

**Metabolome-based Studies  
of Virulence Factors in  
*Pseudomonas aeruginosa***

Von der Naturwissenschaftlichen Fakultät der  
Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades  
Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

Tobias Depke

2021

Referent: Prof. Dr. rer. nat. Mark Brönstrup  
Korreferentin: Prof. Dr. med. Susanne Häußler  
Tag der Promotion: 08.02.2021



## Abstract

*Pseudomonas aeruginosa* is an opportunistic pathogen and an important causative agent of potentially life-threatening nosocomial infections in predisposed patients. The Gram-negative bacterium produces a large and diverse repertoire of small-molecule secondary metabolites that serve as regulators and effectors of its virulence. In this study, a range of mass spectrometry-based bacterial metabolomics approaches was used to investigate these small-molecule virulence factors and their interplay with pseudomonal metabolism as well as with phenotypic traits related to virulence. The groundwork was laid by exploring the metabolite inventory of *P. aeruginosa* and improving the coverage of its metabolome by the application of a custom software named CluMSID, that clusters analytes based on similarities of their MS<sup>2</sup> spectra. CluMSID led to the annotation of, i.a., 27 novel members of the class of alkylquinolone quorum sensing signalling molecules, which represent crucial players in the highly complex network that regulates pseudomonal virulence. The tool was developed towards a versatile and user-friendly R package hosted on Bioconductor, whose functionalities and benefits are described in detail. The new findings on the alkylquinolone chemodiversity led to further studies with a mechanistic focus that probed the substrate specificity of the enzyme complex PqsBC. It was demonstrated that PqsBC accepts different medium-chain acyl-coenzyme A substrates for the condensation with 2-aminobenzoylacetate and thereby produces alkylquinolones with various side chain lengths, whose distribution is a function of substrate specificity and substrate availability. Moreover, it was shown that PqsBC also synthesises alkylquinolones with unsaturated side chains. The focus was further broadened from metabolite and pathway-centred questions to a more global perspective on pseudomonal virulence and metabolism, which directed attention at PrmC, an enzyme with a partially unknown function indispensable for *in vivo* virulence. An untargeted metabolomics experiment yielded insights into the role of PrmC and its influence on the pseudomonal endo- and exometabolome. Finally, clinical *P. aeruginosa* strains with different virulence phenotypes were examined by untargeted metabolomics in order to disclose metabolic variation and interconnections between virulence and metabolism. The analysis resulted in the discovery of a putative virulence biomarker and enabled the construction of a random forest classification model for certain virulence phenotypes based only on metabolomics data. In summary, this study demonstrated the potential of metabolomics for the investigation of *P. aeruginosa* virulence factors and thereby contributed towards the comprehension of the complex interplay of metabolism and virulence in this important pathogen.

## Keywords

*Pseudomonas aeruginosa*; metabolomics; virulence; alkylquinolones

## Zusammenfassung

*Pseudomonas aeruginosa* ist ein wichtiger opportunistischer Erreger potenziell lebensbedrohlicher nosokomialer Infektionen bei prädisponierten Patienten. Das Gram-negative Bakterium produziert ein vielfältiges Repertoire an niedermolekularen Sekundärmetaboliten, die als Regulatoren und Effektoren seiner Virulenz dienen. In dieser Studie wurde eine Reihe von Massenspektrometrie-basierten Ansätzen der bakteriellen Metabolomik verwendet, um diese niedermolekularen Virulenzfaktoren und ihre Wechselwirkungen mit dem pseudomonalen Metabolismus sowie mit virulenzassoziierten phänotypischen Merkmalen zu untersuchen. Die Grundlage bilden die Untersuchung des Metaboliteninventars von *P. aeruginosa* und die Verbesserung der analytischen Abdeckung des Metaboloms durch die Anwendung einer selbstentwickelten Software namens CluMSID, die MS<sup>2</sup>-Spektren nach Ähnlichkeit clustert. CluMSID führte zur Annotation von u.a. 27 neuen Mitgliedern der Klasse der Alkylchinolone, die als Quorum-Sensing-Signalmoleküle entscheidende Akteure im hochkomplexen Netzwerk der Virulenzregulation darstellen. Das Tool wurde zu einem R-Paket entwickelt, das auf Bioconductor verfügbar ist und dessen Funktionalitäten und Vorteile ausführlich beschrieben werden. Die neuen Erkenntnisse über die Chemodiversität der Alkylchinolone führten zu weiteren Studien mit mechanistischem Schwerpunkt, die die Substratspezifität des Enzymkomplexes PqsBC untersuchten. Es wurde nachgewiesen, dass PqsBC verschiedene mittelkettige Acyl-Coenzym-A-Substrate für die Kondensation mit 2-Aminobenzoylacetat akzeptiert und dadurch Alkylchinolone mit verschiedenen Seitenkettenlängen produziert, deren Verteilung eine Funktion der Substratspezifität und der Substratverfügbarkeit ist. Zudem konnte gezeigt werden, dass PqsBC auch Alkylchinolone mit ungesättigten Seitenketten synthetisiert. Im Weiteren wurde der Fokus von Metaboliten- und Stoffwechselweg-zentrierten Fragen hin zu einer globaleren Perspektive der pseudomonalen Virulenz und des Metabolismus erweitert, was die Aufmerksamkeit auf PrmC lenkte, ein Enzym mit teilweise unbekannter, für die *in vivo*-Virulenz unverzichtbarer Funktion. Ein globales Metabolomik-Experiment lieferte Einblicke in die Rolle von PrmC und seinen Einfluss auf das pseudomonale Endo- und Exometabolom. Schließlich wurden klinische *P. aeruginosa*-Stämme mit unterschiedlichen Virulenzphänotypen mittels ungerichteter Metabolomik untersucht, um metabolische Variationen und Zusammenhänge zwischen Virulenz und Metabolismus aufzudecken. Die Analyse resultierte in der Entdeckung eines putativen Virulenzbiomarkers und ermöglichte die Konstruktion eines Random-Forest-Klassifikationsmodells für bestimmte Virulenzphänotypen, das nur auf Metabolomik-Daten basiert. Zusammenfassend hat diese Studie das Potenzial der Metabolomik für die Untersuchung der Virulenzfaktoren von *P. aeruginosa* aufgezeigt und damit zum Verständnis des komplexen Zusammenspiels von Metabolismus und Virulenz bei diesem wichtigen Pathogen beigetragen.

## Schlagworte zum Inhalt

*Pseudomonas aeruginosa*; Metabolomik; Virulenz; Alkylchinolone

## Preamble

This Thesis is a cumulative dissertation according to §8 (1) of the Common Regulations for the Doctorate of Natural Sciences (Dr. rer. nat.) at the Gottfried Wilhelm Leibniz Universität Hannover (PromO).

The following publications have been cumulated in this dissertation:

1. T. Depke, R. Franke, and M. Brönstrup. “Clustering of MS<sup>2</sup> spectra using unsupervised methods to aid the identification of secondary metabolites from *Pseudomonas aeruginosa*”. In: *Journal of Chromatography B* 1071 (Dec. 2017), pp. 19–28. DOI: 10.1016/j.jchromb.2017.06.002
2. T. Depke, R. Franke, and M. Brönstrup. “CluMSID: an R package for similarity-based clustering of tandem mass spectra to aid feature annotation in metabolomics”. In: *Bioinformatics* (Jan. 2019). ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btz005
3. F. Witzgall, T. Depke, M. Hoffmann, M. Empting, M. Brönstrup, R. Müller, and W. Blankenfeldt. “The Alkylquinolone Repertoire of *Pseudomonas aeruginosa* is Linked to Structural Flexibility of the FabH-like 2-Heptyl-3-hydroxy-4(1*H*)-quinolone (PQS) Biosynthesis Enzyme PqsBC”. in: *ChemBioChem* 19.14 (May 2018), pp. 1531–1544. DOI: 10.1002/cbic.201800153
4. T. Depke, S. Häussler, and M. Brönstrup. “The Peptide Chain Release Factor Methyltransferase PrmC Influences the *Pseudomonas aeruginosa* PA14 Endo- and Exometabolome”. In: *Metabolites* 10.10 (Oct. 2020), p. 417. DOI: 10.3390/metabo10100417
5. T. Depke, J. G. Thöming, A. Kordes, S. Häussler, and M. Brönstrup. “Untargeted LC-MS Metabolomics Differentiates Between Virulent and Avirulent Clinical Strains of *Pseudomonas aeruginosa*”. In: *Biomolecules* 10.7 (July 2020), p. 1041. DOI: 10.3390/biom10071041

### **Ad 1.:**

The article was published in a peer-reviewed scientific journal. The three authors contributed to the work as follows:

**Tobias Depke** Conception of the study, planning and implementation of all experiments, planning and implementation of the entire data analysis as well as the discussion, preparation of all figures and tables, writing of the article, processing of all revisions.

