spectra were search against an *E. coli BL 21 gold* protein database including the sequence of the TRX-Alb15-His₆ construct.
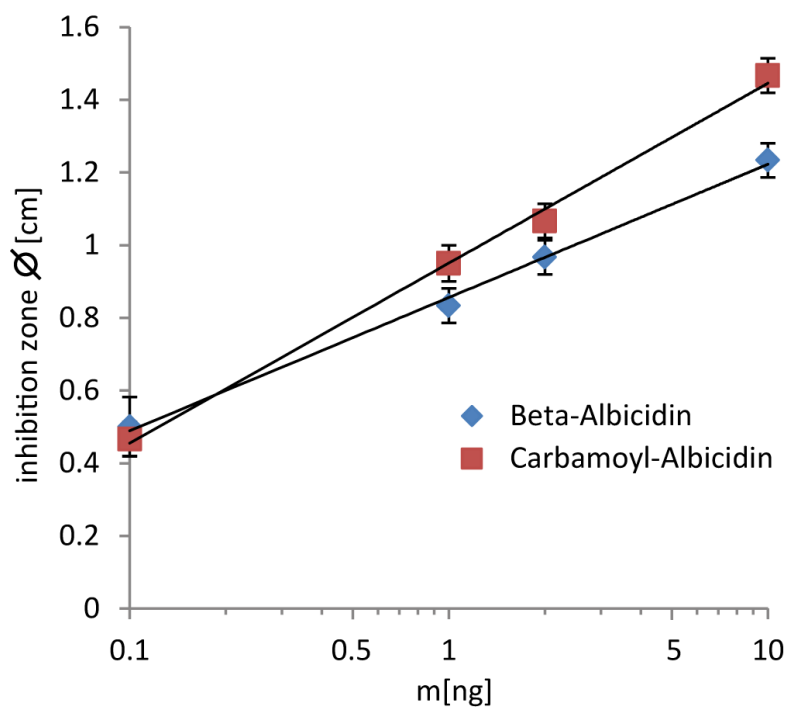
Supplemental Figure 15: Agar inhibition zone assay of *E. coli DH5alpha*. Carbamoyl-albicidin and albicidin were applied on top-agar plates with *E. coli* DH5alpha and incubated overnight. Inhibition zones were measured with a ruler and plotted against the amount of compound /spot. The experiment was performed in duplicates.

C   Appendix to Chapter 5

Supplemental Table 9

Set of *Ophiophagous hannah* venom proteins downloaded from the NCBI databank and used as customary database for MS-based (mass + nº of disulfide linkages), and BLAST assignments of top-down generated sequence tags and bottom-up MS/MS-gathered peptide ion sequences. Sequences of mature proteins are shown in red. Calculated isotope-averaged molecular masses for the fully reduced mature proteins (Mave red) are indicated below the entries.

>sp|Q7ZT98|CRVP_OPHHA Cysteine-rich venom protein ophanin OS=Ophiophagus hannah PE=1 SV=1

MIAFTLLSLAAVLQQSFGNVDFNSESTRRQKKQKEIVDLHNSLRRSVSPTASNMLKMQWYPEAASNA
ERWASNCNLGHSPDYSRVLEGIECGENIYMSSNPRAWTEIIQLWHDEYKNFVYGVGANPPGSVTGHY
TQIVWYKTYRIGCAVNYCPSSEYSYFYVCQYCPSGNMRGSTATPYKSGPTCGDCPSACDNGLCTNPC
TLYNEYTNCDSLVKQSSCQDEWIKSKCPASCFCHNKII

Mave red: 24976.89

>sp|Q53B46|CTX15_OPHHA Cytotoxin 15 OS=*Ophiophagus hannah* PE=1 SV=1

MKTLLLTLVVVTIVCLDLGYTRKCLNTPLPLIYKTCPIGQDKCIKMTIKKLPSKYDVIRGCIDICPK
SSADVEVLCCDTNKCNK

Mave red: 7061.52

>sp|Q69CK0|CTX27_OPHHA Cytotoxin 27 OS=*Ophiophagus hannah* PE=1 SV=1

MKTLLLTLVVVTIVCLDLGYTRKCLNTPLPLIYTTCPIGQDKCVKMTIKKLPSKYDVIRGCIDICPK
SSADVEVLCCDTNKCNK

Mave red: 7020.42

>sp|P01386|NXL2_OPHHA Long neurotoxin 2 OS=*Ophiophagus hannah* PE=1 SV=1

TKCYVTPDATSQTCPDGQDICYTKTWCDGFCSSRGKRIDLGCAATCPKVKPGVDIKCCSTDNCNPFP
TWKRKH

Mave red: 8053.21

>sp|P82662|NXL6_OPHHA Alpha-elapitoxin-Oh2b OS=*Ophiophagus hannah* PE=1 SV=2

MKTLLLTLVVMTIVCLDLGYTLICFISSHDSVTCAPGENVCFLKSWCDAWCGSRGKKLSFGCAATCP
KVNPGIDIECCSTDNCNPHPKLRP

Mave red: 7533.70

>sp|P80966|PA2A1_OPHHA Acidic phospholipase A2 1 OS=Ophiophagus hannah PE=1 SV=2

MNPAHLLVLSAVCVSLLGASSIPPQPLHLIQFGNMIQCTVPGFLSWIKYADYGCYCGAGGSGTPVDK
LDRCCQVHDNCYTQAQKLPACSSIMDSPYVKIYSYDCSERTVTCKADNDECAAFICNCDRVAAHCFA
ASPYNNNNYNIDTTTRC

Mave red: 13733.42

>sp|Q9DF33|PA2A2_OPHHA Acidic phospholipase A2 2 OS=Ophiophagus hannah PE=1 SV=1

MNPAHLLVLSAVCVSLLGASSIPPQPLHLVQFNGMIRCTIPGSIPWWDYSDYGCYCGSGGSGTPVDE
LDRCCQVHDNCYTQAQQLTECSPYSKRYSYDCSEGTLTCKADNDECAAFVCDCDRVAAICFAGAPYN

KENINIDTTTRC

2

Mave red: 13190.60

>sp|A8N286|TML3H_OPHHA Muscarinic toxin-like protein 3 homolog
OS=*Ophiophagus hannah* PE=1 SV=1

MKTLLLTLVVVTIVYLDLGYTTKCYNHQSTTPETTEICPDSGYFCYKSSWIDGREGRIERGCTFTCP
ELTPNGKYVYCCRRDKCNQ

Mave red: 7542.41

>sp|Q2VBN7|CTX14_OPHHA Cytotoxin 14 OS=*Ophiophagus hannah* PE=1 SV=1

MKTLLLTLVVVTIVCLDLGYTRKCLNTPLPLIYTTCPIGQDKCVKMTIKKLPSKYDVIRGCTDICPK
SSADVVVVCCDTNKCDK

Mave red: 6965.34

>sp|Q2VBN5|CTX21_OPHHA Cytotoxin 21 OS=*Ophiophagus hannah* PE=2 SV=1

MKTLLLTLVVVTIVCLDLGYTRKCLNTPLPLIYTTCPIGQDKCVKMTIKKLPSKYDVIRGCTDICPK
SSADVVVVCCDTNKCNK

Mave red: 6964.36

>sp|Q2VBN4|CTX23_OPHHA Cytotoxin 23 OS=*Ophiophagus hannah* PE=1 SV=1

MKTLLLTLVVVTIVCLDLGYTRKCLNTPLPLIYKTCPIGQDKCIKMTIKKLPSKYDVIRGCTDICPK
SSADVVVVCCDTNKCNK

Mave red: 7005.45

>sp|Q2VBN8|CTX9_OPHHA Cytotoxin 9 OS=*Ophiophagus hannah* PE=1 SV=1

MKTLLLTLVVVTIVCLDLGYTRKCLNTPLPLIYKTCPIGQDRCIKMTIKKLPSKYDVIRGCIDICPK
SSADVEVLCCDTNKCNK

Mave red: 7089.53

>sp|Q53B54|NXL17_OPHHA Long neurotoxin OH-17 OS=*Ophiophagus hannah*
PE=1 SV=1

TKCYITPDVKSETCPDGENICYTKSWCDVFCTSRGKRIDLGCAATCPKVKPGVDIKCCSTDNCNPFT
PWKRH

Mave red: 8037.25

>sp|P01387|NXL1_OPHHA Long neurotoxin 1 OS=*Ophiophagus hannah* PE=1
SV=1

TKCYVTPDVKSETCPAGQDICYTETWCDAWCTSRGKRVDLGCAATCPIVKPGVEIKCCSTDNCNPFP
TWRKRP

Mave red: 8106.31

>sp|Q2VBP5|NXL22_OPHHA Long neurotoxin LNTX22 OS=*Ophiophagus hannah*
PE=1 SV=1

MKTLLLTLVVVTIVCLDLGYTTKCYVTPDVKSETCPDGENICYTKSWCEVFCTSRGKRIDLGRAATC
PKVKPGVDIKCCSTDNCNPFTPWKRH

Mave red: 8090.30

>sp|Q2VBP4|NXL28_OPHHA Long neurotoxin LNTX28 OS=*Ophiophagus hannah*
PE=1 SV=1

MKTLLLTLVVMTIVCLDLGYTLICFISSHDSVTCAPGENVCFLKSWCDAWCGSRGKKLSFGCAATCP
RVNPGIDIECCSTDNCNPHPKLRP

Mave red: 7561.71

3

>sp|A8N285|NXL2H_OPHHA Long neurotoxin LNTX-2 homolog
OS=*Ophiophagus hannah* PE=3 SV=1

MKTLLLTLVVVTIVCLDLGYTTKCYVTPDATSQTCPDGENICYTKSWCDVFCSSRGKRIDLGCAATC
PKVKPGVDIKCCSTDNCNPFTPWKRH

Mave red: 7953.09

>sp|Q53B55|NXL31_OPHHA Long neurotoxin-like OH-31 OS=Ophiophagus
hannah PE=2 SV=1

MKTLLLTLVVVTILCLDLGLELTNAPDSWSSRRTCLCPAWVPLRSRPVAGHSKQCGSRGRRVDLGCA
ATCPIVKPGVNINCCSTDNCNPFPKRS

Mave red: 8113.33

>sp|Q53B53|NXL34_OPHHA Long neurotoxin OH-34 (Fragment)
OS=*Ophiophagus hannah* PE=1 SV=1

CXXXVI

KTLLLTLVVVTILCLDLGYTTKCYITPDVKSETCPDGENICYTKTWCDVWCGSRGRRVDLGCAATCP
IVKPGVNINCCSTDNCNPFPKRS
Mave red: 7692.80
>sp|Q53B59|NXL37_OPHHA Long neurotoxin OH-37 OS=*Ophiophagus hannah*
PE=1 SV=1
MKTLLLTLVVMTIVCLDLGYSLICFISPHDSVTCAPGENVCFLKSWCDAWCGSRGKKLSFGCAATCP
KVNPGIDIECCSTDNCNPHPKLRP
Mave red: 7543.74
>sp|Q2VBP3|NXL3X_OPHHA Long neurotoxin LNTX37 OS=*Ophiophagus hannah*
PE=2 SV=1
MKTLLLTLVVVTIMCLDLGYTTKCYKTGERIISETCPPGQDLCYMKTWCDVFCGSRGRVIELGCTAT
CPTVKHHEQITCCSTDNCNPHPKMKQR
Mave red: 8240.53
>sp|P80156|NXL4_OPHHA Long neurotoxin 4 OS=*Ophiophagus hannah* PE=1
SV=1
TKCYKTGDRIISEACPPGQDLCYMKTWCDVFCGTRGRVIELGCTATCPTVKPHEQITCCSTDNCNPH
PKMKQ
Mave red: 8014.29
>sp|Q53B58|NXL55_OPHHA Long neurotoxin OH-55 OS=*Ophiophagus hannah*
PE=1 SV=1
MKTLLLTLVVVTIVCLDLGYTTKCYVTPDVKSETCPAGQDICYTETWCDAWCTSRGKRVNLGCAATC
PIVKPGVEIKCCSTDNCNPFPTRKRP
Mave red: 7919.12
>sp|Q53B57|NXL56_OPHHA Long neurotoxin OH-56 OS=*Ophiophagus hannah*
PE=1 SV=1
MKTLLLTLVVVTIMCLDLGYTTKCYVTPDVTSQTCPDGQNICYTETWCDAWCGSRGKRVNLGCAATC
PKVNPGVDIICCSTDNCNPFPKRS
Mave red: 7594.61
>sp|Q53B56|NXL57_OPHHA OH-57 OS=*Ophiophagus hannah* PE=2 SV=1
MKTLLLTLVVVTIVCLDLGYTRICHKSSFISETCPDGQNLCYLKSWCDIFCGSRGERLEFGCAATCP
EVKPGVNIECCSTDNCNPHPKLRP
4
Mave red: 7766.88
>sp|Q2VBP8|NXLX1_OPHHA Long neurotoxin LNTX1 OS=*Ophiophagus hannah*
PE=2 SV=1
MKILLLTLVVVTIMCLDLGYTTKCYKTGERIISETCPPGQDLCYMKTWCDVFCGSRGRVIELGCTAT
CPTVKPHEQITCCSTDNCNPHPKMKQR
Mave red: 8200.51
>sp|Q2VBP6|NXLX8_OPHHA Long neurotoxin LNTX8 OS=*Ophiophagus hannah*
PE=2 SV=1
MKTLLLTLVVVTIMCLDLGYTTKCYKTGERIISETCPPGQDLCYMKTWCDVFCGSRGRVVELGCTAT
CPTVKPHEQITCCSTDNCNPHPKMKQR
Mave red: 8186.48
>sp|Q53B49|NXS35_OPHHA Short neurotoxin OH-35 OS=*Ophiophagus hannah*
PE=1 SV=1
MKTLLLTLVVVTIVCLDLGHTLICVKQYTIFGVTPEICADGQNLCYKTWHMVYPGGYDHTRGCAATC
PKMKNHDTVHCCTTDKCNL
Mave red: 7299.44
>sp|Q9DF56|PA2A_OPHHA Acidic phospholipase A2 OS=*Ophiophagus hannah*
PE=1 SV=1
MNPAHLLVLSAVCVSLLGASSIPPQPLNLLQFNYMIQCTIPGSRPFLDYMDYGCYCGTGVAGHPVDE
LDRCCQTHDLCYSKAEEQPKCSSLLNSPLMKKYSYTCSGGTLTCNDDNDECGAFICNCDRAARICFA
GAPYNKENKELDIATRCQ
Mave red: 13930.71
>sp|Q2VBN0|TXM38_OPHHA Muscarinic toxin 38 OS=*Ophiophagus hannah*
PE=2 SV=1
MKTLLLTLVVVTIVCLDLGYTMTCYTQYSLSPPTTKTCPDGQNLCYKRWFAFIPHGNKFFRGCAAAC

PKAEHNEVVRCCARDKCNL
Mave red: 7393.55
>sp|P01412|TXW1_OPHHA Weak toxin DE-1 OS=*Ophiophagus hannah* PE=1
SV=2
MKTLLLTLVVVTIVCLDLGYSLICFNQETYRPETTTTCPDGEDTCYSTFWNDHHGVKIERGCGCPRV
NPPISIICCKTDKCNN
Mave red: 7026.89
>sp|Q2VBN2|TXW34_OPHHA Weak neurotoxin WNTX34 OS=*Ophiophagus hannah*
PE=1 SV=1
MKTLLLTLVVVTIVCLDLGYSLTCLNCPEQYCKRIHTCRNGENVCFKRFYEGKLLCKQFRRGCAATC
PEAKSREIVQCCSTDECNH
Mave red: 7537.71
>sp|Q2VBN3|TXW33_OPHHA Weak neurotoxin WNTX33 OS=*Ophiophagus hannah*
PE=2 SV=1
MKTLLLTLVVVTIVCLDLGYSLICFNQETYRPETTTTCPDGEDTCYSTFWNDHRGVKIERGCGCPRV
NPGISIICCKTDKCNN
Mave red: 7005.87
>sp|Q53B61|TXW72_OPHHA Weak neurotoxin OH-72 (Fragment)
5
OS=*Ophiophagus hannah* PE=2 SV=1
LTLVVVTIVCLDLGYTLTCLICPEEYCKRIHTCRDGENVCFKGFYEGKQLGKQFRRGCAATCPEGKP
NEIVQCCSTDECNH
Mave red: 7045.01
>sp|P83234|VESP_OPHHA Ohanin OS=*Ophiophagus hannah* PE=1 SV=2
MLLFTLCFFADQENGGKALASPPGNWQKADVTFDSNTAFESLVVSPDKKTVENVGVPKGVPDSPERF
SSSPCVLGSPGFRSGKHFFEVKYGTQREWAVGLAGKSVKRKGYLRLVPEERIWQKGLWWLRRLETDS
DKLQKGSGKIIVFLDYDEGKVIFDLDGEVTTIQANFNGEEVVPFYYIGARVSLANL
Mave red: 11951.59
>sp|B6RLX2|VKTCT_OPHHA Kunitz-type serine protease inhibitor TCI
OS=*Ophiophagus hannah* PE=1 SV=1
MSSGRLLLLLGLLTLWAELTPVSGLGRPKFCELPAVSGFCKAYIPSFYYNPDASACQKFIYGGCGGN
ANKFKTIEECHRTCVG
Mave red: 6346.27
>sp|P82966|VKTCI_OPHHA Kunitz-type serine protease inhibitor
OS=*Ophiophagus hannah* PE=1 SV=2
MSSGRLLLLLGLLTLWAELTPVSGLGRPKFCELPPEPGLCNARKTFFYYSLHSHACQKFIYGGCGGN
ANKFKTIDECHRTCVG
Mave red: 6499.46
>sp|C5ILC5|NXL2A_OPHHA Alpha-elapitoxin-Oh2a OS=*Ophiophagus hannah*
PE=2 SV=1
MKTLLLTLVVVTIVCLDLGYTLLCYKTPSPINAETCPPGENLCYTKMWCDAWCSSRGKVIELGCAAT
CPSKKPYEEVDCCSTDNCNPHPKLRP
Mave red: 7856.03
>sp|P07526|NXL3_OPHHA Long neurotoxin 3 OS=*Ophiophagus hannah* PE=1
SV=1
TKCYVTPDVKSETCPAGQDLCYTETWCVAWCTVRGKRVSLTCAAICPIVPPKVSIKCCSTDACGPFP
TWPNVR
Mave red: 7957.34
>sp|P80965|NXL5_OPHHA Long neurotoxin OH-5 OS=*Ophiophagus hannah*
PE=1 SV=1
TKCYKTGDRIISEACPPGQDLCYMKTWCDVFCGTRGRVIELGCTATCPTVKPHEQITCCSTDNCDPH
HKMLQ
Mave red: 8040.29
>sp|P83302|NXOH9_OPHHA Neurotoxin Oh9-1 OS=*Ophiophagus hannah* PE=1
SV=1
LICHRVHGLQTCEPDQKFCFRKTTMFFPNHPVLLMGCTSSCPTEKYSVCCSTDKCNK
Mave red: 6515.65

CXXXVIII

>sp|Q2VBP1|NXS11_OPHHA Short neurotoxin SNTX11 OS=Ophiophagus
hannah PE=2 SV=1
MKTLLLTFLVVTIVCLDLGYTLICHRVHGLQTCEPDQKFCFRKTTMFFPNHPVLLMGCTYSCPTEKY
SVCCSTDKCNK
Mave red: 6591.75
6
>sp|Q2VBP0|NXS14_OPHHA Short neurotoxin SNTX14 OS=Ophiophagus
hannah PE=2 SV=1
MKTLLLTFLVVTIVCLDLGYTLICHQLHGLQTCEPAQKFCQKRTTMFFPNHPVLLMGCTYNCPTERY
SVCCSTDKCNK
Mave red: 6569.70
>sp|Q53B52|NXS26_OPHHA Short neurotoxin OH-26 OS=*Ophiophagus hannah*
PE=1 SV=1
MKNLLLTFLVVTIVCLDLGYTLICHQRHGLQTCEPAQKFCFAQTVMPFPNHPLTLMGCTYSCPTEKN
AVCCSTDKCNR
Mave red: 6404.46
>sp|Q2VBN9|NXS2X_OPHHA Short neurotoxin SNTX26 OS=Ophiophagus
hannah PE=2 SV=1
MKTLLLTFLVVTIVCLDLGYTLICHQVHGLQTCEPAQKFCQKRTTMFFPNHPVLLMGCTYNCPTERY
SVCCSTDKCNK
Mave red: 6555.67
>sp|Q53B50|NXS32_OPHHA Short neurotoxin OH-32 OS=*Ophiophagus hannah*
PE=2 SV=1
MKNLLLTFLVVTIVCLDLGYTLICNRVHGLQTCEPAHKFCFSKTVMPFPNHPLTLMGCTYSCPTERN
AVCCSTDKCN
Mave red: 6249.31
>sp|Q53B48|NXS46_OPHHA Short neurotoxin OH-46 OS=*Ophiophagus hannah*
PE=2 SV=1
MKNLLLTFLVVTIVCLDLGYTLICHQVHGLQTCEPAQKFCQIRTTMFFPNHPVLLMGCTYNCPTERY
SVCCSTDKCNK
Mave red: 6540.66
>sp|Q53B47|NXS5_OPHHA Short neurotoxin OH-5 OS=*Ophiophagus hannah*
PE=1 SV=1
MKNLLLTFLVVTIVCLDLGYTLICHRVHGLQTCEPDQKFCFRKTTMFFPNHPVLLMGCTYSCPTEKY
SVCCSTDKCNK
Mave red: 6591.75
>sp|Q2VBP2|NXS6_OPHHA Short neurotoxin SNTX6 OS=*Ophiophagus hannah*
PE=2 SV=1
MKTLLLTFLVVTIVCLDLGYTLICHQLHGLQTCEPAQKFCQKRTTMFSPNHPVLLMGCTYNCPTERY
SVCCSTDKCNK
Mave red: 6509.60
>sp|Q2VBN1|TXM6_OPHHA Muscarinic toxin MTX6 OS=*Ophiophagus hannah*
PE=2 SV=1
MKTLLLTLVVVTILCLDLGYTLTCLTHESLFFETTETCSDGQNLCYAKWFAVFPGARRPDGGCAATC
PDKVPLEIVNCCTTDKCNL
Mave red: 7107.10
>sp|Q69CJ8|TXW1H_OPHHA Weak toxin DE-1 homolog 1 OS=*Ophiophagus
hannah* PE=2 SV=1
MKPVLLTLVVVTIVCLDLGYTRICLKQEPFQPETTTTCPEGEDACYNLFWSDHSEIKIEMGCGCPKT
EPYTNLYCCKIDSCNK
7
Mave red: 7114.06
>gi|82193164|sp|Q53B61.1|TXW72_OPHHA RecName: Full=Weak neurotoxin
OH-72; Flags: Precursor
LTLVVVTIVCLDLGYTLTCLICPEEYCKRIHTCRDGENVCFKGFYEGKQLGKQFRRGCAATCPEGKP
NEIVQCCSTDECNH
Mave red: 7362.42

