

Justus-Liebig-Universität Gießen

Department 08 - Biology & Chemistry



Dissertation

Metabolomics-Guided Discovery and Characterization of five new Cyclic Lipopeptides from Freshwater Isolate *Pseudomonas* sp.

Michael MARNER

Gießen, March 2020

1st examiner: PD Dr. Jens Glaeser

2nd examiner: Prof. Dr. Peter Hammann

Contents

1	Abstract	1
2	Introduction	2
2.1	Antibiotic Resistance	2
2.2	Natural product research and Metabolomics	3
3	Development and Evaluation of a Metabolomics platform	6
3.1	Introduction	6
3.2	Material and Methods	12
3.2.1	Cultivation of bacteria	12
3.2.2	Extract preparation	13
3.2.3	Bioactivity assessment	14
3.2.4	Analytics	15
3.2.5	Data bucketing and visualization	16
3.2.6	Variable dereplication via molecular networking	16
3.3	Results	18
3.3.1	Bioactivity	18
3.3.2	Chemical diversity assessment and automatic annotation	19
3.3.3	Molecular networking and variable dereplication	25
3.3.4	Linking bioactivity to causative agent	30
3.4	Discussion	38
3.4.1	Metabolomics	38
3.4.2	Bioactivity	41
4	Bioprospecting and characterization of the bacterial community of Lake Stechlin	44
4.1	Intoduction	44
4.2	Material and Methods	45
4.2.1	Sampling of microorganisms from Lake Stechlin	45
4.2.2	Sample preparation	46
4.2.3	Cell enumeration via fluorescence microscopy	47
4.2.4	Microbiome analysis	47
4.2.5	Cultivation and conservation	49
4.2.6	Bioactivity assessment via quick supernatant <i>lux</i> assay	49
4.2.7	Phylogenetic identification based on 16S rRNA gene sequencing	50
4.3	Results	51
4.3.1	Microbiome analysis	51
4.3.2	Cultivation and bioactivity assessment	55
4.4	Discussion	59
4.4.1	Microbiome analysis	59
4.4.2	Cultivation and Bioactivity assessment	62
5	Metabolomics-guided discovery of new cyclic lipopeptides from <i>Pseudomonas</i> sp. with anti-Gram-negative activity	65
5.1	Abstract	65
5.2	Introduction	65

5.3	Materials and Methods	66
5.3.1	Isolation <i>Pseudomonas</i> sp. FhG100052	66
5.3.2	Bioactivity assessment	66
5.3.3	Screening for chemical novelty	67
5.3.4	Genome sequencing and biosynthetic gene cluster annotation	68
5.3.5	Optimization of production	69
5.3.6	Purification of compounds	69
5.3.7	Structure elucidation using NMR	70
5.3.8	Determination of absolute configuration	70
5.3.9	Optical rotation	71
5.4	Results	72
5.4.1	Bioactivity of crude extract	72
5.4.2	Molecular Network cluster analysis	72
5.4.3	Optimization of production	73
5.4.4	Compound purification and structure elucidation	75
5.4.5	Optical rotation	84
5.4.6	Genome analysis and biosynthetic gene cluster identification	84
5.4.7	Minimum inhibitory concentrations	86
5.5	Discussion	89
5.5.1	Bioactivity	89
5.5.2	Biosynthesis	91
6	Perspective	94
7	Supplements	116
7.1	Metabolomics platform	116
7.1.1	Data processing	116
7.1.2	Bucket intensities	121
7.1.3	Molecular networking cluster	122
7.2	Sampling Lake Stechlin	128
7.3	Microbiome analysis	129
7.4	Bioluminescence Assay layout	130
7.5	μ -fractionation FhG100052 extract <i>C.albicans</i>	131
7.6	MIC determination Assay layout	132
7.7	<i>steABC</i> flanking regions	133
7.8	Optimization of CLP production	134
7.8.1	Media variation	134
7.8.2	Gas exchange	135
7.8.3	Incubation period	137
7.9	Stechlisins Marfey's Analysis	138
7.10	Stechlisins MS/MS fragmentation	140
7.11	Stechlisins NMR spectra	142
7.11.1	Stechlisin B2	142
7.11.2	Stechlisin C3	151
7.11.3	Stechlisin D3	160
7.11.4	Tensin	169
7.11.5	Stechlisin E2	178
7.11.6	Stechlisin F	187
8	Project contributions	196

9 Declaration of Originality/Eigenständigkeitserklärung	197
10 Acknowledgments	198

List of Figures

1	Recalculation of Line spectra schematic	7
2	MS/MS spectral vector comparison schematic	10
3	Example of a MS/MS network schematic	11
4	Metabolomics platform: Bioactivity Actinobacteria	19
5	Metabolomics platform: Principle component analysis of Actinobacteria bucket table	20
6	Metabolomics platform: Metabolic Heatmap of Actinobacteria bucket table	23
7	Metabolomics platform: MS-networking ClusterA Echocide A . . .	26
9	Metabolomics platform: MS-networking ClusterB - Resomycins . .	26
10	Metabolomics platform: MS-networking ClusterC - Anguinomycines	26
8	Metabolomics platform: MS-Network overview	27
11	Metabolomics platform: MS-networking Cluster D - Naphtocyclinones	29
12	Metabolomics platform: MS-networking Cluster E - Scopafungines	30
13	Metabolomics platform: μ -fractionation of ST101789 in 5315 . . .	31
14	Metabolomics platform: μ -fractionation of ST106693 in 5315 . . .	33
15	Metabolomics platform: μ -fractionation ST107645	35
16	Metabolomics platform: μ -fractionation ST107165	37
17	Lake Stechlin: Krona chart of plankton associated bacterial community	53
18	Lake Stechlin: Krona chart of bacterial community in lake water . .	54
19	Lake Stechlin: Phyla distribution associated to plankton and in water	55
20	Lake Stechlin: OTU distribution associated to plankton and in water	56
21	Lake Stechlin: Stechlin cultivates exhibiting bioactivity against <i>E. coli</i> DH5 α	57
22	Lake Stechlin: Dereplication of Aerobactin in FhG100039 extracts	58
23	Stechlisins: Molecular Network cluster analysis of FhG100052 . .	73
24	Stechlisins: Media optimization for increased CLP production . .	74
25	Stechlisins: Structure and absolute configuration of isolated Stechlisins and Tensin	81
26	Stechlisins: Proposed structures of minor compounds based on MS/MS	83
27	Stechlisins: Phylogenetic analysis aminoacid sequences of A-domain of <i>steABC</i>	85
28	Stechlisins: Biosynthetic gene cluster and proposed biosynthesis of Stechlisins	86
S1	Metabolomics platform: Custom script used for LCMS data processing	117
S2	Metabolomics platform: Custom script used for metabolic heatmap generation	118
S3	Metabolomics platform: Absolute intensity of selected buckets . .	121
S4	Metabolomics platform: MS-networking Cluster - Conglobatin . .	122
S5	Metabolomics platform: Echocide A Reference comparison	124

S6	Metabolomics platform: Anguinomycin A Reference comparison .	125
S7	Metabolomics platform: β -Naphthocyclinone-epoxide Reference comparison	126
S8	Metabolomics platform: Scopafungin Reference comparison	127
S9	Lake Stechlin: Biomass maxima Lake Stechlin	128
S10	Lake Stechlin: Phyla distribution in percent	129
S11	Lake Stechlin: Bioluminescence assay layout	130
S12	Lake Stechlin: μ -fractionation of FhG10052 extract	131
S13	MIC plate design	132
S14	Stechlisins: <i>steABC</i> flanking regions	133
S15	Stechlisins: Optimization of C-source and cultivation duration .	134
S16	Influence of culture vessel on CLP biosynthesis	135
S17	Influence of culture vessel on cell density	136
S18	Stechlisins: Kinetics of Tensin biosynthesis by strain FhG100052 .	137
S19	Stechlisins: Marfey's derivatization	138
S20	Stechlisins: Marfey's derivatization continued	139
S21	Stechlisins: Proposed structures of minor Stechlisins based on MS/MS	141
S22	NMR spectra Stechlisin B2	142
S31	NMR spectra Stechlisin C3	151
S40	NMR spectra Stechlisin D3	160
S49	NMR spectra Tensin	169
S58	NMR spectra Stechlisin E2	178
S67	NMR spectra Stechlisin F	187

1 Abstract

In order to fill the discovery void of novel anti Gram-negative substances, a selection of metabolomic tools was established for industrial routine application. The herein described platform was tested with a set of *Streptomyces* strains, before environmental isolates were analyzed. First, two methods for quality control and chemical diversity assessment were compared. It appeared that a vector space model reduces the effect of outliers, compared to a principle component analysis, and is thereby most suitable to describe heterogeneous data such as metabolome comparisons. Second, the pipeline of data generation yielded unsupervised annotations of many literature known as well as structurally related, yet not described, molecules (variable dereplication) in the bacterial compound mixtures. The majority of compounds was detected by both applied methods, mass spectrometry (MS) based bucket annotation and tandem MS based molecular networking.

In the following, the survey for new anti Gram negative compound producers was extended to environmental isolates. In that sense, the potential of the bacterial community of Lake Stechlin (Brandenburg, Germany) was evaluated. Different sample types, namely plankton associated and free living microorganisms, were retrieved from the lake and analyzed on the basis of their microbiome composition. The observed OTU distribution pattern was comparable to previously published observations, thus was considered to genuinely reflect the natural bacterial composition of the lake. The different samples types were shown contain distinct bacterial communities, thus made a great overall biodiversity available for further experiments.

Prioritization of isolated bacteria from Lake Stechlin was carried out by cell free supernatant assays against *E.coli* DH5 α . A metabolom analysis led to the discovery of a undescribed group of cyclic lipopeptides (CLPs) in the culture broth of *Pseudomonas* sp. FhG100052. Culture condition optimization facilitated the isolation and subsequent structure elucidation of the five new compounds. Characterization comprised extensive MS/MS and NMR experiments in combination with Marfey's analysis. The data were in agreement with *in silico* analysis of the corresponding biosynthetic gene cluster (BGC). The new compounds resemble members of the Amphisin group [147] as they are constructed of a 3-hydroxy fatty acid linked to the N-terminus of an undecapeptide core. Most strikingly, the length of the incorporated fatty acid seems to define the moderate growth inhibitory effects against the Gram negative pathogen *Moraxella catarrhalis* FH6810 as observed by MIC values ranging from no inhibition ($> 128 \mu\text{g/mL}$) to $4 \mu\text{g/mL}$.

2 Introduction

2.1 Antibiotic Resistance

Globally, infectious diseases remain top ten killers of humans. In 2016, lower respiratory tract infections, diarrheal diseases and tuberculosis claimed about 6 million lives, translating to 10 % of all death worldwide [133]). In the near future, emergence and spread of antimicrobial resistance (AMR) among pathogenic bacteria might intensify the situation by impairing the lifesaving potency of antibiotics [92] [104] [132] [10] [162]. The U.S. Centers for Disease Control and Prevention (CDC) estimated more than 2.8 million infections by antibiotic-resistant microorganisms and >35.000 death following such infections in the U.S. in 2019 [134]. While bacteria evolve towards drug resistance, discovery and development of novel antibiotics continues to challenge the scientific community [28] [4].

The need for novel antibiotics to protect human health is not a new phenomenon in developed countries. Since Prontosil was introduced as the first antibacterial drug in 1935 [61], discovery and clinical use of an antibiotic was always followed by resistance development of the treated pathogens [96]. Besides unregulated prescription, the irresponsible usage of antibiotics in livestock and aquaculture treatment, as well as in horticulture, food preservation and industrial processes like ethanol production are relevant factors contributing to this problematic issue [114]. Dynamic resistance development is not surprising, as wide-spread application of antibiotics does in fact select for resistant mutants: Even though, proofreading increases the fidelity of DNA polymerases during replication by a factor of 10-100 [90], mutations occur frequently and give rise to an enormous genetic variability and by chance resistance development. In addition to spontaneous mutation, resistance to any natural antibiotic is intrinsically encoded in the genome of the producer strain. Therefore, it is only a matter of time until resistance spreads in response to the strong selection pressure during antibiotic treatment. Common resistance genes mediate drug target modification or target overexpression, bypassing pathways, efflux systems or direct enzymatic inactivation of the antibiotic [19]. Besides, AMR genes are frequently located on extra chromosomal elements such as plasmids, which significantly facilitate gene transfer, thus distribution, even across species borders [145].

The situation is especially severe for Gram-negative bacteria as they are additionally protected by an outer membrane composed of a asymmetrical lipid bilayer: The outer leaflet is mainly constructed of amphiphilic lipopoly- and oligosaccharides (LPS). While the lipophilic part (lipid A) is tightly anchored in the membrane, the hydrophilic part extends away from the cell restricting

diffusion or incorporation of hydrophobic molecules over/in the membrane [157]. Phosphate and acid moieties account for the overall negatively charged membrane surface. On the other side, the inner leaflet of the outer membrane is mostly composed of phospholipids and lipoproteins responsible for connection to the peptidoglycan layer [24]. Within the OM, the most abundant proteins are β -barrel forming channels, referred to as porines. These, usually water filled, pores allow passive diffusion of small, polar solutes ($< 600\text{Da}$) and determine the general diffusion properties of the OM [51] [77]. This specific diffusion barrier prevents many antibiotics from reaching their molecular target within the cell and thereby dramatically reduces the strains susceptibility [116] [113]. Being surrounded by the inner and outer membrane, the periplasm represents a further multipurpose (incl. defense) compartment. Some antibiotics, like the 'last resort' carbapenems might be able to pass through the porines to reach their molecular target (peptidoglycan biosynthesis), but are immediately rendered harmless by specifically designed enzymes, such as metallo- β -lactamases, allocated within the periplasm [102].

In that sense, it was demonstrated that screening of an unbiased compound library produces roughly 10 to 100 times more hits against Gram-positive bacteria, such as methicillin-resistant *Staphylococcus aureus*, compared to Gram-negative strains [48]. The World Health Organization (WHO) defined multi-drug resistant (MDR) Gram-negative bacteria as a critical threat to global health and emphasized that research effort should focus on antibiotics active against those [151]. A traditional but nevertheless valuable source of antimicrobial compounds are natural products isolated from bacteria or fungi.

2.2 Natural product research and Metabolomics

Natural products (NPs) or secondary metabolites are commonly defined as naturally derived, low molecular weight molecules, which are not directly involved in the primary metabolism of the producer. These specialized molecules do generally not play a role in growth, development and reproduction of an organism, but are rather a result of adaptation to specific environments [14]. Evolutionary shaped features range from trace element allocation over intra and inter species communication to chemical defense or deterrence. Hence, NP biosynthesis is tightly connected to environmental stimuli and precisely regulated. Numerous examples demonstrate successful clinical application of the intrinsic therapeutic character of specific NPs, most notably anti-infectives. Curiosity and medical need paired with economic interests fueled extensive drug discovery programs and led to the discovery of 17 of the 21 antibiotic classes, most of them isolated from bacteria or fungi [28].

In fact, unaltered natural products or structures derived from natural products (e.g. semi-synthetically modified) contribute $\sim 75\%$ to all approved antibacterial agents (1981-2014)[125] [124]. Besides the evolutionary aspect and broad structural diversity, a higher degree of heteroatoms and increased average polarity [129] compared to synthetic libraries might explain the importance of NPs. A potent antimicrobial agent, in contrast to other therapeutics, does not necessarily follow "Lipinski's rule of five" [101] and exhibits weak lipophilicity as for instance reflected by low/negative clogD values and greater polar surface area (PSA). In general it appears that synthetic libraries and corporate archives do not sufficiently cover the specific physicochemical space of antibacterials [129]. Up today, NPs remain a prolific source for novel chemical entities suitable for pharmaceutical development [28] [1] [121], although all recent research campaigns often suffer from high re-discovery rates.

Due to the intensive bioassay-guided exploration of the microbial biosphere, the discovery of antimicrobial substances became increasingly challenging. However, evidence accumulates that this is not caused by an exhausted pool of structures, but rather a lack of novel harvesting strategies. Non traditional cultivation approaches, such as specifically designed diffusion chambers [97] or droplet microfluidics [169], help to expand the *in-vitro* biodiversity and to potentially access new producer strains. Additionally, the increasing availability of whole genome data demonstrated that the genetic capacity of an organisms is usually considerably greater than the number of reported compounds. Finally, technological progress enabled the design of untargeted secondary metabolite surveys [87]:

Innovation of analytical instrumentation and methodology (e.g. invention of Ultra-high performance liquid chromatography in line with high resolution tandem mass spectrometry UPLC-HRMS/MS) allowed to increase the number of newly characterized microbial natural products from a few in the 1940s-50s to an average of ~ 1600 compounds per year [136] since the 1990s. The opportunity to study the microbial metabolite output in greater detail, including low intensity signals, profoundly supported the discovery, while simultaneously creating new challenges. The instrumentational performance is only as powerful as the downstream data mining processes, especially when dealing with gigantic datasets like UPLC-HRMS/MS files of complex environmental samples. In this context, secondary metabolomics [4] [61] [87] based methods help to identify signals of interest, enable automatic annotation against library compounds and facilitates structural characterization of unknown compounds. Tandem Mass spectrometry networks (MS/MS networks) have proven to be particularly valuable tool for data visualization and interpretation [166]. Comparison of vector orientation based on MS/MS fragmentation patterns is used to group compounds according to their

structural similarity. Measured spectra are correlated to each other, amended with related items from spectral libraries and mapped together in one network. Thereby, new derivatives are automatically connected to their already described relative(s).

Although new structures are constantly discovered, the structural diversity among them and their published predecessors declined over time (Tanimoto similarity median of newly described structures p.a. >0.65) [136]. However, in natural product research, scientific value is not necessarily linked to novelty. Truly novel scaffolds might be more appealing than variants of known molecules, but it is known that even small structural alternations can determine the degree of biological activity or toxicity. Hence, it is important to realize the substantial value of derivative structures. Besides the possibility to exhibit greater potency, derivatives might contribute to mode of action studies by establishing structure - activity relationships (SARs).

Public-Private-Partnerships In order to support the discovery of novel bioactive substances, continuous effort and new research strategies are urgently needed. Besides scientific creativity, economic innovation and modern business models are required to carry on the expensive search for new NPs. The concept of *private public partnerships* comprises one of the most promising approaches: By sharing costs and knowledge between experienced pharmaceutical companies and academic research facilities, the financial risk is reduced and a stimulating environment for idea exchange created. In 2014, the Fraunhofer Institute for Molecular Biology and Applied Ecology (IME) and Sanofi established the *Natural product center of excellence* [50] and later the *Fraunhofer-Evotec Natural product excellence center*. These unique partnerships bring together innovation-inclined academic research, state of the art equipment, as well as far-reaching expertise, and is, thus, forming a promising drug discovery platform.

3 Development and Evaluation of a Metabolomics platform

3.1 Introduction

Natural product discovery programs usually rely on the analysis of complex compound mixtures (extracts), routinely carried out by LC-MS. While a MS experiment of a single sample can easily generate thousands of spectra, a manual analysis of a complete set of extracts is almost impossible to realize in an adequate period of time. Many working groups contributed to the development of chemoinformatic tools to effectively deal with "Big Data" generated by increasingly sensitive mass spectrometers. In that sense, computerized chemical profiling has already proven its value when dealing with complex mixtures of secondary metabolites within biological extracts. The term *secondary metabolomics* comprises a range of algorithm aided MS data mining approaches applicable in various fields of research. For instance, metabolomics analysis can rapidly elucidate changes in the metabolite output resulting from altered gene expression by comparing ion intensities across samples.

LC-MS based The basis of most metabolomics techniques involves a dimension reduction of multivariate LC-MS data to allow the application of sophisticated statistics. *Data bucketing* describes a process converting raw three dimensional LC-MS data (retention time R_t , mass to charge ratio m/z , ion intensity I) into two dimensional data matrices [49]. The first step when analyzing metabolite profile spectra, often includes the separation of actual information from background noise (denoising) [149] by peak detection and picking algorithms. Depending on the sample and the instrument type used, specific peak finder algorithms are applied. In commercially available software packages such as DataAnalysis 4.4 ©(Bruker Daltonik GmbH), multiple peak finder algorithms are available to recalculate line spectra from the recorded profile spectra or simply generate a denoised list of masses. For instance, when working with broad peaks (e.g. protein samples) a common principle among peak picking tools is the detection of peak centroids followed by the analysis of flanking regions. Here, a data point is considered a real (not a noise) signal, if it features a large m/z distance to its nearest neighbor on the m/z axis, but a small distance to its nearest neighbor in the following spectrum [149]. Another widely used method, especially when working with time-of-flight data, is the sum peak finder algorithm. The major parameter of the sum peak finder to discriminate analyte and noise peaks is the definition of a suitable full width at half maximum (FWHM) threshold. Usually, intense peaks extend over

a larger m/z range (larger FWHM) compared to noise signals. Denoised data contain mostly desired peak signals (above selected FWHH, minimum I and S/N threshold) and are the basis for further processing (s. Figure 1). The molecular feature algorithm assumes a high time correlation of ions belonging to the same compound and interprets the mass distances between them. Thereby signals belonging to the same analyte (isotopes, different adduction or charging states) are subsequently linked and reported as a single compound or molecular feature. Finally, around each (still three dimensional) item in the list of molecular features a defined area (*bucket*) within the $R_t - m/z$ space is created (for example [49] or [88]). Thereby, both values (R_t and m/z) of a given signal are stored in one single artificial unit, the bucket, without the loss of any information. Besides dimension reduction by binning of information, bucketing can additionally reduce the the impact of peak shifting across samples by carefully selecting an adequate ΔR_t [149]. If a bucket is generated in one sample, all other samples of the experiment are searched at the particular $R_t - m/z$ area and peak intensities found are included in the final aligned data matrix.

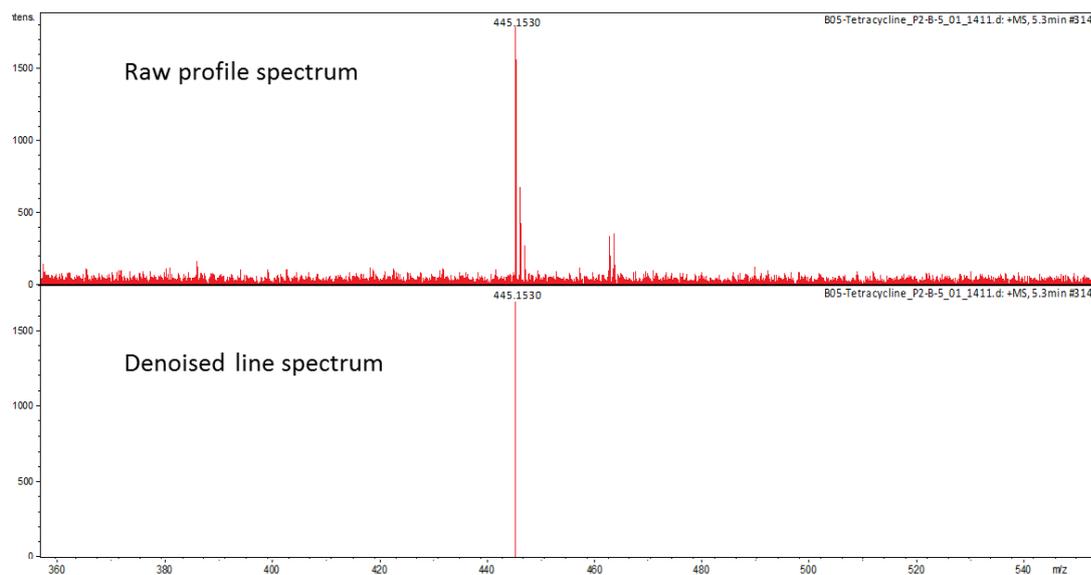


Figure 1: Signals above the selected absolute I (10^3) and $S/N(10)$ threshold are considered data points and used for further processing. Explanatory spectra generated by injection of 1 μL of a 10 $\mu\text{g}/\text{mL}$ tetracycline solution in MeOH.

Metabolic fingerprinting of biological extracts by LC-MS data bucketing represents a classical "Large K , small N problem". For the limited amount of samples N (extracts measured) thousand of variables K (buckets) are generated. A way to determine the structure of such unbalanced data ($N \gg K$) are multivariate analysis like partial least squares (PLS) or principle component analysis (PCA). These help to visualize complex data sets by determination of the informative

dimensionality within the multidimensional data.

In an iterative process, the directions contributing most to the overall variance in the data (highest eigenvalue) are determined (= Principle component 1 (PC1) or eigenvector 1). The direction explaining the second most of the variance in the data set is called PC2 and located perpendicular to PC1 [134] [133]. Dimension reduction is achieved by weighing the influence of each variable of a particular sample on the principle components. The sum of the products of each variable x and its weight b (loadings) is called score u (Equation 1) and represents one numerical value for the sample. Hence, the score describes the combined original variables by one new lateral variable. Essentially a lateral variable can be defined as a formal combination (a mathematical function) of measured variables of a given sample. The calculated score values can be plotted in a two dimensional scatter plot, in which the axis describe the most variation in the data (PC1 and PC2).

$$u = b_1x_1 + b_2x_2 + \dots + b_mx_m \quad (1)$$

By visualizing a metabolomic dataset, the PCA allows exploratory data analysis: Within the *scores plot*, samples with similar bucket intensity distribution, thus chemical composition, would cluster together (similar u). Extracts containing different influential or characteristic buckets would produce distinct PC1 and PC2 scores, thus would cluster away from the group of similar extracts. The scores plot is consequently the primary result of a PCA and helps to determine the underlying structure, in this case the chemical composition of the extracts, as well as the identification of outliers or unique extracts at one glance.

LC-MS/MS based In addition, an in-house semi-automatic dereplication platform based on MS/MS fragmentation signature comparison was implemented. This includes offline comparison of experimental MS/MS spectra against in-house databases (like Sanofi pure compound libraries) amended with *in silico* fragmented compounds [6] from commercial databases such as Antibase [91]. The molecular networking [138] [161] [166] workflow represents an straightforward method to simplify and visualize extensive amounts of data. Molecular networking helps the scientist to focus on relevant signals (chemical novelty), identify background signals like medium components, annotate already known compounds and pinpoint structural relationships of known and unknown molecules. In principle, fragmentation signatures of all measured parent ions are pairwise compared to each other and a spectral library of reference compounds. Each precursor ion is expressed as a vector in an n -dimensional space with its specific fragments being

the attributes of that vector ($n =$ number of fragments) (s. Figure 2). Essentially, the specificity of the fragmentation is translated into the direction of the precursor vector in space. Therefore, the vectors can be normalized to unit vectors without the loss of information relevant for the analysis. Comparison between two precursor ions is then carried out by calculating the cosine similarity ($\cos \theta$) between the two (unit)vectors. Vectors, with $\cos \theta = |1|$ would have the same direction in space and thereby would have identical attributes. Vector with $\cos \theta = 0$ are perpendicular to each other, thus their attributes are completely different. In summary, a pair of molecules, which have a similar fragmentation signature, thus share structural features, have a $\cos \theta$ close to 1, while the relationship between distinct molecules is expressed by $\cos \theta$ values close to 0.

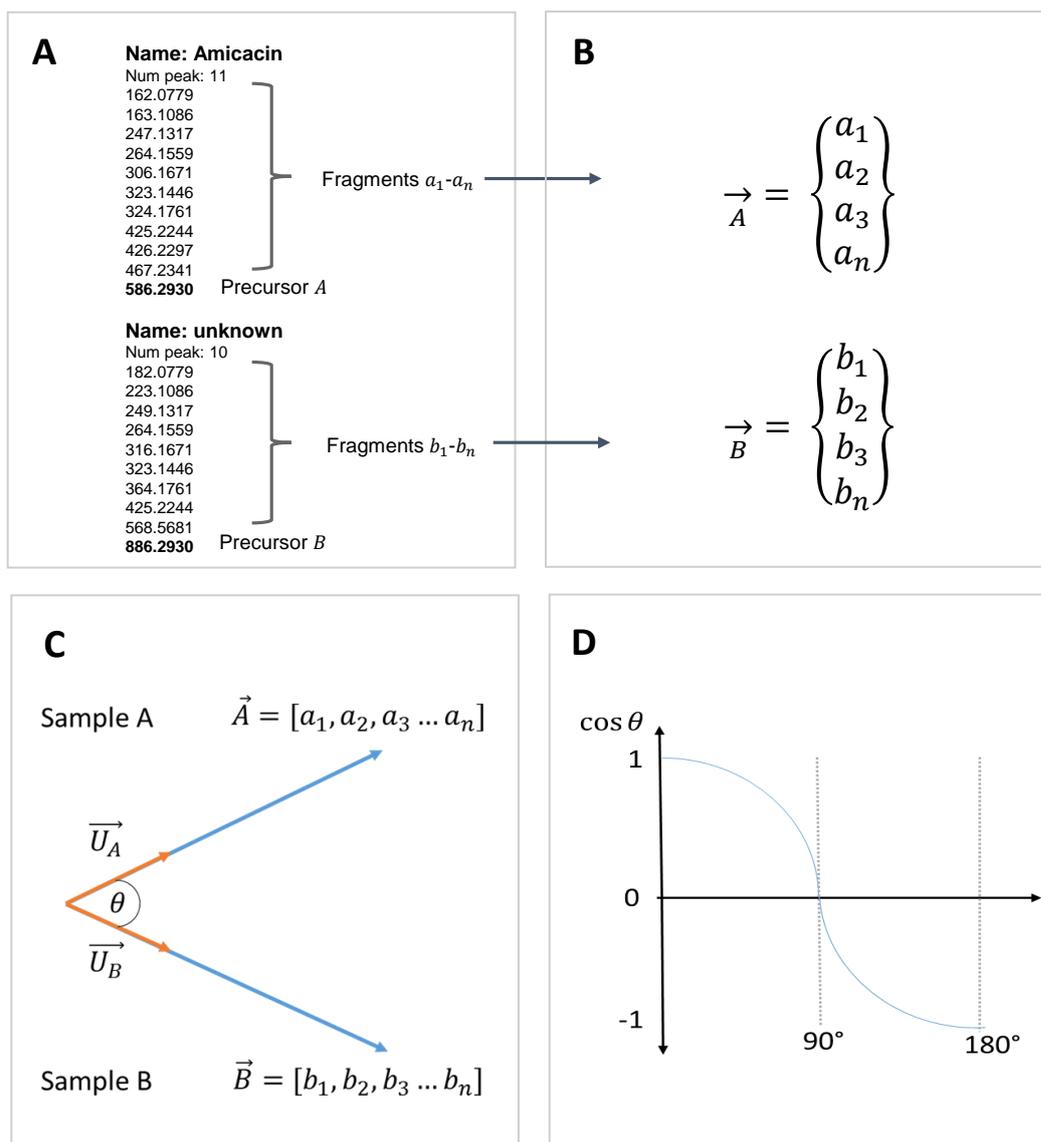


Figure 2: Simplified MS/MS spectral vector comparison of Amicacin and hypothetical unknown molecule B: A) Masslist of the two fragmented molecules. B) Precursors are expressed as vector with their specific fragments as attributes. Thereby the specificity of the fragmentation patterns is conserved in the direction of the vector in space. C) Fragmentation signatures of molecules are compared by calculation of \cos between the unit vectors in space. D) Relationship of \cos between compared molecular vectors and $\cos \theta$.

In conclusion, a molecular network is a map of all MS/MS signals in a given set of samples, which satisfy selected parameters like minimum amount of fragments or clustering partners (s. Figure 3). Each node within the network represents a precursor ion labeled with its m/z . A connection (edge) is generated between two nodes if they share a certain $\cos \theta$ value (usually ≥ 0.7). Thereby, molecule families of similar structure form clusters within the network. By calculation of $\cos \theta$ values for all measured signals and reference compounds, precursors get

automatically annotated ($\cos \theta \geq 0.95$) and structural relationships to of library compounds and unknown signals are expressed in the network. Once all $\cos \theta$ values are calculated, the network can be explored using visualization tools like Cytoscape [146].

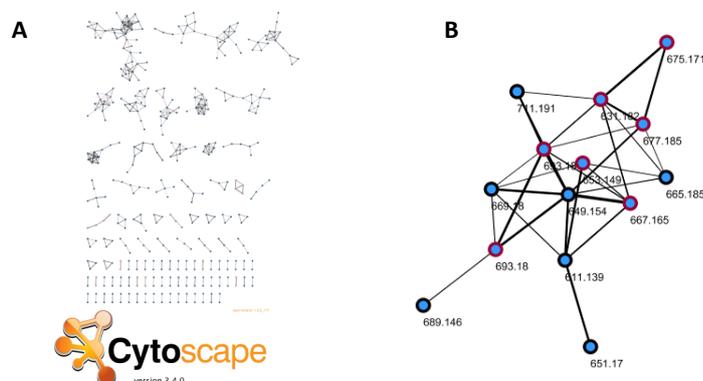


Figure 3: Example of a MS/MS network A) Overview of a complete network visualized with Cytoscape. B) Magnified cluster of the network. Precursor ions are represented as nodes. Red border of node indicates high structural similarity with a spectral reference library compound ($\cos \theta \geq 0.95$). After one-to-one comparison, precursor ions which share many fragments ($\cos \theta \geq 0.7$) are connected by edges to form compound families. Thickness of edge is a proxy for similarity among the connected nodes, thus represents the $\cos \theta$ value

Aim of this study In the herein reported experiment, a set of Actinobacteria (*Streptomyces* sp.) from the Sanofi strain collection [50] was chosen to construct and evaluate an *in house* metabolomics platform for industrial purposes. The platform should help to simplify and visualize UPLC-HRMS/MS data to get an first impression of the chemical diversity within the data set. Second, it should help to focus on relevant signals (chemical novelty) by identifying background signals and annotating already known compounds within the crude extracts. Furthermore, it should pinpoint structural relationships between annotated database compounds and unknown, not yet investigated molecules. Preferably, this work flow should operate as unsupervised as possible.

The genus *Streptomyces* was selected as affiliated bacteria are famous natural product producer and were broadly investigated in the past [163], hence antimicrobial NPs isolated from this genus are well represented in public databases as well corporate chemical libraries. The success of any (semi-) automatic dereplication approach is tightly connected to the size and quality of the spectral reference library accessible. If a lot of information is available, the chance to recognize one piece of that information in an unknown context might be higher. In favor of that,

we were able to expand our database with unpublished NPs obtained from the chemical library of our cooperation partner Sanofi. The fact that *Streptomyces* are talented NP producers and were already extensively studied, contributes to challenge the herein proposed hypothesis. This study tries to demonstrate the unbowed value of Actinobacteria in NP research. It is hypothesized, that advances in both, instrumentation and downstream data analysis, will help to see what was overseen in the past.

3.2 Material and Methods

3.2.1 Cultivation of bacteria

A selected set of *Streptomyces* strains (ST106693, ST101789, ST106693 and ST107645) was fermented under different nutrient regimes. Strains were activated from cryostocks by incubation on 5254 agar at 28 °C until colony formation could be observed by eye. After quality control by stereo microscopy, pure strains were transferred into submerge pre-culture II in 5254 broth. After 5 days, main cultures (50 mL in 300 mL Erlenmeyer flasks) were inoculated in 5315, 5294 or 5254 broth, using pre-culture II (2% v/v inoculum). Main-cultures were incubated for seven days at 28 °C and 180 rpm. Each cultivation was carried out in triplicate.

5315

Oatmeal 20 g * L⁻¹
2.5 mL trace element solution 5314
pH 7.2

5314

3 g * L⁻¹ CaCl₂ * 2H₂O
1 g * L⁻¹ Fe(III)-citrate
0.2 g * L⁻¹ MnSO₄ * H₂O
0.1 g * L⁻¹ ZnCl₂
0.025 g * L⁻¹ CuSO₂ * 5H₂O
0.02 Na₂B₄O₂ * 10H₂O
0.004 g * L⁻¹ CoCl * 6H₂O
0.01 g * L⁻¹ Na₂MoO₄

5294

5 g * L⁻¹ starch (soluble)
10g * L⁻¹ glucose
5 g * L⁻¹ peptone
2 g * L⁻¹ yeast extract

1 g * L⁻¹ NaCl
3 g * L⁻¹ CaCO₃
10 mL glycerin (99.5%)
2.5 mL corn steep (liquid)
pH 7.0

5254

15 g * L⁻¹ glucose
15 g * L⁻¹ soy flour
5 g * L⁻¹ corn steep (solid)
2 g * L⁻¹ CaCO₃
5 g * L⁻¹ NaCl
optional: 18 g * L⁻¹ agar
pH 7.0

3.2.2 Extract preparation

Cultivation was stopped by cooling bacterial cultures as well as medium controls to - 50 °C. Frozen samples were lyophilized (Christ delta 2-24 LSCplus) and subsequently subject of extraction. First, 50 mL of methanol was added to the dried samples and incubated for 2h at 180 rpm. The redissolved suspension was transferred into a 50 mL polypropylene tube (Greiner) and centrifuged at 3320 x g for 15 minutes. Supernatants were filtered over a 30 µm filter (Miltenyi Biotec) into a new 50 mL tube and evaporated to dryness (SpeedVac. Thermo). Dried extracts were concentrated in 1 mL methanol (2h at 28 °C, 4 °C overnight), centrifuged (3320 x g, 30 min) and finally transferred into a 96 deep-well-plate (Masterblock®, Greiner). From this plate, 60 µL were copied into a 96-well V-bottom plate (Greiner) and sealed with piercable cap mats (Micronic, Netherlands). A total of 600 µL was transferred into storage tubes (Micronic) arrayed in 96-well format. The remaining extract (~ 150 µL) was again dried *in vacuo*, before 75 µL dimethylsulfoxide (DMSO, Sigma) was used to further concentrate the extracts. After centrifugation (3320 x g, 10 min), supernatants were copied to a new V-bottom plate (=Assay master plate). An automatic liquid handling system (CyBi®, Jena Analytics) was used to distribute extracts from the assay master plate to 384-well assay plates (Greiner). A three point dilution series (0.5 µL, 0.25 µL and 0.125 µL twice) was prepared for each extract.

3.2.3 Bioactivity assessment

Microbroth dilution assay for extract screening The 100x concentrated methanolic crude extracts were screened for growth inhibitory activity against a set of clinically relevant human pathogens, including *Escherichia coli* ATCC35218, *Pseudomonas aeruginosa* ATCC27853, *Staphylococcus aureus* ATCC25923, *Mycobacterium smegmatis* ATCC607 and *Candida albicans* FH2173. Briefly, an overnight culture (37°C, 180rpm) in cation adjusted Mueller Hinton II medium (BD) was adjusted to $2 * 10^4$ cells/mL or for *C. albicans* FH2173 to $1 * 10^5$ cells/mL. For all strains, the adjusted cell suspension was prepared in Mueller Hinton II as assay medium. In addition, *E.coli* ATCC 35218 was screened in Mueller Hinton II medium supplemented with physiological concentrations of bicarbonate ($3.7 \text{ g} * \text{L}^{-1}$; MHC) and minimal mineral medium (M9). The extract ($0.5\mu\text{L}$, $0.25\mu\text{L}$ and $0.125 \mu\text{L}$ twice) aliquots within in the 384 well microtiter assay plates were supplemented with $50\mu\text{L}$ cell suspension representing each test strain. Gentamycin (*E. coli* and *P. aeruginosa*, *S. aureus*), Nystatin (*C. albicans*) or Isoniazid (*M. smegmatis*) were added as a positive control. A dilution series of the antibiotic was prepared ($256\text{-}0.078 \mu\text{g/mL}$) to ensure that concentrations achieve a range of effects from complete to no growth inhibition. Cell suspensions without the extract and antibiotic were used as negative controls. After incubation (18h, 37°C, 180rpm, 95% rH) cell growth was assessed by measuring the turbidity with a microplate spectrophotometer at 590 nm (LUMIstar®Omega BMG Labtech). *C.albicans* and *M.smegmatis* were incubated for two days before microbial viability assays (BacTiter-Glo™, Promega) were carried out to assess extract potency. The positive control containing the highest antibiotic concentration represents complete inhibition of microbial growth and was considered *blank* or *low count*, while the negative control was considered to exhibit maximal microbial growth (*High count*). The percent growth inhibition was calculated from the absorption units (AU) or luminescence units (LU):

$$\text{Growth inhibition}[\%] = 100 * \left[1 - \frac{AU_{\text{Sample}} - AU_{\text{Low}}}{AU_{\text{High}} - AU_{\text{Low}}}\right] \quad (2)$$

μ-fractionation of crude extracts Extracts showing at least 85% growth inhibition were considered bioactive. These extracts were partitioned into 159 fractions (~ 9s) by reversed-phase liquid chromatography using a BEH C₁₈ column (Agilent 1290 Infinity®LC) and were recorded by QTOF-MS/MS (maXis II™Bruker Daltonics). The fractions were collected in 384-well plates using a custom fraction collector (μFRACS, Zinsser Analytics) and rescreened against the same test strain. In addition to turbidity assays based on optical density, we conducted microbial

viability assays (BacTiter-Glo™, Promega) on the fractionated extracts according to the manufacturer's instructions, applying the same positive controls, negative controls and growth inhibition calculation (Equation 2).

3.2.4 Analytics

Acquisition of mass spectra was carried out by ultra-high performance liquid chromatography - (tandem) mass spectrometry - photodiode array - evaporative light scattering detector (UHPLC-MS/MS-PDA-ELSD) measurements using a maXisII™ (Bruker Daltonics) high resolution mass spectrometer in line with an Agilent 1290 infinity LC system. The column (Waters, Acquity UPLC BEH C₁₈, 30 Å, 1.7 µm, 2.1 mm * 100 mm) was kept at a constant temperature of 40 °C during all measurements. A sample volume of 1 µL was injected. A linear gradient of water (A) and acetonitrile (B), both supplemented with 0.1 % formic acid, at a constant flow of 600 µL * min⁻¹ was used to separate the analytes by reverse phase chromatography.

UV spectra of elutes were recorded via photodiode array (PDA) at 205-640nm. Subsequently samples were splitted: 90 % of the sample volume was analyzed via ELSD (Agilent 1290 Infinity ELSD G4261B) and 10 % by mass spectrometry. Mass accuracy was guaranteed by direct injection of a 50 % sodium formate calibration solution (Sigma) into the MS immediately before the first experiment. The same sodium formate solution was used as internal calibration standard at 0.05 mL * min⁻¹). Additionally, a quality control solution composed of 100 mg * mL⁻¹ Reserpine (m/z 609.2807 [$M + H$]⁺), m/z Rifampicin (698,317 [$M + H$]⁺), Oligomycin-A (m/z 791.5304 [$M + H$]⁺) and Genistein (m/z 271.0601 [$M + H$]⁺) was included into the sample sequence to monitor mass accuracy and reproducibility of chromatography over time. A deviation of $\Delta\text{ppm} = 2$ to theoretical masses and $\Delta\text{sec} = 12 R_t$ was tolerated. Gaseous ion formation was achieved by electrospray ionization at 4.5kV (capillary) and spray shield offset of -0.5 kV in positive mode. Nebulizer gas (N₂) was supplied at constant pressure of 1.6 bar. Heated drying gas (N₂ at 250 °C) was supplied at 7.5 L * min⁻¹. Spectra of cationic analytes were recorded at 1 scan/sec. During tandem MS experiments (MS/MS), fragmentation of analytes was carried out by collision induced fragmentation (collision energy of 28.0-35.05 eV and collision gas (N₂) at 10⁻² mbar).

3.2.5 Data bucketing and visualization

Data Buckting and Annotation Scripted data processing (Figure S1) included recalculation of line spectra as well as molecular feature finding and was carried out using DataAnalysis 4.4 ©(Bruker Daltonik GmbH). Recalculation of line spectra (*sum peak* finder), thus separation of real signals and background noise was achieved by implementing a FWHM threshold of 3 points and an absolute ion intensity (I) cutoff of 10.000 relative intensity units. Subsequently, the molecular feature finder (S/N = 5; minimal time-correlation coefficient = 0.7; minimum compound length = 8 spectra) was used to correlate mass list entries belonging to the same molecule. Based on the molecular feature list, data bucketing was performed in ProfileAnalysis 2.3 (Bruker Daltonik GmbH). Buckets were generated from 100 - 1600 m/z and R_t 0.5 -18 min. Bucket size was set to Δ_{sec} of 12 and Δ_{ppm} of 5. The generated list of buckets was exported to MetaboScape 3.0 (Bruker Daltonik GmbH) and annotation with a in house reference data base. Quality of automatic annotation was guaranteed by allowing narrow deviations of m/z ($\Delta_{\text{ppm}} = 2$), retention time ($\Delta_{\text{sec}} = 12$) and a maximum mSigma score of 10.

Principle component analysis Primary data visualization was done by PCA in MetaboScape 3.0(Bruker Daltonik GmbH). The model was plotted without scaling algorithm. Grouping was done on the basis of strain identity and cultivation medium used.

Metabolomic heatmap Calculation of metabolomic heatmaps represents a complementary approach based on similarity rather than differences in the data structure (like PCA). Essentially cosine similarity of bucket vectors of all samples were compared one to one. Thereby, each extract pair was assigned a $\cos \theta$ value as a measure of similarity with respect to their bucket distribution. Extract pairs sharing an overlapping pattern of filled and empty buckets are considered related ($\cos \theta \geq 0.7$) and form groups in the calculated dendrogram and heatmap. Calculation and plotting was carried out in a custom R script (s. Figure S2). The latest version of the script can be found at github.com/christoph-hartwig-ime-br/cosine-V2 or doi.org/10.5281/zenodo.3932968.

3.2.6 Variable dereplication via molecular networking

The UHPLC-QTOF-MS/MS data of the *Streptomyces* extracts were additionally analyzed using molecular networking to allow the variable dereplication of known

and unknown metabolites. First, the raw data (*.d files) was converted to plain text files (*.mgf) containing MS/MS peak lists using MSConvert (ProteoWizard package [31]), wherein each parent ion is represented by a list of fragment mass/intensity value pairs. Following, the molecular networking algorithm converted each precursor ion into a vector in an n -dimensional space, with n being the number of fragment ions. The vectors were compared pairwise using dot product calculations based on the cosine between the two (= cosine similarity). Each vector pair was thus assigned a cosine similarity score of 0.0 - 1.0, where 0.0 represents an angle of 90° between the two vectors and 1.0 either 0° or 180°. Perpendicular parent ion vectors share no fragments and are entirely different, whereas a cosine score close to 1.0 indicates shared fragments, thus a putative structural relationship between the compared precursor ions. Pairs with a cosine similarity score greater than 0.7 were defined as related and were thus connected in the network. Additionally, ions need a minimum of six shared fragments (tolerance Δ ppm 0.05) with at least one partner ion to be included in the final network. *In silico* fragmented compounds [7] of a commercial database (AntiBase 2017 [91]) as well as our in-house pure compound MS/MS database were included in the network as reference substances to narrow down the molecular formula to highlight compounds of interest. CytoScape v3.4.0 was used to visualize the data as a network consisting of nodes and edges, wherein each node represents a parent ion and its color reflects the sample from which the MS/MS file was obtained. The thickness of the edges represents the cosine similarity score between nodes (thick edges indicate high similarity). Structures of successfully annotated compounds were automatically generated using the add-in application chemViz2 (v. 1.1.0) on the basis of the SMILES information deposited in the respective data base.

3.3 Results

3.3.1 Bioactivity

The 100fold concentrated organic extracts of *Streptomyces* strains ST101789, ST107645, ST106693 and ST107165 were screened for growth inhibitory effects. Essentially, only two extracts showed activity against Gram-negative test organisms (ST107165 in 5294 and ST101789 in 5315, s. Figure 4). Reduced growth was only observed when *E.coli* was screened in minimal M9 medium or in the other case in MHI supplemented with bicarbonate. On the other hand, all tested extracts, except ST101789 in 5254 and 5294, showed bioactivity towards at least one Gram-positive test strain. Especially *S.aureus* ATCC25923 was strongly inhibited by these extracts in almost all tested dilutions. *M.smegmatis* ATCC607 was mainly inhibited by extracts from ST107165 and ST107645 obtained from cultivation in 5254 and 5294. Extracts obtained from fermentations in medium 5315 were excluded from the analysis, due to high medium background activity against *M.smegmatis*. Finally, all tested crude extracts of ST107165 inhibited the growth *C.albicans*, while extracts from the other Actinobacteria showed medium specific activity. Only extracts of strains ST107193 and ST107645 generated from fermentations in 5254 reduced the growth of *C.albicans*, whereas ST101789 did not show inhibition in any condition.

	Medium	<i>Ecol</i>			<i>Paer</i>	<i>Saur</i>	<i>Msme</i>	<i>Calb</i>
		MHII	MHC	M9	MHII	MHII	MHII	MHII
ST101789	5254							
	5294							
	5315			0.5		0.125		
ST106693	5254							
	5294					0.5		
	5315					0.125		
ST107165	5254					0.125	0.125	0.125
	5294	0.125				0.125	0.125	0.125
	5315					0.125		0.125
ST107645	5254					0.125	0.125	0.125
	5294					0.125	0.25	
	5315					0.125		

Figure 4: Screening results of extracts from ST101789, ST106693, ST107165, ST107645. Extracts were tested in micro broth dilution assays against *E.coli* ATCC35218 (*Ecol*) in Mueller Hinton II broth (MHII), MHII supplemented with bicarbonate (MHC) and minimal medium (M9), *P.aeruginosa* ATCC27853, *S.aureus* ATCC25923, *M.smegmatis* ATCC607 and *C.albicans* FH2173. Activity is given in the lowest volume of extract causing at least 85 % rel. growth inhibition of the test strain in 50 μ L assay volume. Assay results with diagonal line were invalidated due to activity of medium controls.

3.3.2 Chemical diversity assessment and automatic annotation

Besides the assessment of antimicrobial potency, all extracts were subject to UPLC-HRMS measurements. As a starting point of data exploration, the chemical diversity within the set of extracts was investigated. To do so, the compound distribution within the *Streptomyces* extracts was compared on the basis of a PCA of the bucket matrix (Figure 5). The two most significant factors (PC1 and PC2) account for 49.5 % of the total variance in the data set. The three dimensional model describes 60.3 %.

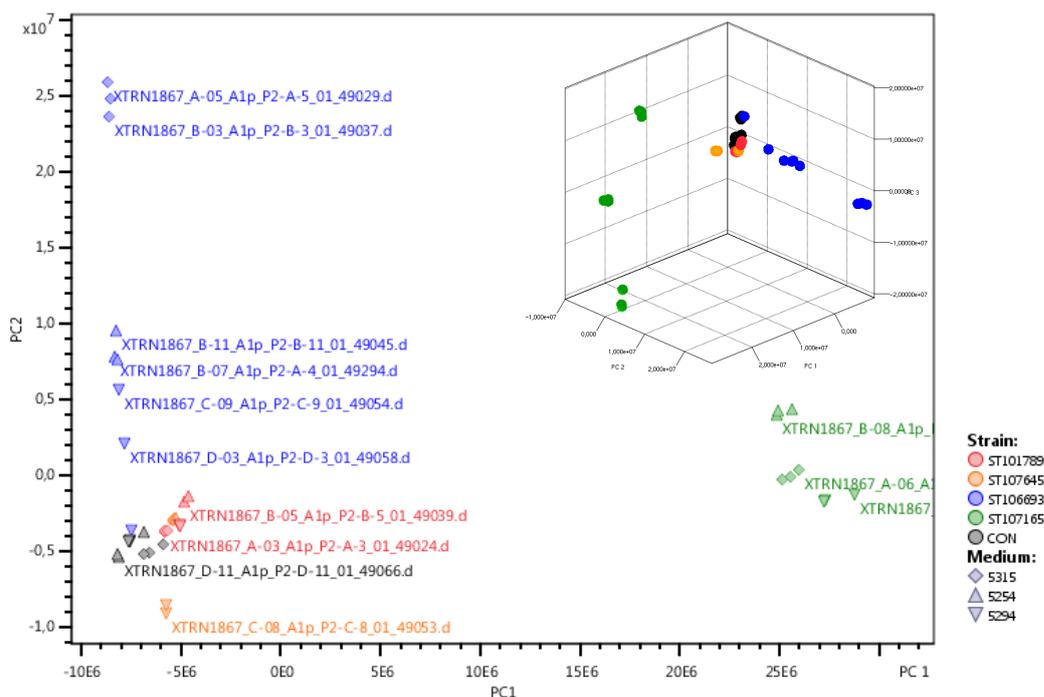


Figure 5: PCA based on the bucket matrix of strains ST101789 (red), ST107645 (orange), ST106693 (blue), ST107165 (green) and the media controls (black). Culture media are represented in different shapes (5315 diamonds, 5254 upwards triangle, 5294 downwards triangle) in the 2D scores plot. Top right: 3D PCA of the same data.

Biological replicates representing the different strain and media combinations cluster closely together in the scores plot. Strain ST106693 cultured in 5294 (blue reversed triangle in Figure 5) represents the sole exception to this observation: One sample of this triplicate formed a group with the medium controls of 5294, apart from the other two replicates. Score values of ST101789 (red) and ST107645 (orange) lay close to each other and the media controls (black), while ST107165 (green) and ST106693 (blue) form distant clusters. Extracts obtained from ST106693 fermented in 5315 medium are clearly separated from fermentation of the same strain in the other two media. Bucket composition of ST107165 medium triplicates are different from each other and all other investigated extracts and thereby form distinct groups in the two and three dimensional plot.

Commonly observed microbial compounds (*frequent hitters*) were automatically annotated using the in house analyte list containing over 1600 entries. In addition to the quality criteria (s. section 3.2.5), an annotation was only considered valid if found in all samples of an triplicate (ST106693 in 5294 in duplicate). In total, seven microbial metabolites were annotated in that way.

Table 1: Bucket annotation ST107645.

Medium	m/z [M + H] ⁺	Rt [min]	Formula	Name	Annotation quality		
					Δ ppm	Δ Rt	$m\sigma$
5294	461.260	3.74	C ₂₀ H ₃₆ N ₄ O ₈	Desferri-ferrioxamine H	0.51	0.06	1.7

Annotating the bucket table of ST107645 produced one single hit in the analyte library: Desferri-ferrioxamine H [3] (Table 1). The molecule was only detected in the extracts generated from fermentations in 5294 medium. For ST101789, the bucket m/z 693.182@11.31min was automatically annotated as β -naphthocyclinone epoxide [86] (Table 2). In this case, the compound could only be detected cultivation carried out in 5315 medium.

Table 2: Bucket annotation ST101789.

Medium	m/z [M + H] ⁺	Rt [min]	Formula	Name	Annotation quality		
					Δ ppm	Δ Rt	$m\sigma$
5315	693.182	11.31	C ₃₅ H ₃₂ O ₁₅	β -naphthocyclinone epoxide	0.29	0.08	3.5

Within the extracts of *Streptomyces* sp. ST100693, three related compounds, Anguinomycin A and B [21] as well as Leptomycin B [21][154] were detected (Table 3). Interestingly, Anguinomycin A was found in media 5294 and 5315, whereas the B derivative and Leptomycin B were only observed in 5315 (s. top left and right Figure S3).

Table 3: Bucket annotation ST106693.

Medium	m/z [M + H] ⁺	Rt [min]	Formula	Name	Annotation quality		
					Δ ppm	Δ Rt	$m\sigma$
5294/5315	513.321	12.24	C ₃₁ H ₄₄ O ₆	Anguinomycin-A	0.25	0.06	3.9
5315	527.336	13.92	C ₃₂ H ₄₆ O ₆	Anguinomycin-B	0.07	0.06	3.5
5315	663.334	14.35	C ₃₃ H ₄₈ O ₆	Leptomycin B	0.21	0.13	0.5

Extracts of ST1070165 produced two hits during automatic database inquiry. Most strikingly, Scopafungin (aka Niphimycin) [78] was well present in all extracts. Although the compound was biosynthesized in all media, production titer varied across media: Production was observed to be 3 times higher in 5254 compared to 5315 (s. bottom left Figure S3).

Table 4: Bucket annotation ST107165.

Medium	m/z [M + H] ⁺	Rt [min]	Formula	Name	Annotation quality		
					Δ ppm	Δ Rt	$m\sigma$
All	1142.730	9.75	C ₅₉ H ₁₀₃ N ₃ O ₁₈	Scopafungin	0.30	0.05	6.2
5294	469.149	6.69	C ₂₅ H ₂₄ O ₉	Echoside A	0.31	0.07	2.0

Furthermore, the bucket m/z 469.149@6.69 min was annotated as Echoside A [40], a glycosidated terphenyle chromophore [103]. The compound was only observed in the UPLC-HRMS records of ST107165 triplicates fermented in 5254 medium.

Chemical fingerprinting - metabolomic heatmaps To evaluate the similarity of the metabolite composition within the *Streptomyces* extracts from a different perspective, the same bucket matrix (s. subsection 3.3.2) was visualized by a metabolic heatmap. The dendrogram as well as the heatmap itself was constructed based on one-to-one comparisons ($\cos \theta$) of bucket distribution across extracts. In total, 16 metabolic families were identified. A metabolic family was defined as a group of extracts sharing a high cosine similarity score ($\cos \theta \geq 0.7$, dark blue) among each other and a low one with any other extract ($\cos \theta \leq 0.35$, white). Values were derived from the group of quality controls (QC). QC samples formed a homogeneous family at the bottom of the heatmap ($\cos \theta = 0.89 - 0.73$), clearly distinct from all other analyzed samples. Triplicates of strain and media combinations exhibit a high degree of similarity (mostly $\cos \theta \geq 0.85$), thus form distinct branches in the dendrogram and lay in close proximity in the heatmap. Thereby, these samples are structured in 12 distinct metabolomic families. The media controls do not cluster together (as observed by PCA), but form remote groups of triplicates apart from each other.

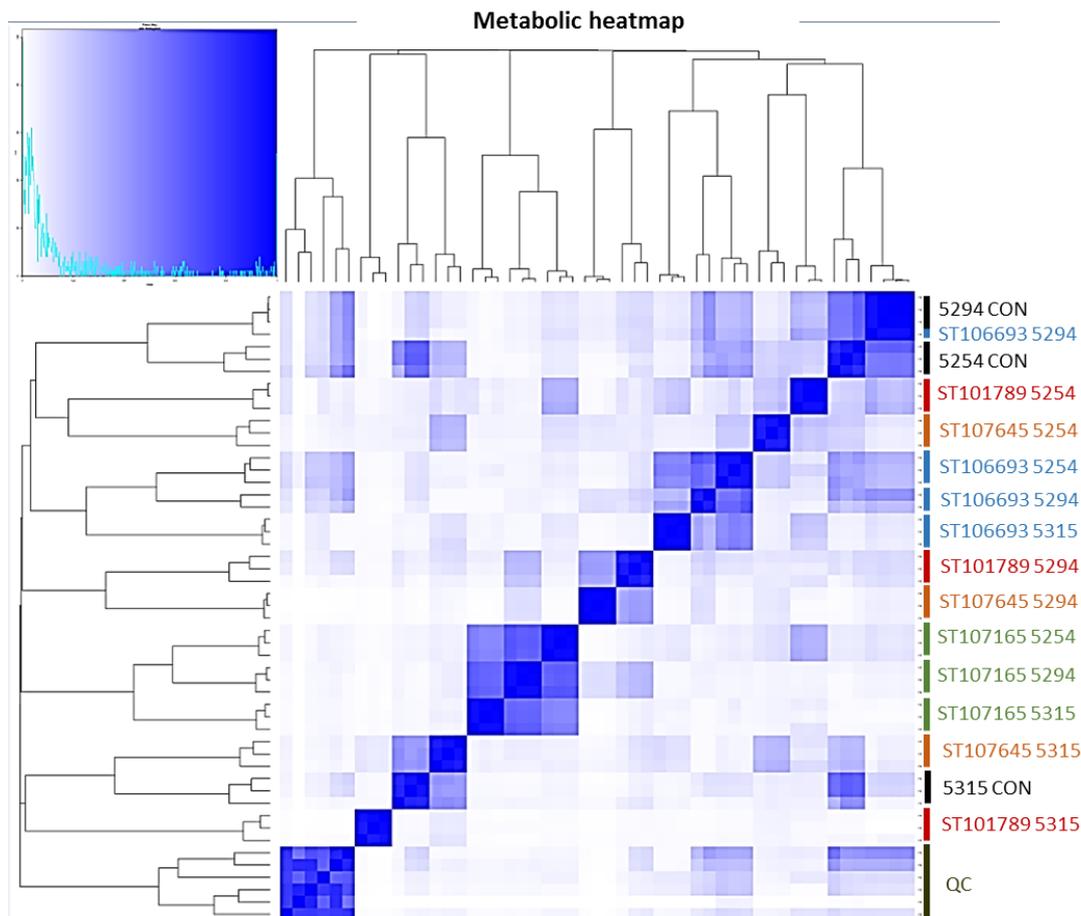


Figure 6: Metabolic heatmap based on cosine similarity of bucket matrix of strains ST101789 (red), ST107645 (orange), ST106693 (blue), ST107165 (green) and the media controls as well as quality controls (black).

As the PCA indicated (s. Figure 5), the heatmap shows that one replicate of ST106693 cultured in 5294 medium is highly similar ($\cos \theta = 0.93$) to the respective medium control, forming a four membered family, while the remaining duplicate clusters next to (but not together with, $\cos \theta = 0.54$) the metabolomic families of ST106693 cultivations in 5254 and 5315.

In terms of buckets distribution, the PCA showed only little differences between all ST101789 and ST107645 fermentations and the media controls (subsubsection 3.3.2). Remarkably, the similarity analysis implies a low amount of shared buckets between these samples: Although, fermentations of ST101789 in 5254 are overall the most similar extracts compared to the 5254 medium control, the two groups exhibit a very small cosine similarity score ($\cos \theta = 0.15$). The same holds true for ST107645 cultivated in 5315 medium: Despite being the most similar sample, the actual similarity value remains rather low ($\cos \theta = 0.42$). ST101789 and ST107645 cultured in 5294 formed families apart from the medium control. Even

though the $\cos \theta$ calculation demonstrates a low one-to-one similarity, ST101789 and ST107645 clustering seem to be influenced by the cultivation medium, as all nine extracts obtained from one strain are less similar to each other than the strain and the respective medium controls.

On the other hand, ST107165 and ST106693 exhibited a different behavior: The most prominent cluster in the heatmap is comprised of the nine ST107165 fermentations. By definition, the triplicates cultured in the different media form metabolic families by themselves, however the similarity between these families is, compared to the rest of the data, rather high ($\cos \theta = 0.60 - 0.98$). Comparably, ST106693 also forms a cluster consisting of three metabolic groups (corresponding to the media used).

3.3.3 Molecular networking and variable dereplication

Based on UPLC-QTOF-HRMS/MS data, structural relationships between compounds within the set of extracts and reference compounds were investigated. Each precursor ion was automatically compared, one-to-one, with all other precursors in the dataset and reference libraries. In total 3930 precursors and library items fulfilled the selected parameter (s. subsection 3.2.6) and were plotted in one single MS-network (Figure 8). Nodes are labels with its m/z and the edges with the respective m/z difference. In the following, five clusters are described in detail (additional cluster in supplements).

Cluster A - Echoside A In agreement with the bucketing approach, the precursor ion m/z 469.149 was automatically annotated (Echoside A) in the extracts of ST1017165 (s. Figure 7). Interestingly, Echoside A was not only detected in 5294 extracts (purple) but also, to a lower extent, in the other two cultivation regimes (5254 (yellow) and 5315 (green)). The precursor of Echoside A was located in a cluster of minimal size (two interacting nodes). The two binding partner share a m/z difference of ($1.8 * 10^{-4} m/z$). The automatic annotation was validated by manual comparison of the MS/MS spectrum of m/z 469.149 in the crude extract and the respective spectrum of pure Echoside A (s. Figure S5).

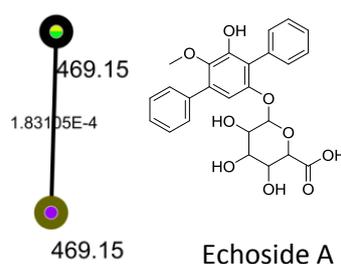


Figure 7: ClusterA Echoside A

Cluster B - Resomycins Using the *in silico* fragmented Antibase library, the precursors m/z 365.102 and m/z 383.113, detected in the extracts of ST107645, were annotated as Resomycin A and B [111] (s. Figure 9). The two compounds exhibit a different ring substitution pattern: while Resomycin A is hydroxylated at at the C-9 position, this moiety is absent in Resomycin B. The characteristic m/z difference of 18.0105 is indicated on the edge between the two derivatives. Resomycin B was only present in 5315 extracts (pink), while Resomycin A was additionally detected in 5254 extracts (green). As observed for Echoside A, the ion corresponding to the single protonated Resomycin A was included twice in the network.