**Raimo Franke** Conception of the study, support in the planning and implementation of the data analysis and the discussion, revision of the article.

**Mark Brönstrup** Support in the conception of the study, support in the discussion, revision of the article.

Elsevier explicitly permits the use of articles in dissertations. From the “Permission Guidelines” (<https://www.elsevier.com/about/our-business/policies/copyright/permissions>):

“Can I include/use my article in my thesis/dissertation?”

Yes. Authors can include their articles in full or in part in a thesis or dissertation for non-commercial purposes.”

### **Ad 2.:**

The article was published in a peer-reviewed scientific journal. The three authors contributed to the work as follows::

**Tobias Depke** Conception of the project, planning and execution of all programming work and experiments, creation of figures and supplementary information, writing of the article.

**Raimo Franke** Support in the conception of the project, assistance in planning some programming work, revision of the article.

**Mark Brönstrup** Support in the conception of the project, revision of the article.

The journal specifically permits the publication of articles as part of a dissertation. According to the “Author FAQs” ([https://academic.oup.com/journals/pages/authors/authors\\_faqs/online\\_licensing#three](https://academic.oup.com/journals/pages/authors/authors_faqs/online_licensing#three)):

“What rights do I retain as an Oxford Journal author?”

[...] The right to include the article in full or in part in a thesis or dissertation, provided that this not published commercially;

For the uses specified here, please note that there is no need for you to apply for written permission from Oxford University Press in advance. Please go ahead with the use ensuring that a full acknowledgment is made to the original source of the material including the journal name, volume, issue, page numbers, year of publication, title of article and to Oxford University Press and/or the learned society.”

### **Ad 3.:**

The article was published in a peer-reviewed scientific journal. The seven authors contributed to the work as follows:

**Florian Witzgall** Conception of the structural biology part of the study, implementation and evaluation of all structural biology experiments (sections “Structure determination of PqsBC in four different crystal forms”, “Analysis of PqsBC in five different crystal forms reveals flexibility around the active site of PqsC”, “PqsC adopts open, intermediate and closed conformations”, “The acyl-binding site acts as a molecular ruler that determines the acyl-CoA specificity of PqsBC”) including the creation of the respective figures and writing of the relevant sections in the Results and Materials and Methods section, discussion of the results, writing of the article, editing of revisions to the above sections.

**Tobias Depke** Conception of the targeted metabolomics part of the study, performance and evaluation of all targeted metabolomics experiments (sections “The diversity of AQs/AQNOs produced by *P. aeruginosa* depends strictly on PqsBC”, “The AQ/AQNO spectrum of *P. aeruginosa* depends on acyl-CoA availability to PqsBC”, “PqsBC also produces mono-unsaturated AQs/AQNOs”) including the preparation of the corresponding figures and writing of the corresponding sections

in the Results and Materials and Methods part, discussion of the results, support in writing the article, processing of revisions to the above-mentioned sections.

**Michael Hoffmann** Planning and execution of the protein-analytical part of the study, execution and evaluation of all protein-analytical experiments (section “PqsBC accepts a broad spectrum of acyl-CoAs”) including preparation of the corresponding figures and writing of the corresponding sections in the Results and Materials and Methods section, revision of the article.

**Martin Empting** Planning and execution of the molecular modelling work including preparation of the corresponding figures and writing of the corresponding sections in the Materials and Methods section, support in discussing the results, revision of the article.

**Mark Brönstrup** Support in planning and discussion of the targeted metabolomics experiments, revision of the article

**Rolf Müller** Support in the conception of the study, revision of the article.

**Wulf Blankenfeldt** Conception of the study, discussion of the results, support in writing the article, revision of the article, support in processing all revisions.

Wiley expressly permits the use of articles in dissertations, as can be read in the information on licenses and copyright (<https://authorservices.wiley.com/author-resources/Journal-Authors/licensing-open-access/licensing/licensing-info-faqs.html>):

“Contributors may use the articles in teaching duties and in other works such as theses.”

#### **Ad 4.:**

The article was published in a peer-reviewed scientific journal. The three authors contributed to the work as follows:

**Tobias Depke** Conception of the study, planning and implementation of all experiments, planning and implementation of the entire data analysis and discussion, preparation of all figures and tables, writing of the article.

**Susanne Häußler** Conception of the study, provision of the *P. aeruginosa* strains, support in the discussion, revision of the article.

**Mark Brönstrup** Conception of the study, support in the discussion, revision of the article.

The article was published in open access with a Creative-Commones-License and can therefore be reused if referenced correctly (<https://www.mdpi.com/2218-1989/10/10/417>):

“This is an open access article distributed under the **Creative Commons Attribution License** which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.”

#### **Ad 5.:**

The article was published in a peer-reviewed scientific journal. The five authors contributed to the work as follows:

**Tobias Depke** Conceptual design of the study; planning, implementation and evaluation of all metabolomics experiments; preparation of illustrations and supplementary material; writing of the article.

**Janne Gesine Thöming** Selection and provision of clinical strains; planning, implementation and evaluation of transcriptomics and phenotyping experiments; support in the preparation of the corresponding figure; assistance in writing the article.

**Adrian Kordes** Planning, implementation and evaluation of transcriptomics and phenotyping experiments.

**Susanne Häußler** Support in the conception of the study; discussion of the results; revision of the article.

**Mark Brönstrup** Support in the conception of the study; discussion of the results; revision of the article.

The article was published in open access with a Creative Commons licence and the same conditions apply as for Publication 4 (<https://www.mdpi.com/2218-273X/10/7/1041>).





General Introduction .....	1	1
Aims of the thesis .....	41	2
Publication 1: Clustering of MS <sup>2</sup> Spectra to Aid Metabolite Identification .....	45	3
Publication 2: CluMSID: an R package for clustering of tandem mass spectra .....	95	4
Publication 3: PqsBC determines the AQ repertoire of <i>P. aeruginosa</i> .....	153	5
Publication 4: The influence of PrmC on metabolism and virulence.....	215	6
Publication 5: Metabolomics of virulent and avirulent clinical <i>P. aeruginosa</i> strains .	241	7
General Discussion and Outlook .....	301	8
Supplementary Tables for Publication 1 .....	321	A

# Contents

<b>Abstract</b>	<b>III</b>
<b>Zusammenfassung</b>	<b>IV</b>
<b>Preamble</b>	<b>V</b>
<b>List of Figures</b>	<b>XVII</b>
<b>List of Tables</b>	<b>XXII</b>
<b>List of Abbreviations</b>	<b>XXIV</b>
<b>1. General Introduction</b>	<b>1</b>
1.1. <i>Pseudomonas aeruginosa</i> . . . . .	1
1.1.1. Relevance as an opportunistic pathogen . . . . .	1
1.1.2. Metabolic versatility . . . . .	2
1.1.3. Quorum sensing . . . . .	3
1.1.4. Virulence . . . . .	6
1.2. Metabolomics . . . . .	9
1.2.1. Systems biology from an analytical chemistry perspective . . . . .	9
1.2.2. Targeted and untargeted metabolomics . . . . .	11
1.2.3. Metabolomics technologies . . . . .	12
1.2.4. (Pre-)Processing and data analysis . . . . .	20
1.2.5. Metabolite identification . . . . .	22
1.2.6. Microbial metabolomics . . . . .	23
1.2.7. The <i>P. aeruginosa</i> metabolome . . . . .	24
References . . . . .	25
<b>2. Aims of the Thesis</b>	<b>41</b>