>gi|238734464|gb|ACR55626.1| long chain neurotoxin precursor
[*Ophiophagus hannah*]
MKTLLLTLVVVTIVCLDLGYTLLCYKTPSPINAETCPPGENLCYTKMWCDAWCSSRGKVIELGCAAT
CPSKKPYEEVDCCSTDNCNPHPKLRP
Mave red: 7969.19
>gi|1584763|prf||2123384A phospholipase A2
DLIQFGNMIQCTVPGFLSWIKYADYGCYCGAGGSGTPVDKLDRCCQVHDNCYTQAQKLPACSSIMDS
PYVKIYSYDESCRVAVTCKADNDECAAFICNCDRVAAYCFAASPYNNNNYNIDT(TTRC)
Mave red: 13806.51
>gi|229634|prf||771760A polypeptide DE1
LICFNQETYRPETTTTCPDGENCYSTFWHNDGHVKIERGCGCPRVNPPISICCKTDKCNN
Mave red: 6811.64
>gi|ETE69273|Cystatin-B [*Ophiophagus hannah*]
MRCLVLTSSGYAPNRGQLIQGLDCEGEETLSLWFAQPVVAATSFPSPAMSCGGLSEPKAATAETQQI
TQEIKSQLEEKESRNFDIFNAVSYKTQVVAGINYFIKIHVGNDEYFHVRVYKRLPHENKPVELTNYQ
SKKEKHEDLTYF
Mave red: 9559.71 (Mave nat: 9557.71)
>sp|D9IX98|66-104
MVGPYNPAGGGGGHPSSCFGHKIDRISHSSGMGCGRRPN
Mmono red: 3937.99
>sp|D9IX98|157-192
ELAKKDQHNNCFGRRIDRISHSTDLGCRRRPNPPPAPTAAPLAVA
Mave red: 4124.68
>sp|D9IX98|154-190
QQELAKKDQHNNCFGRRIDRISHSTDLGCRRRPNPPP
Mmono red: 4339.18 (Mmono nat: 4337.18)
>sp|ETE58964| hypothetical protein
MKTLLLTLVVVTIVCLDLGHTRICLTDYSKVSETIEICPDGQNFCFKKFPKGIPFLPWVNRGCAATC
PKPEPKVYVDCCARDKCNR
Mave red: 7401.68
8
> sp|ETE61012| Insulin-like growth factor I
MLTATSIHFFYFGLCLLTLTGSSVAAAGHETLCGAELVDALQFVCKPAGYGGNRRPSSSRGIVEECC
FQSCDLVRLEMYCAPVKPPPKSARSLRAQRHTDMPKAQKVAKVLSNLI
Mave red: 7590.82
>sp|A3R0T9| Zinc metalloproteinase-disintegrin-like VM3_OPHHA
MIQVLLVTICLVVFPYQGSSIILESGKVNDYEVVYPQKIPVLPKSKIQRREQKMYEDTMKYEFKVNG
EPVVLHLERNKELFSKDYTETHYSPDGREITTSPPVEDHCYYHGYIQSDIDSTAILNACNGLKGYFR
HHGEAYHIEPLKFSDSEAHAVYKYENIEKEDETPKICGVKHSTWESDEPIEKISQKKDFLEEKKYLE
LYIVADYVMFRKYGRNVTTIRMRVFDMVNYITVVYKALNIHVALIGFEIWSLKDKFVINASTKNNLL
HFSIWRSTVLRKRNDNAQLLTGVDLNGYTLGSAYLKAMCDVLQSVGIVQDYSKSPYLVGAAMAHEIG
HNLGMEHDTKTCSCMRGNCIMSPEEEGSDFPMEFSSCSLYDFQNYMLTDTPQCLINKPSNTSIIKNA
VCGNYVEEEGEECDCGSPEQCENNCCEAATCKLKPGAKCAKGACCKKCQFKKAGAECRAARNECDLP
EFCIGQSAECPMDRFHKNGHSCQNDQGYCFRGYCPTLAKQCITLWGSDAKVAPDECFQNNTNGNEYD
YCKKTNNVIIPCKPTDVKCGRLYCTGGTENPSEGEKISSDPCKASYSEIEDIGMVDHRTKCGEKMVC
SDGKCIPL
>sp|ETE62121|hypothetical Zinc metalloproteinase-disintegrin-like
L345_12124
MIQVLLVTICLVVFPYQGSSIILESGKVNDYEVVYPQKIPVLPKSKIQRREQKMYEDTMKYEFKVNG
EPVVLHLERNKELFSKDYTETHYSPDGREITTSPPVEDHCYYHGYIQSDIDSTAILNACNGLKHHGE
AYHIEPLKFSDSEAHAVYKYENIEKEDETPKICGVKHSTWESDEPIEKISQVSITSEEFRKYSRNVT
AIRMRVFDMVNYITVVYKALNIRVALIGFEIWSLKDKFVINASTKNNLLHFSIWRSTVLRKRNDNAQ
LLTGVDLNGYTLGSAYLKAMCDVLQSVGIVQDYSKSPYLVGAAMAHEIGHNLGMEHDTKTCSCMRGN
CIMSPEEEGSDFPMEFSSCSLYDFQNYMLTETPQCLINKPSNTSIIKNAVCGNYVEEEGEECDCGSP
EQCENNCCEAATCKLKPGAKCAKGACCKKCQFKKAGAECRAARNECDLPEFCIGQSAECPMDRFHKN
GHSCQNNQGYCFRGYCPTLAKQCITLWGSDAKVAPDECFQNNTNGNEYDYCKKTNNVIIPCKPT
>sp|P81383| L-amino-acid oxidase OXLA_OPHHA

CXL

MNDFLLLLLVLFLGVPRSENHVINLEECFQEPEYENWLATASHGLTKTLNPKKIVIVGAGISGLTAA
KLFREAGHEVVILEASDRVGGRIKTHREDGWYVDVGPMRVPQTHRIVREYIKKFNISLNPFRQTDEN
AWYLIKHVRQKMSANNPENFGYQLNPNERGKSASQLFDETLDKVTDDCTLQKEKYDSFSTKEYLIKE
GKLSTGAVEMIGDFLNEEAGFHNSFLISVMDHFLFLNNSFDEITGGFDQLPERFFKDMDSIVHLNST
VEKIVHINNKVTVFYEGLSTNMRLVADYVLITATARATRLIKFVPPLSIPKTRALRSLIYASATKII
LVCTDKFWEKDGIHGGRSITDLPSRVIYYPNHDFTNGIGVLLASYTWYSDSEFYTTLSDEKCVDVVM
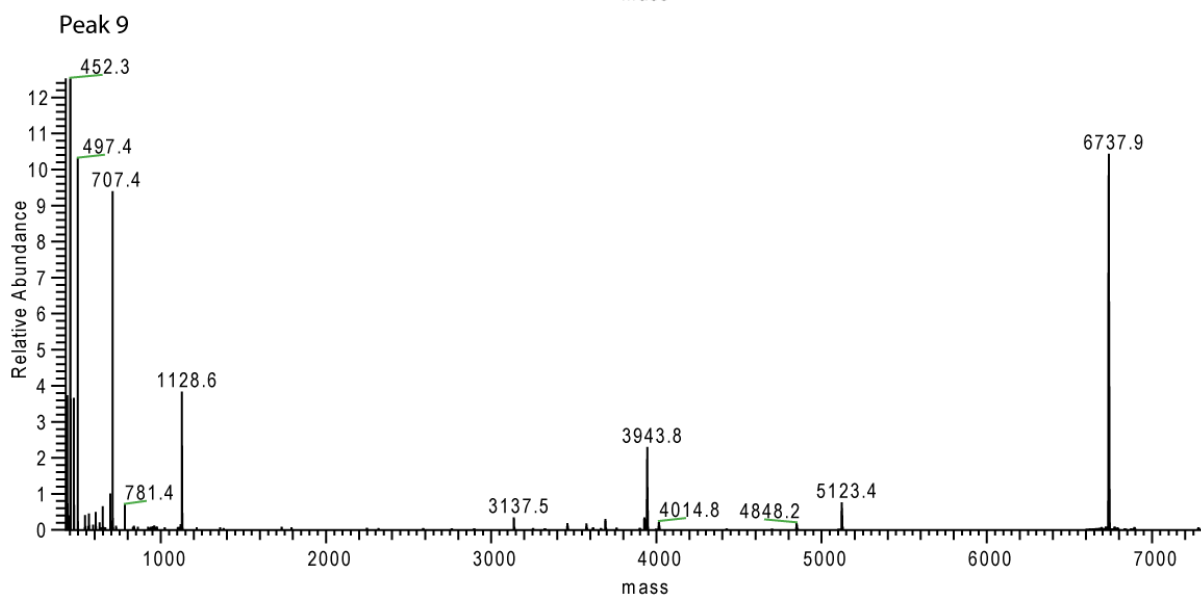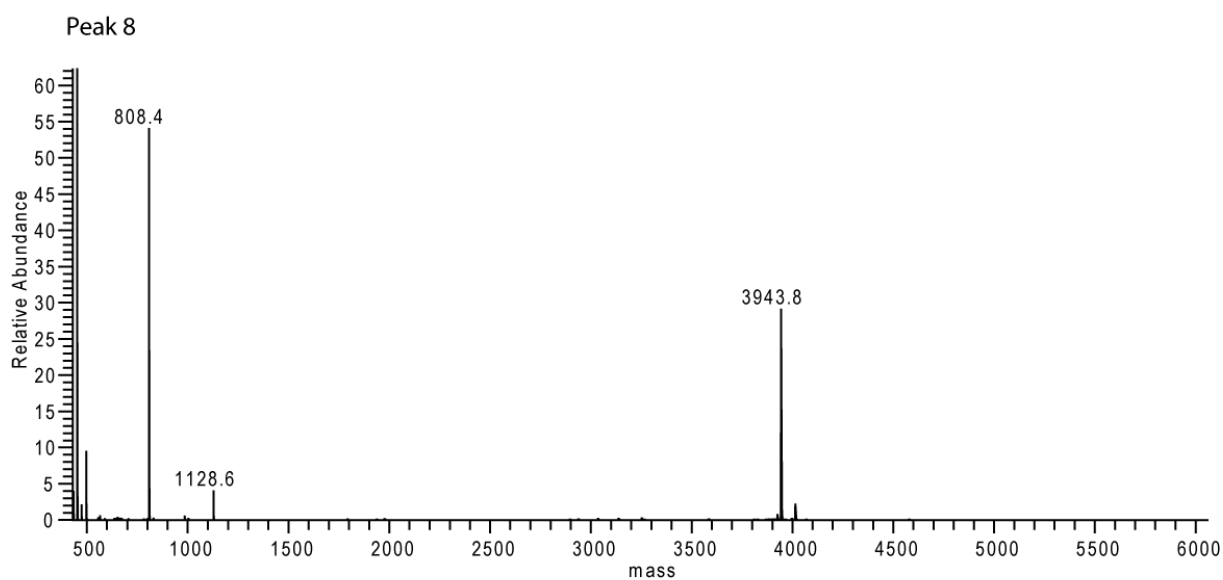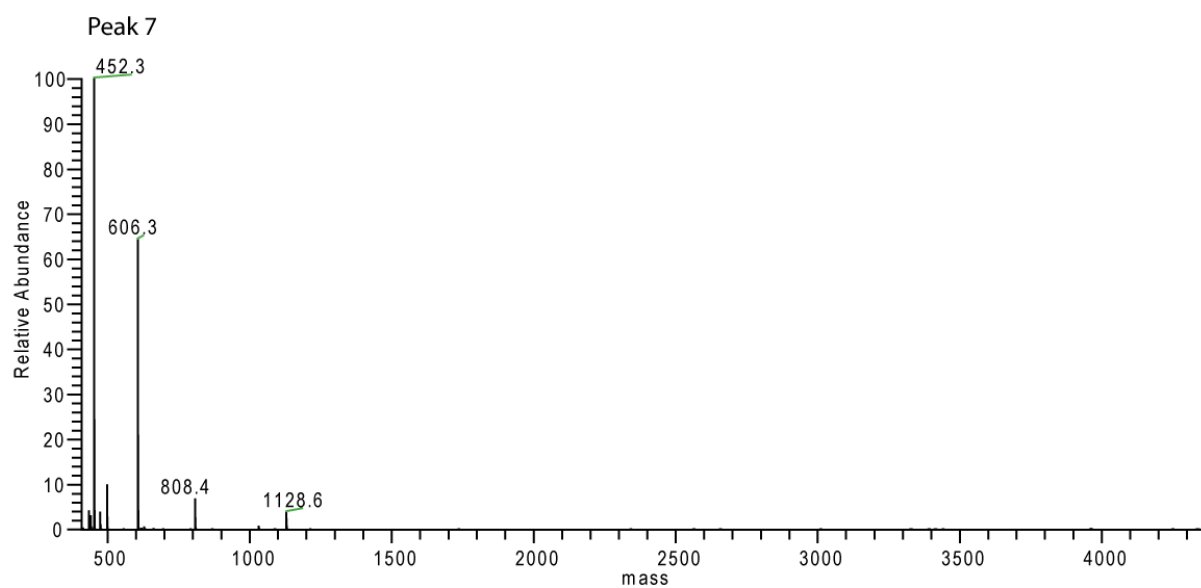DDLVEIHNVSKDYLKSVCGKHVVQKWALDQYSMGAFSTYTPYQITHYSQMLAQNEGRIYFAGEYTAH
PHGWIETSMKSAIREAINIHNA

D   Appendix to Chapter 6

Peak 1

Peak 2
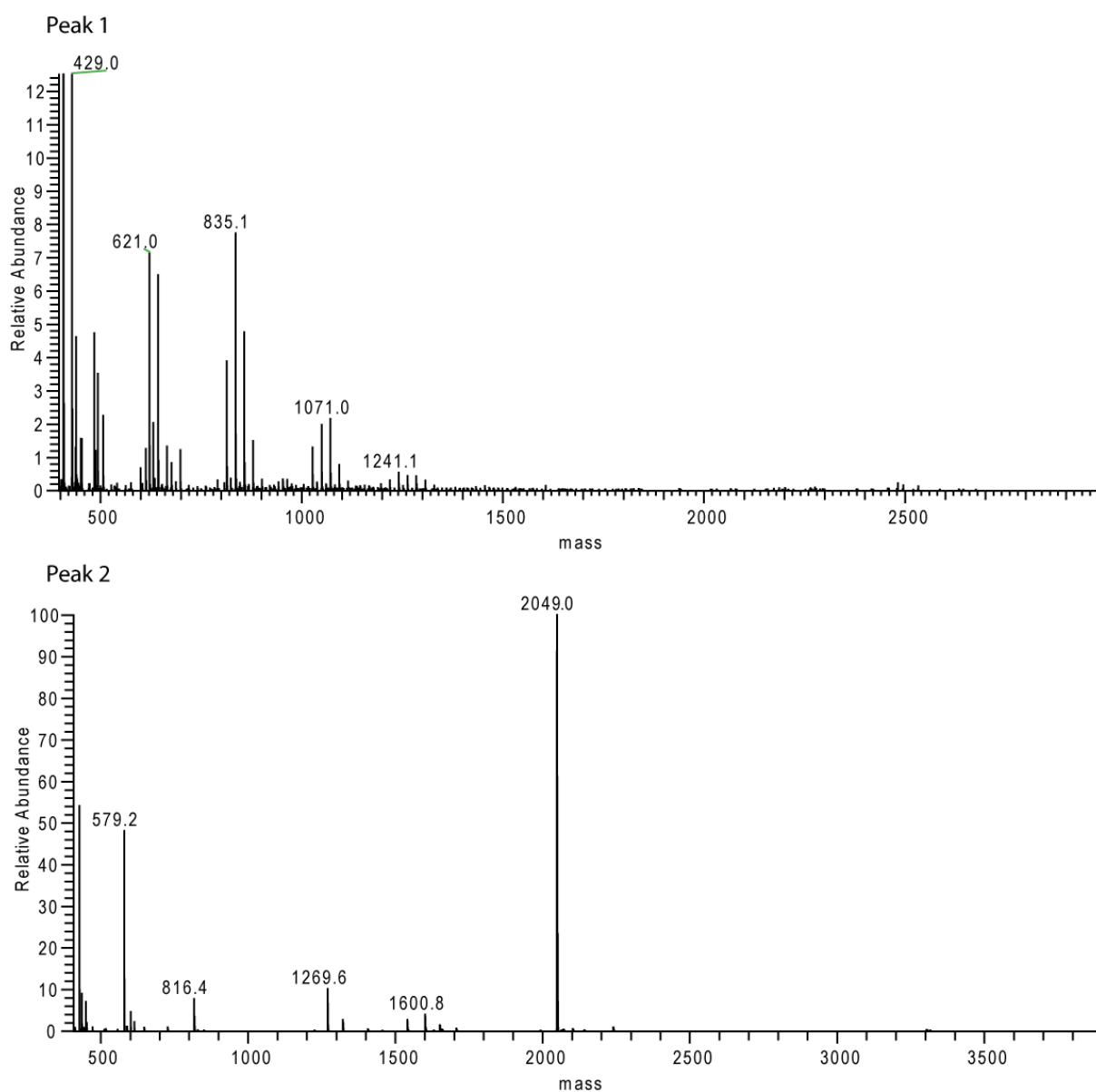
Supplemental Figure 16: Deconvoluted top-down mass spectra from *Vipera anatolica* (Peaks 1-34)
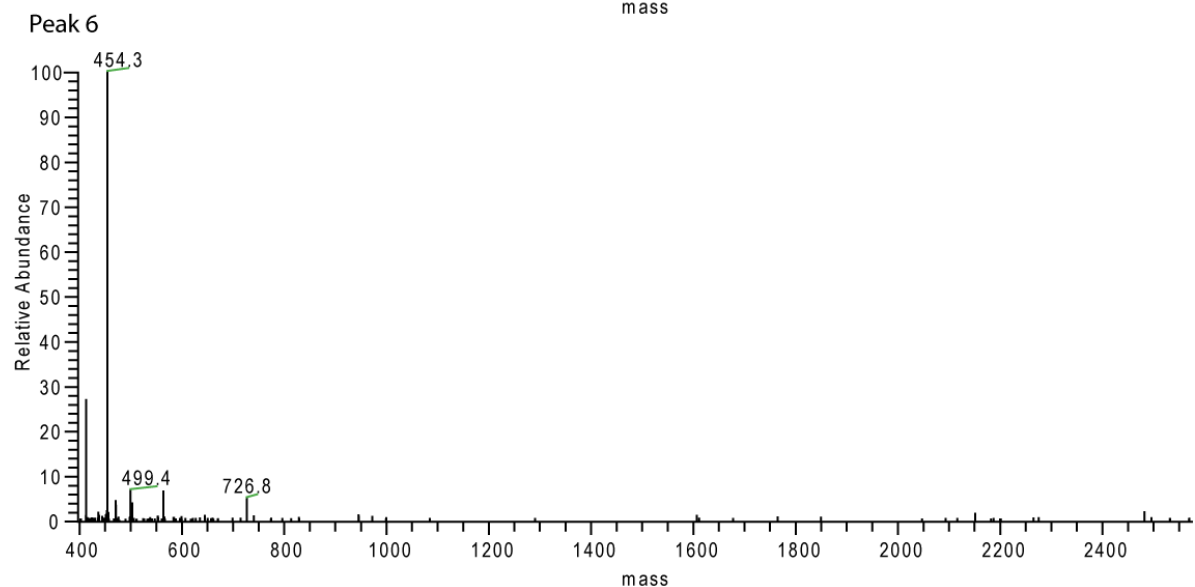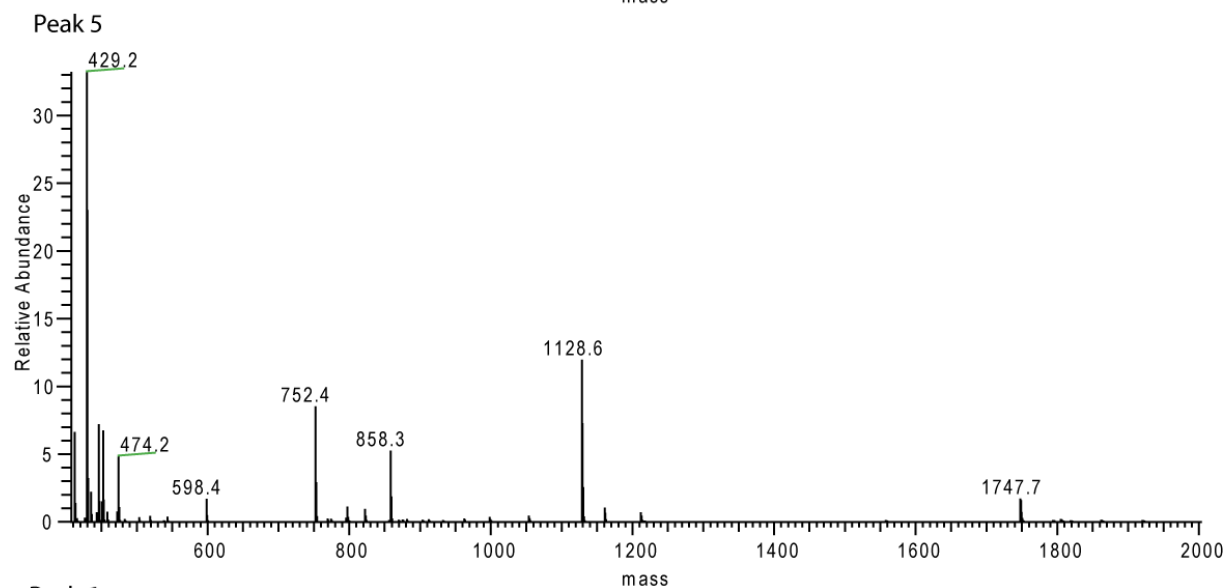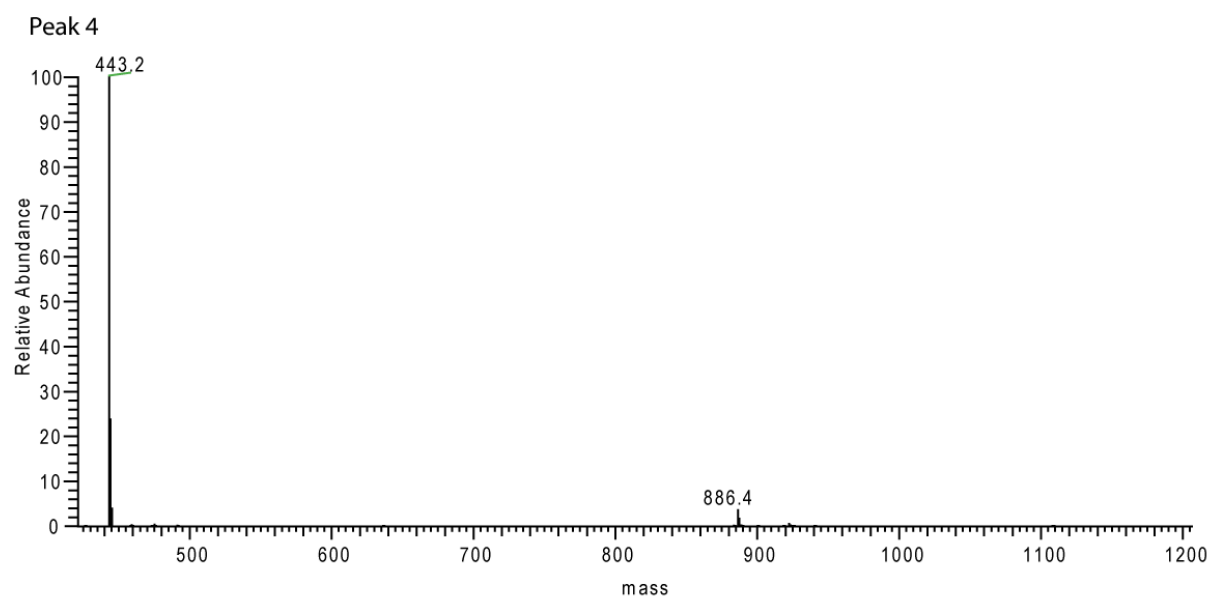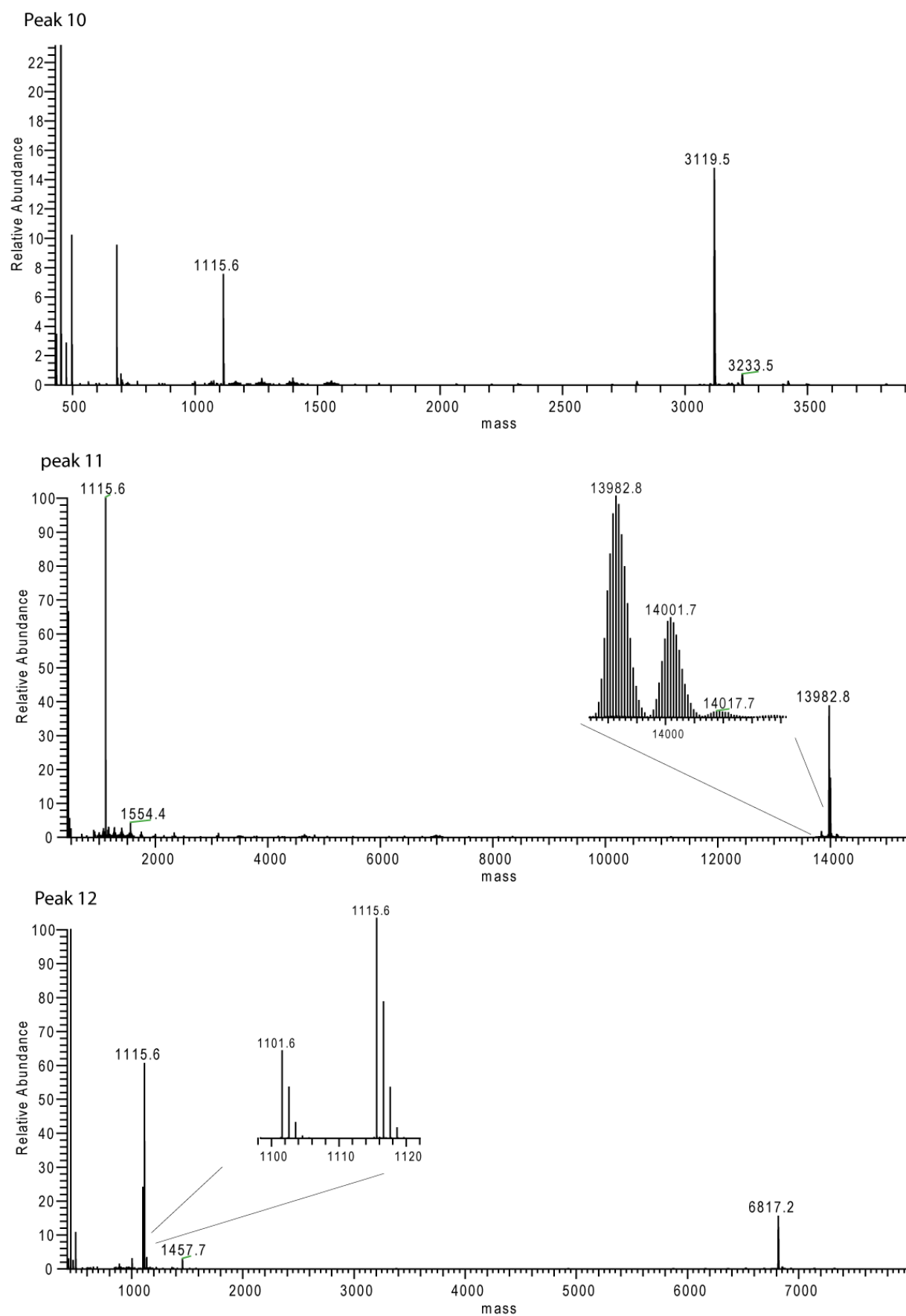
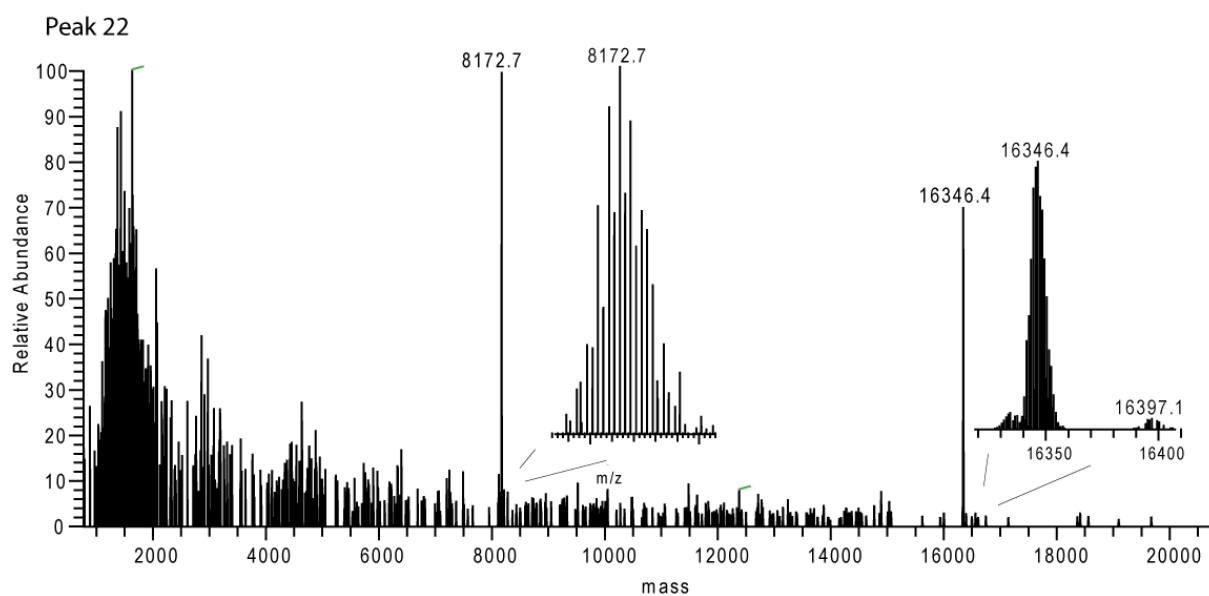Peak 4