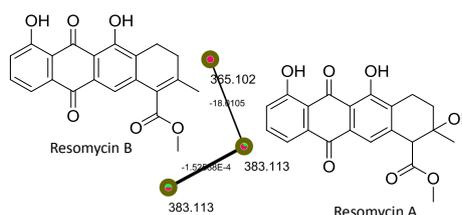


Figure 9: ClusterB Resomycins

Cluster C - Anguinomycines Cluster C is composed of four structurally related precursor ions with the m/z values of 513.32, 495.32, 509.325 and 497.327. All ions, expect 497.327, were detected in extracts of strain ST106693 cultured in 5294 (green) and 5315 (blue)

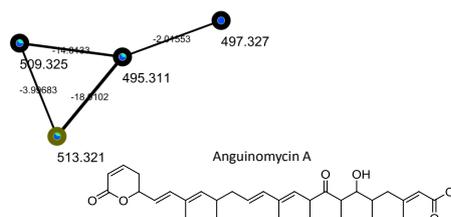


Figure 10: ClusterC Anguinomycines

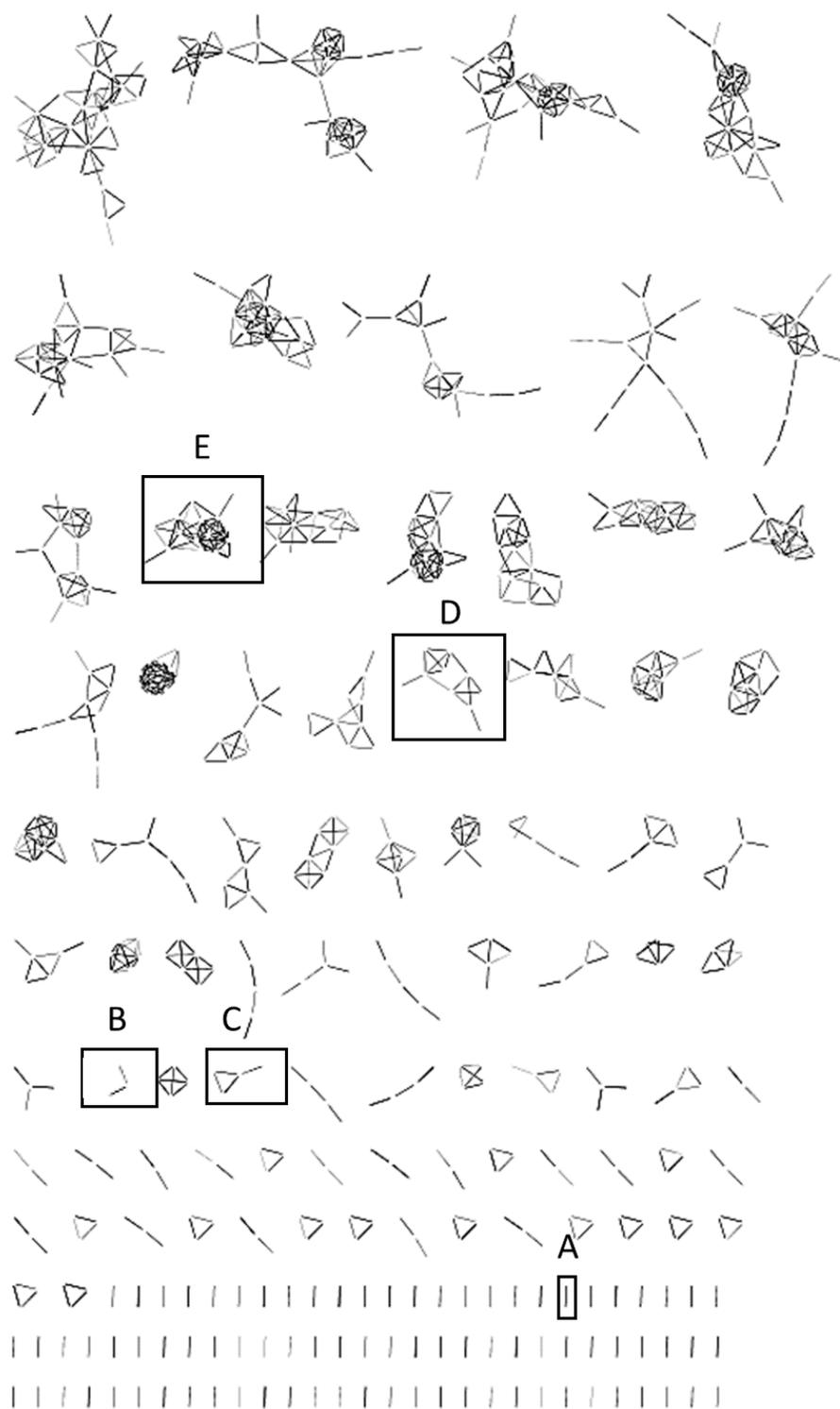


Figure 8: Molecular network constructed from *Streptomyces* sp. extracts. Each node represents an precursor ion. Edges link nodes corresponding to ions with similar fragmentation pattern to form clusters of molecule families. Representative cluster are analysed in detail: **A**: Echocide A cluster; **B**: Resomycin cluster; **C**: Anguinomycines cluster; **D**: Naphtocyclinones cluster; **E**: Scopafungines cluster

(s. Figure 10). While 513.32 was annotated as Anguinomycin A, the other signals remained unexplained by the automatic comparison with the spectral libraries. The automatic annotation was verified by manual comparison of MS/MS spectra within the crude extract and a measurement of pure Anguinomycin A (s. Figure S6). However, the characteristic m/z differences between Anguinomycin and its binding partners indicate the presence of an in source dehydrated variant of Anguinomycin A (m/z 495.311 $[M + H - H_2O]^+$). Ion 509.325 likely corresponds to Anguinomycin B: Again, protonation and in source dehydrogenation might explain the mass shift observed. The last signal (497.327) might correspond to an dehydroxlated Anguinomycin A $C_{31}H_{43}O_5$, which could not be found in the consulted data bases.

Cluster D - Naphthocyclinones Cluster D illustrates the structural relationship of ten precursor ions, five of which were annotated as members of the naphthocyclinone family (s. Figure 11) - among them, β -naphthocyclinone epoxide, which was already predicted by annotation via data bucketing. The automatic annotation was validated by manual comparison of fragmentation signatures of β -naphthocyclinone within the crude extract and an authentic standard (s. Figure S7). In addition to the epoxide, β -naphthocyclinone and the chlorohydrin variant, as well as γ - and α - naphthocyclinone were found. In accordance to the bucketing based observations, the group of molecules was only observed in extracts of strain ST101789, if fermented in 5315 medium (pink). Remarkably, half of the precursors within the cluster were not identified by the automatic data base queries. Bioactivity guided fractionation identified α - Naphthocyclinone, the putative demethylated variant of α - Naphthocyclinone and β -naphthocyclinone epoxide as growth inhibition causing agents against *S.aureus* in extract ST101789(5315) (s. Figure 13). The moderate growth inhibition against *E.coli* screened in M9 medium could not be validated by confirmatory screens and μ -fractionation.

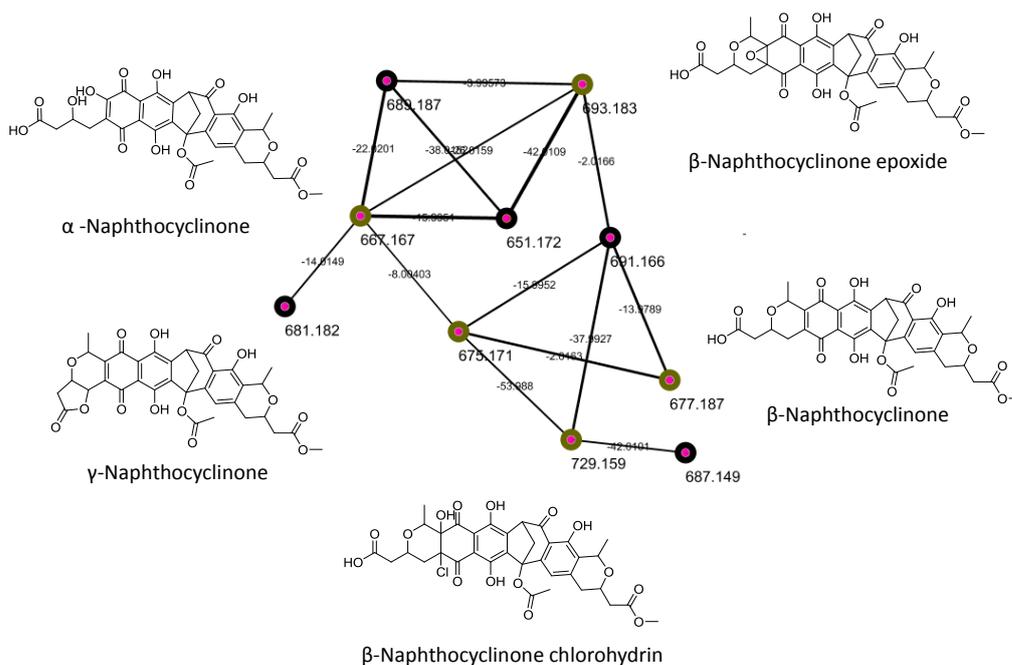


Figure 11: Cluster D from the molecular network constructed from *Streptomyces* sp. extracts.

Cluster E - Scopafungines The largest analyzed cluster within the network originated from the UPLC-QTOF HRMS/MS records of from ST107165 extracts. A total of 13 precursor ions were observed to possess a similar fragmentation signature, based on $\cos\theta$ calculation between their molecular vectors in space (Figure 12). Remarkably, three of them could be identified as N²-methylniphimycin (m/z 1156.75) [84] [13], Amycin A (m/z 1228.73)[57] and Scopafungin (a.k.a. Niphimycin, m/z 1142.73) [78]. The latter was found in both, the *in silico* and the measured in house MS/MS data base. The automatic annotation of Scopafungin within the crude extract was verified by comparison of the respective MS/MS spectra to an authentic standard (s. Figure S8). Besides compounds, eight other structurally related ions were present in the investigated extracts and could not be found in the data bases. However, literature research focused on this group of molecules revealed the identity of (m/z 1142.75 [$M + H$]⁺) as Niphimycin C [68]. These ions were not annotated automatically as Niphimycin C-E were published in january 2018, hence were not included in the data base used for *in silico* fragmentation (Antibase 2017). Further not annotated ions within the cluster might be explained by the m/z differences between explained and unexplained compounds. These indicate, for instance, the presence of an unknown dehydroxylated (m/z 1124.72) and a demethyl-dehydroxy (m/z 1110.71) Scopafungin derivative.

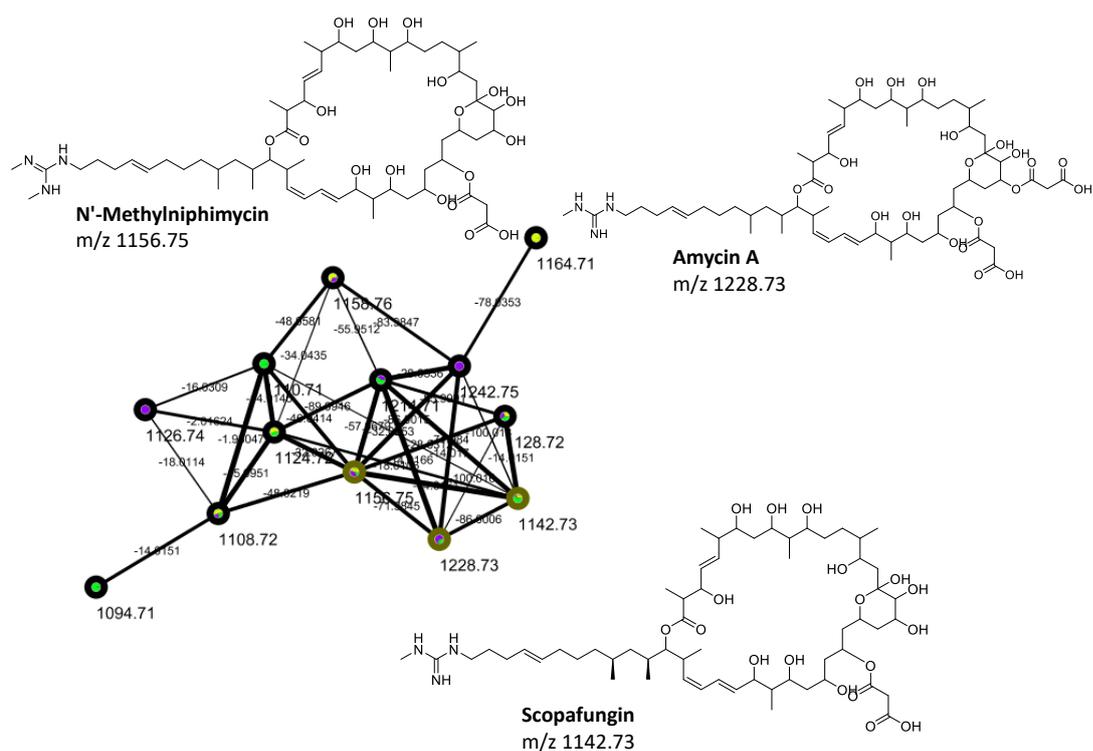
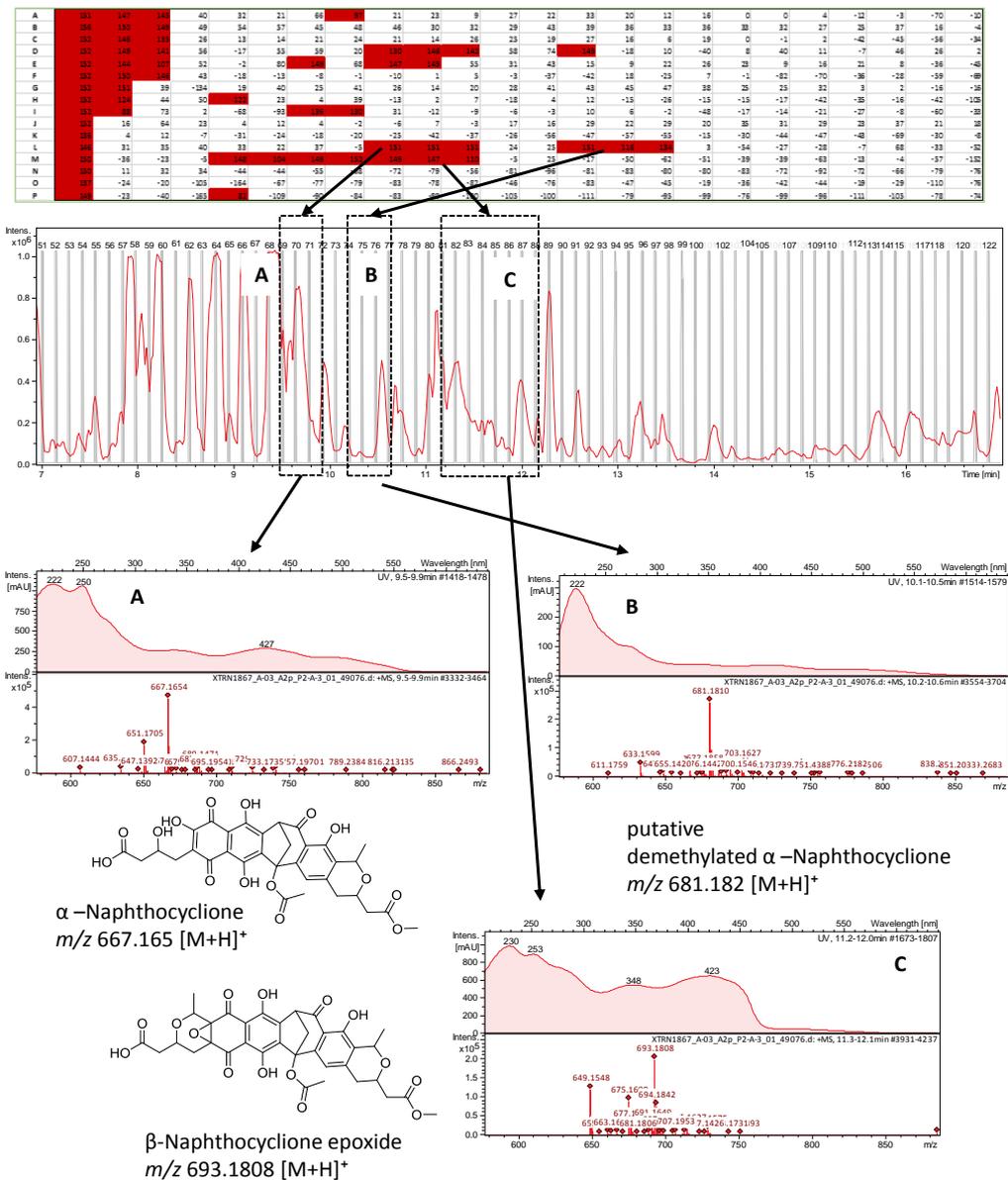


Figure 12: Cluster E from the molecular network constructed from *Streptomyces* sp. extracts.

3.3.4 Linking bioactivity to causative agent

Extracts of strain ST101789 cultured in 5315 inhibited the growth *S.aureus* ATCC25923. To identify the causative agent within the compound mixture at hand, the extract was fractionated into 159 fractions (μ -fractionation, s. section 3.2.3) and re-screened (Figure 13). Fractions 69-71, 74-76 and fractions 81-87 inhibited the test strain. Growth inhibition causing components of the extract could be dereplicated as α -Naphthocyclinone, the putative demethylated variant of α -Naphthocyclinone and β -naphthocyclinone epoxide.



Similar to extracts obtained from ST101789 fermented in 5315 medium, extracts of strain ST106693 cultured in the same medium showed bioactivity against *S.aureus* ATCC25923. Fractionation and subsequent re-screening yielded four groups of growth inhibiting fractions (Figure 14). Dereplication was carried out by comparing the major ions within these bioactive fractions to the annotated precursors in the respective molecular network (Figure S4). Fractions 87 and 91-94 contained mainly ions corresponding a group of polyketide macrolides, the Conglobatins [165]. Besides Conglobatin (m/z 499.2803), a de-methylated (m/z 485.2649) and a de-dimethyl variant (m/z 471.2490) were detected. All three compounds were found in single and double charged state. Fractions 97-98 were mainly composed of Anguinomycin A (protonated and in source dehydrated). Remarkably, the first group of growth inhibitory fractions (35-36) contained one major ion m/z 330.2382 which could not be found in any database.

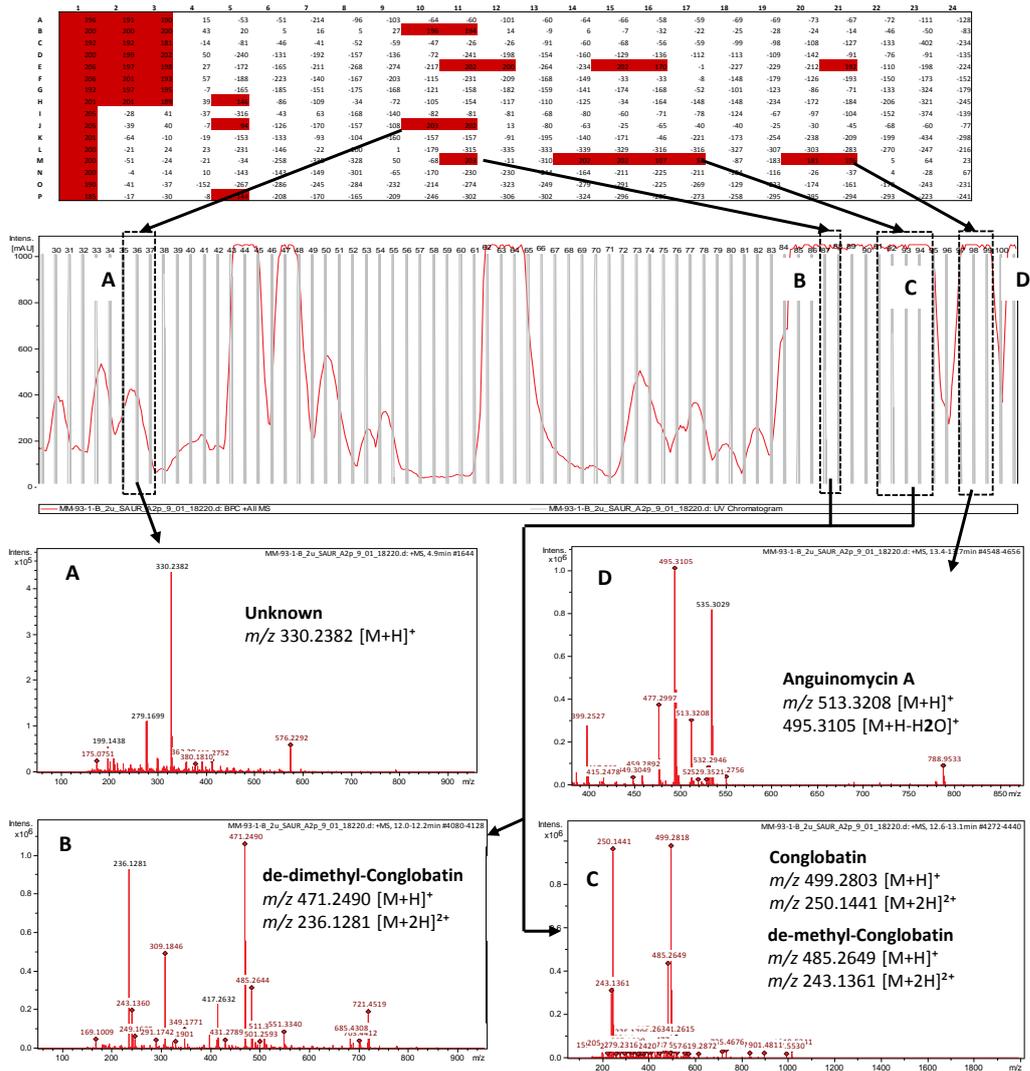


Figure 14: μ -fractionation of extract ST106693 cultured in 5315 against *S. aureus* ATCC25923. Top: Bioassay readout of μ -fractionated extracts against *S. aureus*. The extract was fractionated twice on the same plate. 2 μ L of extract were collected in wells A-H5 to A-H24. Collection of 5 μ L injected extract was done in well I-P5 to I-P24. Numbers indicate the relative growth inhibition of the each fraction relative to the negative control. Fractions 35-36, 87, 91-94 fractions 97-98 inhibited the test strain (indicated in red). Middle: Chromatogram of 2 μ L injection. Peaks corresponding to growth inhibitory effects are highlighted (A-D). Bottom: Mass spectra of fractions A-D and major ions detected. Growth inhibition causing components of the extract could be dereplicated as Conglobatin (C), demethyl-Conglobatin (B, C) a putative de-dimethyl variant of Conglobatin (B) and Anguinomycin A (D). Besides, one ion, m/z 333.2382, could not be found in the consulted databases.

Strain ST107645 showed strongest bioactivity when cultured in 5254 medium. Yet, compounds dereplicated in these extracts do only partly explain this pattern: desferrio-ferrioxamine H was only detected in 5294 and the Resomycins in 5254 and 5315 medium. If Desferrio-ferrioxamine H would be the causative agent, only the 5294 extracts should be active. If Resomycin would be responsible for the growth inhibition of *S.aureus*, *M.smegmatis* and *C.albicans* as observed in 5254 extracts, the bioactivity pattern should be identical in extracts generated from 5315 cultivations (resomycin production level in both cultivation regimes similar). To identify the causative agent within the compound mixture at hand, the extract was fractionated in 159 fractions (μ -fractionation, s. section 3.2.3). Rescreening of the fractionated extract against *S.aureus* produced three sets of consecutive growth inhibitory wells/fractions (Figure 15) at both tested injection volumes. Zone A contained a set of minor and one major ion (m/z 901.4764 @ 7.6 min). These ions were also detected in the medium controls, hence bioactivity observed in these fractions is most likely not caused by a bacterial metabolite. Fraction 75 was composed of one sharp peak containing one ion (m/z 383.1129 @ 10.5 min) with UV maxima at 259 and 430 nm. The ion m/z 383.1129 $[M + H]^+$ compares to the annotated Resomycin A in terms of exact mass, UV absorption and described bioactivity [111]. Resomycin B was also detected in the fractions at low abundance. Finally, fractions 63-66 (zone B) essentially contained one single ion m/z 567.1765 eluting in two distinct peaks at 8.9 - 9.3 min. The predicted molecular formula ($C_{32}H_{26}N_2O_8$) could not be found in any data base.

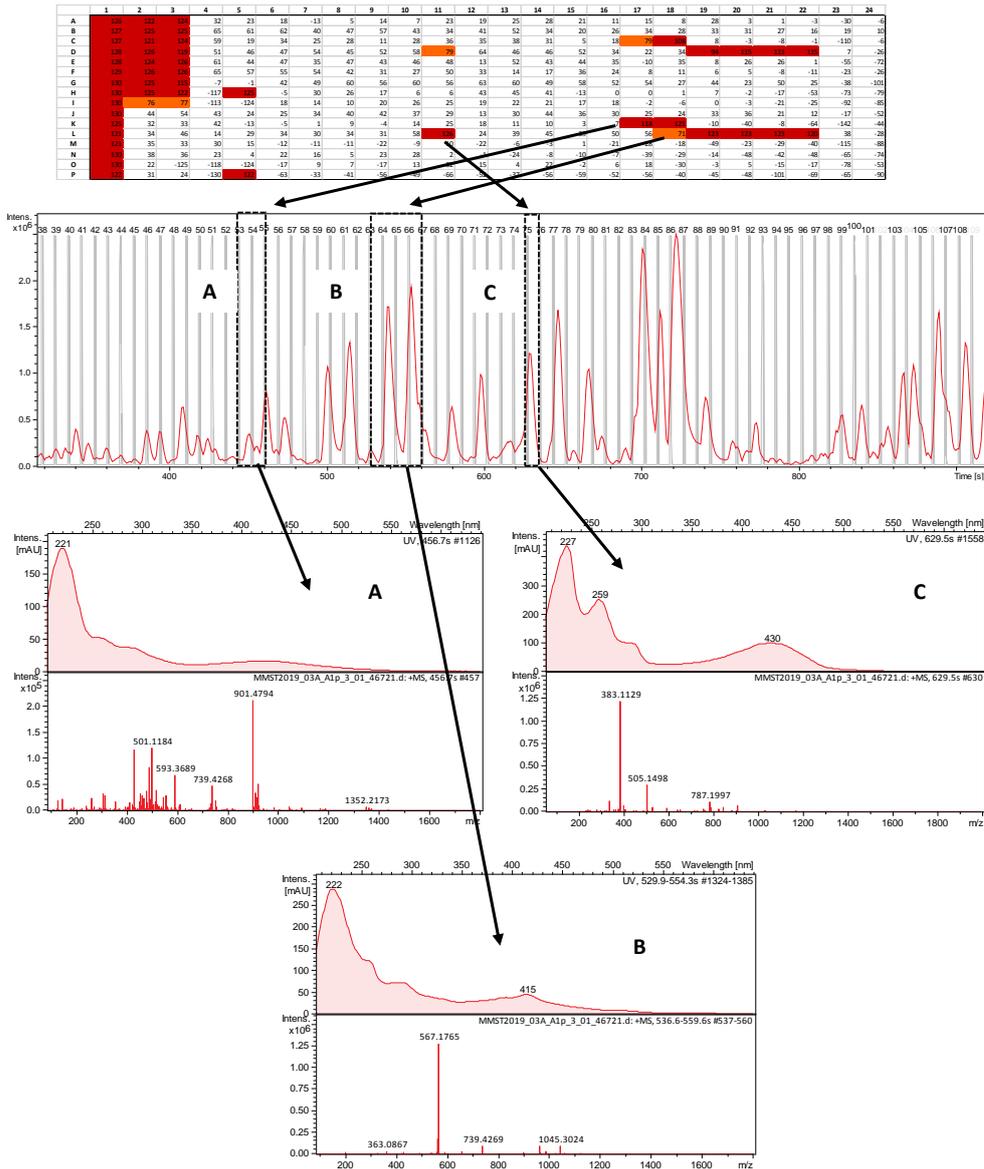


Figure 15: μ -fractionation of extract ST107645 cultured in 5254 against *S. aureus* ATCC25923. Top: Bioassay readout of μ - fractionated extracts against *S. aureus*. The extract was fractionated twice on the same plate. 2 μ L of extract were collected in wells A-H5 to A-H24. Collection of 5 μ L injected extract was done in well I-P5 to I-P24. Numbers indicate the growth inhibition of the each fraction relative to the negative control. Fractions 53-54, 63-66 and fraction 75 inhibited the test strain (indicated in red). Middle: Chromatogram of 2 μ L injection. Peaks corresponding to growth inhibitory effects are highlighted (A-C). Bottom: UV and mass spectra of fractions A-C and major ions within

Extract ST107165 cultured in 5244 showed growth inhibitory effects against Gram-positive bacteria, yeast and Gram-negative bacterium *E. coli* (the latter only if the screening medium was supplemented with bicarbonate = MHC medium).

Fractionation was carried out for retesting against *S.aureus*, *C.albicans* and *E.coli* in MHC medium. Assay results showed identical growth inhibition: Fraction 64-84 inhibited the growth of all test strains in both injection volumina. For *S.aureus*, the 5 μ L injection additionally inhibited the growth in fractions 85-88 (s. Figure 16). Within the bioactive fractions three major protonated molecular ions were present: m/z 1142.7300 $[M + H]^+$ eluting at 9.6 min (A1, blue), m/z 1156.7452 $[M + H]^+$ at 10.7 min (A2, green) and m/z 1228.7303 $[M + H]^+$ at 11.0 min (A3, black). The double protonated molecular ions of these three compounds were detected at the same intensity as the single protonated molecules. The identity of these compounds was previously determined as Scopafungin, Amycin A and N'-methylniphimycin (s. Figure 12).

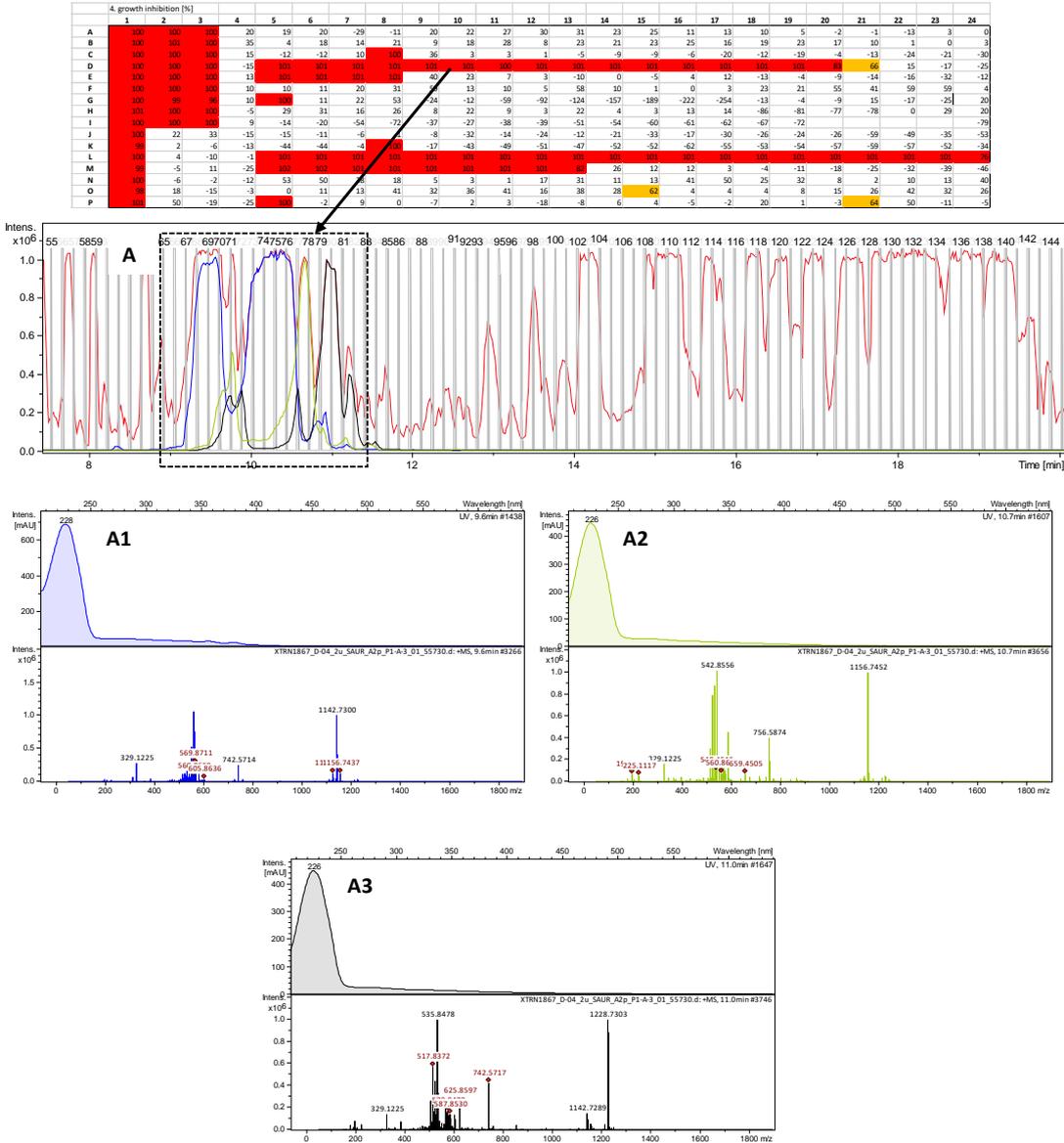


Figure 16: μ -fractionation of extract ST107165 cultured in 5294 against *S. aureus* ATCC25923. Top: Bioassay readout of μ - fractionated extracts against *S. aureus*. The extract was fractionated twice on the same plate. 2 μ L of extract were collected in wells A-H5 to A-H24. Collection of 5 μ L injected extract was done in well I-P5 to I-P24. Numbers indicate the growth inhibition of the each fraction relative to the negative control. Fractions 64-88 inhibited the test strain (indicated in red). Middle: Chromatogram of 2 μ L injection. Peaks corresponding to growth inhibitory effects are highlighted (A1-A3). Bottom: UV and mass spectra of fractions A1-A3 and major ions within

3.4 Discussion

3.4.1 Metabolomics

In an industrial setting, hundreds of strain-media combinations are routinely used to increase the probability of success. Small to medium scale fermentation in multiwell plates might increase the throughput but complicate quality control by visual inspection. Therefore a robust retrospective data analysis is crucial to properly evaluate the conducted experiments. Data bucketing was used to convert the three-dimensional UHPLC-QTOF-HRMS data into two-dimensional matrices. Reduced dimensionality allowed the evaluation of cultured chemical diversity within the *Streptomyces* extracts. Two different models were selected for data exploration. A principle component analysis was used to determine differences in the chemical composition of the extracts, whereas the metabolic heatmap (MSH) approach helped to visualized similarity between samples.

Outlier effects First of all it appeared that the nature of the experiment, ergo the data structure, dictates whether PCA or MSH is more appropriate for chemical diversity assessment. An elementary step during PCA is the calculation of latent variables (score values) for each sample to allow two- or three-dimensional plotting (Equation 1) [151]. Differently composed samples would cluster apart from each other, while similar would form groups in the scatter plot [134] [133]. The structure of the presented data set is rather heterogeneous, which is certainly not surprising since different strains are in fact expected to produce distinct metabolite signatures. Clearly, a group of drastically distinct samples would reduce the resolution of the scores plot. In the presented example, extracts of strain ST107165 and ST106693 seem to be profoundly different compared to all other extracts. When plotting a data set containing highly diverse score values in a linear PC system, samples which are less different, like ST101789 and ST107645 extracts, would cluster together in the scatter plot even though being evidently different. Conceptually, extreme samples tend to "pull" the PC model towards themselves underlining the sensitivity of a PCA towards outliers. In other words, in a heterogeneous data set a PCA illustrates the most profound (outlier) effects, but fails to point out minor to intermediate differences between extracts. Generally, data pre-treatment such as outlier screening (e.g. OutlierD [34]) or scaling (e.g. Pareto or log transformation [12]), could help to reduce outlier effects, but proofed useless in the data set at hand.

Strikingly, the comparison of the same bucket lists in a vector space model helped to distinguish ST101789 and ST107645 extracts from each other and the medium.

Greater resolution and thereby increased explanatory power of MSH compared to PCA seems to significantly facilitate information retrieval from heterogeneous data structures, such as metabolomics data. Essentially, the cosine similarity compares samples one-to-one and the obtained value is independent from the overall data structure including outlier effects.

Retrospective quality control Besides chemical diversity analysis, the PCA and MSH were primarily applied for quality (growth and contamination) control purposes. Essentially, both models allowed similar observations: Triplicated strain and media combinations formed distinctive cluster in the scores plot and the heatmap, indicating similar behavior of biological replicates and proper sample handling. One replicate of ST106693 cultured in 5294 medium clustered with the respective media control, apart from the remaining suggesting no or only weak microbial growth in this particular flask. In contrast, all other inoculated flasks exhibited bacterial growth. Assuming active bacterial metabolism is reflected by high difference (and low similarity) to media controls, two different metabolic behaviors could be observed:

Strain ST106693 and ST107165 cultivates lay far apart from each other and the medium in the 3D-PCA model, implying an extreme conversion of medium components to strain specific metabolites. Accordingly, these two strains formed each one cluster of high similarity in the heatmap, independently of the cultivation media. On the other hand, metabolic signatures of strains ST101789 and ST107645 are less pronounced: These extracts clustered closely to the medium controls within the scatter plot and metabolic groups are located in proximity to the respective controls. A low degree of altered medium composition might indicate weak growth either due to unfavorable conditions or a generally slow growth rate of the particular strain. In that sense, strains might not have utilized the provided medium to an extent posing nutrient shortage. In Actinobacteria, secondary metabolism is usually tightly connected to environmental stimuli, like nutrient depletion, and might not have been maximized by the applied media composition and incubation duration. Therefore, remaining medium components were well present in inoculated samples and only a few specialized bacterial metabolites were detected, leaving the samples rather similar.

Automatic Annotation Automatic annotation using either the bucket data base or molecular networking yielded similar observations when analyzing the provided microbial compound mixtures. Echoside A, Scopafungin, Anguiniomycin A and B as well as β -Naphthocyclinone epoxide were dereplicated by both, MS based spectral comparison and fragmentation signature alignment. Successful

annotation requires careful parameter selection. Usually this is an iterative process in which a balance between annotation accuracy and experimental mass precision is determined. Threshold definition leads to the philosophical decision whether an experiment should be susceptible to either false positive or false negative results. In this study, identification of uniformity among signals, regardless whether distribution across samples or data base comparison, was done conservatively. It was accepted to rather recognize, in reality, identical signals as different than suffer from incorrect annotation results. Thereby, annotation results of both methods seem to be quite robust, but in turn bucket lists as well as networks are inflated with duplicate signals. Besides, both methodologies exhibited further benefits as well as weaknesses:

A considerable draw back of automatic annotation using the bucketing approach would be the limitations concerning the reference databases. Each item in a bucket data base needs to be acquired in house using the exact same instrumental conditions as the samples to be analyzed. Obviously, any automatic annotation approach is only as powerful as the underlying data base and in this regard, commercially availability and high expense of many natural products substantially reduce feasibility. The molecular networking approach on the other hand, allows to incorporate *in silico* fragmented compounds from public data bases. Cluster B contained ions 365.102 $[M + H]^+$ and 383.113 $[M + H]^+$, which could both be annotated by spectral alignment, without having the reference compounds physically available. Both, Resomycin A and B would not have been dereplicated by bucket annotation. Besides, bucket annotation heavily relies on R_t comparison of analyte and reference compounds making chromatography inflexible. For adequate library correlation, each experimental extract needs to be prepared using the same solvent, gradient and column.

One major short coming of the MS/MS network is the exclusion of single compounds sharing no structural relationship with any other compound in the investigated set of extracts or databases . The precursor of Echoside A was located in a cluster of minimal size (two interacting nodes, Figure 7). The two binding partner share a m/z difference below the instrument sensitivity suggesting the same identity ($1.8 * 10^{-4} m/z$). If the algorithm would not accidentally have distinguished between these identical signals (recorded in different samples and due to conservative parameter selection), the node Echoside A would not have been included in the network as no other binding partner was present in the samples. The same might have happened to Desferri-ferrioaxamine H, which was identified by bucket annotation (ST107645 in 5294) but was absent in the network. The Naphthocyclinone cluster nicely illustrates the advantage of the networking

approach. Unknown signals, which have never been described can be explained by fragmentation signature correlation. One bucket, generated in the extracts of strain ST101789 cultured in 5315, was automatically annotated as β -Naphthocyclinone epoxide [86] [168] using the in house bucket database. Besides the β epoxide variant, molecular networking analysis allowed the dereplication of four additional naphthocyclinones (α , β , β chlorohydrin and γ) by comparison to *in silico* fragments. However most interestingly, additional precursor ions were observed in the naphthocyclinone cluster. These were not annotated and might represent putative new structures (Figure 11). Ion m/z 691.166 ($C_{35}H_{31}O_{15}$) might be explained by double bond formation within the structure of β -Naphthocyclinone epoxide. The mass difference between precursor ion m/z 687.149 and its closest relative (β -Naphthocyclinone chlorohydrin) in the network indicates the loss of an Acetyl group (42Da) resulting in a molecular formula of $C_{33}H_{31}ClO_{14}$. β -Naphthocyclinone epoxide deacetylation at the same position would explain the presence of precursor 651.172 ($C_{33}H_{30}O_{14}$) in the center of the network. Furthermore, a putative demethylated α -Naphthocyclinone was detected m/z 681.182 ($C_{32}H_{28}O_{15}$). In all cases, the derivative eluted 1.5 - 2 minutes apart from the described structure, excluding in source fragmentation. Strikingly, the undescribed demethylated α -Naphthocyclinone produced a pronounced sequence of inhibited fractions when tested against *S.aureus*.

Ultimately, it appears that both methodologies are required for maximized information retrieval.

3.4.2 Bioactivity

The annotated compounds listed in Table 1, Table 2 Table 3 and Table 4 plus the extension via molecular networking might explain the observed bioactivity pattern of the crude extracts.

Extracts of strain ST101789 cultured in 5315 showed growth inhibition of *S.aureus*. Automatic annotation and μ -fractionation could correlate the growth inhibitory effect of the crude extract with the Naphthocyclinones, a group of known anti Gram-positive compounds [168] targeting the leucin aminoacyl-tRNA-synthase. The enzyme inhibition was demonstrated to be competitive and reversible as increased concentrations of amino acids in the assay broth reduced the effect. According to the herein presented data, no effect on Gram-negative bacteria was observed. Growth inhibition comparison of wild type and L-form *E.coli* indicated resistance due to restricted penetration of the antibiotic into the cells [86].

Strain ST106693 was shown to be a potent producer of Anguinomycin A and B (in 5315 and 5294) and various variants of Conglobatin [165](in 5315). The respective crude extracts exhibited specific growth inhibition against *S.aureus*. Anguinomycins are primarily described as anti tumor compounds showing cytotoxicity towards murine P388 leukemia cells by inhibition of CRM-1 mediated nuclear protein export [21]. As of today, no antimicrobial bioactivity is postulated. Considering the pronounced lipophilic polyketide chain of Anguinomycin A and B an unspecific membrane integration/perturbation effect might cause the growth inhibition of *S.aureus*, while the LPS layer might have protected the Gram-negative test strains against this effect. Conglobatins belong to the Elaiophylin family of compounds [63]. Although Conglobatin was primarily described as antitumor agent, other members of this molecular family were observed to possess growth inhibitory activity against Gram-positive bacteria. Elaiophylins showed antimicrobial activity against *S.aureus*, *B.subtilis* and *E.faecium*, whereas Gram-negative pathogens, yeast and fungi were not susceptible [94] [65]. The name giving compound, elaiophylin tends to form cation selective ion channels in lipid bilayers [58], a mechanism which might explain the observed antibiotic activity of Conglobatin against *S.aureus*.

Extracts of strain ST1017165 exhibited growth inhibitory effects against Gram-positive bacteria, yeast and *E.coli* if tested in medium supplemented with physiological concentrations of bicarbonate. This pattern might be explained by a group of macrolides (Scopafungin, Niphimycin C, Amycin A and N'-Methylniphimycin) observed in essentially all extracts of the strain and within the bioactive fractions of extracts obtained from cultivations in 5294 medium. Scopafungin was extensively screened for antimicrobial and antifungal effects, yielding, among others, activity against *S.aureus* (MIC = 8 µg/mL [78]), *M.smegmatis* [118] as well as *C.albicans* (MIC = 16 µg/ml [143]). Closely related analoges Niphimycin C, Amycin A and N'-Methylniphimycin exhibit the same activity pattern [57]. Interestingly, ST107165 extracts obtained from 5294 cultivates additionally exhibited activity against *E.coli* if screened in bicarbonate buffered medium. Bicarbonate increases the cAMP concentration within bacterial cells, which reduces the availability of initiation factors IF1-3 and ultimately impairs the protein biosynthesis [41]. Damaged protein biosynthesis, is thought to heavily impact the outer membrane structure in Gram-negative bacteria, as crucial transport enzymes and major outer membrane proteins are missing. In addition, lacking or poorly assembled Lipid A - LPS modules leave the destabilized phospholipid bilayer unprotected so that the lipophilic fatty acid side chain of the Niphimycins might integrate and disrupt the membranes leading to cell leakage and death. The antitumor compound Echoside A (DNA Topoisomerase inhibitor ([40])) were mostly biosynthesized

in 5294 medium and might have contributed to the unspecific growth inhibition observed.

Finally, extracts of ST107645 were observed to inhibit the growth of Gram-positive bacteria and yeast, whereas the Gram-negative test strains remained unaffected. Annotated compounds Resomycin A and B [111] were correlated to the observed bioactivity of the crude extract by μ -fractionation against *S.aureus*. Interestingly, an additional bioactive metabolite, which could not be correlated to a literature known compound, was identified by μ -fractionation of the extract exhibiting the most pronounced activity. A follow up study should be targeting the isolation, structural characterization and bioactivity profiling of this molecule.

Summary This study indicates the value of a multipurpose secondary metabolomics platform for semi-automatic analysis of huge data sets. Data bucketing followed by PCA or MSH substantially helped to evaluate the data structure and observe underlying trends concerning growth behavior of the investigated microorganisms on the basis of their metabolite output. Furthermore, automatic annotation proved to be an elegant way to characterize and prioritize large (microbial)extract libraries in a time effective manner. Compound titer can be quickly compared across samples. Bucket annotation and in particular molecular networking reduced distraction by known metabolites and thereby direct research attention to chemical novelty. I am convinced that further advances in instrumentation and data analysis approaches will revitalize NP drug discovery. However, I believe that even the most sophisticated algorithm cannot substitute the judgment and intuition of a scientist.

Finally, μ -fractionation connected bioactivity to single automatic annotated compounds and lead to the discovery of two putatively unknown anti Gram-positive compounds ('567' and '330'). Because this project is focused on the discovery of new structures with bioactivity against Gram-negative pathogens, these two compounds were not further investigated herein, but nevertheless represent intriguing starting points for follow up projects. In pursuit of main objective, the described methodology was used to analyze the metabolite output of environmental isolates obtained from Lake Stechlin (section 4).

4 Bioprospecting and characterization of the bacterial community of Lake Stechlin

4.1 Introduction

Natural occurring chemicals mediate a microorganism's reaction to environmental stimuli, as well as inter and intra species signaling and communication. While competing for resources like nutrients and space, microbes biosynthesize bioactive molecules with potential pharmaceutical value. This observation leads to the approach of biosprospecting "high value", nutrient rich microbial habitats, as not only diversity but also immense competition and communication among strains can be expected. Strains, which acquired the advantageous capacity to biosynthesize antibiotic agents, might be favored in such environments.

Filter feeding zooplankton In this context, symbiotic micro-macroorganismal relationships in freshwater ecosystems represent a surprisingly underexplored source for novel bioactive substances. Crustacean zooplankton species are regarded as valuable bacterial habitats, as they provide bacterial attachment sites [120] and accumulate organic matter in their guts and fecal pellets [152]. Evidently, microorganisms are not randomly recruited from the environment, but form a specialized functional community in close interaction with the host [120][38]. The microbiome provides metabolic flexibility and substantially contributes to the genetic diversity of the host [107]. Most zooplankton species are filter feeders, hence their microbiome is in constant contact with potentially pathogenic microorganisms from the surrounding freshwater body [155][112]. A robust chemical defense system is likely to decrease the hosts susceptibility and thereby protect the integrity of the associated microbial community.

Due to their sensitivity to chemical contaminants plankton genera like *Daphnia* are appreciated by ecotoxicologist as natural markers for the evaluation of freshwater quality in aquatic environments. Accordingly, it is quite unlikely that zooplankton associated microorganisms defend their habitat by means of cytotoxic molecules, but rather employ specialized antibacterial agents. Here, it is hypothesized that the associated microbiome prevents successful colonization, thus infection of their host, by pathogenic microorganisms. While competing with other microorganisms, associated and symbiotic strains might optimize the host's chemical defense system against pathogens by biosynthesizing specific antimicrobial compounds.

Particulate organic matter (POM) in freshwater ecosystems Another promising source for microbial bioactive scaffold architects are organic particles, as they represent carbon- and nutrient-rich micro zones in otherwise meso- or oligotrophic aquatic environments [9]. By creating a spatial heterogeneity of nutrient availability, these particles form specific bacterial niches, and thereby influence abundance as well as species composition of microbes. It is known that attached-living bacteria may be physiologically and metabolically distinct from strictly planktonic, free-living species [9], [140]. "Particle specialists" might possess unique adaptations for an attached lifestyle like high expression activity of hydrolytic enzymes to facilitate nutrient utilization of particles [79]. Because particulate organic substrate is obviously of great value in the microbial world, it is not surprising that particle-colonizing strains were shown to exhibit growth inhibitory activity against competing bacteria [105]. It is likely that the biosynthesis of antimicrobial compounds is involved in the integrity protection of the short-term community on the particle surface. This chemically mediated antagonism between microbes within the particle matrix might be considered a valuable source for novel antimicrobial molecules.

Aim of this study The following experiments represent the first steps of a NP discovery survey from the environmental sample to new substances. The primary goal was to sample distinct bacterial communities from plankton, organic particles and the surrounding freshwater itself to bring a large diversity of specialized bacteria into laboratory culture. At the same time, the phylogenetic composition of these communities was analyzed on the basis of 16S rRNA gene amplicon sequencing. Sequencing was primarily carried out to evaluate the bacterial communities associated to plankton organisms and the surrounding water column of Lake Stechlin. Evaluation followed three major aspects: First, the authenticity of the starting material retrieved from the environment was analyzed by comparison with already published data from the same and similar habitats. Second, the communities were compared among each other to exclude unnecessary redundancy. Ultimately, the overall community was explored with respect to genera of well known NP biosynthesizing bacteria.

4.2 Material and Methods

4.2.1 Sampling of microorganisms from Lake Stechlin

Sampling was carried out at August 19th 2016 at 9 am from the mesocosm facility of the Leibniz-Institute of Freshwater Ecology and Inland Fisheries at lake Stechlin

(53°09' 3.35" N, 13°01' 20.53" E). Plankton samples were taken with two different plankton nets: A 250 µm mesh size was used to capture only large zooplankton organisms, while the second net (90 µm mesh size) was used for additional sampling of smaller phytoplankton specimen. Integral hauls of the water column from 15 m to the surface were carried out four times by each net type. The collected organisms were immediately transferred from the concentrator unit into sterile 50 mL Falcon tubes. Samples were constantly kept cool (4°C) until processing (48h). Water samples were taken at the same position as the plankton samples by lowering a custom water trap to the biomass maximum of 13 m below the surface. Cryptophyta biomass [µg/L] was used as a proxy to determine the biomass maximum depth (data provided by the IGB, Figure S9). The water trap was lowered 5 times to collect 15 L water.

4.2.2 Sample preparation

Water and Plankton samples were pre-processed immediately after sampling at the laboratory facilities of the IGB. The water samples were filtered sequentially:

- 1) pore size 5 µm (Durapore membrane PVDF, 25 mm SVLP02500)
- 2) pore size 1.2 µm (EMD Millipore MF Millipore, mixed cellulose ester 25 mm, RAWP02500)
- 3) pore size 0.1 µm (EMD Millipore Durapore PVDF, 25 mm HVLP02500)

The 5 µm filter membrane was used to remove large particles as well as zoo- and phytoplankton specimen. The membranes were not used for further experiments. The filtrate was passed through the 1.2 µm membrane to retain organic particles and the bacterial community attached. Finally the filtrate was filtered by the 0.1 µm membrane to retain free living, planktonic microorganisms. In total, 600 mL lake water was filtered on three sets of filters. The filter membranes were kept in sterile lake water (0.1 µm flow through) at 4°C). Three filter of each pore size were frozen at -80°C for microbiome analysis. The remaining lake water was filter sterilized (from here on referred to as sterile lake water) and used for culture medium preparation.

Zoo- and phytoplankton samples were diluted 1:1 with sterile lake water to guarantee survival and minimize shifts in the microbial community. All samples were transferred in a portable cooling box at 4 °C to the laboratory facilities at the Industry Park Frankfurt Hoechst.

Particles and cells retained on the 1.2 µm and 0.1 µm membranes were re-suspended

by vortexing in 20 mL sterile lake water. The plankton organisms were concentrated with 30 µm cell filter (Miltanyi Biotec,130-101-812). The plankton on the filter was washed thoroughly with sterile lake water to exclude "unattached" microorganisms. Ultimately the cleaned plankton organisms were re-suspended in 20 mL sterile lake water. This washed plankton solution was homogenized on ice with a sterile Ultra Turrax (IKA) at a motor speed of 13.500 rpm. The plankton homogenate was separated from residual debris by filtration over a 5 µm syringe filter disc (Sartorius sedim, Minisart 17594). Each preparation was carried out in triplicate and an aliquot of each sample type was used for cell enumeration and for microbiome analysis.

4.2.3 Cell enumeration via fluorescence microscopy

To estimate the number of cells present in the six environmental sample filtrates, a 1 mL aliquot was filtered over a 0.1 µm polycarbonate filter membrane (Satorius, 23007-25N) and air dried. To lower background fluorescence, the filter were stained black in 2% acetic acid supplemented with 2% w/v of acid black 52 (Cas 5610-64-0, Sigma) overnight. Excess dye was removed from the filter by an acetic acid bath (2%). Five µL of 1:50 diluted SybrGreen1 nucleic acid stain (Cas 163795-75-3, Molecular Probes) was pipetted on the dry sample and incubated for 10 min in the dark. A fluorescence microscope (DM2500LED,Leica) was used to excite fluorescence of the cells on the filter. Five pictures were taken of different areas of each filter and analyzed using Adobe Photoshop (v.10). Each picture covered 0.004% (0.018mm²) of the total filter surface. An even distribution of cells on the filter was assumed and the mean of the cell counts was multiplied by a factor of 26737 to extrapolate to the cell titer x/mL. In order to cultivate environmental cells at the desired concentration the estimated cell numbers were used to dilute the water and plankton samples with "Stechlin medium" (330mg yeast extract, 330mg glucose, 330mg peptone in 1L lake Stechlin water from 13m depth, sterile filtered) to a final volume of 200 mL.

4.2.4 Microbiome analysis

DNA Extraction The phylogenetic composition of the bacterial community within the different sample types was analyzed on the basis of 16S rRNA gene amplicon sequencing. First, DNA was retrieved from the environmental samples by using the NucleoSpin®Soil DNA Kit (Machery-Nagel,Germany). Cell lysis, precipitation of contaminants, DNA binding, cleaning and finally elution was carried out according to the manufacturer protocol. Three aliquots from all

plankton samples (2mL each) were processed and later combined on one single NucleoSpin®Soil DNA binding column to maximize the DNA amount. DNA elution was carried out twice using 25 µL SE Buffer (60°C) The 1.2 µm and 0.1 µm membrane filter which were prepared during sequential filtration of the water samples, were transferred to the DNA bead tube using two sets of sterile forceps. The filter membrane was rolled into a cylinder with the top side facing inward. Following steps were done as described above. Obtained DNA extracts were stored at -50 °C prior to PCR and amplicon sequencing.

16S rRNA gene amplicon sequencing and processing The DNA extracts were processed by LGC Genomics (Berlin, Germany). Forward primer U341F (5'-CCT AYG GGR BGC ASC AG-3') and reverse primer U806R (5'-GGA CTA CNN GGG TAT CTA AT-3') were used to generate clonal clusters of the hypervariable V3-V4 region (U341F-U860R) on the flow cell. Amplicon clusters were sequenced by 300-bp paired-end read sequencing (Illumina MiSeq V3 system). Reads were distinguished based on their index sequences (demultiplexing) with the Illumina bcl2fastq 1.8.4 software and oriented according to their direction. Barcode and adaptor regions as well as primer sequences were trimmed. Short reads (≤ 100 b) and reads with incorrect or missing barcode(s) were excluded from the analysis. Forward and reverse reads were combined by BMerge 34.48. Remaining sequences were processed by the SILVA rRNA gene database project (SILVAngs 1.3)[137]. Reads were aligned (SINA v1.2.10 for ARB SVN, revision 21008) [135] against the curated SILVA SSU rRNA SEED database. During the quality check [137] sequences containing more than 2% of ambiguities or homopolymers were discarded. PCR artifacts, contaminations and reads with low alignment quality were identified based on SINA alignment scores (50 alignment identity, 40 alignment score) and excluded. Read redundancy was reduced (dereplicated) and data clustering was carried out using cd-hit-est (version 3.1.2; bioinformatics.org)[98] running in *accurate mode*. Sequence overhangs were disregarded and identity criteria of 1.00 and 0.98 applied. The longest sequence of each OTU was regarded as reference and classified by local nucleotide BLAST query against the non-redundant SILVA SSU Ref dataset (release 128; arb-silva.de) using blastn (version 2.2.30+; Blast.ncbi.gov) with standard settings [29]. Finally, taxonomic information was quantified by mapping the classification of all reference OTUs on all reads assigned to the respective OTU. Reads with no or weak BLAST hits remained unclassified and were assigned to the meta group "No relative" in the SILVAngs fingerprint and Krona charts (RIBOCON GmbH, Bremen, Germany) [130]. Reads were defined as weak if the Blast quality (θ_{Bq}) ≤ 93

$$\theta_{Bq} = \frac{x_{id} - y_{cov}}{2} \quad (3)$$

where x_{id} is the sequence identity [%] and y_{cov} the alignment coverage [%] of the given OTU. The described work flow was first carried out and published by Ionescu [72] and Klindworth [83] This study focused on the taxonomic composition of the different bacterial communities, thus reads assigned to archaea, mitochondria and chloroplasts were neglected. Results were visualized using Krona charts by RIBOCON GmbH.

Statistical analysis Comparison between sample types was carried out by principle component analysis using the statistics software package Past (v3.14). For further analysis, the OTUs detected within the triplicates of one sample type were pooled. OTU richness and diversity indices of the individual samples were calculated using the same software package.

4.2.5 Cultivation and conservation

Based on the cell number estimates, the environmental cells were cultured in 96 well plates at two different concentrations: pure cultures (0.3 cells/well) and synthetic communities (10 cells/well). The concentrations were adjusted by dilution with "Stechlin medium" (sterile Lake Stechlin water from 9m depth supplemented with 0.3 g Glucose, Yeast extract and peptone).

The culture volume of 150 μ L/well was distributed to 15 microtiter plates per sample type and concentration by a semi-automatic liquid dispenser (Multidrop, MTX Lab Systems). As soon as cultured growth was observed (turbidity determination by eye compared to medium control), the culture plates were splitted: 75 μ L were used to inoculate 1.5mL artificial lake water (ALW) medium in a 96 deep well plate (Greiner, 780280). Plates were incubated for 14 days at 24 °C and 180 rpm. The other half of the cultures was mixed with glycerol 80% and stored at -80°C. Medium distribution as well as glycerol supplementation was carried out using a liquid handling robot (Matrix, ThermoScientific).

4.2.6 Bioactivity assessment via quick supernatant *lux* assay

After incubation, the bioactivity of the environmental cultures was assessed by quick supernatant *lux* assay in order to prioritize the large amount of cultivates. The cultivates (in deep well plates) were centrifuged (3320 x g, 30min) and 100 μ L cell free supernatant was transferred to the assay test plate (Corning, 3603).

Assay plates were dried *in vacuo* before a two point detection was carried out for each environmental culture. The test plates were prepared according to the plate design (see Figure S11).

The assay itself was conducted following a previously established *in-house* protocol [71] Briefly, an overnight culture (Mueller Hinton II(BD) + 100 µg/mL ampicillin, 28°C, 180rpm) of the bio-luminescent screening strain *E. coli* DH5α[pFU 166][158] was diluted to an OD_{600nm} of 0.02. To each test well 50 µL adjusted screening culture was added. During the assay the incubation parameters were kept constant at 37°C, 180rpm and 95% humidity. A dilution series of gentamycin sulfate (Sigma Aldrich, cas 1405-41-0) was used as positive control. Concentrations of 16-0.125 µg/mL were used to achieve effects ranging from complete growth inhibition to no inhibition. *E. coli* cell suspensions without supernatant and antibiotic were used as negative controls (negative CON)

Bio-luminescence, as a proxy for cell viability, was assessed after 6 hours (t_{6h}) using a microplate spectrophotometer (Wallac Victor² 1490, Perkin Elmer). The percent growth inhibition was calculated from the luminescence units (LU):

$$Cell\ viability\ inhibition[\%] = 100 * [1 - \frac{LU_{sample} - LU_{negativeCON}}{LU_{positiveCON} - LU_{negativeCON}}] \quad (4)$$

Culture supernatants achieving at least 85% inhibition of test strain viability were considered bioactive and selected for further testing. Follow up experiments included strain cultivation in a larger culture volume (4mL in 24 well plates), methanolic extract preparation and screening against a broader panel of test strains.

4.2.7 Phylogenetic identification based on 16S rRNA gene sequencing

All environmental cultures exhibiting growth inhibitory effects against *E. coli* DH5α[pFU 166] (based on supernatant *lux* assay) were subject to 16S rRNA gene sequencing. DNA was extracted by mechanical cell disruption using a TissueLyser II bead mill (Qiagen, Hilden, Germany). Amplification was carried out using oligonucleotide primer E8F (5'-AGAGTTTGATCCTGGCTCAG-3') and EUB1492R (5'-ACGGYTACCTTGTACGACTT-3')[164]. Subsequently sequencing was done by LGC Genomics (Berlin, Germany). Sequences were processed using Geneious R10 (Biomatters Ltd.) and aligned against the NCBI database (Blast.ncbi.gov) using the Megablast algorithm [123].

4.3 Results

4.3.1 Microbiome analysis

Overview After demultiplexing, adapter and primer region clipping a total of 323817 combined reads with an average length of 422 bp were obtained from the environmental samples. A proportion of 99,6 % (322441) reads was successfully classified into 578 different OTUs organized in 26 phyla. Interestingly, more than half of all sequences were obtained from the sample containing only larger plankton organisms, which were captured with the 250 μm mesh size net (see Table 5). The lowest number of OTUs was observed in the free living water community (W 0.1). Taxomic assignment led to identification of the 4 major phyla ($\geq 5\%$ of all analysed sequences) Proteobacteria (58.5 %), Bacteroidetes (19.3 %), Actinobacteria (8.4 %) and Verrucomicrobia (6.6 %). The other 22 phyla were present at lower levels across the samples or in some cases exclusively in one sample type. A total of 753 sequences remained unclassified due to weak or no BLAST hits and were assigned to the meta group "No relative". In addition, 588 sequences were rejected during quality control. A first visualization of phylum distribution was provided by RIBOCON GmbH in the form of Krona charts (s. Figure 17, Figure 18).

Table 5: Microbiome overview

Sample ID	Combined reads	OTS	Phyla	Richness and diversity indices				
				Dominance	Simpson	Shannon	Evenness	Chao-1
PL90	48096	218	21	0.10	0.88	3.21	0.12	218
PL250	173321	283	18	0.13	0.87	2.70	0.05	284
W1.2	60551	172	19	0.04	0.96	4.00	0.31	173
W0.1	41849	115	17	0.06	0.94	3.54	0.30	115

Bacterial community structure The community composition of plankton and water samples differs with respect to the detected bacterial phyla. Both plankton samples were dominated by members of the Proteobacteria (64-78 %), while Actinobacteria were the most abundant phylum in the 0.1 μm water sample (41 %). The community retained on the 1.2 μm filter was substantially composed of the phyla Bacteroidetes (26.9 %) and Verrucomicrobia (22.3 %)(see Figure 19). Interestingly, the community structure within the Proteobacteria differs among the two plankton samples: PL250 was composed of equal shares of Gammaproteobacteria (48 %) and Betaproteobacteria (49 %), while a considerable abundance of Alphaproteobacteria (27 %) and reduced presence of Gammaproteobacteria was detected in the 90 μm samples. The notable presence of Gammaproteobacteria in the larger plankton samples is to a large extent attributed to genera *Aeromonas* and *Rheinheimera*. Almost all Bacteroidetes OTUs found in the plankton samples were assigned to the family Flavobacteriaceae (94-98 %), while in the communities retained by the 0.1 and 1.2 μm filter the family Chitinophagaceae (up to 20 %) was also well represented. In addition to the 4 major phyla (see above) Firmicutes, Epsilonbacteraeota, Cyanobacteria, Patescibacteria, Chloroflexi, Planktomyces and Acidobacteria were found in all samples, but in varying abundances and different OTU compositions. Only 39% of all detected OTUs are shared among plankton and water samples, while the rest is either only detected in plankton samples (15%) or in the two water samples (46%). The phyla Aegiribacteria, Hydrogenedente and Elusimicrobia were exclusively present in the 90 μm plankton sample. The Modulibacteria were only seen in the 250 μm sample. Furthermore, the two phyla Tenericute and Fibrobacteres were shared across the two different plankton communities, but not found in the water samples. In contrast, sequences assigned to the phyla Deinococcus-Thermus and Dependientiae were solely observed in the water communities.