<b>3. Publication 1:</b>	
<b>Clustering of MS<sup>2</sup> Spectra to Aid Metabolite Identification</b>	<b>45</b>
Abstract . . . . .	45
3.1. Introduction . . . . .	46
3.2. Materials and methods . . . . .	49
3.2.1. Chemicals and analytical standards . . . . .	49
3.2.2. Bacterial strains and growth conditions . . . . .	49
3.2.3. Metabolite extraction . . . . .	50
3.2.4. Liquid chromatography—tandem mass spectrometry . . . . .	50
3.2.5. Data processing . . . . .	51
3.3. Results and discussion . . . . .	54
3.3.1. Data pre-processing and manual assignment . . . . .	54
3.3.2. Data exploration by multidimensional scaling . . . . .	54
3.3.3. Density-based clustering of MS <sup>2</sup> spectra . . . . .	55
3.3.4. Hierarchical clustering of MS <sup>2</sup> spectra . . . . .	57
3.3.5. Hierarchical clustering of neutral loss patterns . . . . .	64
3.3.6. Semi-targeted analysis with preferred mass lists . . . . .	65
3.4. Conclusion . . . . .	65
Acknowledgements . . . . .	67
References . . . . .	67
Supplementary Material . . . . .	74
<b>4. Publication 2:</b>	
<b>CluMSID: an R package for clustering of tandem mass spectra</b>	<b>95</b>
Abstract . . . . .	95
4.1. Introduction . . . . .	96
4.2. Implementation and main functions . . . . .	96
4.2.1. Data import, pre-processing and similarity calculation . . . . .	96
4.2.2. Mining and visualisation of similarity data . . . . .	98
4.2.3. Additional functionalities . . . . .	98
4.2.4. Comparison to existing tools . . . . .	98
4.3. Conclusion . . . . .	99
Funding . . . . .	99
References . . . . .	99
Supplementary information . . . . .	100

<b>5. Publication 3:</b>	
<b>PqsBC determines the AQ repertoire of <i>P. aeruginosa</i></b>	<b>153</b>
Abstract . . . . .	153
5.1. Introduction . . . . .	154
5.2. Results . . . . .	157
5.2.1. The diversity of AQs/AQNOs produced by <i>P. aeruginosa</i> depends strictly on PqsBC . . . . .	157
5.2.2. PqsBC accepts a broad spectrum of acyl-CoAs . . . . .	158
5.2.3. Structure determination of PqsBC in four different crystal forms . . . . .	160
5.2.4. Analysis of PqsBC in five different crystal forms reveals flexibility around the active site of PqsC . . . . .	161
5.2.5. PqsC adopts open, intermediate and closed conformations . . . . .	162
5.2.6. The acyl-binding site acts as a molecular ruler that determines the acyl-CoA specificity of PqsBC . . . . .	162
5.2.7. The AQ/AQNO spectrum of <i>P. aeruginosa</i> depends on acyl-CoA availability to PqsBC . . . . .	168
5.2.8. PqsBC also produces mono-unsaturated AQs/AQNOs . . . . .	168
5.3. Discussion . . . . .	169
5.4. Experimental Section . . . . .	175
Acknowledgements . . . . .	178
Abbreviations . . . . .	179
References . . . . .	179
Supporting Information . . . . .	187
<b>6. Publication 4:</b>	
<b>The influence of PrmC on metabolism and virulence</b>	<b>215</b>
Abstract . . . . .	215
Keywords . . . . .	216
6.1. Introduction . . . . .	216
6.2. Materials and methods . . . . .	216
6.3. Results and discussion . . . . .	219
6.4. Conclusion . . . . .	224
Author contributions . . . . .	224
Funding . . . . .	224
Acknowledgements . . . . .	224

Conflicts of interest . . . . .	225
Abbreviations . . . . .	225
References . . . . .	225
Supporting Information . . . . .	227
<b>7. Publication 5:</b>	
<b>Metabolomics of virulent and avirulent clinical <i>P. aeruginosa</i> strains</b>	<b>241</b>
Abstract . . . . .	241
Keywords . . . . .	242
7.1. Introduction . . . . .	242
7.2. Materials and methods . . . . .	244
7.2.1. Bacterial strains . . . . .	244
7.2.2. Transcriptomics . . . . .	244
7.2.3. Untargeted metabolomics . . . . .	246
7.2.4. Data analysis and model building . . . . .	247
7.3. Results . . . . .	248
7.3.1. Virulent cluster A and avirulent cluster B strains have different metabolic profiles	249
7.3.2. Metabolic differences between virulent cluster A and avirulent cluster B strains manifest in differential abundance of virulence-associated secondary metabolites	254
7.3.3. An unknown metabolite is a potential biomarker for virulent phenotypes . . . . .	257
7.3.4. Virulent and avirulent strains with distinct biofilm phenotypes can be differen- tiated based on untargeted metabolomics data by machine learning . . . . .	259
7.4. Discussion . . . . .	261
7.5. Conclusions . . . . .	263
Author contributions . . . . .	264
Funding . . . . .	264
Acknowledgements . . . . .	264
Conflicts of interest . . . . .	264
Abbreviations . . . . .	265
References . . . . .	265
Supporting Information . . . . .	272
<b>8. General Discussion and Outlook</b>	<b>301</b>
8.1. Discussion . . . . .	301
8.1.1. Chemical inventorying and extended coverage of the <i>P. aeruginosa</i> metabolome	301
8.1.2. Mechanistic insights into <i>P. aeruginosa</i> virulence factor biosynthesis . . . . .	304

8.1.3. Exploration of the interplay of metabolism and virulence in a genetically modified laboratory strain . . . . .	305
8.1.4. Exploration of the interplay of metabolism and virulence in clinical <i>P. aeruginosa</i> strains . . . . .	307
8.1.5. Contribution of a feature annotation tool for the metabolomics and natural product chemistry community . . . . .	310
8.2. Conclusions . . . . .	311
8.3. Outlook . . . . .	312
References . . . . .	313
<b>A. Supplementary Tables for Publication 1</b>	<b>321</b>
<b>Acknowledgements</b>	<b>371</b>
<b>Curriculum vitae and List of Publications</b>	<b>372</b>

## List of Figures

1.1	The general principle of quorum sensing. . . . .	3
1.2	Schematic representation of the three quorum sensing systems in <i>Pseudomonas aeruginosa</i> . . . . .	4
1.3	Most important QSSMs in <i>Pseudomonas aeruginosa</i> . . . . .	5
1.4	Alkylquinolone bioynthesis in <i>Pseudomonas aeruginosa</i> . . . . .	7
1.5	Interrelation of omics experiments and systems biology. . . . .	10
1.6	The Omics Cascade. . . . .	11
1.7	A typical metabolomics workflow. . . . .	12
1.8	Schematic of the general setup of a mass spectrometer. . . . .	16
1.9	Schematic of the electrospray ionisation process. . . . .	17
1.10	Design and function of a quadrupole time-of-flight mass analyser. . . . .	19
3.1	Chemical structures of selected secondary metabolites produced by <i>Pseudomonas aeruginosa</i> . . . . .	47
3.2	Reachability plot of OPTICS-clustered MS <sup>2</sup> spectra. . . . .	56
3.3	Results of the hierarchical clustering with average linkage as agglomeration method of 518 consensus MS <sup>2</sup> spectra visualised as circular dendrogram. . . . .	58
3.4	Most important alkyl quinolone fragmentation reactions according to Lepine <i>et al.</i> 2004. . . . .	59
3.5	Possible structures of an alkyl quinolone species with the sum formula C <sub>20</sub> H <sub>27</sub> NO <sub>3</sub> . . . . .	62
3.S1	MDS plot displaying cosine differences of 518 MS <sup>2</sup> spectra. . . . .	74
3.S2	Heatmap displaying cosine similarity between 518 consensus spectra and dendrogram as result of hierarchical clustering with average linkage. . . . .	75
3.S3	MS <sup>2</sup> spectrum of M503.33T797.17, a putative proton-bound mixed dimer of HHQ and HQNO. . . . .	76
3.S4	MS <sup>2</sup> spectra of M202.12T578.44 (A) and M200.11T578.31 (B) that have been putatively annotated as C4-HQ and C4:1-HQ. . . . .	77
3.S5	MS <sup>2</sup> spectrum of M312.2T794.57, putative C11:2-QNO. . . . .	78