Peak 5

Peak 6

Supplemental Figure 16 continued

Peak 7

Peak 8

Peak 9

Supplemental Figure 16 continued

Peak 10

peak 11

Peak 12

Supplemental Figure 16 continued

CXLVI

Peak 13

Peak 14

Peak 15

Supplemental Figure 16 continued

Peak 16

Peak 17

Peak 18

Supplemental Figure 16 continued

Peak 19

Peak 20/21
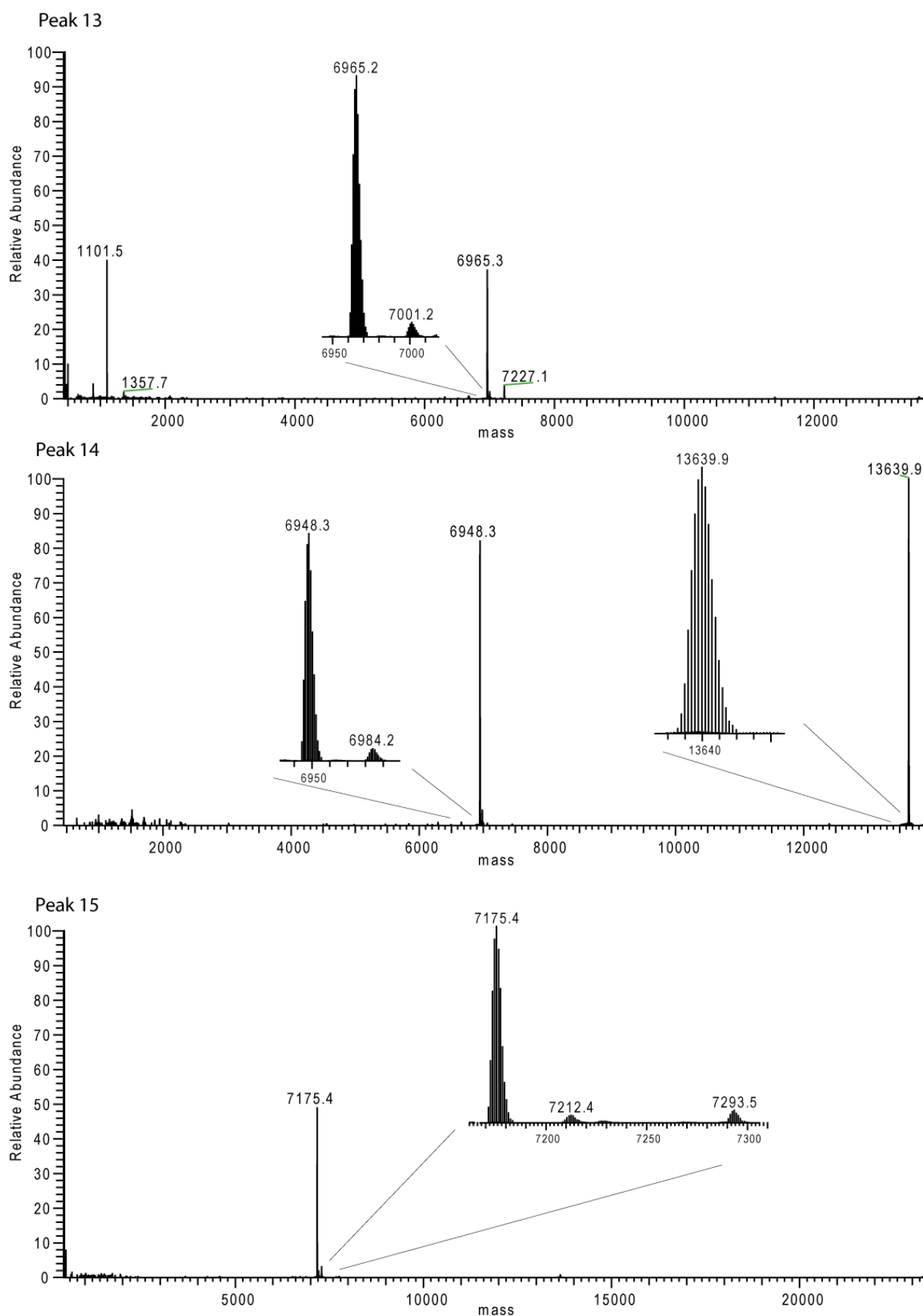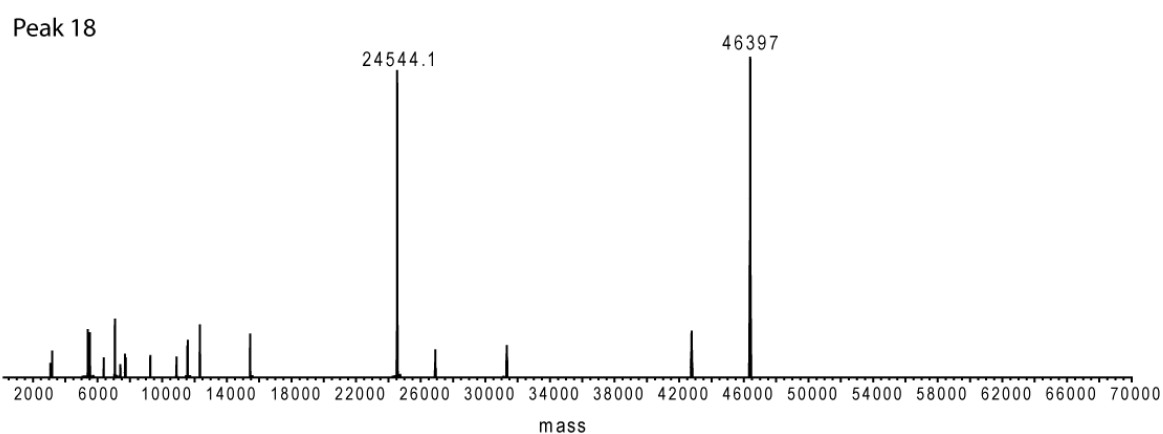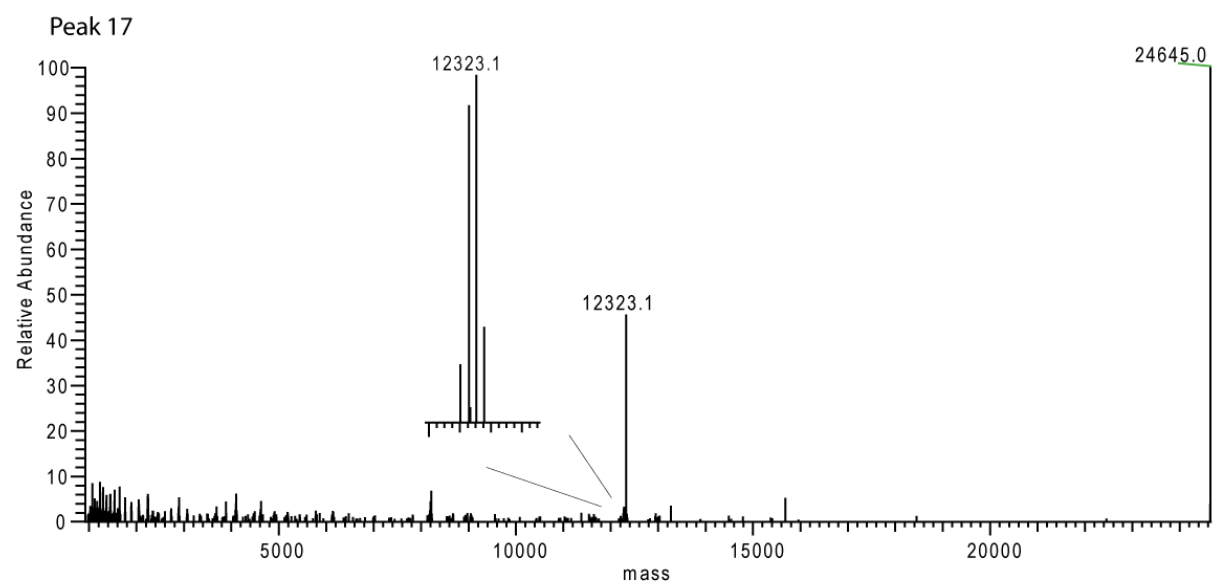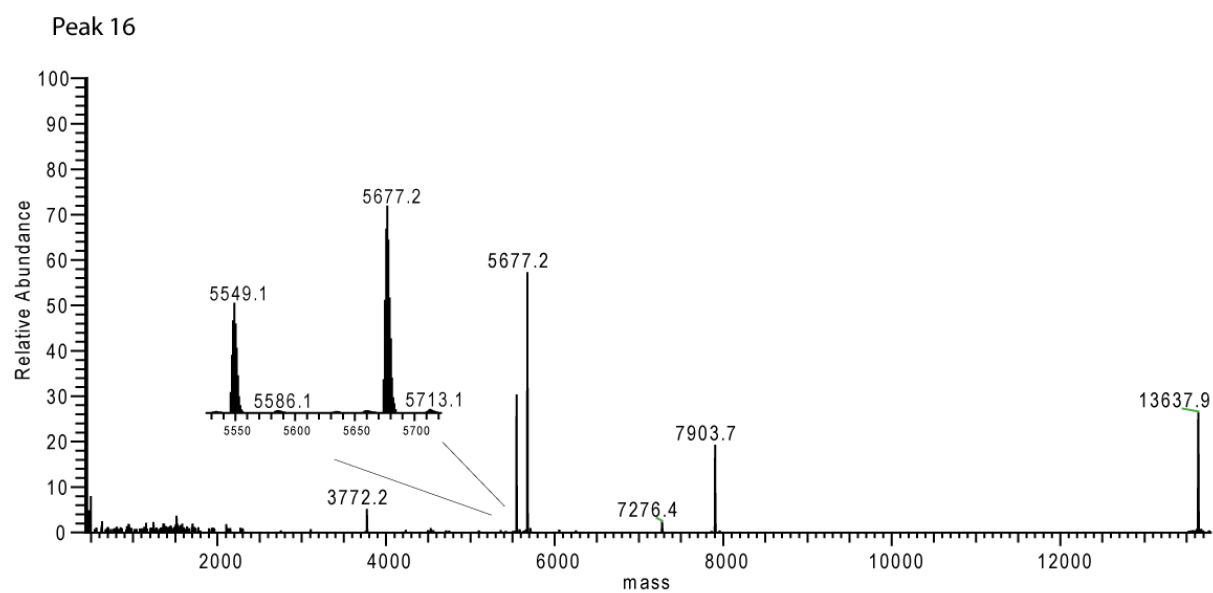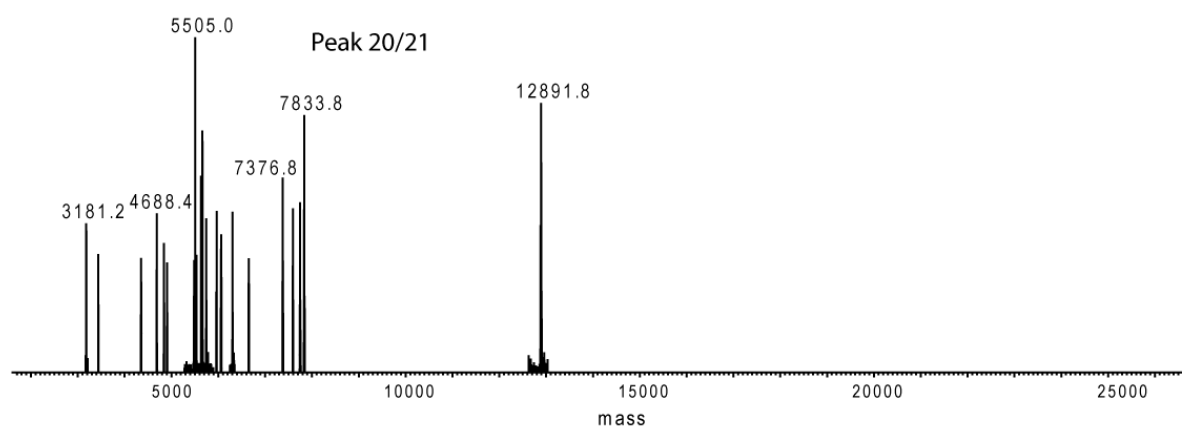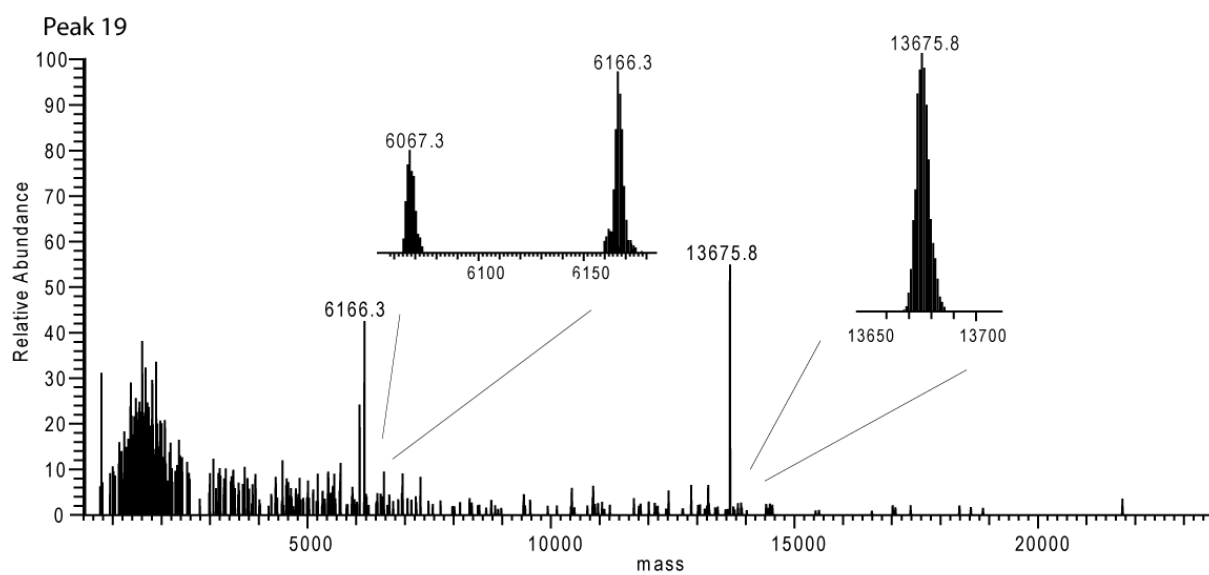
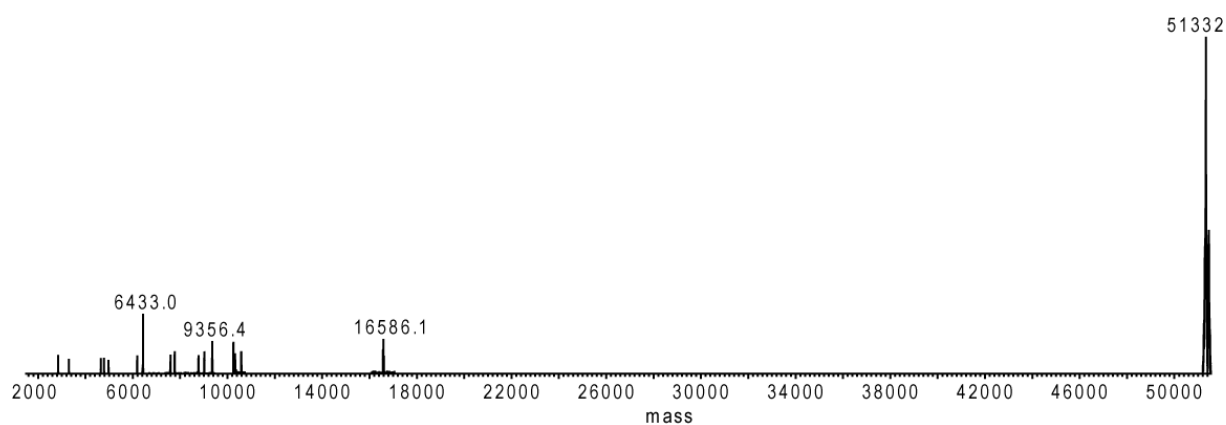
Peak 22