Richness and Diversity The relative OTU distribution within the environmental samples was compared by principle component analysis (see Figure 20). The two most significant factors (Component 1 and Component 2) accounted for 72% (cumulative) of the total variance. Most importantly, score values belonging to different environmental sample types are clearly separated in the scatter plot, whereas samples of the same sample type (triplicates) cluster together. The plankton sample sets are both unique and form defined cluster apart from each other and the water samples. Generally, all water samples lay closer together, but groups composed of the 1.2 μm and 0.1 μm membrane filter triplicates are well differentiated within the cluster.

Bacterial richness and taxa diversity within the analyzed samples is summarized in

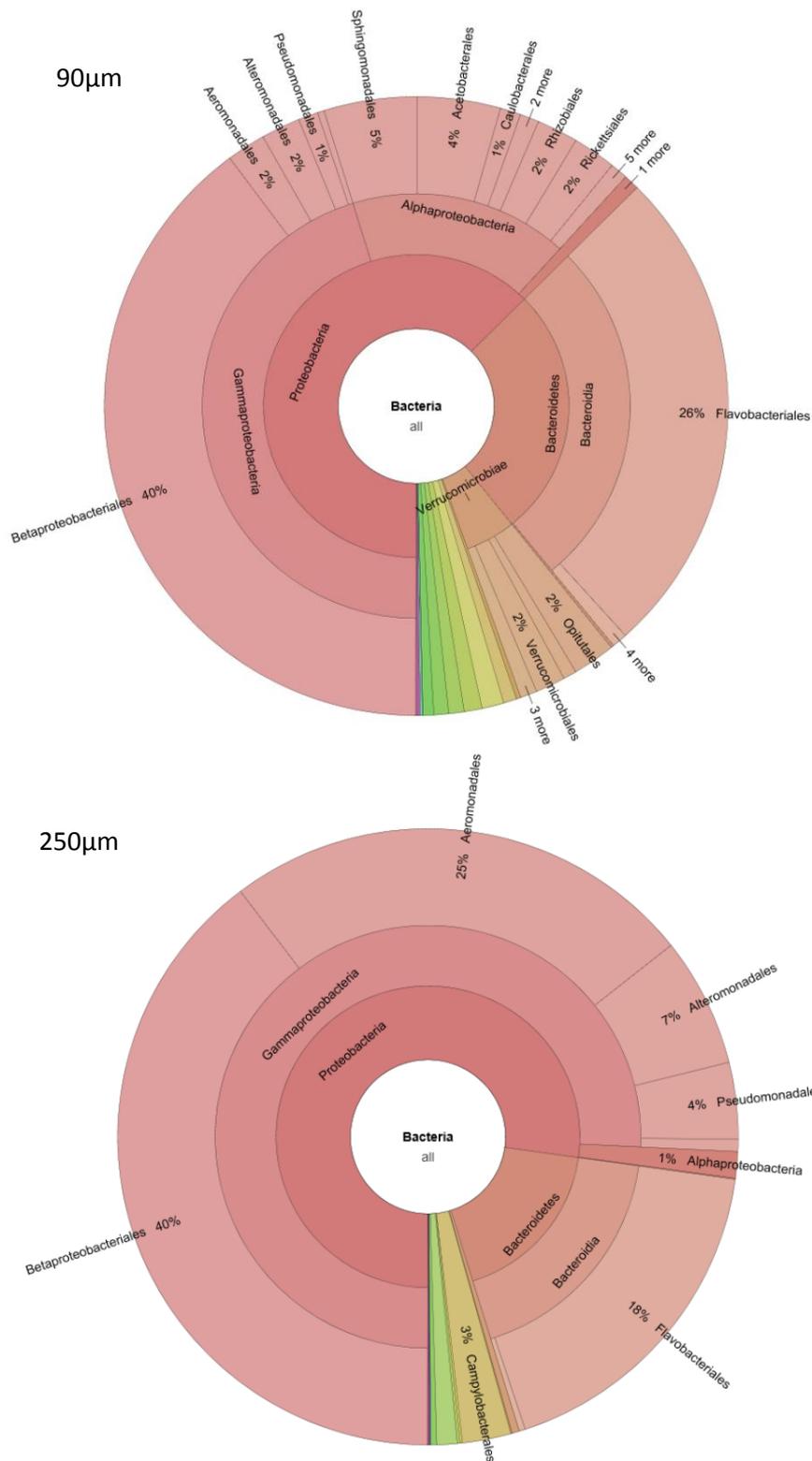


Figure 17: Top: Visualization of OTU distribution within the 90 µm sample on phylum, class and order level. Bottom: OTU distribution within 250 µm samples. Both communities are dominated by the phylum Proteobacteria, but differ in phylum composition: The proteobacterial community of PL250 is composed of equal shares of Gamma and Betaproteobacteria, while a notable abundance of Alphaproteobacteria and reduced a number of Gammaproteobacteria was observed in PL90

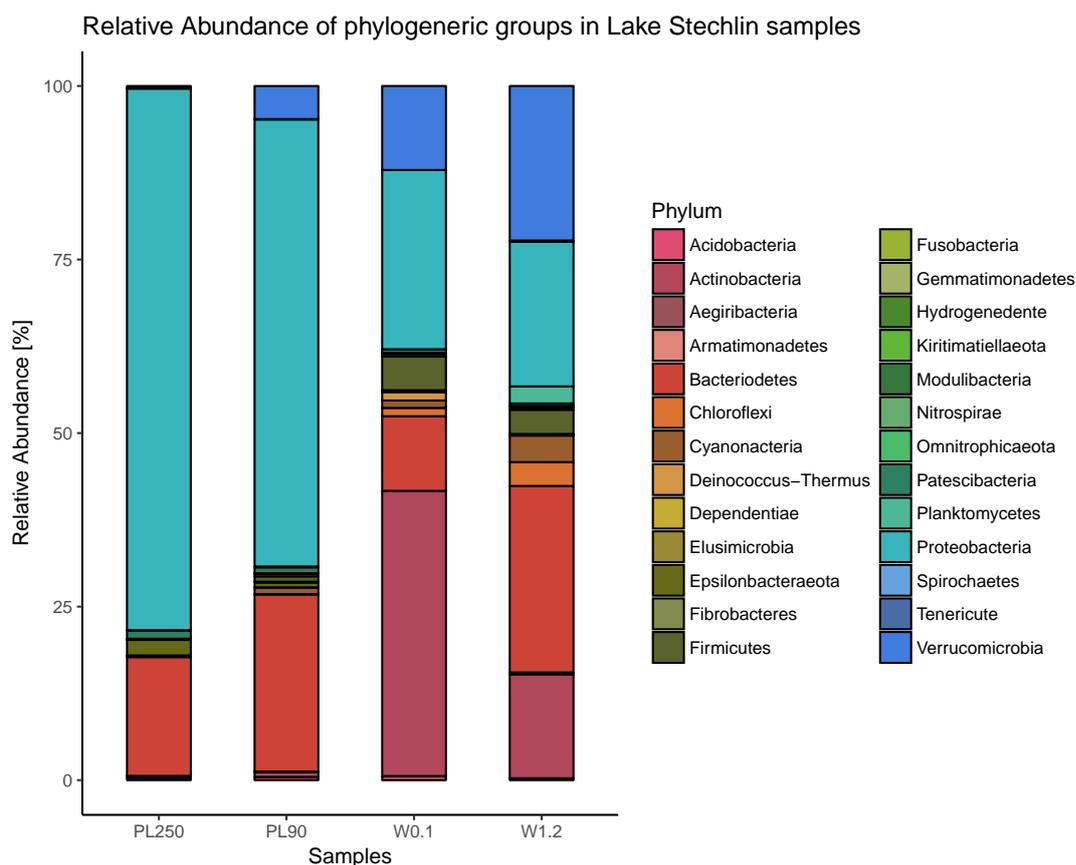


Figure 19: Relative abundance of phylogenetic groups in plankton (PL250, PL90) and water (w0.1,w1.2) bacterial communities retrieved from Lake Stechlin (s. also Figure S10)

Table 5. Most strikingly, the plankton samples (PL90 and PL250) are less diverse with respect to OTU distribution compared to the water samples (W1.2 and W0.1). This observation is reflected by a lower Simpson diversity index (0.87-0.88 compared to 0.94-0.96) and higher Dominance values (0.10-0.13 compared to 0.04-0.06). Additionally, W1.2 and W0.1 share higher Evenness values suggesting a homogeneous distribution of sequences.

4.3.2 Cultivation and bioactivity assessment

Cell enumeration and cultivation The cell numbers estimated by fluorescence microscopy assisted counting are summarized in Table 6. The highest cell numbers were observed in association with the larger plankton organisms. Notably, both plankton samples showed cell numbers of at least one order of magnitude higher than the water filter samples. Samples were diluted accordingly and after incubation a total of 9136 wells exhibited microbial growth.

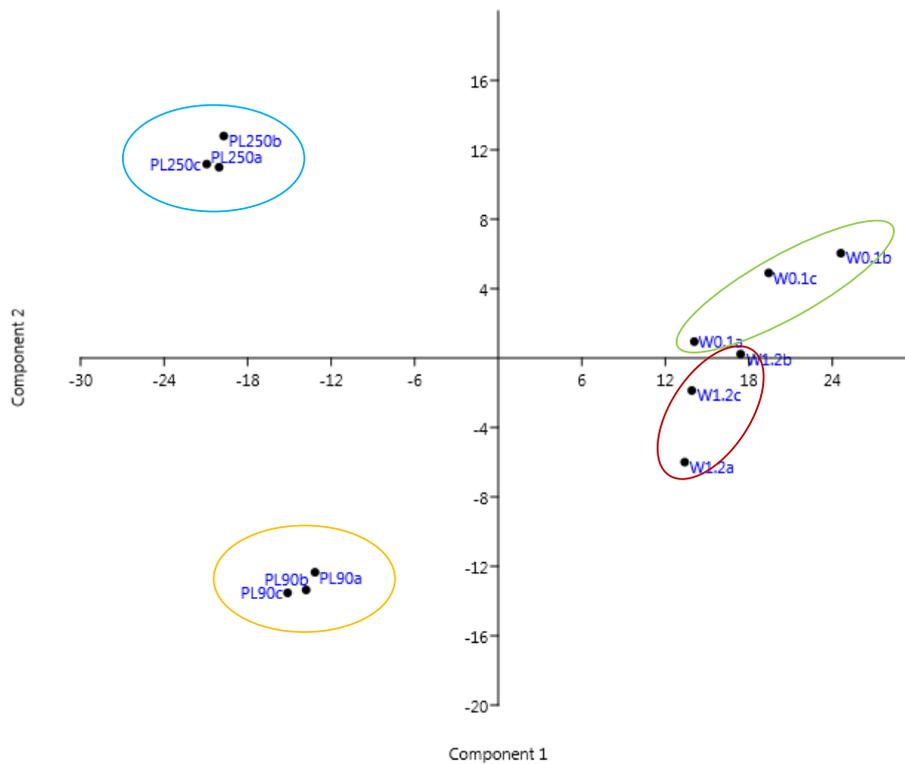


Figure 20: Principle component analysis visualized as scores plot of OTU abundance in bacterial communities retrieved from Lake Stechlin. For each sample type, three samples were retrieved: Score values of the OTU abundance distribution within the different bacterial communities are represented in a 2D-scatter plot wherein principle component 1 and 2 account for 72 % of the total variance of the data. Scores of communities retrieved from the same source (different plankton nets or water filter) form distinct cluster: w0.1 samples; red cluster: w1.2 samples.

Table 6: Estimated cell numbers in Lake Stechlin samples: Cell number of plankton net samples was obtained by concentrating and homogenizing plankton net hauls before cell counting. Cells within water samples were concentrated on either 1.2 or 0.1 μm filter discs.

Sample	ID	Cells * mL ⁻¹
Plankton net 90 μm	PL90	$1.4 * 10^7$
Plankton net 250 μm	PL250	$8.5 * 10^7$
1.2 μm filter disc	W1.2	$2.5 * 10^6$
0.1 μm filter disc	W0.1	$1.2 * 10^6$

Bioactivity evaluation Culture broth supernatants obtained from Lake Stechlin cultivates (s. section 4.3.2) were screened for inhibitory effects against the test strain *E.coli* DH5 α [pFU 166][158]. In total, 40 supernatants (0.35%) exhibited a cell viability decreasing effect on the test strain of at least 85 % compared

to the untreated growth control (s. Equation 4). Three-quarters of the active supernatants were obtained from the bacterial cultures from the plankton samples (PL90 52.5 %; PL250 22.5 %). The phylogenetic classification of these cultures based on 16S rRNA gene sequencing is summarized in Figure 21.

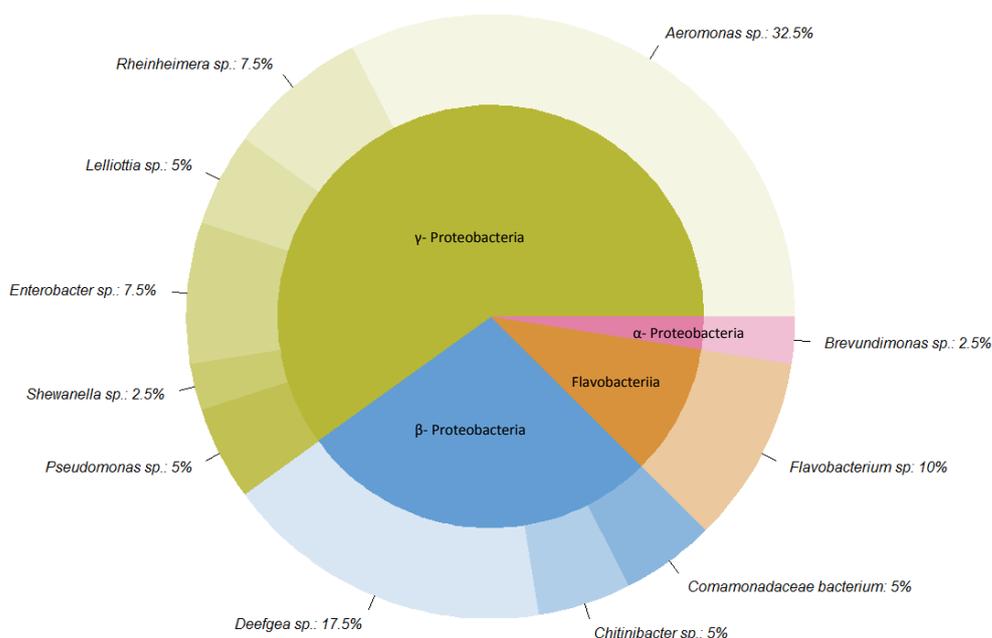
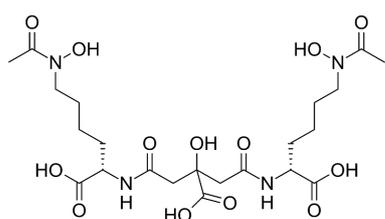


Figure 21: Phylogenetic classification of Stechlin cultivates exhibiting bioactivity against *E. coli* DH5α[pFU 166] as determined by supernatant *lux*-assay

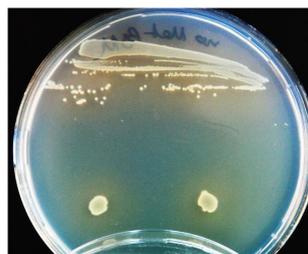
Follow up analysis of MeOH-extracts To allow further analysis and prioritization, the cultivates (Figure 21) were re-fermented and an organic extract (methanol) was prepared. Extracts of bacteria identified as *Shewanella* spp., *Deefgea* sp., *Brevundimonas* sp., *Lelliottia* sp. and *Rheinheimera* sp. showed no growth inhibition against any of the test strains after extraction with methanol. The methanolic extracts of half of the four Flavobacteriia strains (*Flavobacterium* sp. FhG100040 and FhG100042) did inhibit *E. coli* ATCC 35218 TEM1 in MHII medium supplemented with physiological concentrations of sodium-bicarbonate. Both extracts were μfractionated (subsubsection 5.3.2). Screening of obtained fractions showed growth inhibition corresponding to the injection peak (in fractions 3-4).

Extracts of *Enterobacter* sp. FhG100039 were active against *E. coli* ATCC 35218 TEM1 and weakly active against *P. aeruginosa* ATCC 27853 in MHII + bicarbonate. After μ-fractionation extract FhG100039 was retested against *E. coli*

ATCC 35218 TEM1 in MHC medium. Both extract controls (2 μ L and 5 μ L) as well as fraction 22 of the 5 μ L injection inhibited the growth of the test strain. Semi-automated dereplication via MS/MS networking suggested the presence of the siderophore Aerobactin [55] in the extract. Remarkably, the corresponding mass 565.235 $[M + H]^+$ could be detected in the active fraction. Manual analysis of the MS fragmentation pattern within the raw data could proof the result of the networking experiment. The CAS-agar assay of strain FhG100039 indicated the biosynthesis of a siderophore (s. Figure 22).



(a) Aerobactin



(b) CAS assay FhG100039

Figure 22: Left: Structure of Aerobactin. The compound was found in the bioactive fraction of the extract FhG100039 when tested against *E.coli* ATCC 35218 TEM1 in MHC medium. Right: Overnight culture of strain FhG100039 grown on BSM medium in iron depleted conditions and overlaid with CAS agar.

***Pseudomonas* spp.** A share of 5 % (2 cultures) of the environmental samples causing growth inhibition of the test strain in the supernatant *lux*-assay could be identified as members of the genus *Pseudomonas*. The methanolic extract of one of them, *Pseudomonas* sp. FhG100052, exhibited moderate growth inhibition against *E.coli* ATCC35218 tested in Mueller Hinton II medium supplemented bicarbonate and *E.coli* DH5 α [pFU 166] in regular Mueller Hinton medium. Additionally a strong strong activity against *C.albicans* FH2173 and *M.smegmatis* ATCC 607 was detected in MTT and BTG assays. Accordingly, the extract was μ -fractionated against these 4 test strains. In the case of *M. smegmatis* no fraction was active except fractions 3-4 (both injection volumina), which corresponds to the injection peak.

In addition to fractions 3-4, screening for reduced microbial viability (BacTiter-Glo™, Promega) against *C. albicans* revealed bioactivity in fractions 103-112 (s. subsection 5.3.2). The same fractions were seen to exhibit moderate growth inhibition (in *lux*-assay) against *E. coli* DH5 α [pFU 166]. Molecular networking analysis revealed a high abundance of the cyclic lipopeptide Tensin [8] within the

active fraction. Besides, at least five undescribed, structurally related molecules were found in the cluster (s. Figure 23).

4.4 Discussion

4.4.1 Microbiome analysis

Overall, this study is entitled to evaluate the taxonomic diversity of the bacterial community associated to plankton and particles as well as the free living community in the water column with respect to their value in natural product research. However, apart from recognizing the microorganisms solely for their ability to produce NPs, this data set allows curious insights into the ecology of freshwater plankton associated bacteria.

First, both the free living bacteria (W 0.1) and the particle associated water community (W 1.2) exhibited a broader diversity of rare OTUs (higher Shannon and Simpson index, lower dominance and higher evenness scores) compared to the two plankton samples. Overall, these ecological diversity estimators indicate a rather low OTU diversity when compared to reports from other environments, such as rhizosphere soil (Shannon 9-10 [95]).

A high diversity of unique OTUs (46% of total OTUs only found in water samples) in low individual abundance is frequently observed in freshwater environments. These rare OTUs in ambient water might provide a reservoir of microorganisms adapted to conditions different than the prevailing ones [37]. The same data also suggest a rather specific plankton microbiome (only 39% shared OTUs), dominated by a limited set of adapted taxa (15% OTUs only found in plankton samples). Several studies suggested a constant exchange of bacteria between ambient water and zooplankton communities [60], but limited to specific taxa [37]. Apparently, a major community shaping aspect is the capacity to metabolize material synthesized by the plankton [38], such as moults and digestion (by)-products.

An earlier study has already shown that the bacterial community associated with zooplankton organisms from Lake Stechlin, in contrast to the free living water community, is dominated by Alpha- and Betaproteobacteria [60]. In particular the a large number of reported Sphingomonadaceae bacteria could also be observed in the herein analyzed plankton microbiomes (37 % of all Alphaproteobacteria detected in the two plankton samples combined). The conformity between these two studies might indicate a close and rather stable association of Sphingomonadaceae bacteria and the zooplankton community of Lake Stechlin. In contrast, the authors highlighted the genus *Porphyrobacter* (Alphaproteobacteria) and contributed its

pronounced appearance to its chitinolytic capacity, thus adaptation to grow on zooplankton bodies and carcasses [60] in the otherwise oligotrophic Lake Stechlin. Remarkably, in this study no OTU could be assigned to this genus. Instead, a variety of other literature known chitinolytic genera were enriched in the plankton samples. In this context, both genera, *Aeromonas* as well as *Flavobacterium* were highly abundant. Association of these genera on chitin containing particles is a well described phenomenon: While *A. hydrophila* is able to degrade chitin by extracellular chitinases, the *Flavobacterium* was shown to "steal" degradation products using cell-associated enzymes [74]. Besides, other taxa, like Chitinibacteraceae (*Chitinibacter*, *Chitinilyticum*, *Deefgea* and *Idobacter*) members of the class Chitinivibrionia were highly enriched in plankton samples. The exoskeleton of zooplankton represents a major source of chitin in aquatic ecosystems, thus bacteria possessing the enzymatic repertoire to make chitin bioavailable for other organisms play an important role in the carbon and nitrogen cycling of the lake [153] [38]. The phylogenetic discrepancy, but functional similarity of the presented and the previously reported data from lake Stechlin might be caused by the region of attachment. Bacteria, which are metabolically adapted to chitin utilization might accumulate at the outer surface rather than the protective interior of the zooplankton organisms. As a consequence they might be more susceptible to other ecological and biogeochemical factors and ultimately subject to (seasonal) phylogenetic fluctuation.

In this study, bacteria belonging to *Aeromonas* spp. as well as representatives of *E. coli*, *Klebsiella* spp. and *Pseudomonas* spp. were enriched in the community obtained from the larger (zoo)plankton organisms. Opportunistic pathogens are often isolated from aquatic ecosystems [112]. Genera like *Aeromonas* spp., *Rheinheimera* spp. or members of the Neisseriaceae are frequently reported to colonize marine [112] as well as freshwater plankton [60]. Another study compared *Aeromonas* spp. concentrations attached to zooplankton and the surrounding water column and found a difference of six orders of magnitude [155]. The interaction between hygienically relevant microorganisms like *Vibrio* spp. and marine zooplankton is already well described (review [44]). Huq and coworkers could even establish a correlation between the outbreaks of cholera epidemics in Bangladesh and seasonal plankton blooms in the Bay of Bengal [70] [81]. In this context other projects are also interested in the role of freshwater plankton as vectors for pathogen dispersal (e.g. Prof. Flemming, University of Duisburg) and hopefully illuminate further details of this relationship. Whether the described group of Gram negative bacteria actually represent a threat to man remains uncertain, as no strain identification was carried out. Nonetheless, it is intriguing to argue that other members of the associated microbiome might somehow be

able to confine the spread of these opportunists on the plankton organism.

These examples demonstrate, that the observed community structure is comparable to earlier observations and might therefore genuinely reflect the natural state at the time of sampling. It was quite important to acquire this seemingly trivial result to exclude contamination or other sampling induced effects. Also, the bacterial community retrieved from Lake Stechlin was of course essentially profiled to evaluate its value for natural product research. Importantly, the applied sampling strategies helped to sample distinct bacterial communities (see Figure 20) and thereby increased the overall biodiversity available for further experiments. It was assumed that a broad biodiversity corresponds to a high genetic diversity, which in turn might translate into a large enzymatic repertoire and ultimately a diverse secondary metabolite output. Another important aspect is that the community structure consisted of a compelling mix of prolific NP producing phyla (Proteobacteria and Bacteroidetes (see Figure 19) and rather underexplored taxa, like Acidobacteria, Plancomycetes or freshwater Actinobacteria.

Freshwater Actinobacteria, unlike their marine and terrestrial relatives, are rarely considered as source for novel NPs. In part, this might be due to their rather small genome sizes (1.2-1.4Mb)[54], thus reduced enzymatic capacity. Microorganisms belonging to this group are usually characterized by very small cell volumes and [64] are known to be among the most abundant bacteria in freshwater habitats. In line with that, Actinobacteria were the dominating phyla on the 0.1 μm water filter and not detected in association with the zooplankton organisms. The lifestyle of freshwater Actinobacteria might be substantially different from their rather complex terrestrial counterparts. Hence, freshwater Actinobacteria might have adapted to oligotrophic habitats and lack the capacities to compete (for instance, on the basis of NPs) in nutrient rich environments. Only few representatives were successfully brought into axenic culture so far, already indicating rather specific requirements for growth and explaining absent reports of NPs isolated from these organisms.

On the other hand, the phylum Bacteroidetes represents a well recognized source for NPs [23]. Especially bacteria belonging to the genus *Flexibacter* and *Chitinophaga* have proven their value in NP research as indicated by the discovery of the β -lactam Formadicine [80], a group of macrolide antibiotics, the elansolides [76] and the lantibiotic Pinensin [119]. Interestingly, in the presented data set, the *Chitinophagales* are among the most prominent orders (8 % of all OTUs) within the bacterial community of Lake Stechlin, retained by the 1.2 μm filter.

Among the highly abundant Proteobacteria, the genus *Pseudomonas* was well present in all samples (up to 5% of all sequenced OTUs). Apart from its clinical

relevance, members of *Pseudomonas* spp. are considered key player in many microbial communities by sharing a tremendous diversity with respect to their genomic repertoire and metabolic capacity. Genome size varies from 4.5 - 7.1 Mb with 4237-6396 predicted genes [59]. Particularly interesting in this context is the relatively small core genome of *Pseudomonas* spp.: Comparative genomics studies revealed that only 40% of genes are conserved among different strains within this genus, conferring to a steadily growing pan-genome. Core genes usually code for central enzymes of the primary metabolism, essential for the survival of the bacterium, whereas the flexible part of the genome is believed to reflect adaptation to ecological requirements - including secondary metabolite production [59]. In general, pseudomonad bacteria are a remarkably rich source for novel chemistry, structurally distinct from other famous NP producer such as Actinomycetes, Bacilli, Cyano- and Myxobacteria (for example [126]).

A valuable addition to the present data set would be a metagenome analysis with a focus on conserved regions of known NP biosynthetic gene cluster, such as NRPS adenylation or PKS acyltransferase domains, to estimate the biosynthetic potential of the bacterial communities. All in all, the results obtained from this simplified analysis provided a solid basis to confidently move forward and conduct a NP discovery project.

4.4.2 Cultivation and Bioactivity assessment

After incubation, more wells than expected exhibited microbial growth (grown 9136, expected 7680). That might be due to difficulties during cell number estimation thus cell distribution in culture broth. Small cells and cells with low fluorescence signal might have been overlooked during counting by eye. Additionally, cells might not have been distributed evenly on the filter and thereby cell number extrapolation might have been inaccurate. Other methodologies like FACS assisted cell counting should be considered for future experiments.

Due to that, most obtained cultures were considered synthetic communities rather than pure cultures. Conveniently, most cultures exhibiting growth inhibitory effects in the culture broth supernatant *lux* assay were in fact pure cultures or at least largely dominated by one strain (according to visual inspection and 16S rRNA gene sequencing). Obviously, the selection of culture medium is a critical factor in this context. The dominant strains might have been favored by the medium composition, while others were not able to survive and/or replicate. It is also possible that the identified strains actively inhibited the growth of the other microorganisms within the well and in the following the growth of the screening strain. Metabolite synthesis and transport as a consequence of

resource competition or mutualism among cells are well accepted concept in the NP community [156]. Either direct cell contact or the recognition of "foreign" signals might up-regulate otherwise repressed transcription cascades. Known examples of shifted gene expression levels as a result of co cultivation comprise upregulation of genes associated with antibiotic pathways, such as pyocyanin in *Pseudomonas* sp., or elevated expression of (metallo-)beta-lactamase genes like *BetaLact* in *Roseobacter denitrificans* [36]. A series of experiments was carried out to investigate stimulus dependent metabolite production. During cultivation, a set of described signaling molecules such as N-hexanoyl-L-homoserine lactone [122] were supplemented to the culture broth, but no significant change in the metabolome of target strains was observed (data not shown).

This finding might also partly explain why most results of the primary supernatant assay could not be reproduced after re-fermentation. Purification and cultivation of evidently axenic cultures offers the advantage of reduced complexity, thus increased reproducibility, but might have restricted the metabolic output of target strains. Not only signals from a co-cultured microorganisms, but also other environmental stimuli were neglected during refermentation. In that sense it might be considered that the primary cultivation was carried out with sterile filtered lake water, while following cultivation were carried out in standardized *in vitro* media.

Another aspect contributing to the low number of growth inhibitory effects in follow up assays, might be due to the nature of the agents causing the initially observed inhibition. The conducted organic extraction and MTT-assay is primarily designed to investigate small, hydrophilic molecules which are stable in methanolic solution. For instance, one third of cultures exhibiting growth inhibitory activity in the supernatant assay were identified as members of the genus *Aeromonas*. Aeromonads are known to produce the proteinogenic toxin Aerolysin [27] [26]. The over 50kDa molecule is primarily released as protoxin, before being converted into its active form by proteases also released by the same bacteria [67]. Likewise, members of the genus *Rheinheimera* and *Pseudoalteromonas* were among the cultures passing the primary supernatant assay. Both genera are known to produce L-aminoacid oxidases (71kDa), possessing a wide antimicrobial activity spectrum, mainly due to generation of hydrogen peroxide [33]. Members of the genus *Shewanella* were reported to possess type II Toxin-Antitoxin loci and were proven to produce SO_ 3166, a potent toxin belonging to nucleotide-binding (HEPN) superfamily [167].

Of course, many strains are capable to produce both, proteinogenic toxins and small bioactive molecules, but still it appears worthwhile to consider the nature

of the molecules, which one would like to detect by the assay response. While the supernatant assay might have been able to depict such effects, the screening of organic extracts did not. During compound extraction the physio-chemical properties of the solution change dramatically, potentially leading to protein precipitation or degradation, thus exclusion from the assay. Depending on the research question, future experiments should include a modified assay: If large peptides and proteins are of interest, the cell free supernatant should be used directly for profiling and analytics. Otherwise, if working with small molecules, the lyophilized supernatants should be resolved in organic solvent once to exclude all sensitive agents before the first assay. After evaporation the assay could be carried out as described (s. subsection 4.2.6).

Despite that, some interesting bioactive molecules could be identified. First, bioactivity guided μ -fractionation in line with tandem MS analysis using molecular networking led to the discovery of Aerobactin in the methanolic extract of *Enterbacter* sp. FhG100039. The effect was only observed when the assay against *E.coli* ATCC35218, was carried out in Mueller Hinton II medium supplemented with physiological concentration of sodium bicarbonate.

The identification of Tensin and related unknown molecules, in the methanolic extract of *Pseudomonas* sp. FhG100052 represents the arguably most intriguing finding. Tensin itself was already described to exhibit anti fungal properties [128], but bioactivity against Gram-negative indicator strains was not postulated so far. Interestingly, the group of cyclic lipopeptides (CLPs) include a variety of famous molecules with activity against human pathogens [82], like Colistin and Daptomycin (s. subsection 5.2). Obviously, this group of similar molecules eluted in close proximity from the column, while using the standard gradient. Thereby a baseline separation via μ -fractionation was not achieved and the bioactivity could not clearly be addressed to one (or more) of the molecules. Strongest growth inhibitory effects were observed in fraction where Tensin was by far the most abundant ion, but this observation could also be contributed to concentration effects: The unknown CLPs were detected, in some cases, in one order of magnitude lower intensities, compared to Tensin within the extract. To distinguish between active and inactive variant a normalization by screening of defined concentrations should be conducted. The small alterations in the molecules architecture can determine the degree of activity or toxicity. Attracted by these observations, a follow up study targeting the isolation, structural characterization and bioactivity profiling of this group of molecules was carried out and is described in the next chapter (s. section 5).

5 Metabolomics-guided discovery of new cyclic lipopeptides from *Pseudomonas* sp. with anti-Gram-negative activity

5.1 Abstract

Bioactivity guided fractionation, followed by detailed metabolome analysis using molecular networking led to the discovery of five new cyclic lipopeptides (CLPs) in the culture broth of the γ -proteobacterium *Pseudomonas* sp. FHG100052. The new compounds resemble members of the Amphisin group [147] as they are constructed of a 3-hydroxy fatty acid linked to the N-terminus of an undecapeptide core. Additionally, a macrocycle, formed by lactonization of the C-terminus and D-*allo*-Thr4, is conserved in the new CLPs and all other members of the Amphisins group. Culture condition optimization led to the isolation and subsequent structure elucidation of five new and one known derivatives by extensive MS/MS and NMR experiments in combination with Marfey's analysis. The data were in agreement with *in silico* analysis of the corresponding biosynthetic gene cluster (BGC). Most strikingly, the length of the incorporated fatty acid seems to define the moderate growth inhibitory effects against *Moraxella catarrhalis* FH6810 as observed by MIC values ranging from no inhibition ($> 128 \mu\text{g/mL}$) to $4 \mu\text{g/mL}$.

The work described herein was published in the Journal of Natural Products and can be accessed at Marner et al. 2020 [110].

5.2 Introduction

The cyclic lipopeptides (CLPs) are an interesting group of specialized bacterial metabolites because they include a variety of molecules with activity against human pathogens [82]. Prominent representatives with clinical applications include the polymyxin Colistin (anti Gram-negative) [150] and the calcium dependent antibiotic Daptomycin (anti Gram-positive) [11]. Certain CPLs are also active against *Mycobacterium tuberculosis* [53], phytopathogenic fungi [128] and protists [56]. The amphipathicity of CLPs confers functional dualism, combining antimicrobial and biosurfactant properties [127].

CLPs are structurally composed of a peptide macrocycle (often lactam or lactone) linked to a fatty acid side chain. Despite these common features, this class of natural products exhibits considerable structural diversity, most notably because most CLPs are non-ribosomal peptides, allowing the incorporation of

non-proteinogenic amino acids. Post-translational modification further contributes to the structural complexity of the cyclic peptide moiety [108, 109, 141]. This combination of features reduces the susceptibility of these compounds to ubiquitous peptidases [144]. Multiple bacterial genera including *Streptomyces*, *Actinoplanes* [89], *Bacillus* [131] and *Pseudomonas* [139] have been used for the isolation of CPLs. Here, we set out to identify and isolate five new and one known CLP from *Pseudomonas* sp. FhG100052, to carry out a comprehensive structural analysis of each molecule, and to determine their antimicrobial activity against a panel of clinical isolates representing common human pathogens. Finally, compound profiling was completed by a set of economically relevant phytopathogens.

5.3 Materials and Methods

5.3.1 Isolation *Pseudomonas* sp. FhG100052

Pseudomonas sp. strain FhG100052 was purified following a bioprospecting campaign at Lake Stechlin (Brandenburg, Germany) and identified by 16S rRNA gene sequence analysis. The strain was cultured in basal salt medium (BSM) comprising 4,25 g * L⁻¹ K₂ HPO₄*3 H₂ O; 1,0 g * L⁻¹ NaH₂ HPO₄*H₂ O; 2,0 g * L⁻¹ NaH₄ Cl; 0,20g * L⁻¹ MgSO₄*7 H₂ O; 0,012g * L⁻¹ FeSO₄*7 H₂ O; 0,003g * L⁻¹ MnSO₄*H₂ O; 0,003 g * L⁻¹ ZnSO₄*7 H₂ O; 0,001 g * L⁻¹ CoCl*6 H₂ O; 0,1 g * L⁻¹ N[CH₂ COOH]₃; 0,5 g * L⁻¹ yeast extract and 4,0 g * L⁻¹ C(glycerin). After 72h at 28 C°and 180 rpm, the cultivation was stopped by cooling the bacterial cultures as well as the medium controls to -50 C°.

5.3.2 Bioactivity assessment

Primary screening Primary anti-bacterial screening and fractionation of extracts obtained from FhG100052 was carried out as described before (subsubsection 3.2.3).

Minimum inhibitory concentration of pure compounds Determination of the minimum inhibitory concentrations (MIC) of compounds purified from microbial extracts against the test panel was carried according to Figure S13. An overnight culture was adjusted to McFarland 1 and subsequently diluted 1:600 in Mueller Hinton II medium (BD). The assay volume of 100 µL was distributed to each test well and incubation was done at 37°C, 180rpm, and 95% rH for 18h. Antimicrobial activity was evaluated by turbidity measurements. All compounds were tested in triplicates. Inhibition was calculated using Equation 2.

Anti-fungal MICs were determined based on the EUCAST recommendations [46] [45]. Briefly, spore solutions of *A.flavus* ATCC9170 and *S.tritici* MUCL45407 were diluted to $\sim 10^5$ spores/mL. *Aspergillus* dilution was done in MHII and incubation was carried out at 37 °C and 180 rpm for 48h. Spores of *Septoria* were diluted in YM medium (yeast extract 4 g * L⁻¹, malt extract 4 g * L⁻¹, Sucrose 4 g * L⁻¹) and incubation was carried out at 24 °C for 48h. Nystatin, Amphotericin B and Tebuconazole were used as positive controls. Cell growth was determined by application of 50 μ L BacTiter-Glo™(Promega) and subsequent luminescence read out.

5.3.3 Screening for chemical novelty

The UHPLC-QTOF-MS/MS data from the methanolic *Pseudomonas* sp. FhG100052 extract were analyzed using molecular networking to allow the variable dereplication of known and unknown metabolites. First, the raw data (*.d files) was converted to plain text files (*.mgf) containing MS/MS peak lists using MSConvert (ProteoWizard package [31]), wherein each parent ion is represented by a list of fragment mass/intensity value pairs. These were computed with the molecular networking algorithm by converting each precursor ion into a vector in an n -dimensional space, with n being the number of fragment ions. The vectors were compared pairwise using dot product calculations based on the cosine between the two (= cosine similarity). Each vector pair was thus assigned a cosine similarity score of 0.0-1.0, where 0.0 represents an angle of 90° between the two vectors and 1.0 either 0° or 180°. Perpendicular parent ion vectors share no fragments and are entirely different, whereas a cosine score close to 1.0 indicates shared fragments, thus a putative structural relationship between the compared precursor ions. Pairs with a cosine similarity score greater than 0.7 were defined as related and were thus connected in the network. Additionally, ions need a minimum of six shared fragments (tolerance Δ ppm 0.05) with at least one partner ion to be included in the final network. In silico fragmented compounds [7] of a commercial database (AntiBase 2017 [91]) as well as our in-house pure compound MS/MS database were included in the network as reference substances to narrow down the molecular formula and highlight compounds of interest. CytoScape v3.4.0 was used to visualize the data as a network consisting of nodes and edges, wherein each node represents a parent ion and its color reflects the sample from which the MS/MS file was obtained. The thickness of the edges represents the cosine similarity score between nodes (thick edges indicate high similarity) and size of the nodes the abundance of the respective parent ion.

5.3.4 Genome sequencing and biosynthetic gene cluster annotation

Extraction of genomic high molecular weight DNA Genomic high molecular weight DNA of strain FhG100052 was retrieved following a protocol based on Sambrook and Russell [73] and adapted by Josh Quick (Nanopore WGS Consortium). Briefly, a cell pallet obtained from an overnight culture of FhG100052 was resuspended in 100 μ L sterile PBS, supplemented with 7.5 mL TLB buffer and incubated at 37 °C in a 50mL falcon tube. After 1h the temperature was increased to 50 °C for 3h. Each hour during incubation the solution was mixed by slowly inverting the tube. Afterwards, the viscous cell lysate was distributed to PhaseLock tubes by splitting it into 10 x 750 μ L aliquots. The same volume of saturated phenol (Thermo Fisher) was added to each tube to denature enzymes and other proteins. The tubes were incubated on a rotor (\sim 10 min, 40rpm) until a very fine emulsion was formed. After centrifugation (3320 x g, 10min) the aqueous phase (\sim 600 μ L) was evenly distributed into 10 new Phaselock tubes. To each tube 600 μ L saturated phenol and 600 μ L chloroform were added. Tubes were incubated and centrifugated as described before. Again, the aqueous phase was removed (\sim 6 mL) and collected in a 50 mL falcon tube. Genomic DNA was precipitated by addition of 4 mL 5M ammonium acetate. After addition of 30 mL ethanol (100 %, -20 °C) and inversion, the DNA was submerged in a fresh falcon tube containing 70 % ethanol by a hook made from a melted glass capillary. The DNA formed an opaque pellet, which was ultimately transferred into an Eppendorf tube. Again, 70% ethanol was added and subsequently removed by centrifugation and evaporation at 40 °C. Finally, 150 μ L EB buffer was added to the DNA and stored at 4 °C over night until a gelatinous consistency was observed.

DNA quantification was carried out with a 100fold diluted sample using the QunantIt PicoGreen Kit (Promega) according to the manufacturer's instructions. Quality control was done by puls field gel electrophoresis (CHEF-DR II®, Bio Rad) following the protocol of Bio-Rad Laboratories Inc [17] [18] [16]. The purified DNA was provided at a concentration of 50 ng/ μ L. Sequencing was carried out using Illumina NextSeq 500V2 by LGC genomics (Berlin, Germany). The quality control report (fastqc) was also provided by LGC. Finally, assembly, annotation and biosynthetic genecluster analysis was carried out by the *in house* bioinformatics platform: Adapter trimmed sequences (SeqPrep) were assembled using masurca 3.2.8 [170] . Genome annotation and gene cluster prediction was done on the basis of Gendb [115] and the antiSMASH (v. 4.2.0)\Arts genome mining tools [5]. The nucleotide sequence of the biosynthetic gene cluster *steABC* was deposited at GenBank (MT080808).

5.3.5 Optimization of production

Media variation Strain FhG100052 was cultured in BSM supplemented with 10 mmol * L⁻¹, 50 mmol * L⁻¹, 100 mmol* L⁻¹ or 150 mmol * L⁻¹ of glucose, glycerol, mannitol or arabinose, in order to maximize CPL production, thus facilitating the isolation process. In these experiments, the optical density was used as a proxy for culture growth, and was monitored over time. After incubation for 72 h (180 rpm, 28°C) the cultures were lyophilized, extracted with methanol and analyzed by UHPLC-QTOF-MS. CLP levels were determined by extracted ion chromatogram peak integration. All experiments were carried out in triplicate. The harvesting time point was chosen based on a previous kinetic study monitoring cultivation parameters (pH and OD_{600nm}) as well as relative amount of target CLPs over time.

Gas exchange In addition to the media variation, the effect of increased mixing and oxygenation of the culture broth on CLP production was investigated. FhG100052 was cultured in baffled and regular 300mL Erlenmeyer flasks (culture volume = 50mL) in BSM supplemented with 100 mmol glycerol (72h, 180rpm, 28 °C). Similar to the media variation experiments, the cell density and the CLP levels were determined. All experiments were carried out in triplicate.

Incubation duration In order to determine the ideal cultivation duration in terms of maximal compound yield, the CLP production was monitored over time. A volume of 500mL culture broth (BSM + 100mM glycerin) was inoculated with FhG100052 (0.01 %, OD 1.0). An aliquot of 11 mL of culture broth was removed each day for one week: 1mL was used to determine pH and OD_{600nm}, while the remaining broth was utilized to determine he CLP levels via LC/MS.

5.3.6 Purification of compounds

To isolate the new CLPs, 36 L culture broth (BSM + 100 mmol *L⁻¹ glycerol) was inoculated with *Pseudomonas* sp. FhG100052. After three days of incubation at 28°C and 180 rpm, the culture broth was lyophilized and extracted with methanol. Extracts were concentrated *in vacuo* and fractionated by sequential elution over a XAD 16N column. Fractions containing target CLPs were merged. After liquid-liquid extraction with ethyl acetate, the concentration of the organic fraction was adjusted to 200 mg/mL in methanol. Further fractionation was carried out by preparative HPLC equipped with a Phenomenex Synergi 4u Fusion-RP 80A column and external Gilson fraction collector. After manual injection, separation

was achieved by eluting in a linear gradient increasing from from 25% acetonitrile (+0.1 % formic acid) to 95 % acetonitrile in 28 min. The fractions of interest were concentrated to 30 mg/mL (in methanol), injected into a semi-preparative HPLC system (Hewlett Packard Model 1100 with NUCLEODUR C18 Gravity-SB column and Gilson fraction collector) and eluted in a gradient increasing from 25 - 95 % acetonitrile (+0.1 % formic acid) in 24 min. Final purity was achieved by UHPLC microfractionation (Agilent 1290 Infinity LC with ACQUITY UPLC BEH C18 Column, 130Å, 1.7 µm, 2.1 mm X 100 mm column) in a gradient of 57 - 70 % acetonitrile (+0.1% formic acid) in 18 min.

5.3.7 Structure elucidation using NMR

NMR spectra were recorded on a Bruker AVANCE 500 spectrometer operating at a proton frequency of 500.30 MHz and a ¹³C-carbon frequency of 125.82 MHz. The instrument was equipped with a 5 mm TCI cryo probe head. All experiments were carried out using samples of 2 - 3 mg compound dissolved in 600 µL *d*₆-acetone at 300 K. ¹H-chemical shifts were referenced to the solvent signals (1H: 2.04 ppm, 13C: 29.80 ppm). Two-dimensional homonuclear experiments (DQF-COSY, TOCSY and ROESY), were performed with a spectral width of 10 ppm. Spectra were recorded with 1024 increments in t1 and 4096 complex data points in t2. For each t1 value 8 transients were averaged. For multiplicity edited-HSQC spectra 1024 increments with 2048 complex data points in t2 were collected using a sweep width of 10 ppm in the proton and 90 ppm in the carbon dimension. For each t1 value 8 transients were averaged. The HMBC spectra were acquired with a sweep width of 10 ppm in the proton and 200 ppm in the carbon dimension using a defocusing delay of 62 ms (optimized for coupling constants of 8 Hz). A total of 32 transients were averaged for each of 1024 increments in t1, and 4096 complex points in t2 were recorded.

5.3.8 Determination of absolute configuration

The absolute configuration of the isolated CLPs was partially determined by derivatization using Marfey's reagent [15]. An amino acid LC/MS reference library was constructed to allow retention time comparison with amino acids featuring unknown stereo chemistry. First, aqueous standard solutions (50 µM) of at least one enantiomer of each amino acid present in the isolated group of CLPs were prepared. Derivatisation of standard amino acids was conducted using 70 µM FDVA (1-fluoro-2,4-dinitrophenyl-5-L-valinamid) in acetone. The reaction

was carried out in three separated vials in a ratio of 1:1:0.4 (aminoacid standard : FDVA : NaHCO₃)

Vial 1: L-Glu, L-Asp, L-Ser, L-Thr, L-Val, L-Ile

50 µL of each amino acid (50 µM aqueous solution) was mixed with 120 µL 1M NaHCO₃ and 300 µL 70 µM FDVA in acetone. The solution was stirred for 3h at 40 °C. The reaction was stopped by neutralization with 120 µL 1M HCL.

Vial 2: D-Thr, D-Glu, D-Leu, D-*allo*-Ile, D-Gln

50 µL of each amino acid (50 µM aqueous solution) was mixed with 100 µL 1M NaHCO₃ and 250 µL 70 µM FDVA in acetone. The solution was stirred for 3h at 40 °C. The reaction was stopped by neutralization with 100 µL 1M HCL.

Vial 3: L-*allo*-Thr and L-*allo*-Ile

50 µL of each amino acid (50 µM aqueous solution) was mixed with 40 µL 1M NaHCO₃ and 100 µL 70 µM FDVA in acetone. The solution was stirred for 3h at 40 °C. The reaction was stopped by neutralization with 40 µL 1M HCL.

A 100 µL aliquot of each of the three reaction mixtures was individually evaporated to dryness. Residues were re-dissolved in 15 µL DMSO and submitted to UHPLC-HRMS (Agilent 1290 Infinity®LC in line with maXis II™). Total hydrolysis of CLPs was carried out by dissolving 200 µg of the peptide in 6M deuteriohydrochloric acid (DCl in D₂O) and stirring for 7h at 160 °C. DCl was evaporated under nitrogen flow followed by lyophilization. The resulting amino acid mixture (~ 1.5 µmol of each) was dissolved in 300 µL water, mixed with 300 µL FDVA and 100 µL 1M NaHCO₃ and stirred for 2h at 40 °C. The reaction was quenched with 100 µL 1M HCl and subsequently freeze dried. Dry residue was re-dissolved in 15 µL DMSO and submitted to UHPLC-HRMS.

5.3.9 Optical rotation

The optical rotation of all isolated CLPs was determined using a polarimeter (P3000, Knüss Optronic Germany) and a 100mm flow through glass cell (PRG-100-DT, Knüss Optronic Germany). The measurements were carried out in LC-MS grade methanol at 23.7 °C and a wave length of 589 nm. Values were calculated using:

$$[\alpha]_{\lambda}^T = \frac{\alpha_{measured}}{l \cdot c} \quad (5)$$

$\alpha_{measured}$ = the experimental rotation angle [°]

l = the path length [dm]

c = compound concentration in solution [g/mL]

T = temperature [C°]

λ = wavelength [nm]

5.4 Results

5.4.1 Bioactivity of crude extract

The growth inhibitory effects of FhG100052 extract against the test strains were determined by micro broth dilution assays. Essentially, the extract strongly inhibited the growth of *C. albicans* FH2173 at all tested concentrations. Additionally, the extract exhibited strong growth inhibitory effects against *E.coli* DH5 α , but showed no effect against other Gram-positive (incl.Mycobacteria) or Gram-negative bacteria. According to the results of the primary screening, 5 μ L of FhG100052 crude extract was microfractionated and retested against *E.coli* DH5 α and *C.albicans* FH2173 using a luminescence-based cell viability assay. Fractions 103 to 110 showed unambiguously strong inhibition of cell viability (see Figure S12, bottom). Analysis of the UHPLC-QTOF-MS/MS raw data (Figure S12, top) revealed that these fractions were mainly composed of two double-protonated pseudo-molecular ions: 698.4210 $[M + 2H]^{2+}$ and 705.4290 $[M + 2H]^{2+}$. The single protonated ions of the same compounds (1395.8340 $[M + H]^+$ and 1409.8495 $[M + H]^+$) were also detected, but at much lower intensities (data not shown). Furthermore, 705.4290 $[M + 2H]^{2+}$ eluted after 14.7 min and 15.1 min, forming two distinct peaks in the chromatogram.

5.4.2 Molecular Network cluster analysis

The UHPLC-QTOF-MS/MS data files of the FhG100052 crude extract were investigated in detail by molecular networking analysis. The pseudomolecular ions present in the bioactive fractions were found in two distinct clusters (see Figure 23). The double-protonated ions 698.4210 $[M + 2H]^{2+}$ and 705.4290 $[M + 2H]^{2+}$ were found in cluster A, whereas the single protonated ions 1395.8340 $[M + H]^+$ and 1409.8495 $[M + H]^+$) were found in cluster B. Spectral library and literature search identified 1409.8495 $[M + H]^+$ as Tensin [128]. At this point, the identity of the Tensin was confirmed by comparing the amino acid sequence of the peptide moiety (inferred from the MS/MS fragmentation signatures) to the published structures (data not shown). Interestingly, three additional precursor ions were found in cluster B (1381.8187 $[M + 2H]^{2+}$, 1423.8667 $[M + 2H]^{2+}$, 1437.8794

$[M + 2H]^{2+}$. The nodes representing the double-charged variant of these putative new molecules were located in cluster A, tightly connected to the identified CLP Tensin.

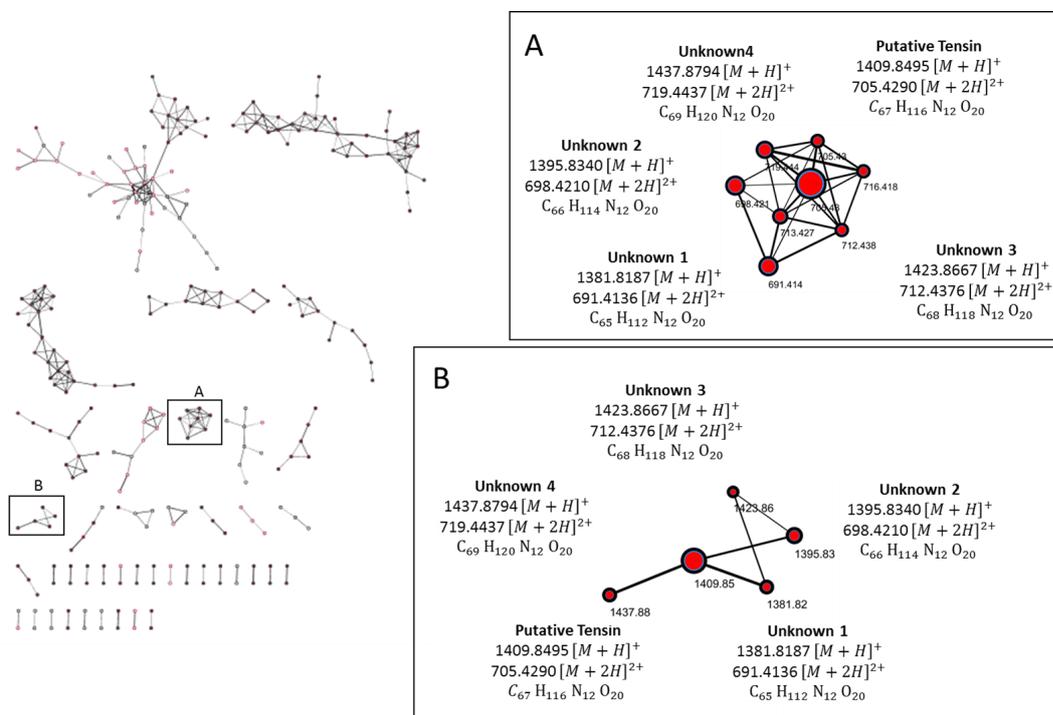


Figure 23: Molecular Network cluster analysis of FhG100052.

Masses of ions in cluster B differ in multiples of 14.0152 ± 0.0005 Da. This characteristic mass shift among derivatives is indicative for a series of $\pm CH_2$ analogues. Consequently, the mass relationship between the double-protonated ions in cluster A were multiples of 7.0076 Da and molecular formulas of the unknown compounds were predicted accordingly:

unknown 1: 1381.8187 $[M + H]^+$; 691.4136 $[M + 2H]^{2+}$; $C_{65}H_{112}N_{12}O_{20}$

unknown 2: 1395.8340 $[M + H]^+$; 698.4210 $[M + 2H]^{2+}$; $C_{66}H_{114}N_{12}O_{20}$

unknown 3: 1423.8667 $[M + H]^+$; 712.4376 $[M + 2H]^{2+}$; $C_{68}H_{118}N_{12}O_{20}$

unknown 4: 1437.8794 $[M + H]^+$; 719.4437 $[M + 2H]^{2+}$; $C_{69}H_{120}N_{12}O_{20}$

5.4.3 Optimization of production

The integrated peak area of the extraction ion chromatogram (EIC) of Tensin 1409.8495 $[M_2 + H]^+$ as a proxy for compound production was compared across the different cultivation conditions (see Figure 24). EICs of the other CLPs

were detected close to the background and thereby not quantified. The highest amount of Tensin was detected in fermentations carried out in carbon free BSM supplemented with 50 - 150 mM glycerol. Lowest abundance of Tensin was found in arabinose cultivates. Regardless of the supplemented carbon source, the compound production was higher in 50 mM than in 10 mM, but not significantly different comparing 50-150mM samples.

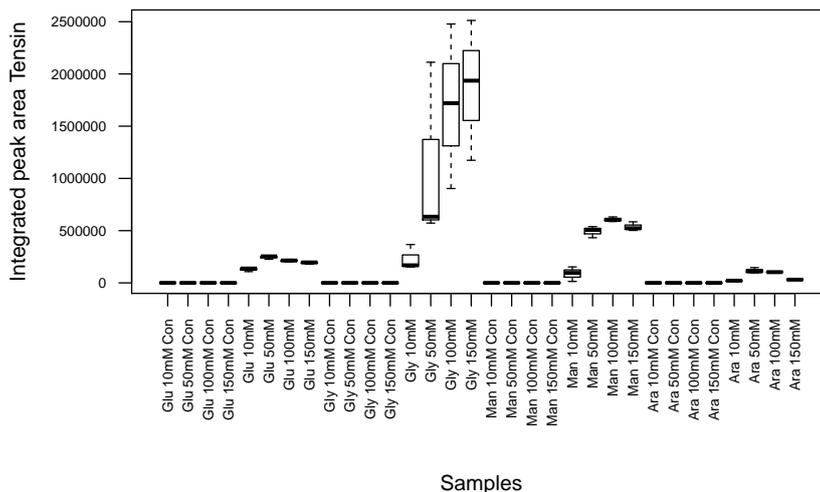


Figure 24: Integrated peak areas of EICs 1409.8495 $[M_2 + H]^+$ detected in FhG100052 extracts cultured BSM broth supplemented with different carbon sources (Glu = glucose; Gly = glycerol; Man = mannitol; Ara = arabinose)

Interestingly, the metabolite detection does not correlate with the OD measurements when comparing the different carbon sources. Bacteria cultured in glucose, glycerol or mannitol supplemented medium exhibited similar cell densities (see Figure S15), while the detected CLP titer varied drastically.

Gas exchange The influence of increased mixing, thus oxygenation, on cell density and CLP production was investigated by cultivation of FhG100052 in baffled and regular flasks. Cultures reared in baffled flasks exhibited 1.4 fold higher mean cell densities compared to the cultures in regular flasks (mean $OD_{600regular} = 5.7$; mean $OD_{600baffled} = 7.9$; see Figure S17). Similar to the previous observations, the detected amount of CLPs and the measured cell density did not correlate linearly. CLP detection in the oxygenated cultures exceeded the regular flask

cultures by 13fold (as determined by integrated peak area of EIC 1409.8495 \pm 0.05 at 15.0 \pm 0.2 min) (s.Figure S16).

Incubation duration Ultimately the CLP production was monitored over time to determine the ideal harvesting time point with maximum CLP titer in the culture broth. Essentially, the CLP production plateaued after 72h and remained constant for the remaining time investigated (s. Figure S18)

5.4.4 Compound purification and structure elucidation

Compound purification started from 36L FhG100052 culture broth obtained by fermentation under optimized conditions (subsubsection 5.4.3). Isolation was carried out by XAD and various HPLC as well as UPLC C₁₈ fractionations to ultimately yield 0.5 - 6.5 mg of pure compound (Stechlisin B2 = 6.5mg; Stechlisin C1 = 0.8mg; Stechlisin C3 = 2.3mg; Tensin = 4.3mg; Stechlisin D3 = 3.3mg; Stechlisin E1 = 0.5mg; Stechlisin E2 = 2.9mg; Stechlisin F = 2.2mg). The structures of the CLPs were elucidated using HRMS/MS and NMR analyses. Structure elucidation of compounds (\geq 1mg) was achieved by 1-dimensional (1D, ¹H and ¹³C-spectra) and 2-dimensional (2D, multiplicity edited-HSQC, HMBC and TOCSY spectra) nuclear magnetic resonance (NMR) (Figure 25). The sequential assignment of the amino acids was based on correlations obtained in the ROESY spectrum (Figure S11). The NMR data of the major fermentation product, which was assigned to the structure of Tensin, as well as five new CLPs are summarized in Table 7 Table 8 Table 9 Table 10 and Table 11.

Table 7: ^1H - chemical shifts of Stechlisins. NMR experiments were performed on an AVANCE500 instrument in d_6 -acetone at 300 K. Chemical shifts are referenced to the solvent signals (1H: 2.04 ppm). Abbreviations: (b) = broad signal; (n.a.) = not assigned.

FA	Stechlisin B2	Stechlisin C3	Stechlisin D3	Tensin	Stechlisin E2	Stechlisin F
1'	175.58	175.67	175.61	175.62	175.55	~175.6 (b)
2'	44.14	44.18	44.16	44.17	44.14	44.17
3'	70.20	70.20	70.21	70.21	70.38	70.20
4'	38.47	38.55	38.54	38.55	38.57	38.57
5'	25.97	26.32	26.33	26.33	26.32	26.34
6'	32.53	~30.3 (a)	~30.3 (a)	~30.3 (a)	~30.3 (a)	~30.0-30.3(a) (a)
7'	23.25	~30.0 (a)	~30.0(a)	~30.0 (a)	~30.0 (a)	~30.0-30.3(a)
8'	14.30	32.55	32.54	32.55	32.53	~30.0-30.3(a)
9'		23.29	23.29	23.29	23.29	~30.0-30.3(a)
10'		14.35	14.35	14.35	14.35	32.59
11'						23.31
12'						14.35
Leu1						
α	~55.8 (b)	~55.9 (b)	~55.9 (b)	55.83	~55.9 (b)	~55.8 (b)
β	40.62	40.61	40.60	40.62	40.62	40.62
γ	25.41	25.39	25.42	25.39	25.38	25.42
δ	22.94	~23.0 (b)	22.96	23.00	22.81	~23.00 (b)
δ'	22.12	22.09	22.11	22.1	22.22	22.09
C'	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Asp2/ Glu2						
α	~54.1 (b)	~54.2 (b)	~54.2 (b)	~54.2 (b)	~57.4 (b)	~54.2 (b)
β	~35.8 (b)	~35.9 (b)	~35.9 (b)	~35.9 (b)	~26.4 (b)	~35.9 (b)
γ	n.a.	n.a.	n.a.	n.a.	~31.2 (b)	n.a.
C'	n.a.	n.a.	n.a.	n.a.	~174.4 (b) ~176.1 (b)	n.a.
Thr3						
α	~60.3 (b)	~60.5 (b)	~60.4 (b)	~60.4 (b)	~60.7 (b)	~60.4 (b)
β	70.30	~70.2 (b)	70.32	70.28	~70.3 (b)	70.29
γ	17.88	18.03	17.8	17.92	18.02	17.91
C'	n.a.	n.a.	n.a.	n.a.	172.56	n.a.
Leu4						
α	~56.0 (b)	~56.1 (b)	~56.0 (b)	~56.1 (b)	~56.3 (b)	~56.0 (b)
β	~41.4 (b)	41.25	~41.5 (b)	41.34	41.25	41.35
γ	25.31	25.31	25.26	25.32	25.31	25.32
δ	23.07	23.06	23.11	23.05	22.97	23.06
δ'	22.58	22.55	22.52	22.62	22.77	22.61
C'	~174.0 (b)	174.18	n.a.	174.08	173.88	174.06

Table 8: ¹H- chemical shifts of Stechlisins continued

FA	Stechlisin B2	Stechlisin C3	Stechlisin D3	Tensin	Stechlisin E2	Stechlisin F
Ser6						
NH	7.32	7.31	7.27	7.32	7.33	7.32
α	4.49	4.45	4.50	4.48	4.49	4.48
β	4.04/3.79	4.03/3.82	4.04/3.77	4.04/3.80	4.03/3.81	4.04/3.80
Leu7						
NH	7.63	7.59	7.68	7.63	7.71	7.63
α	4.33	4.30	4.37	4.32	4.33	4.32
β	1.92/1.67	1.91/1.64	1.93/1.69	1.92/1.66	1.94/1.67	1.92/1.66
γ	1.85	1.86	1.85	1.86	1.86	1.85
δ	0.95	0.95	0.95	0.95	0.95	0.95
δ'	0.85	0.85	0.85	0.85	0.85	0.85
Gln8						
NH	8.03	8.01	8.04	8.03	8.09	8.03
α	4.31	4.32	4.26	4.3	4.25	4.3
β	2.03/1.94	2.04/1.93	2.00/1.94	2.03/1.93	2	2.02/1.93
γ	2.28	2.27	2.30	2.28	2.28	2.27
NH ₂	7.03/6.31	7.02/6.28	7.04/6.32	7.03/6.29	7.05/6.32	7.02/6.28
Leu9						
NH	7.93	7.88	7.99	7.91	7.93	7.91
α	4.21	4.21	4.20	4.21	4.18	4.21
β	1.68	1.68	1.72/1.64	1.69/1.65	1.67	1.67
γ	1.80	1.79	1.83	1.79	1.80	1.80
δ	0.92	0.92	0.92	0.92	0.92	0.92
δ'	0.87	0.88	0.88	0.87	0.87	0.87
Ile10/ Val10						
NH	7.39	7.34	7.76	7.38	7.37	7.39
α	4.03	4.03	4.22	4.04	3.99	4.03
β	2.04	2.25	1.82/1.62	2.03	2.04	2.04
β-Me	0.94	-	1.77	0.94	0.94	0.94
γ	1.55/1.23	0.96	0.94	1.55/1.22	1.55/1.23	1.55/1.22
γ'		0.95				
δ	0.88	-	0.86	0.88	0.89	0.88
Glu11						
NH	-6.84 (b)	-6.95 (b)	-6.77 (b)	-6.87 (b)	6.83	-6.86 (b)
α	4.61	4.59	4.6	4.59	4.60	4.60
β	1.97/1.82	1.97/1.84	1.94/1.83	1.97/1.83	1.94/1.83	1.97/1.82
γ	2.35/2.30	2.36/2.30	2.33/2.24	2.36/2.30	2.36/2.29	2.35/2.29

Table 9: ^{13}C - chemical shifts of Stechlisins. NMR experiments were performed on an AVANCE500 instrument in d_6 -acetone at 300 K. Chemical shifts are referenced to the solvent signals (^{13}C : 29.8 ppm). Abbreviations: (a) = ^{13}C -signal below solvent signal; (b) = broad signal; (n.a.) = not assigned.