3.S6	MS <sup>2</sup> spectrum of M330.21T844.12, a putative side chain-oxidation product of C11:1-QNO. . . . .	78
3.S7	MS <sup>2</sup> spectrum of M304.19T717.91, a putative side chain-oxidation product of C9-QNO. . . . .	79
3.S8	MS <sup>2</sup> spectrum of M304.19T786.75, another putative side chain-oxidation product of C9-QNO. . . . .	79
3.S9	MS <sup>2</sup> spectrum of M358.24T929.17, a putative side chain-oxidation product of C13:1-QNO. . . . .	80
3.S10	Extracted ion chromatogram of <i>m/z</i> 260.164, the exact mass of the [M+H] <sup>+</sup> of both HQNO and PQS. . . . .	80
3.S11	Cluster 43 containing various alkyl quinolone species. . . . .	81
3.S12	Cluster 48 from the dendrogram obtained by hierarchical clustering of MS <sup>2</sup> spectra similarities. . . . .	81
3.S13	Cluster 35 from the dendrogram obtained by hierarchical clustering of MS <sup>2</sup> spectra similarities. . . . .	81
3.S14	MS <sup>2</sup> spectra of pyocyanin (upper pane) and M215.12T626.24, putatively annotated as tetrahydropyocyanin (lower pane). . . . .	82
3.S15	MS <sup>2</sup> spectrum of M254.09T400.89, annotated as aeruginosin A. . . . .	82
3.S16	MS <sup>2</sup> spectra of pyocyanin (upper pane) and M255.08T482.73, putatively annotated as pyocyanin carboxylic acid (lower pane). . . . .	83
3.S17	Dendrogram obtained by hierarchical clustering of neutral loss pattern similarities. . . . .	84
3.S18	Dendrogram obtained by hierarchical clustering of product ion spectra similarities of the spectra acquired in the alkyl quinolone-biased semi-targeted analysis. . . . .	85
4.1	Schematic of a CluMSID workflow. . . . .	97
5.1	Scheme: Current understanding of AQ and AQNO biosynthesis in <i>Pseudomonas aeruginosa</i> , highlighting the role of PqsBC. . . . .	156
5.1	AQ/AQNO profile of <i>Pseudomonas aeruginosa</i> PA14. . . . .	158
5.2	Acylation of PqsBC. . . . .	159
5.3	Structural comparison of 22 individual PqsBC heterodimers in crystal forms 1 to 5 with the Protein Structural Statistics Web Server (PSSweb). . . . .	163
5.4	Conformational changes in PqsBC. . . . .	164
5.5	Modeling of PqsBC-acyl-CoA complexes. . . . .	166



5.6	Transition from the open <b>A</b> ) to the closed state <b>B</b> ) reduces the size of the acyl-binding pocket from 708 to 256 Å <sup>3</sup> as a consequence of changes in the indicated residues. . . . .	167
5.7	The AQ/AQNO spectrum of <i>P. aeruginosa</i> depends on the availability of fatty acids and 2-octenoic acid is the precursor of C7:1-HQ and C7:1-QNO. . . . .	170
5.8	A structural, hypothetical model of the catalytic cycle of PqsBC. . . . .	174
5.S1	Deconvoluted mass spectra of PqsBC competitive acyl-CoA loading experiments. . . . .	200
5.S2	Crystal morphologies of different PqsBC variants. . . . .	201
5.S3	Crystal packing of different PqsBC variants. . . . .	201
5.S4	Crystal contacts of the TEV-protease cleavable His <sub>6</sub> -tag at the <i>N</i> -terminus of PqsC with symmetry related molecules in crystal form 1. . . . .	202
5.S5	Ligands bound to the active site of PqsC. . . . .	203
5.S6	Multiple sequence alignment of PqsC (Uniprot ID: Q9I4X1), PqsB (Uniprot ID: Q9I4X2) and FabH enzymes from <i>Mycobacterium tuberculosis</i> ( <i>mtFabH</i> ; Uniprot ID: P9WNG3), <i>Escherichia coli</i> ( <i>ecFabH</i> ; Uniprot ID: P0A6R0) and from <i>Staphylococcus aureus</i> ( <i>saFabH</i> ; Uniprot ID: Q8NXE2). . . . .	204
5.S7	Conformational states of PqsBC. . . . .	205
5.S8	Stereo plots of the V241-P242 peptide bond in <i>cis</i> (A, open PqsBC) and <i>trans</i> configuration (B, closed PqsBC). . . . .	206
5.S9	Effect of exogenous fatty acid addition on the AQ profile of <i>P. aeruginosa</i> . . . . .	207
5.S10	Mono-unsaturated AQs/AQNOs in combined cell and supernatant extracts of exponentially growing wildtype <i>P. aeruginosa</i> PA14. . . . .	208
5.S11	Addition of octanoic acid increased the levels of C7:1-HQ in <i>P. aeruginosa</i> PA14. . . . .	209
5.S12	Addition of fully deuterated octanoic-d15 acid led to deuterated C7-HQ/C7-QNO and C7:1-HQ/C7:1-QNO species. . . . .	210
6.1	Volcano plots of GC-MS metabolomics data, LC-MS endometabolomics data and LC-MS exometabolomics data. . . . .	220
6.2	Directional fold changes of phenazines and AQs in the LC-MS endo- and exometabolomics data. . . . .	222
7.1	Schematic of the experimental and data analysis workflow of the metabolomics part of this study. . . . .	250
7.1	PCA scores plot of the discovery data set. . . . .	251

7.2	Regulation of identified metabolites in the discovery data set. . . . .	252
7.3	Box plots of identified metabolites in the discovery data set. Data distribution for all identified metabolites with an absolute fold change of $\geq 1.5$ is shown as box plots. For each metabolite, one boxplot shows the abundances in each group. Red – virulent cluster A strains, blue – avirulent cluster B strains. . . . .	253
7.4	PCA scores plot of the discovery data set with only virulence-associated secondary metabolites included into the calculation. . . . .	256
7.5	Levels of the two non-overlapping features in the discovery data set. . . .	258
7.6	ROC curve for random forest classification model based on all features and on identified features only. . . . .	260
7.S1	PCA scores plot of the discovery data set. . . . .	272
7.S2	Volcano plot of the discovery data set. . . . .	273
7.S3	PCA scores plot with only annotated features considered in the analysis. . .	274
7.S4	PCA loadings plot of the discovery data set. . . . .	274
7.S5	Transcriptional profiles reveal no gene expression pattern associated with the virulence phenotype. . . . .	275
7.S6	Levels of the two phenazines pyocyanin and phenazine-1-carboxylic acid, in the different strains of the discovery data set. . . . .	275
7.S7	Rhamnolipid levels in the different strains of the discovery data set. . . . .	276
7.S8	Structures of annotated rhamnolipids. . . . .	276
7.S9	Credentialed peak pair of M187T6_2 and its $^{13}\text{C}$ -labeled derivative. . . .	277
7.S10	Full scan and $\text{MS}^2$ spectrum of the feature M187T6_2 in the discovery data set. . . . .	278
7.S11	Pearson's correlation of selected feature intensities to those of M187T6_2. . .	279
7.S12	Boxplots of feature intensities for M187T6_2 in the discovery and validation data set. . . . .	280
7.S13	PCA scores plot of the validation data set. . . . .	281
7.S14	Area under the ROC curve for a logistic regression model using the feature intensity of M187T6_2 to discriminate virulence phenotypes in the validation data set. . . . .	281
7.S15	Intra-group correlation of M187T6_2 with 48h survival in the <i>Galleria mellonella</i> assay. . . . .	282
7.S16	Multidimensional scaling plot visualizing tree distances between the samples of the discovery data set. . . . .	282