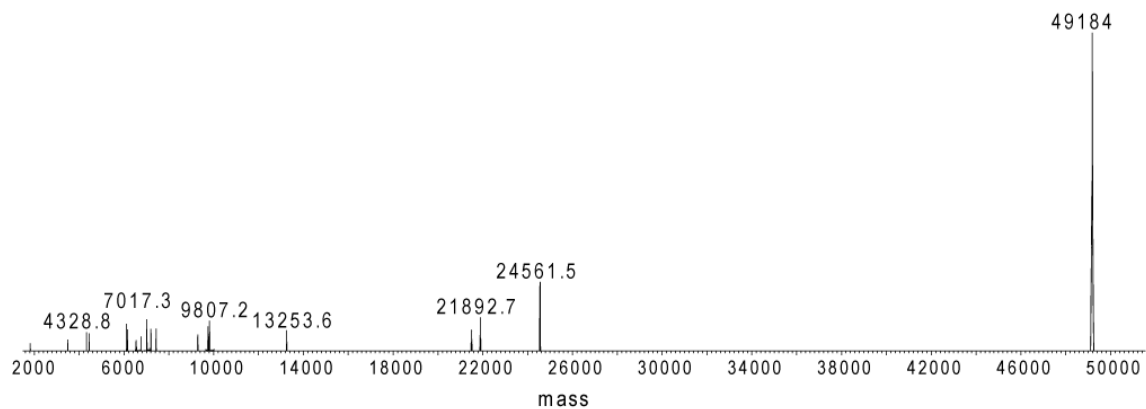**Supplemental Figure 16 continued**

Peak 23 -25
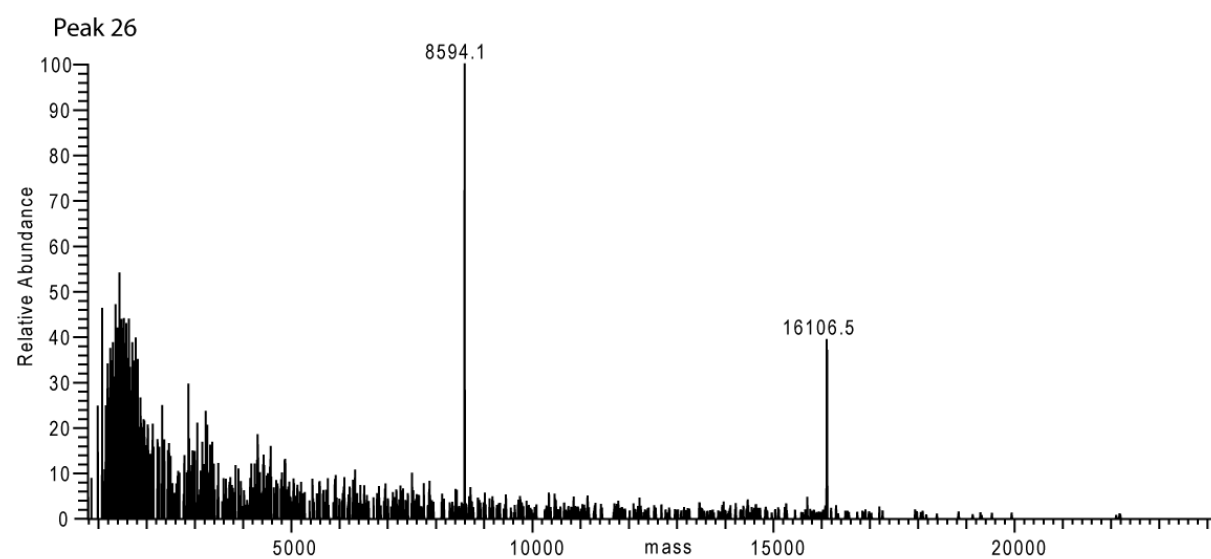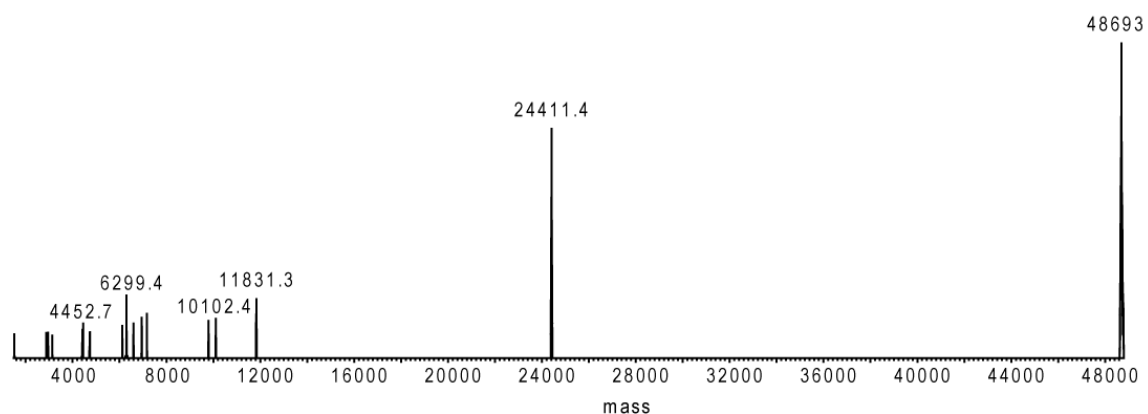
Peak 26

Peak 27

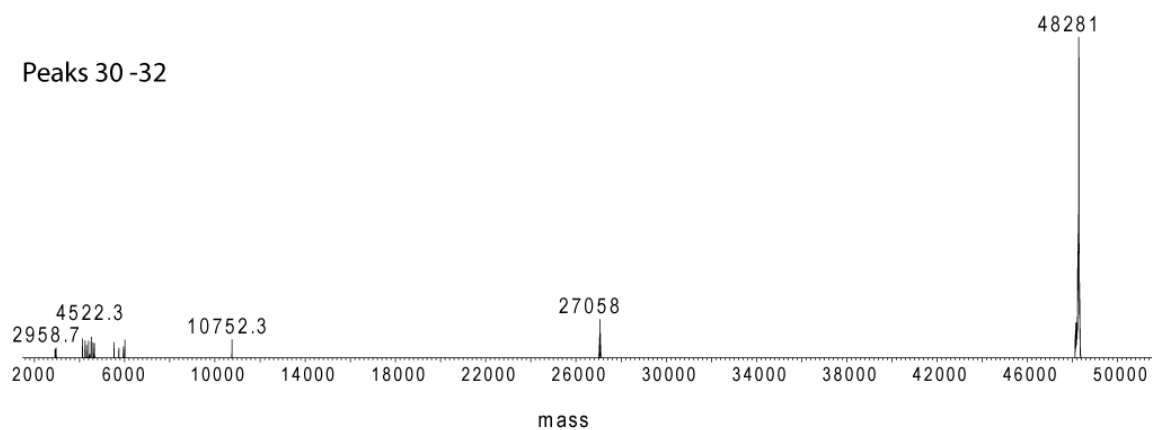**Supplemental Figure 16 continued**

CL

Peak 28-29

Peaks 30 -32

Peak 33
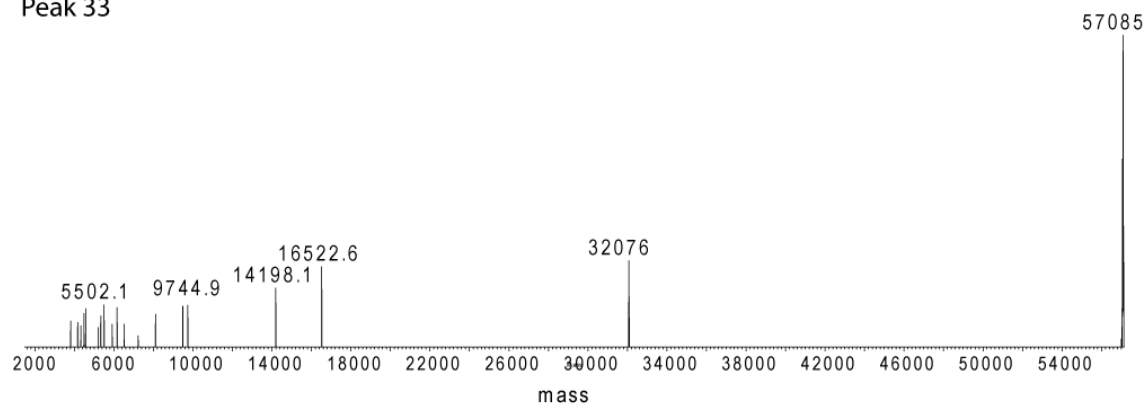
Peak 34

Supplemental Figure 16 continued

Peak 28-29

49184

9807.2 13253.6 21892.7 24561.5

mass
2000 6000 10000 14000 18000 22000 26000 30000 34000 38000 42000 46000 50000

Peaks 30 -32

48281

10752.3 27058

mass
2000 6000 10000 14000 18000 22000 26000 30000 34000 38000 42000 46000 50000

Peak 33

57085

14198.1 16522.6 32076

mass
2000 6000 10000 14000 18000 22000 26000 30000 34000 38000 42000 46000 50000 54000

Peak 34

27603

10669.6 26398

mass
2000 6000 10000 14000 18000 22000 26000 30000 34000 38000 42000 46000 50000 54000

Peaks 35 -36

25225

5217.2 7579.7 13594.2 32257

mass
2000 6000 10000 14000 18000 22000 26000 30000 34000 38000 42000 46000 50000 54000

Supplemental Figure 16 continued

CLII

Peak 4 - 443.2 Da
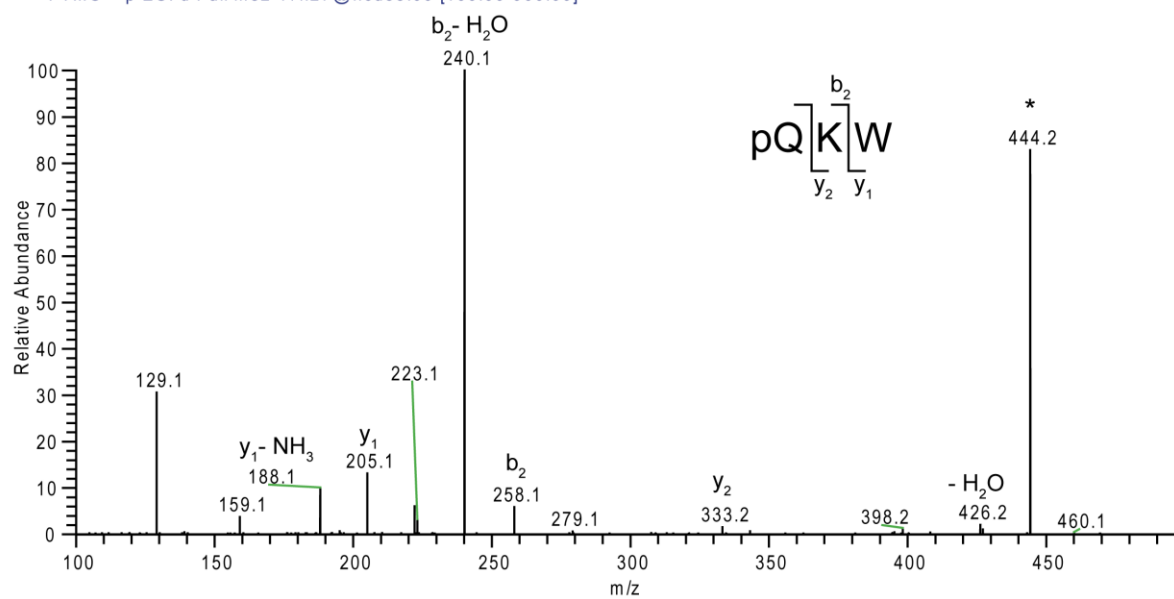
Supplemental Figure 17: MS and MSMS spectrum of the snake venom metalloprotease inhibitor.

Peak 14 - 5461.6 Da
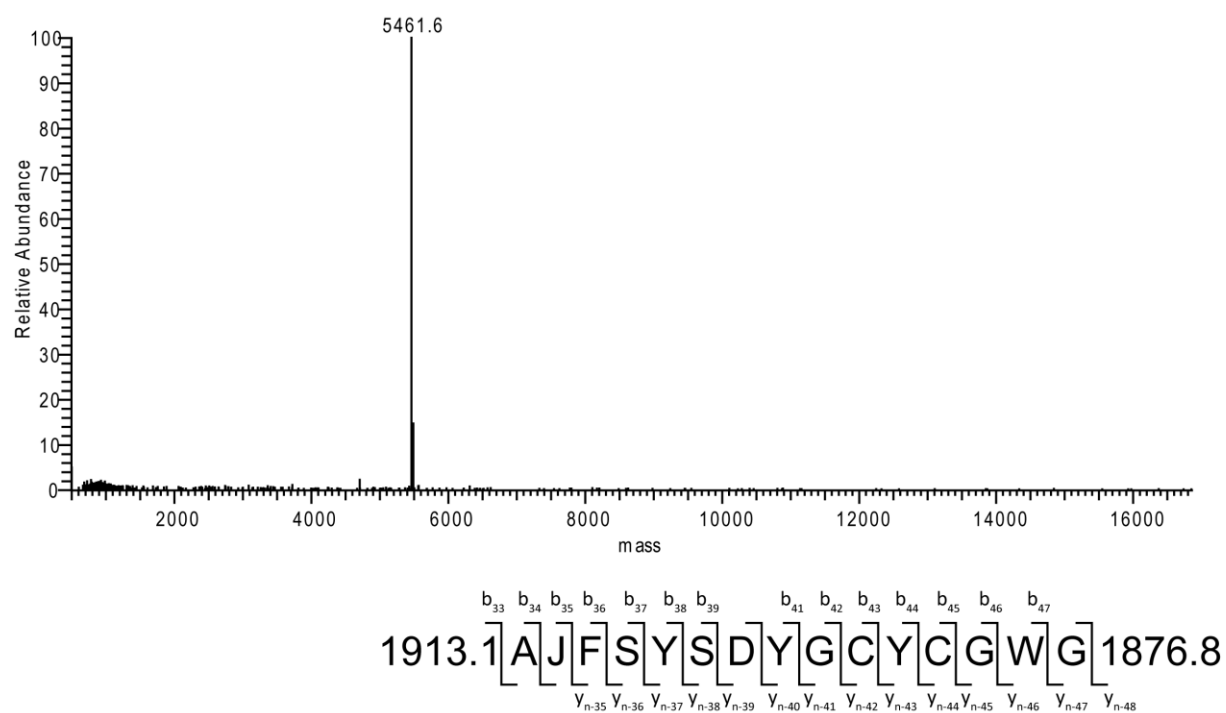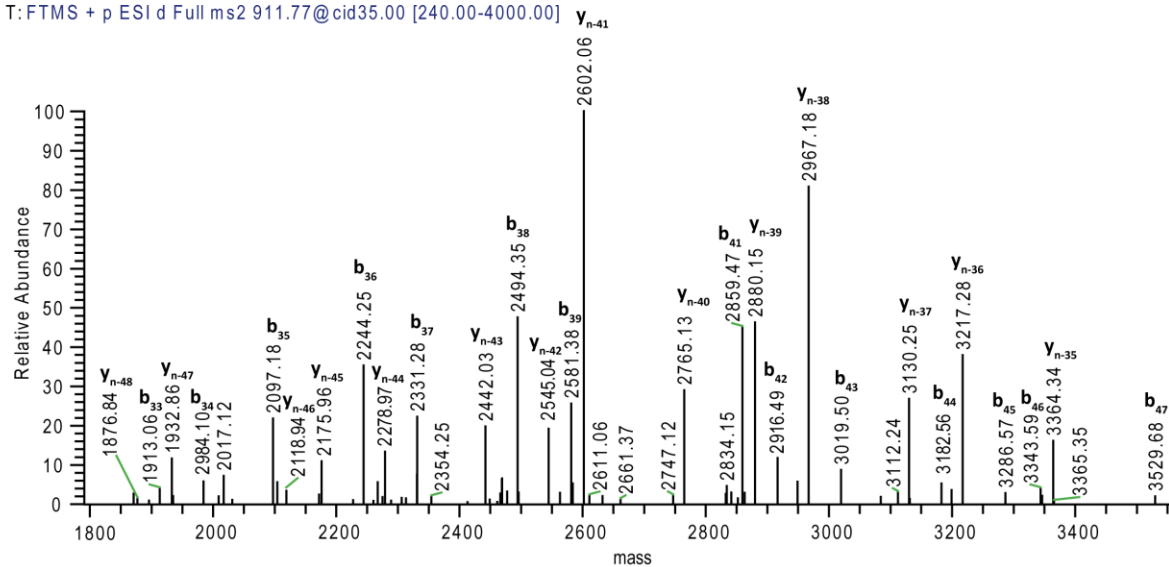
Supplemental Figure 18: Deconvoluted MS and MSMS spectrum (zoomed) of a PLA2 fragment.
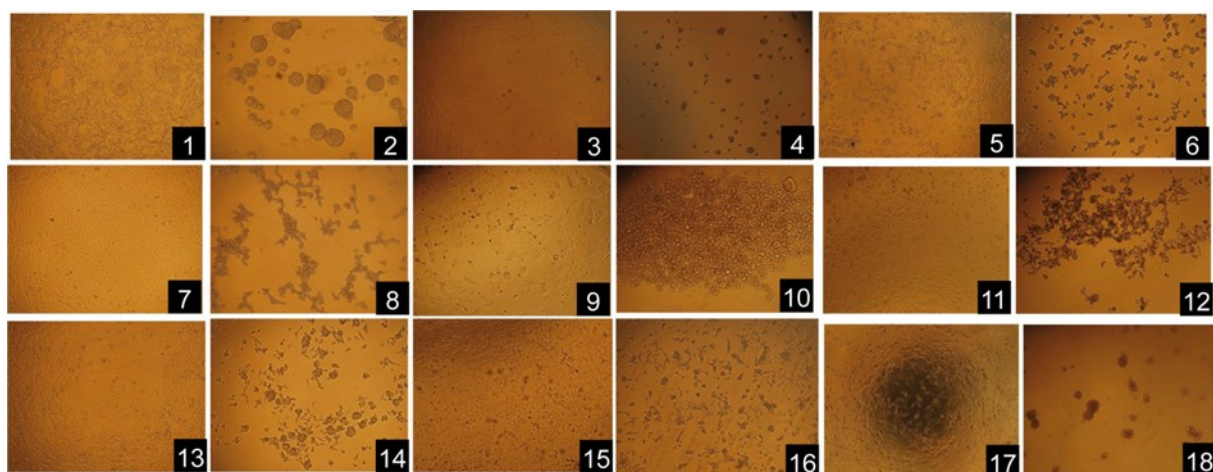
Supplemental Figure 19 Effect of crude venom on various cancer and non-cancerous cell lines. Cells were treated with crude venom for 48 h at 37° C. **1**: HEK293 untreated, **2:** HEK293, 10 µg/ml, **3:** Vero, untreated, **4:** Vero, 10 µg/ml, **5:** MPanc-96 untreated, **6:** MPanc-96, 10 µg/ml, **7:** A549, untreated, **8:** A549, 10 µg/ml, **9:** CACO-2, untreated, **10:** CACO-2, 10 µg/ml, **11:** PC-3, untreated, **12:** PC-3, 10 µg/ml, **13:** HeLa, untreated, **14:** HeLa, 10 µg/ml, **15:** MCF-7, untreated, **16:** MCF-7, 10 µg/ml, **17:** U87MG, untreated, **18:** U87MG, 10 µg/ml.

References

1. Brade, H., *Endotoxin in health and disease*1999: CRC Press.
2. Calvete, J.J., *Next-generation snake venomics: protein-locus resolution through venom proteome decomplexation.* Expert review of proteomics, 2014. **11**(3): p. 315-329.
3. Cimmino, A., M. Masi, M. Evidente, and A. Evidente, *Fungal Phytotoxins with Potential Herbicidal Activity to Control Chenopodium album.* Natural product communications, 2015. **10**(6): p. 1119-1126.
4. Reid, D.S. and L.J. Harris, *Microorganisms and microbial toxins.* Advances in experimental medicine and biology, 1999. **459**: p. 9-21.
5. Yamaguchi, Y., J.H. Park, and M. Inouye, *Toxin-antitoxin systems in bacteria and archaea.* Annual review of genetics, 2011. **45**: p. 61-79.
6. Vetter, I., J.L. Davis, L.D. Rash, R. Anangi, M. Mobli, P.F. Alewood, R.J. Lewis, and G.F. King, *Venomics: a new paradigm for natural products-based drug discovery.* Amino acids, 2011. **40**(1): p. 15-28.
7. King, G.F., *Venoms as a platform for human drugs: translating toxins into therapeutics.* Expert opinion on biological therapy, 2011. **11**(11): p. 1469-1484.
8. Calvete, J.J., *Venomics: digging into the evolution of venomous systems and learning to twist nature to fight pathology.* Journal of proteomics, 2009. **72**(2): p. 121-126.
9. McCleary, R.J. and R.M. Kini, *Non-enzymatic proteins from snake venoms: a gold mine of pharmacological tools and drug leads.* Toxicon : official journal of the International Society on Toxinology, 2013. **62**: p. 56-74.
10. Kalia, J., M. Milescu, J. Salvatierra, J. Wagner, J.K. Klint, G.F. King, B.M. Olivera, and F. Bosmans, *From foe to friend: using animal toxins to investigate ion channel function.* Journal of molecular biology, 2015. **427**(1): p. 158-175.
11. Trautmann, A., *Curare can open and block ionic channels associated with cholinergic receptors.* Nature, 1982. **298**(5871): p. 272-275.
12. Harvey, A.L., *Toxins and drug discovery.* Toxicon : official journal of the International Society on Toxinology, 2014. **92**: p. 193-200.
13. Griffith, H.R. and G. ENID, *The use of curare in general anesthesia.* Survey of Anesthesiology, 1985. **29**(6): p. 361.
14. Dewar, G.H., I.G. Marshall, S.S. Patel, and R. Waigh, *Chemodegradable neuromuscular blocking agents-I: bis 3, 4-dihydroisoquinolinium salts.* Pharmaceutical Science Communications, 1993. **4**: p. 3-3.
15. Stenlake, J., *Chance, coincidence and atracurium.* Pharm J, 2001. **267**(7167): p. 430-441.
16. Organization, W.H., *WHO model list of essential medicines: 18th list, April 2013.* 2013.

17. Endo, Y., K. Mitsui, M. Motizuki, and K. Tsurugi, *The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. The site and the characteristics of the modification in 28 S ribosomal RNA caused by the toxins.* The Journal of biological chemistry, 1987. **262**(12): p. 5908-5912.

18. Lord, M.J., N.A. Jolliffe, C.J. Marsden, C.S. Pateman, D.C. Smith, R.A. Spooner, P.D. Watson, and L.M. Roberts, *Ricin. Mechanisms of cytotoxicity.* Toxicological reviews, 2003. **22**(1): p. 53-64.

19. Uckun, F.M., L.M. Chelstrom, L. Tuel-Ahlgren, I. Dibirdik, J.D. Irvin, M.C. Langlie, and D.E. Myers, *TXU (anti-CD7)-pokeweed antiviral protein as a potent inhibitor of human immunodeficiency virus.* Antimicrobial agents and chemotherapy, 1998. **42**(2): p. 383-388.