FA	Stechlisin B2	Stechlisin C3	Stechlisin D3	Tensin	Stechlisin E2	Stechlisin F
1'	175.58	175.67	175.61	175.62	175.55	~175.6 (b)
2'	44.14	44.18	44.16	44.17	44.14	44.17
3'	70.20	70.20	70.21	70.21	70.38	70.20
4'	38.47	38.55	38.54	38.55	38.57	38.57
5'	25.97	26.32	26.33	26.33	26.32	26.34
6'	32.53	~30.3 (a)	~30.3 (a)	~30.3 (a)	~30.3 (a)	~30.0- 30.3(a) (a)
7'	23.25	~30.0 (a)	~30.0(a)	~30.0 (a)	~30.0 (a)	~30.0- 30.3(a)
8'	14.30	32.55	32.54	32.55	32.53	~30.0- 30.3(a)
9'		23.29	23.29	23.29	23.29	~30.0- 30.3(a)
10'		14.35	14.35	14.35	14.35	32.59
11'						23.31
12'						14.35
Leu1						
α	~55.8 (b)	~55.9 (b)	~55.9 (b)	55.83	~55.9 (b)	~55.8 (b)
β	40.62	40.61	40.60	40.62	40.62	40.62
γ	25.41	25.39	25.42	25.39	25.38	25.42
δ	22.94	~23.0 (b)	22.96	23.00	22.81	~23.00 (b)
δ'	22.12	22.09	22.11	22.1	22.22	22.09
C'	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Asp2/ Glu2						
α	~54.1 (b)	~54.2 (b)	~54.2 (b)	~54.2 (b)	~57.4 (b)	~54.2 (b)
β	~35.8 (b)	~35.9 (b)	~35.9 (b)	~35.9 (b)	~26.4 (b)	~35.9 (b)
γ	n.a.	n.a.	n.a.	n.a.	~31.2 (b)	n.a.
C'	n.a.	n.a.	n.a.	n.a.	~174.4 (b) ~176.1 (b)	n.a.
Thr3						
α	~60.3 (b)	~60.5 (b)	~60.4 (b)	~60.4 (b)	~60.7 (b)	~60.4 (b)
β	70.30	~70.2 (b)	70.32	70.28	~70.3 (b)	70.29
γ	17.88	18.03	17.8	17.92	18.02	17.91
C'	n.a.	n.a.	n.a.	n.a.	172.56	n.a.
Leu4						
α	~56.0 (b)	~56.1 (b)	~56.0 (b)	~56.1 (b)	~56.3 (b)	~56.0 (b)
β	~41.4 (b)	41.25	~41.5 (b)	41.34	41.25	41.35
γ	25.31	25.31	25.26	25.32	25.31	25.32
δ	23.07	23.06	23.11	23.05	22.97	23.06
δ'	22.58	22.55	22.52	22.62	22.77	22.61
C'	~174.0 (b)	174.18	n.a.	174.08	173.88	174.06

Table 10: ^{13}C - chemical shifts of Stechlisins continued

FA	Stechlisin B2	Stechlisin C3	Stechlisin D3	Tensin	Stechlisin E2	Stechlisin F
Leu5						
α	54.02	53.99	54.03	54.01	53.99	54.01
β	40.86	40.74	40.85	40.81	40.82	40.82
γ	25.38	25.43	25.38	25.42	25.34	25.39
δ	23.74	23.76	23.78	23.76	23.78	23.76
C'	173.11	173.15	n.a.	173.15	173.21	173.15
Ser6						
α	57.26	57.33	57.24	57.33	57.43	57.31
β	64.73	~64.6 (b)	~64.9 (b)	~64.7 (b)	~65.0 (b)	~64.7 (b)
C'	172.23	172.3	172.11	172.24	172.19	172.24
Leu7						
α	53.46	53.61	53.31	~53.5 (b)	53.36	~53.4 (b)
β	39.79	39.84	39.70	39.81	39.7	~39.8 (b)
γ	25.25	25.26	25.26	25.26	25.25	25.25
δ	24.06	24.00	24.13	24.06	24.11	24.06
δ'	21.46	21.43	21.42	21.45	21.43	21.45
C'	173.81	173.82	n.a.	173.81	173.88	~173.8 (b)
Gln8						
α	54.41	54.25	~54.7 (b)	54.39	~54.5 (b)	~54.4 (b)
β	28.93	~29.0 (b)	29.17	28.93	~28.8 (b)	~28.9 (b)
γ	32.48	32.48	32.60	32.48	~32.5 (b)	~32.5 (b)
δ	175.52	175.51	175.53	175.49	n.a.	~175.5 (b)
C'	n.a.	n.a.	n.a.	n.a.	~174.0 (b)	173.79
Leu9						
α	53.91	53.99	53.86	53.91	53.91	53.91
β	41.03	41.17	40.85	41.06	40.96	41.05
γ	25.52	25.52	25.52 (d)	25.52	25.49	25.52
δ	23.53	23.52	23.53	23.53	23.52	23.53
δ'	21.40	21.44	21.39	21.41	21.36	21.41
C'	n.a.	n.a.	n.a.	n.a.	173.01	173.07

Table 11: ^{13}C - chemical shifts of Stechlisins continued

FA	Stechlisin B2	Stechlisin C3	Stechlisin D3	Tensin	Stechlisin E2	Stechlisin F
Ile10/ Val10						
α	59.95	~60.6 (b)	53.75	~59.9 (b)	~60.1 (b)	~59.9 (b)
β	36.52	30.55	40.69	36.58	36.48	36.55
β -Me	16.06		25.53 (d)	16.06	16.05	16.06
γ	26.22	19.92	23.47	26.21	26.32	26.21
γ'		19.17				
δ	11.34		21.01	11.34	11.32	11.34
C'	171.9	n.a.	n.a.	171.95	172	n.a.
Glu11						
α	53.09	53.09	53.16	53.15	~53.3 (b)	53.14
β	~29.2 (b)	~29.0 (b)	~29.7 (b)	29.17	~29.5 (a)	29.17
γ	30.45	30.45	30.41	30.47	30.53	30.46
δ	173.42	173.49	n.a.	173.43	173.37	173.39
C'	169.84	169.98	169.53	169.87	169.75	169.85
C'	171.9	n.a.	n.a.	171.95	172	n.a.

Comparison of the 2D datasets of all isolated compounds strongly suggested the same relative stereochemistry in all derivatives. Therefore, the absolute configuration was determined by total hydrolysis of the major fermentation product Tensin, followed by chemical derivatization with with N_α -(2,4-dinitro-5-fluorophenyl)-L-valinamide (Marfey reagent) and LC/MS comparison to reference substrates (s. Figure S19 and Figure S20). Based on these results, the new compounds resemble members of the Amphisin group as they are all constructed of a 3-hydroxy fatty acid linked to the N-terminus of an undecapeptide moiety. Interestingly, Stechlisin B2, Tensin and Stechlisin F share the same amino acid sequence but carry different fatty acids chains (ranging from C_8 to C_{12} in length, Figure 25, red). Stechlisin E2 is distinct from all other described members of the Amphisin group in the sense that it possesses a glutamic instead of an aspartic acid at the second position of the peptide chain (Figure 25, green) . Further distinctions of the new compounds and Tensin are located at the tenth position of the peptide moiety (Figure 25, blue): At this position either leucine (Stechlisin D3), valine (Stechlisin C3) or isoleucine (all other cases) is incorporated.

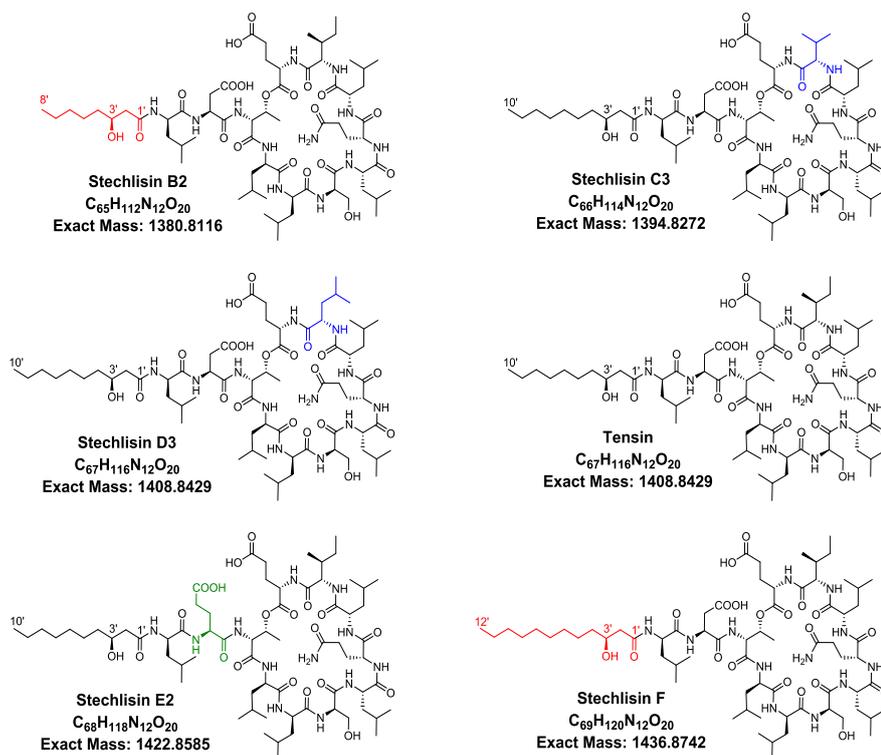


Figure 25: Structure and absolute configuration of isolated Stechlisins and Tensin. Differences to Tensin are highlighted. Red: differences in the fatty acid side chain. Blue: differences in the cyclic peptide part. Green: differences in the linear amino acid sequence.

Apart from the six isolated and fully characterized compounds (Figure 25), 28 other minor derivatives were observed in the extract of FhG100052 (Table S1). For eleven of these compounds, structure propositions were made based on manual annotation of their MS/MS fragmentation (Table 12, Figure 26), leaving open the aspects of stereochemistry as well as the identity of the amino acid residues in positions 4,5,7,9 and 10, where isobaric Leu/Ile (Xle) residues were assigned. For the remaining 17 compounds, only molecular formula determination was possible, as low intensity MS/MS signals did not allow total sequence determination.

Table 12: Proposed structure of Stechlisins A-F: Purified and fully characterized compounds are bold. For the remaining compounds, structures were proposed on the basis of their MS/MS fragmentation pattern. Based on MS, no distinction between leucin and isoleucin was possible (= Xle). Other amino acids written in conventional three letter code: Abu = aminobutyric acid; Leu = leucin; Asp = aspartic acid; Thr = threonine; Ile = Isoleucin; Ser = serine; Val = valine; Glu = glutamic acid; Glu-O-CH₃ = methyl ester of glutamic acid; Gln = glutamine)

Name	Molecular formula	Amino acid sequence
Stechlisin A	$C_{64}H_{110}N_{12}O_{20}$	3-OH-C ₈ -acid-Leu-Asp- <i>allo</i> -Thr-Xle-Xle-Ser-Xle-Gln-Xle-Val-Glu
Stechlisin B1	$C_{65}H_{112}N_{12}O_{20}$	3-OH-C ₁₀ -acid-Leu-Asp- <i>allo</i> -Thr-Xle-Xle-Ser-Xle-Gln-Abu-Xle-Glu
Stechlisin B2	$C_{65}H_{112}N_{12}O_{20}$	3-OH-C₈-acid-D-Leu-D-Asp-D-<i>allo</i>-Thr-D-Leu-D-Leu-D-Ser-L-Leu-D-Gln-L-Leu-L-Ile-L-Glu
Stechlisin B3	$C_{65}H_{112}N_{12}O_{20}$	3-OH-C ₁₀ -acid-Leu-Asp- <i>allo</i> -Thr-Xle-Xle-Ser-Xle-Val-Val-Xle-Glu
Stechlisin B4	$C_{65}H_{112}N_{12}O_{20}$	3-OH-C ₁₀ -acid-Leu-Asp-Ser-Xle-Xle-Ser-Xle-Gln-Xle-Val-Glu
Stechlisin C1	$C_{66}H_{114}N_{12}O_{20}$	3-OH-C ₀₈ -acid-Leu-Glu- <i>allo</i> -Thr-Xle-Xle-Ser-Xle-Gln-Xle-Xle-Glu
Stechlisin C2	$C_{66}H_{114}N_{12}O_{20}$	3-OH-C ₁₀ -acid-Leu-Asp- <i>allo</i> -Thr-Xle-Xle-Ser-Xle-Gln-Val-Xle-Glu
Stechlisin C3	$C_{66}H_{114}N_{12}O_{20}$	3-OH-C₁₀-acid-D-Leu-D-Asp-D-<i>allo</i>-Thr-D-Leu-D-Leu-D-Ser-L-Leu-D-Gln-L-Leu-L-Val-L-Glu
Stechlisin C4	$C_{66}H_{114}N_{12}O_{20}$	3-OH-C ₁₀ -acid-Leu-Asp- <i>allo</i> -Thr-Xle-Val-Ser-Xle-Gln-Xle-Xle-Glu
Stechlisin C5	$C_{66}H_{114}N_{12}O_{20}$	3-OH-C ₁₀ -acid-Leu-Asp- <i>allo</i> -Thr-Val-Xle-Ser-Xle-Gln-Xle-Xle-Glu
Stechlisin D1	$C_{67}H_{116}N_{12}O_{20}$	3-OH-C ₁₀ -acid-Leu-Asp- <i>allo</i> -Thr-Xle-Xle-Ser-Xle-Gln-Xle-Xle-Glu
Tensin	$C_{67}H_{116}N_{12}O_{20}$	3-OH-C₁₀-acid-D-Leu-D-Asp-D-<i>allo</i>-Thr-D-Leu-D-Leu-D-Ser-L-Leu-D-Gln-L-Leu-L-Ile-L-Glu
Stechlisin D3	$C_{67}H_{116}N_{12}O_{20}$	3-OH-C₁₀-acid-D-Leu-D-Asp-D-<i>allo</i>-Thr-D-Leu-D-Leu-D-Ser-L-Leu-D-Gln-L-Leu-L-Leu-L-Glu
Stechlisin E1	$C_{68}H_{118}N_{12}O_{20}$	3-OH-C ₁₀ -acid-Leu-Asp- <i>allo</i> -Thr-Xle-Xle-Ser-Xle-Gln-Xle-Xle-Glu-O-CH ₃
Stechlisin E2	$C_{68}H_{118}N_{12}O_{20}$	3-OH-C₁₀-acid-D-Leu-D-Glu-D-<i>allo</i>-Thr-D-Leu-D-Leu-D-Ser-L-Leu-D-Gln-L-Leu-L-Ile-L-Glu
Stechlisin E3	$C_{68}H_{118}N_{12}O_{20}$	3-OH-C ₁₁ -acid-Leu-Asp- <i>allo</i> -Thr-Xle-Xle-Ser-Xle-Gln-Xle-Xle-Glu
Stechlisin F	$C_{69}H_{120}N_{12}O_{20}$	3-OH-C₁₂-acid-D-Leu-D-Asp-D-<i>allo</i>-Thr-D-Leu-D-Leu-D-Ser-L-Leu-D-Gln-L-Leu-L-Ile-L-Glu

Alterations within the Stechlisin peptide moieties are mostly due to exchange of the aliphatic amino acids leucin, isoleucin and valine among each other at one or more position of the sequence. For instance, Stechlisin C4 carries a (iso-)leucin (Xle) at position four and subsequently a valine at position five, whereas Stechlisin C5 contains Val4 and Xle5.

Further distinguishing features were observed in three Stechlisin variants: Stechlisin B2 carries an aminobutyric acid at position nine of the sequence. No other detected compound or elsewhere described member of the amphisin group contains Abu. In addition, Stechlisins E1 and E3 exhibited peculiar features: Stechlisin E1 is the only CLP possessing a methyl ester side chain at the C-terminal acid and E3 the only variant showing a unique fatty acid moiety composed of an odd number of carbon atoms (C₁₁) (s. also Figure 26).

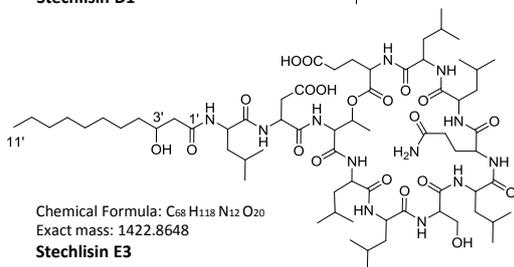
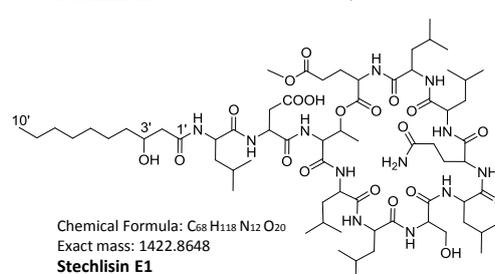
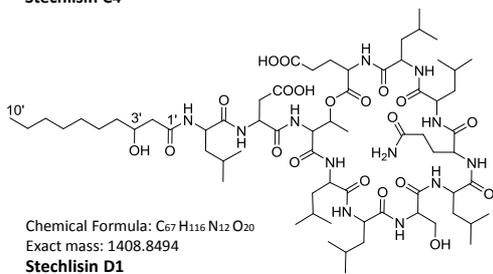
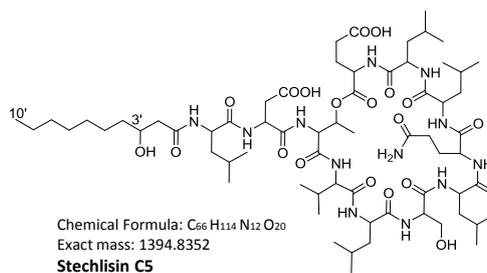
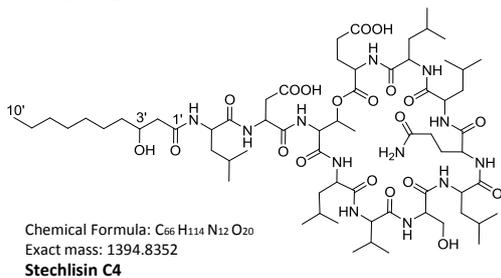
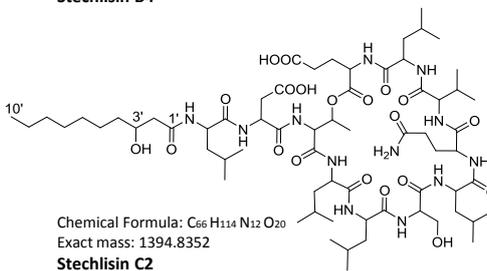
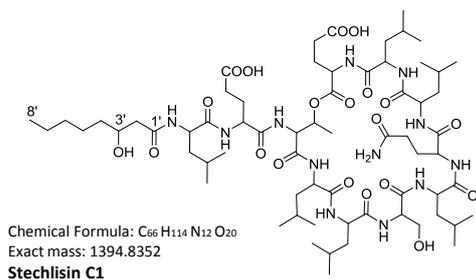
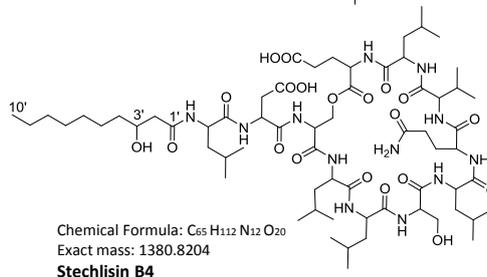
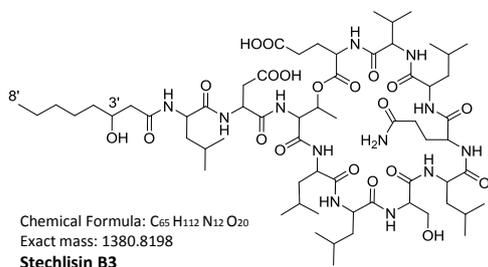
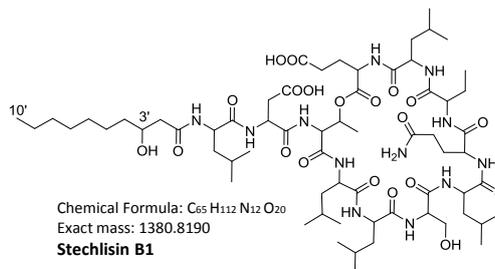
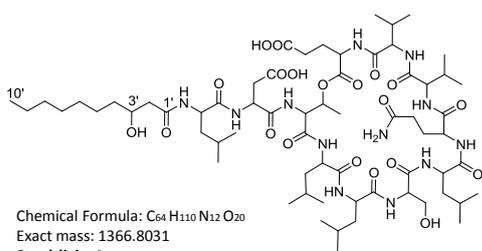


Figure 26: Stechlisins: Proposed structures based on MS/MS

5.4.5 Optical rotation

Results from optical rotation determination are summarized in Table 13. Essentially all analyzed compounds exhibit negative angles ranging from - 25 °, as determined for the smallest compound Stechlisin B2, to values of -37.5 to 53.7 ° for the larger compounds.

Table 13: Specific rotation values of isolated Stechlisins

Compound	c [mg/mL]	$\alpha_{measured}$	$[\alpha]_D^{23.6}$
Stechlisin B2	0.8	-0.02	-25,0 (c 0.08 MeOH)
Stechlisin C3	1,5	-0.07	-46,6 (c 0.15 MeOH)
Stechlisin D3	1,6	-0.06	-37,5 (c 0.16 MeOH)
Tensin	0,7	-0.03	-42,8 (c 0.07 MeOH)
Stechlisin E2	1,5	-0.08	-53,7 (c 0.15 MeOH)
Stechlisin F	0,6	-0.03	-50,0 (c 0.06 MeOH)

5.4.6 Genome analysis and biosynthetic gene cluster identification

To get insights into the biosynthesis of the Stechlisins, *Pseudomonas* sp. FhG100052 was genome sequenced. Illumina sequencing of genomic DNA of strain FhG100052 yielded 13104566 total raw reads, thus 6552283 read pairs. Sequence length ranged from 20 - 150 bp with a GC-content of 59 % and an average Phred score [47] of 34. After alignment and assembly, the draft genome (153 contigs, predicted genome size of 6.1 Mbp) was analysed using the publicly available antiSMASH pipeline. Most interestingly one complete NRPS type gene cluster was annotated on contig 6, ranging from nucleotide 12017 - 89530. The cluster is constructed of 11 modules organized in three core genes (6.5, 13, and 18 kbp in length, s. Figure 28, Figure S14). Similar to other *Pseudomonas* CLP gene clusters, the first 10 modules are each composed of a condensation (C-), one adenylation (A-) and one peptide carrier protein (PCP-) domain or thiolation (T-) domain, while the last domain harbors two additional termination or thio-esterase (TE-) domains. In close proximity of the core genes, a *luxR* type transcription factor [25] and homologs of *macA* and *macB* [117] coding for membrane fusion and transporter proteins were found (s. Figure S14).

sequences biosynthesized by *steABC* matches the peptide moiety of Tensin and the prediction of the pHMM algorithm [20]. Hence the genes, in the following termed *steA*, *steB* and *steC*, are proposed to code for the enzymatic machinery responsible for the biosynthesis of a range of CLPs, including Tensin. Based on the molecular network analysis and manual annotation of MS/MS signals, at least five major and numerous minor compounds (s. Figure 23, Table S1) are likely to be synthesized by *steABC* under the used cultivation conditions. No other gene cluster with 11 A-domains was found in the genome of FhG100052.

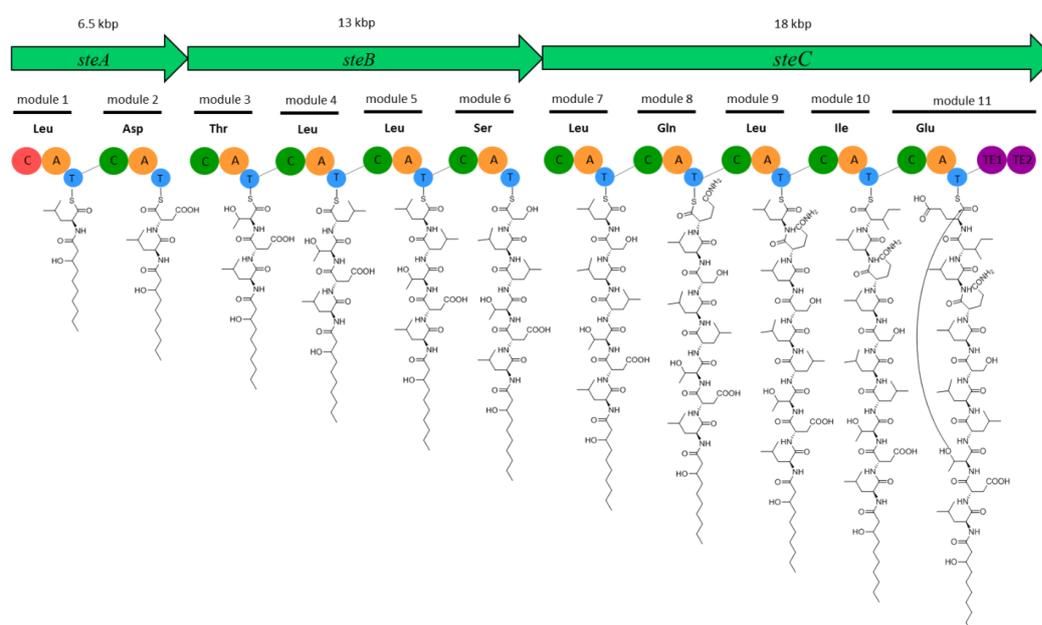


Figure 28: Biosynthetic gene cluster and proposed biosynthesis of Stechlisin and Tensin: In total the nrps type gene cluster has a length of 37.5 kbp. The 11 modules, responsible for peptide elongation, are organized the three core genes *steA*, *steB* and *steC*. The first 10 modules are each composed of a condensation (C-), one adenylation (A-) and one peptide carrier protein or thiolation (T-) domain, while the last domain harbors two additional termination or thio-esterase (TE-) domains.

5.4.7 Minimum inhibitory concentrations

In addition to the six fully characterized CLPs (Figure 25), Stechlisin C1 was subject to antimicrobial MIC determination. C1 was purified in low amounts, restricting spectral analysis but still allowed antimicrobial testing (0.5 mg). Pure substances were tested against a bacterial and a smaller fungal panel. Interestingly, all tested substances showed growth inhibitory effects in the *lux* assay against

E. coli DH5 α (primary screening section 4.3.2) with MICs ranging from 8 to 2 $\mu\text{g}/\text{mL}$.

Most remarkably, a clear structure-activity relationship (SAR) could be observed when the purified compounds were screened against the Gram-negative bacterium *Moraxella catarrhalis* FH6810: While Stechlisin B2 did not inhibit the growth of the test strain at any tested concentration, Tensin showed moderate (32 $\mu\text{g}/\text{mL}$) and Stechlisin F pronounced activity (4 $\mu\text{g}/\text{mL}$, Table 14). All other substances did not inhibit the growth of any bacterial or fungal indicator strain at the concentrations tested (>128 $\mu\text{g}/\text{mL}$) (s. Table 14)

Table 14: Antibacterial and antifungal minimum inhibitory concentrations of purified Stechlisins. Values are given in $\mu\text{g}/\text{mL}$. Highest tested concentration was $128 \mu\text{g}/\text{mL}$. Tensin and Stechlisin F inhibited the growth of *M. catarrhalis* at 32 and $4 \mu\text{g}/\text{mL}$. All other compounds exhibited no growth inhibitory effect towards the tested pathogens. MHC = Mueller Hinton II broth supplemented with bicarbonate ($3.7 \text{ g} * \text{L}^{-1}$); Ca²⁺ = Mueller Hinton II broth supplemented with $50 \text{ mg} * \text{L}^{-1} \text{ Ca}^{+2}$.

	MIC [$\mu\text{g}/\text{mL}$]						
	Stechlisin B2	Stechlisin C1	Stechlisin C3	Stechlisin D3	Tensin	Stechlisin E2	Stechlisin F
<i>E. coli</i> ATCC 25922 wild type	>128	>128	>128	>128	>128	>128	>128
<i>E. coli</i> ATCC 25922 ΔTolC	>128	>128	>128	>128	>128	>128	>128
<i>E. coli</i> ATCC 25922 wild type MHC	>128	>128	>128	>128	>128	>128	>128
<i>E. coli</i> DH5a [pFU166]	8	4	4	4	2	4	2-1
<i>M. catarrhalis</i> FH 6810	>128	>128	>128	>128	32	>128	4
<i>M. smegmatis</i> ATCC607	>128	>128	>128	>128	>128	>128	>128
<i>S. aureus</i> ATCC 25923	>128	>128	>128	>128	>128	>128	>128
<i>S. aureus</i> ATCC 25923 + Ca ²⁺	>128	>128	>128	>128	>128	>128	>128
<i>B. subtilis</i> DSM 10	>128	>128	>128	>128	>128	>128	>128
<i>C. albicans</i> FH2173	>128	>128	>128	>128	>128	>128	>128
<i>S. tritici</i> MUCL45407	>128	>128	>128	>128	>128	>128	>128
<i>A. flavus</i> ATCC9170	>128	>128	>128	>128	>128	>128	>128

5.5 Discussion

This section represents the application of the established metabolomics platform (section 3) to uncharacterized microorganisms (section 4). To this end, the overall goal was to investigate the metabolite output of prioritized environmental strains in detail.

The presented metabolomics analysis, driven by the search for chemical novelty, substantially facilitated the identification of the caustive agent responsible for the primary observed bioactivity of *Pseudomonas* sp. FhG100052. In particular the molecular networking analysis provided fundamental value to the project by automatic dereplication of the known CLP Tensin and the identification of five structurally related, yet unknown CLPs in proximity of the bioactive extract fractions. This finding caught the research enthusiasm and thereby directed the focus to carefully investigate the raw data. Manual annotation of fragmentation patterns finally led to the discovery of 33 previously not described compounds. These compounds, in the following termed Stechlisins, are members of the Amphisins and closely related to Tensin. For the six most abundant compounds isolation, structure elucidation and biological profiling was carried out. In the following, results regarding bioactivity and biosynthesis of the Stechlisins are discussed.

5.5.1 Bioactivity

First of all, the MICs determined for the isolated CLPs might explain the growth inhibitory effect against *E.coli* DH5 α observed in the primary screening. Cell free supernatant, μ -fractionated MeOH extracts and finally the purified compounds showed pronounced growth inhibitory effects against the test strain. In contrast, the observed *C.albicans* activity of the crude and fractionated MeOH extract is not reflected in the activity profile of the Stechlisins. This discrepancy might be due to exceptionally high CLP concentration within the extract, hence the results of the primary screening might represent a rather unspecific growth inhibitory effect.

These findings emphasize a general challenge in NP research: Extracts are comprised of a mixture of various substances at dramatically different concentrations and potencies. In this context, it is important to realize that almost each substance becomes unspecifically toxic at high concentrations, hence producing a positive assay read out. Discrimination between specific and unspecific effects during primary screening might come at the price of insensitivity. In this study, a trade off in the direction of false positive instead of false negative was chosen.

Internal standards or calibration lines might help to estimate the abundance of a certain compound during LCMS-UV analysis, but disrespect effects like ion suppression, the limit of detection or detector saturation. Moreover, the nature of NP discovery programs restricts the availability of authentic standards as usually unknown compounds are the target of interest. In the end, it is hardly feasible to normalize the substance concentration within an assay, without the effort of compound isolation.

Of course, this instance also bears chances: Compounds present at low abundances might not be recognized by strictly bioactivity guided extract characterization approaches. In that sense, these NPs might have been overlooked in the past - even in already extensively studied genera such as *Streptomyces* spp. or *Pseudomonas* spp. Additional screening for chemical novelty potentially helps to appreciate these compounds. For example, not the already characterized and vastly abundant Tensin, but the lower abundant Stechlisin F exhibited the strongest growth inhibitory effect against the Gram-negative bacterium *Moraxella catarrhalis* FH6810 when test in adjusted concentrations.

Structure activity relationship Interestingly, the SAR study indicates that the length of the side chain seems to determine the degree of growth inhibition given an amino acid sequence of D-Leu-D-Asp-D-*allo*-Thr-D-Leu-D-Leu-D-Ser-L-Leu-D-Gln-L-Leu-L-Ile-L-Glu. Compounds with a different sequence of amino acids in the peptide moiety exhibit no bioactivity against the tested pathogens. Stechlisin B2 possesses the shortest side chain (3-OH-C₈-acid) and exhibited no growth inhibitory effects, whereas Tensin is constructed of an 3-OH-C₁₀-acid and showed moderate activity (32 µg/mL). Stechlisin F also carries the same amino acid sequence as B2 and Tensin, but exhibits an even longer side chain (3-OH-C₁₂-acid) and also increased bioactivity (4 µg/mL). The length of the lipophilic side chain heavily influences the overall amphipathicity of the molecule. The ecological functions of CLPs are a consequence of the amphipathicity, giving the compounds emulsifier and surfactant properties and might additionally determine antibiotic potency or toxicity [11] [85]. Thereby the length of the incorporated fatty acid might confer functional dualism, combining antimicrobial and biosurfactant properties [82]. Likewise, the amino acid composition seems to be a critical delimiter of the observed antimicrobial activity. Stechlisins C3, D3 and E2 contain the same C₁₀ side chain, but each differ at one position in amino acid sequence compared to Tensin (C3: L-Val11 instead of L-Ile11; D3: L-Leu11 instead of L-Ile11; E2: D-Glu2 instead of D-Asp2). Apparently, even the small modifications, for instance an interchange of valine, leucine or isoleucine mediate in- or decreased antimicrobial activity.

Specificity The antimicrobial activity of other biosurfactant CLPs is often attributed to their membrane destabilizing behavior [66]. It is believed that the fatty acid chain may insert into the lipid bilayer of cell membranes thereby inflicting membrane perturbation. Besides, the cyclic peptide part of the molecules might support this activity, as polar residues (like aspartic or glutamic acid) are thought to extend towards the extra cellular aqueous medium, while hydrophobic residues reach into the center of the membrane [66]. Strikingly, the herein described group of Stechlisins exhibit a specific and narrow anti Gram-negative activity. While *E.coli* DH5 α is strongly inhibited, other *E.coli* strains were unaffected by the compounds in the tested sensitivity range. A potential reason for the increased susceptibility could be the composition of the outer membrane (OM) of DH5 α . It was shown that the outer lipooligosaccharides of DH5 α are anchored in the phospholipid membrane by lipid A and do possess inner and outer core structures, but lack the outermost polysaccharide chain (rough-type LPS)[32]. Usually, this specific polysaccharide region is thought to contribute to protecting the structural integrity of the OM and restricting diffusion of hydrophobic molecules over the membrane. The absence of O-polysaccharides renders the OM hydrophobic, thus being a possible cause for the facilitated insertion of the fatty acid moiety of the herein characterized CLPs. In that sense, it is not surprising, that some *Moraxella catarrhalis* strains are considered to possess a 'semi-rough' OM exhibiting either only one repeating type of O-polysaccharide [160] or none at all [157]. In both cases increased OM permeability towards hydrophobic agents such as macrolides was observed [157]. It is hypothesized here that the absence of the outer polysaccharide chain within the protective LPS layer is responsible for the increased susceptibility of *E.coli* DH5 α and *M. catarrhalis* against the investigated Stechlisins.

5.5.2 Biosynthesis

The biosynthetic gene cluster (BGC) responsible for the assembly of Tensin was identified by A-domain comparison (amino acid sequence) of *steABC* found in the genome of FhG100052 and 11 representative CLPs (Table 12). This analysis indicated a substrate activation specificity of *steABC*, which fits the structure of the major metabolite Tensin. However, not only one compound, but a group of 5 major and numerous minor compounds were discovered. It is worth to mention that *steABC* is the only NRPS type gene cluster with the correct number of A-domains found in the genome of FhG100052. Thereby the CLP biosynthesis in FhG100052 follows the typical NRPS collinearity between number of A-domains and amino acids. In close proximity of the core genes, a luxR-type transcription factor, which is also highly conserved among *Pseudomonas* spp. CLPs gene clusters [25] [126], was annotated. Downstream of *steABC*, homologs of *macA*

and *macB* [117] coding for membrane fusion and transporter proteins were found. These or equivalent genes are frequently encountered in the flanking region of CLP gene cluster in *Pseudomonas* spp. and are likely to be involved in the transport of CLPs outside of the cell [42] [100]. In the present BGC, no specific epimerization domain was found, suggesting external racemase conversion of amino acids as already described for other members of the Amphisin group, such as Arthrofactin [142] or Anikasin [56].

Analogous to the aminoacyl tRNA complex translating the genetic code into peptides at the ribosomes, the NRPS use a specificity determining 'Stachelhaus' code [35] [109] for correct peptide production. For the ribosomal system, energy consuming proof reading mechanisms have been discovered and helped to understand the astonishing fidelity of the translation process. In contrast, no similarly efficient editing mechanisms are known for NRPS [62]. The lack of proof reading might explain the encountered inhomogeneous Tensin/Stechlisin mixture observed in FhG 100052 extracts. In fact, residue variation within the Stechlisin group was mainly observed in structurally related amino acids.

Most of the observed structural diversity is explained by interchange of the aliphatic (Ile, Leu, Val) or the acidic (Asp, Gln) residues among each other at one or more position in the product CLP. A-domain inspecificity with respect to these amino acids is a common phenomenon as for instance postulated for Gramicidin S [2] or Surfactin [52] [22] [30] biosynthesis. Both examples demonstrate adenylation domain affinity to Ile, Leu and Val resulting in a group of analogous CLPs. Interestingly, the name giving member of described group of CLP, Amphisin, differs from Tensin only at the eleventh position of the peptide sequence (Asp11 instead of Gln11). As shown in this study, *steA* A2 can incorporate either acidic residue, suggesting that *steC* A11 is not only capable of Gln activation but might also show affinity to aspartic acid. Therefore both literature known molecules, Amphisin and Tensin, might be biosynthesized by the herein investigated BGC, although Amphisin was not observed in this study.

The Stechlisins vary with respect to the length of the N-terminal fatty acid side chain (C₈ - C₁₂). Similar observations were made, for instance, during the discovery and isolation of the calcium dependent antibiotic Daptomycin. The clinically applied drug was first described as a minor signal within a set of CLP exhibiting different fatty acid moieties (A21978C complex). Feeding of decanoic acid significantly optimized production of the desired C₁₀ Daptomycin [39]. Studies on Surfactin biosynthesis reported enzymes involved in 3-hydroxy fatty acid activation and transport to the initial C-domain. Further, the study demonstrated their external (of the NRPS BGC) location [85] on the genome. Primary fatty

acid metabolism might yield the substrate for these activating fatty acyl CoA ligases [85] [11], implying a relationship between fatty acid structure provided to CPL production and available carbon source [43] during fermentation. Matching these observations, no genetic information corresponding to the biosynthesis of the lipophilic side chain could be found within or in proximity of the *steABC* cluster (s. Figure S14. Accordingly, cultivation of FhG100052 in basal salt medium supplemented with different carbon sources produced similar cell densities, but varying Tensin titer (Figure S16 Figure S17).

6 Perspective

Classical antibiotic therapy, based on small molecules, is a main pillar of our health care system and involved in numerous medicinal procedures. In the future, the necessity of antibiotic treatment is likely to become even more substantial for human health. To keep up in the Red Queen's race, efficient methodologies for novel antibiotic discovery are needed. The presented metabolomics approaches might help to identify already known compounds and allow value estimations early in the discovery process. The task at hand is challenging and can not be tackled by individual research facilities. Public data bases such as the Global Natural Products Social Molecular Networking initiative shares knowledge globally and might help to spend research funds and effort more effectively. First steps are made, but along way lays ahead.

The microbiome analysis of the different bacterial communities obtained from Lake Stechlin revealed a promising composition of potential NP producers. However, cultivation dependent approaches such as the presented study, will always suffer from a great discrepancy of presence and culturability of a particular strains [148]. In the future, bioinformatic approaches might help to identify biosynthetic genes (gene cluster) of interest within metagenome data sets and reliably predict resulting structures. Advances in artificial intelligence algorithms might be able to foresee the bioactivity and even the mode of action of a certain structure before the wet lab work started. Prioritized or modified gene clusters might then be produced by heterologous expression in a easily cultivable strain.

Yet, the presented study yielded a group of previously uncharacterized bioactive derivative structures in the already extensively studied genus *Pseudomonas* spp. In this context it would be of greatest interest to study the remaining 28 compounds for which either only a molecular formula or a structure based on MS/MS fragmentation could be assigned. Further experiments should include a comprehensive structure elucidation and bioactivity assessment of purified compounds. As observed for Stechlisin F, the variants of greatest intensity might not be the most interesting ones.

Certainly, solutions to present and future challenges in this field require intensified knowledge exchange as well as continuous dedication and creativity.

References

- [1] Generic. 2005. DOI: 10.1351/pac200577010007. URL: <https://www.degruyter.com/view/j/pac.2005.77.issue-1/pac200577010007/pac200577010007.xml>.
- [2] Kjell Aarstad, Trine-Lise Zimmer, and Søren G. Laland. “The Fidelity of Gramicidin S Synthetase”. In: *European Journal of Biochemistry* 112.2 (1980), pp. 335–338. ISSN: 0014-2956. DOI: 10.1111/j.1432-1033.1980.tb07209.x. URL: <https://febs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1432-1033.1980.tb07209.x>.
- [3] Srinivas Adapa, Peter Hubtr, and Walter Keller-Schierlein. “Stoffwechselprodukte von Mikroorganismen. 216. Mitteilung. Isolierung, Strukturaufklärung und Synthese von Ferrioxamin H”. In: *Helvetica Chimica Acta* 65.6 (1982), pp. 1818–1824. ISSN: 1522-2675. DOI: 10.1002/hlca.19820650617. URL: <https://doi.org/10.1002/hlca.19820650617>.
- [4] Food Administration and Drug. “New Drugs at FDA: CDER’s New Molecular Entities and New Therapeutic Biological Products.” In: (2018). URL: <https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugInnovation/%20default.htm>.
- [5] Mohammad Alanjary et al. “The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery”. In: *Nucleic acids research* 45.W1 (2017), W42–W48. ISSN: 1362-4962 0305-1048. DOI: 10.1093/nar/gkx360. URL: <https://www.ncbi.nlm.nih.gov/pubmed/28472505%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5570205/>.
- [6] Pierre-Marie Allard et al. “Integration of Molecular Networking and In-Silico MS/MS Fragmentation for Natural Products Dereplication”. In: *Analytical Chemistry* 88.6 (2016), pp. 3317–3323. ISSN: 0003-2700. DOI: 10.1021/acs.analchem.5b04804. URL: <http://dx.doi.org/10.1021/acs.analchem.5b04804>.
- [7] Felicity Allen, Russ Greiner, and David Wishart. “Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification”. In: *Metabolomics* 11.1 (2015), pp. 98–110. ISSN: 1573-3890. DOI: 10.1007/s11306-014-0676-4. URL: <https://doi.org/10.1007/s11306-014-0676-4>.
- [8] Bianca Audrain et al. “Role of bacterial volatile compounds in bacterial biology”. In: *FEMS Microbiology Reviews* 39.2 (2015), pp. 222–233. ISSN:

- 0168-6445. DOI: 10.1093/femsre/fuu013. URL: <http://dx.doi.org/10.1093/femsre/fuu013>.
- [9] F. Azam et al. “Bacteria-organic matter coupling and its significance for oceanic carbon cycling”. In: *Microbial Ecology* 28.2 (1994), pp. 167–179. ISSN: 1432-184X. DOI: 10.1007/bf00166806. URL: <https://doi.org/10.1007/BF00166806>.
- [10] Stephen Baker. “A return to the pre-antimicrobial era?” In: *Science* 347.6226 (2015), p. 1064. DOI: 10.1126/science.aaa2868. URL: <http://science.sciencemag.org/content/347/6226/1064.abstract>.
- [11] Richard H. Baltz, Vivian Miao, and Stephen K. Wrigley. “Natural products to drugs: daptomycin and related lipopeptide antibiotics”. In: *Natural product reports* 22.6 (2005), pp. 717–741. ISSN: 0265-0568. DOI: 10.1039/b416648p. URL: <http://europepmc.org/abstract/MED/16311632%20https://doi.org/10.1039/b416648p>.
- [12] Robert A. van den Berg et al. “Centering, scaling, and transformations: improving the biological information content of metabolomics data”. In: *BMC Genomics* 7.1 (2006), p. 142. ISSN: 1471-2164. DOI: 10.1186/1471-2164-7-142. URL: <http://dx.doi.org/10.1186/1471-2164-7-142>.
- [13] Roberto G. S. Berlinck. “Natural guanidine derivatives”. In: *Natural Product Reports* 19.5 (2002), pp. 617–649. ISSN: 0265-0568. DOI: 10.1039/A901981B. URL: <http://dx.doi.org/10.1039/A901981B>.
- [14] S. Bernardini et al. “Natural products for human health: an historical overview of the drug discovery approaches”. In: *Natural Product Research* 32.16 (2018), pp. 1926–1950. ISSN: 1478-6419. DOI: 10.1080/14786419.2017.1356838. URL: <https://doi.org/10.1080/14786419.2017.1356838>.
- [15] R. Bhushan and H. Brückner. “Marfey’s reagent for chiral amino acid analysis: A review”. In: *Amino Acids* 27.3 (2004), pp. 231–247. ISSN: 1438-2199. DOI: 10.1007/s00726-004-0118-0. URL: <https://doi.org/10.1007/s00726-004-0118-0>.
- [16] BioRad. “Pulsed Field Electrophoresis Bulletin 6228”. In: ().
- [17] BioRad. “Pulsed Field Gel Electrophoresis Bulletin 6224”. In: ().
- [18] BioRad. “Pulsed Field Gel Electrophoresis Bulletin 6226”. In: ().
- [19] Jessica M. A. Blair et al. “Molecular mechanisms of antibiotic resistance”. In: *Nature Reviews Microbiology* 13 (2014), p. 42. DOI: 10.1038/nrmicro3380. URL: <https://doi.org/10.1038/nrmicro3380>.

- [20] Kai Blin et al. “antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification”. In: *Nucleic Acids Research* 45.W1 (2017), W36–W41. ISSN: 0305-1048. DOI: 10.1093/nar/gkx319. URL: <https://doi.org/10.1093/nar/gkx319>.
- [21] Simone Bonazzi et al. “Anguinomycins and Derivatives: Total Syntheses, Modeling, and Biological Evaluation of the Inhibition of Nucleocytoplasmic Transport”. In: *Journal of the American Chemical Society* 132.4 (2010), pp. 1432–1442. ISSN: 0002-7863. DOI: 10.1021/ja9097093. URL: <https://doi.org/10.1021/ja9097093>.
- [22] Jean-Marc Bonmatin, Olivier Laprevote, and Françoise Peypoux. “Diversity Among Microbial Cyclic Lipopeptides: Iturins and Surfactins. Activity-Structure Relationships to Design New Bioactive Agents”. In: *Combinatorial Chemistry High Throughput Screening* 6.6 (2003), pp. 541–556. DOI: 10.2174/138620703106298716. URL: <https://www.ingentaconnect.com/content/ben/cchts/2003/00000006/00000006/art00006%20https://doi.org/10.2174/138620703106298716>.
- [23] Chiara Borsetto et al. “Microbial community drivers of PK/NRP gene diversity in selected global soils”. In: *Microbiome* 7.1 (2019), p. 78. ISSN: 2049-2618. DOI: 10.1186/s40168-019-0692-8. URL: <https://doi.org/10.1186/s40168-019-0692-8>.
- [24] V. Braun. “Covalent lipoprotein from the outer membrane of *Escherichia coli*”. In: *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes* 415.3 (1975), pp. 335–377. ISSN: 0304-4157. DOI: [https://doi.org/10.1016/0304-4157\(75\)90013-1](https://doi.org/10.1016/0304-4157(75)90013-1). URL: <http://www.sciencedirect.com/science/article/pii/0304415775900131>.
- [25] I. de Bruijn and J. M. Raaijmakers. “Diversity and functional analysis of LuxR-type transcriptional regulators of cyclic lipopeptide biosynthesis in *Pseudomonas fluorescens*”. In: *Applied and environmental microbiology* 75.14 (2009), pp. 4753–4761. ISSN: 1098-5336 0099-2240. DOI: 10.1128/AEM.00575-09. URL: <https://www.ncbi.nlm.nih.gov/pubmed/19447950%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2708414/>.
- [26] J. T. Buckley and S. P. Howard. “The cytotoxic enterotoxin of *Aeromonas hydrophila* is aerolysin”. In: *Infection and immunity* 67.1 (1999), pp. 466–467. ISSN: 0019-9567 1098-5522. URL: <https://www.ncbi.nlm.nih.gov/pubmed/9925450%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC96339/>.

- [27] J. Thomas Buckley et al. “Purification and some properties of the hemolytic toxin aerolysin”. In: *Canadian Journal of Biochemistry* 59.6 (1981), pp. 430–435. ISSN: 0008-4018. DOI: [10.1139/o81-059](https://doi.org/10.1139/o81-059). URL: <https://doi.org/10.1139/o81-059>.
- [28] Mark S. Butler and Antony D. Buss. “Natural products — The future scaffolds for novel antibiotics?” In: *Biochemical Pharmacology* 71.7 (2006), pp. 919–929. ISSN: 0006-2952. DOI: <https://doi.org/10.1016/j.bcp.2005.10.012>. URL: <http://www.sciencedirect.com/science/article/pii/S0006295205006635>.
- [29] Christiam Camacho et al. “BLAST+: architecture and applications”. In: *BMC Bioinformatics* 10.1 (2009), p. 421. ISSN: 1471-2105. DOI: [10.1186/1471-2105-10-421](https://doi.org/10.1186/1471-2105-10-421). URL: <https://doi.org/10.1186/1471-2105-10-421>.
- [30] Gregory L. Challis, Jacques Ravel, and Craig A. Townsend. “Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains”. In: *Chemistry Biology* 7.3 (2000), pp. 211–224. ISSN: 1074-5521. DOI: [https://doi.org/10.1016/S1074-5521\(00\)00091-0](https://doi.org/10.1016/S1074-5521(00)00091-0). URL: <http://www.sciencedirect.com/science/article/pii/S1074552100000910>.
- [31] Matthew C. Chambers et al. “A cross-platform toolkit for mass spectrometry and proteomics”. In: *Nature Biotechnology* 30 (2012), p. 918. DOI: [10.1038/nbt.2377](https://doi.org/10.1038/nbt.2377)<https://www.nature.com/articles/nbt.2377#supplementary-information>. URL: <http://dx.doi.org/10.1038/nbt.2377>.
- [32] H. Chart et al. “An investigation into the pathogenic properties of *Escherichia coli* strains BLR, BL21, DH503b1 and EQ1”. In: *Journal of Applied Microbiology* 89.6 (2000), pp. 1048–1058. ISSN: 1364-5072. DOI: [10.1046/j.1365-2672.2000.01211.x](https://doi.org/10.1046/j.1365-2672.2000.01211.x). URL: <https://sfamjournals.onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2672.2000.01211.x>.
- [33] Wen Ming Chen, Chang Yi Lin, and Shih Yi Sheu. “Investigating antimicrobial activity in *Rheinheimera* sp. due to hydrogen peroxide generated by l-lysine oxidase activity”. In: *Enzyme and Microbial Technology* 46.6 (2010), pp. 487–493. ISSN: 0141-0229. DOI: <https://doi.org/10.1016/j.enzmictec.2010.01.006>. URL: <http://www.sciencedirect.com/science/article/pii/S0141022910000189>.

- [34] HyungJun Cho et al. “OutlierD: an R package for outlier detection using quantile regression on mass spectrometry data”. In: *Bioinformatics* 24.6 (2008), pp. 882–884. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btn012. URL: <https://doi.org/10.1093/bioinformatics/btn012>.
- [35] E. Conti et al. “Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S”. In: *The EMBO journal* 16.14 (1997), pp. 4174–4183. ISSN: 0261-4189 1460-2075. DOI: 10.1093/emboj/16.14.4174. URL: <https://www.ncbi.nlm.nih.gov/pubmed/9250661%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1170043/>.
- [36] Crystal A. Conway, Nwadiuto Esiobu, and Jose V. Lopez. “Co-cultures of *Pseudomonas aeruginosa* and *Roseobacter denitrificans* reveal shifts in gene expression levels compared to solo cultures”. In: *TheScientific-WorldJournal* 2012 (2012), pp. 120108–120108. ISSN: 1537-744X 2356-6140. DOI: 10.1100/2012/120108. URL: <https://www.ncbi.nlm.nih.gov/pubmed/22566756%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3330761/>.
- [37] D. De Corte et al. “Linkage between copepods and bacteria in the North Atlantic Ocean”. In: *Aquatic Microbial Ecology* 72.3 (2014), pp. 215–225. URL: <https://www.int-res.com/abstracts/ame/v72/n3/p215-225/>.
- [38] Daniele De Corte et al. “Metagenomic insights into zooplankton-associated bacterial communities”. In: *Environmental Microbiology* 20.2 (2018), pp. 492–505. ISSN: 1462-2912. DOI: 10.1111/1462-2920.13944. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1462-2920.13944>.
- [39] Morgan deBono et al. “A21978C, a complex of new acidic peptide antibiotics: Isolation, chemistry, and mass spectral structure elucidation”. In: *The Journal of antibiotics* 40 (1987), pp. 761–77. DOI: 10.7164/antibiotics.40.761.
- [40] Jingjing Deng et al. “p-Terphenyl O-03b2-glucuronides, DNA topoisomerase inhibitors from *Streptomyces* sp. LZ350394gdmAI”. In: *Bioorganic Medicinal Chemistry Letters* 24.5 (2014), pp. 1362–1365. ISSN: 0960-894X. DOI: <https://doi.org/10.1016/j.bmcl.2014.01.037>. URL: <http://www.sciencedirect.com/science/article/pii/S0960894X14000675>.
- [41] Orsolya Dobay et al. “Bicarbonate Inhibits Bacterial Growth and Biofilm Formation of Prevalent Cystic Fibrosis Pathogens”. In: *Frontiers in microbiology* 9 (2018), pp. 2245–2245. ISSN: 1664-302X. DOI: 10.3389/fmicb.2018.02245. URL: <https://pubmed.ncbi.nlm.nih.gov/30283433%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6157313/>.

- [42] Jean-Frédéric Dubern et al. “Genetic and functional characterization of the gene cluster directing the biosynthesis of putisolvin I and II in *Pseudomonas putida* strain PCL1445”. In: *Microbiology* 154.7 (2008), pp. 2070–2083. ISSN: 1350-0872. DOI: <https://doi.org/10.1099/mic.0.2008/016444-0>. URL: <https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.2008/016444-0>.
- [43] Gerrit Eggink, Gern N. M. Huijberts, and Pieter de Waard. “Biosynthesis of Poly(3-Hydroxyalkanoates) by *Pseudomonas Putida*”. In: *Frontiers of Polymers and Advanced Materials*. Ed. by Paras N. Prasad. Boston, MA: Springer US, 1994, pp. 607–612. ISBN: 978-1-4615-2447-2. DOI: 10.1007/978-1-4615-2447-2_57. URL: https://doi.org/10.1007/978-1-4615-2447-2_57.
- [44] Martina Erken, Carla Lutz, and Diane McDougald. “Interactions of *Vibrio* spp. with Zooplankton”. In: *Microbiology Spectrum* 3.3 (2015). DOI: doi: 10.1128/microbiolspec.VE-0003-2014. URL: <http://www.asmscience.org/content/journal/microbiolspec/10.1128/microbiolspec.VE-0003-2014>.
- [45] EUCAST. “Method for the determination of broth dilution minimum inhibitory concentrations of antifungal agent for yeasts”. In: v 7.3.1 valid from 15 January, 2017 (2017).
- [46] EUCAST. “Susceptibility testing in moulds”. In: v. 9.3 (2009).
- [47] Brent Ewing and Phil Green. “Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities”. In: *Genome Research* 8.3 (1998), pp. 186–194. DOI: 10.1101/gr.8.3.186. URL: <http://genome.cshlp.org/content/8/3/186.abstract>.
- [48] Attilio Fabbretti, Claudio O. Gualerzi, and Letizia Brandi. “How to cope with the quest for new antibiotics”. In: *FEBS Letters* 585.11 (2011), pp. 1673–1681. ISSN: 0014-5793. DOI: 10.1016/j.febslet.2011.04.029. URL: <https://febs.onlinelibrary.wiley.com/doi/abs/10.1016/j.febslet.2011.04.029>.
- [49] David Forner et al. “Chemical dereplication of marine actinomycetes by liquid chromatography–high resolution mass spectrometry profiling and statistical analysis”. In: *Analytica Chimica Acta* 805 (2013), pp. 70–79. ISSN: 0003-2670. DOI: <https://doi.org/10.1016/j.aca.2013.10.029>. URL: <http://www.sciencedirect.com/science/article/pii/S0003267013013378>.
- [50] Jeffrey L Fox. “Fraunhofer to min Sanofi microbial collection”. In: *Nature Biotechnology* 32.4 (2014), p. 305.

- [51] Stefania Galdiero et al. “Microbe-host interactions: structure and role of Gram-negative bacterial porins”. In: *Current protein peptide science* 13.8 (2012), pp. 843–854. ISSN: 1875-5550 1389-2037. DOI: 10.2174/138920312804871120. URL: <https://pubmed.ncbi.nlm.nih.gov/23305369/><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3706956/>.
- [52] Giuliano Galli et al. “Characterization of the surfactin synthetase multi-enzyme complex”. In: *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* 1205.1 (1994), pp. 19–28. ISSN: 0167-4838. DOI: [https://doi.org/10.1016/0167-4838\(94\)90087-6](https://doi.org/10.1016/0167-4838(94)90087-6). URL: <http://www.sciencedirect.com/science/article/pii/0167483894900876>.
- [53] Jeff Gerard et al. “Massetolides A–H, Antimycobacterial Cyclic Depsipeptides Produced by Two Pseudomonads Isolated from Marine Habitats”. In: *Journal of Natural Products* 60.3 (1997), pp. 223–229. ISSN: 0163-3864. DOI: 10.1021/np9606456. URL: <http://dx.doi.org/10.1021/np9606456>.
- [54] Rohit Ghai et al. “Key roles for freshwater Actinobacteria revealed by deep metagenomic sequencing”. In: *Molecular Ecology* 23.24 (2014), pp. 6073–6090. ISSN: 0962-1083. DOI: 10.1111/mec.12985. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.12985>.
- [55] F. Gibson and D. I. Magrath. “The isolation and characterization of a hydroxamic acid (aerobactin) formed by *Aerobacter aerogenes* 62-1”. In: *Biochimica et Biophysica Acta (BBA) - General Subjects* 192.2 (1969), pp. 175–184. ISSN: 0304-4165. DOI: [https://doi.org/10.1016/0304-4165\(69\)90353-5](https://doi.org/10.1016/0304-4165(69)90353-5). URL: <http://www.sciencedirect.com/science/article/pii/0304416569903535>.
- [56] Sebastian Götze et al. “Structure, Biosynthesis, and Biological Activity of the Cyclic Lipopeptide Anikasin”. In: *ACS Chemical Biology* 12.10 (2017), pp. 2498–2502. ISSN: 1554-8929. DOI: 10.1021/acscchembio.7b00589. URL: <http://dx.doi.org/10.1021/acscchembio.7b00589>.
- [57] Susanne Grabley et al. “Secondary metabolites by chemical screening. II. Amycins A and B, two novel niphimycin analogs isolated from a high producer strain of elaiophylin and nigericin”. In: *The Journal of antibiotics* 43 (1990), pp. 639–47. DOI: 10.7164/antibiotics.43.639.
- [58] P. A. Grigoriev, R. Schlegel, and U. Gräfe. “Cation selective ion channels formed by macrodiolide antibiotic elaiophylin in lipid bilayer membranes”. In: *Bioelectrochemistry* 54.1 (2001), pp. 11–15. ISSN: 1567-5394. DOI: [https://doi.org/10.1016/S0302-4598\(01\)00102-7](https://doi.org/10.1016/S0302-4598(01)00102-7). URL: <http://www.sciencedirect.com/science/article/pii/S0302459801001027>.

- [59] Harald Gross and Joyce E. Loper. “Genomics of secondary metabolite production by *Pseudomonas* spp”. In: *Natural Product Reports* 26.11 (2009), pp. 1408–1446. ISSN: 0265-0568. DOI: 10.1039/B817075B. URL: <http://dx.doi.org/10.1039/B817075B>.
- [60] Hans-Peter Grossart, Claudia Dziallas, and Kam W. Tang. “Bacterial diversity associated with freshwater zooplankton”. In: *Environmental Microbiology Reports* 1.1 (2009), pp. 50–55. ISSN: 1758-2229. DOI: 10.1111/j.1758-2229.2008.00003.x. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1758-2229.2008.00003.x>.
- [61] E. Grundmann. “Gerhard Domagk Ein Pathologe besiegt die bakteriellen Infektionskrankheiten”. In: *Der Pathologe* 22.4 (2001), pp. 241–251. ISSN: 0172-8113. DOI: 10.1007/s002920100469. URL: <https://doi.org/10.1007/s002920100469>.
- [62] Jan Grünewald and Mohamed A. Marahiel. “Chapter 21 - Nonribosomal Peptide Synthesis”. In: *Handbook of Biologically Active Peptides (Second Edition)*. Ed. by Abba J. Kastin. Boston: Academic Press, 2013, pp. 138–149. ISBN: 978-0-12-385095-9. DOI: <https://doi.org/10.1016/B978-0-12-385095-9.00021-X>. URL: <http://www.sciencedirect.com/science/article/pii/B978012385095900021X>.
- [63] Min Gui et al. “Natural Occurrence, Bioactivity and Biosynthesis of Elaiophylin Analogues”. In: *Molecules (Basel, Switzerland)* 24.21 (2019), p. 3840. ISSN: 1420-3049. DOI: 10.3390/molecules24213840. URL: <https://www.ncbi.nlm.nih.gov/pubmed/31731388><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6864862/>.
- [64] Martin W. Hahn et al. “Isolation of Novel Ultramicrobacteria Classified as Actinobacteria from Five Freshwater Habitats in Europe and Asia”. In: *Applied and Environmental Microbiology* 69.3 (2003), pp. 1442–1451. DOI: 10.1128/aem.69.3.1442-1451.2003. URL: <https://aem.asm.org/content/aem/69/3/1442.full.pdf>.
- [65] P Hammann, G Kretzschmar, and G Seibert. “Secondary metabolites by chemical screening 7, Elaiophylin derivatives and their biological activities”. In: *The Journal of Antibiotics* 43.11 (1990), pp. 1431–1440.
- [66] H. Heerklotz and J. Seelig. “Detergent-like action of the antibiotic peptide surfactin on lipid membranes”. In: *Biophysical Journal* 81.3 (2001), pp. 1547–1554. ISSN: 0006-3495 1542-0086. DOI: 10.1016/S0006-3495(01)75808-0. URL: <https://www.ncbi.nlm.nih.gov/pubmed/11509367><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1301632/>.

- [67] S P Howard and J T Buckley. “Activation of the hole-forming toxin aerolysin by extracellular processing”. In: *Journal of Bacteriology* 163.1 (1985), pp. 336–340. URL: <https://jlb.asm.org/content/jlb/163/1/336.full.pdf>.
- [68] Yuanyuan Hu et al. “Identification and Proposed Relative and Absolute Configurations of Niphimycins C–E from the Marine-Derived Streptomyces sp. IMB7-145 by Genomic Analysis”. In: *Journal of Natural Products* 81.1 (2018), pp. 178–187. ISSN: 0163-3864. DOI: 10.1021/acs.jnatprod.7b00859. URL: <https://doi.org/10.1021/acs.jnatprod.7b00859>.
- [69] Gia Khuong Hoang Hua and Monica Höfte. “The involvement of phenazines and cyclic lipopeptide sessilin in biocontrol of Rhizoctonia root rot on bean (*Phaseolus vulgaris*) by *Pseudomonas* sp. CMR12a is influenced by substrate composition”. In: *Plant and Soil* 388.1 (2015), pp. 243–253. ISSN: 1573-5036. DOI: 10.1007/s11104-014-2327-y. URL: <https://doi.org/10.1007/s11104-014-2327-y>.
- [70] Anwar Huq et al. “Critical Factors Influencing the Occurrence of *Vibrio cholerae* in the Environment of Bangladesh”. In: *Applied and Environmental Microbiology* 71.8 (2005), pp. 4645–4654. DOI: 10.1128/aem.71.8.4645-4654.2005. URL: <https://aem.asm.org/content/aem/71/8/4645.full.pdf>.
- [71] Mareike Imhoff. “Validation of chemical elicitors and use of a heterologous expression system for the identification of new antibiotics from actinobacteria”. In: *Master Thesis* (2016).
- [72] Danny Ionescu et al. “Microbial and Chemical Characterization of Underwater Fresh Water Springs in the Dead Sea”. In: *PLOS ONE* 7.6 (2012), e38319. DOI: 10.1371/journal.pone.0038319. URL: <https://doi.org/10.1371/journal.pone.0038319>.
- [73] Sambrook J.F and D. W. Russell. *Molecular Cloning: A Laboratory Manual (3-Volume Set)*. Vol. 1. 2001. ISBN: 978-087969577-4.
- [74] Nina Jagmann, Katharina Styp von Rekowski, and Bodo Philipp. “Interactions of bacteria with different mechanisms for chitin degradation result in the formation of a mixed-species biofilm”. In: *FEMS Microbiology Letters* 326.1 (2012), pp. 69–75. ISSN: 0378-1097. DOI: 10.1111/j.1574-6968.2011.02435.x. URL: <https://doi.org/10.1111/j.1574-6968.2011.02435.x>.