7.S17	Variable importance plot displaying mean decrease in accuracy and mean decrease in impurity (Gini impurity) of the random forest model constructed from the discovery data set. . . . .	283
7.S18	Percentage of correctly predicted virulence phenotype in the validation set if run 100 times independently. . . . .	283
7.S19	PCA scores plot of the cluster C data set. . . . .	284
7.S20	Percentage of correctly predicted virulence phenotype in the validation set if run 100 times independently . . . . .	284
8.1	Circle of research scope in this study. . . . .	302

## List of Tables

3.1	Alkyl quinolones identified by semi-targeted analysis. . . . .	66
5.1	Acyl-CoA substrate specificity of PqsBC. . . . .	160
5.S1	Final crystallization conditions and cryoprotectants for different PqsBC variants. . . . .	195
5.S2	Crystallographic data collection and refinement statistics for different PqsBC variants. . . . .	196
5.S3	Clustering of PqsBC heterodimers into open, closed and intermediate states.	197
5.S4	Oligonucleotides for amplification of <i>pqsB</i> and <i>pqsC</i> and for site-directed mutagenesis of <i>pqsC</i> . . . . .	198
5.S5	PqsBC variants purified in this study. . . . .	199
6.S1	Feature table of the GC-MS data. . . . .	228
6.S2	Feature table of all annotated features of the LC-MS endometabolomics data. . . . .	233
6.S3	Feature table of all annotated features of the LC-MS exometabolomics data.	236
7.1	Strains used in this study. . . . .	245
7.S1	Harvesting data for discovery batch. . . . .	285
7.S2	Harvesting data for the validation batch. . . . .	286
7.S3	Harvesting data for the additional batch. . . . .	286
7.S4	Metabolite identifications. . . . .	287
7.S5	XCMS online parameters. . . . .	294
7.S6	Transcriptomic fold changes of proteins associated with phenazine production. . . . .	295
7.S7	Transcriptomic fold changes of proteins associated with pyochelin, rhamnolipid and alkylquinolone production. . . . .	296
7.S8	Primary metabolites annotated in this study. . . . .	297

3.S1	List of theoretical $m/z$ of alkyl quinolone from C1 to C21 with saturated, mono-unsaturated and doubly unsaturated side chain used as preferred mass list in the semi-targeted LC-MS/MS run. . . . .	322
3.S2	List of features that have been identified by comparison of exact mass, retention time and MS <sup>2</sup> fragmentation to our in-house library or putatively annotated by comparison of exact mass and MS <sup>2</sup> fragmentation to metabolite databases. . . . .	323
3.S3	List of cluster ID assignments resulting from density based clustering using the OPTICS algorithm along with the colour coding used in Figure 4. . . .	331
3.S4	List of cluster ID assignments resulting from hierarchical clustering with average linkage of product ion spectra similarities. . . . .	340
3.S5	List of cluster ID assignments resulting from hierarchical clustering with average linkage of neutral loss pattern similarities. . . . .	348
3.S6	List of cluster ID assignments resulting from hierarchical clustering with average linkage of product ion spectra similarities of the spectra acquired in the alkyl quinolone-biased semi-targeted analysis. . . . .	357
3.S7	List of putatively identified features, their sum formula and isotopic pattern fit in mSigma as determined by the SmartFormula functionality of the Bruker DataAnalysis software. . . . .	369

## List of Abbreviations

<b>2-AA</b>	2-Aminoacetophenone	<b>CMP</b>	Cytidine monophosphate
<b>2-ABA</b>	2-Aminobenzoylacetate	<b>CoA</b>	Coenzyme A
<b>2-HABA</b>	2-Hydroxylaminobenzoylacetate	<b>CoASH</b>	Free reduced coenzyme A
<b>3-oxo-C12-HSL</b>	3-Oxo-dodecanoylhomoserine lactone	<b>CRISPR</b>	Clustered regularly interspaced short palindromic repeats
<b>ACP</b>	Acyl carrier protein	<b>CSV</b>	Comma-separated values (file format)
<b>AHL</b>	<i>N</i> -acylhomoserine lactone	<b>DAHPh</b>	3-Deoxy-arabino-heptulosonate 7-phosphate
<b>AI</b>	Autoinducer	<b>DDA</b>	Data-dependent acquisition
<b>AIDS</b>	Acquired immunodeficiency syndrome	<b>DHQ</b>	2,4-Dihydroxyquinoline
<b>AMP</b>	Adenosine monophosphate	<b>DI</b>	Direct infusion
<b>AQ</b>	2-Alkyl-4(1 <i>H</i> )-quinolone	<b>DNA</b>	Desoxyribonucleic acid
<b>AQNO</b>	2-Alkyl-4-hydroxyquinoline- <i>N</i> -oxide	<b>DTNB</b>	5,5'-dithiobis-2-nitrobenzoic acid
<b>ASU</b>	Asymmetric unit	<b>DTT</b>	Dithiothreitol
<b>A.U.</b>	Arbitrary units	<b>ecFabH</b>	<i>Escherichia coli</i> FabH
<b>AUC</b>	Area und the curve	<b>EDTA</b>	Ethylenediaminetetraacetic acid
<b>BAF</b>	Bruker analysis file (supposedly)	<b>EI</b>	Electron impact ionisation
<b>Bis-Tris</b>	Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane	<b>EMBRIC</b>	European Marine Biological Research Infrastructure Cluster
<b>BM2</b>	...	<b>ESI</b>	Electrospray ionisation
<b>C<sub>n:m</sub></b>	Hydrocarbon of chain length <i>n</i> with <i>m</i> double bonds	<b>Fab</b>	Fatty acid biosynthesis
<b>C4-HSL</b>	Butanoylhomoserine lactone	<b>Fad</b>	Fatty acid degradation
<b>Cas</b>	CRISPR-associated	<b>GC</b>	Gas chromatography
<b>CC</b>	Correlation coefficient	<b>GMP</b>	Guanosine monophosphate
<b>CE</b>	Capillary electrophoresis	<b>GNPS</b>	Global Natural Product Social Molecular Networking
<b>CF</b>	Cystic fibrosis	<b>HAI</b>	Healthcare-associated infection
<b>CID</b>	Collision-induced dissociation	<b>HEPES</b>	4-(2-Hydroxyethyl)-1-piperazine-ethanesulfonic acid
<b>CluMSID</b>	Clustering of MS <sup>2</sup> spectra for metabolite identification	<b>HGF</b>	Helmholtz-Gemeinschaft deutscher Forschungszentren