20. Falini, B., A. Bolognesi, L. Flenghi, P.L. Tazzari, M.K. Broe, H. Stein, H. Durkop, F. Aversa, P. Corneli, G. Pizzolo, and et al., *Response of refractory Hodgkin's disease to monoclonal anti-CD30 immunotoxin.* Lancet, 1992. **339**(8803): p. 1195-1196.

21. Stirpe, F., S. Olsnes, and A. Pihl, *Gelonin, a new inhibitor of protein synthesis, nontoxic to intact cells. Isolation, characterization, and preparation of cytotoxic complexes with concanavalin A.* The Journal of biological chemistry, 1980. **255**(14): p. 6947-6953.

22. Shapira, A. and I. Benhar, *Toxin-based therapeutic approaches.* Toxins, 2010. **2**(11): p. 2519-2583.

23. Wagner, P.L. and M.K. Waldor, *Bacteriophage control of bacterial virulence.* Infection and Immunity, 2002. **70**(8): p. 3985-3993.

24. Turturro, F., *Denileukin diftitox: a biotherapeutic paradigm shift in the treatment of lymphoid-derived disorders.* Expert review of anticancer therapy, 2007. **7**(1): p. 11-17.

25. Jiang, Y.H., C.H. Liao, and H.C. Kuo, *Current and potential urological applications of botulinum toxin A.* Nature reviews. Urology, 2015. **12**(9): p. 519-533.

26. Karsai, S., R. Adrian, S. Hammes, J. Thimm, and C. Raulin, *A randomized double-blind study of the effect of Botox and Dysport/Reloxin on forehead wrinkles and electromyographic activity.* Archives of dermatology, 2007. **143**(11): p. 1447-1449.

27. Kuehne, S.A., S.T. Cartman, J.T. Heap, M.L. Kelly, A. Cockayne, and N.P. Minton, *The role of toxin A and toxin B in Clostridium difficile infection.* Nature, 2010. **467**(7316): p. 711-713.

28. Stuckey, D.W., S.D. Hingtgen, N. Karakas, B.E. Rich, and K. Shah, *Engineering toxin-resistant therapeutic stem cells to treat brain tumors.* Stem cells, 2015. **33**(2): p. 589-600.

29. Zarins-Tutt, J.S., T.T. Barberi, H. Gao, A. Mearns-Spragg, L. Zhang, D.J. Newman, and R.J. Goss, *Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products.* Natural product reports, 2015.

30. Jestoi, M., *Emerging fusarium-mycotoxins fusaproliferin, beauvericin, enniatins, and moniliformin: a review.* Critical reviews in food science and nutrition, 2008. **48**(1): p. 21-49.

31. Sy-Cordero, A.A., C.J. Pearce, and N.H. Oberlies, *Revisiting the enniatins: a review of their isolation, biosynthesis, structure determination and biological activities.* The Journal of antibiotics, 2012. **65**(11): p. 541-549.

32. Levy, D., A. Bluzat, M. Seigneuret, and J.L. Rigaud, *Alkali cation transport through liposomes by the antimicrobial fusafungine and its constitutive enniatins.* Biochemical pharmacology, 1995. **50**(12): p. 2105-2107.

33. Fleming, A., *On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae.* British journal of experimental pathology, 1929. **10**(3): p. 226.

34. Haight, T.H. and M. Finland, *Laboratory and clinical studies on erythromycin.* The New England journal of medicine, 1952. **247**(7): p. 227-232.

35. Griffith, R.S. and F.B. Peck, Jr., *Vancomycin, a new antibiotic. III. Preliminary clinical and laboratory studies.* Antibiotics annual, 1955. **3**: p. 619-622.

36. Boucher, H.W., G.H. Talbot, D.K. Benjamin, Jr., J. Bradley, R.J. Guidos, R.N. Jones, B.E. Murray, R.A. Bonomo, and D. Gilbert, *10 x '20 Progress--development of new drugs active against gram-negative bacilli: an update from the Infectious Diseases Society of America.* Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 2013. **56**(12): p. 1685-1694.

37. Lewis, K., *Platforms for antibiotic discovery.* Nature reviews. Drug discovery, 2013. **12**(5): p. 371-387.

38. Beveridge, T.J., *Structures of gram-negative cell walls and their derived membrane vesicles.* Journal of bacteriology, 1999. **181**(16): p. 4725-4733.

39. Medema, M.H. and M.A. Fischbach, *Computational approaches to natural product discovery.* Nature chemical biology, 2015. **11**(9): p. 639-648.

40. Sharon, G., N. Garg, J. Debelius, R. Knight, P.C. Dorrestein, and S.K. Mazmanian, *Specialized metabolites from the microbiome in health and disease.* Cell metabolism, 2014. **20**(5): p. 719-730.

41. Donia, M.S., P. Cimermancic, C.J. Schulze, L.C. Wieland Brown, J. Martin, M. Mitreva, J. Clardy, R.G. Linington, and M.A. Fischbach, *A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics.* Cell, 2014. **158**(6): p. 1402-1414.

42. Qin, J., R. Li, J. Raes, M. Arumugam, K.S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D.R. Mende, J. Li, J. Xu, S. Li, D. Li, J. Cao, B. Wang, H. Liang, H. Zheng, Y. Xie, J. Tap, P. Lepage, M. Bertalan, J.M. Batto, T. Hansen, D. Le Paslier, A. Linneberg, H.B. Nielsen, E. Pelletier, P. Renault, T. Sicheritz-Ponten, K. Turner, H. Zhu, C. Yu, M. Jian, Y. Zhou, Y. Li, X. Zhang, N. Qin, H. Yang, J. Wang, S. Brunak, J. Dore, F. Guarner, K. Kristiansen, O. Pedersen, J. Parkhill, J. Weissenbach, P. Bork, and S.D. Ehrlich, *A human gut microbial gene catalogue established by metagenomic sequencing.* Nature, 2010. **464**(7285): p. 59-65.

43. Saleem, M., M. Nazir, M.S. Ali, H. Hussain, Y.S. Lee, N. Riaz, and A. Jabbar, *Antimicrobial natural products: an update on future antibiotic drug candidates.* Natural product reports, 2010. **27**(2): p. 238-254.

44. Ling, L.L., T. Schneider, A.J. Peoples, A.L. Spoering, I. Engels, B.P. Conlon, A. Mueller, T.F. Schaberle, D.E. Hughes, S. Epstein, M. Jones, L. Lazarides, V.A. Steadman, D.R. Cohen, C.R. Felix, K.A. Fetterman, W.P. Millett, A.G. Nitti, A.M. Zullo, C. Chen, and K. Lewis, *A new antibiotic kills pathogens without detectable resistance.* Nature, 2015. **517**(7535): p. 455-459.

45. von Nussbaum, F. and R.D. Sussmuth, *Multiple attack on bacteria by the new antibiotic teixobactin.* Angewandte Chemie, 2015. **54**(23): p. 6684-6686.

46.   Wright, G., *Antibiotics: An irresistible newcomer.* Nature, 2015. **517**(7535): p. 442-444.

47.   Servick, K., *The drug push.* Science, 2015. **348**(6237): p. 850-853.

48.   Zündorf, I., T. Dingermann, and R. Fürst, *Teixobactin–ein neuer Stern am Antibiotikahimmel?* Gire, 2014. **7**: p. 17-23.

49.   Tetsch, L., *Teixobactin: Reservemunition aus dem Boden.* Biologie in unserer Zeit, 2015. **45**(3): p. 141-143.

50.   Harrer, R., *Teixobactin: Antibiotikum vom iChip.* Chemie in unserer Zeit, 2015. **49**(3): p. 156-157.

51.   Choices, N., *New'game-changing'antibiotic discovered-Health News-NHS Choices.* 2015.

52.   Wiedemann, I., E. Breukink, C. van Kraaij, O.P. Kuipers, G. Bierbaum, B. de Kruijff, and H.G. Sahl, *Specific binding of nisin to the peptidoglycan precursor lipid II combines pore formation and inhibition of cell wall biosynthesis for potent antibiotic activity.* The Journal of biological chemistry, 2001. **276**(3): p. 1772-1779.

53.   Hasper, H.E., N.E. Kramer, J.L. Smith, J.D. Hillman, C. Zachariah, O.P. Kuipers, B. de Kruijff, and E. Breukink, *An alternative bactericidal mechanism of action for lantibiotic peptides that target lipid II.* Science, 2006. **313**(5793): p. 1636-1637.

54.   Arias, C.A. and B.E. Murray, *A new antibiotic and the evolution of resistance.* The New England journal of medicine, 2015. **372**(12): p. 1168-1170.

55.   Birch, R.G. and S.S. Patil, *Preliminary characterization of an antibiotic produced by Xanthomonas albilineans which inhibits DNA synthesis in Escherichia coli.* Journal of general microbiology, 1985. **131**(5): p. 1069-1075.

56.   Royer, M., L. Costet, E. Vivien, M. Bes, A. Cousin, A. Damais, I. Pieretti, A. Savin, S. Megessier, M. Viard, R. Frutos, D.W. Gabriel, and P.C. Rott, *Albicidin pathotoxin produced by Xanthomonas albilineans is encoded by three large PKS and NRPS genes present in a gene cluster also containing several putative modifying, regulatory, and resistance genes.* Molecular plant-microbe interactions : MPMI, 2004. **17**(4): p. 414-427.

57.   Cociancich, S., A. Pesic, D. Petras, S. Uhlmann, J. Kretz, V. Schubert, L. Vieweg, S. Duplan, M. Marguerettaz, J. Noell, I. Pieretti, M. Hugelland, S. Kemper, A. Mainz, P. Rott, M. Royer, and R.D. Sussmuth, *The gyrase inhibitor albicidin consists of p-aminobenzoic acids and cyanoalanine.* Nature chemical biology, 2015. **11**(3): p. 195-197.

58.   Birch R. G., P.S.S., *The relation of blocked chloroplast differentiation to sugarcane leaf scald disease.* Phytopathology, 1983. **73**: p. 1368-1374

59.   Pieretti, I., A. Pesic, D. Petras, M. Royer, R.D. Süssmuth, and S. Cociancich, *What makes Xanthomonas albilineans unique amongst xanthomonads?* Frontiers in Plant Science, 2015. **6**.

60.   Hashimi, S.M., M.K. Wall, A.B. Smith, A. Maxwell, and R.G. Birch, *The phytotoxin albicidin is a novel inhibitor of DNA gyrase.* Antimicrobial agents and chemotherapy, 2007. **51**(1): p. 181-187.

61.   Kretz, J., D. Kerwat, V. Schubert, S. Gratz, A. Pesic, S. Semsary, S. Cociancich, M. Royer, and R.D. Sussmuth, *Total synthesis of albicidin: a lead structure from Xanthomonas albilineans for potent antibacterial gyrase inhibitors.* Angewandte Chemie, 2015. **54**(6): p. 1969-1973.

62.   Maxwell, A. and D.M. Lawson, *The ATP-binding site of type II topoisomerases as a target for antibacterial drugs.* Current topics in medicinal chemistry, 2003. **3**(3): p. 283-303.

63.   Baumann, S., J. Herrmann, R. Raju, H. Steinmetz, K.I. Mohr, S. Huttel, K. Harmrolfs, M. Stadler, and R. Muller, *Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity.* Angewandte Chemie, 2014. **53**(52): p. 14605-14609.

64.   Calvete, J.J. and B. Lomonte, *A bright future for integrative venomics.* Toxicon : official journal of the International Society on Toxinology, 2015.

65.   Smith, C.G. and J.R. Vane, *The discovery of captopril.* FASEB journal : official publication of the Federation of American Societies for Experimental Biology, 2003. **17**(8): p. 788-789.

66.   Ferreira, S.H. and J.R. Vane, *The disappearance of bradykinin and eledoisin in the circulation and vascular beds of the cat.* British journal of pharmacology and chemotherapy, 1967. **30**(2): p. 417-424.

67.   Hashemzadeh, M., M. Furukawa, S. Goldsberry, and M.R. Movahed, *Chemical structures and mode of action of intravenous glycoprotein IIb/IIIa receptor blockers: A review.* Experimental and clinical cardiology, 2008. **13**(4): p. 192-197.

68.   Folkers, P.J., G.M. Clore, P.C. Driscoll, J. Dodt, S. Kohler, and A.M. Gronenborn, *Solution structure of recombinant hirudin and the Lys-47----Glu mutant: a nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing study.* Biochemistry, 1989. **28**(6): p. 2601-2617.

69.   Food, U. and D. Administration, *CDER drug and biologic approvals for calendar year 2005*, 2005.

70.   Twede, V.D., G. Miljanich, B.M. Olivera, and G. Bulaj, *Neuroprotective and cardioprotective conopeptides: an emerging class of drug leads.* Current opinion in drug discovery & development, 2009. **12**(2): p. 231-239.

71.   Vetter, I. and R.J. Lewis, *Therapeutic potential of cone snail venom peptides (conopeptides).* Current topics in medicinal chemistry, 2012. **12**(14): p. 1546-1552.

72.   Durek, T. and D.J. Craik, *Therapeutic conotoxins: a US patent literature survey.* Expert opinion on therapeutic patents, 2015. **25**(10): p. 1159-1173.

73.   Skov, M.J., J.C. Beck, A.W. de Kater, and G.M. Shopp, *Nonclinical safety of ziconotide: an intrathecal analgesic of a new pharmaceutical class.* International journal of toxicology, 2007. **26**(5): p. 411-421.

74.   Prommer, E., *Ziconotide: a new option for refractory pain.* Drugs of today, 2006. **42**(6): p. 369-378.

75.   Diochot, S., A. Baron, M. Salinas, D. Douguet, S. Scarzello, A.S. Dabert-Gay, D. Debayle, V. Friend, A. Alloui, M. Lazdunski, and E. Lingueglia, *Black mamba venom peptides target acid-sensing ion channels to abolish pain.* Nature, 2012. **490**(7421): p. 552-555.

76.   Laustsen, A.H., B. Lomonte, B. Lohse, J. Fernandez, and J.M. Gutierrez, *Unveiling the nature of black mamba (Dendroaspis polylepis) venom through venomics and antivenom immunoprofiling: Identification of key toxin targets for antivenom development.* Journal of proteomics, 2015. **119**: p. 126-142.

77.   Baron, A., S. Diochot, M. Salinas, E. Deval, J. Noel, and E. Lingueglia, *Venom toxins in the exploration of molecular, physiological and pathophysiological functions of acid-*

sensing ion channels. Toxicon : official journal of the International Society on Toxinology, 2013. **75**: p. 187-204.

78. Pu, X.C., P.T. Wong, and P. Gopalakrishnakone, *A novel analgesic toxin (hannalgesin) from the venom of king cobra (Ophiophagus hannah).* Toxicon : official journal of the International Society on Toxinology, 1995. **33**(11): p. 1425-1431.

79. Petras, D., P. Heiss, R.D. Sussmuth, and J.J. Calvete, *Venom Proteomics of Indonesian King Cobra, Ophiophagus hannah: Integrating Top-Down and Bottom-Up Approaches.* Journal of proteome research, 2015.

80. Vonk, F.J., N.R. Casewell, C.V. Henkel, A.M. Heimberg, H.J. Jansen, R.J. McCleary, H.M. Kerkkamp, R.A. Vos, I. Guerreiro, J.J. Calvete, W. Wuster, A.E. Woods, J.M. Logan, R.A. Harrison, T.A. Castoe, A.P. de Koning, D.D. Pollock, M. Yandell, D. Calderon, C. Renjifo, R.B. Currier, D. Salgado, D. Pla, L. Sanz, A.S. Hyder, J.M. Ribeiro, J.W. Arntzen, G.E. van den Thillart, M. Boetzer, W. Pirovano, R.P. Dirks, H.P. Spaink, D. Duboule, E. McGlinn, R.M. Kini, and M.K. Richardson, *The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(51): p. 20651-20656.

81. Gopalakrishnakone, P., X.C. Pu, P.T.-H. Wong, M.C.E. Gwee, and R.M. Kini, *Therapeutic molecules derived from snake venom*, 2003, Google Patents.

82. Minea, R., C. Helchowski, B. Rubino, K. Brodmann, S. Swenson, and F. Markland, Jr., *Development of a chimeric recombinant disintegrin as a cost-effective anti-cancer agent with promising translational potential.* Toxicon : official journal of the International Society on Toxinology, 2012. **59**(4): p. 472-486.

83. Calvete, J.J., *The continuing saga of snake venom disintegrins.* Toxicon : official journal of the International Society on Toxinology, 2013. **62**: p. 40-49.

84. Gocmen, B., P. Heiss, D. Petras, A. Nalbantsoy, and R.D. Sussmuth, *Mass spectrometry guided venom profiling and bioactivity screening of the Anatolian Meadow Viper, Vipera anatolica.* Toxicon : official journal of the International Society on Toxinology, 2015.

85. Pagel, W., *Paracelsus: an introduction to philosophical medicine in the era of the Renaissance*1982: Karger Medical and Scientific Publishers.

86. Paracelsus, T., *: Werke. Bd. 2, Darmstadt 1965, S. 497.Entstanden 1538. Erstdruck in lateinischer Übersetzung: Argentorati (Mylius) 1566. Erste deutsche Ausgabe: Basel (Perna) 1574.* Theophrast Paracelsus: Werke. . Vol. Bd. 2. 1574, Darmstadt: Erste deutsche Ausgabe: Basel (Perna).

87. Ihde, A.J., *The development of modern chemistry*1970: Courier Corporation.

88. Sobel, J., *Botulism.* Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 2005. **41**(8): p. 1167-1173.

89. Sobel, J., N. Tucker, A. Sulka, J. McLaughlin, and S. Maslanka, *Foodborne botulism in the United States, 1990-2000.* Emerging infectious diseases, 2004. **10**(9): p. 1606-1611.

90. Erbguth, F.J., *Historical notes on botulism, Clostridium botulinum, botulinum toxin, and the idea of the therapeutic use of the toxin.* Movement disorders : official journal of the Movement Disorder Society, 2004. **19 Suppl 8**: p. S2-6.

91.   Schep, L.J., W.A. Temple, G.A. Butt, and M.D. Beasley, *Ricin as a weapon of mass terror--separating fact from fiction.* Environment international, 2009. **35**(8): p. 1267-1271.

92.   Koepsell, D.R. and R. Arp, *Breaking Bad and Philosophy: Badder Living Through Chemistry.* Vol. 67. 2012: Open Court Publishing.

93.   Chambers, S., *Walter White is a Bad Teacher: Pedagogy, Partage, and Politics in Season 4 of Breaking Bad.* Theory & Event, 2014. **17**(1).

94.   Tachjian, A., V. Maria, and A. Jahangir, *Use of herbal products and potential interactions in patients with cardiovascular diseases.* Journal of the American College of Cardiology, 2010. **55**(6): p. 515-525.

95.   Gutierrez, J.M., D. Williams, H.W. Fan, and D.A. Warrell, *Snakebite envenoming from a global perspective: Towards an integrated approach.* Toxicon : official journal of the International Society on Toxinology, 2010. **56**(7): p. 1223-1235.

96.   Warrell, D.A., *Snake bite.* Lancet, 2010. **375**(9708): p. 77-88.

97.   Williams, D., J.M. Gutierrez, R. Harrison, D.A. Warrell, J. White, K.D. Winkel, and P. Gopalakrishnakone, *The Global Snake Bite Initiative: an antidote for snake bite.* Lancet, 2010. **375**(9708): p. 89-91.

98.   Gutierrez, J.M., T. Burnouf, R.A. Harrison, J.J. Calvete, U. Kuch, D.A. Warrell, and D.J. Williams, *A multicomponent strategy to improve the availability of antivenom for treating snakebite envenoming.* Bulletin of the World Health Organization, 2014. **92**(7): p. 526-532.

99.   Harrison, R.A., A. Hargreaves, S.C. Wagstaff, B. Faragher, and D.G. Lalloo, *Snake envenoming: a disease of poverty.* PLoS neglected tropical diseases, 2009. **3**(12): p. e569.

100.   Kasturiratne, A., A.R. Wickremasinghe, N. de Silva, N.K. Gunawardena, A. Pathmeswaran, R. Premaratna, L. Savioli, D.G. Lalloo, and H.J. de Silva, *The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths.* PLoS medicine, 2008. **5**(11): p. e218.

101.   Chippaux, J.P., *Snake-bites: appraisal of the global situation.* Bulletin of the World Health Organization, 1998. **76**(5): p. 515-524.

102.   Williams, D.J., J.M. Gutierrez, J.J. Calvete, W. Wuster, K. Ratanabanangkoon, O. Paiva, N.I. Brown, N.R. Casewell, R.A. Harrison, P.D. Rowley, M. O'Shea, S.D. Jensen, K.D. Winkel, and D.A. Warrell, *Ending the drought: new strategies for improving the flow of affordable, effective antivenoms in Asia and Africa.* Journal of proteomics, 2011. **74**(9): p. 1735-1767.

103.   Theakston, R.D., D.A. Warrell, and E. Griffiths, *Report of a WHO workshop on the standardization and control of antivenoms.* Toxicon, 2003. **41**(5): p. 541-557.

104.   Theakston, R.D. and D.A. Warrell, *Crisis in snake antivenom supply for Africa.* Lancet, 2000. **356**(9247): p. 2104.

105.   Warrell, D.A., *Unscrupulous marketing of snake bite antivenoms in Africa and Papua New Guinea: choosing the right product--'what's in a name?'.* Trans R Soc Trop Med Hyg, 2008. **102**(5): p. 397-399.

106.   Sanchez, L.V., D. Pla, M. Herrera, J.P. Chippaux, J.J. Calvete, and J.M. Gutierrez, *Evaluation of the preclinical efficacy of four antivenoms, distributed in sub-Saharan Africa, to neutralize the venom of the carpet viper, Echis ocellatus, from Mali, Cameroon, and Nigeria.* Toxicon : official journal of the International Society on Toxinology, 2015. **106**: p. 97-107.

107. Bortoli, S. and D.A. Volmer, *Account: characterization and identification of microcystins by mass spectrometry.* European journal of mass spectrometry, 2014. **20**(1): p. 1-19.

108. Whitehouse, C.M., R.N. Dreyer, M. Yamashita, and J.B. Fenn, *Electrospray interface for liquid chromatographs and mass spectrometers.* Analytical chemistry, 1985. **57**(3): p. 675-679.

109. Hillenkamp, F. and M. Karas, *Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization.* Methods in enzymology, 1990. **193**: p. 280-295.

110. Hu, Q., R.J. Noll, H. Li, A. Makarov, M. Hardman, and R. Graham Cooks, *The Orbitrap: a new mass spectrometer.* Journal of mass spectrometry : JMS, 2005. **40**(4): p. 430-443.

111. Scigelova, M. and A. Makarov, *Orbitrap mass analyzer--overview and applications in proteomics.* Proteomics, 2006. **6 Suppl 2**: p. 16-21.

112. Makarov, A. and M. Scigelova, *Coupling liquid chromatography to Orbitrap mass spectrometry.* Journal of chromatography. A, 2010. **1217**(25): p. 3938-3945.