- [75] Gahzaleh Jahanshah et al. “Discovery of the Cyclic Lipopeptide Gacamide A by Genome Mining and Repair of the Defective GacA Regulator in *Pseudomonas fluorescens* Pf0-1”. In: *Journal of Natural Products* 82.2 (2019), pp. 301–308. ISSN: 0163-3864. DOI: 10.1021/acs.jnatprod.8b00747. URL: <https://doi.org/10.1021/acs.jnatprod.8b00747>.
- [76] Rolf Jansen et al. “Elansolid2005A3, a Unique p-Quinone Methide Antibiotic from *Chitinophaga sancti*”. In: *Chemistry – A European Journal* 17.28 (2011), pp. 7739–7744. ISSN: 0947-6539. DOI: 10.1002/chem.201100457. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/chem.201100457>.
- [77] B. K. Jap and P. J. Walian. “Structure and functional mechanism of porins”. In: *Physiological Reviews* 76.4 (1996), pp. 1073–1088. ISSN: 0031-9333. DOI: 10.1152/physrev.1996.76.4.1073. URL: <https://doi.org/10.1152/physrev.1996.76.4.1073>.
- [78] L. E. Johnson and A. Dietz. “Scopafungin, a crystalline antibiotic produced by *Streptomyces hygroscopicus* var. *enhygrus* var. *nova*”. In: *Applied microbiology* 22.3 (1971), pp. 303–308. ISSN: 0003-6919. URL: <https://www.ncbi.nlm.nih.gov/pubmed/4940870><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC376304/>.
- [79] Markus Karner and Gerhard J. Herndl. “Extracellular enzymatic activity and secondary production in free-living and marine-snow-associated bacteria”. In: *Marine Biology* 113.2 (1992), pp. 341–347. ISSN: 1432-1793. DOI: 10.1007/bf00347289. URL: <https://doi.org/10.1007/BF00347289>.
- [80] Nami et al. Katayama. “Formadicins, new monocyclic beta-lactam antibiotics of bacterial origin. I. Taxonomy, fermentation and biological activities.” In: *The Journal of antibiotics* 38.9 (1985), pp. 1117–27.
- [81] Thomas J. Kirn, Brooke A. Jude, and Ronald K. Taylor. “A colonization factor links *Vibrio cholerae* environmental survival and human infection”. In: *Nature* 438.7069 (2005), pp. 863–866. ISSN: 1476-4687. DOI: 10.1038/nature04249. URL: <https://doi.org/10.1038/nature04249>.
- [82] Laurens H.J. Kleijn and Nathaniel I. Martin. “The Cyclic Lipopeptide Antibiotics”. In: Berlin, Heidelberg: Springer Berlin Heidelberg, 2017, pp. 1–27. DOI: 10.1007/7355_2017_9. URL: https://doi.org/10.1007/7355_2017_9.
- [83] Anna Klindworth et al. “Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies”. In: *Nucleic Acids Research* 41.1 (2013), e1–e1. ISSN: 0305-1048.

- DOI: 10.1093/nar/gks808. URL: <http://dx.doi.org/10.1093/nar/gks808>.
- [84] Veneta Ivanova; Victoria Gesheva; Mariana Kolarova. “Dihydroniphimycin: New Polyol Macrolide Antibiotic Produced by *Streptomyces hygrosopicus* 15”. In: *The Journal of Antibiotics* 53.6 (2000), pp. 627–632. DOI: <https://doi.org/10.7164/antibiotics.53.627>, .
- [85] Femke I. Kraas et al. “Functional Dissection of Surfactin Synthetase Initiation Module Reveals Insights into the Mechanism of Lipoinitiation”. In: *Chemistry Biology* 17.8 (2010), pp. 872–880. ISSN: 1074-5521. DOI: <https://doi.org/10.1016/j.chembiol.2010.06.015>. URL: <http://www.sciencedirect.com/science/article/pii/S1074552110002590>.
- [86] Bernd Krone and Axel Zeeck. “Neue Isochromanchinon-Antibiotica der Naphthocyclinon-Reihe”. In: *Liebigs Annalen der Chemie* 1983.3 (1983), pp. 471–502. ISSN: 0170-2041. DOI: 10.1002/jlac.198319830313. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jlac.198319830313>.
- [87] Daniel Krug and Rolf Muller. “Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products”. In: *Natural Product Reports* 31.6 (2014), pp. 768–783. ISSN: 0265-0568. DOI: 10.1039/C3NP70127A. URL: <http://dx.doi.org/10.1039/C3NP70127A>.
- [88] Daniel Krug et al. “Efficient mining of myxobacterial metabolite profiles enabled by liquid chromatography–electrospray ionisation–time-of-flight mass spectrometry and compound-based principal component analysis”. In: *Analytica Chimica Acta* 624.1 (2008), pp. 97–106. ISSN: 0003-2670. DOI: <https://doi.org/10.1016/j.aca.2008.06.036>. URL: <http://www.sciencedirect.com/science/article/pii/S0003267008012439>.
- [89] Johannes H. Kügler et al. “Surfactants tailored by the class Actinobacteria”. In: *Frontiers in Microbiology* 6.212 (2015). ISSN: 1664-302X. DOI: 10.3389/fmicb.2015.00212. URL: <https://www.frontiersin.org/article/10.3389/fmicb.2015.00212>.
- [90] T. A. Kunkel. “Evolving views of DNA replication (in)fidelity”. In: *Cold Spring Harbor symposia on quantitative biology* 74 (2009), pp. 91–101. ISSN: 1943-4456 0091-7451. DOI: 10.1101/sqb.2009.74.027. URL: <https://www.ncbi.nlm.nih.gov/pubmed/19903750><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3628614/>.
- [91] Hartmut Laatsch. *AntiBase: the natural compound identifier*. Wiley-Vch, 2017. ISBN: 3527343598.

- [92] Markus Lakemeyer et al. “Thinking Outside the Box—Novel Antibacterials To Tackle the Resistance Crisis”. In: *Angewandte Chemie International Edition* 57.44 (2018), pp. 14440–14475. ISSN: 1433-7851. DOI: 10.1002/anie.201804971. URL: <https://doi.org/10.1002/anie.201804971>.
- [93] Maurice V. Laycock et al. “Viscosin, a potent peptidolipid biosurfactant and phytopathogenic mediator produced by a pectolytic strain of *Pseudomonas fluorescens*”. In: *Journal of Agricultural and Food Chemistry* 39.3 (1991), pp. 483–489. ISSN: 0021-8561. DOI: 10.1021/jf00003a011. URL: <https://doi.org/10.1021/jf00003a011>.
- [94] S.D. Lee S.Y. Hong and H.S. et al. Kim M.S. Kim. “Structure determination and biological activities of elaiophylin produced by *Streptomyces* sp. MCY-846”. In: *Journal of Microbiology and Biotechnology* 6.4 (1997), pp. 245–249.
- [95] Shaonan Lei et al. “Analysis of the community composition and bacterial diversity of the rhizosphere microbiome across different plant taxa”. In: *MicrobiologyOpen* 8.6 (2019), e00762–e00762. ISSN: 2045-8827. DOI: 10.1002/mbo3.762. URL: <https://www.ncbi.nlm.nih.gov/pubmed/30565881%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6562120/>.
- [96] Kim Lewis. “Platforms for antibiotic discovery”. In: *Nature Reviews Drug Discovery* 12 (2013), p. 371. DOI: 10.1038/nrd3975. URL: <http://dx.doi.org/10.1038/nrd3975>.
- [97] Kim Lewis et al. “Uncultured microorganisms as a source of secondary metabolites”. In: *The Journal Of Antibiotics* 63 (2010), p. 468. DOI: 10.1038/ja.2010.87. URL: <https://doi.org/10.1038/ja.2010.87>.
- [98] Weizhong Li and Adam Godzik. “Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences”. In: *Bioinformatics* 22.13 (2006), pp. 1658–1659. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btl158. URL: <http://dx.doi.org/10.1093/bioinformatics/btl158>.
- [99] Wen Li et al. “The Antimicrobial Compound Xantholysin Defines a New Group of *Pseudomonas* Cyclic Lipopeptides”. In: *PLOS ONE* 8.5 (2013), e62946. DOI: 10.1371/journal.pone.0062946. URL: <https://doi.org/10.1371/journal.pone.0062946>.
- [100] S.P. Lim et al. “Flexible exportation mechanisms of arthrofactin in *Pseudomonas* sp. MIS38”. In: *Journal of Applied Microbiology* 107.1 (2009), pp. 157–166. ISSN: 1364-5072. DOI: 10.1111/j.1365-2672.2009.04189.x. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2672.2009.04189.x>.

- [101] Christopher A. Lipinski et al. “Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings”. In: *Advanced Drug Delivery Reviews* 46.1 (2001), pp. 3–26. ISSN: 0169-409X. DOI: [https://doi.org/10.1016/S0169-409X\(00\)00129-0](https://doi.org/10.1016/S0169-409X(00)00129-0). URL: <http://www.sciencedirect.com/science/article/pii/S0169409X00001290>.
- [102] María-Natalia Lisa et al. “A general reaction mechanism for carbapenem hydrolysis by mononuclear and binuclear metallo- β -lactamases”. In: *Nature Communications* 8.1 (2017), p. 538. ISSN: 2041-1723. DOI: [10.1038/s41467-017-00601-9](https://doi.org/10.1038/s41467-017-00601-9). URL: <https://doi.org/10.1038/s41467-017-00601-9>.
- [103] Ji-Kai Liu. “Natural Terphenyls:2009 Developments since 1877”. In: *Chemical Reviews* 106.6 (2006), pp. 2209–2223. ISSN: 0009-2665. DOI: [10.1021/cr050248c](https://doi.org/10.1021/cr050248c). URL: <https://doi.org/10.1021/cr050248c>.
- [104] Yi-Yun Liu et al. “Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study”. In: *The Lancet Infectious Diseases* 16.2 (2016), pp. 161–168. ISSN: 1473-3099. DOI: [https://doi.org/10.1016/S1473-3099\(15\)00424-7](https://doi.org/10.1016/S1473-3099(15)00424-7). URL: <http://www.sciencedirect.com/science/article/pii/S1473309915004247>.
- [105] Richard A. Long and Farooq Azam. “Antagonistic Interactions among Marine Pelagic Bacteria”. In: *Applied and Environmental Microbiology* 67.11 (2001), pp. 4975–4983. DOI: [10.1128/aem.67.11.4975-4983.2001](https://doi.org/10.1128/aem.67.11.4975-4983.2001). URL: <https://aem.asm.org/content/aem/67/11/4975.full.pdf>.
- [106] Zongwang Ma et al. “Biosynthesis, Chemical Structure, and Structure-Activity Relationship of Orfamide Lipopeptides Produced by *Pseudomonas protegens* and Related Species”. In: *Frontiers in Microbiology* 7.382 (2016). ISSN: 1664-302X. DOI: [10.3389/fmicb.2016.00382](https://doi.org/10.3389/fmicb.2016.00382). URL: <https://www.frontiersin.org/article/10.3389/fmicb.2016.00382>.
- [107] Emilie Macke et al. “Host-genotype dependent gut microbiota drives zooplankton tolerance to toxic cyanobacteria”. In: *Nature Communications* 8.1 (2017), p. 1608. ISSN: 2041-1723. DOI: [10.1038/s41467-017-01714-x](https://doi.org/10.1038/s41467-017-01714-x). URL: <https://doi.org/10.1038/s41467-017-01714-x>.
- [108] Mohamed A. Marahiel. “A structural model for multimodular NRPS assembly lines”. In: *Natural Product Reports* 33.2 (2016), pp. 136–140. ISSN:

- 0265-0568. DOI: 10.1039/C5NP00082C. URL: <http://dx.doi.org/10.1039/C5NP00082C>.
- [109] Mohamed A. Marahiel, Torsten Stachelhaus, and Henning D. Mootz. “Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis”. In: *Chemical Reviews* 97.7 (1997), pp. 2651–2674. ISSN: 0009-2665. DOI: 10.1021/cr960029e. URL: <https://doi.org/10.1021/cr960029e>.
- [110] Michael Marner et al. “Molecular Networking-Guided Discovery and Characterization of Stechlisins, a Group of Cyclic Lipopeptides from a *Pseudomonas* sp”. In: *Journal of Natural Products* (2020). ISSN: 0163-3864. DOI: 10.1021/acs.jnatprod.0c00263. URL: <https://doi.org/10.1021/acs.jnatprod.0c00263>.
- [111] Rajendra Maskey, Iris Grün-Wollny, and Hartmut Laatsch. “Resomycins A-C: New Anthracyclinone Antibiotics Formed by a Terrestrial *Streptomyces* sp”. In: *The Journal of antibiotics* 56 (2003), pp. 795–800. DOI: 10.7164/antibiotics.56.795.
- [112] T.L. Maugeri et al. “Distribution of potentially pathogenic bacteria as free living and plankton associated in a marine coastal zone”. In: *Journal of Applied Microbiology* 97.2 (2004), pp. 354–361. ISSN: 1364-5072. DOI: 10.1111/j.1365-2672.2004.02303.x. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2672.2004.02303.x>.
- [113] Kerrie L. May and Marcin Grabowicz. “The bacterial outer membrane is an evolving antibiotic barrier”. In: *Proceedings of the National Academy of Sciences* 115.36 (2018), pp. 8852–8854. DOI: 10.1073/pnas.1812779115. URL: <https://www.pnas.org/content/pnas/115/36/8852.full.pdf>.
- [114] Richard William Meek, Hrushi Vyas, and Laura Jane Violet Piddock. “Nonmedical Uses of Antibiotics: Time to Restrict Their Use?” In: *PLOS Biology* 13.10 (2015), e1002266. DOI: 10.1371/journal.pbio.1002266. URL: <https://doi.org/10.1371/journal.pbio.1002266>.
- [115] Folker Meyer et al. “GenDB—an open source genome annotation system for prokaryote genomes”. In: *Nucleic Acids Research* 31.8 (2003), pp. 2187–2195. ISSN: 0305-1048. DOI: 10.1093/nar/gkg312. URL: <https://doi.org/10.1093/nar/gkg312>.
- [116] Samuel I. Miller. “Antibiotic Resistance and Regulation of the Gram-Negative Bacterial Outer Membrane Barrier by Host Innate Immune Molecules”. In: *mBio* 7.5 (2016), e01541–16. ISSN: 2150-7511. DOI: 10.1128/mBio.01541-16. URL: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5040116/>.

- [117] Sita D. Modali and Helen I. Zgurskaya. “The periplasmic membrane proximal domain of MacA acts as a switch in stimulation of ATP hydrolysis by MacB transporter”. In: *Molecular Microbiology* 81.4 (2011), pp. 937–951. ISSN: 1365-2958. DOI: 10.1111/j.1365-2958.2011.07744.x. URL: <https://doi.org/10.1111/j.1365-2958.2011.07744.x>.
- [118] Tatsushi Mogi et al. “Identification of new inhibitors for alternative NADH dehydrogenase (NDH-II)”. In: *FEMS Microbiology Letters* 291.2 (2009), pp. 157–161. ISSN: 0378-1097. DOI: 10.1111/j.1574-6968.2008.01451.x. URL: <https://doi.org/10.1111/j.1574-6968.2008.01451.x>.
- [119] Kathrin I. Mohr et al. “Pinensins: The First Antifungal Lantibiotics”. In: *Angewandte Chemie International Edition* 54.38 (2015), pp. 11254–11258. ISSN: 1433-7851. DOI: 10.1002/anie.201500927. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201500927>.
- [120] Eva Møller, Lasse Riemann, and Morten Søndergaard. *Bacteria associated with copepods: Abundance, activity and community composition*. Vol. 47. 2007, pp. 99–106. DOI: 10.3354/ame047099.
- [121] Mark G. Moloney. “Natural Products as a Source for Novel Antibiotics”. In: *Trends in Pharmacological Sciences* 37.8 (2016), pp. 689–701. ISSN: 0165-6147. DOI: <https://doi.org/10.1016/j.tips.2016.05.001>. URL: <http://www.sciencedirect.com/science/article/pii/S0165614716300232>.
- [122] Fernando Morgan-Sagastume et al. “Production of acylated homoserine lactones by *Aeromonas* and *Pseudomonas* strains isolated from municipal activated sludge”. In: *Canadian Journal of Microbiology* 51.11 (2005), pp. 924–933. ISSN: 0008-4166. DOI: 10.1139/w05-077. URL: <https://doi.org/10.1139/w05-077>.
- [123] Aleksandr Morgulis et al. “Database indexing for production MegaBLAST searches”. In: *Bioinformatics* 24.16 (2008), pp. 1757–1764. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btn322. URL: <https://doi.org/10.1093/bioinformatics/btn322>.
- [124] David J. Newman and Gordon M. Cragg. “Natural Products as Sources of New Drugs from 1981 to 2014”. In: *Journal of Natural Products* 79.3 (2016), pp. 629–661. ISSN: 0163-3864. DOI: 10.1021/acs.jnatprod.5b01055. URL: <https://doi.org/10.1021/acs.jnatprod.5b01055>.
- [125] David J. Newman and Gordon M. Cragg. “Natural Products as Sources of New Drugs over the Last 25 Years”. In: *Journal of Natural Products* 70.3 (2007), pp. 461–477. ISSN: 0163-3864. DOI: 10.1021/np068054v. URL: <https://doi.org/10.1021/np068054v>.

- [126] Don D. Nguyen et al. “Indexing the *Pseudomonas* specialized metabolome enabled the discovery of poaeamide B and the bananamides”. In: *Nature Microbiology* 2 (2016), p. 16197. DOI: 10.1038/nmicrobiol.2016.197<https://www.nature.com/articles/nmicrobiol2016197#supplementary-information>. URL: <https://doi.org/10.1038/nmicrobiol.2016.197>.
- [127] T. H. Nielsen et al. “Antibiotic and Biosurfactant Properties of Cyclic Lipopeptides Produced by Fluorescent *Pseudomonas* spp. from the Sugar Beet Rhizosphere”. In: *Applied and Environmental Microbiology* 68.7 (2002), pp. 3416–3423. DOI: 10.1128/aem.68.7.3416-3423.2002. URL: <http://aem.asm.org/content/68/7/3416.abstract>.
- [128] T. H. Nielsen et al. “Structure, production characteristics and fungal antagonism of tensin – a new antifungal cyclic lipopeptide from *Pseudomonas fluorescens* strain 96.578”. In: *Journal of Applied Microbiology* 89.6 (2000), pp. 992–1001. ISSN: 1365-2672. DOI: 10.1046/j.1365-2672.2000.01201.x. URL: <http://doi.org/10.1046/j.1365-2672.2000.01201.x>.
- [129] Rosemarie O’Shea and Heinz E. Moser. “Physicochemical Properties of Antibacterial Compounds: Implications for Drug Discovery”. In: *Journal of Medicinal Chemistry* 51.10 (2008), pp. 2871–2878. ISSN: 0022-2623. DOI: 10.1021/jm700967e. URL: <https://doi.org/10.1021/jm700967e>.
- [130] Brian D. Ondov, Nicholas H. Bergman, and Adam M. Phillippy. “Interactive metagenomic visualization in a Web browser”. In: *BMC Bioinformatics* 12.1 (2011), p. 385. ISSN: 1471-2105. DOI: 10.1186/1471-2105-12-385. URL: <https://doi.org/10.1186/1471-2105-12-385>.
- [131] Marc Ongena and Philippe Jacques. “*Bacillus* lipopeptides: versatile weapons for plant disease biocontrol”. In: *Trends in Microbiology* 16.3 (2008), pp. 115–125. ISSN: 0966-842X. DOI: <https://doi.org/10.1016/j.tim.2007.12.009>. URL: <http://www.sciencedirect.com/science/article/pii/S0966842X08000383>.
- [132] World Health Organization. “Antibiotic resistance”. In: *Fact sheets* (2018). URL: <https://www.who.int/en/news-room/fact-sheets/detail/antibiotic-resistance>.
- [133] World Health Organization. “Global Health Estimates 2016: Death by Cause, Age, Sex, by Country and Region, 2000-2016”. In: *WHO Fact sheet* (2018). URL: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>.
- [134] Center for Disease Control and Prevention. “Antibiotic Resistance Threats in the United States 2019”. In: (2019). URL: <https://www.cdc.gov/drugresistance/biggest-threats.html>.

- [135] Elmar Pruesse, Jörg Peplies, and Frank Oliver Glöckner. “SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes”. In: *Bioinformatics* 28.14 (2012), pp. 1823–1829. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bts252. URL: <http://dx.doi.org/10.1093/bioinformatics/bts252>.
- [136] Cameron R. Pye et al. “Retrospective analysis of natural products provides insights for future discovery trends”. In: *Proceedings of the National Academy of Sciences* 114.22 (2017), pp. 5601–5606. DOI: 10.1073/pnas.1614680114. URL: <https://www.pnas.org/content/pnas/114/22/5601.full.pdf>.
- [137] Christian Quast et al. “The SILVA ribosomal RNA gene database project: improved data processing and web-based tools”. In: *Nucleic Acids Research* 41.D1 (2013), pp. D590–D596. ISSN: 0305-1048. DOI: 10.1093/nar/gks1219. URL: <http://dx.doi.org/10.1093/nar/gks1219>.
- [138] Robert A. Quinn et al. “Molecular Networking As a Drug Discovery, Drug Metabolism, and Precision Medicine Strategy”. In: *Trends in Pharmacological Sciences* 38.2 (2017), pp. 143–154. ISSN: 0165-6147. DOI: <https://doi.org/10.1016/j.tips.2016.10.011>. URL: <http://www.sciencedirect.com/science/article/pii/S0165614716301559>.
- [139] Jos M. Raaijmakers, Irene de Bruijn, and Maarten J. D. de Kock. “Cyclic Lipopeptide Production by Plant-Associated *Pseudomonas* spp.: Diversity, Activity, Biosynthesis, and Regulation”. In: *Molecular Plant-Microbe Interactions* 19.7 (2006), pp. 699–710. ISSN: 0894-0282. DOI: 10.1094/MPMI-19-0699. URL: <https://doi.org/10.1094/MPMI-19-0699>.
- [140] L. Riemann and A. Winding. “Community Dynamics of Free-living and Particle-associated Bacterial Assemblages during a Freshwater Phytoplankton Bloom”. In: *Microbial Ecology* 42.3 (2001), pp. 274–285. ISSN: 1432-184X. DOI: 10.1007/s00248-001-0018-8. URL: <https://doi.org/10.1007/s00248-001-0018-8>.
- [141] Niran Roongsawang, Kenji Washio, and Masaaki Morikawa. “Diversity of Nonribosomal Peptide Synthetases Involved in the Biosynthesis of Lipopeptide Biosurfactants”. In: *International Journal of Molecular Sciences* 12.1 (2011), pp. 141–172. ISSN: 1422-0067. DOI: 10.3390/ijms12010141. URL: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3039948/>.
- [142] Niran Roongsawang et al. “Cloning and Characterization of the Gene Cluster Encoding Arthrofactin Synthetase from *Pseudomonas* sp. MIS38”. In: *Chemistry Biology* 10.9 (2003), pp. 869–880. ISSN: 1074-5521. DOI:

- <https://doi.org/10.1016/j.chembiol.2003.09.004>. URL: <http://www.sciencedirect.com/science/article/pii/S1074552103001960>.
- [143] Daniel Samain, J. Carter Cook, and Kenneth L. Rinehart. “Structure of scopafungin, a potent nonpolyene antifungal antibiotic”. In: *Journal of the American Chemical Society* 104.15 (1982), pp. 4129–4141. ISSN: 0002-7863. DOI: 10.1021/ja00379a015. URL: <https://doi.org/10.1021/ja00379a015>.
- [144] Tanja Schneider et al. “Cyclic lipopeptides as antibacterial agents – Potent antibiotic activity mediated by intriguing mode of actions”. In: *International Journal of Medical Microbiology* 304.1 (2014), pp. 37–43. ISSN: 1438-4221. DOI: <https://doi.org/10.1016/j.ijmm.2013.08.009>. URL: <http://www.sciencedirect.com/science/article/pii/S1438422113001331>.
- [145] Constance Schultz and Suzanne Geerlings. “Plasmid-Mediated Resistance in Enterobacteriaceae”. In: *Drugs* 72.1 (2012), pp. 1–16. ISSN: 1179-1950. DOI: 10.2165/11597960-000000000-00000. URL: <https://doi.org/10.2165/11597960-000000000-00000>.
- [146] Paul Shannon et al. “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks”. In: *Genome Research* 13.11 (2003), pp. 2498–2504. DOI: 10.1101/gr.1239303. URL: <http://genome.cshlp.org/content/13/11/2498.abstractN2>.
- [147] Dan Sorensen et al. “Cyclic lipoundecapeptide amphisin from *Pseudomonas* sp. strain DSS73”. In: *Acta Crystallographica Section C* 57.9 (2001), pp. 1123–1124. ISSN: 0108-2701. DOI: doi:10.1107/S0108270101010782. URL: <https://doi.org/10.1107/S0108270101010782>.
- [148] James T. Staley and Allan Konopka. “MEASUREMENT OF IN SITU ACTIVITIES OF NONPHOTOSYNTHETIC MICROORGANISMS IN AQUATIC AND TERRESTRIAL HABITATS”. In: *Annual Review of Microbiology* 39.1 (1985), pp. 321–346. ISSN: 0066-4227. DOI: 10.1146/annurev.mi.39.100185.001541. URL: <https://doi.org/10.1146/annurev.mi.39.100185.001541>.
- [149] Ragnar Stolt et al. “Second-Order Peak Detection for Multicomponent High-Resolution LC/MS Data”. In: *Analytical Chemistry* 78.4 (2006), pp. 975–983. ISSN: 0003-2700. DOI: 10.1021/ac050980b. URL: <https://doi.org/10.1021/ac050980b>.
- [150] D. R. Storm, K. S. Rosenthal, and P. E. Swanson. “Polymyxin and Related Peptide Antibiotics”. In: *Annual Review of Biochemistry* 46.1 (1977), pp. 723–763. ISSN: 0066-4154. DOI: 10.1146/annurev.bi.46.070177.

003451. URL: <https://doi.org/10.1146/annurev.bi.46.070177.003451>.
- [151] Evelina Tacconelli et al. “Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis”. In: *The Lancet Infectious Diseases* 18.3 (2018), pp. 318–327. ISSN: 1473-3099. DOI: [https://doi.org/10.1016/S1473-3099\(17\)30753-3](https://doi.org/10.1016/S1473-3099(17)30753-3). URL: <http://www.sciencedirect.com/science/article/pii/S1473309917307533>.
- [152] K. W. Tang, V. Turk, and H. P. Grossart. “Linkage between crustacean zooplankton and aquatic bacteria”. In: *Aquatic Microbial Ecology* 61.3 (2010), pp. 261–277. URL: <https://www.int-res.com/abstracts/ame/v61/n3/p261-277/>.
- [153] Kam Tang, K. M. L. Hutalle, and Hans-Peter Grossart. *Microbial abundance, composition and enzymatic activity during decomposition of copepod carcasses*. Vol. 45. 2006, pp. 219–227. DOI: [10.3354/ame045219](https://doi.org/10.3354/ame045219).
- [154] Teruhiko Beppu Tetsuo Hamamoto Haruo Seto. “Leptomycins A and B, new antifungal antibiotics”. In: *The Journal of Antibiotics* 36.6 (1983), pp. 646–650.
- [155] Miriam Tewes. “Association of hygienically relevant microorganisms with freshwater plankton”. In: *Dissertation* (2012).
- [156] Matthew F. Traxler and Roberto Kolter. “Natural products in soil microbe interactions and evolution”. In: *Natural Product Reports* 32.7 (2015), pp. 956–970. ISSN: 0265-0568. DOI: [10.1039/C5NP00013K](https://doi.org/10.1039/C5NP00013K). URL: <http://dx.doi.org/10.1039/C5NP00013K>.
- [157] Hideto Tsujimoto, Naomasa Gotoh, and Takeshi Nishino. “Diffusion of macrolide antibiotics through the outer membrane of *Moraxella catarrhalis*”. In: *Journal of Infection and Chemotherapy* 5.4 (1999), pp. 196–200. ISSN: 1341-321X. DOI: <https://doi.org/10.1007/s101560050034>. URL: <http://www.sciencedirect.com/science/article/pii/S1341321X99713450>.
- [158] Frank Uliczka et al. “Monitoring of Gene Expression in Bacteria during Infections Using an Adaptable Set of Bioluminescent, Fluorescent and Colorigenic Fusion Vectors”. In: *PLOS ONE* 6.6 (2011), e20425. DOI: [10.1371/journal.pone.0020425](https://doi.org/10.1371/journal.pone.0020425). URL: <https://doi.org/10.1371/journal.pone.0020425>.

- [159] Isabelle Vallet-Gely et al. “Association of Hemolytic Activity of *Pseudomonas entomophila*, a Versatile Soil Bacterium, with Cyclic Lipopeptide Production”. In: *Applied and Environmental Microbiology* 76.3 (2010), pp. 910–921. DOI: 10.1128/aem.02112-09. URL: <https://aem.asm.org/content/aem/76/3/910.full.pdf>.
- [160] Cees M. Verduin et al. “*Moraxella catarrhalis*: from emerging to established pathogen”. In: *Clinical microbiology reviews* 15.1 (2002), pp. 125–144. ISSN: 0893-8512 1098-6618. DOI: 10.1128/cmr.15.1.125-144.2002. URL: <https://www.ncbi.nlm.nih.gov/pubmed/11781271><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC118065/>.
- [161] Mingxun Wang et al. “Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking”. In: *Nat Biotech* 34.8 (2016), pp. 828–837. ISSN: 1087-0156. DOI: 10.1038/nbt.3597<http://www.nature.com/nbt/journal/v34/n8/abs/nbt.3597.html#supplementary-information>. URL: <http://dx.doi.org/10.1038/nbt.3597>.
- [162] Grant W. Waterer and Richard G. Wunderink. “Increasing threat of Gram-negative bacteria”. In: *Critical Care Medicine* 29.4 (2001), N75–N81. ISSN: 0090-3493. URL: http://journals.lww.com/ccmjournal/Fulltext/2001/04001/Increasing_threat_of_Gram_negative_bacteria.4.aspx.
- [163] Milind G. Watve et al. “How many antibiotics are produced by the genus *Streptomyces*?” In: *Archives of Microbiology* 176.5 (2001), pp. 386–390. ISSN: 1432-072X. DOI: 10.1007/s002030100345. URL: <https://doi.org/10.1007/s002030100345>.
- [164] W G Weisburg et al. “16S ribosomal DNA amplification for phylogenetic study”. In: *Journal of Bacteriology* 173.2 (1991), pp. 697–703. DOI: 10.1128/jb.173.2.697-703.1991. URL: <https://jb.asm.org/content/jb/173/2/697.full.pdf>.
- [165] J. W. Westley et al. “Conglobatin, a novel macrolide dilactone from *Streptomyces conglobatus* ATCC 31005”. In: *The Journal of antibiotics* 32.9 (1979), pp. 874–877. ISSN: 0021-8820. DOI: 10.7164/antibiotics.32.874. URL: <http://europepmc.org/abstract/MED/511778><https://doi.org/10.7164/antibiotics.32.874>.
- [166] Jane Y. Yang et al. “Molecular Networking as a Dereplication Strategy”. In: *Journal of Natural Products* 76.9 (2013), pp. 1686–1699. ISSN: 0163-3864. DOI: 10.1021/np400413s. URL: <http://dx.doi.org/10.1021/np400413s>.

- [167] Jianyun Yao et al. “Identification and characterization of a HEPN-MNT family type II toxin-antitoxin in *Shewanella oneidensis*”. In: *Microbial biotechnology* 8.6 (2015), pp. 961–973. ISSN: 1751-7915. DOI: 10.1111/1751-7915.12294. URL: <https://www.ncbi.nlm.nih.gov/pubmed/26112399%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4621449/>.
- [168] Axel Zeeck, Hans Zähler, and Mithat Mardin. “Stoffwechselprodukte von Mikroorganismen, 1291) Isolierung und Konstitution der Isochromanchinon-Antibiotica beta- und gamma-Naphthocyclinon”. In: *Justus Liebigs Annalen der Chemie* 1974.7 (1974), pp. 1100–1125. DOI: doi : 10 . 1002 / jlac . 197419740707. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jlac.197419740707>.
- [169] Karsten Zengler et al. “Cultivating the uncultured”. In: *Proceedings of the National Academy of Sciences* 99.24 (2002), p. 15681. DOI: 10.1073/pnas.252630999. URL: <http://www.pnas.org/content/99/24/15681.abstract>.
- [170] Aleksey V. Zimin et al. “The MaSuRCA genome assembler”. In: *Bioinformatics* 29.21 (2013), pp. 2669–2677. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btt476. URL: <https://doi.org/10.1093/bioinformatics/btt476>.

7 Supplements

7.1 Metabolomics platform

7.1.1 Data processing

Figure S1: Custom script used for LCMS data processing in Data Analysis 4.4

```
1 Analysis.Compounds.Clear
2
3 Analysis.RecalculateLineSpectra
4
5 Analysis.Save
6
7 Analysis.AddChromatogramRangeSelection 0.1,22.5,0,0
8
9 CalCheck
10
11 Analysis.FindMolecularFeatures
12
13 Analysis.Save
14 Form.Close
15
16 '*****
17 Function CalCheck
18   If not Analysis.RecalibrateAutomatically Then
19     Dim MsgText
20     MsgText = "Calibration of the analysis '"+Analysis.Name+"' failed."
21
22     Dim fso, f, ts
23     Set fso = CreateObject("Scripting.FileSystemObject")
24     fso.CreateTextFile "C:\BDalSystemData\Calib-Msg.txt" ' Create a file.
25     Set f = fso.GetFile("C:\BDalSystemData\Calib-Msg.txt")
26     Set ts = f.OpenAsTextStream(8)
27     ts.Write MsgText
28     ts.Close
29     MsgBox "Calibration of the analysis '"+Analysis.Name+"' failed."
30     'in File "C:\BDalSystemData\Calib-Msg.txt" sind calib-Fails zu lesen.
31   End If
32 End Function
```

Figure S2: Custom script used for metabolomic heatmap generation based on bucket vectors

```
1 # Version 2.0 - https://dx.doi.org/10.5281/zenodo.3932968  
2 https://github.com/christoph-hartwig-ime-br/cosine-V2  
3 # Written 2020 by Dipl.-Ing.(FH) Christoph Hartwig (Fraunhofer IME-BR) # Based on  
4 first implementation of the workflow by Dr. Florian Zubeil (Fraunhofer IME-BR) in  
5 2017.  
6 #  
7 # Version history:  
8 # V 1.0 2017 Florian Zubeil (Fraunhofer IME-BR) -  
9 https://dx.doi.org/10.5281/zenodo.3911715  
10 https://github.com/fzubeil-IME-BR/metabolomics\_cosine - FeatureFinding and Bucketing  
11 via xcmsSet and xSet, cosine calculation with own code in C, heatmap as widget with  
12 d3heatmap and saveWidget # V 2.0 2019 Christoph Hartwig (Fraunhofer IME-BR) -  
13 outsourcing of Feature Finding and Bucketing to Bruker Software - moving all  
14 calculations to R-packets - including export of results # # Data Processing is  
15 handled in DataAnalysis (Bruker) for feature finding [Reprocessed files are stored  
16 and can be used for future bucketings, so that reprocessig is necessary only once] #  
17 Bucketing is done via ProfileAnalysis (Bruker) and that output file "XYZ_HPlus.txt"  
18 is the input for this script [ProfileAnalysis allows huge sample numbers, only  
19 limited by memory.] # Additionally to the cosine similarity calculation (using the  
20 same distance function as above) and generation of the heatmap the following  
21 functions were added by Christoph Hartwig:  
22 # Export of the cosine similarity matrix [allowing further processing in e.g. Excel]  
23 # Statistics about buckets and samples numbers, uniqueness etc. [feedback  
24 information to evaluate feature finding and bucketing parameters] # Readout of  
25 clusterin sequence and pairwise similarities [to find "jumps" and therefore find  
26 borders of metabolomics clades]  
27  
28 library(readr)  
29 library(coop)  
30 library(gplots)  
31 require(data.table)  
32 library(parallelDist)  
33  
34 #Define functions for graphics generation my_palette <-  
35 colorRampPalette(c("white","blue"), bias=10)(n = 500) #color scheme  
36  
37 calcall<-function(project){  
38  
39   heatpng<-function(mt, graphname){  
40     png(filename=paste(c(project,graphname), collapse=""), # create PNG for the  
41         heat map  
42         width = 50*400, # 100 x 300 pixels  
43         height = 50*400,  
44         res = 1000, # 300 pixels per inch  
45         pointsize = 5) # smaller font size  
46     test2<-heatmap.2(mt, symm=TRUE, distfun=function(x) as.dist((1-x)/2),  
47         notecol="black", #as.dist((1-x)/2) inverts similarity to dissimilarity  
48         main = paste(c(project,graphname), collapse=""), # heat map title  
49         col=my_palette,  
50         # breaks=col breaks, # enable color transition at  
51         specified limits  
52         margins =c(25,25), # widens margins around plot  
53         trace="none")  
54     dev.off()  
55     write.table(mt[rev(test2$rowInd),0],  
56         file=paste(c(project,"_cosclust_result.txt"), collapse=""),  
57         dec=".",sep="\t",row.names=TRUE,col.names=TRUE) #safing the order of similarity  
58     matrix in an .txt file  
59     clusstable2<<-(mt[rev(test2$rowInd),0])  
60     clusstable<<-rev(test2$rowInd)  
61   }  
62  
63   table.gesamt<-read_tsv(paste(c(project,".txt"), collapse=""),col_names=TRUE,  
64     col_types = NULL, na = c("", "NA"), trim_ws = FALSE, skip = 0, n_max = Inf,  
65     progress = show_progress(), skip_empty_rows = TRUE) #load project file  
66  
67   table.gesamt<-as.data.frame(table.gesamt) #converts imported project file to data  
68   frame  
69   names<-table.gesamt[,1] #reading the filenames aka sample names  
70   n<-ncol(table.gesamt) #table length, necessary for matrix generation
```

```

44 table.buckets<-table.gesamt[,2:n] #cutting out the filenames
45
46 #readout of all bucketnames for search and export
47 bucketnames<-colnames(table.gesamt) #readout of bucketnames
48
49 matrix.bucketnames<-as.matrix(bucketnames) #transforming bucket name table to
matrix
50 rm(table.gesamt) #frees up memory, helpful for large datasets
51
52 table.buckets<-as.matrix(table.buckets) #transform buckettable to matrix
53 n2<-nrow(table.buckets) #calculates the number of samples
54 rownames(table.buckets)<-names #Adding the filenames as rownames
55 bucketnames2<-matrix.bucketnames[-1,] #extrats the bucketnames
56 colnames(table.buckets)<-bucketnames2 #adding the bucketnames as colnames
57 table.rotated<-t(table.buckets) #transformation of the buckettable for cosine
calculation
58 n.buckets.filled<-nrow(counter<-as.matrix(table.buckets[!table.buckets==0]))
#write filled buckets in vector and count lines
59 ausgabe<-summary(table.buckets[!table.buckets==0]) #saving results in variable for
output in summary
60
61 rm(table.buckets) #frees up memory, helpful for large datasets
62
63 #Analysis of filled buckets per sample
64 table.buckets.logical<-table.rotated
65 table.buckets.logical[table.buckets.logical > 1] <- 1
66 n.buckets.sample<-as.matrix(colSums(table.buckets.logical)) #How many buckets are
filled for every sample
67 n.samples.bucket<-as.matrix(rowSums(table.buckets.logical)) #How many samples
contain a certain feature, for every feature
68
69 #for every sample the number of buckets which are only present in this sample
70 n.uniquebuckets.sample<-as.matrix(table.buckets.logical)
71 n.uniquebuckets.sample[rowSums(n.uniquebuckets.sample) > 1] <-0
72 n.uniquebuckets.sample.result<-as.matrix(colSums(n.uniquebuckets.sample))
73 n.uniquebuckets.bucket.result<-as.matrix(rowSums(n.uniquebuckets.sample))
74
75 #Calculation of similarities
76 table.cosine<-cosine(table.rotated) #calculation of cosinus similarities based on
bucket table
77 rm(table.rotated)
78
79 #Export of tables
80 write.table(table.cosine, file=paste(c(project,"_cossim_result.txt"),
collapse=""), dec=".",sep="\t",row.names=TRUE,col.names=NA) #saving the cosine
similarity matrix in an .txt file
81 write.table(n.buckets.sample, file=paste(c(project,"_numfilledbuckets.txt"),
collapse=""), dec=".",sep="\t",row.names=TRUE,col.names=NA) #saving the number of
filled buckets per sample in an .txt file
82 write.table(matrix.bucketnames, file=paste(c(project,"_bucketnames.txt"),
collapse=""), dec=".",sep="\t",row.names=TRUE,col.names=NA) #saving the bucket
names in an .txt file
83 write.table(n.samples.bucket, file=paste(c(project,"_numsamplesperbucket.txt"),
collapse=""), dec=".",sep="\t",row.names=TRUE,col.names=NA) #saving the number of
sample containing each bucket in an .txt file
84
85 heatmap(table.cosine,"_cossim_result.png") #generates the heatmap output
86
87 clusttable3<-as.matrix(clusttable) #grabs the clustering sequence from the
heatmap and uses that to read the pairwise similarities in that sequence. Output
allows the generation of metabolic groups (based on clustering)
88 groupingoutput <- matrix(0 , ncol=2 ,nrow=nrow(clusttable3))
89 for(i in 1:nrow(clusttable3)){
90   groupingoutput[i,1]<- (colnames(table.cosine, do.NULL)[clusttable3[i]])
91 }
92 for(i in 2:nrow(clusttable3)){
93   groupingoutput[i,2]<- (table.cosine[clusttable3[i-1],clusttable3[i]])
94 }
95 }
96 write.table(groupingoutput, file=paste(c(project,"_grouping_result.txt"),
collapse=""), dec=".",sep="\t",row.names=FALSE,col.names=FALSE) #saving the cosine
similarity matrix in an .txt file
97

```

```

98 #generates a summary of the different outputs
99 capture.output(
100   print(project),
101   print("summary(table.buckets[!table.buckets==0])"),
102   ausgabe,
103   print(""),
104   print("Number of Buckets"),
105   n,
106   print(""),
107   print("Number of Samples"),
108   n2,
109   print(""),
110   print("total number of filled buckets"),
111   n.buckets.filled,
112   print(""),
113   print("Number of Samples per bucket"),
114   summary(n.samples.bucket),
115   print(""),
116   print("Number of Buckets per sample"),
117   summary(n.buckets.sample),
118   print(""),
119   print("Number of unique buckets pers sample"),
120   summary(n.uniquebuckets.sample.result),
121   print(""),
122   print("Number of unique buckets pers bucket"),
123   summary(n.uniquebuckets.bucket.result),
124
125   file=paste(c(project,"_summary.txt"), collapse=""), append=FALSE, type="output",
126   split=FALSE )
127
128 setwd("E:/") # Working directory
129 experiment<-"Dummy-Data" #Experiment name, has to be Filename_HPlus.txt without
_HPlus.txt as input project<-(paste(c(experiment,"_HPlus"), collapse="")) #adds the
_HPlus to adress the correct filename. Allows adresssing different bucket-tables
based on same dataset (used in other versions of this script)
130
131 calcall(paste(c(experiment,"_HPlus"), collapse="")) #starts the complete script
132
133

```

7.1.2 Bucket intensities

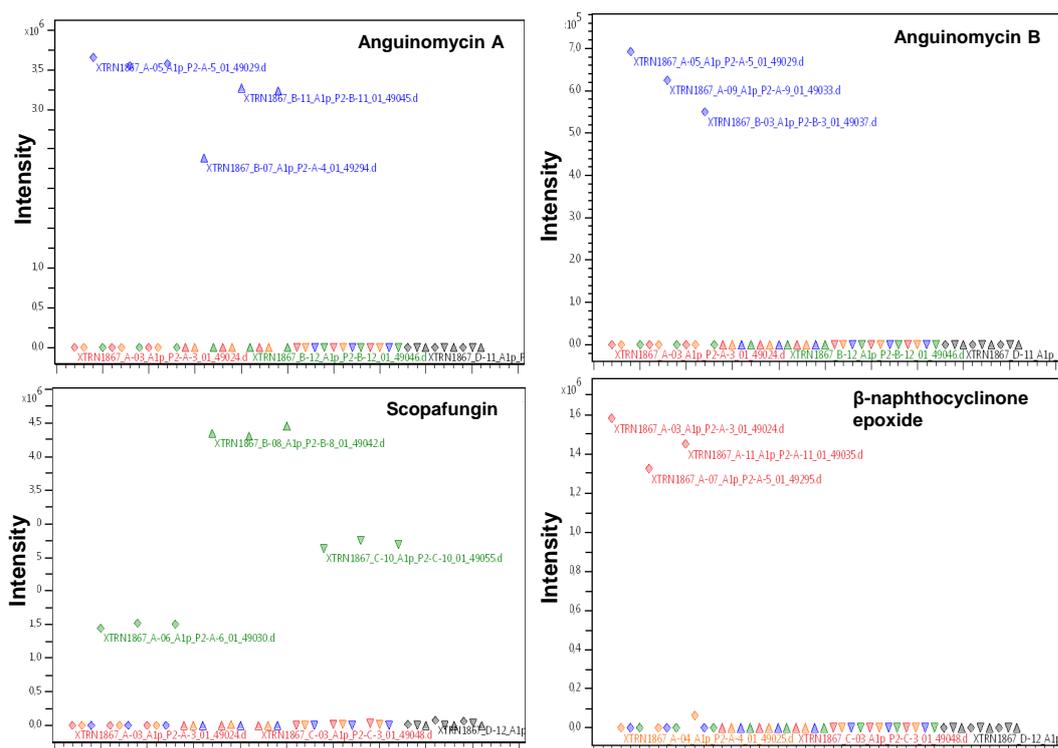


Figure S3: Absolute intensity of selected buckets: Top left: Anguinomycin A produced by ST100639 (blue) in media 5294 (triangles) and 5315 (diamonds). Top right: Anguinomycin B only produced in 5315. Bottom left: Scopafungin production by *Streptomyces* sp. ST1070165 cultured in different media. Production was observed to be 3 times higher in 5254 compared to 5315. Bottom right: Production of β -naphthocyclinone epoxide by strain ST101789 was only detected in 5315.

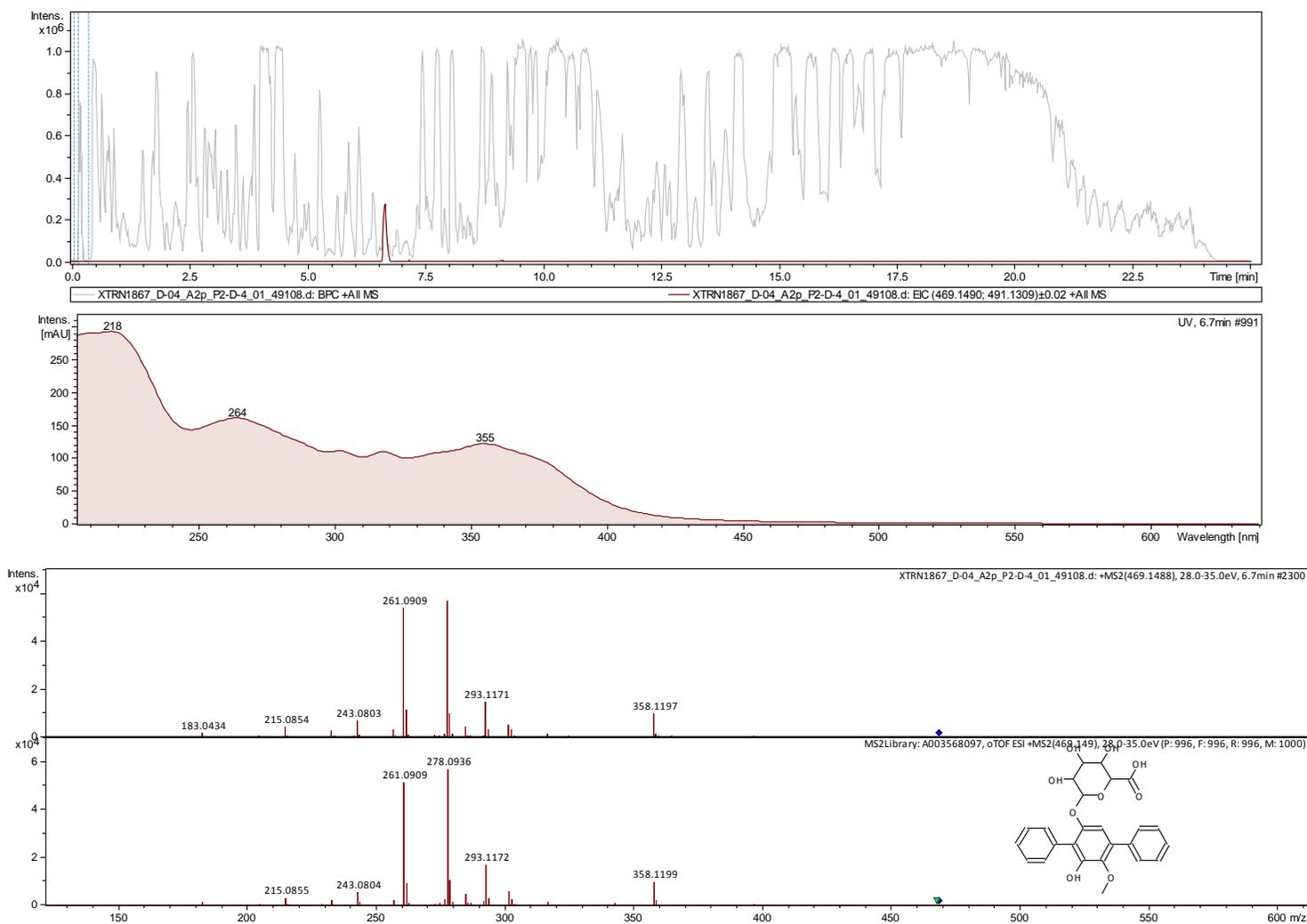


Figure S5: Echoside A: comparison of MS/MS fragmentation measured in ST1017165 extracts and pure compound. Top: Base peak Chromatogram of ST1017165 extract in grey and extracted ion chromatogram of m/z 469.149 at 6.7 min in red. Middle: UV absorption at 6.7 min. Bottom: Fragmentation signature of m/z 469.149 in ST1017165 extract and below from pure Echoside A from in house library and structures of major fragments. Fragmentation of compound in bacterial extract and reference is identical.

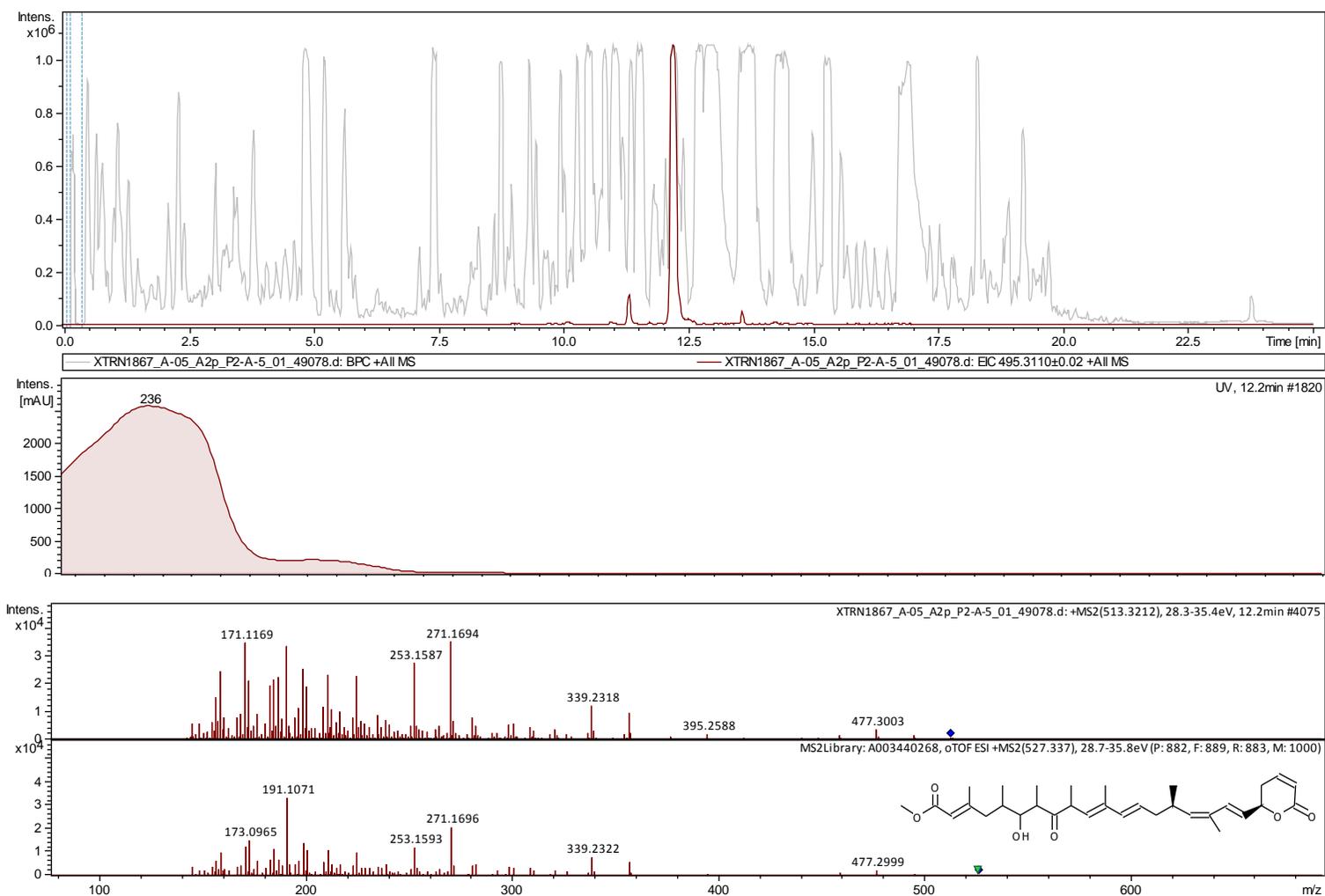


Figure S6: Anguinomycin A: comparison of MS/MS fragmentation measured in ST106693 extracts and pure compound. Top: Base peak Chromatogram of ST106693 extract in grey and extracted ion chromatogram of m/z 495.133; 513.3212 at 12.2 min in red. Middle: UV absorption at 12.2 min. Bottom: Fragmentation signature of m/z 513.3212 in ST106693 extract and below from pure Anguinomycin A from in house library. Fragmentation of compound in bacterial extract and reference is identical.

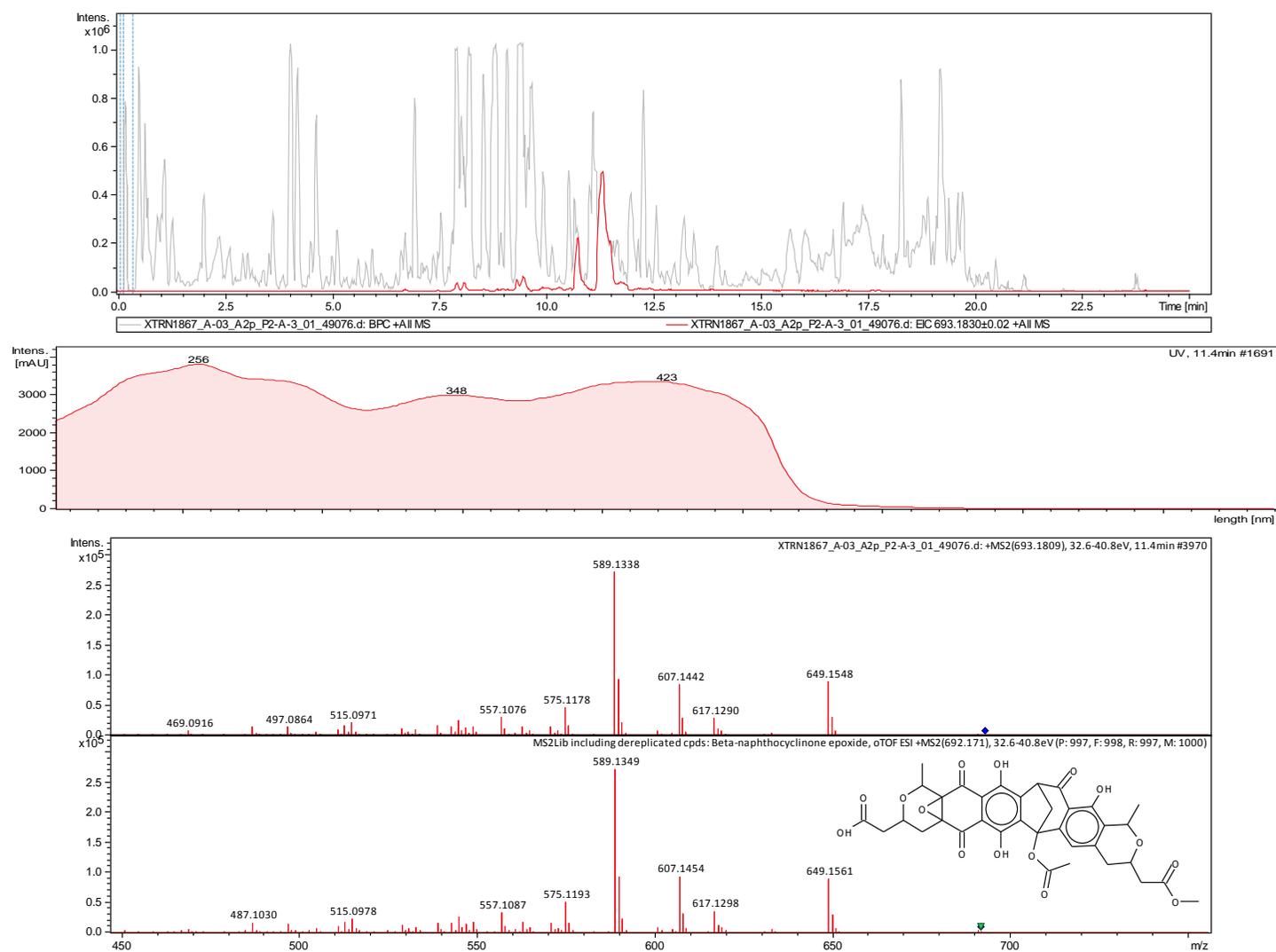


Figure S7: β -Naphthocyclinone-epoxide: comparison of MS/MS fragmentation measured in ST101789 extracts and pure compound. Top: Base peak Chromatogram of ST101789 extract in grey and extracted ion chromatogram of m/z 693.1809 at 11.4 min in red. Middle: UV absorption at 11.4 min. Bottom: Fragmentation signature of m/z 693.1809 in ST101789 extract and below from pure β -Naphthocyclinone-epoxide from in house library. Fragmentation of compound in bacterial extract and reference is identical.

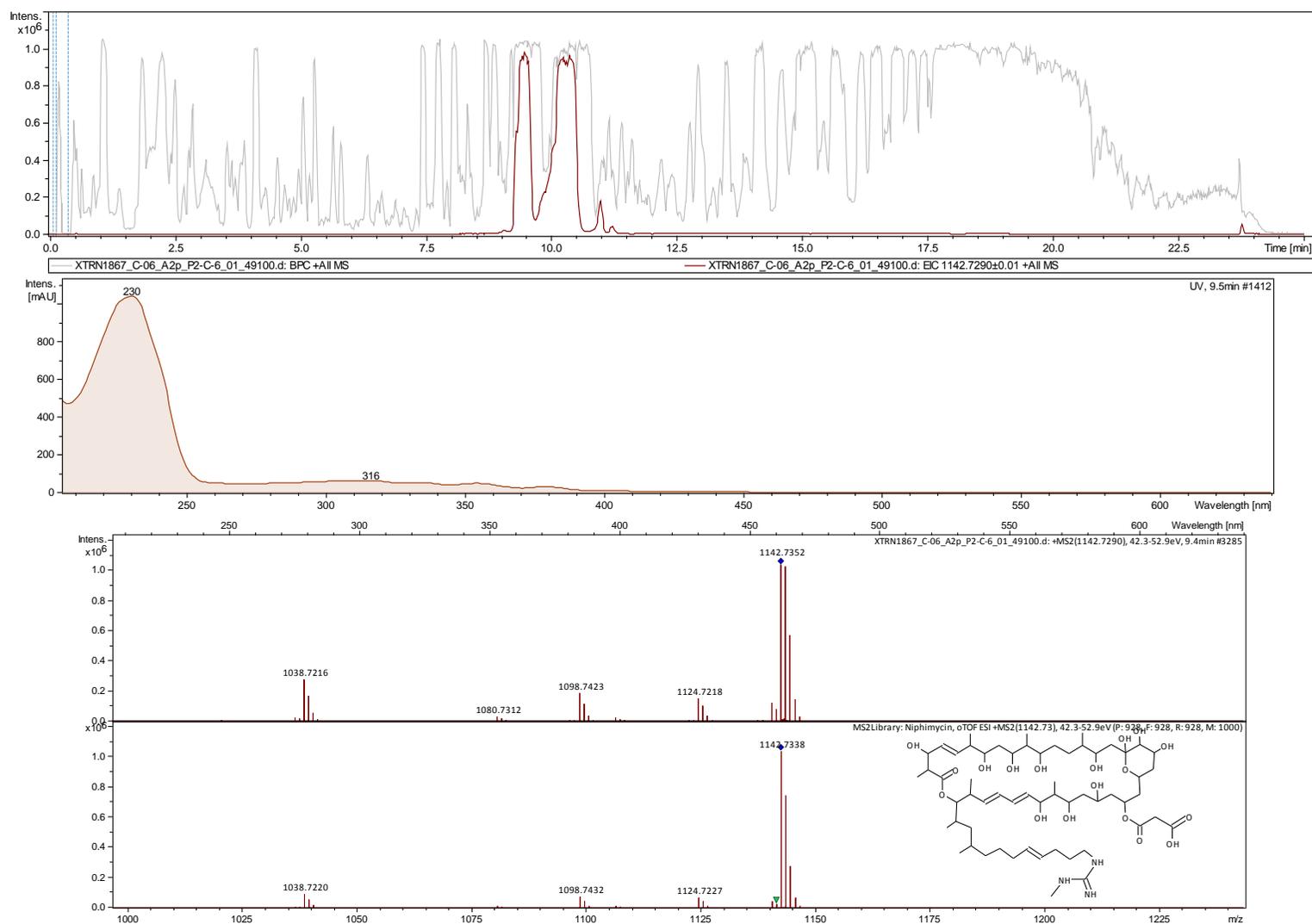


Figure S8: Scopafungin: comparison of MS/MS fragmentation measured in ST1017165 extracts and pure compound. Top: Base peak Chromatogram of ST1017165 extract in grey and extracted ion chromatogram of m/z 1142.7290 at 9.5 min in red. Middle: UV absorption at 9.5 min. Bottom: Fragmentation signature of m/z 1142.7290 in ST1017165 extract and below from pure Scopafungin from in house library. Fragmentation of compound in bacterial extract and reference is identical.