**HHQ** 2-Heptyl-4-hydroxyquinoline

**HPLC** High performance liquid chromatography

**HQ** Hydroxyquinoline

**HQNO**  
2-Heptyl-[1*H*]-4-quinolone-*N*-oxide

**HSL** Homoserine lactone

**HTML** Hypertext markup language

**HZI** Helmholtz-Zentrum für Infektionsforschung

**ID** Identifier

**IMAC** Immobilised metal ion affinity chromatography

**IPTG** Isopropyl- $\beta$ -D-thiogalactopyranoside

**Kan** Kanamycin

**LB** Lysogeny broth

**LC** Liquid chromatography

**M** Parent molecule

**MASST** Mass Spectrometry Search Tool

**MD** Molecular dynamics

**MDS** Multidimensional scaling

***m*/FabH** *Micrococcus luteus* FabH

**MOPS** 3-Morpholinopropane-1-sulfonic acid

**MPD** 2-Methyl-2,4-pentanediol

**MS** Mass spectrometry

**MS<sup>2</sup>** Tandem mass spectrometry

**MSMLS** Mass Spectrometry Metabolite Library of Standards

**MS/MS** Tandem mass spectrometry

**MSTFA** *N*-Methyl-*N*-(trimethylsilyl)trifluoroacetamide

***mt*FabH** *Mycobacterium tuberculosis* FabH

**MvFR** Multiple virulence factor regulator

**MWCO** Molecular weight cut off

***m/z*** Mass-to-charge ratio

**mzML** Mass-to-charge ratio markup language

**mzXML** Mass-to-charge ratio extended markup language

**NAD** Nicotinamide adenine dinucleotide

**n.d.** Not detectable/not determined

**netCDF** Network common data format

**NMR** Nuclear magnetic resonance (spectroscopy)

**nrpg** Normalised reads per gene

**OD<sub>600</sub>** Optical density at 600 nm

**OPTICS** Ordering points to identify the clustering structure

***p*<sub>adj</sub>** Adjusted *p*-value

**PCA** Principal component analysis

**PCR** Polymerase chain reaction

**PDB** Protein Data Bank

**PDF** Portable document format

**PE** Phosphatidylethanolamine

**PEG** Polyethyleneglycol

**PERMANOVA** Permutational multivariate analysis of variation

**PETRA** Positron-electron tandem ring accelerator

**PNG** Portable network graphics

**ppm** Parts per million

**PQS** *Pseudomonas* quinolone signal

**PSSweb** Protein Structural Statistics Web Server

**QNO** Quinolone-*N*-oxide

**QS** Quorum sensing

**QSSM** Quorum sensing signalling molecule

**QTOF** Quadrupole time-of-flight (mass spectrometer)

<b>RF</b> Release factor	<b>TMS</b> Trimethylsilyl (group)
<b>Rha</b> Rhamnose, rhamnosyl	<b>Tn</b> Transposon
<b>ROC</b> Receiver operating characteristics	<b>Tris</b> Tris(hydroxymethyl)aminomethane
<b>RP</b> Reversed phase	<b>TSV</b> Tab-separated values (file format)
<b>rpm</b> Revolutions per minute	<b>UDP</b> Uridine diphosphate
<b>RT</b> Retention time	<b>UMP</b> Uridine monophosphate
<b>SDS</b> Sodium dodecyl sulfate	<b>UPLC</b> Ultra-performance liquid chromatography
<b>SEC</b> Size exclusion chromatography	<b>VAP</b> Ventilator-associated pneumonia
<b>SLS</b> Swiss Light Source	<b>VIP</b> Variable importance in projection
<b>TB</b> Terrific broth	<b>v/v</b> Volume per volume
<b>TCEP</b> Tris(2-carboxyethyl)phosphine	<b>WHO</b> World Health Organisation
<b>TEV</b> Tobacco etch virus	



# 1 | General Introduction

## 1.1. *Pseudomonas aeruginosa*

*Pseudomonas aeruginosa* is a Gram-negative rod-shaped  $\gamma$ -proteobacterium of high scientific and medical relevance as a pathogen, model organism and ubiquitous environmental microbe. It is the type species of the *Pseudomonas* genus, one of the most complex and with more than 140 species the largest of the Gram-negative genera [1]. Walter Migula first described the genus *Pseudomonas* at the end of the 19<sup>th</sup> century and named it based on the cellular morphology of the genus' members [2–4]. While Migula himself did not elaborate on its etymology, the name *Pseudomonas* is thought to derive from the greek words  $\psi\epsilon\upsilon\delta\acute{\eta}\varsigma$  (*pseudēs*; wrong, false) and  $\mu\omicron\nu\acute{\alpha}\varsigma$  (*monás*; unit, individual) [5]. However, it is not clear whether *-monas* really refers to 'unit' as in unicellular organism and an alternative explanation says that it alludes to the *Monas* genus, a group of nanoflagellate protists from the Crysophyceae family that display a certain resemblance to *Pseudomonas* species in terms of size and motility [4]. The epithet *aeruginosa*, from Latin *aerūgō* for 'verdigris', was coined by Julius Schröter in 1872, alluding to the blue-green colour of *P. aeruginosa* colonies and cultures [6]. In the early days of microbiology, it also went by the names of *Pseudomonas pyocyaneus* and *Bacillus pyocyaneus* which also refer to the pigments produced by the bacterium [7]. Although *P. aeruginosa* has been identified as an infectious agent around the turn of the last century [8, 9], it has only been considered a significant human pathogen since the 1950s [7, 10, 11].

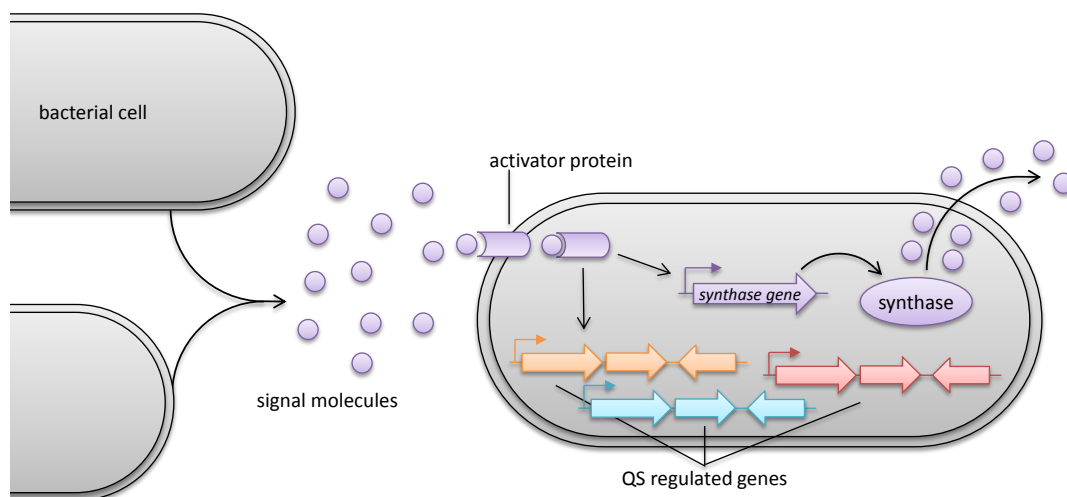
### 1.1.1. Relevance as an opportunistic pathogen

*P. aeruginosa* is able to infect a multitude of different plant and animal species, including the model organisms *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Galleria mellonella*, *Drosophila melanogaster* and *Mus musculus* [12], but it has gained most importance as a human pathogen and causative agent of nosocomial infections, ranking among the most frequent germs associated with healthcare-associated infections (HAIs) [13]. While mostly harmless for healthy humans, *P. aeruginosa* is capable of infecting nearly

all human tissues as a facultative pathogen in predisposed individuals [14]. Along with general immunocompromise, e. g. caused by AIDS or neutropenia under chemotherapy, severe burn wounds as well as cystic fibrosis (CF) and other lung pathologies are the most eminent risk factors for *P. aeruginosa* infections [15]. The most typical clinical manifestation is pneumonia, with or without bacteremia, which is often associated with (contaminated) medical devices [16, 17]. For instance, in intensive care patients on artificial ventilation, intubation damages natural barriers and enables *P. aeruginosa* to cause ventilator-associated pneumonia (VAP) which is difficult to manage and often fatal [18]. As *P. aeruginosa* is a ubiquitous environmental bacterium that can persist on various hospital surfaces [19, 20], the number of infections in susceptible populations is high [15]. In 80-95% of CF patients, chronic (co-)infections with *P. aeruginosa* lead to respiratory failure, the major cause of death for these patients which can only be treated by lung transplantation [21]. Effective pathogen control or elimination is essential to prevent irreversible lung damage; however, antibiotic therapy is problematic as *P. aeruginosa* is endowed with several natural and additional acquired drug resistances and with an outer membrane that impairs drug uptake by its poor permeability [22]. Furthermore, *P. aeruginosa* is able to form biofilms that render it even less susceptible to antibiotics, disinfectants and the host immune response [23, 24]. As the elderly are more susceptible to *P. aeruginosa* colonisation and infection, the pathogen is expected to further gain clinical importance in aging populations, i. e. in most of the world [25].