113. Gao, P. and G. Xu, *Mass-spectrometry-based microbial metabolomics: recent developments and applications.* Analytical and bioanalytical chemistry, 2015. **407**(3): p. 669-680.

114. Beck, S., A. Michalski, O. Raether, M. Lubeck, S. Kaspar, N. Goedecke, C. Baessmann, D. Hornburg, F. Meier, I. Paron, N.A. Kulak, J. Cox, and M. Mann, *The Impact II, a Very High-Resolution Quadrupole Time-of-Flight Instrument (QTOF) for Deep Shotgun Proteomics.* Molecular & cellular proteomics : MCP, 2015. **14**(7): p. 2014-2029.

115. Petras, D., L. Sanz, A. Segura, M. Herrera, M. Villalta, D. Solano, M. Vargas, G. Leon, D.A. Warrell, R.D. Theakston, R.A. Harrison, N. Durfa, A. Nasidi, J.M. Gutierrez, and J.J. Calvete, *Snake venomics of African spitting cobras: toxin composition and assessment of congeneric cross-reactivity of the pan-African EchiTAb-Plus-ICP antivenom by antivenomics and neutralization approaches.* J Proteome Res, 2011. **10**(3): p. 1266-1280.

116. Lomonte, B., J. Escolano, J. Fernandez, L. Sanz, Y. Angulo, J.M. Gutierrez, and J.J. Calvete, *Snake venomics and antivenomics of the arboreal neotropical pitvipers Bothriechis lateralis and Bothriechis schlegelii.* J Proteome Res, 2008. **7**(6): p. 2445-2457.

117. Gutierrez, J.M., B. Lomonte, G. Leon, A. Alape-Giron, M. Flores-Diaz, L. Sanz, Y. Angulo, and J.J. Calvete, *Snake venomics and antivenomics: Proteomic tools in the design and control of antivenoms for the treatment of snakebite envenoming.* J Proteomics, 2009. **72**(2): p. 165-182.

118. Calvete, J.J., L. Sanz, Y. Angulo, B. Lomonte, and J.M. Gutierrez, *Venoms, venomics, antivenomics.* FEBS Lett, 2009. **583**(11): p. 1736-1743.

119. Calvete, J.J., A. Borges, A. Segura, M. Flores-Diaz, A. Alape-Giron, J.M. Gutierrez, N. Diez, L. De Sousa, D. Kiriakos, E. Sanchez, J.G. Faks, J. Escolano, and L. Sanz, *Snake venomics and antivenomics of Bothrops colombiensis, a medically important pitviper of the Bothrops atrox-asper complex endemic to Venezuela: Contributing to its taxonomy and snakebite management.* J Proteomics, 2009. **72**(2): p. 227-240.

120. Gutierrez, J.M., L. Sanz, M. Flores-Diaz, L. Figueroa, M. Madrigal, M. Herrera, M. Villalta, G. Leon, R. Estrada, A. Borges, A. Alape-Giron, and J.J. Calvete, *Impact of regional variation in Bothrops asper snake venom on the design of antivenoms:*

integrating antivenomics and neutralization approaches. J Proteome Res, 2010. **9**(1): p. 564-577.

121. Boldrini-Franca, J., C. Correa-Netto, M.M. Silva, R.S. Rodrigues, P. De La Torre, A. Perez, A.M. Soares, R.B. Zingali, R.A. Nogueira, V.M. Rodrigues, L. Sanz, and J.J. Calvete, *Snake venomics and antivenomics of Crotalus durissus subspecies from Brazil: assessment of geographic variation and its implication on snakebite management.* J Proteomics, 2010. **73**(9): p. 1758-1776.

122. Calvete, J.J., P. Cid, L. Sanz, A. Segura, M. Villalta, M. Herrera, G. Leon, R. Harrison, N. Durfa, A. Nasidi, R.D. Theakston, D.A. Warrell, and J.M. Gutierrez, *Antivenomic assessment of the immunological reactivity of EchiTAb-Plus-ICP, an antivenom for the treatment of snakebite envenoming in sub-Saharan Africa.* Am J Trop Med Hyg, 2010. **82**(6): p. 1194-1201.

123. Eichberg, S., L. Sanz, J.J. Calvete, and D. Pla, *Constructing comprehensive venom proteome reference maps for integrative venomics.* Expert review of proteomics, 2015. **12**(5): p. 557-573.

124. Michalski, A., E. Damoc, O. Lange, E. Denisov, D. Nolting, M. Muller, R. Viner, J. Schwartz, P. Remes, M. Belford, J.J. Dunyach, J. Cox, S. Horning, M. Mann, and A. Makarov, *Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes.* Molecular & cellular proteomics : MCP, 2012. **11**(3): p. O111 013698.

125. Catherman, A.D., O.S. Skinner, and N.L. Kelleher, *Top Down proteomics: facts and perspectives.* Biochemical and biophysical research communications, 2014. **445**(4): p. 683-693.

126. Dang, X., J. Scotcher, S. Wu, R.K. Chu, N. Tolic, I. Ntai, P.M. Thomas, R.T. Fellers, B.P. Early, Y. Zheng, K.R. Durbin, R.D. Leduc, J.J. Wolff, C.J. Thompson, J. Pan, J. Han, J.B. Shaw, J.P. Salisbury, M. Easterling, C.H. Borchers, J.S. Brodbelt, J.N. Agar, L. Pasa-Tolic, N.L. Kelleher, and N.L. Young, *The first pilot project of the consortium for top-down proteomics: a status report.* Proteomics, 2014. **14**(10): p. 1130-1140.

127. Khoury, G.A., R.C. Baliban, and C.A. Floudas, *Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database.* Scientific reports, 2011. **1**.

128. Dischinger, J., S. Basi Chipalu, and G. Bierbaum, *Lantibiotics: promising candidates for future applications in health care.* International journal of medical microbiology : IJMM, 2014. **304**(1): p. 51-62.

129. Jones, A.W. and H.J. Cooper, *Dissociation techniques in mass spectrometry-based proteomics.* The Analyst, 2011. **136**(17): p. 3419-3429.

130. Seidler, J., N. Zinn, M.E. Boehm, and W.D. Lehmann, *De novo sequencing of peptides by MS/MS.* Proteomics, 2010. **10**(4): p. 634-649.

131. Eng, J.K., A.L. McCormack, and J.R. Yates, *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.* Journal of the American Society for Mass Spectrometry, 1994. **5**(11): p. 976-989.

132. Condurso, H.L. and S.D. Bruner, *Structure and noncanonical chemistry of nonribosomal peptide biosynthetic machinery.* Natural product reports, 2012. **29**(10): p. 1099-1110.

133. Bouslimani, A., L.M. Sanchez, N. Garg, and P.C. Dorrestein, *Mass spectrometry of natural products: current, emerging and future technologies.* Natural product reports, 2014. **31**(6): p. 718-729.

134.   Tietz, J.I. and D.A. Mitchell, *Using Genomics for Natural Product Structure Elucidation.* Current topics in medicinal chemistry, 2015.

135.   Kersten, R.D., Y.L. Yang, Y. Xu, P. Cimermancic, S.J. Nam, W. Fenical, M.A. Fischbach, B.S. Moore, and P.C. Dorrestein, *A mass spectrometry-guided genome mining approach for natural product peptidogenomics.* Nature chemical biology, 2011. **7**(11): p. 794-802.

136.   Mohimani, H., R.D. Kersten, W.T. Liu, M. Wang, S.O. Purvine, S. Wu, H.M. Brewer, L. Pasa-Tolic, N. Bandeira, B.S. Moore, P.A. Pevzner, and P.C. Dorrestein, *Automated genome mining of ribosomal peptide natural products.* ACS chemical biology, 2014. **9**(7): p. 1545-1551.

137.   Mohimani, H., W.T. Liu, R.D. Kersten, B.S. Moore, P.C. Dorrestein, and P.A. Pevzner, *NRPquest: Coupling Mass Spectrometry and Genome Mining for Nonribosomal Peptide Discovery.* Journal of natural products, 2014. **77**(8): p. 1902-1909.

138.   Medema, M.H., Y. Paalvast, D.D. Nguyen, A. Melnik, P.C. Dorrestein, E. Takano, and R. Breitling, *Pep2Path: automated mass spectrometry-guided genome mining of peptidic natural products.* PLoS computational biology, 2014. **10**(9): p. e1003822.

139.   Little, J.L., C.D. Cleven, and S.D. Brown, *Identification of "known unknowns" utilizing accurate mass data and chemical abstracts service databases.* Journal of the American Society for Mass Spectrometry, 2011. **22**(2): p. 348-359.

140.   Dettmer, K., P.A. Aronov, and B.D. Hammock, *Mass spectrometry-based metabolomics.* Mass spectrometry reviews, 2007. **26**(1): p. 51-78.

141.   Horai, H., M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M.Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, and T. Nishioka, *MassBank: a public repository for sharing mass spectral data for life sciences.* Journal of mass spectrometry : JMS, 2010. **45**(7): p. 703-714.

142.   Zhu, Z.J., A.W. Schultz, J. Wang, C.H. Johnson, S.M. Yannone, G.J. Patti, and G. Siuzdak, *Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of metabolites guided by the METLIN database.* Nature protocols, 2013. **8**(3): p. 451-460.

143.   Stein, S., *Mass spectral reference libraries: an ever-expanding resource for chemical identification.* Analytical chemistry, 2012. **84**(17): p. 7274-7282.

144.   Watrous, J., P. Roach, T. Alexandrov, B.S. Heath, J.Y. Yang, R.D. Kersten, M. van der Voort, K. Pogliano, H. Gross, J.M. Raaijmakers, B.S. Moore, J. Laskin, N. Bandeira, and P.C. Dorrestein, *Mass spectral molecular networking of living microbial colonies.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(26): p. E1743-1752.

145.   Moss, N.A., M.J. Bertin, K. Kleigrewe, T.F. Leao, L. Gerwick, and W.H. Gerwick, *Integrating mass spectrometry and genomics for cyanobacterial metabolite discovery.* Journal of industrial microbiology & biotechnology, 2015.

146.   Teta, R., G. Della Sala, E. Glukhov, L. Gerwick, W.H. Gerwick, A. Mangoni, and V. Costantino, *A combined LC-MS/MS and molecular networking approach reveals new cyanotoxins from the 2014 cyanobacterial bloom in Green Lake, Seattle.* Environmental science & technology, 2015.

147. Winnikoff, J.R., E. Glukhov, J. Watrous, P.C. Dorrestein, and W.H. Gerwick, *Quantitative molecular networking to profile marine cyanobacterial metabolomes.* The Journal of antibiotics, 2014. **67**(1): p. 105-112.

148. Vizcaino, M.I. and J.M. Crawford, *The colibactin warhead crosslinks DNA.* Nature chemistry, 2015. **7**(5): p. 411-417.

149. Klitgaard, A., J.B. Nielsen, R.J. Frandsen, M.R. Andersen, and K.F. Nielsen, *Combining Stable Isotope Labeling and Molecular Networking for Biosynthetic Pathway Characterization.* Analytical chemistry, 2015. **87**(13): p. 6520-6526.

150. Saleh, H.P., D; Mainz, A; Kerwat, D; Nalbantsoy, A; Erzurumlu, Y; Süssmuth, RD, *Deuterium-labeled precursor feeding reveals a new pABA-containing meroterpenoid from the mango pathogen Xanthomonas citri pv. mangiferaeindicae* Journal of Natural Products, 2015. **submitted**.

151. Birch, R.G., *Xanthomonas albilineans and the antipathogenesis approach to disease control.* Molecular plant pathology, 2001. **2**(1): p. 1-11.

152. Rott, P. and M.J. Davis, *Leaf scald.* A guide to sugarcane diseases, 2000: p. 38-44.

153. Autrey, L., S. Saumtally, and A. Dookun. *Aerial transmission of the leaf scald pathogen, Xanthomonas albilineans*. in *Proceedings XXI Congress of ISSCT, Bangkok (Thailand), 5-14 Mar 1992*. 1995.

154. Daugrois, J., V. Dumont, P. Champoiseau, L. Costet, R. Boisne-Noc, and P. Rott, *Aerial contamination of sugarcane in Guadeloupe by two strains of Xanthomonas albilineans.* European journal of plant pathology, 2003. **109**(5): p. 445-458.

155. Champoiseau, P., P. Rott, and J.-H. Daugrois, *Epiphytic populations of Xanthomonas albilineans and subsequent sugarcane stalk infection are linked to rainfall in Guadeloupe.* Plant disease, 2009. **93**(4): p. 339-346.

156. Young, J.M., D.C. Park, H.M. Shearman, and E. Fargier, *A multilocus sequence analysis of the genus Xanthomonas.* Systematic and applied microbiology, 2008. **31**(5): p. 366-377.

157. Studholme, D.J., A. Wasukira, K. Paszkiewicz, V. Aritua, R. Thwaites, J. Smith, and M. Grant, *Draft Genome Sequences of Xanthomonas sacchari and Two Banana-Associated Xanthomonads Reveal Insights into the Xanthomonas Group 1 Clade.* Genes, 2011. **2**(4): p. 1050-1065.

158. Studholme, D.J., A. Wasukira, K. Paszkiewicz, V. Aritua, R. Thwaites, J. Smith, and M. Grant, *Correction: Studholme et al., Draft Genome Sequences of Xanthomonas sacchari and Two Banana-Associated Xanthomonads Reveal Insights into the Xanthomonas Group 1 clade. Genes 2011, 2, 1050–1065.*

159. Rodriguez, R.L., A. Grajales, M.L. Arrieta-Ortiz, C. Salazar, S. Restrepo, and A. Bernal, *Genomes-based phylogeny of the genus Xanthomonas.* BMC microbiology, 2012. **12**: p. 43.

160. Naushad, H.S. and R.S. Gupta, *Phylogenomics and molecular signatures for species from the plant pathogen-containing order xanthomonadales.* PloS one, 2013. **8**(2): p. e55216.

161. Simpson, A.J.G., F. Reinach, P. Arruda, F. Abreu, M. Acencio, R. Alvarenga, L.C. Alves, J. Araya, G. Baia, and C. Baptista, *The genome sequence of the plant pathogen Xylella fastidiosa.* Nature, 2000. **406**(6792): p. 151-157.

162. Pieretti, I., M. Royer, V. Barbe, S. Carrere, R. Koebnik, S. Cociancich, A. Couloux, A. Darrasse, J. Gouzy, M.A. Jacques, E. Lauber, C. Manceau, S. Mangenot, S. Poussier, B. Segurens, B. Szurek, V. Verdier, M. Arlat, and P. Rott, *The complete genome sequence*

of Xanthomonas albilineans provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. BMC genomics, 2009. **10**: p. 616.

163. Pieretti, I., M. Royer, V. Barbe, S. Carrere, R. Koebnik, A. Couloux, A. Darrasse, J. Gouzy, M.A. Jacques, E. Lauber, C. Manceau, S. Mangenot, S. Poussier, B. Segurens, B. Szurek, V. Verdier, M. Arlat, D.W. Gabriel, P. Rott, and S. Cociancich, *Genomic insights into strategies used by Xanthomonas albilineans with its reduced artillery to spread within sugarcane xylem vessels.* BMC genomics, 2012. **13**: p. 658.

164. Marguerettaz, M., I. Pieretti, P. Gayral, J. Puig, C. Brin, S. Cociancich, S. Poussier, P. Rott, and M. Royer, *Genomic and evolutionary features of the SPI-1 type III secretion system that is present in Xanthomonas albilineans but is not essential for xylem colonization and symptom development of sugarcane leaf scald.* Molecular plant-microbe interactions : MPMI, 2011. **24**(2): p. 246-259.

165. Royer, M., R. Koebnik, M. Marguerettaz, V. Barbe, G.P. Robin, C. Brin, S. Carrere, C. Gomez, M. Hugelland, G.H. Voller, J. Noell, I. Pieretti, S. Rausch, V. Verdier, S. Poussier, P. Rott, R.D. Sussmuth, and S. Cociancich, *Genome mining reveals the genus Xanthomonas to be a promising reservoir for new bioactive non-ribosomally synthesized peptides.* BMC genomics, 2013. **14**: p. 658.

166. Pieretti, I., S. Bolot, S. Carrere, V. Barbe, S. Cociancich, P. Rott, and M. Royer, *Draft Genome Sequence of Xanthomonas sacchari Strain LMG 476.* Genome announcements, 2015. **3**(2).

167. Vandroemme, J., B. Cottyn, S. Baeyen, P. De Vos, and M. Maes, *Draft genome sequence of Xanthomonas fragariae reveals reductive evolution and distinct virulence-related gene content.* BMC genomics, 2013. **14**: p. 829.

168. Roper, M.C., L.C. Greve, J.G. Warren, J.M. Labavitch, and B.C. Kirkpatrick, *Xylella fastidiosa requires polygalacturonase for colonization and pathogenicity in Vitis vinifera grapevines.* Molecular plant-microbe interactions : MPMI, 2007. **20**(4): p. 411-419.

169. Chatterjee, S., R.P. Almeida, and S. Lindow, *Living in two worlds: the plant and insect lifestyles of Xylella fastidiosa.* Annual review of phytopathology, 2008. **46**: p. 243-271.

170. Perez-Donoso, A.G., Q. Sun, M.C. Roper, L.C. Greve, B. Kirkpatrick, and J.M. Labavitch, *Cell wall-degrading enzymes enlarge the pore size of intervessel pit membranes in healthy and Xylella fastidiosa-infected grapevines.* Plant physiology, 2010. **152**(3): p. 1748-1759.

171. Rott, P., L. Fleites, G. Marlow, M. Royer, and D.W. Gabriel, *Identification of new candidate pathogenicity factors in the xylem-invading pathogen Xanthomonas albilineans by transposon mutagenesis.* Molecular plant-microbe interactions : MPMI, 2011. **24**(5): p. 594-605.

172. Wichmann, F., F.J. Vorholter, L. Hersemann, F. Widmer, J. Blom, K. Niehaus, S. Reinhard, C. Conradin, and R. Kolliker, *The noncanonical type III secretion system of Xanthomonas translucens pv. graminis is essential for forage grass infection.* Molecular plant pathology, 2013. **14**(6): p. 576-588.

173. Gardiner, D.M., N.M. Upadhyaya, J. Stiller, J.G. Ellis, P.N. Dodds, K. Kazan, and J.M. Manners, *Genomic analysis of Xanthomonas translucens pathogenic on wheat and barley reveals cross-kingdom gene transfer events and diverse protein delivery systems.* PloS one, 2014. **9**(1): p. e84995.

174. Stevens, M.P., M.W. Wood, L.A. Taylor, P. Monaghan, P. Hawes, P.W. Jones, T.S. Wallis, and E.E. Galyov, *An Inv/Mxi-Spa-like type III protein secretion system in*

Burkholderia pseudomallei modulates intracellular behaviour of the pathogen. Molecular microbiology, 2002. **46**(3): p. 649-659.

175. Alavi, S.M., S. Sanjari, F. Durand, C. Brin, C. Manceau, and S. Poussier, *Assessment of the genetic diversity of Xanthomonas axonopodis pv. phaseoli and Xanthomonas fuscans subsp. fuscans as a basis to identify putative pathogenicity genes and a type III secretion system of the SPI-1 family by multiple suppression subtractive hybridizations.* Applied and environmental microbiology, 2008. **74**(10): p. 3295-3301.

176. Schikora, A., I. Virlogeux-Payant, E. Bueso, A.V. Garcia, T. Nilau, A. Charrier, S. Pelletier, P. Menanteau, M. Baccarini, P. Velge, and H. Hirt, *Conservation of Salmonella infection mechanisms in plants and animals.* PloS one, 2011. **6**(9): p. e24112.

177. Kube, M., A.M. Migdoll, I. Muller, H. Kuhl, A. Beck, R. Reinhardt, and K. Geider, *The genome of Erwinia tasmaniensis strain Et1/99, a non-pathogenic bacterium in the genus Erwinia.* Environmental microbiology, 2008. **10**(9): p. 2211-2222.

178. Jennings, M.E., L.N. Quick, N. Ubol, S. Shrom, N. Dollahon, and J.W. Wilson, *Characterization of Salmonella type III secretion hyper-activity which results in biofilm-like cell aggregation.* PloS one, 2012. **7**(3): p. e33080.

179. Correa, V.R., D.R. Majerczak, D. Ammar el, M. Merighi, R.C. Pratt, S.A. Hogenhout, D.L. Coplin, and M.G. Redinbaugh, *The bacterium Pantoea stewartii uses two different type III secretion systems to colonize its plant host and insect vector.* Applied and environmental microbiology, 2012. **78**(17): p. 6327-6336.

180. Katzen, F., D.U. Ferreiro, C.G. Oddo, M.V. Ielmini, A. Becker, A. Puhler, and L. Ielpi, *Xanthomonas campestris pv. campestris gum mutants: effects on xanthan biosynthesis and plant virulence.* Journal of bacteriology, 1998. **180**(7): p. 1607-1617.

181. Kim, J.G., X. Li, J.A. Roden, K.W. Taylor, C.D. Aakre, B. Su, S. Lalonde, A. Kirik, Y. Chen, G. Baranage, H. McLane, G.B. Martin, and M.B. Mudgett, *Xanthomonas T3S Effector XopN Suppresses PAMP-Triggered Immunity and Interacts with a Tomato Atypical Receptor-Like Kinase and TFT1.* The Plant cell, 2009. **21**(4): p. 1305-1323.

182. Galvan, E.M., M.V. Ielmini, Y.N. Patel, M.I. Bianco, E.A. Franceschini, J.C. Schneider, and L. Ielpi, *Xanthan chain length is modulated by increasing the availability of the polysaccharide copolymerase protein GumC and the outer membrane polysaccharide export protein GumB.* Glycobiology, 2013. **23**(2): p. 259-272.

183. Potnis, N., K. Krasileva, V. Chow, N.F. Almeida, P.B. Patil, R.P. Ryan, M. Sharlach, F. Behlau, J.M. Dow, M. Momol, F.F. White, J.F. Preston, B.A. Vinatzer, R. Koebnik, J.C. Setubal, D.J. Norman, B.J. Staskawicz, and J.B. Jones, *Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper.* BMC genomics, 2011. **12**: p. 146.

184. Darrasse, A., S. Carrere, V. Barbe, T. Boureau, M.L. Arrieta-Ortiz, S. Bonneau, M. Briand, C. Brin, S. Cociancich, K. Durand, S. Fouteau, L. Gagnevin, F. Guerin, E. Guy, A. Indiana, R. Koebnik, E. Lauber, A. Munoz, L.D. Noel, I. Pieretti, S. Poussier, O. Pruvost, I. Robene-Soustrade, P. Rott, M. Royer, L. Serres-Giardi, B. Szurek, M.A. van Sluys, V. Verdier, C. Verniere, M. Arlat, C. Manceau, and M.A. Jacques, *Genome sequence of Xanthomonas fuscans subsp. fuscans strain 4834-R reveals that flagellar motility is not a general feature of xanthomonads.* BMC genomics, 2013. **14**: p. 761.

185. Filloux, A., *The rise of the Type VI secretion system.* F1000prime reports, 2013. **5**: p. 52.

186. Russell, A.B., S.B. Peterson, and J.D. Mougous, *Type VI secretion system effectors: poisons with a purpose.* Nature reviews. Microbiology, 2014. **12**(2): p. 137-148.