7.2 Sampling Lake Stechlin

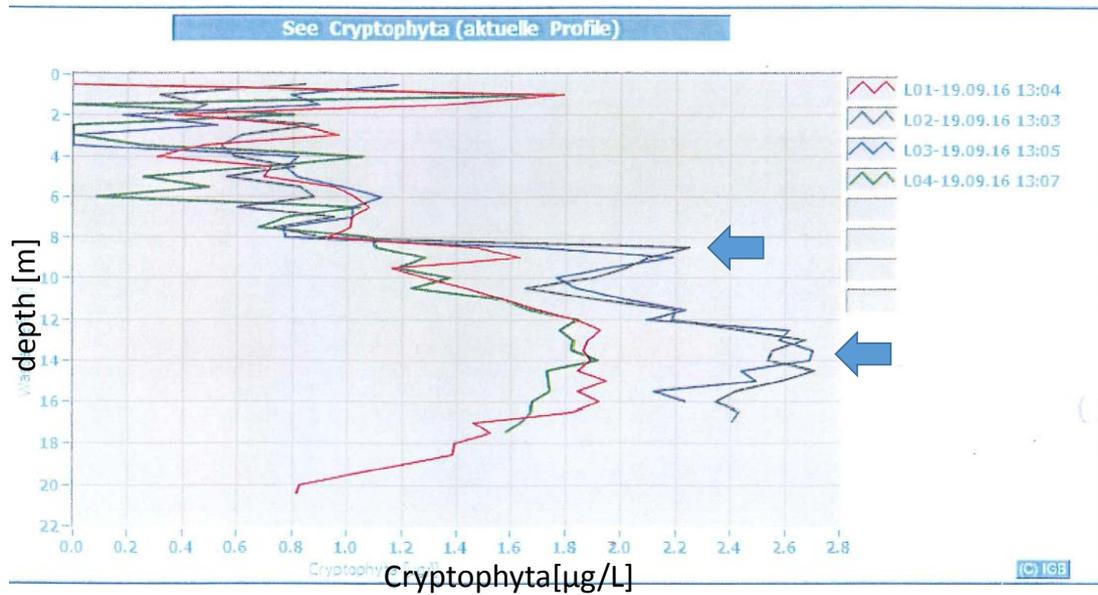


Figure S9: Biomass maxima Lake Stechlin. Shown are the depth profiles of cryptophyta distribution in the water column 0-22m of Lake Stechlin at the 19.09.2016, shortly before sampling was carried out. Cryptophyta maxima as a proxy for biomass were observed at 9 and 13 m

7.3 Microbiome analysis

	Plankton		water	
	90µm	250µm	0.1µm	1.2µm
Acidobacteria	0,463	0,339	0,590	0,273
Actinobacteria	0,711	0,279	41,087	14,992
Aegiribacteria	0,065	0,000	0,000	0,000
Armatimonadetes	0,011	0,009	0,000	0,253
Bacteriodetes	25,509	17,098	10,753	26,863
Chloroflexi	0,046	0,134	1,193	3,437
Cyanobacteria	0,978	0,111	1,073	3,821
Deinococcus-Thermus	0,000	0,000	1,200	0,050
Dependentiae	0,000	0,000	0,163	0,121
Elusimicrobia	0,002	0,000	0,000	0,000
Epsilonbacteraeota	0,684	2,235	0,122	0,063
Fibrobacteres	0,120	0,017	0,000	0,000
Firmicutes	0,782	0,126	4,850	3,482
Fusobacteria	0,000	0,013	0,031	0,264
Gemmatimonadetes	0,389	0,000	0,349	0,286
Hydrogenedente	0,013	0,000	0,000	0,000
Kiritimatiellaeota	0,040	0,000	0,000	0,008
Modulibacteria	0,000	0,006	0,000	0,000
Nitrospirae	0,000	0,009	0,146	0,000
Omnitrophicaeota	0,015	0,000	0,000	0,012
Patescibacteria	0,831	1,225	0,466	0,349
Planktomycetes	0,124	0,005	0,060	2,449
Proteobacteria	64,400	78,028	25,811	20,842
Spirochaetes	0,000	0,005	0,000	0,174
Tenericute	0,050	0,050	0,000	0,000
Verrucomicrobia	4,762	0,307	12,104	22,262

Figure S10: Phyla distribution across bacterial communities retrieved from Lake Stechlin [%]

7.4 Bioluminescence Assay layout

		1	2	3	4	5	6	7	8	9	10	11	12
A ($\mu\text{L/ml}$)	Gen 16	NEG CON	A1	A1	A2	A2	A3	A3	A4	A4	A5	A5	
B ($\mu\text{g/ml}$)	Gen 8	NEG CON	B1	B1	B2	B2	B3	B3	B4	B4	B5	B5	
C ($\mu\text{g/ml}$)	Gen 4	NEG CON	C1	C1	C2	C2	C3	C3	C4	C4	C5	C5	
D ($\mu\text{g/ml}$)	Gen 2	NEG CON	D1	D1	D2	D2	D3	D3	D4	D4	D5	D5	
E ($\mu\text{g/ml}$)	Gen 1	NEG CON	E1	E1	E2	E2	E3	E3	E4	E4	E5	E5	
F ($\mu\text{g/ml}$)	Gen 0.5	NEG CON	F1	F1	F2	F2	F3	F3	F4	F4	F5	F5	
G ($\mu\text{g/ml}$)	Gen 0.25	NEG CON	G1	G1	G2	G2	G3	G3	G4	G4	G5	G5	
H ($\mu\text{g/ml}$)	Gen 0.125	NEG CON	H1	H1	H2	H2	H3	H3	H4	H4	H5	H5	

		1	2	3	4	5	6	7	8	9	10	11	12
A ($\mu\text{L/ml}$)	Gen 16	NEG CON	A6	A6	A7	A7	A8	A8	A9	A9	A10	A10	
B ($\mu\text{g/ml}$)	Gen 8	NEG CON	B6	B6	B7	B7	B8	B8	B9	B9	B10	B10	
C ($\mu\text{g/ml}$)	Gen 4	NEG CON	C6	C6	C7	C7	C8	C8	C9	C9	C10	C10	
D ($\mu\text{g/ml}$)	Gen 2	NEG CON	D6	D6	D7	D7	D8	D8	D9	D9	D10	D10	
E ($\mu\text{g/ml}$)	Gen 1	NEG CON	E6	E6	E7	E7	E8	E8	E9	E9	E10	E10	
F ($\mu\text{g/ml}$)	Gen 0.5	NEG CON	F6	F6	F7	F7	F8	F8	F9	F9	F10	F10	
G ($\mu\text{g/ml}$)	Gen 0.25	NEG CON	G6	G6	G7	G7	G8	G8	G9	G9	G10	G10	
H ($\mu\text{g/ml}$)	Gen 0.125	NEG CON	H6	H6	H7	H7	H8	H8	H9	H9	H10	H10	

		1	2	3	4	5	6	7	8	9	10	11	12
A ($\mu\text{L/ml}$)	Gen 16	NEG CON	A11	A11	A12	A12	MED	MED	MED	MED	MED	MED	
B ($\mu\text{g/ml}$)	Gen 8	NEG CON	B11	B11	B12	B12	MED	MED	MED	MED	MED	MED	
C ($\mu\text{g/ml}$)	Gen 4	NEG CON	C11	C11	C12	C12	MED	MED	MED	MED	MED	MED	
D ($\mu\text{g/ml}$)	Gen 2	NEG CON	D11	D11	D12	D12	MED	MED	MED	MED	MED	MED	
E ($\mu\text{g/ml}$)	Gen 1	NEG CON	E11	E11	E12	E12	MED	MED	MED	MED	MED	MED	
F ($\mu\text{g/ml}$)	Gen 0.5	NEG CON	F11	F11	F12	F12	MED	MED	MED	MED	MED	MED	
G ($\mu\text{g/ml}$)	Gen 0.25	NEG CON	G11	G11	G12	G12	MED	MED	MED	MED	MED	MED	
H ($\mu\text{g/ml}$)	Gen 0.125	NEG CON	H11	H11	H12	H12	MED	MED	MED	MED	MED	MED	

Figure S11: Bioluminescence assay layout: Each plate of environmental cultivates is separated into three assay plates to accommodate a two point detection for each cultivate A1-A12: supernatants with coordinates on cultivation plate; Gen: gentamycin; NEGCON: *E. coli* DH5 α [pFU 166] in Mueller Hinton II without any supplementation; MED: supernatant of cultivation medium of environmental samples

7.5 μ -fractionation FhG10052 extract *C.albicans*

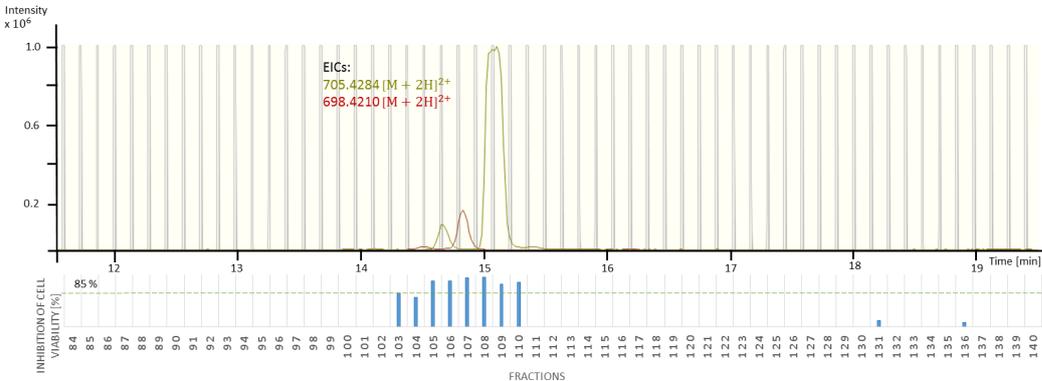


Figure S12: μ -fractionation of FhG10052 extract

7.6 MIC determination Assay layout

	1	2	3	4	5	6	7	8	9	10	11	12
A [μ L]	DMSO 1	DMSO 0.5	DMSO 0.25	DMSO 0.125	DMSO 0.06	DMSO 0.03	DMSO 0.015	DMSO 0.007	DMSO 0.003	DMSO 0.002	DMSO 0.001	DMSO 0.0005
B [μ g]	Cmp 128	Cmp 64	Cmp 32	Cmp 16	Cmp 8	Cmp 4	Cmp 2	Cmp 1	Cmp 0.5	Cmp 0.25	Cmp 0.125	Cmp 0.06
C [μ g]	Cmp 128	Cmp 64	Cmp 32	Cmp 16	Cmp 8	Cmp 4	Cmp 2	Cmp 1	Cmp 0.5	Cmp 0.25	Cmp 0.125	Cmp 0.06
D [μ g]	Cmp 128	Cmp 64	Cmp 32	Cmp 16	Cmp 8	Cmp 4	Cmp 2	Cmp 1	Cmp 0.5	Cmp 0.25	Cmp 0.125	Cmp 0.06
E [μ g]	Rif 64	Rif 32	Rif 16	Rif 8	Rif 4	Rif 2	Rif 1	Rif 0.5	Rif 0.25	Rif 0.125	Rif 0.06	Rif 0.03
F [μ g]	Tet 64	Tet 32	Tet 16	Tet 8	Tet 4	Tet 2	Tet 1	Tet 0.5	Tet 0.25	Tet 0.125	Tet 0.06	Tet 0.03
G [μ g]	Gen 64	Gen 32	Gen 16	Gen 8	Gen 4	Gen 2	Gen 1	Gen 0.5	Gen 0.25	Gen 0.125	Gen 0.06	Gen 0.03
H [μ g]	Blank	Blank	Blank	Blank	Blank	NEG CON	NEG CON	NEG CON	NEG CON	NEG CON	NEG CON	NEG CON

Figure S13: Plate design for determining MIC values of pure compounds against the microbial test panel. Positive controls in rows E-G: Rif = Rifamycin; Tet = Tetracycline; Gen = Gentamycin; numbers indicate tested concentration 64 - 0.03 μ g/mL. All tested compounds and antibiotic controls were applied solved in DMSO. Therefore, pure DMSO was tested in row A (solvent control). Numbers indicate μ L DMSO /100 μ L assay solution. Sample test wells in rows B-D: each compound (cmp) was tested in triplicate in concentration ranging from 128-0.06 μ g/mL. Blank = medium control ('LOW'); Neg CON = cells in medium without any supplementation ('HIGH').

7.7 *steABC* flanking regions

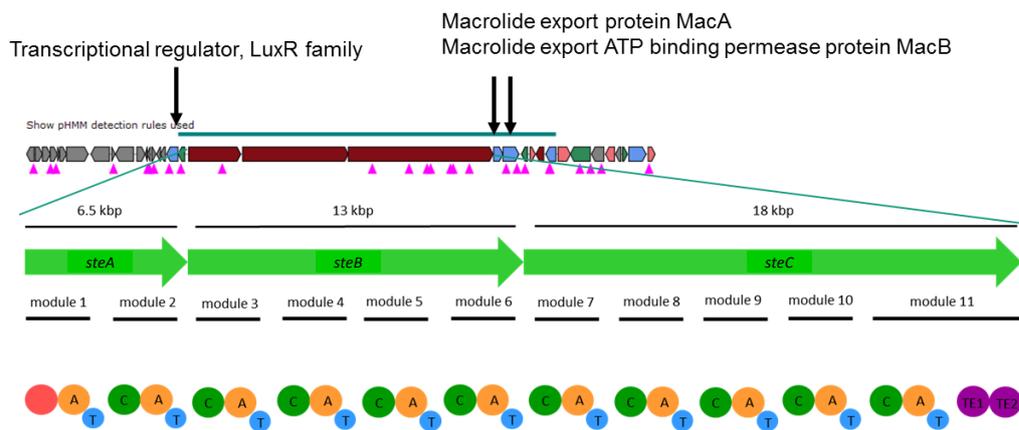


Figure S14: Biosynthetic gene cluster of Tensin/Steclisins. Upper part taken from Antismash annotation algorithm. Core genes *steABC* in red. Relevant flanking and accessory genes in blue. Middle part: Detailed representation of the core genes. Eleven modules are organized in three core genes.

7.8 Optimization of CLP production

7.8.1 Media variation

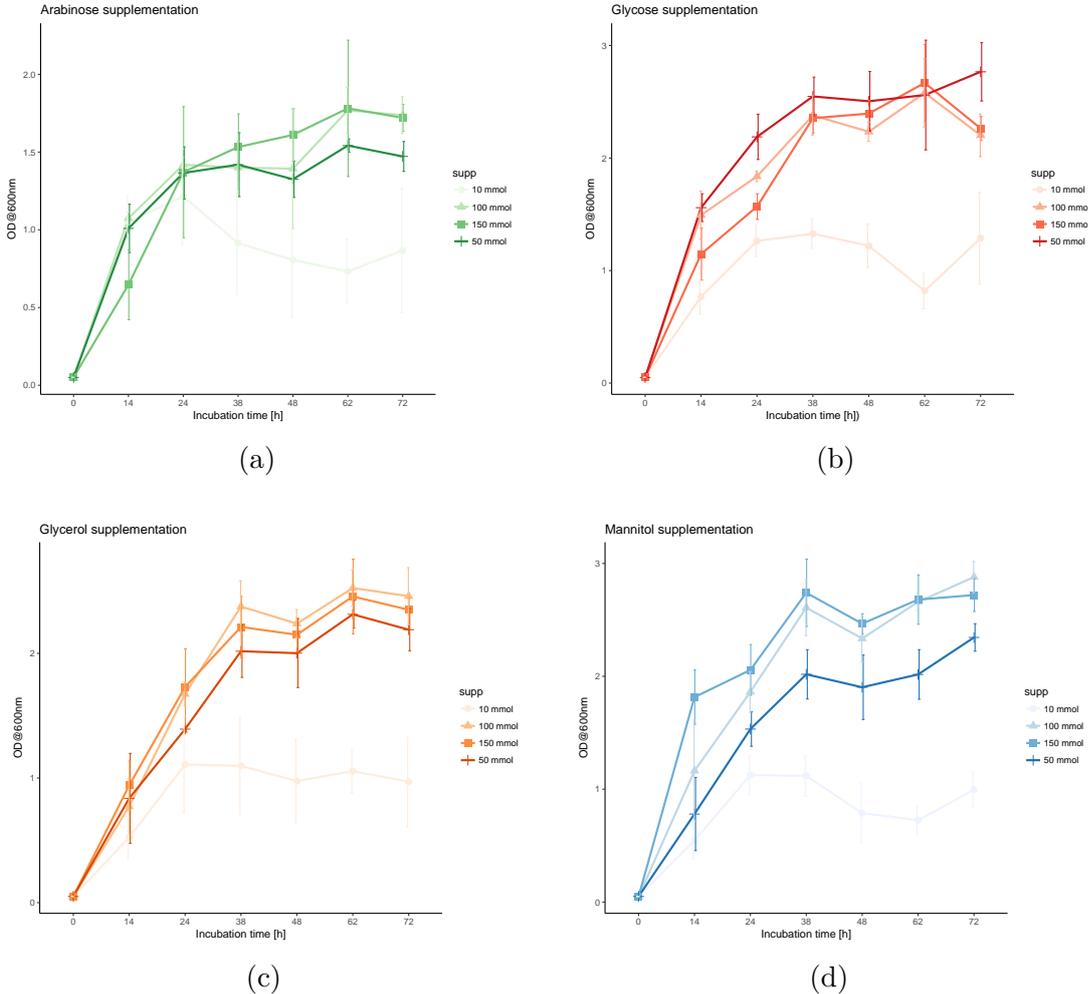


Figure S15: [Cell densities of FhG100052 cultured in media supplemented with different carbon sources and concentrations overtime] Cell densities of FhG100052 cultured in media supplemented with different carbon sources and concentrations over time

7.8.2 Gas exchange

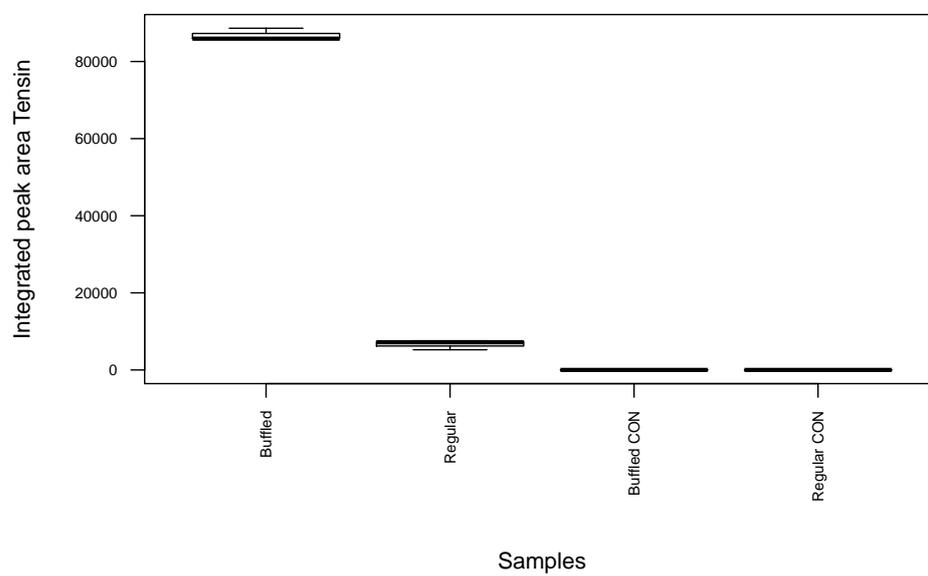


Figure S16: Integrated peak areas of EICs 1409.8495 $[M_2 + H]^+$ detected in FhG100052 extracts cultured in either regular or baffled flasks

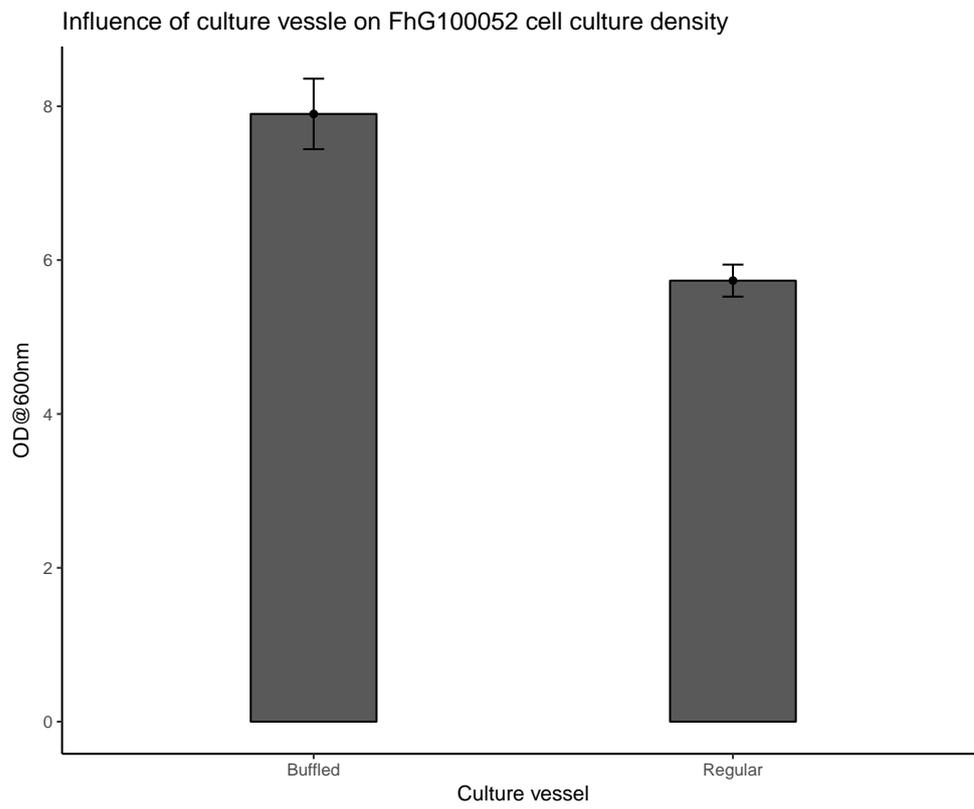


Figure S17: Cell densities of FhG100052 cultured in either regular or buffled flasks

7.8.3 Incubation period

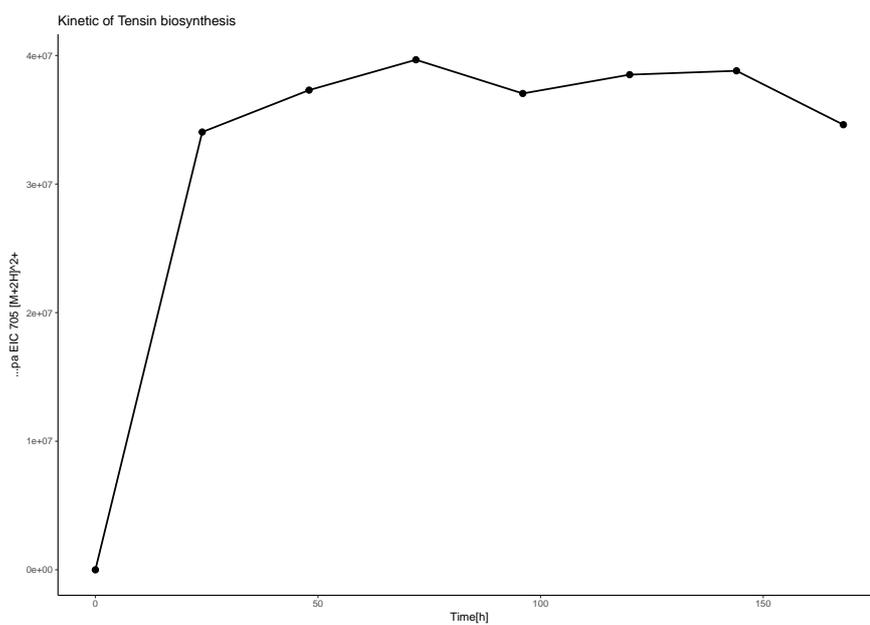


Figure S18: Kinetic of Tensin biosynthesis by FhG100052: Integrated peak areas of EICs 705.4288 $[M_2 + 2H]^{2+}$ detected in FhG100052 extracts over time

7.9 Stechlisins Marfey's Analysis

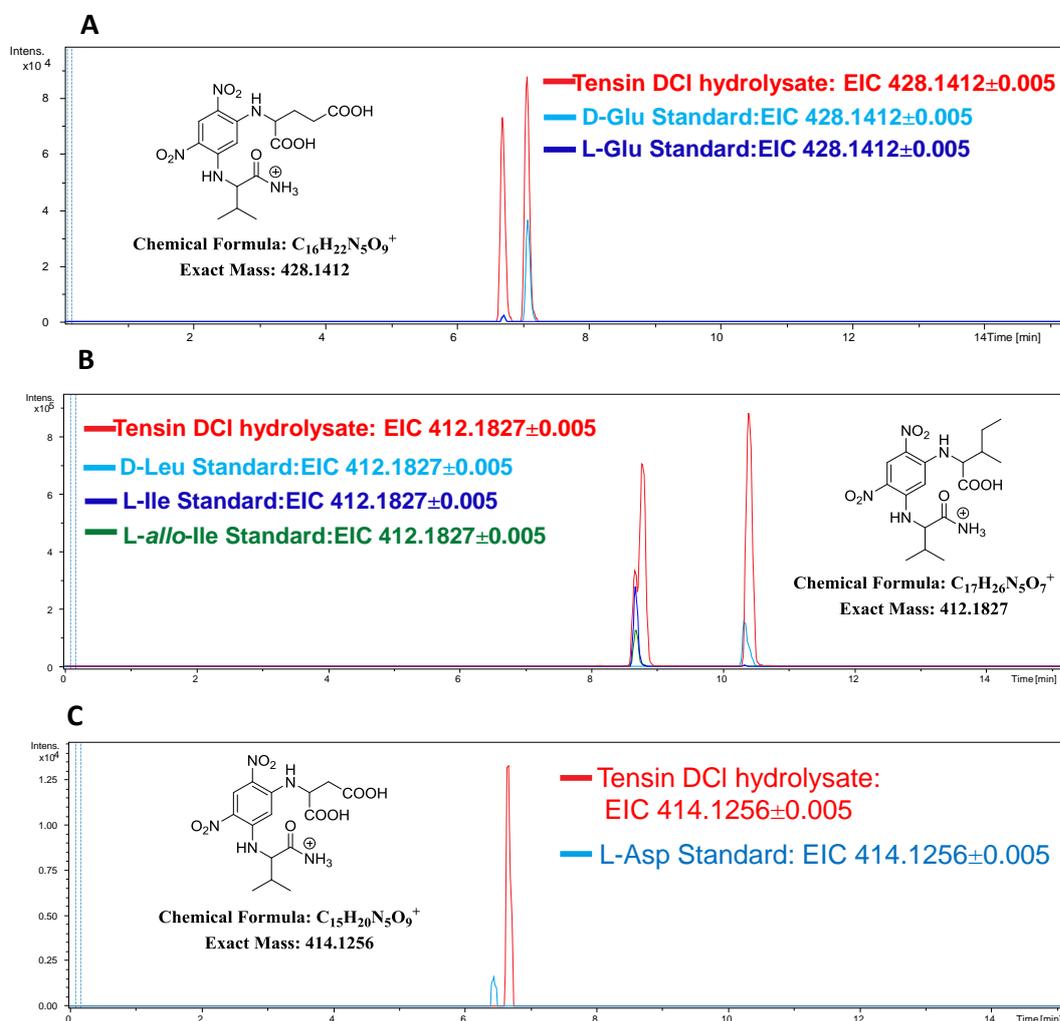


Figure S19: Comparison of the Marfey derivatization products of the Tensin DCI hydrolysate and amino acid standards. A: L-Glutamic acid and D-Glutamic acid standards. Note: D-Glutamine present in the molecule is entirely converted to D-Glutamic acid during the acid hydrolysis, therefore showing only the signal corresponding to the acid derivative. B: Tensin DCI hydrolysate and D-Leucine, L-Leucine, L-Isoleucine, L-allo-Isoleucine and D-allo-Isoleucine standards. C: L-Aspartic acid standard.

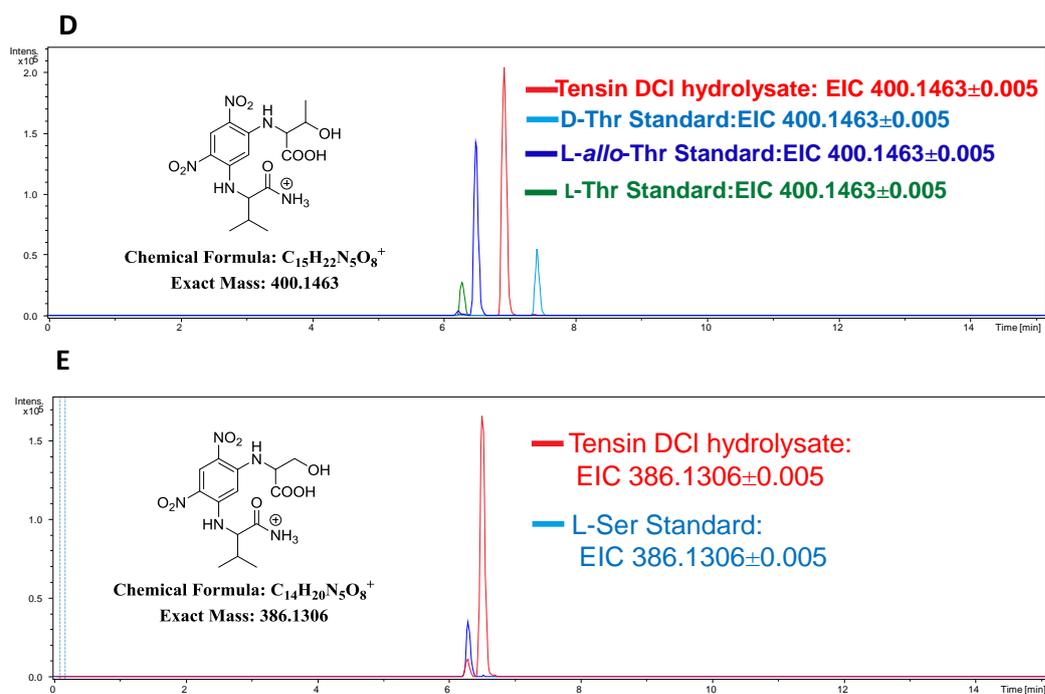


Figure S20: Comparison of the Marfey derivatization products of the Tensin DCl hydrolysate and amino acid standards continued. D: Comparison of the Tensin DCl hydrolysate and D-threonine, L-*allo*-threonine and L-Threonine standards. E: Comparison with L-Serine standard

7.10 Stechlisins MS/MS fragmentation

Table S1: Overview Stechlisins. In total 34 different CLPs were observed in LC-MS/MS records of FhG100052 extracts of which six compounds were isolated. Furthermore, eleven putative structures were assigned via MS/MS fragmentation analysis, leaving 17 compounds unexplained. Compounds are sorted according to their molecular formula and state of investigation. Numbers indicate the number of different isomers per formula and investigation status.

Molecular formula	m/z [$M + 2H$] ²⁺ (Δ ppm)	# of isomers			
		Observed	Isolated	MS/MS predicted	unknown
C ₆₄ H ₁₁₀ N ₁₂ O ₂₀	684.4060 (-1.1)	1	0	1	0
C ₆₅ H ₁₁₂ N ₁₂ O ₂₀	691.4140 (-1.3)	4	1	2	1
C ₆₆ H ₁₁₄ N ₁₂ O ₂₀	691.4214 (-0.8)	5	1	4	0
C ₆₇ H ₁₁₆ N ₁₂ O ₂₀	705.4293 (-0.9)	7	2	1	4
C ₆₈ H ₁₁₈ N ₁₂ O ₂₀	712.4370 (-0.9)	5	1	2	2
C ₆₉ H ₁₂₀ N ₁₂ O ₂₀	719.4458 (-2.0)	5	1	0	4
C ₆₉ H ₁₁₉ N ₁₁ O ₂₁	719.9375 (-1.8)	2	0	1	1
C ₇₀ H ₁₂₂ N ₁₂ O ₂₀	726.4227 (-0.8)	2	0	0	2
C ₆₉ H ₁₁₈ N ₁₂ O ₂₁	726.4337 (0.4)	3	0	0	3
	Total	34	6	11	17

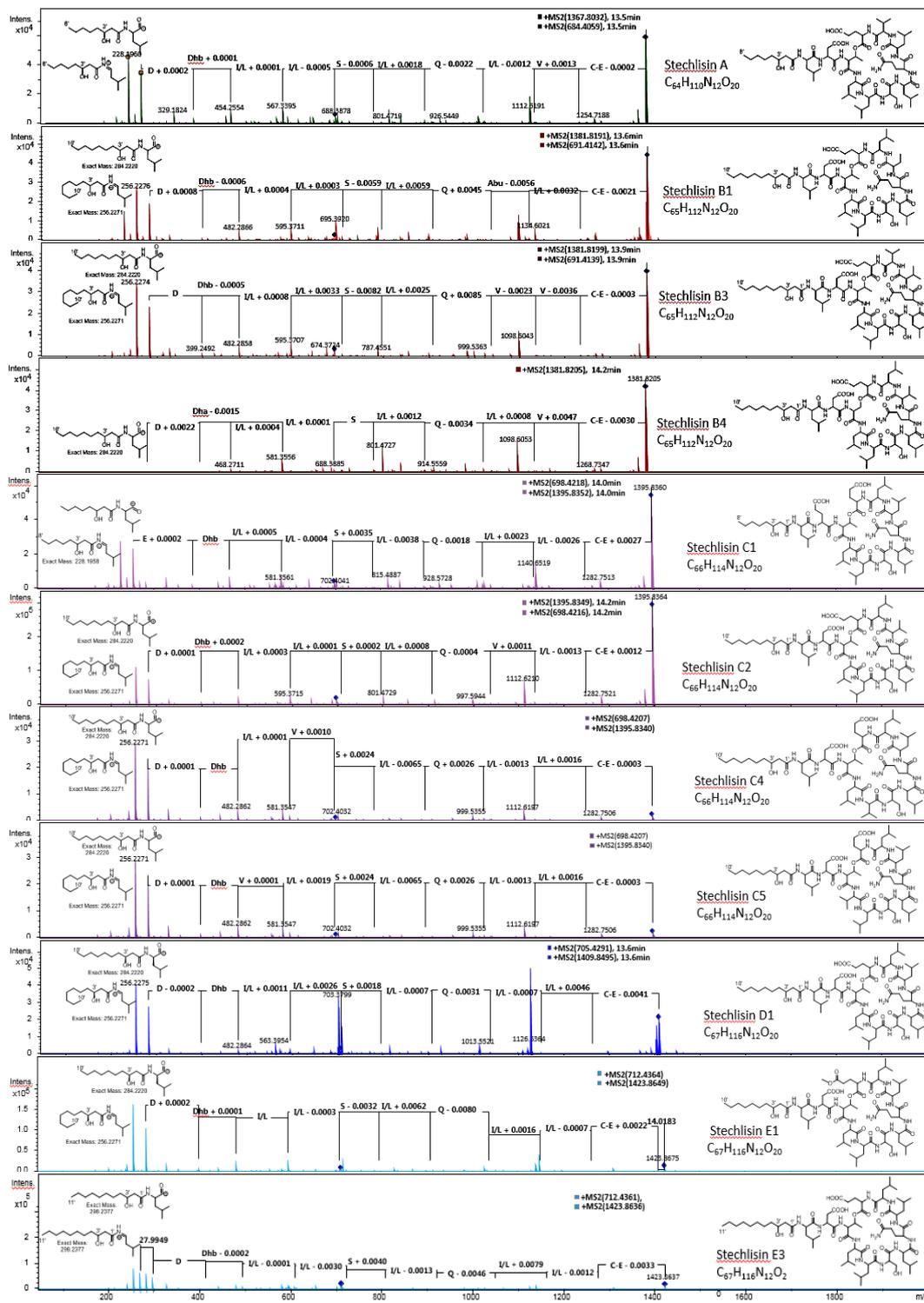


Figure S21: Proposed structures based on MS/MS

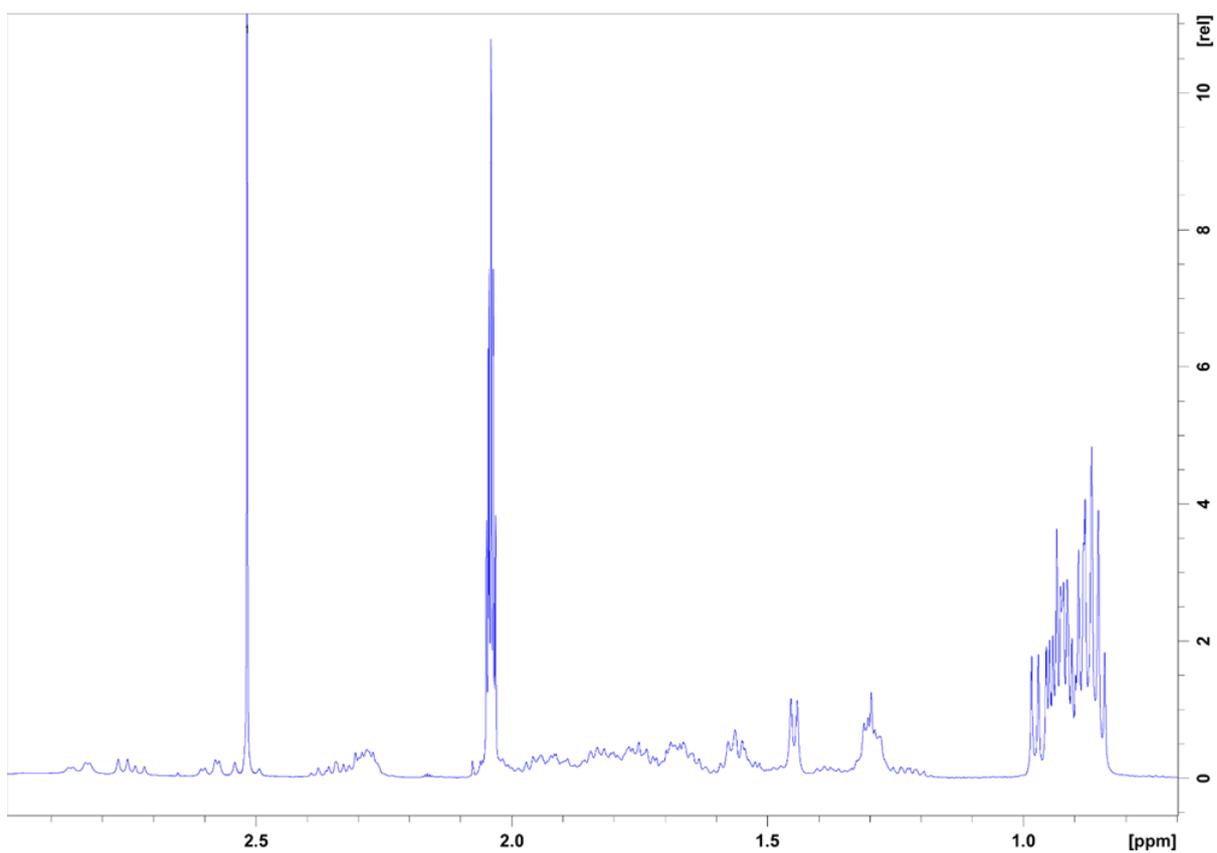


Figure S23: ^1H NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum zoom 0.8-3.5 ppm of Stechlisin B2

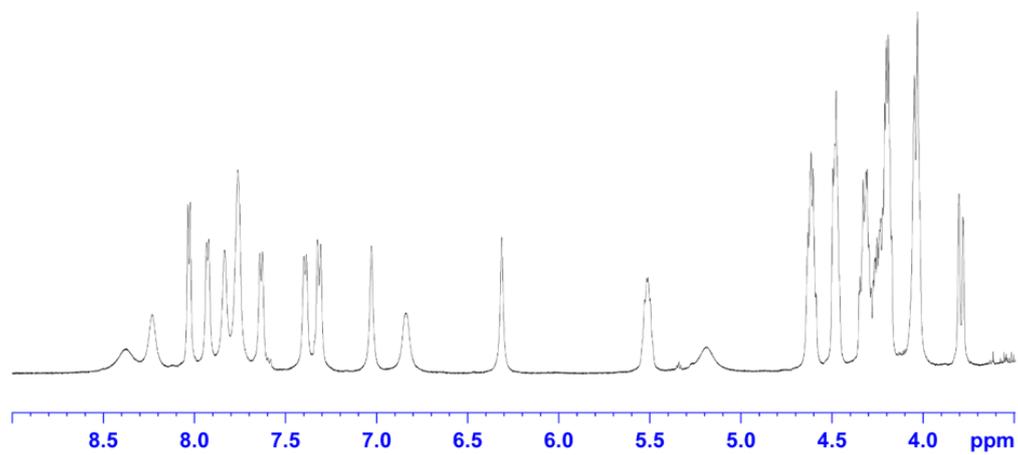


Figure S24: ^1H NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum zoom 3.5-9.0 ppm of Stechlisin B2

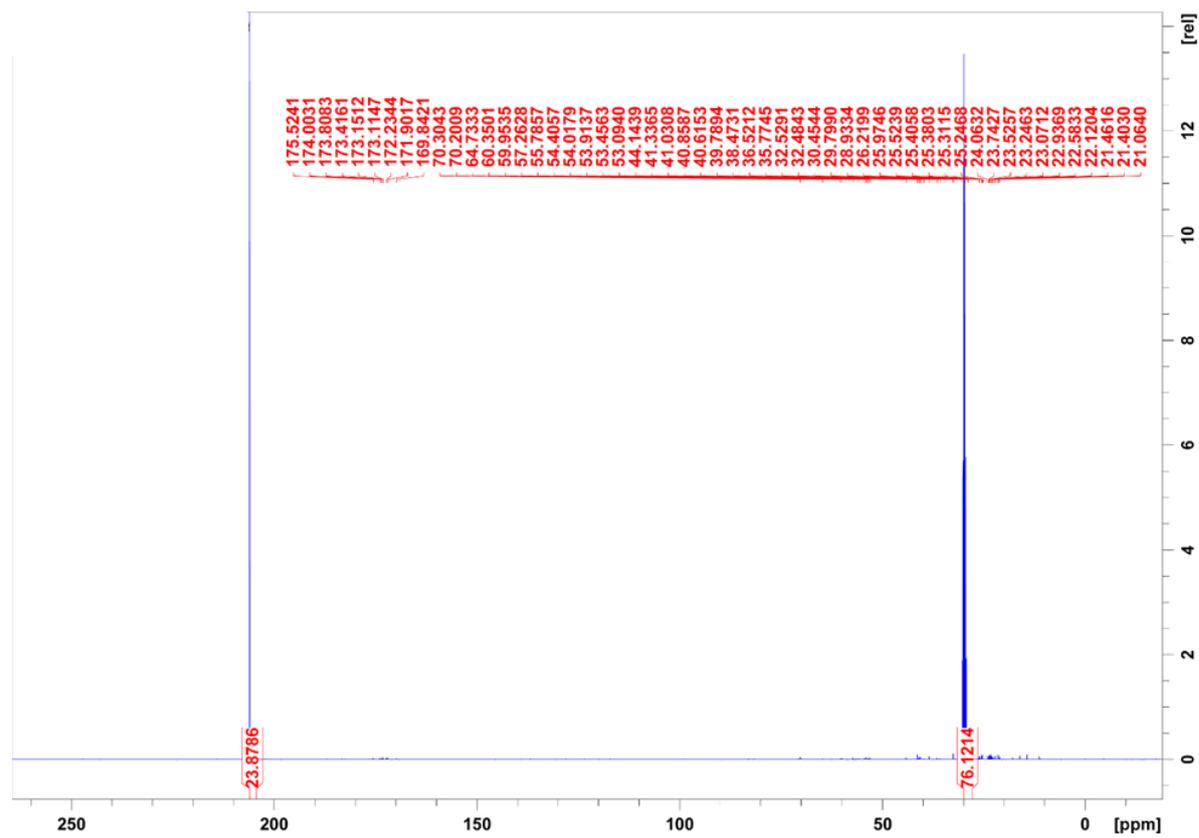


Figure S25: ^{13}C NMR (125 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum of Stechlisin B2

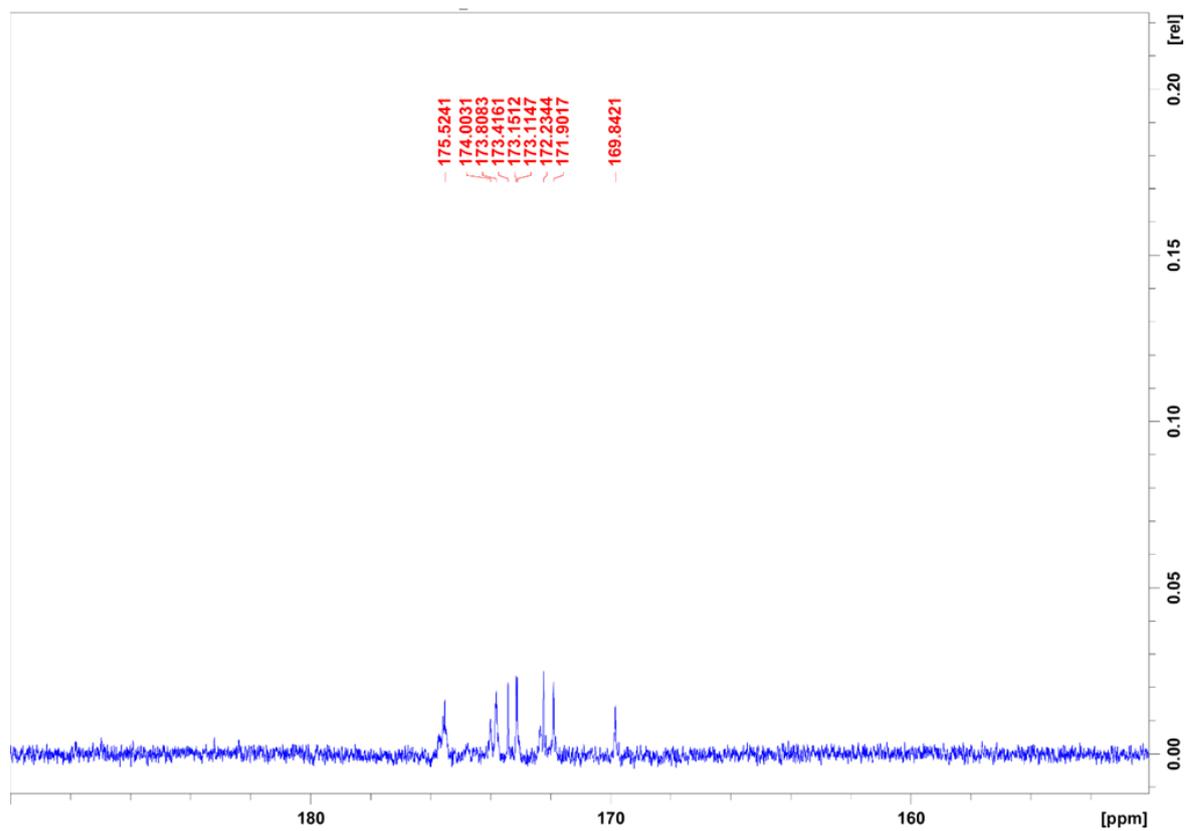


Figure S26: ¹³C NMR (125 MHz, (CD₃)₂CO) spectrum of the carbonyl region of Stechlisin B2

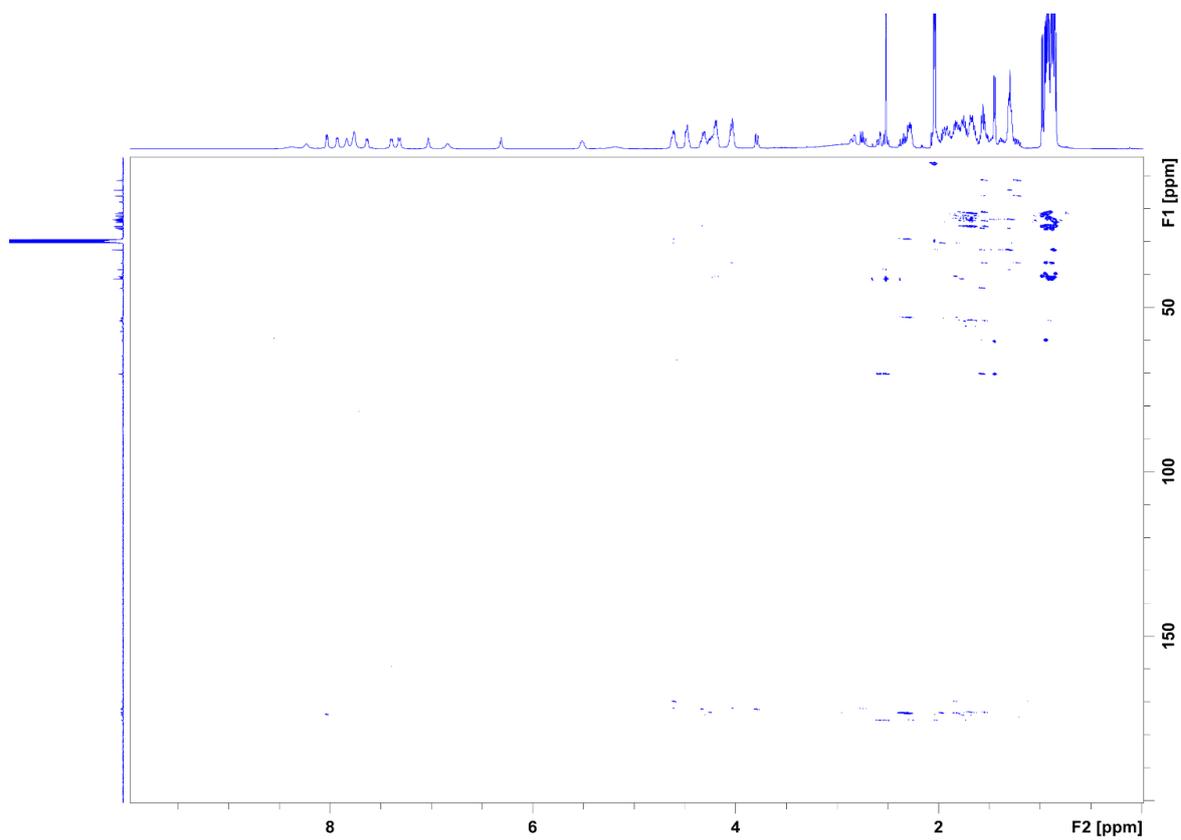


Figure S27: HMBC NMR (500 MHz, (CD₃)₂CO) of Stechlisin B2

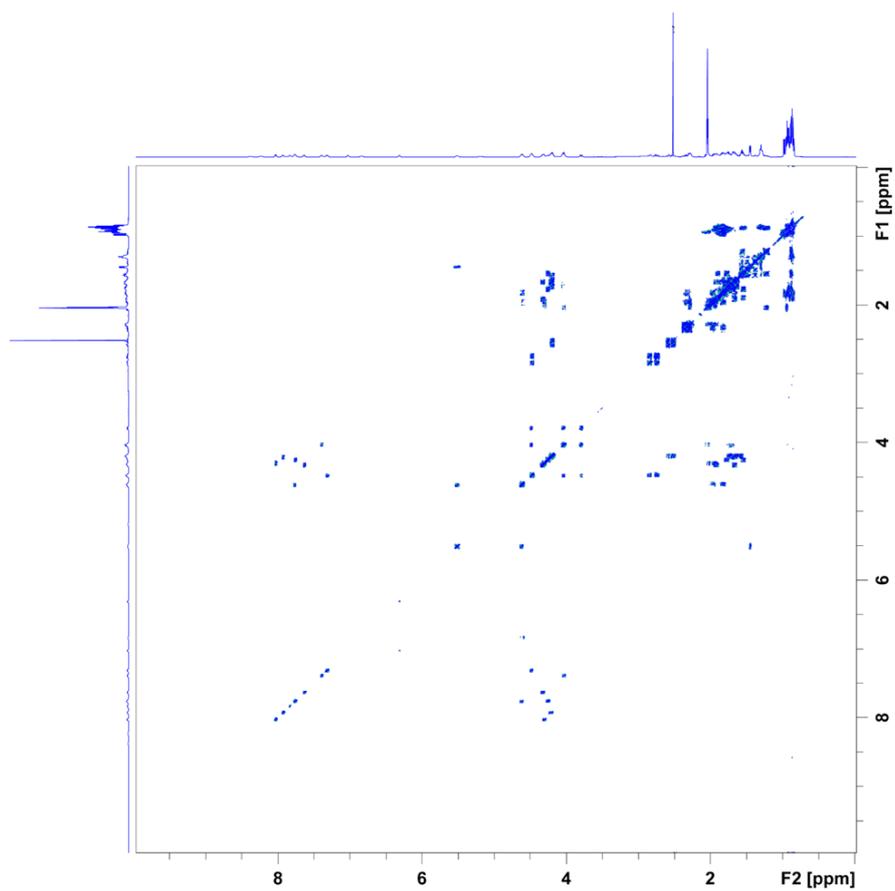


Figure S28: COSY NMR (500 MHz, (CD₃)₂CO) of Stechlisin B2

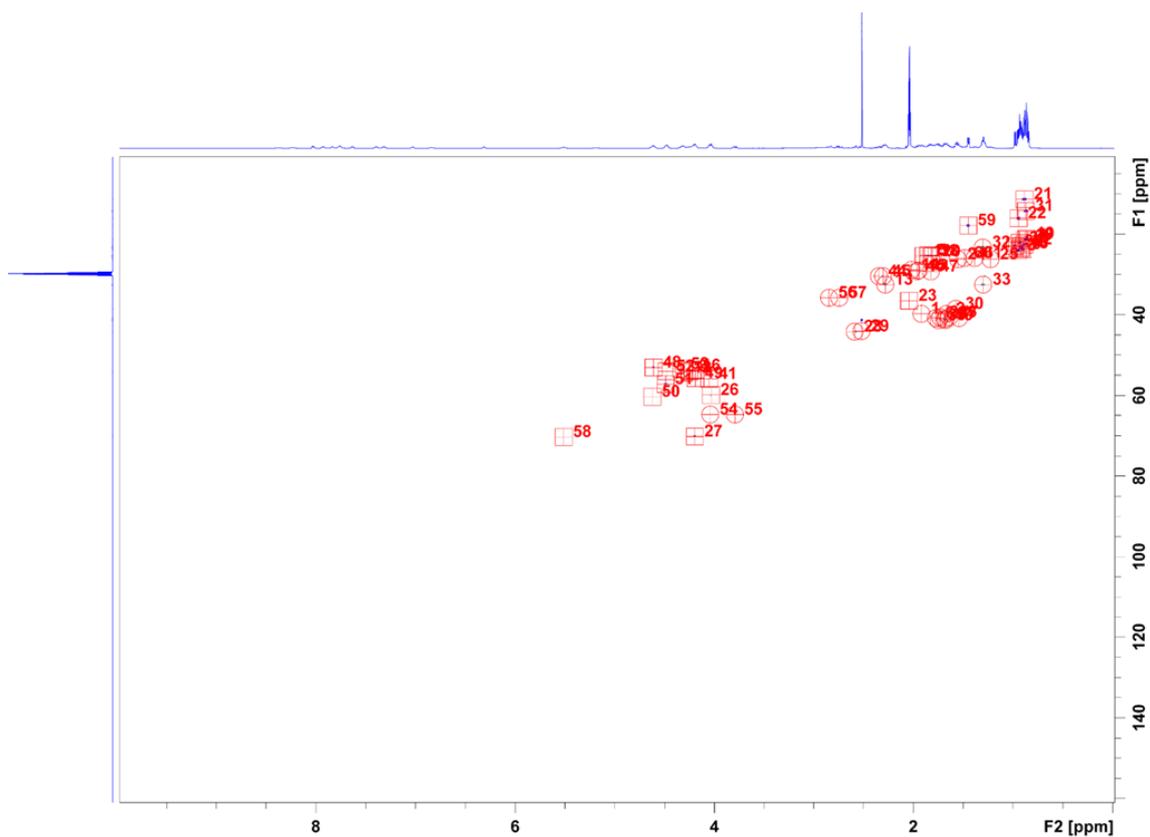


Figure S29: HSQC NMR (500 MHz, (CD₃)₂CO) of Stechlisin B2

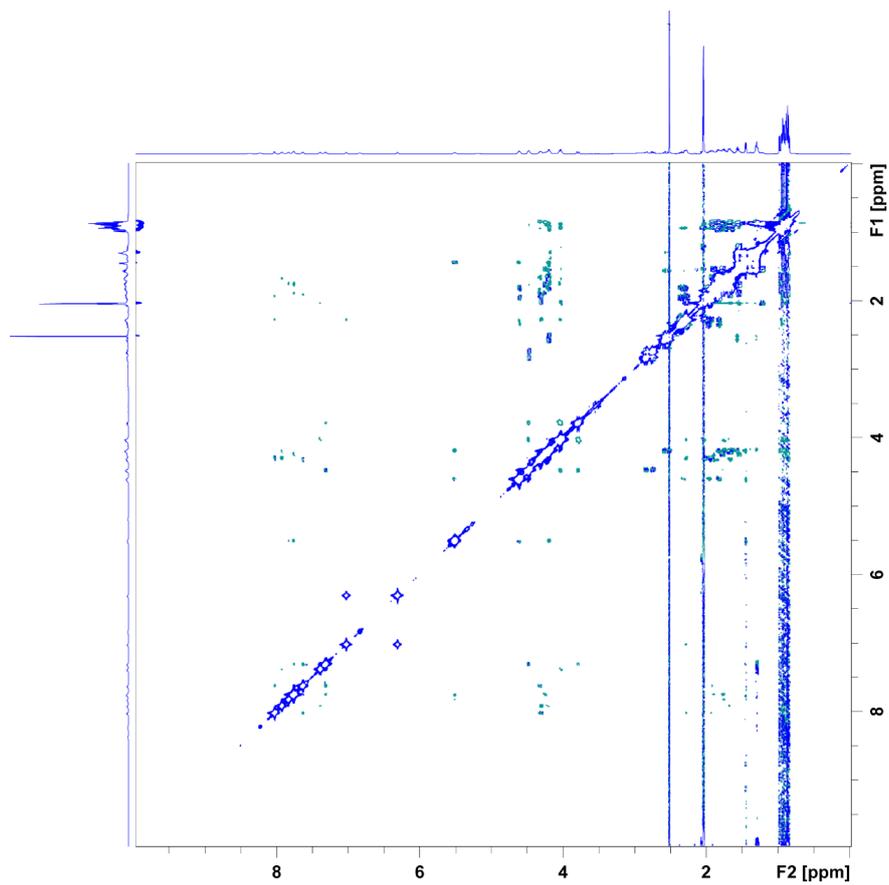


Figure S30: ROESY NMR (500 MHz, (CD₃)₂CO) of Stechlisin B2

7.11.2 Stechlisin C3

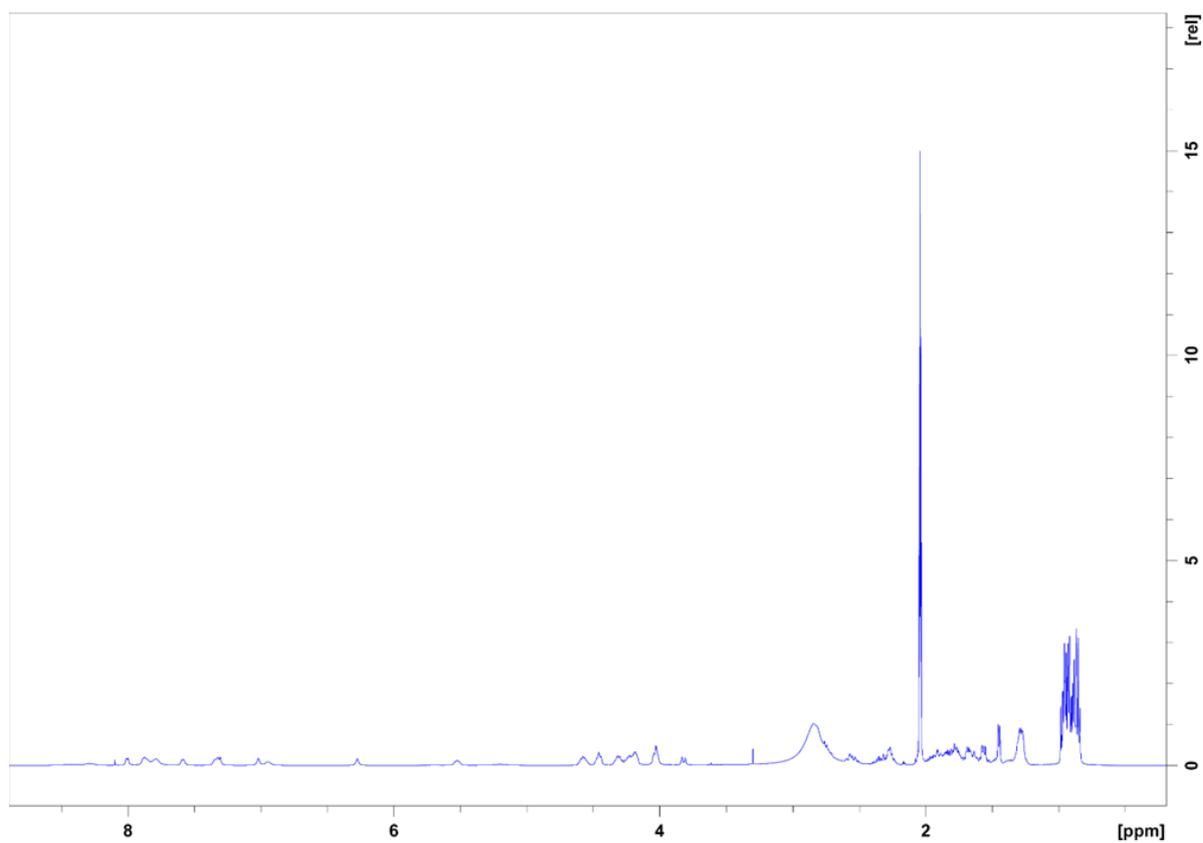


Figure S31: ¹H NMR (500 MHz, (CD₃)₂CO) spectrum of Stechlisin C3

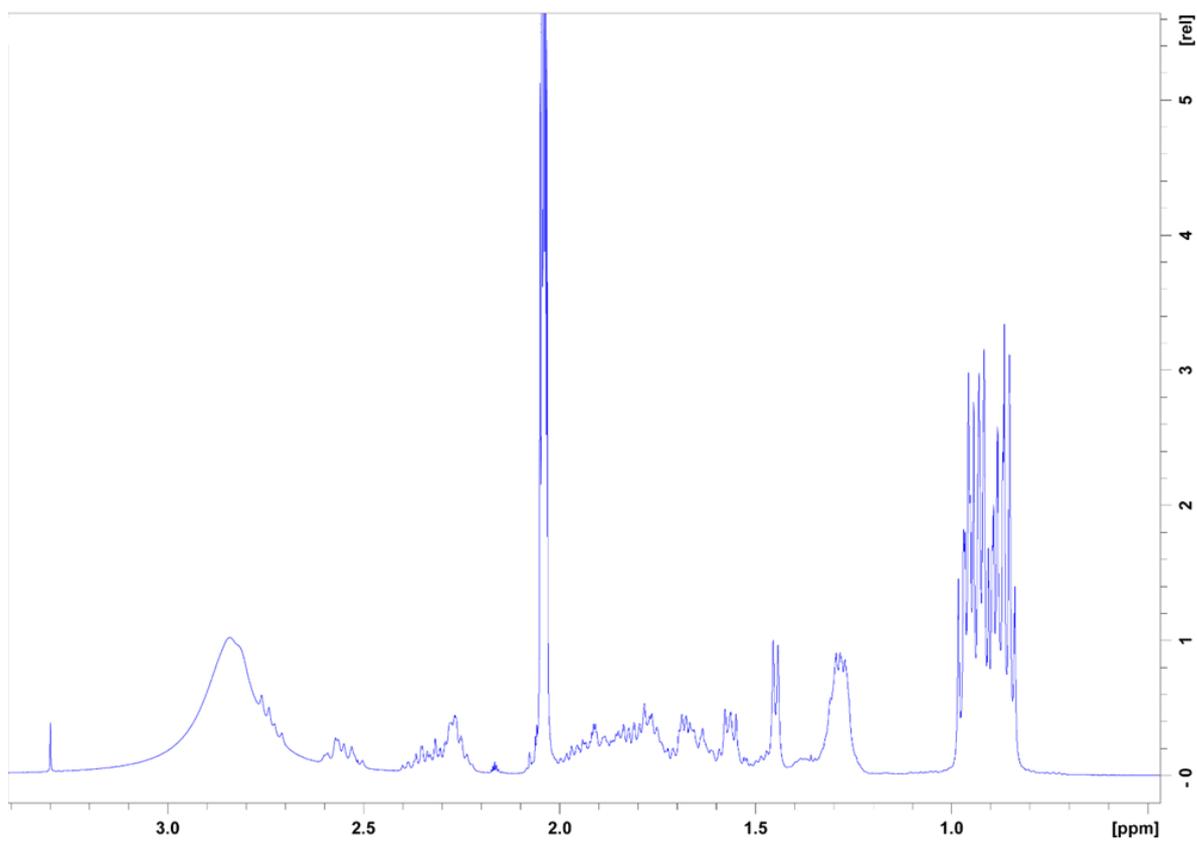


Figure S32: ^1H NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum zoom 0.8-3.5 ppm of Stechlisin C3