### 1.1.2. Metabolic versatility

*P. aeruginosa* is not only a human pathogen but also infects other animals and plants [26]. Moreover, it can be found in almost all environments inhabited by humans and thrives in various environmental niches as different as sweet melon rhizosphere and dolphin gastric juice [26, 27]. To survive under such diverse circumstances, the bacterium has developed a broad range of metabolic features that enable it to metabolise a multitude of organic substrates [28]. This metabolic versatility of *Pseudomonas* species has been described as early as 1926, when L. E. den Dooren de Jong found that they are capable of degrading various organic compounds [29, 30]. Today it is known that *P. aeruginosa* is able to use a variety of carbon sources [31] and that it can flexibly adapt its central carbon metabolism to use a combination of the Entner-Doudoroff, Embden-Meyerhof-Parnas and pentose phosphate pathways [32]. Multiple metabolic pathways are modified to meet the needs of a specific situation such as multi-species competition in CF airway infections [33].

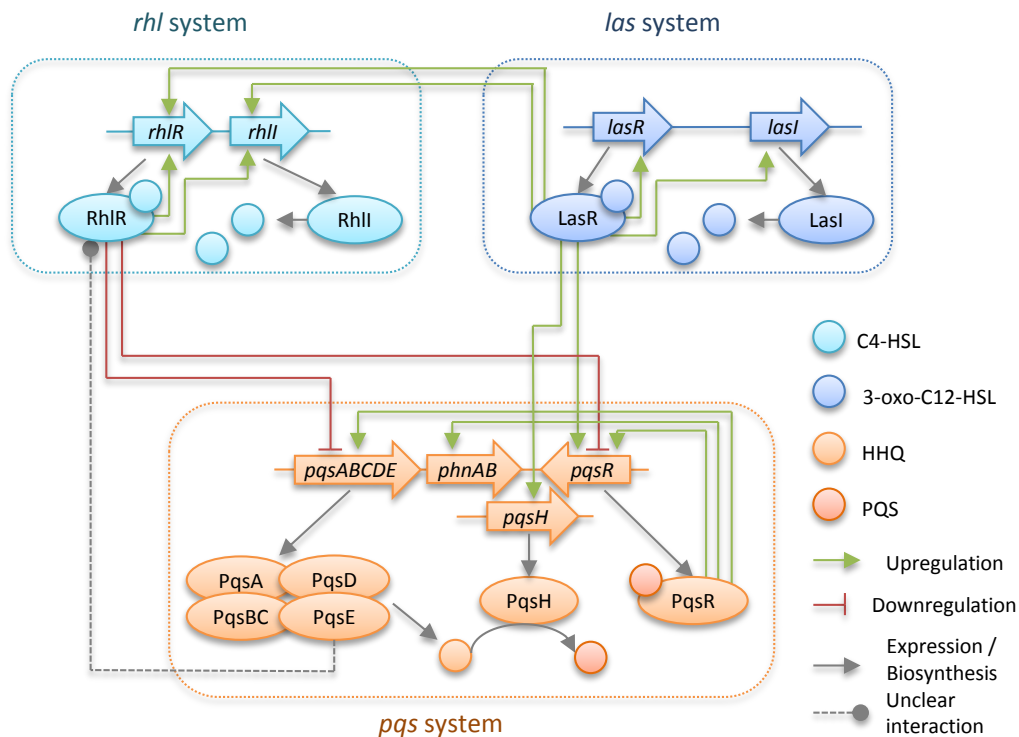


**Figure 1.1.:** The general principle of quorum sensing. QSSM synthases produces QSSMs that transition to the extracellular space in the environment of the bacterium. The molecules reach other bacterial individuals by means of diffusion. Once a concentration threshold is passed, they are sensed by specific receptor proteins, which, in turn, serve as activators or repressors for the gene expression of diverse QS regulated genes. The expression of synthase genes is activated as well, resulting in autoinduction and increased QSSM production.

Metabolic versatility is not limited to catabolic pathways, though. *P. aeruginosa* also produces a very rich arsenal of secondary metabolites [26]. Many of them are associated with virulence and in turn interconnected with other metabolic processes [34]. For instance, the chemodiversity of small molecules that *P. aeruginosa* uses to communicate with conspecific individuals is unparalleled.

### 1.1.3. Quorum sensing

Quorum sensing (QS) denotes the bacterial interindividual communication by means of diffusible secreted small molecule secondary metabolites [35]. The production of quorum sensing signalling molecules (QSSM) remains on a basal level until a concentration threshold (the quorum) is reached, whereupon the expression of various genes is influenced, including the QS system itself (autoinduction) [35, 36]. This population density dependent process serves diverse regulatory purposes (Figure 1.1). In *P. aeruginosa*, which harbours one of the most complex and most intensively studied QS systems, multiple interconnected signalling systems regulate, among other phenomena, virulence and persistence on the gene expression level, thus enabling the bacteria to flexibly adapt to their host [37]. The three main QS systems and their interdependences are depicted in Figure 1.2. The existence of a fourth QS systems which relies on 2-(2-hydroxyphenyl)-

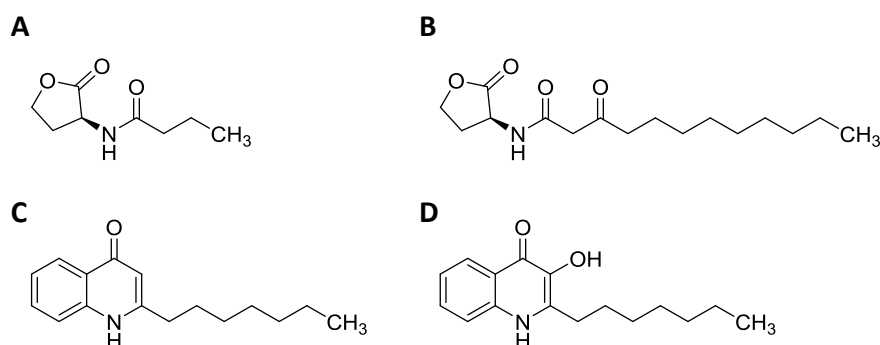


**Figure 1.2.:** Schematic representation of the *las*, *rhl* and *pqs* quorum sensing systems in *Pseudomonas aeruginosa* and their interconnections. Modified from [42] with information from [43] and [37]. Large arrows, large ellipses and small circles represent genes, proteins and QS signal molecules, respectively, and are colour-coded according to the QS system they belong to. Green arrows signify up-regulation and red blocked lines down-regulation of genes by proteins. Gray dashed lines symbolise unclear interactions. Grey arrows indicate protein expression (gene to protein) or QS signal molecule biosynthesis (protein to small molecule).

thiazole-4-carbaldehyde, called IQS, has been postulated but not universally accepted [38, 39]. Considering the importance of QS, QS inhibition is believed to be a promising approach for novel anti-virulence drugs, so-called pathoblockers, that do not kill the bacteria but limit the harm they can cause to the host [40–42].