187. Magnani, G.S., L.M. Cruz, H. Weber, J.C. Bespalhok, E. Daros, V. Baura, M.G. Yates, R.A. Monteiro, H. Faoro, F.O. Pedrosa, and E.M. Souza, *Culture-independent analysis of endophytic bacterial communities associated with Brazilian sugarcane.* Genetics and molecular research : GMR, 2013. **12**(4): p. 4549-4558.

188. Monteiro-Vitorello, C.B., L.E. Camargo, M.A. Van Sluys, J.P. Kitajima, D. Truffi, A.M. do Amaral, R. Harakava, J.C. de Oliveira, D. Wood, M.C. de Oliveira, C. Miyaki, M.A. Takita, A.C. da Silva, L.R. Furlan, D.M. Carraro, G. Camarotte, N.F. Almeida, Jr., H. Carrer, L.L. Coutinho, H.A. El-Dorry, M.I. Ferro, P.R. Gagliardi, E. Giglioti, M.H. Goldman, G.H. Goldman, E.T. Kimura, E.S. Ferro, E.E. Kuramae, E.G. Lemos, M.V. Lemos, S.M. Mauro, M.A. Machado, C.L. Marino, C.F. Menck, L.R. Nunes, R.C. Oliveira, G.G. Pereira, W. Siqueira, A.A. de Souza, S.M. Tsai, A.S. Zanca, A.J. Simpson, S.M. Brumbley, and J.C. Setubal, *The genome sequence of the gram-positive sugarcane pathogen Leifsonia xyli subsp. xyli.* Molecular plant-microbe interactions : MPMI, 2004. **17**(8): p. 827-836.

189. Zhang, L. and R.G. Birch, *The gene for albicidin detoxification from Pantoea dispersa encodes an esterase and attenuates pathogenicity of Xanthomonas albilineans to sugarcane.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(18): p. 9984-9989.

190. Strieker, M., A. Tanovic, and M.A. Marahiel, *Nonribosomal peptide synthetases: structures and dynamics.* Current opinion in structural biology, 2010. **20**(2): p. 234-240.

191. Stachelhaus, T. and M.A. Marahiel, *Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis.* FEMS microbiology letters, 1995. **125**(1): p. 3-14.

192. Weissman, K.J. and P.F. Leadlay, *Combinatorial biosynthesis of reduced polyketides.* Nature reviews. Microbiology, 2005. **3**(12): p. 925-936.

193. Vivien, E., D. Pitorre, S. Cociancich, I. Pieretti, D.W. Gabriel, P.C. Rott, and M. Royer, *Heterologous production of albicidin: a promising approach to overproducing and characterizing this potent inhibitor of DNA gyrase.* Antimicrobial agents and chemotherapy, 2007. **51**(4): p. 1549-1552.

194. Royer, M., L. Costet, E. Vivien, M. Bes, A. Cousin, A. Damais, I. Pieretti, A. Savin, S. Megessier, M. Viard, R. Frutos, D.W. Gabriel, and P.C. Rott, *Albicidin pathotoxin produced by Xanthomonas albilineans is encoded by three large PKS and NRPS genes present in a gene cluster also containing several putative modifying, regulatory, and resistance genes.* Mol. Plant Microbe Interact., 2004. **17**: p. 414-427.

195. Dal-Bianco, M., M.S. Carneiro, C.T. Hotta, R.G. Chapola, H.P. Hoffmann, A.A. Garcia, and G.M. Souza, *Sugarcane improvement: how far can we go?* Current opinion in biotechnology, 2012. **23**(2): p. 265-270.

196. Birch, R.G. and S.S. Patil, *Evidence that an albicidin-like phytotoxin induces chlorosis in sugarcane leaf scald disease by blocking plastid DNA replication.* Physiological and Molecular Plant Pathology, 1987. **30**(2): p. 207-214.

197. Birch, R.G. and S.S. Patil, *Correlation between albicidin production and chlorosis induction by Xanthomonas albilineans, the sugarcane leaf scald pathogen.* Physiological and Molecular Plant Pathology, 1987. **30**(2): p. 199-206.

198. Birch, R.G. and S.S. Patil, *Antibiotic and process for the production thereof.* US patent, 1985: p. 4,525,354.

199. Corbett, K.D. and J.M. Berger, *Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases.* Annual review of biophysics and biomolecular structure, 2004. **33**: p. 95-118.

200. Vivien, E., D. Pitorre, S. Cociancich, I. Pieretti, D.W. Gabriel, P.C. Rott, and M. Royer, *Heterologous production of albicidin: a promising approach to overproducing and characterizing this potent Inhibitor of DNA Gyrase.* Antimicrob. Agents Chemother., 2007. **51**: p. 1549-1552.

201. Dosselaere, F. and J. Vanderleyden, *A metabolic node in action: chorismate-utilizing enzymes in microorganisms.* Critical reviews in microbiology, 2001. **27**(2): p. 75-131.

202. Stachelhaus, T., H.D. Mootz, and M.A. Marahiel, *The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases.* Chemistry & biology, 1999. **6**(8): p. 493-505.

203. Du, L., C. Sanchez, M. Chen, D.J. Edwards, and B. Shen, *The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase.* Chemistry & biology, 2000. **7**(8): p. 623-642.

204. Aravind, L., V. Anantharaman, and E.V. Koonin, *Monophyly of class I aminoacyl tRNA synthetase, USPA, ETFP, photolyase, and PP-ATPase nucleotide-binding domains: implications for protein evolution in the RNA.* Proteins, 2002. **48**(1): p. 1-14.

205. Huang, G., L. Zhang, and R.G. Birch, *Analysis of the genes flanking xabB: a methyltransferase gene is involved in albicidin biosynthesis in Xanthomonas albilineans.* Gene, 2000. **255**(2): p. 327-333.

206. Koglin, A., M.R. Mofid, F. Lohr, B. Schafer, V.V. Rogov, M.M. Blum, T. Mittag, M.A. Marahiel, F. Bernhard, and V. Dotsch, *Conformational switches modulate protein interactions in peptide antibiotic synthetases.* Science, 2006. **312**(5771): p. 273-276.

207. Tanovic, A., S.A. Samel, L.O. Essen, and M.A. Marahiel, *Crystal structure of the termination module of a nonribosomal peptide synthetase.* Science, 2008. **321**(5889): p. 659-663.

208. Zhang, L., J. Xu, and R.G. Birch, *Factors affecting biosynthesis by Xanthomonas albilineans of albicidin antibiotics and phytotoxins.* Journal of applied microbiology, 1998. **85**(6): p. 1023-1028.

209. Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, *Basic local alignment search tool.* Journal of molecular biology, 1990. **215**(3): p. 403-410.

210. McWilliam, H., W. Li, M. Uludag, S. Squizzato, Y.M. Park, N. Buso, A.P. Cowley, and R. Lopez, *Analysis Tool Web Services from the EMBL-EBI.* Nucleic acids research, 2013. **41**(Web Server issue): p. W597-600.

211. Bachmann, B.O. and J. Ravel, *Chapter 8. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data.* Methods in enzymology, 2009. **458**: p. 181-217.

212. Zhang, Y., *I-TASSER server for protein 3D structure prediction.* BMC bioinformatics, 2008. **9**: p. 40.

213. Bogomolovas, J., B. Simon, M. Sattler, and G. Stier, *Screening of fusion partners for high yield expression and purification of bioactive viscotoxins.* Protein expression and purification, 2009. **64**(1): p. 16-23.

214. Calendar, R. and P. Berg, *Purification and physical characterization of tyrosyl ribonucleic acid synthetases from Escherichia coli and Bacillus subtilis.* Biochemistry, 1966. **5**: p. 1681-1690.

215. Rott, P., *A guide to sugarcane diseases*2000: Editions Quae.

216. Champoiseau, P., P. Rott, and J.H. Daugrois, *Epiphytic Populations of Xanthomonas albilineans and Subsequent Sugarcane Stalk Infection Are Linked to Rainfall in Guadeloupe.* Plant Disease, 2009. **93**(4): p. 339-346.

217. Daugrois, J.H., V. Dumont, P. Champoiseau, L. Costet, R. Boisne-Noc, and P. Rott, *Aerial Contamination of Sugarcane in Guadeloupe by Two Strains of Xanthomonas albilineans.* European Journal of Plant Pathology, 2003. **109**(5): p. 445-458.

218. Mensi, I., M.S. Vernerey, D. Gargani, M. Nicole, and P. Rott, *Breaking dogmas: the plant vascular pathogen Xanthomonas albilineans is able to invade non-vascular tissues despite its reduced genome.* Open biology, 2014. **4**: p. 130116.

219. Huang, G., L. Zhang, and R.G. Birch, *A multifunctional polyketide-peptide synthetase essential for albicidin biosynthesis in Xanthomonas albilineans.* Microbiology, 2001. **147**(Pt 3): p. 631-642.

220. Birch, R.G. and S.S. Patil, *Antibiotic and process for the production thereof*, 1985, Google Patents.

221. Brewer, S.J., P.M. Taylor, and M.K. Turner, *An adenosine triphosphate-dependent carbamoylphosphate--3-hydroxymethylcephem O-carbamoyltransferase from Streptomyces clavuligerus.* The Biochemical journal, 1980. **185**(3): p. 555-564.

222. Freel Meyers, C.L., M. Oberthur, H. Xu, L. Heide, D. Kahne, and C.T. Walsh, *Characterization of NovP and NovN: completion of novobiocin biosynthesis by sequential tailoring of the noviosyl ring.* Angewandte Chemie, 2004. **43**(1): p. 67-70.

223. Parthier, C., S. Gorlich, F. Jaenecke, C. Breithaupt, U. Brauer, U. Fandrich, D. Clausnitzer, U.F. Wehmeier, C. Bottcher, D. Scheel, and M.T. Stubbs, *The O-carbamoyltransferase TobZ catalyzes an ancient enzymatic reaction.* Angewandte Chemie, 2012. **51**(17): p. 4046-4052.

224. Dauvergne, J., K. Wellington, and K. Chibale, *Unprecedented observation of sulfonamides in the transesterification of N-unsubstituted carbamates with sulfonyl chlorides.* Tetrahedron Letters, 2004. **45**(1): p. 43-47.

225. Yu, T.W., L. Bai, D. Clade, D. Hoffmann, S. Toelzer, K.Q. Trinh, J. Xu, S.J. Moss, E. Leistner, and H.G. Floss, *The biosynthetic gene cluster of the maytansinoid antitumor agent ansamitocin from Actinosynnema pretiosum.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(12): p. 7968-7973.

226. Li, Y., P. Zhao, Q. Kang, J. Ma, L. Bai, and Z. Deng, *Dual carbamoylations on the polyketide and glycosyl moiety by asm21 result in extended ansamitocin biosynthesis.* Chemistry & biology, 2011. **18**(12): p. 1571-1580.

227. Castaneda, A., J.D. Reddy, B. El-Yacoubi, and D.W. Gabriel, *Mutagenesis of all eight avr genes in Xanthomonas campestris pv. campestris had no detected effect on pathogenicity, but one avr gene affected race specificity.* Molecular plant-microbe interactions : MPMI, 2005. **18**(12): p. 1306-1317.

228. Shryock, T.R., *Performance standards for antimicrobial disk and dilution susceptibility tests for bacteria isolated from animals: approved standard*2002: NCCLC.

229. Fry, B.G., K. Roelants, D.E. Champagne, H. Scheib, J.D. Tyndall, G.F. King, T.J. Nevalainen, J.A. Norman, R.J. Lewis, R.S. Norton, C. Renjifo, and R.C. de la Vega, *The*

toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. Annual review of genomics and human genetics, 2009. **10**: p. 483-511.

230. Gibbs, H.L., L. Sanz, and J.J. Calvete, *Snake population venomics: proteomics-based analyses of individual variation reveals significant gene regulation effects on venom protein expression in Sistrurus rattlesnakes.* Journal of molecular evolution, 2009. **68**(2): p. 113-125.

231. Gibbs, H.L. and S.P. Mackessy, *Functional basis of a molecular adaptation: prey-specific toxic effects of venom from Sistrurus rattlesnakes.* Toxicon : official journal of the International Society on Toxinology, 2009. **53**(6): p. 672-679.

232. Casewell, N.R., G.A. Huttley, and W. Wuster, *Dynamic evolution of venom proteins in squamate reptiles.* Nature communications, 2012. **3**: p. 1066.

233. Durban, J., A. Perez, L. Sanz, A. Gomez, F. Bonilla, S. Rodriguez, D. Chacon, M. Sasa, Y. Angulo, J.M. Gutierrez, and J.J. Calvete, *Integrated "omics" profiling indicates that miRNAs are modulators of the ontogenetic venom composition shift in the Central American rattlesnake, Crotalus simus simus.* BMC genomics, 2013. **14**: p. 234.

234. Castoe, T.A., A.P. de Koning, K.T. Hall, D.C. Card, D.R. Schield, M.K. Fujita, R.P. Ruggiero, J.F. Degner, J.M. Daza, W. Gu, J. Reyes-Velasco, K.J. Shaney, J.M. Castoe, S.E. Fox, A.W. Poole, D. Polanco, J. Dobry, M.W. Vandewege, Q. Li, R.K. Schott, A. Kapusta, P. Minx, C. Feschotte, P. Uetz, D.A. Ray, F.G. Hoffmann, R. Bogden, E.N. Smith, B.S. Chang, F.J. Vonk, N.R. Casewell, C.V. Henkel, M.K. Richardson, S.P. Mackessy, A.M. Bronikowski, M. Yandell, W.C. Warren, S.M. Secor, and D.D. Pollock, *The Burmese python genome reveals the molecular basis for extreme adaptation in snakes.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(51): p. 20645-20650.

235. Casewell, N.R., S.C. Wagstaff, W. Wuster, D.A. Cook, F.M. Bolton, S.I. King, D. Pla, L. Sanz, J.J. Calvete, and R.A. Harrison, *Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(25): p. 9205-9210.

236. Reyes-Velasco, J., D.C. Card, A.L. Andrew, K.J. Shaney, R.H. Adams, D.R. Schield, N.R. Casewell, S.P. Mackessy, and T.A. Castoe, *Expression of venom gene homologs in diverse python tissues suggests a new model for the evolution of snake venom.* Molecular biology and evolution, 2015. **32**(1): p. 173-183.

237. Diz, A.P., M. Martinez-Fernandez, and E. Rolan-Alvarez, *Proteomics in evolutionary ecology: linking the genotype with the phenotype.* Molecular ecology, 2012. **21**(5): p. 1060-1080.

238. Calvete, J.J., *Snake venomics: from the inventory of toxins to biology.* Toxicon : official journal of the International Society on Toxinology, 2013. **75**: p. 44-62.

239. Brahma, R.K., R.J. McCleary, R.M. Kini, and R. Doley, *Venom gland transcriptomics for identifying, cataloging, and characterizing venom proteins in snakes.* Toxicon : official journal of the International Society on Toxinology, 2015. **93**: p. 1-10.

240. Calvete, J.J., P. Juarez, and L. Sanz, *Snake venomics. Strategy and applications.* Journal of mass spectrometry : JMS, 2007. **42**(11): p. 1405-1414.

241. Escoubas, P., L. Quinton, and G.M. Nicholson, *Venomics: unravelling the complexity of animal venoms with mass spectrometry.* Journal of mass spectrometry : JMS, 2008. **43**(3): p. 279-295.

242. Bandeira, N., K.R. Clauser, and P.A. Pevzner, *Shotgun protein sequencing: assembly of peptide tandem mass spectra from mixtures of modified proteins.* Molecular & cellular proteomics : MCP, 2007. **6**(7): p. 1123-1134.

243. Sousa, L.F., C.A. Nicolau, P.S. Peixoto, J.L. Bernardoni, S.S. Oliveira, J.A. Portes-Junior, R.H. Mourao, I. Lima-dos-Santos, I.S. Sano-Martins, H.M. Chalkidis, R.H. Valente, and A.M. Moura-da-Silva, *Comparison of phylogeny, venom composition and neutralization by antivenom in diverse species of bothrops complex.* PLoS neglected tropical diseases, 2013. **7**(9): p. e2442.

244. Serrano, S.M., J.D. Shannon, D. Wang, A.C. Camargo, and J.W. Fox, *A multifaceted analysis of viperid snake venoms by two-dimensional gel electrophoresis: an approach to understanding venom proteomics.* Proteomics, 2005. **5**(2): p. 501-510.

245. Fox, J.W. and S.M. Serrano, *Exploring snake venom proteomes: multifaceted analyses for complex toxin mixtures.* Proteomics, 2008. **8**(4): p. 909-920.

246. Calvete, J.J., *Proteomic tools against the neglected pathology of snake bite envenoming.* Expert review of proteomics, 2011. **8**(6): p. 739-758.

247. Rokyta, D.R., A.R. Lemmon, M.J. Margres, and K. Aronow, *The venom-gland transcriptome of the eastern diamondback rattlesnake (Crotalus adamanteus).* BMC genomics, 2012. **13**: p. 312.

248. Margres, M.J., K. Aronow, J. Loyacano, and D.R. Rokyta, *The venom-gland transcriptome of the eastern coral snake (Micrurus fulvius) reveals high venom complexity in the intragenomic evolution of venoms.* BMC genomics, 2013. **14**: p. 531.

249. Margres, M.J., J.J. McGivern, K.P. Wray, M. Seavy, K. Calvin, and D.R. Rokyta, *Linking the transcriptome and proteome to characterize the venom of the eastern diamondback rattlesnake (Crotalus adamanteus).* Journal of proteomics, 2014. **96**: p. 145-158.

250. Wagstaff, S.C., L. Sanz, P. Juarez, R.A. Harrison, and J.J. Calvete, *Combined snake venomics and venom gland transcriptomic analysis of the ocellated carpet viper, Echis ocellatus.* Journal of proteomics, 2009. **71**(6): p. 609-623.

251. Paiva, O., D. Pla, C.E. Wright, M. Beutler, L. Sanz, J.M. Gutierrez, D.J. Williams, and J.J. Calvete, *Combined venom gland cDNA sequencing and venomics of the New Guinea small-eyed snake, Micropechis ikaheka.* Journal of proteomics, 2014. **110**: p. 209-229.

252. Mann, M. and M. Wilm, *Error-tolerant identification of peptides in sequence databases by peptide sequence tags.* Analytical chemistry, 1994. **66**(24): p. 4390-4399.

253. Mortz, E., P.B. O'Connor, P. Roepstorff, N.L. Kelleher, T.D. Wood, F.W. McLafferty, and M. Mann, *Sequence tag identification of intact proteins by matching tanden mass spectral data against sequence data bases.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(16): p. 8264-8267.

254. Ginter, J.M., F. Zhou, and M.V. Johnston, *Generating protein sequence tags by combining cone and conventional collision induced dissociation in a quadrupole time-of-flight mass spectrometer.* Journal of the American Society for Mass Spectrometry, 2004. **15**(10): p. 1478-1486.

255. Cantor, T., *Sketch of an Undescribed Hooded Serpent with Fangs and Maxillar Teeth*1836: Printed at the Bengal Military Orphan Press by GH Huttmann.

256. O'Shea, M., *Venomous snakes of the world*2008: New Holland Publishers.

257. Whitaker, R., A. Captain, and F. Ahmed, *Snakes of India*2004: Draco Books.

258. Danpaiboon, W., O. Reamtong, N. Sookrung, W. Seesuay, Y. Sakolvaree, J. Thanongsaksrikul, F. Dong-din-on, P. Srimanote, K. Thueng-in, and W. Chaicumpa, *Ophiophagus hannah venom: proteome, components bound by Naja kaouthia antivenin and neutralization by N. kaouthia neurotoxin-specific human ScFv.* Toxins, 2014. **6**(5): p. 1526-1558.

259. Vellard, J., *Variation gèographique du venin de Bothrox atrox.* CR Acad Sci (Paris), 1937. **204**: p. 1369-1371.

260. Vellard, J.-A., *Variations géographiques du venin de" Crotalus terrificus"*1938: Laboratoire d'évolution des êtres organisés.

261. Chippaux, J.P., V. Williams, and J. White, *Snake venom variability: methods of study, results and interpretation.* Toxicon : official journal of the International Society on Toxinology, 1991. **29**(11): p. 1279-1303.

262. Olsen, J.V., B. Macek, O. Lange, A. Makarov, S. Horning, and M. Mann, *Higher-energy C-trap dissociation for peptide modification analysis.* Nature methods, 2007. **4**(9): p. 709-712.

263. Whitelegge, J., *Intact protein mass spectrometry and top-down proteomics.* Expert review of proteomics, 2013. **10**(2): p. 127-129.

264. Ganthavorn, S., *A case of king cobra bite.* Toxicon : official journal of the International Society on Toxinology, 1971. **9**(3): p. 293-294.

265. Tin, M., M. Rai, C. Maung, P. Tun, and D.A. Warrell, *Bites by the king cobra (Ophiophagus hannah) in Myanmar: successful treatment of severe neurotoxic envenoming.* The Quarterly journal of medicine, 1991. **80**(293): p. 751-762.

266. Karnchanachetanee, C., *King cobra bite.* Journal of the Medical Association of Thailand = Chotmaihet thangphaet, 1994. **77**(12): p. 646-651.

267. Veto, T., R. Price, J. Silsby, and J. Carter, *Treatment of the first known case of king cobra envenomation in the United Kingdom, complicated by severe anaphylaxis.* Anaesthesia, 2007. **62**(1): p. 75-78.

268. Kini, R.M. and R. Doley, *Structure, function and evolution of three-finger toxins: mini proteins with multiple targets.* Toxicon : official journal of the International Society on Toxinology, 2010. **56**(6): p. 855-867.

269. Nirthanan, S. and M.C. Gwee, *Three-finger alpha-neurotoxins and the nicotinic acetylcholine receptor, forty years on.* Journal of pharmacological sciences, 2004. **94**(1): p. 1-17.

270. Utkin, Y.N., V.V. Kukhtina, E.V. Kryukova, F. Chiodini, D. Bertrand, C. Methfessel, and V.I. Tsetlin, *"Weak toxin" from Naja kaouthia is a nontoxic antagonist of alpha 7 and muscle-type nicotinic acetylcholine receptors.* The Journal of biological chemistry, 2001. **276**(19): p. 15810-15815.

271. Utkin, Y.N., V.V. Kukhtina, I.V. Maslennikov, A.V. Eletsky, V.G. Starkov, C. Weise, P. Franke, F. Hucho, and V.I. Tsetlin, *First tryptophan-containing weak neurotoxin from cobra venom.* Toxicon : official journal of the International Society on Toxinology, 2001. **39**(7): p. 921-927.

272. Kukhtina, V.V., C. Weise, T.A. Muranova, V.G. Starkov, P. Franke, F. Hucho, S. Wnendt, C. Gillen, V.I. Tsetlin, and Y.N. Utkin, *Muscarinic toxin-like proteins from cobra venom.* European journal of biochemistry / FEBS, 2000. **267**(23): p. 6784-6789.

273. Yap, M.K., S.Y. Fung, K.Y. Tan, and N.H. Tan, *Proteomic characterization of venom of the medically important Southeast Asian Naja sumatrana (Equatorial spitting cobra).* Acta tropica, 2014. **133**: p. 15-25.