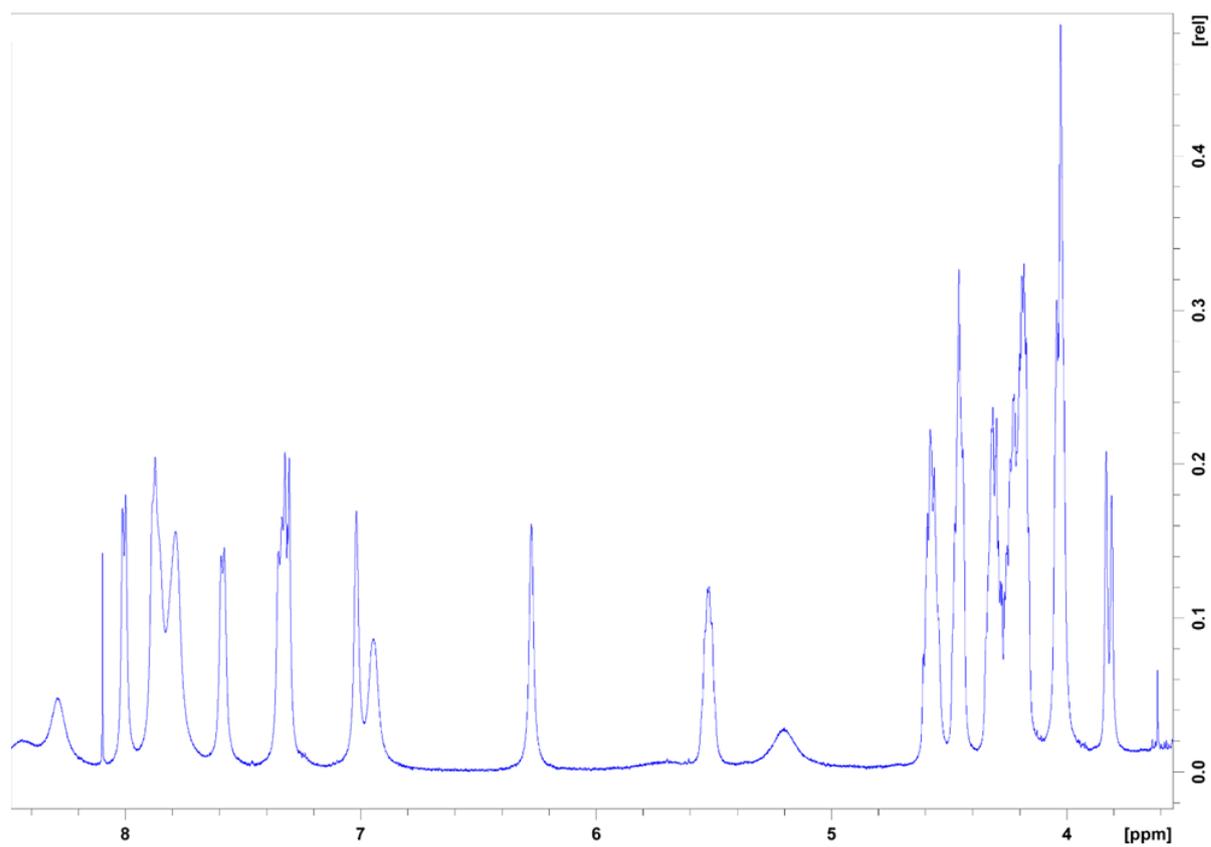


Figure S33: ^1H NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum zoom 3.5-8.5 ppm of Stechlisin C3

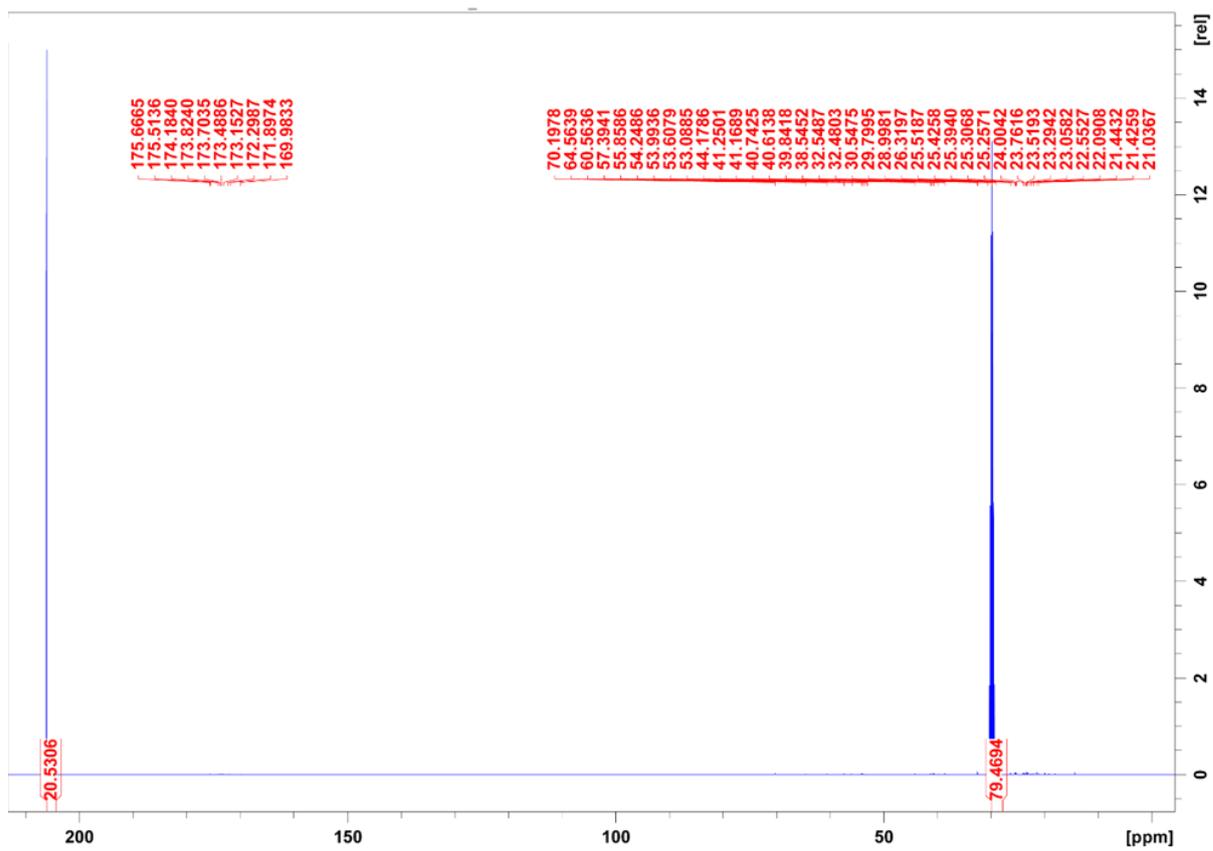


Figure S34: ^{13}C NMR (125 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum of Stechlisin C3

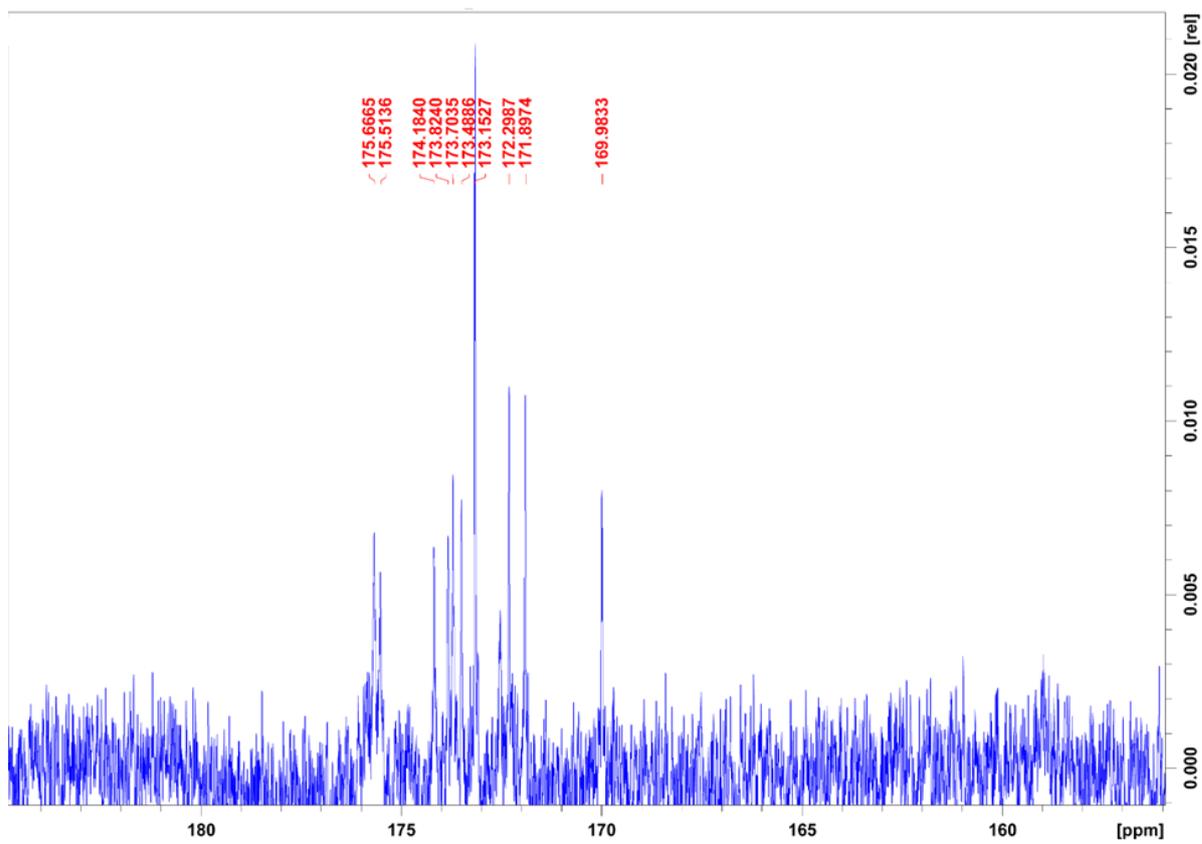


Figure S35: ¹³C NMR (125 MHz, (CD₃)₂CO) spectrum of the carbonyl region of Stechlisin C3

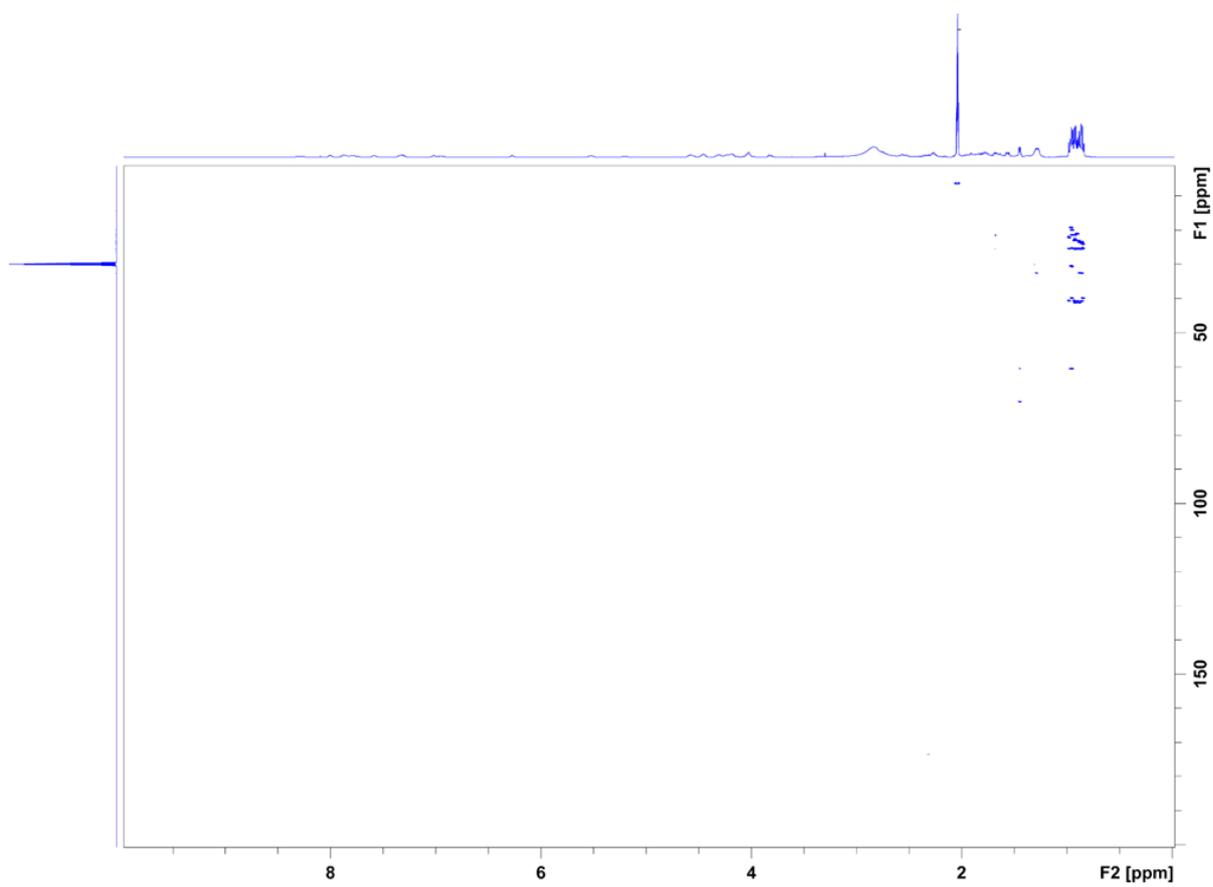


Figure S36: HMBC NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum of Stechlisin C3

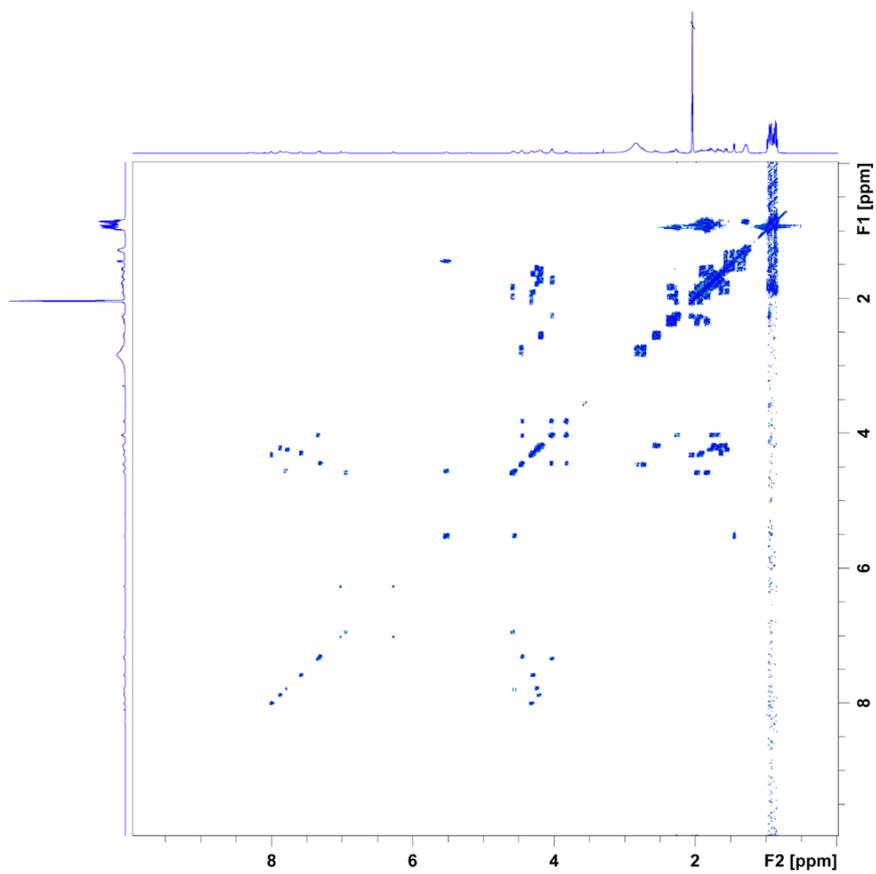


Figure S37: COSY NMR (500 MHz, (CD₃)₂CO) spectrum of Stechlisin C3

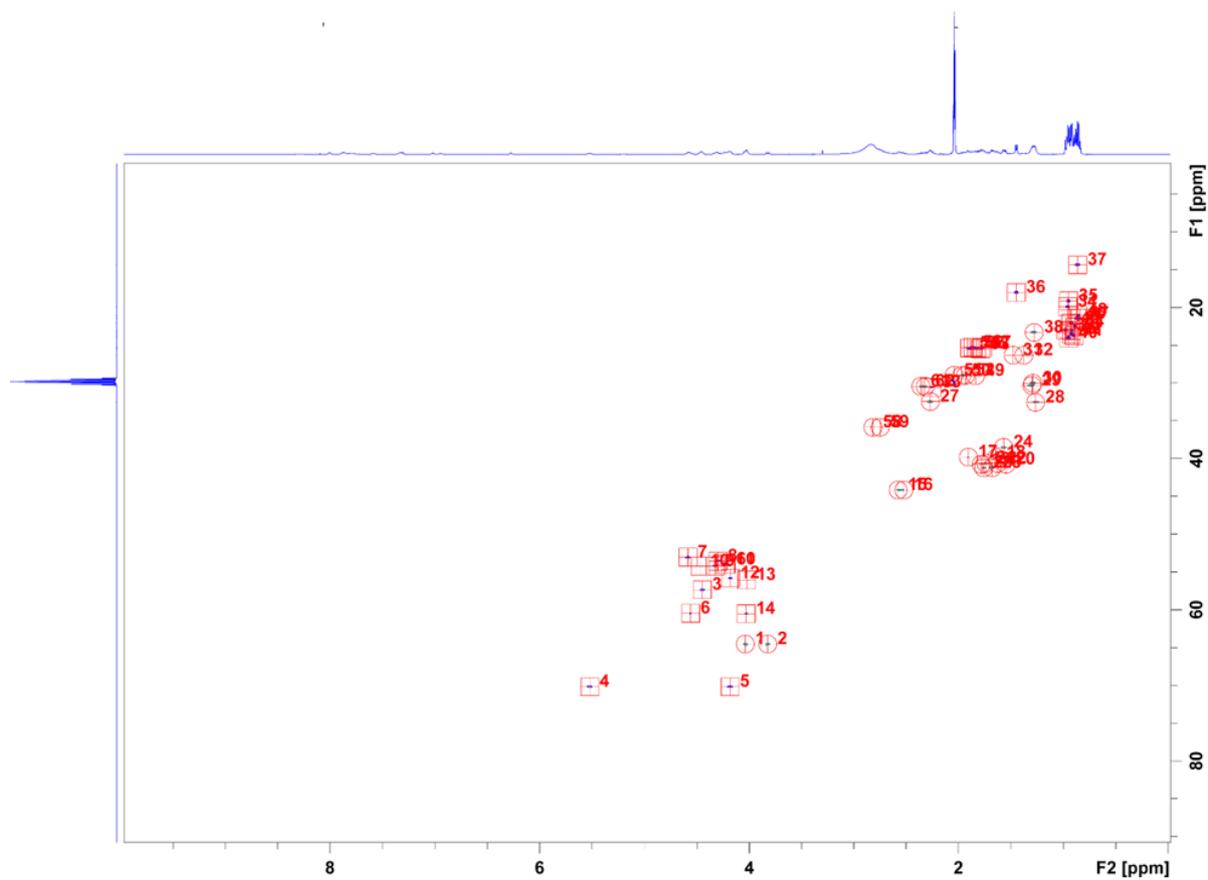


Figure S38: HSQC NMR (500 MHz, (CD₃)₂CO) spectrum of Stechlisin C3

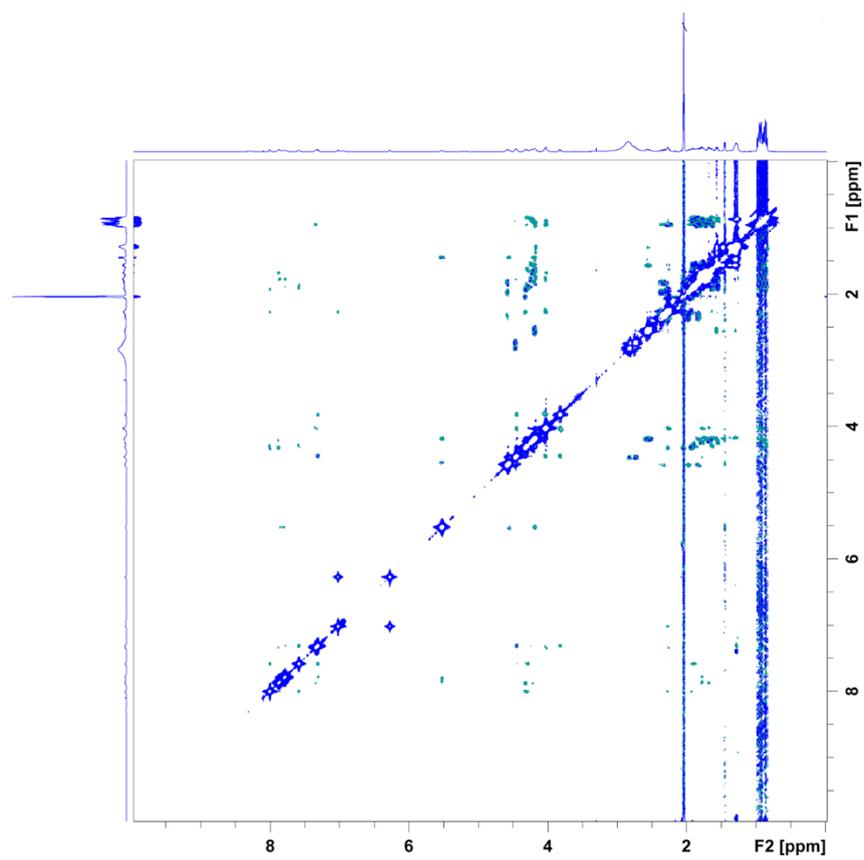


Figure S39: ROESY NMR (500 MHz, (CD₃)₂CO) spectrum of Stechlinin C3

7.11.3 Stechlisin D3

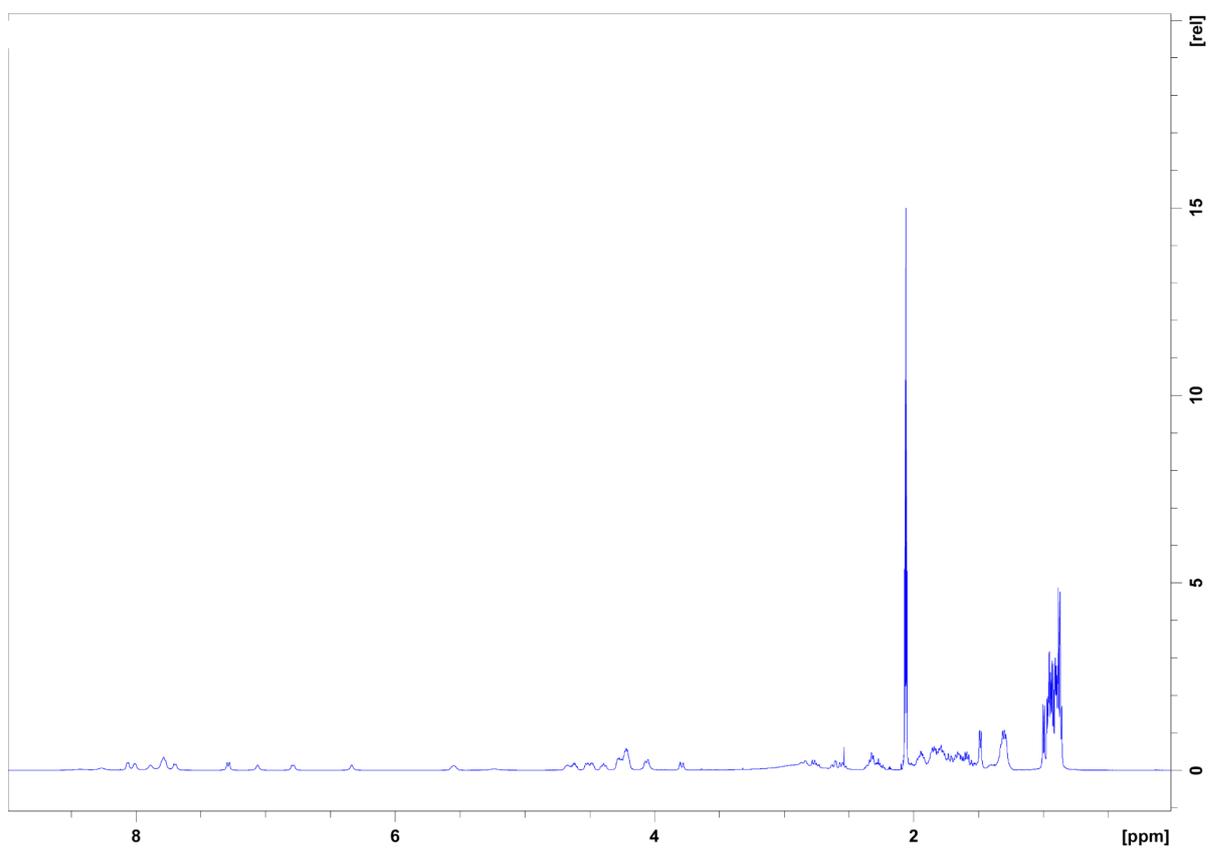


Figure S40: ¹H NMR (500 MHz, (CD₃)₂CO) spectrum of Stechlisin D3

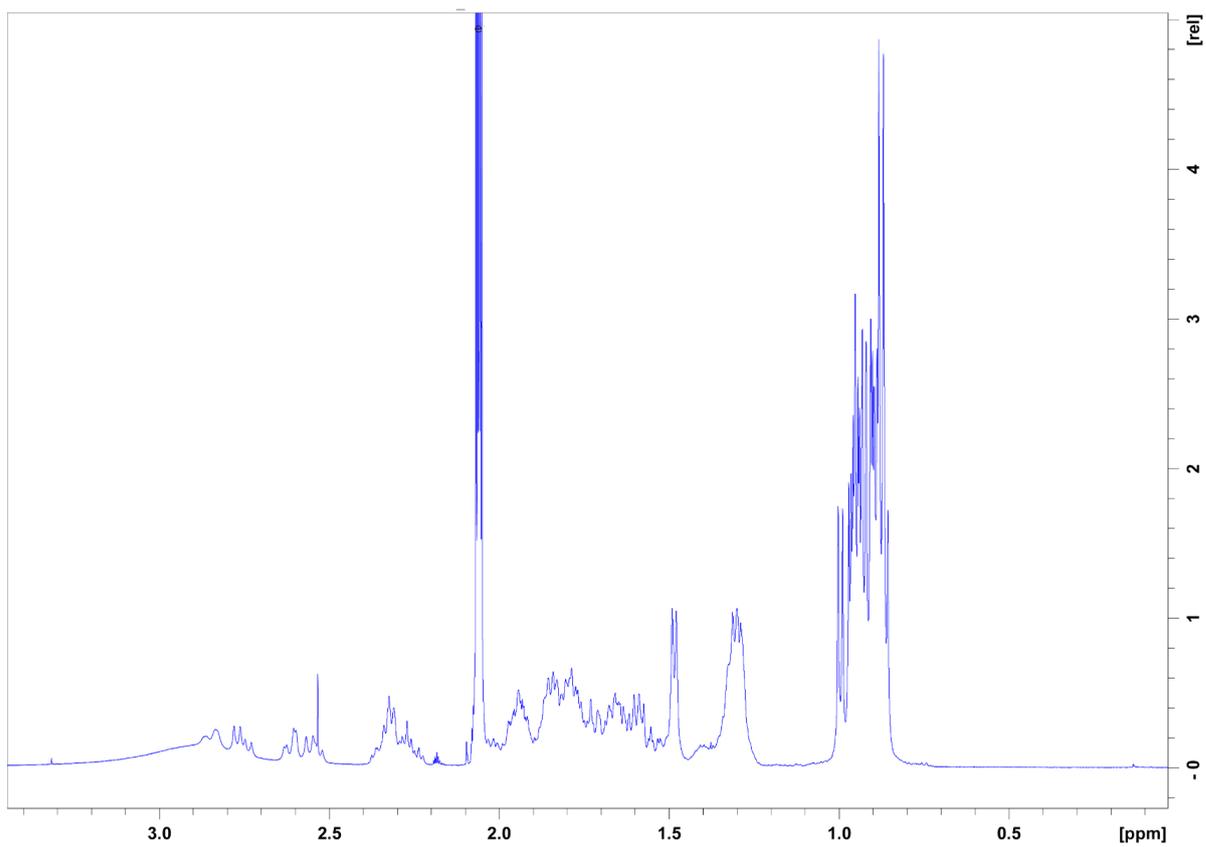


Figure S41: ¹¹H NMR (500 MHz, (CD₃)₂CO) spectrum zoom 0.8-3.5 ppm of Stechlisin D3

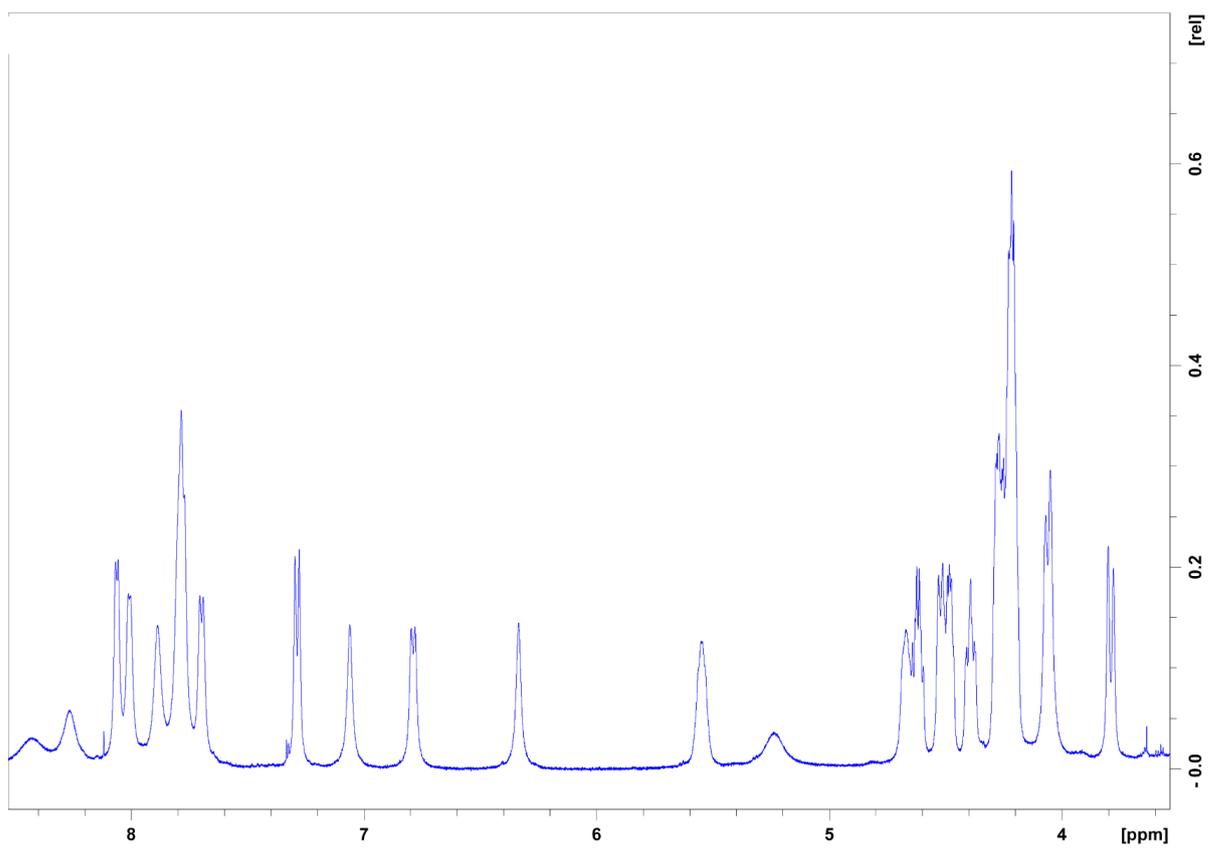


Figure S42: ^1H NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum zoom 3.5-8.5 ppm of Stechlisin D3

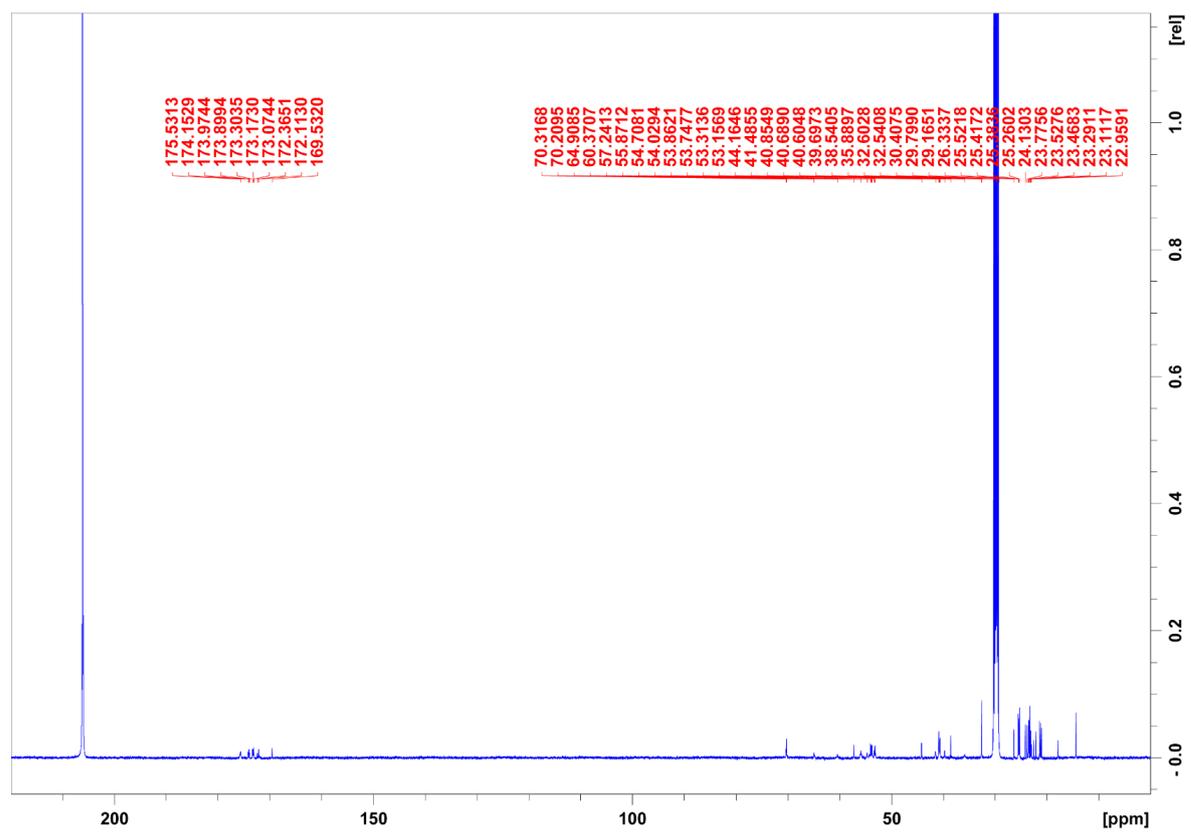


Figure S43: ^{13}C NMR (125 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum of Stechlisin D3

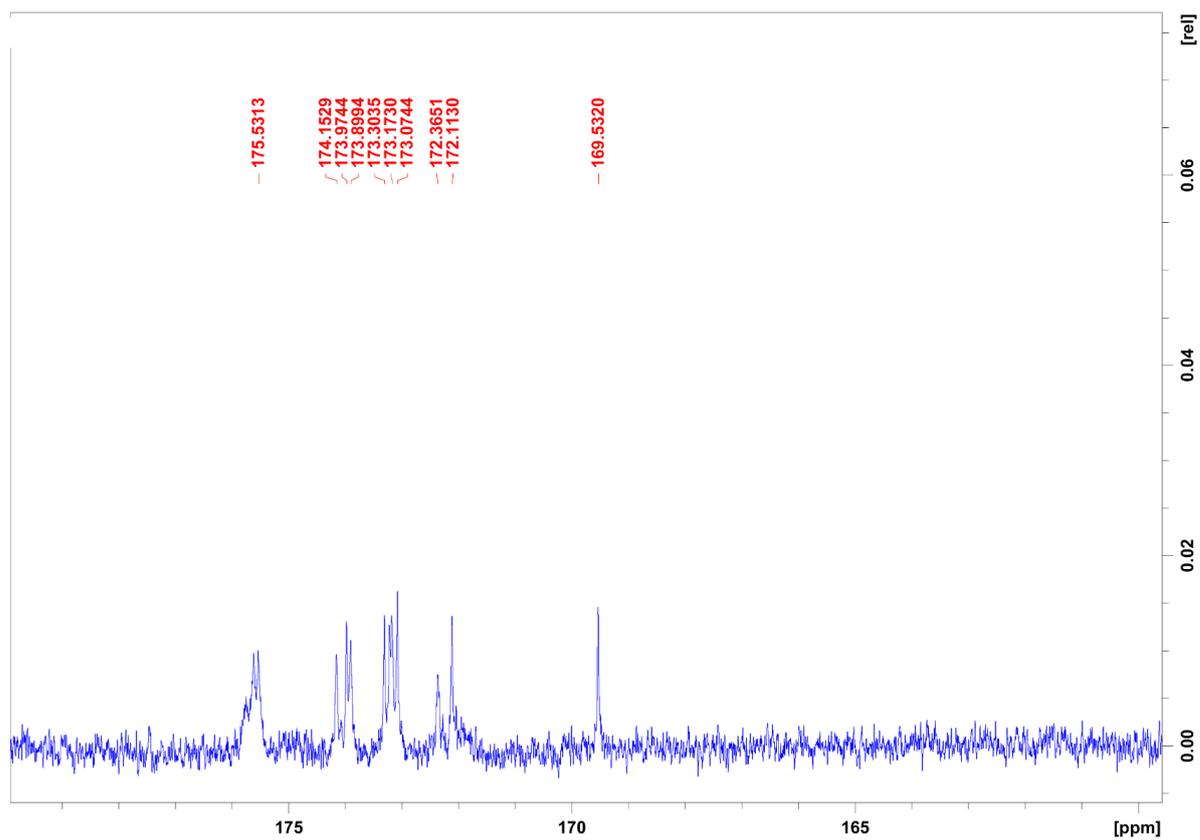


Figure S44: ¹³C NMR (125 MHz, (CD₃)₂CO) spectrum of the carbonyl region of Stechlisin D3

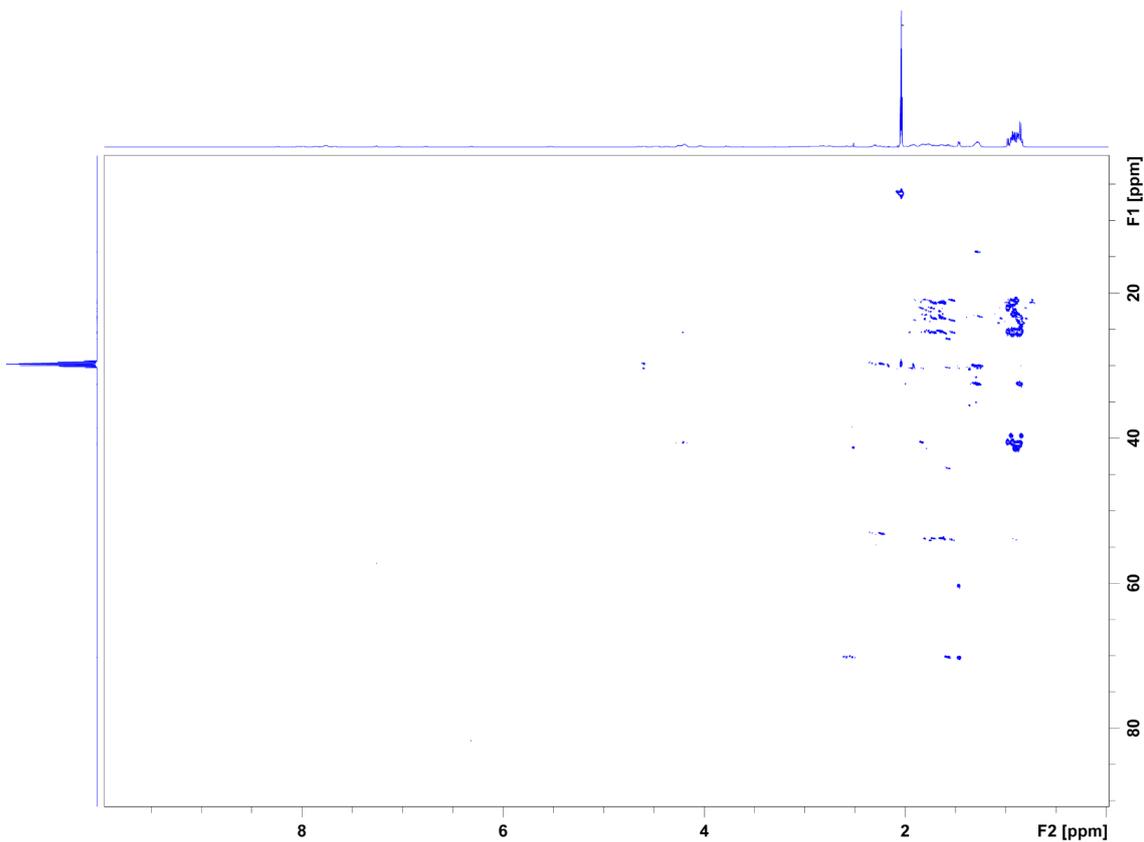


Figure S45: HMBC NMR (500 MHz, (CD₃)₂CO) spectrum of Stechlisin D3

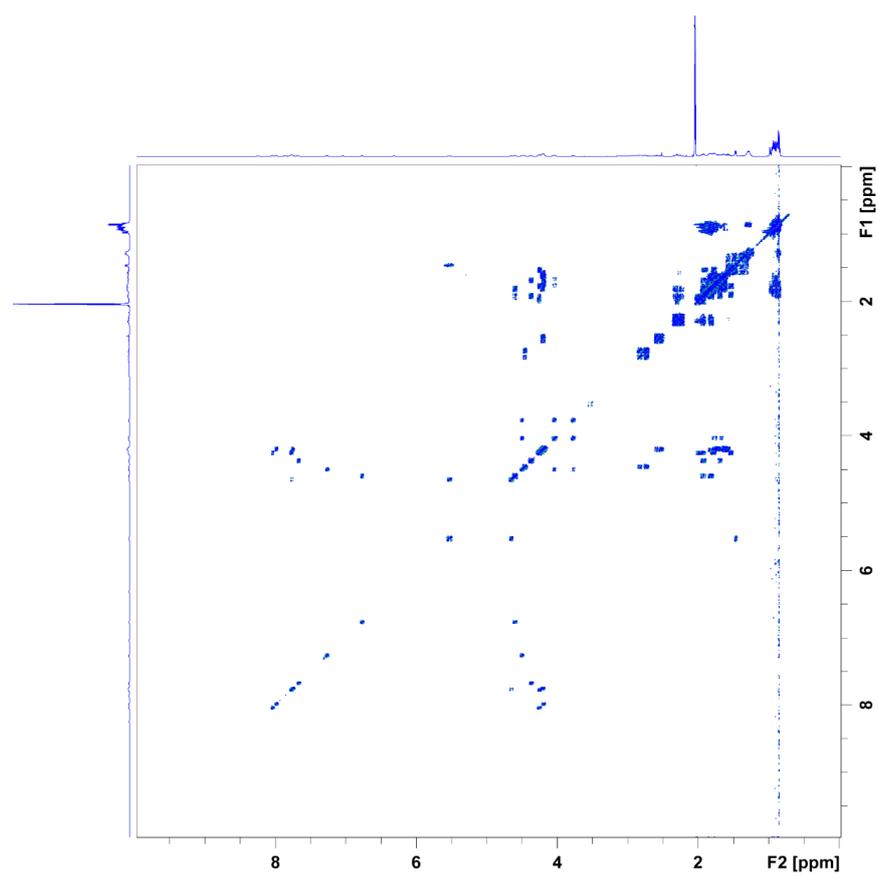


Figure S46: COSY NMR (500 MHz, (CD₃)₂CO) spectrum of Stechlisin D3

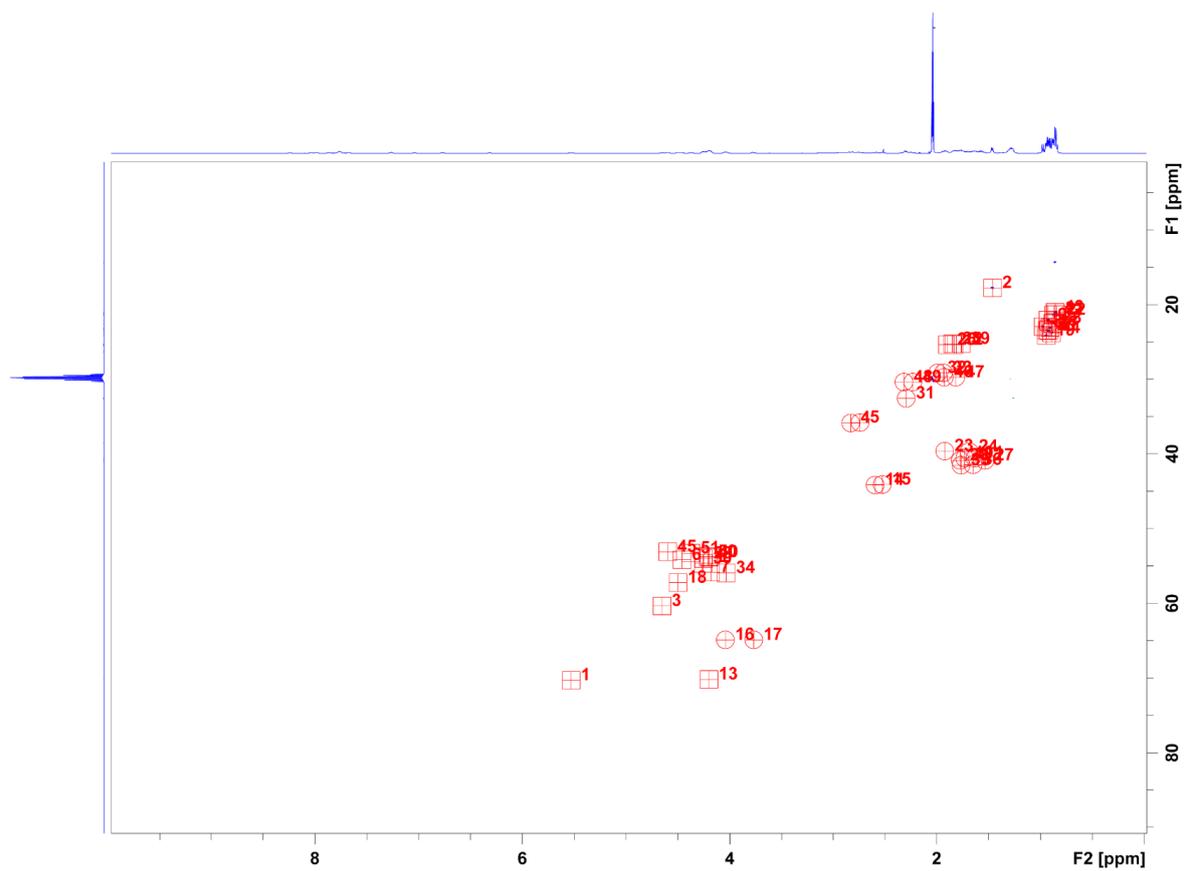


Figure S47: HSQC NMR (500 MHz, (CD₃)₂CO) spectrum of Stechlisin D3

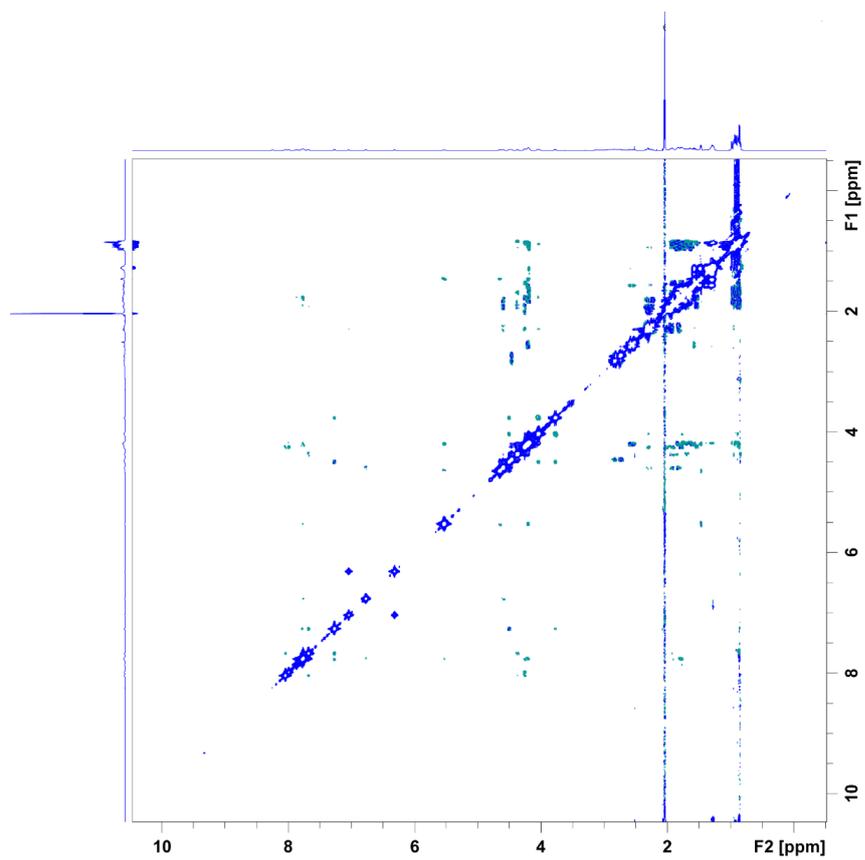


Figure S48: ROESY NMR (500 MHz, (CD₃)₂CO) spectrum of Stechlisin D3

7.11.4 Tensin

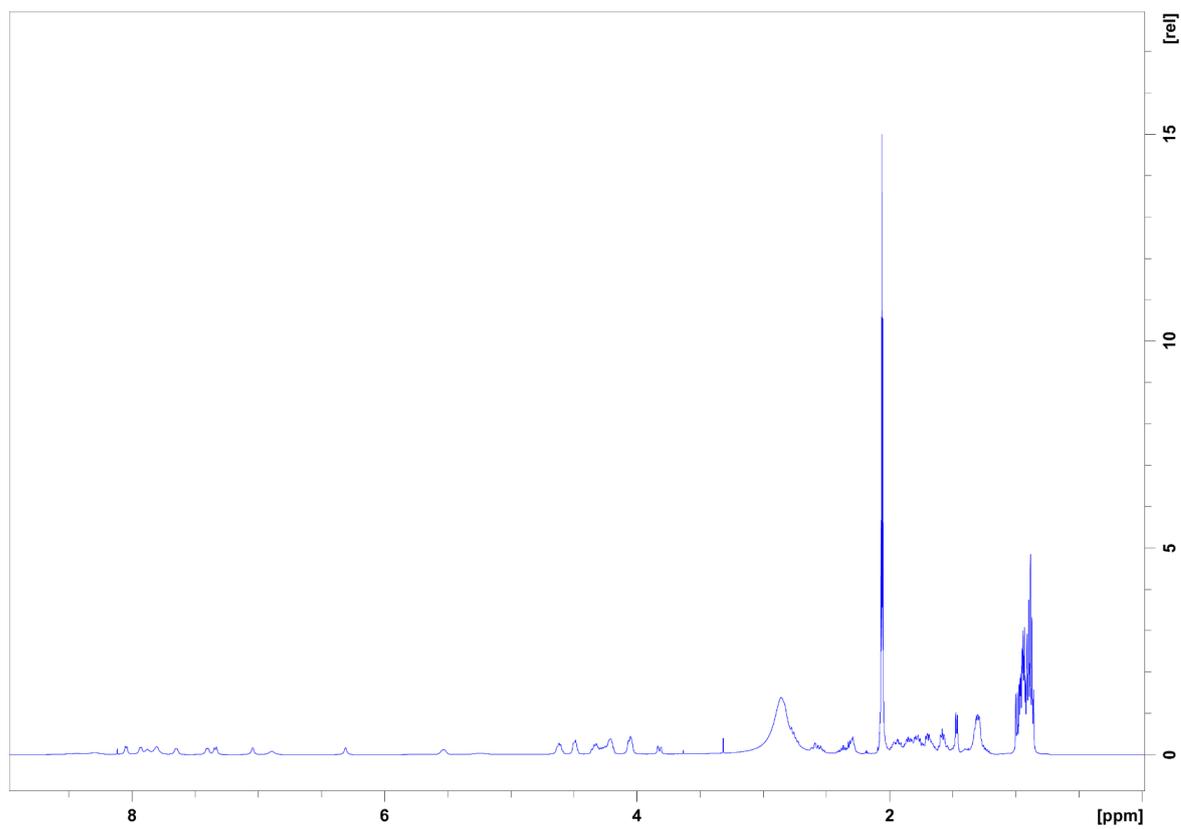


Figure S49: ^1H NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum of Tensin

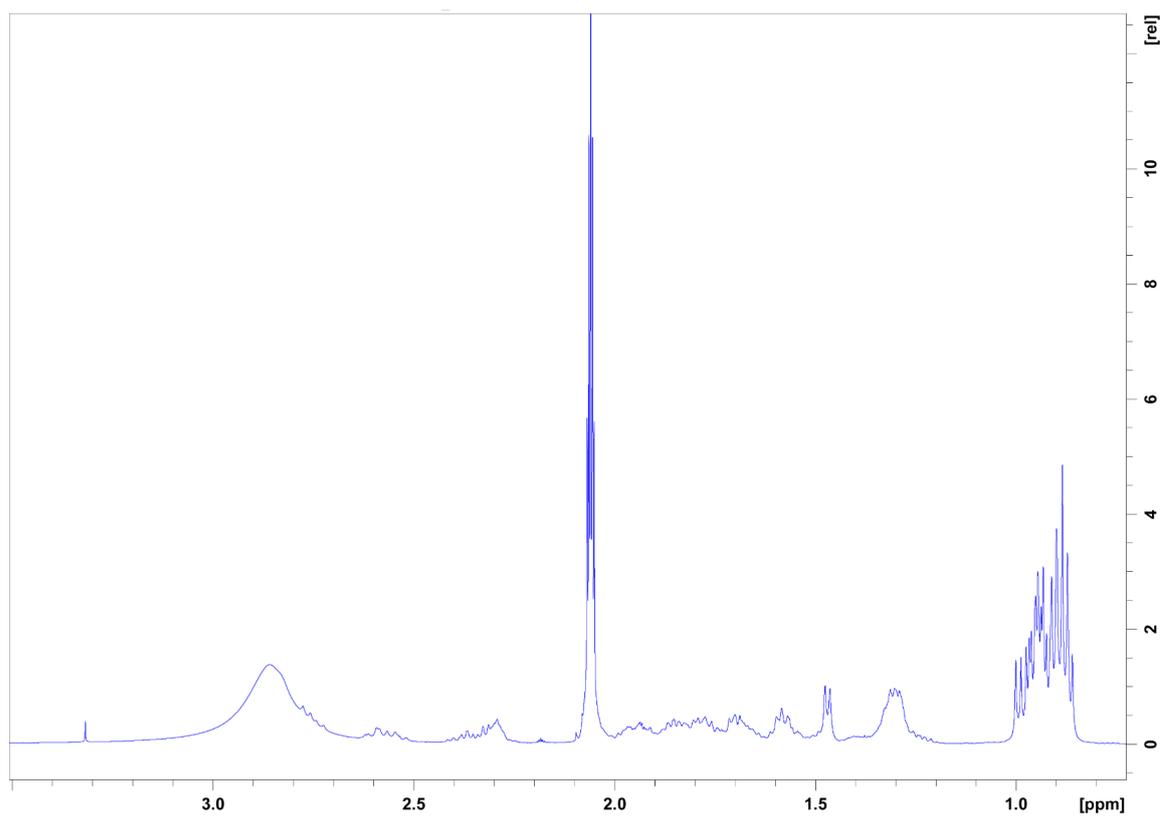


Figure S50: ^1H NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum zoom 0.8-3.5 ppm of Tensin

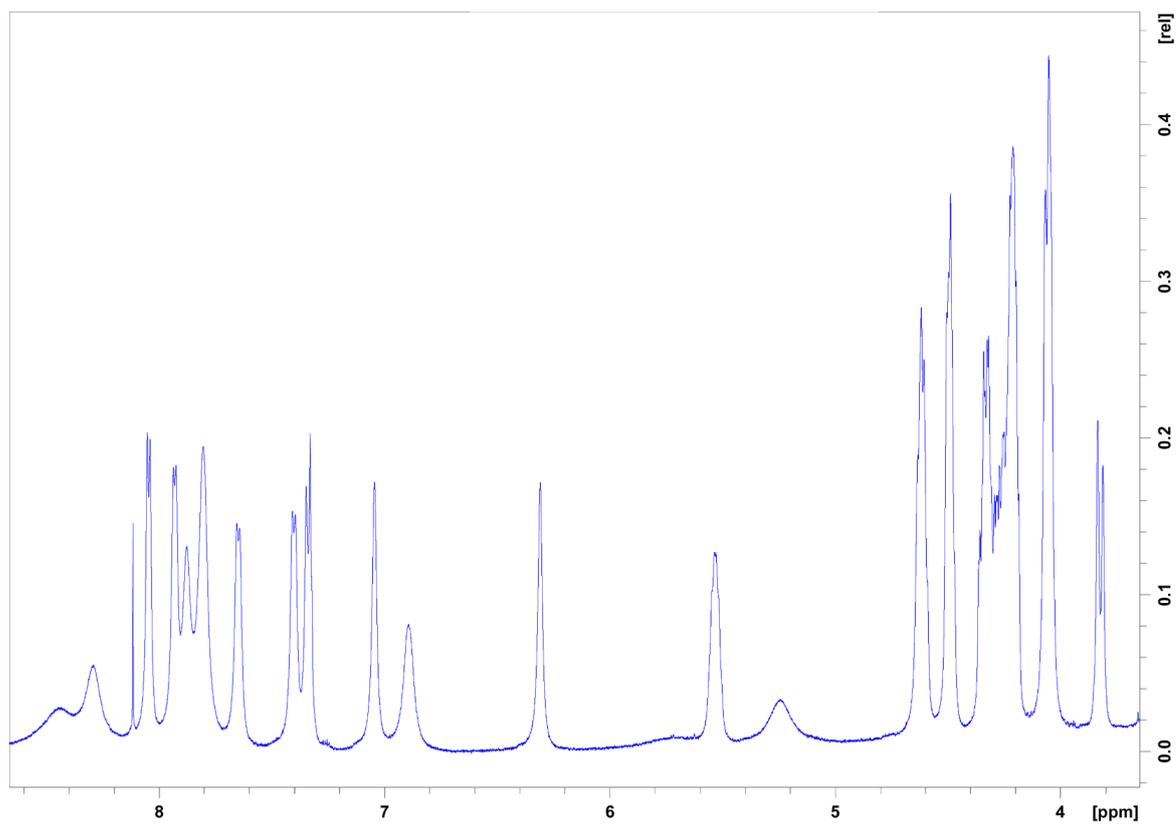


Figure S51: ^1H NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum zoom 3.5-8.5 ppm of Tensin

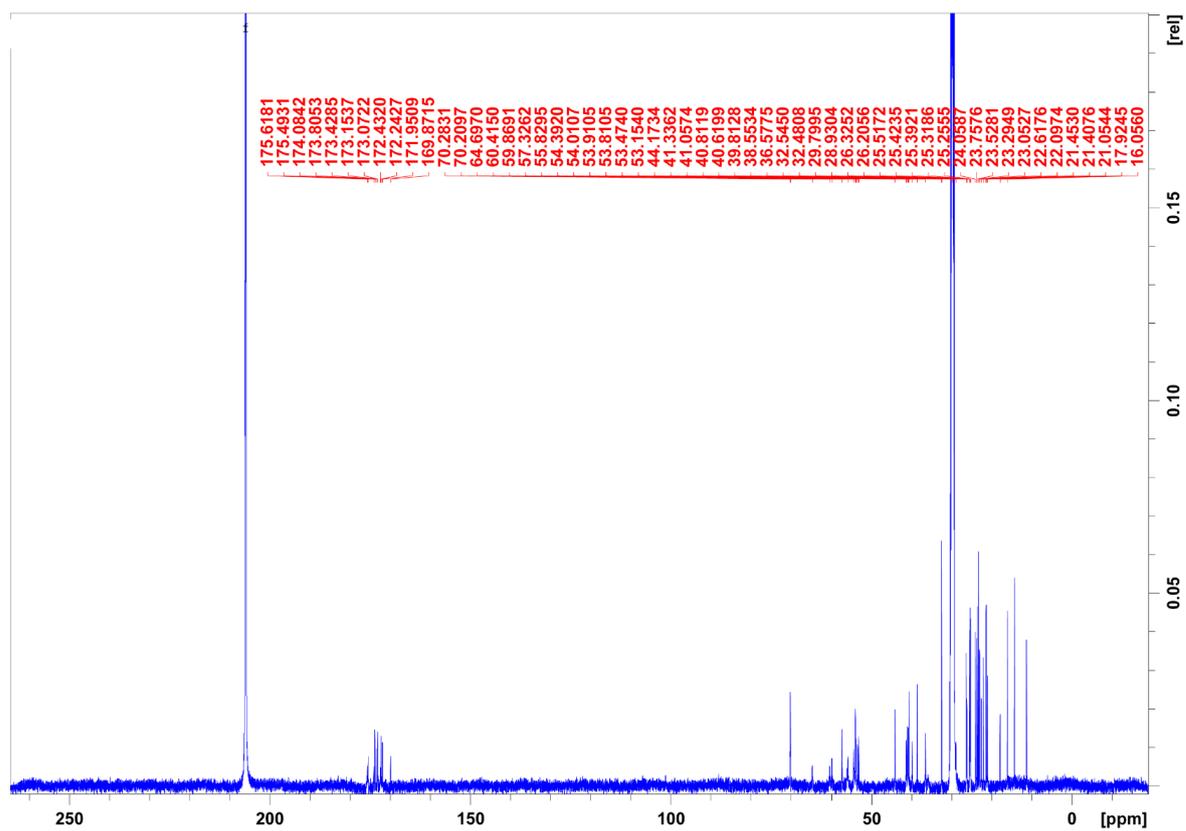


Figure S52: ^{13}C NMR (125 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum of Tensin

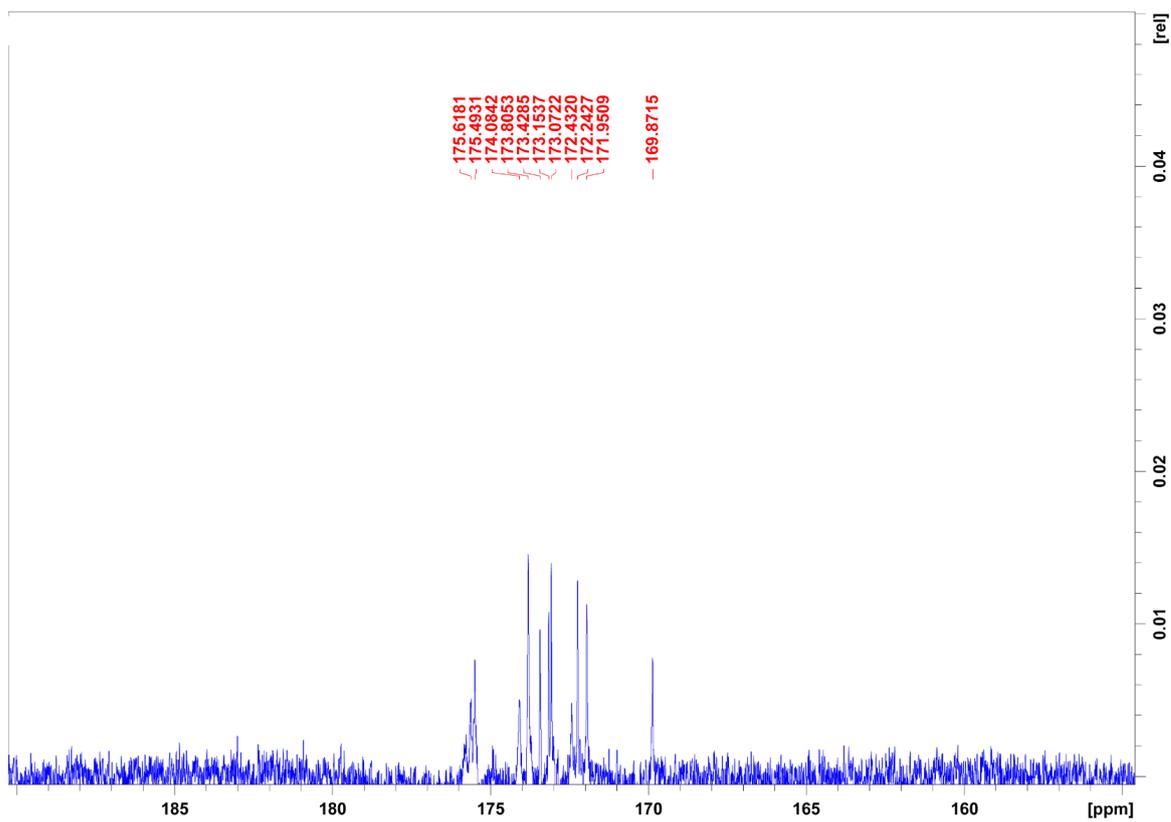


Figure S53: ^{13}C NMR (125 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum of the carbonyl region of Tensin

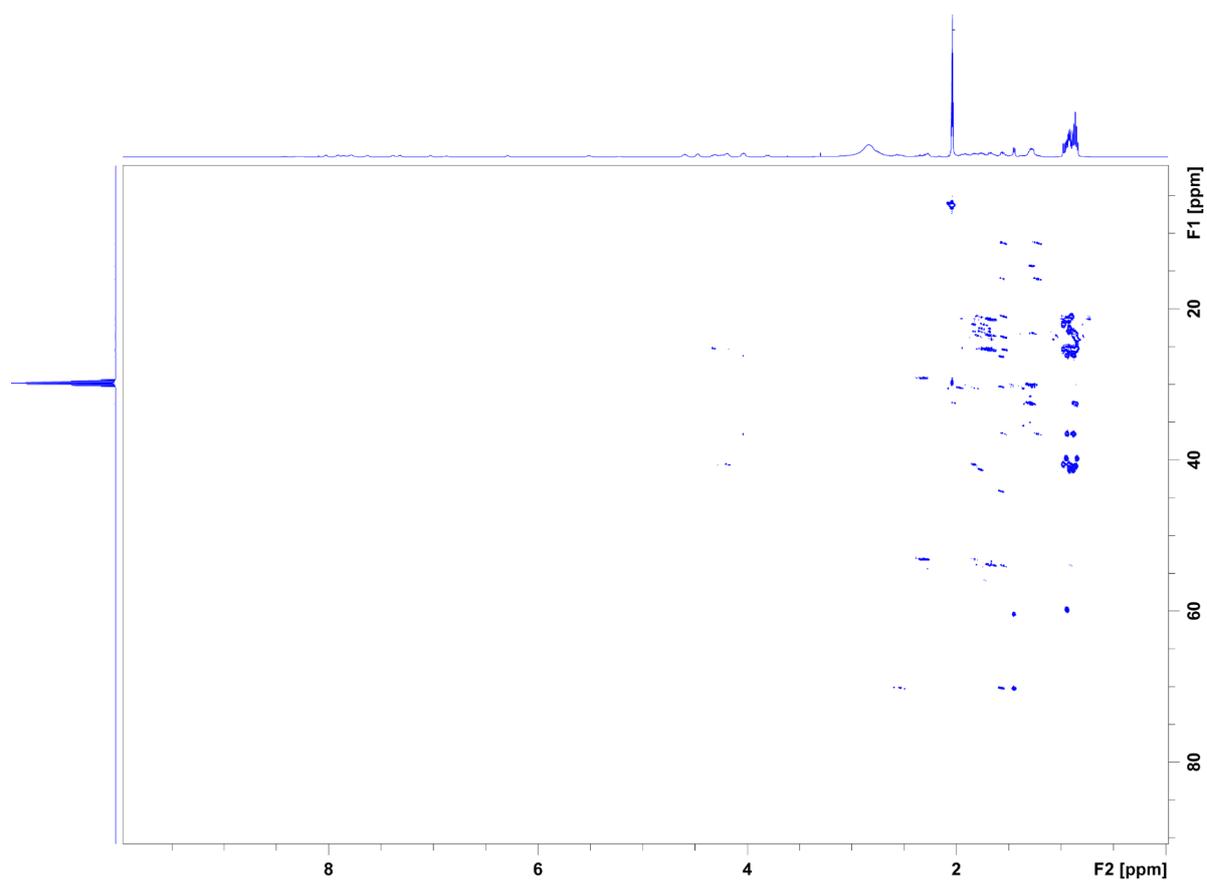


Figure S54: HMBC NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum of Tensin

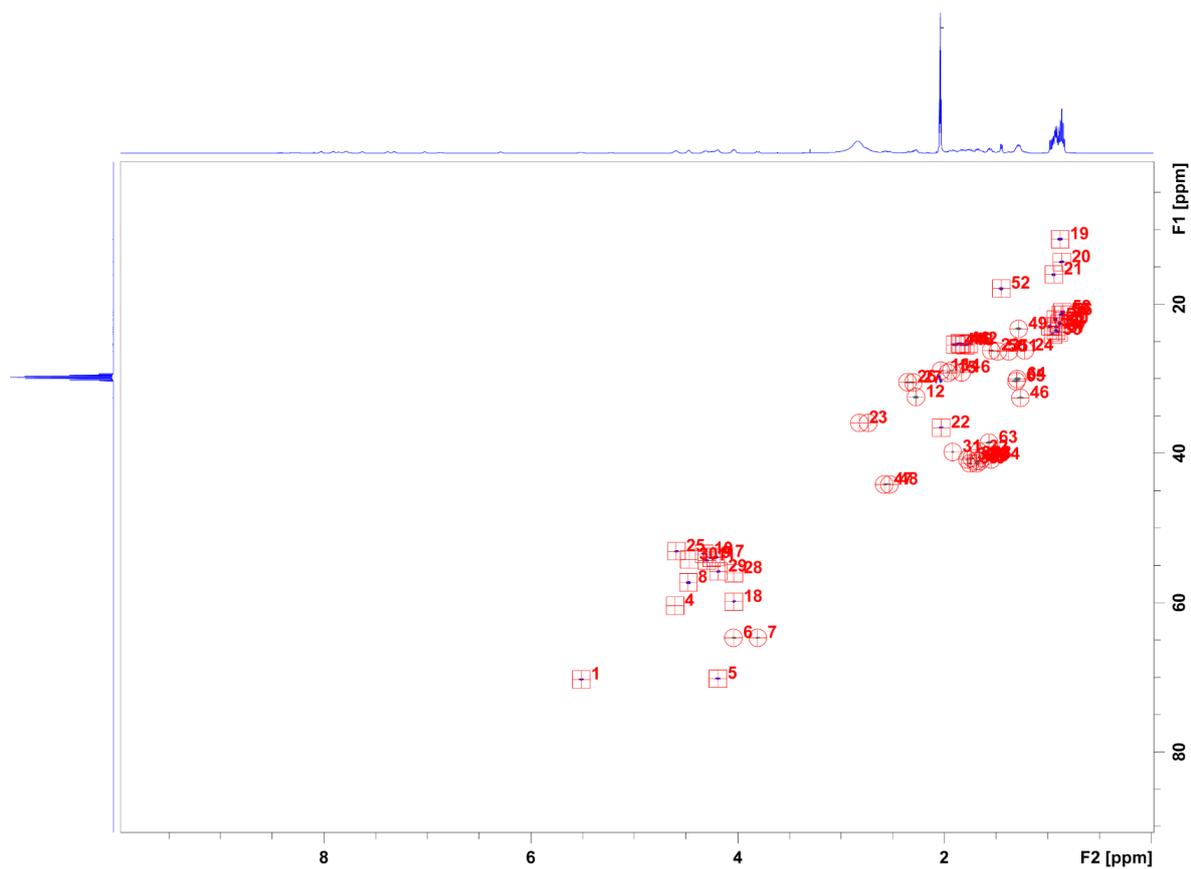


Figure S56: HSQC NMR (500 MHz, (CD₃)₂CO) spectrum of Tensin

7.11.5 Stechlisin E2

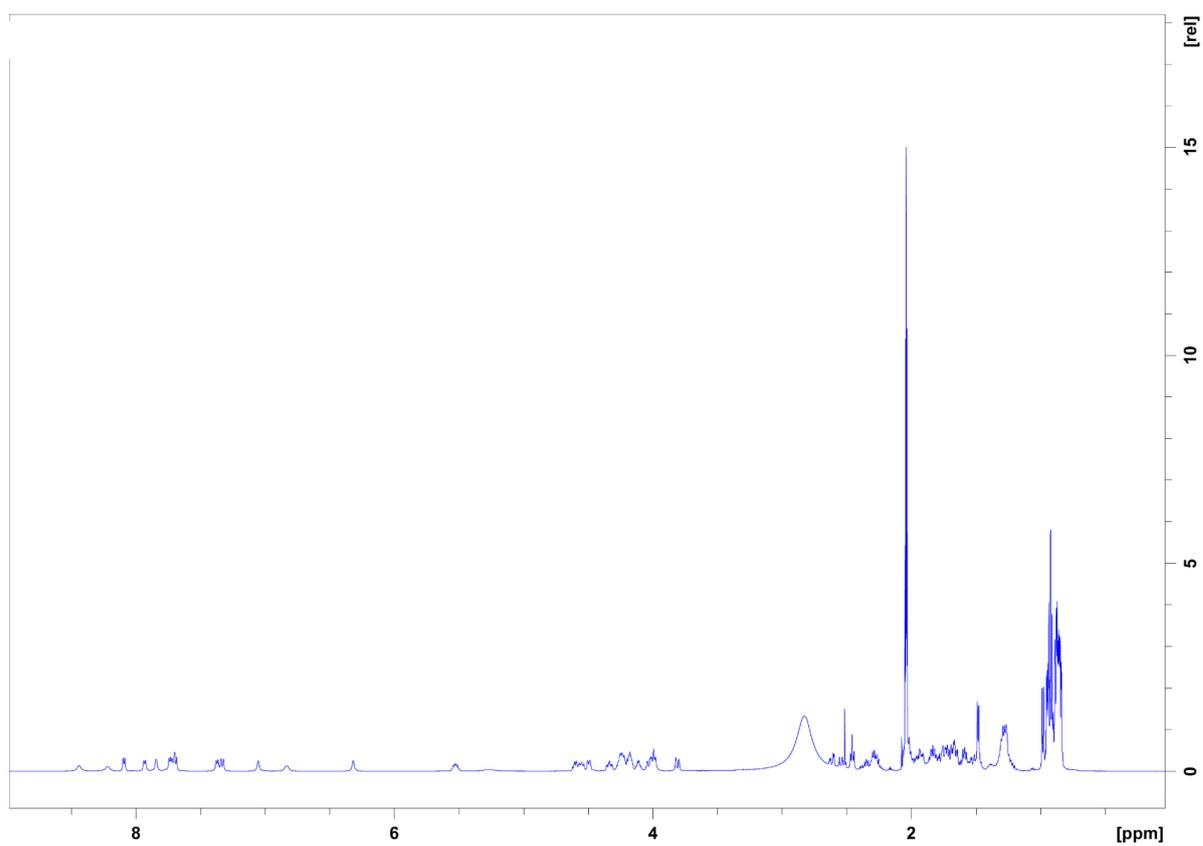


Figure S58: ¹H NMR (500 MHz, (CD₃)₂CO) spectrum of Stechlisin E2

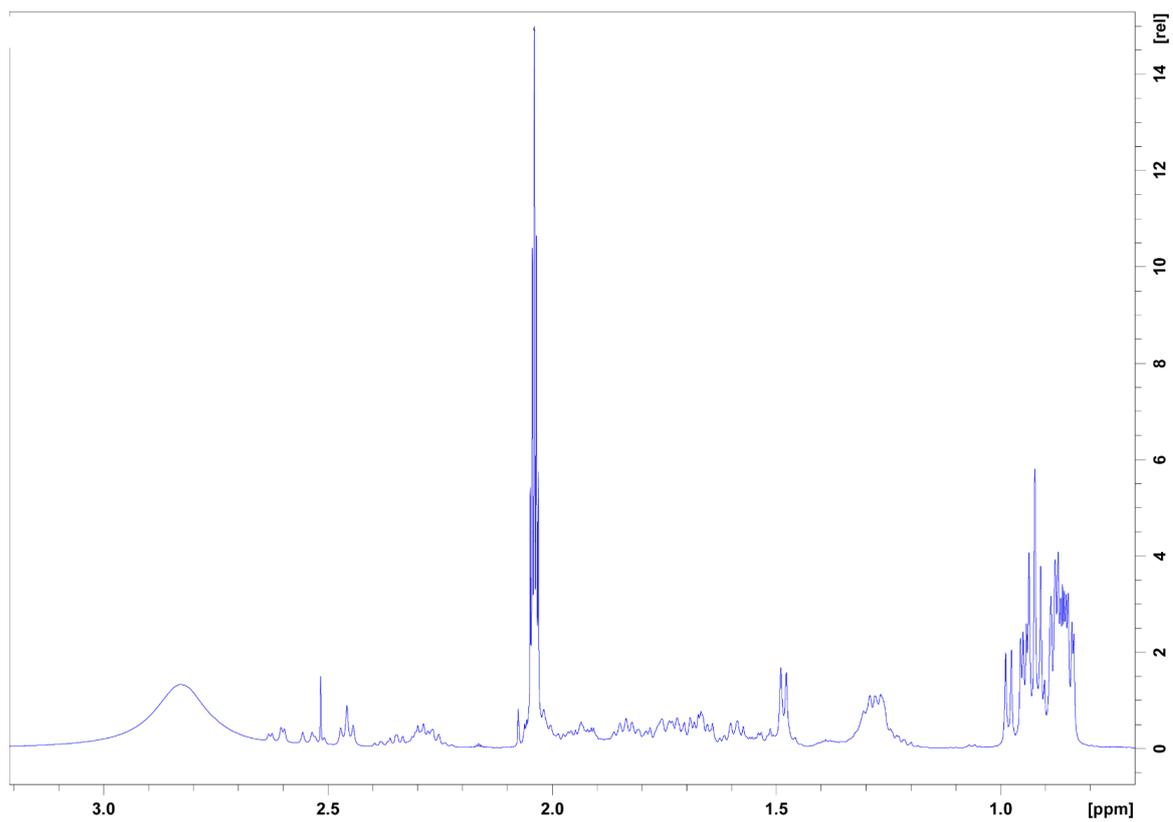


Figure S59: ^1H NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum zoom 0.8-3.5 ppm of Stechlisin E2

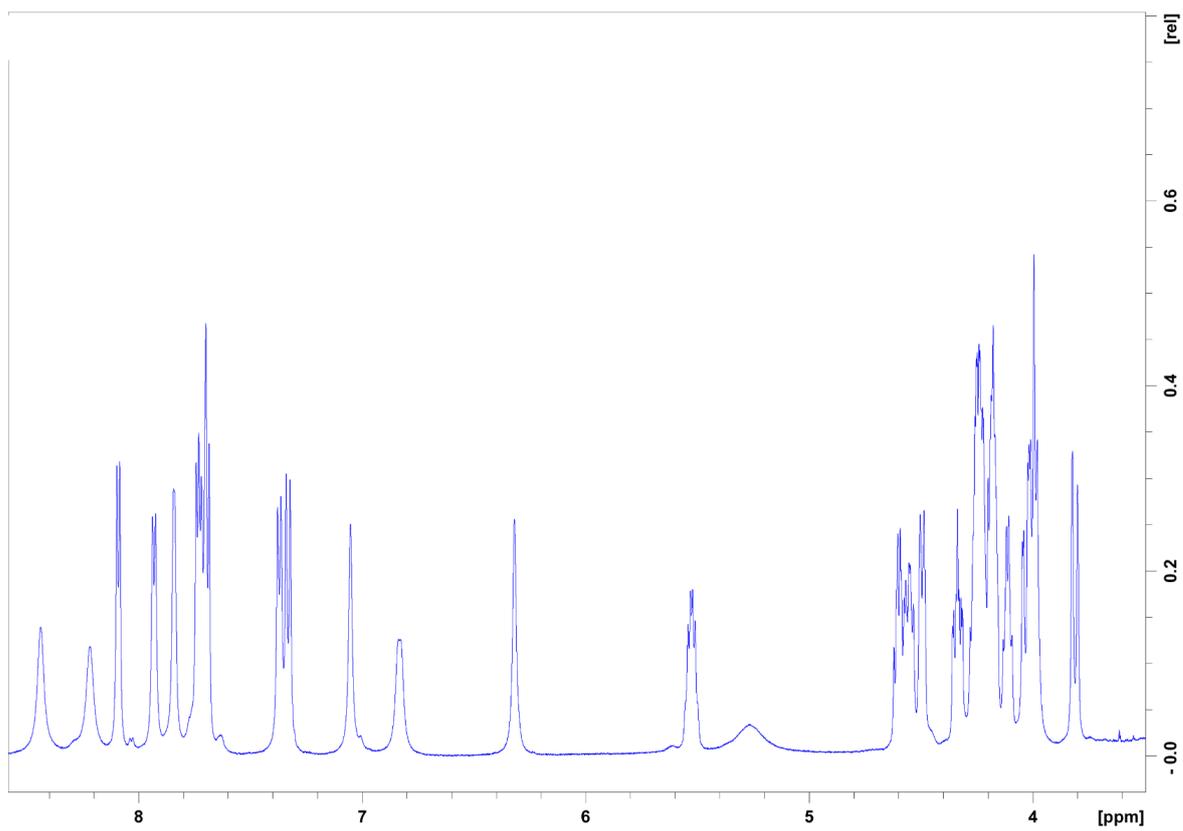


Figure S60: ^1H NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum zoom 3.5-8.5 ppm of Stechlisin E2

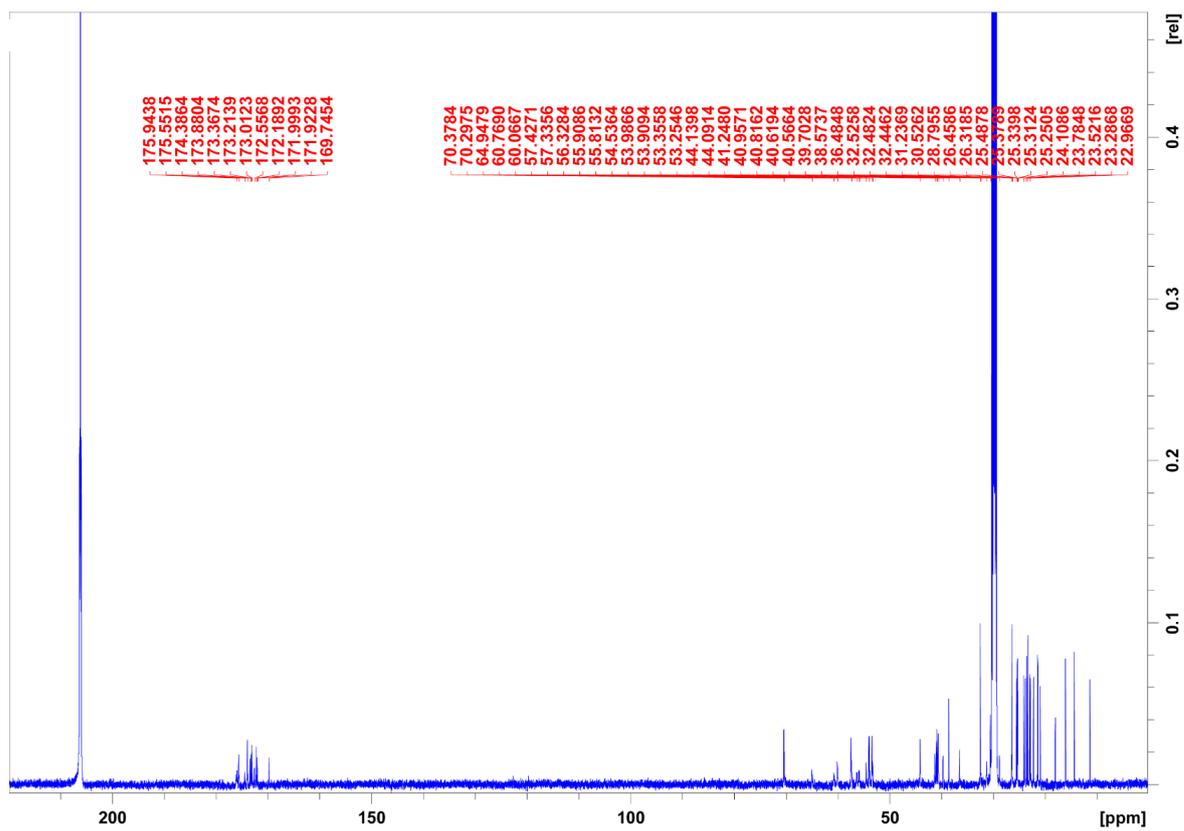


Figure S61: ¹³C NMR (125 MHz, (CD₃)₂CO) spectrum of Stechlisin E2

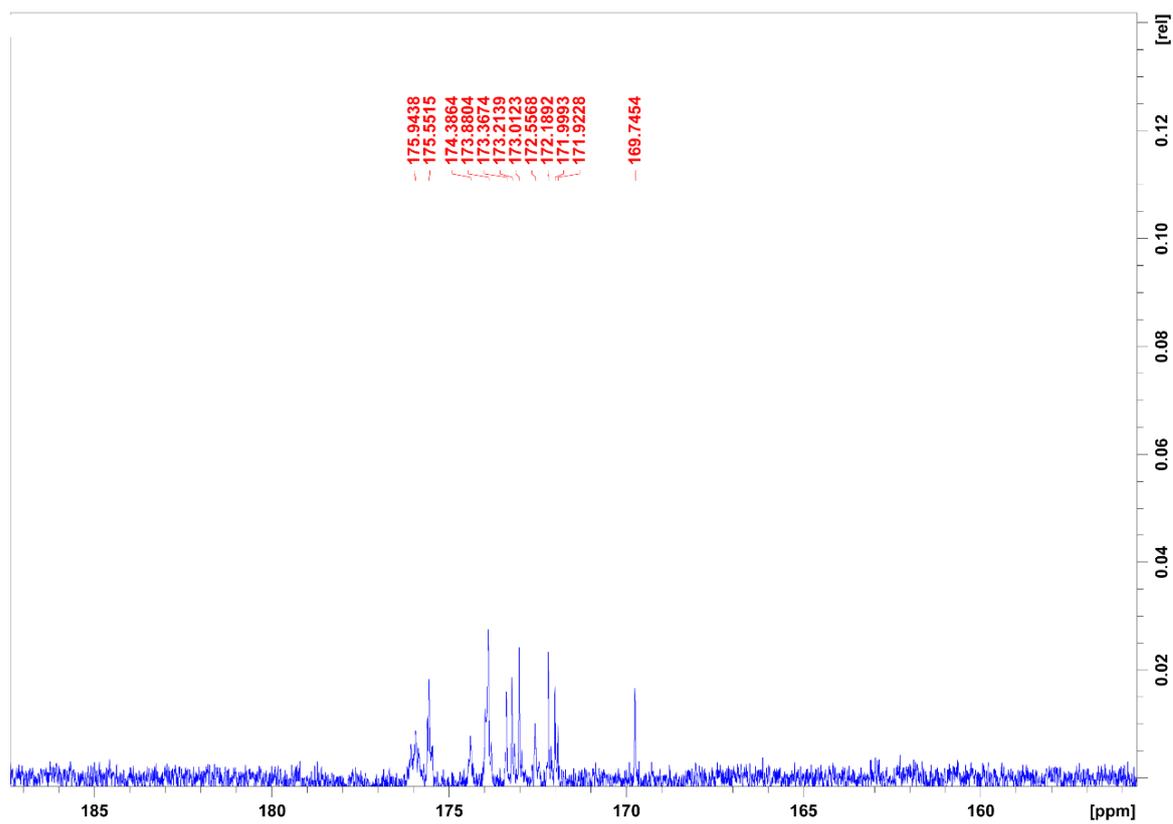


Figure S62: ¹³C NMR (125 MHz, (CD₃)₂CO) spectrum of the carbonyl region of Stechlisin E2

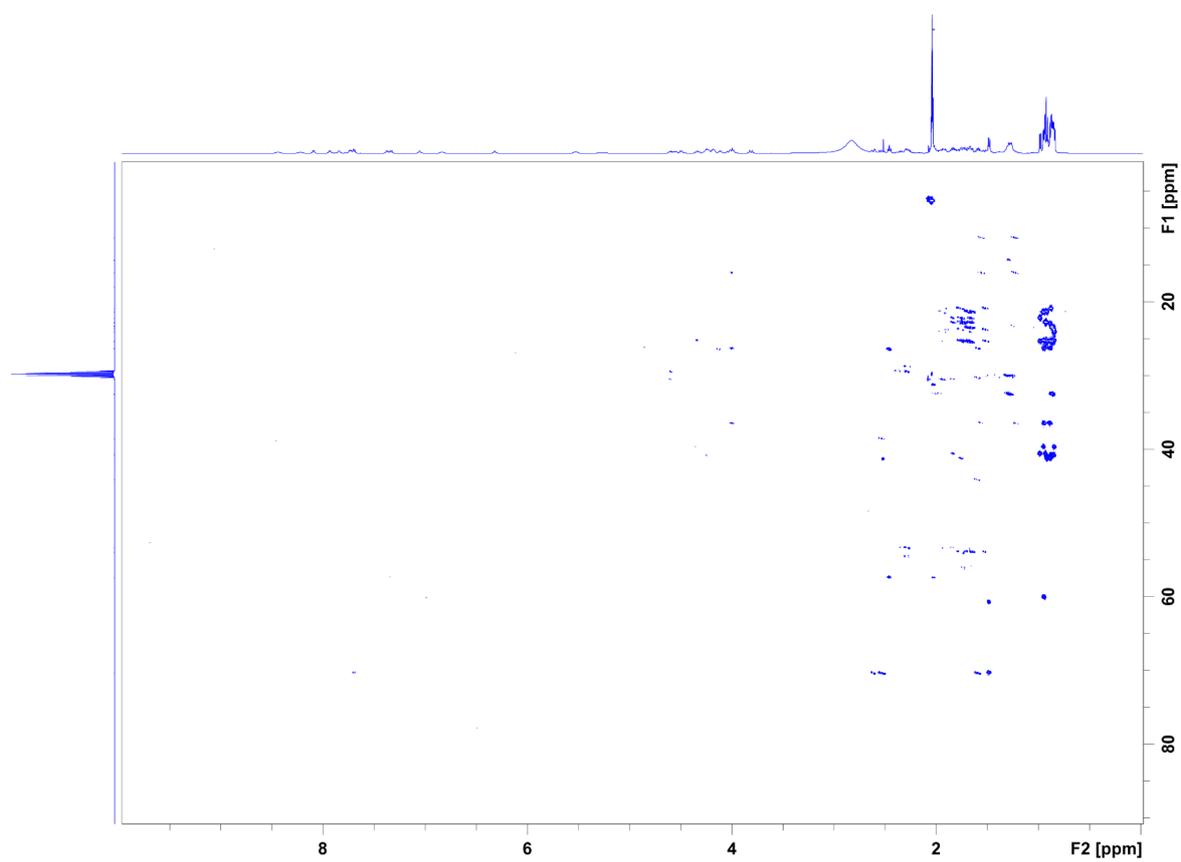


Figure S63: HMBC NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum of Stechlisin E2

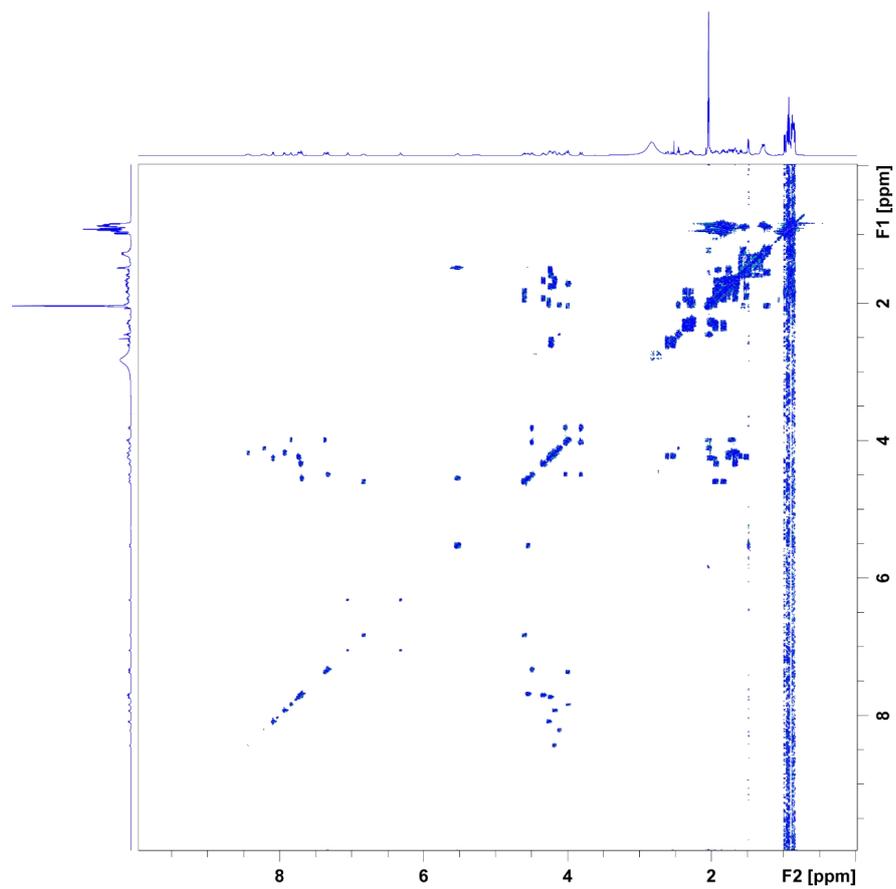


Figure S64: COSY NMR (500 MHz, (CD₃)₂CO) spectrum of Stechlisin E2

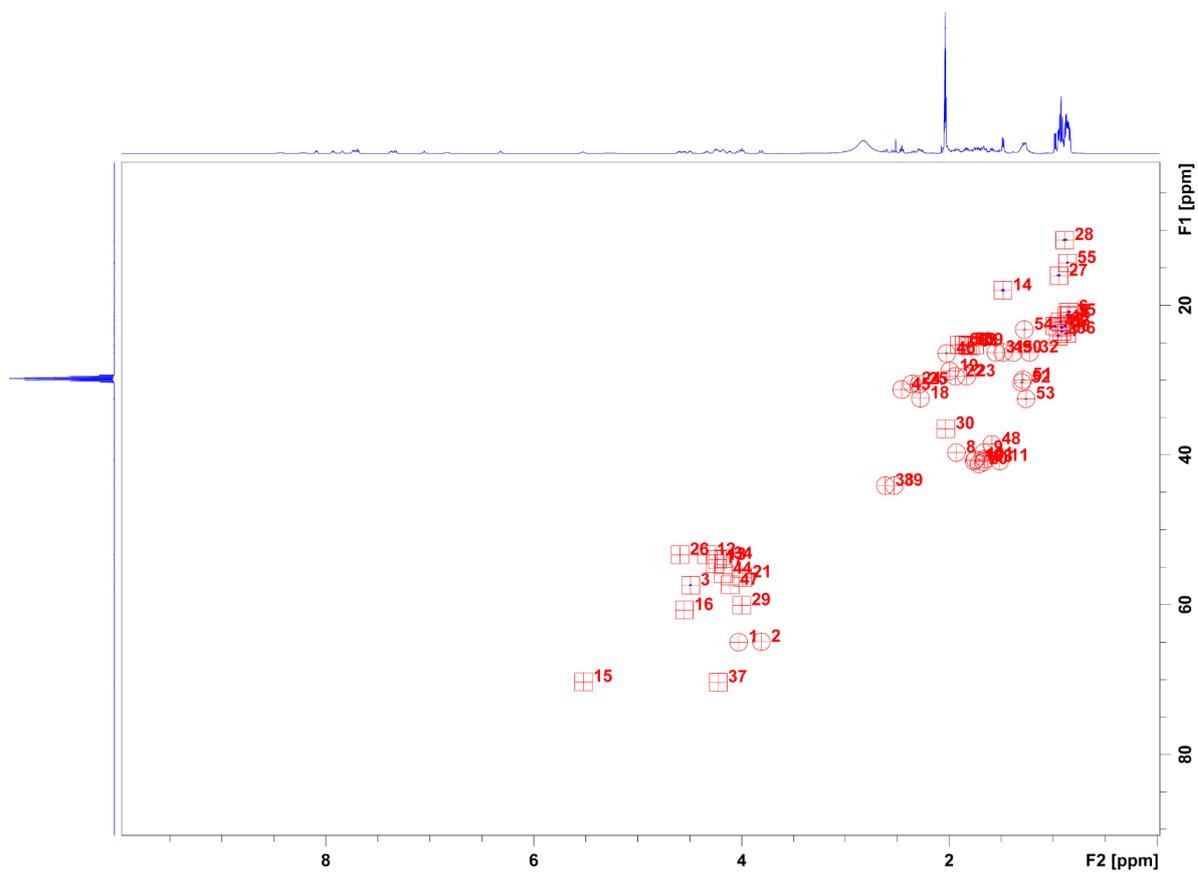


Figure S65: HSQC NMR (500 MHz, (CD₃)₂CO) spectrum of Stechlisin E2

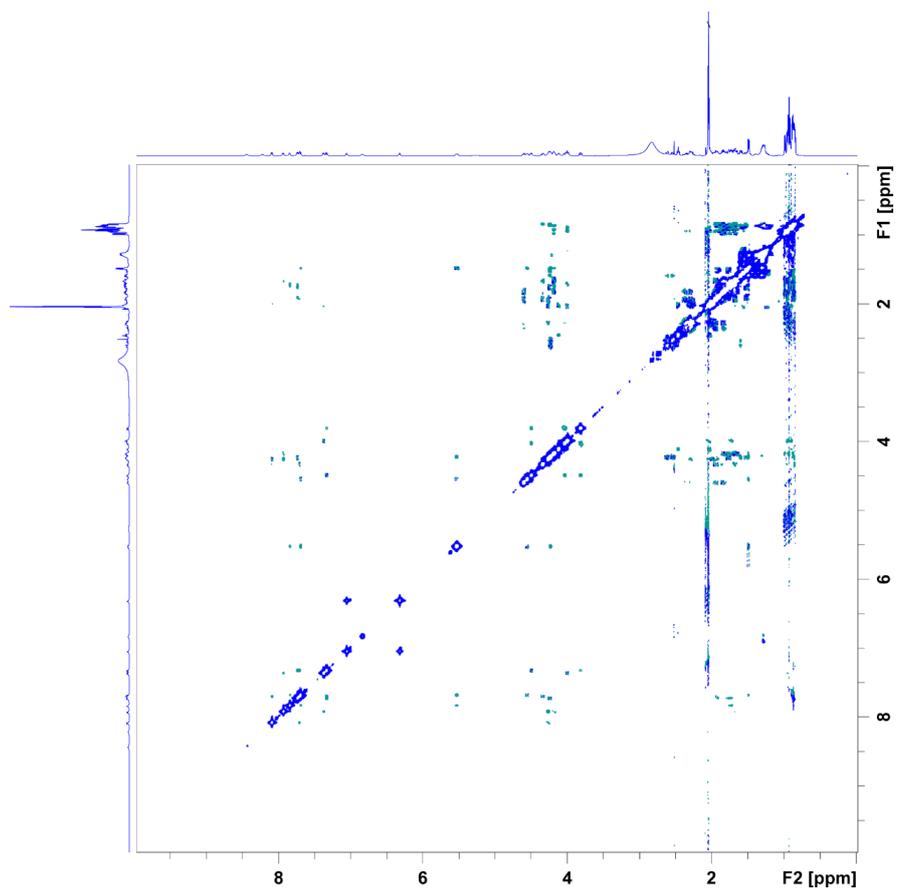


Figure S66: ROESY NMR (500 MHz, (CD₃)₂CO) spectrum of Stechlisin E2

7.11.6 Stechlisin F

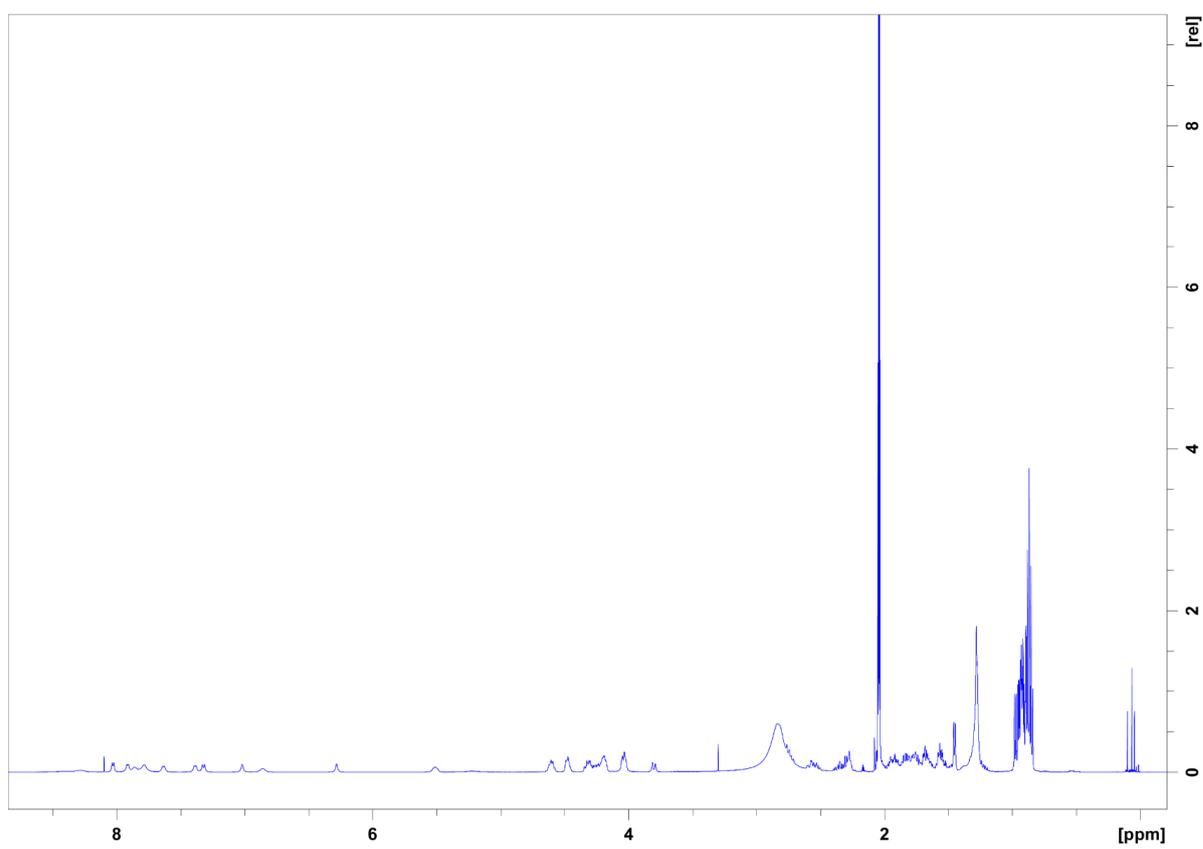


Figure S67: ¹H NMR (500 MHz, (CD₃)₂CO) spectrum of Stechlisin F

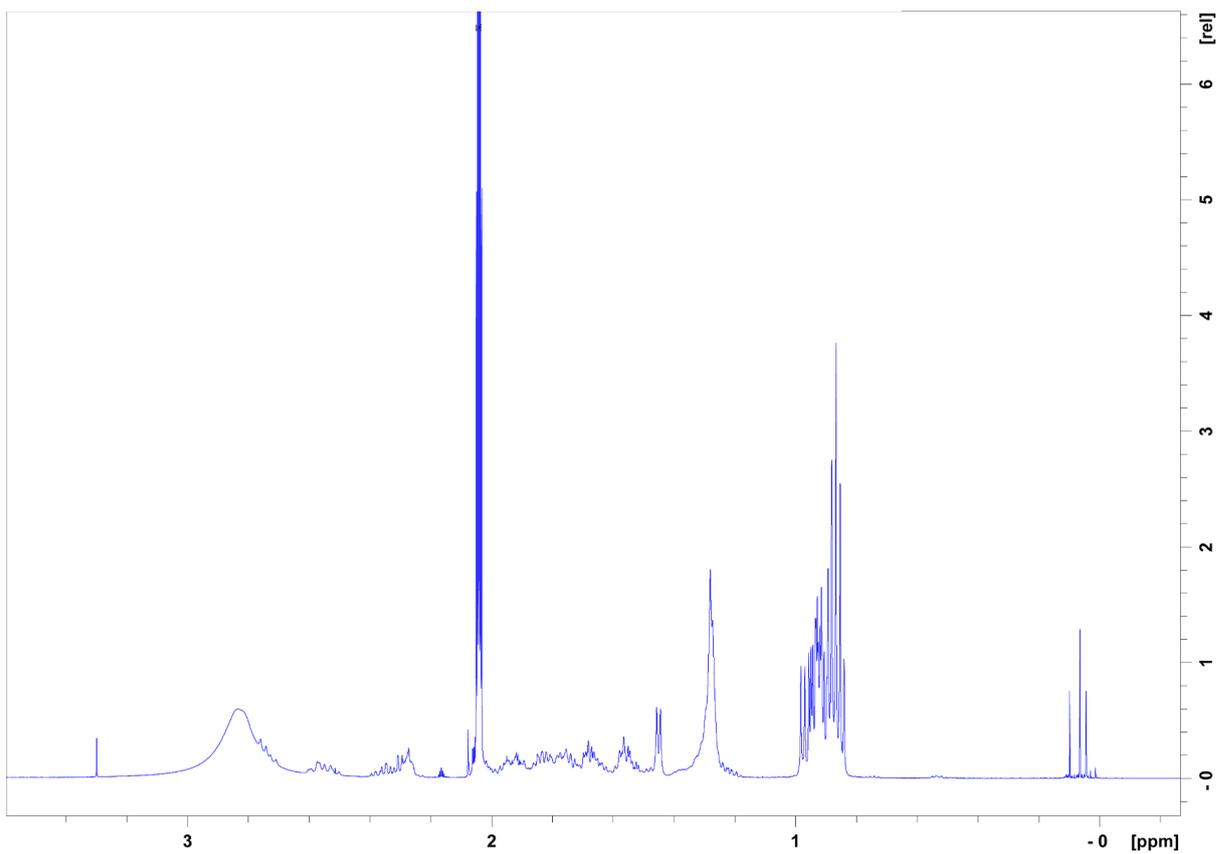


Figure S68: ^1H NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum zoom 0.8-3.5 ppm of Stechlisin F

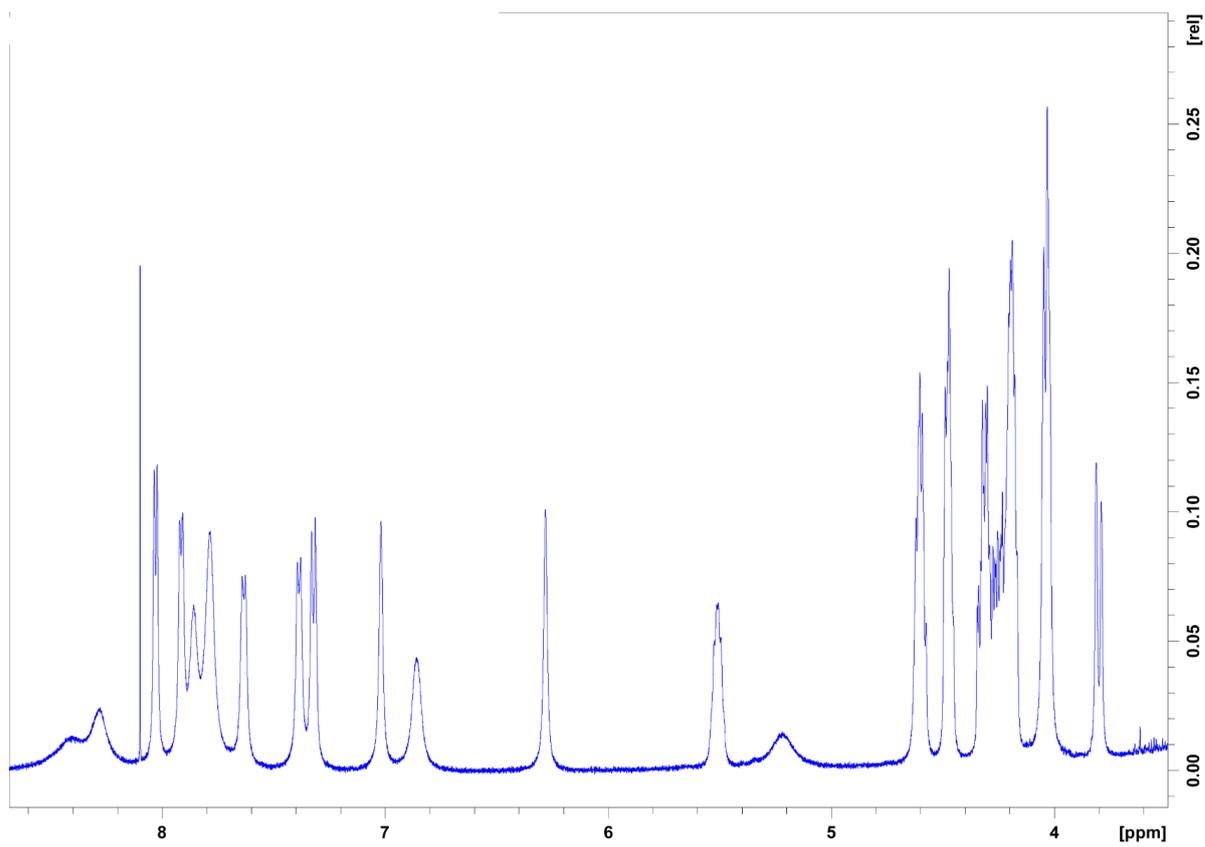


Figure S69: ^1H NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum zoom 3.5-8.5 ppm of Stechlisin F

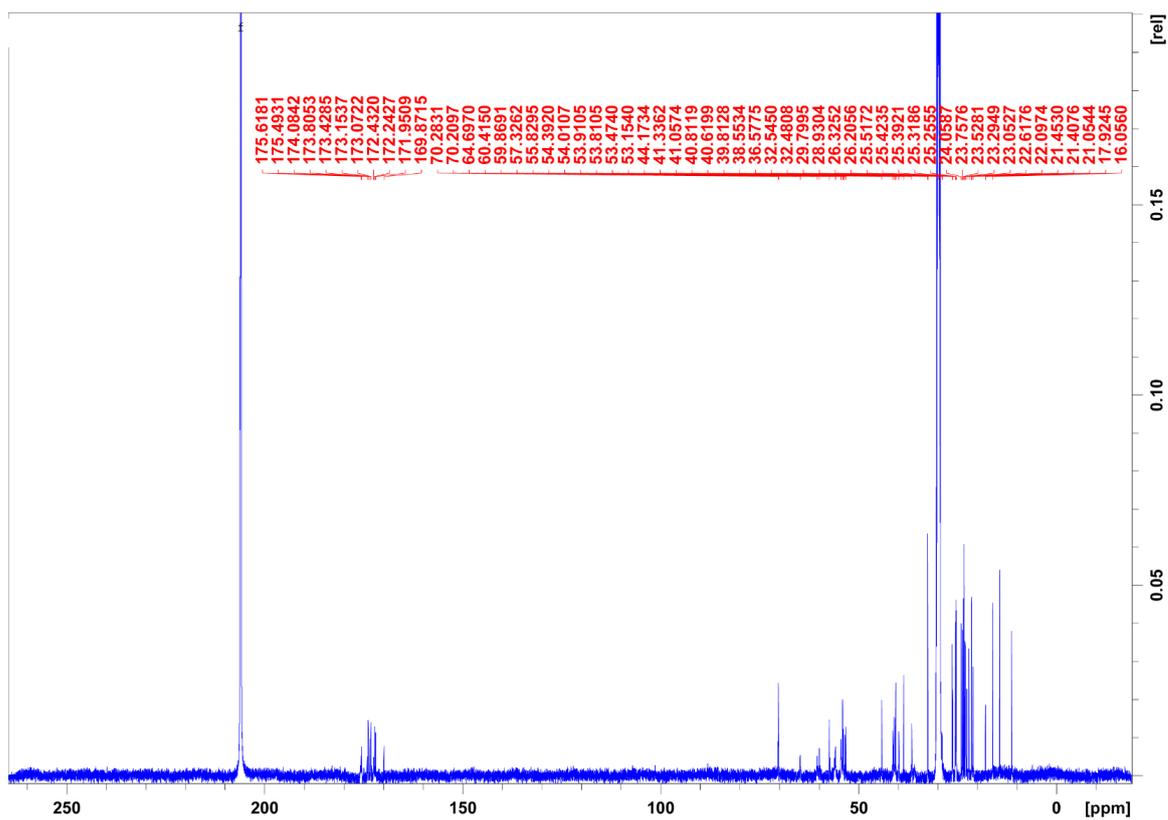


Figure S70: ¹³C NMR (125 MHz, (CD₃)₂CO) spectrum of Stechlisin F

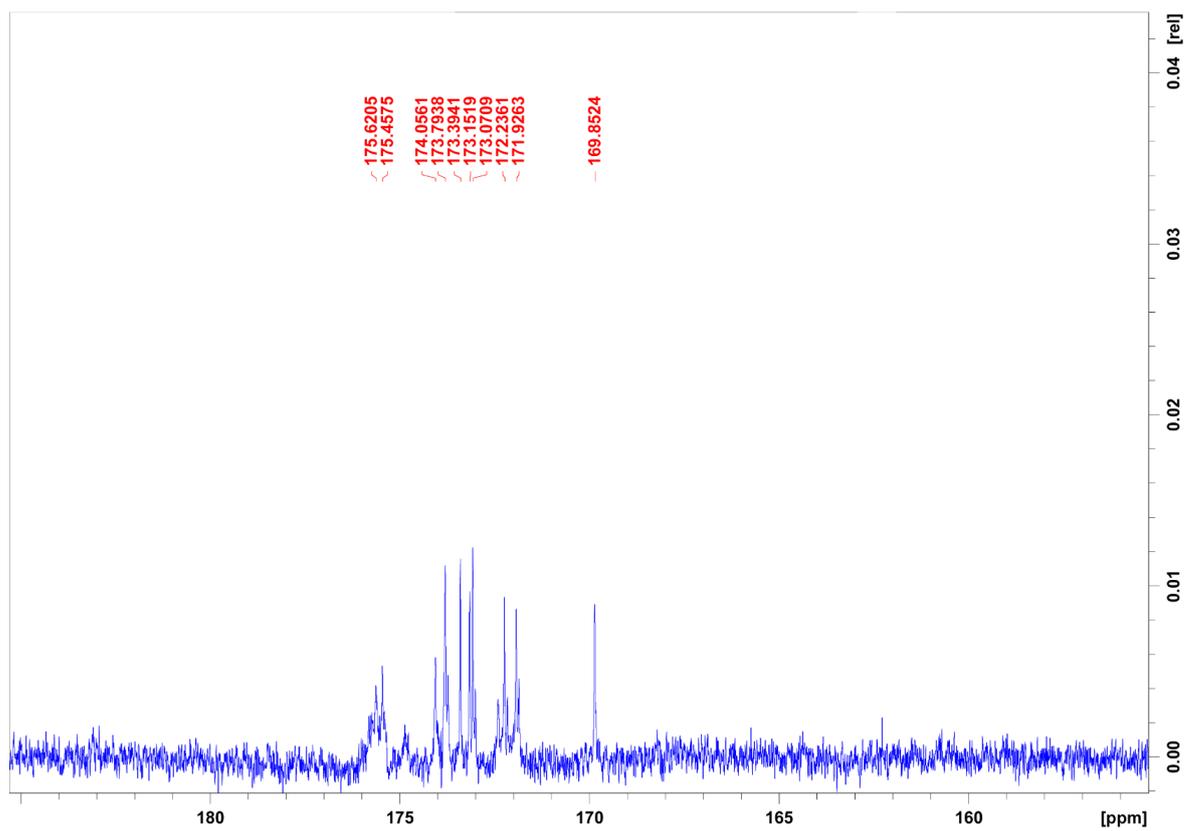


Figure S71: ¹³C NMR (300 MHz, (CD₃)₂CO) spectrum of the carbonyl region of Stechlisin F

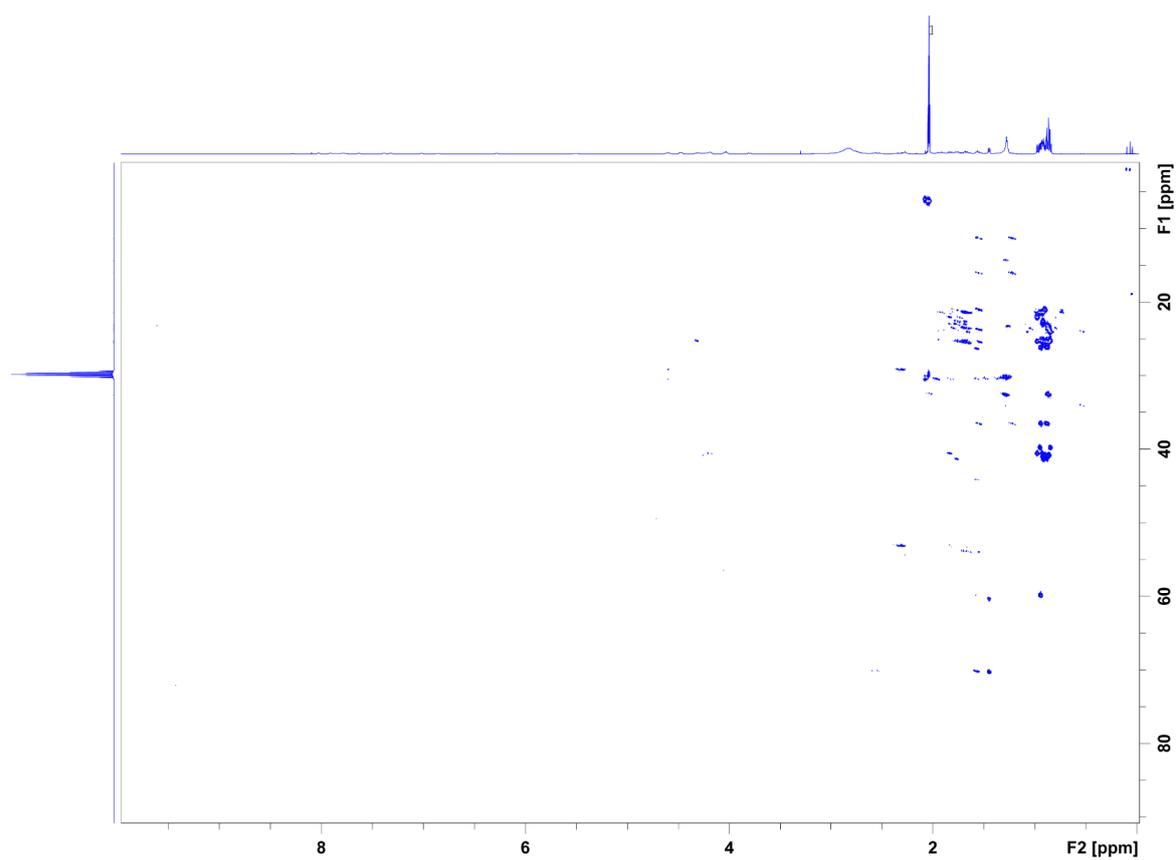


Figure S72: HMBC NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$) spectrum of Stechlisin F

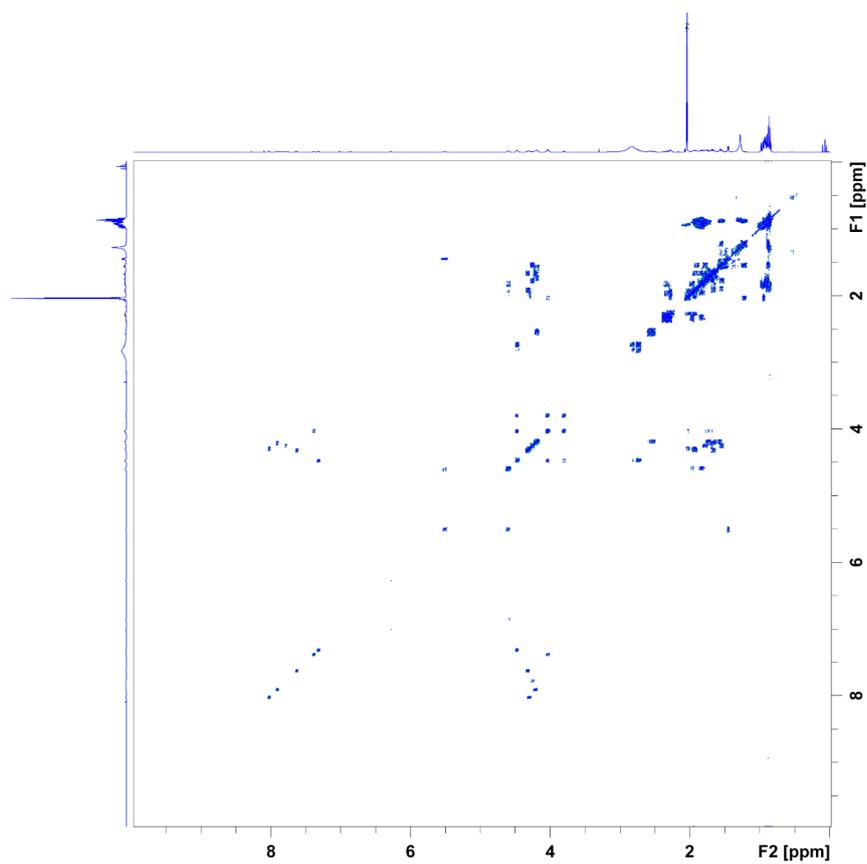


Figure S73: COSY NMR (500 MHz, (CD₃)₂CO) spectrum of Stechlisin F

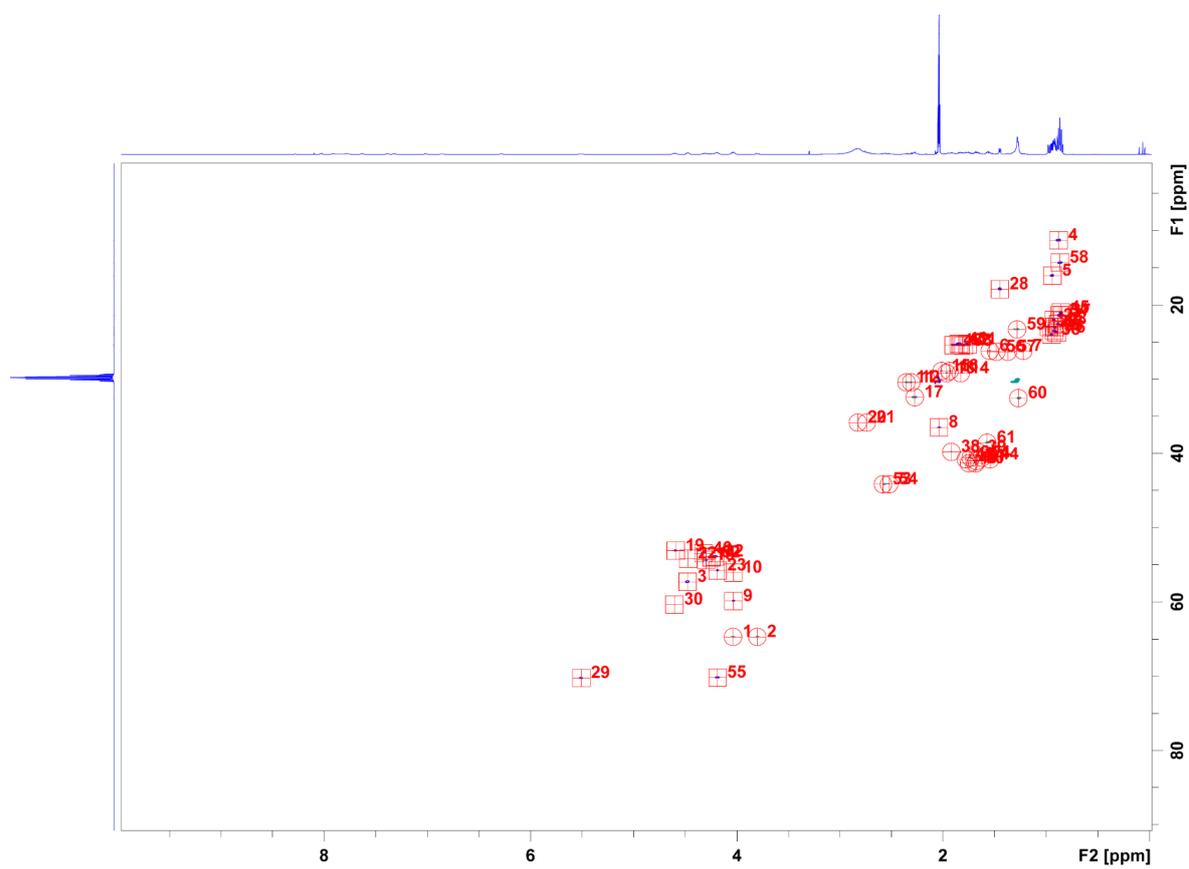


Figure S74: HSQC NMR (500 MHz, (CD₃)₂CO) spectrum of Stechlisin F

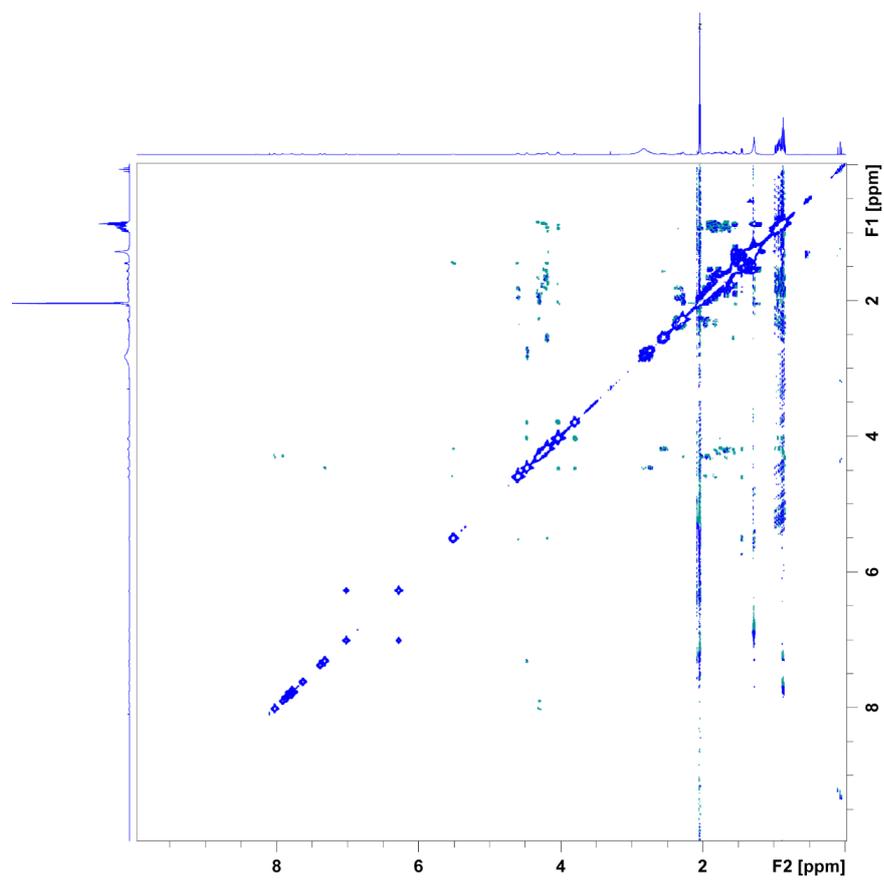


Figure S75: ROESY NMR (500 MHz, (CD₃)₂CO) spectrum of Stechlisin F

8 Project contributions

In the first section of this work (Development and Evaluation of a Metabolomics platform) Dr. Florian Zubeil and Christoph Hartwig contributed by developing and coding key programs for data handling, transformation and visualization. Florian was the driving force to purchase the Molecular networking algorithm to make it available for offline, thus industrial, application. The third part (Metabolomics-guided discovery of new cyclic lipopeptides from *Pseudomonas* sp. with anti-Gram-negative activity) was supported by NMR measurements and data interpretation of Dr. Michael Kurz, Dr. Armin Bauer and Dr. Maria A. Patras. MS/MS Analysis as well as Marfey's derivatization were strongly supported by Dr. Maria A. Patras. Genome assembly and annotation was carried out in cooperation with the bioinformatics group and in particular Dr. Frank Förster.

9 Declaration of Originality/Eigenständigkeitserklärung

Ich erkläre: Ich habe die vorliegende Dissertation selbstständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Ich stimme einer evtl. Überprüfung meiner Dissertation durch eine Antiplagiat-Software zu. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der "Satzung der Justus-Liebig-Universität zur Sicherung guter wissenschaftlicher Praxis" niedergelegt sind, eingehalten.

10 Acknowledgments

I would like to thank PD. Dr. Jens Glaeser, Prof. Dr. Peter Hammann and Prof. Dr. Andreas Vilcinskas for accepting me as a PhD student and for giving me the opportunity to be part of Natural Products group and the Fraunhofer IME-BR. I am thankful for the experiences I could obtain while conducting this project in an applied research environment. Furthermore, I would like to thank PD. Dr. Jens Glaeser for his efforts as first reviewer and his scientific input during the entire duration of the project. I would like to thank Prof. Dr. Peter Hammann for his intriguing insights in industrial antibiotic research and development as well as his commitment as second reviewer. Furthermore I would like to thank Prof. Till Schäberle and Prof. Peter Kämpfer for completing the examination board.

In general I am very thankful for the continuous support and stimulating discussions within the working group of the IME-BR. In particular, I would like to express my gratitude to Dr. Sanja Mihajlovic and Dr. Maria Patras for proof reading the manuscript. I gratefully acknowledge Christoph Hartwig for assistance in LCMS measurements and data analysis. I highly appreciate the scientific input and research enthusiasm of Dr. Luigi Toti, Dr. Armin Bauer, Dr. Marius Spohn.

Finally, I would like thank my family and friends for their mental support. Thank you all!