274. Konshina, A.G., I.A. Boldyrev, Y.N. Utkin, A.V. Omel'kov, and R.G. Efremov, *Snake cytotoxins bind to membranes via interactions with phosphatidylserine head groups of lipids.* PloS one, 2011. **6**(4): p. e19064.

275. Huang, M.Z., P. Gopalakrishnakone, M.C. Chung, and R.M. Kini, *Complete amino acid sequence of an acidic, cardiotoxic phospholipase A2 from the venom of Ophiophagus hannah (King Cobra): a novel cobra venom enzyme with "pancreatic loop".* Archives of biochemistry and biophysics, 1997. **338**(2): p. 150-156.

276. Escalante, T., A. Rucavado, J.W. Fox, and J.M. Gutierrez, *Key events in microvascular damage induced by snake venom hemorrhagic metalloproteinases.* Journal of proteomics, 2011. **74**(9): p. 1781-1794.

277. de Plater, G.M., R.L. Martin, and P.J. Milburn, *A C-type natriuretic peptide from the venom of the platypus (Ornithorhynchus anatinus): structure and pharmacology.* Comparative biochemistry and physiology. Part C, Pharmacology, toxicology & endocrinology, 1998. **120**(1): p. 99-110.

278. Pung, Y.F., S.V. Kumar, N. Rajagopalan, B.G. Fry, P.P. Kumar, and R.M. Kini, *Ohanin, a novel protein from king cobra venom: its cDNA and genomic organization.* Gene, 2006. **371**(2): p. 246-256.

279. Yamazaki, Y. and T. Morita, *Structure and function of snake venom cysteine-rich secretory proteins.* Toxicon : official journal of the International Society on Toxinology, 2004. **44**(3): p. 227-231.

280. Matsunaga, Y., Y. Yamazaki, F. Hyodo, Y. Sugiyama, M. Nozaki, and T. Morita, *Structural divergence of cysteine-rich secretory proteins in snake venoms.* Journal of biochemistry, 2009. **145**(3): p. 365-375.

281. Kelleher, N.L., P.M. Thomas, I. Ntai, P.D. Compton, and R.D. LeDuc, *Deep and quantitative top-down proteomics in clinical and translational research.* Expert review of proteomics, 2014. **11**(6): p. 649-651.

282. Gault, J., C. Malosse, S. Machata, C. Millien, I. Podglajen, M.C. Ploy, C.E. Costello, G. Dumenil, and J. Chamot-Rooke, *Complete posttranslational modification mapping of pathogenic Neisseria meningitidis pilins requires top-down mass spectrometry.* Proteomics, 2014. **14**(10): p. 1141-1151.

283. LeDuc, R.D., R.T. Fellers, B.P. Early, J.B. Greer, P.M. Thomas, and N.L. Kelleher, *The C-score: a Bayesian framework to sharply improve proteoform scoring in high-throughput top down proteomics.* Journal of proteome research, 2014. **13**(7): p. 3231-3240.

284. Durbin, K.R., R.T. Fellers, I. Ntai, N.L. Kelleher, and P.D. Compton, *Autopilot: an online data acquisition control system for the enhanced high-throughput characterization of intact proteins.* Analytical chemistry, 2014. **86**(3): p. 1485-1492.

285. Li, Y., P.D. Compton, J.C. Tran, I. Ntai, and N.L. Kelleher, *Optimizing capillary electrophoresis for top-down proteomics of 30-80 kDa proteins.* Proteomics, 2014. **14**(10): p. 1158-1164.

286. Cannon, J.R., M.B. Cammarata, S.A. Robotham, V.C. Cotham, J.B. Shaw, R.T. Fellers, B.P. Early, P.M. Thomas, N.L. Kelleher, and J.S. Brodbelt, *Ultraviolet photodissociation for characterization of whole proteins on a chromatographic time scale.* Analytical chemistry, 2014. **86**(4): p. 2185-2192.

287. Fellers, R.T., J.B. Greer, B.P. Early, X. Yu, R.D. LeDuc, N.L. Kelleher, and P.M. Thomas, *ProSight Lite: graphical software to analyze top-down mass spectrometry data.* Proteomics, 2015. **15**(7): p. 1235-1238.

288. Correa-Netto, C., L. Junqueira-de-Azevedo Ide, D.A. Silva, P.L. Ho, M. Leitao-de-Araujo, M.L. Alves, L. Sanz, D. Foguel, R.B. Zingali, and J.J. Calvete, *Snake venomics and venom gland transcriptomic analysis of Brazilian coral snakes, Micrurus altirostris and M. corallinus.* Journal of proteomics, 2011. **74**(9): p. 1795-1809.

289. Petras, D., L. Sanz, Á. Segura, M. Herrera, M. Villalta, D. Solano, M. Vargas, G. León, D.A. Warrell, and R.D.G. Theakston, *Snake venomics of African spitting cobras: toxin composition and assessment of congeneric cross-reactivity of the pan-African EchiTAb-Plus-ICP antivenom by antivenomics and neutralization approaches.* Journal of proteome research, 2011. **10**(3): p. 1266-1280.

290. Lomonte, B., D. Pla, M. Sasa, W.-C. Tsai, A. Solórzano, J.M. Ureña-Díaz, M.L. Fernández-Montes, D. Mora-Obando, L. Sanz, and J.M. Gutiérrez, *Two color morphs of the pelagic yellow-bellied sea snake, Pelamis platura, from different locations of Costa Rica: snake venomics, toxicity, and neutralization by antivenom.* Journal of proteomics, 2014. **103**: p. 137-152.

291. King, G.F. and M.C. Hardy, *Spider-venom peptides: structure, pharmacology, and potential for control of insect pests.* Annual review of entomology, 2013. **58**: p. 475-496.

292. Rodriguez de la Vega, R.C., E.F. Schwartz, and L.D. Possani, *Mining on scorpion venom biodiversity.* Toxicon : official journal of the International Society on Toxinology, 2010. **56**(7): p. 1155-1161.

293. Undheim, E.A., A. Jones, K.R. Clauser, J.W. Holland, S.S. Pineda, G.F. King, and B.G. Fry, *Clawing through evolution: toxin diversification and convergence in the ancient lineage Chilopoda (centipedes).* Molecular biology and evolution, 2014. **31**(8): p. 2124-2148.

294. Yang, S., Z. Liu, Y. Xiao, Y. Li, M. Rong, S. Liang, Z. Zhang, H. Yu, G.F. King, and R. Lai, *Chemical punch packed in venoms makes centipedes excellent predators.* Molecular & cellular proteomics : MCP, 2012. **11**(9): p. 640-650.

295. Liu, Z.C., R. Zhang, F. Zhao, Z.M. Chen, H.W. Liu, Y.J. Wang, P. Jiang, Y. Zhang, Y. Wu, J.P. Ding, and W.H. Lee, *Venomic and transcriptomic analysis of centipede Scolopendra subspinipes dehaani.* Journal of proteome research, 2012. **11**(12): p. 6197-6212.

296. Puillandre, N., P. Bouchet, T.F. Duda, Jr., S. Kauferstein, A.J. Kohn, B.M. Olivera, M. Watkins, and C. Meyer, *Molecular phylogeny and evolution of the cone snails (Gastropoda, Conoidea).* Molecular phylogenetics and evolution, 2014. **78**: p. 290-303.

297. Bouzid, W., M. Verdenaud, C. Klopp, F. Ducancel, C. Noirot, and A. Vetillard, *De Novo sequencing and transcriptome analysis for Tetramorium bicarinatum: a comprehensive venom gland transcriptome analysis from an ant species.* BMC genomics, 2014. **15**: p. 987.

298. Aili, S.R., A. Touchard, P. Escoubas, M.P. Padula, J. Orivel, A. Dejean, and G.M. Nicholson, *Diversity of peptide toxins from stinging ant venoms.* Toxicon : official journal of the International Society on Toxinology, 2014. **92**: p. 166-178.

299. Touchard, A., M. Dauvois, M.J. Arguel, F. Petitclerc, M. Leblanc, A. Dejean, J. Orivel, G.M. Nicholson, and P. Escoubas, *Elucidation of the unexplored biodiversity of ant venom peptidomes via MALDI-TOF mass spectrometry and its application for chemotaxonomy.* Journal of proteomics, 2014. **105**: p. 217-231.

300. Cheek, S., S.S. Krishna, and N.V. Grishin, *Structural classification of small, disulfide-rich protein domains.* Journal of molecular biology, 2006. **359**(1): p. 215-237.

301. Zhao, D.S., Z.R. Gregorich, and Y. Ge, *High throughput screening of disulfide-containing proteins in a complex mixture.* Proteomics, 2013. **13**(22): p. 3256-3260.

302. Zhang, Z. and A.G. Marshall, *A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra.* Journal of the American Society for Mass Spectrometry, 1998. **9**(3): p. 225-233.

303. Vaudel, M., J.M. Burkhart, R.P. Zahedi, E. Oveland, F.S. Berven, A. Sickmann, L. Martens, and H. Barsnes, *PeptideShaker enables reanalysis of MS-derived proteomics data sets.* Nature biotechnology, 2015. **33**(1): p. 22-24.

304. Muth, T., L. Weilnbock, E. Rapp, C.G. Huber, L. Martens, M. Vaudel, and H. Barsnes, *DeNovoGUI: an open source graphical user interface for de novo sequencing of tandem mass spectra.* Journal of proteome research, 2014. **13**(2): p. 1143-1146.

305. Vizcaino, J.A., E.W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Rios, J.A. Dianes, Z. Sun, T. Farrah, N. Bandeira, P.A. Binz, I. Xenarios, M. Eisenacher, G. Mayer, L. Gatto, A. Campos, R.J. Chalkley, H.J. Kraus, J.P. Albar, S. Martinez-Bartolome, R. Apweiler, G.S. Omenn, L. Martens, A.R. Jones, and H. Hermjakob, *ProteomeXchange provides globally coordinated proteomics data submission and dissemination.* Nature biotechnology, 2014. **32**(3): p. 223-226.

306. Kang, T.S., D. Georgieva, N. Genov, M.T. Murakami, M. Sinha, R.P. Kumar, P. Kaur, S. Kumar, S. Dey, S. Sharma, A. Vrielink, C. Betzel, S. Takeda, R.K. Arni, T.P. Singh, and R.M. Kini, *Enzymatic toxins from snake venom: structural characterization and mechanism of catalysis.* The FEBS journal, 2011. **278**(23): p. 4544-4576.

307. Ferreira, S.H., *A Bradykinin-Potentiating Factor (Bpf) Present in the Venom of Bothrops Jararca.* British journal of pharmacology and chemotherapy, 1965. **24**: p. 163-169.

308. Rubin, B., M.J. Antonaccio, and Z.P. Horovitz, *Captopril (SQ 14,225) (D-3-mercapto-2-methylpropranoyl-L-proline): a novel orally active inhibitor of angiotensin-converting enzyme and antihypertensive agent.* Progress in cardiovascular diseases, 1978. **21**(3): p. 183-194.

309. Goncalves-Machado, L., D. Pla, L. Sanz, R.J. Jorge, M. Leitao-De-Araujo, M.L. Alves, D.J. Alvares, J. De Miranda, J. Nowatzki, K. de Morais-Zani, W. Fernandes, A.M. Tanaka-Azevedo, J. Fernandez, R.B. Zingali, J.M. Gutierrez, C. Correa-Netto, and J.J. Calvete, *Combined venomics, venom gland transcriptomics, bioactivities, and antivenomics of two Bothrops jararaca populations from geographic isolated regions within the Brazilian Atlantic rainforest.* Journal of proteomics, 2015.

310. Woolf, C.J., *Pain: morphine, metabolites, mambas, and mutations.* The Lancet. Neurology, 2013. **12**(1): p. 18-20.

311. Calderon, L.A., J.C. Sobrinho, K.D. Zaqueo, A.A. de Moura, A.N. Grabner, M.V. Mazzi, S. Marcussi, A. Nomizo, C.F. Fernandes, J.P. Zuliani, B.M. Carvalho, S.L. da Silva, R.G. Stabeli, and A.M. Soares, *Antitumoral activity of snake venom proteins: new trends in cancer therapy.* BioMed research international, 2014. **2014**: p. 203639.

312. Swenson, S., F. Costa, R. Minea, R.P. Sherwin, W. Ernst, G. Fujii, D. Yang, and F.S. Markland, Jr., *Intravenous liposomal delivery of the snake venom disintegrin contortrostatin limits breast cancer progression.* Molecular cancer therapeutics, 2004. **3**(4): p. 499-511.

313.   Zhou, Q., R.P. Sherwin, C. Parrish, V. Richters, S.G. Groshen, D. Tsao-Wei, and F.S. Markland, *Contortrostatin, a dimeric disintegrin from Agkistrodon contortrix contortrix, inhibits breast cancer progression.* Breast cancer research and treatment, 2000. **61**(3): p. 249-260.

314.   Chippaux, J.-P., *Clinic and treatment of envenomations*2006: Krieger Publishing Company.

315.   Fry, B., *Venomous Reptiles and Their Toxins: Evolution, Pathophysiology and Biodiscovery*2015: Oxford University Press.

316.   Göçmen, B., H. Arikan, A. Mermer, B. Langerwerf, and H. Bahar, *Morphological, hemipenial and venom electrophoresis comparisons of the Levantine Viper, Macrovipera lebetina (Linnaeus, 1758), from Cyprus and Southern Anatolia.* Turkish Journal of Zoology, 2006. **30**(2): p. 225-234.

317.   Böhme, W. and U. Joger, *Eine neue Art des Vipera berus-Komplexes aus der Türkei.* Amphibia-reptilia, 1983. **4**(2): p. 265-271.

318.   GÖÇMEN, B., J. MULDER, M. KARIŞ, and M.A. OĞUZ, *The poorly known Anatolian Meadow Viper, Vipera anatolica: new morphological and ecological data.* HERPETOLOGICA, 2014. **8**: p. 1-10.

319.   Varol Tok, I.U., Murat Sevinç, Wolfgang Böhme, Pierre-André Crochet, Ulrich Joger, Yakup Kaska, Yusuf Kumlutaş, Uğur Kaya, Aziz Avci, Nazan Üzüm, Can Yeniyurt, Ferdi Akarsu *Vipera anatolica - The IUCN Red List of Threatened Species. Version 2015.1.* 2015 04 June 2015]; Available from: www.iucnredlist.org.

320.   Ge, Y., B.G. Lawhorn, M. ElNaggar, E. Strauss, J.H. Park, T.P. Begley, and F.W. McLafferty, *Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry.* Journal of the American Chemical Society, 2002. **124**(4): p. 672-678.

321.   Wagstaff, S.C., P. Favreau, O. Cheneval, G.D. Laing, M.C. Wilkinson, R.L. Miller, R. Stocklin, and R.A. Harrison, *Molecular characterisation of endogenous snake venom metalloproteinase inhibitors.* Biochemical and biophysical research communications, 2008. **365**(4): p. 650-656.

322.   Munekiyo, S.M. and S.P. Mackessy, *Presence of peptide inhibitors in rattlesnake venoms and their effects on endogenous metalloproteases.* Toxicon : official journal of the International Society on Toxinology, 2005. **45**(3): p. 255-263.

323.   Krecsak, L., G. Zacher, and T. Malina, *Clinical picture of envenoming with the Meadow Viper (Vipera (Acridophaga) ursinii).* Clinical toxicology, 2011. **49**(1): p. 13-20.

324.   Igci, N. and D.O. Demiralp, *A preliminary investigation into the venom proteome of Macrovipera lebetina obtusa (Dwigubsky, 1832) from Southeastern Anatolia by MALDI-TOF mass spectrometry and comparison of venom protein profiles with Macrovipera lebetina lebetina (Linnaeus, 1758) from Cyprus by 2D-PAGE.* Archives of toxicology, 2012. **86**(3): p. 441-451.

325.   Lucena, S., R. Castro, C. Lundin, A. Hofstetter, A. Alaniz, M. Suntravat, and E.E. Sanchez, *Inhibition of pancreatic tumoral cells by snake venom disintegrins.* Toxicon : official journal of the International Society on Toxinology, 2015. **93**: p. 136-143.

326.   Sanchez, E.E., A. Rodriguez-Acosta, R. Palomar, S.E. Lucena, S. Bashir, J.G. Soto, and J.C. Perez, *Colombistatin: a disintegrin isolated from the venom of the South American snake (Bothrops colombiensis) that effectively inhibits platelet aggregation and SK-Mel-28 cell adhesion.* Archives of toxicology, 2009. **83**(3): p. 271-279.

327. Calvete, J.J., C. Marcinkiewicz, D. Monleon, V. Esteve, B. Celda, P. Juarez, and L. Sanz, *Snake venom disintegrins: evolution of structure and function.* Toxicon : official journal of the International Society on Toxinology, 2005. **45**(8): p. 1063-1074.

328. Macedo, J.A., M.S. Castro, and J.W. Fox, *Disintegrins and their applications in cancer research and therapy.* Current protein & peptide science, 2015.

329. Maurer, G.D., I. Tritschler, B. Adams, G. Tabatabai, W. Wick, R. Stupp, and M. Weller, *Cilengitide modulates attachment and viability of human glioma cells, but not sensitivity to irradiation or temozolomide in vitro.* Neuro-oncology, 2009. **11**(6): p. 747-756.

330. Ray, A.M., F. Schaffner, H. Janouskova, F. Noulet, D. Rognan, I. Lelong-Rebel, L. Choulier, A.F. Blandin, M. Lehmann, S. Martin, T. Kapp, S. Neubauer, F. Rechenmacher, H. Kessler, and M. Dontenwill, *Single cell tracking assay reveals an opposite effect of selective small non-peptidic alpha5beta1 or alphavbeta3/beta5 integrin antagonists in U87MG glioma cells.* Biochimica et biophysica acta, 2014. **1840**(9): p. 2978-2987.

331. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Analytical biochemistry, 1976. **72**: p. 248-254.

332. Mosmann, T., *Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays.* Journal of immunological methods, 1983. **65**(1-2): p. 55-63.

333. Yalcin, H.T., M.O. Ozen, B. Gocmen, and A. Nalbantsoy, *Effect of Ottoman Viper (Montivipera xanthina (Gray, 1849)) Venom on Various Cancer Cells and on Microorganisms.* Cytotechnology, 2014. **66**(1): p. 87-94.

334. Harvey, A.L., R. Edrada-Ebel, and R.J. Quinn, *The re-emergence of natural products for drug discovery in the genomics era.* Nature reviews. Drug discovery, 2015. **14**(2): p. 111-129.

335. Garg, N.L.-K., T; Caraballo-Rodriguez, M; Melnik, AV; Floros, DJ; Petras, D; Gregor, R; Dorrestein, PC; Phelan, VV, *Natural Products as Mediators of Disease.* Natural product reports, 2015. **submitted**.

336. SÜßMUTH, R., J. KRETZ, V. SCHUBERT, A. PESIC, M. HÜGELLAND, M. Royer, S. COCIANCICH, P. ROTT, D. KERWAT, and S. GRÄTZ, *Albicidin derivatives, their use and synthesis*, 2014, Google Patents.

337. Petras, D.D., T.; Cociancich, S.; Royer, M.; Süssmuth, R.D., *Bioactivity guided mass spectrometric networking reveals new albicidin derivatives.* Analytical Chemistry, 2015. **in preparation**.

338. Hempel, B.F.P., D.; Suessmuth, R.D., *In situ methylation of pABA-3-OH at the peptidyl carrier domains during albicidin biosynthesis.* Master thesis, 2015.

339. Weber, T., K. Blin, S. Duddela, D. Krug, H.U. Kim, R. Bruccoleri, S.Y. Lee, M.A. Fischbach, R. Muller, W. Wohlleben, R. Breitling, E. Takano, and M.H. Medema, *antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters.* Nucleic acids research, 2015. **43**(W1): p. W237-243.

340. Smith, L.A., P.F. Reid, F.C. Wang, D.N. Parcej, J.J. Schmidt, M.A. Olson, and J.O. Dolly, *Site-directed mutagenesis of dendrotoxin K reveals amino acids critical for its interaction with neuronal K+ channels.* Biochemistry, 1997. **36**(25): p. 7690-7696.

341. Petras, D.H., P.; Süssmuth, R.D.; Harris, R.; Calvete, J.J., *Top-down venomics of the east African green mamba, Dendroaspis angusticeps, and the black mamba,*

Dendroaspis polylepis, highlights the complexity of their toxin arsenals. Molecular and Cellular Proteomics, 2015. **submitted**.

342. Hashimi, S.M. and R.G. Birch, *Functional analysis of genes for benzoate metabolism in the albicidin biosynthetic region of Xanthomonas albilineans.* Applied microbiology and biotechnology, 2010. **87**(4): p. 1475-1485.

343. Bostock, J.M., G. Huang, S.M. Hashimi, L. Zhang, and R.G. Birch, *A DHA14 drug efflux gene from Xanthomonas albilineans confers high-level albicidin antibiotic resistance in Escherichia coli.* Journal of applied microbiology, 2006. **101**(1): p. 151-160.

344. Bork, P. and E.V. Koonin, *A P-loop-like motif in a widespread ATP pyrophosphatase domain: implications for the evolution of sequence motifs and enzyme activity.* Proteins, 1994. **20**(4): p. 347-355.

345. Conti, E., T. Stachelhaus, M.A. Marahiel, and P. Brick, *Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S.* The EMBO journal, 1997. **16**(14): p. 4174-4183.

Acknowledgments

First of all I would like to thank my supervisor Professor Roderich Süssmuth. Without his support, all of this work would not have been possible.

Then I would like my second referee and former supervisor Professor Juan Calvete. Thanks a lot for introducing me into the fascination world of venoms.

Many thanks are also owned to Professor Stefan Hüttenhain, who supervised me during my undergraduate studies. Thanks for all your guidance, motivation and coffees, which brought me into science.

Furthermore I would like to thank the undergraduate students I supervised during the last years: Benjamin Hempel, Paul Heiß, Tam Dang, Maurits Brandt and Alexander Platz. Thanks a lot!

You helped me a lot and seeing you grow was even more satisfying than the great results you produced.

More thanks are owned to Dr. Andi Mainz, Hesham Saleh and Dr. Rashed Al Toma who proof-read this thesis and gave me helpful feedback and to all my colleagues of the Süssmuth group for the good work, help, patience, inspiration und good fun in the Lab. Especially: Dennis Kerwart, Leonart von Eckardstein, Siamak Semsary, Maria Kunert, Maria Seidel, Simon Boecker, Lennart Richter, Natalia Jungmann, Martin Jasyk, Dr. Ullrich Keller, Dr. Sebastian Reichau, Dr. Fabiana Salm and Dr. Bartlomiej Krawczyk.

Not to be forgotten shall be all current and former cooperation partners for the fruitful work, trust and sharing their ideas, knowledge and results. Thanks a lot Dr. Stéphane Cociancich, Dr. Monique Royer, Professor Ayse Nalbantsoy, Dr. Nicholas Casewell and Professor Elke Dittmann.

Special thanks are also owned to Professor Pieter Dorrestein and his group who hosted me for three month in their lab in San Diego. It was a great and very inspiring time.

Finally I want to thanks my parents, family and all my friends. You make life worth living.