### The *rhl* and *las* QS systems

The canonical QS system in Gram-negative bacteria uses *N*-acylhomoserine lactones (AHLs) as QSSM. AHLs are biosynthesised from fatty acids of variable lengths that are connected to homoserine via a peptidic bond by LuxI-type enzymes and sensed by LuxR-type receptors [36]. The receptor-AHL complex binds to bacterial DNA and



**Figure 1.3.:** Chemical structures of the most important quorum sensing signalling molecules of the *las*, *rhl* and *pqs* quorum sensing systems in *Pseudomonas aeruginosa*. **A:** Butanoylhomoserine lactone (C4-HSL), *rhl* system. **B:** 3-Oxo-dodecanoylhomoserine lactone (3OC12-HSL), *las* system. **C:** 2-Heptyl-4-quinolone (2-heptyl-4-hydroxyquinoline, HHQ), *pqs* system. **D:** 2-Heptyl-3-hydroxy-4-quinolone (*Pseudomonas quinolone signal*, PQS), *pqs* system.

serves as a transcriptional regulator of various virulence genes [37]. In *P. aeruginosa*, the *las* system, consisting of LasI and LasR, controls via the QSSM *N*-3-oxo-dodecanoylhomoserine lactone (3-oxo-C12-HSL) the production of several elastases, exotoxin A and alkaline phosphatase. The *rhl* system, that in turn consists of RhlI and RhlR and employs *N*-butanoylhomoserine lactone as QSSM, regulates i. a. the biosynthesis of rhamnolipids [44–46]. Both systems are autoinductive and connected by LasR, which can act as a transcription factor for RhlR [37]. Beyond this, AHL dependent regulation is much more complex and even after 20 years of research not entirely understood. For example, there are LuxR family proteins for which no corresponding LuxI-type equivalent exists but which are still able to modulate the AHL signal [47] and only recently it has been found that *P. aeruginosa* RhlR has an alternative ligand presumably produced by PqsE, an enzyme from the *pqs* QS system [43]. Moreover, AHLs have been shown to be able to alter gene expression in the absence of any LuxR-type receptors [48], which further illustrates the complexity of AHL-dependent QS.

### The *pqs* QS system

Beyond LuxI/LuxR-type QS, *P. aeruginosa* also possesses another QS system, the *pqs* system, which is based on alkylquinolones (AQs) like 2-heptyl-4-quinolone (HHQ) or 2-heptyl-3-hydroxy-4-quinolone (*Pseudomonas quinolone signal*, PQS) [49]. PQS, HHQ and their congeners exert their influence on gene expression after binding to the response regulator PqsR (synonym: MvfR), thereby up-regulating the expression of

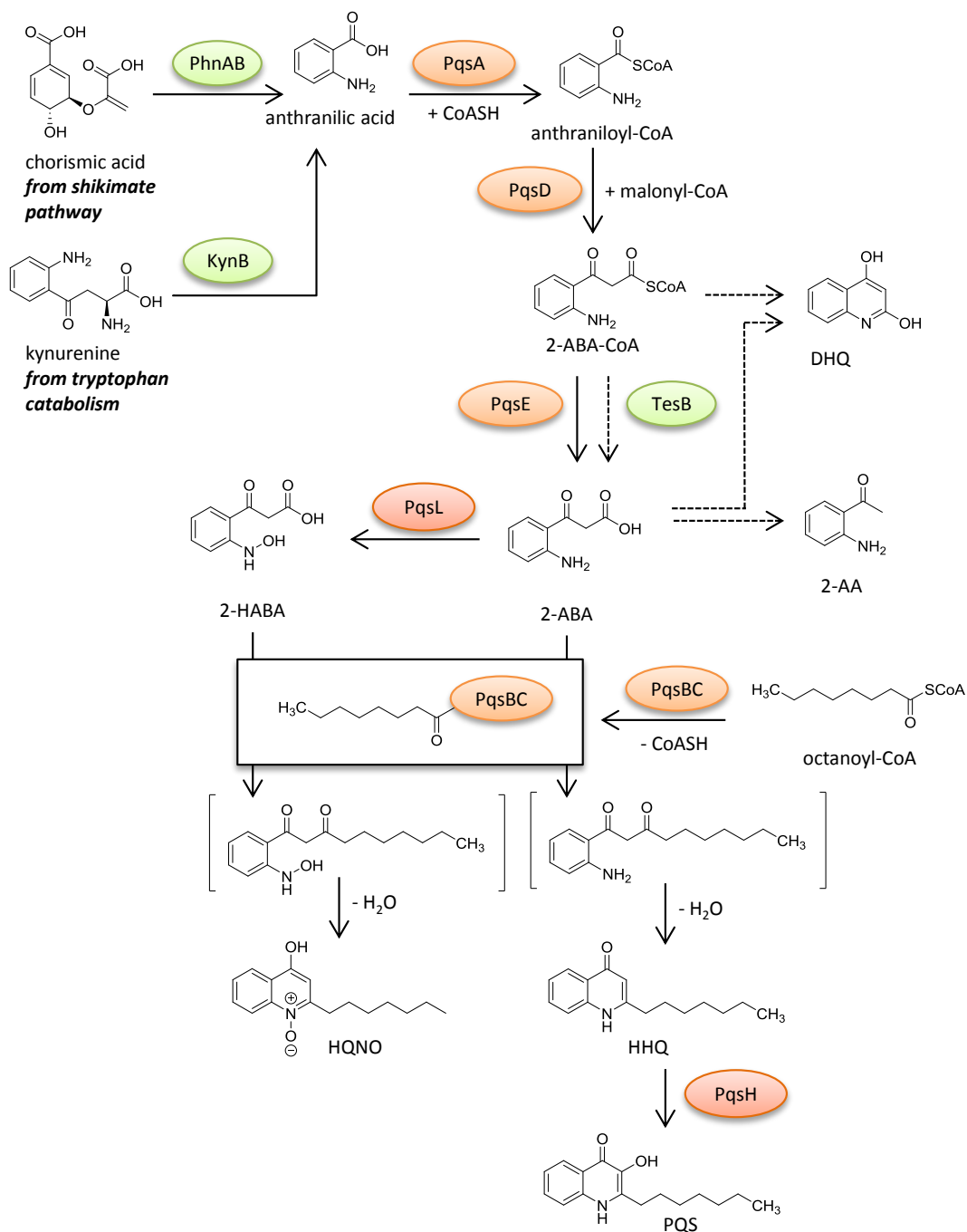
*pqsABCDE* and *phnAB*. Thus, along with autoinduction, the positive regulation of phenazine biosynthesis is one of the main tasks of the *pqs* system [37]. The biosynthesis of PQS is achieved by the proteins encoded in the *pqs* operon, PqsABCDE, as well as PqsH. It has long been thought that PqsABCD catalyse the condensation of anthranilate and  $\beta$ -ketofatty acids to HHQ, which is then oxidised by PqsH to yield PQS [50, 51]. In the meantime, however, it has been demonstrated that the biosynthesis is effected from anthranilate, malonyl-CoA and fatty acids without  $\beta$ -keto group [52]: Anthranilate is condensed with coenzyme A by PqsA and PqsD transfers a malonyl group from malonyl-CoA onto anthraniloyl-CoA, yielding 2-aminobenzoylacetyl-CoA. After hydrolysis, 2-aminobenzoylacetate (2-ABA) can be *N*-oxidised by PqsL to direct it into quinolone-*N*-oxide (QNO) biosynthesis. Both 2-ABA and its congener 2-HABA are condensed with a CoA-activated fatty acid by PqsBC, followed by spontaneous ring formation yielding compounds of the HHQ and HQNO family, respectively (Figure 1.4). While AQs with heptyl (HHQ, HQNO) or nonyl (NHQ, NQNO) chains are most frequently reported, a multitude of related compounds with varying alkyl chain length have been found in *P. aeruginosa* [53, 54]. The question why and how this biosynthetic pathway yields so many similar but different AQ congeners is still not fully answered.

The role of the fifth protein encoded in the *pqs* operon, PqsE, has long been elusive, although it had been identified early on as an important player in the production of virulence factors such as rhamnolipids, lectins, pyocyanin and HCN [37]. Drees and Fetzner could show that it takes part in AQ biosynthesis by acting as a thioesterase for the intermediate 2-aminobenzoylacetyl-coenzyme A [55]. Recent studies suggest that PqsE might be involved in the production of an unidentified alternative ligand of RhIR thereby forming another crosslink in *P. aeruginosa*'s complex QS machinery [43, 56].

AQs are not only QSSM but also serve different purposes [57]. For instance, QNOs possess antibiotic properties and have been shown to interfere with growth and virulence of other bacterial species [58]. Moreover, PQS takes part in the formation of membrane vesicles [59] and in iron homeostasis, where it chelates  $\text{Fe}^{3+}$  [60]. To summarise, AQs are important regulators and effectors in many aspects of pseudomonal physiology, especially with regard to virulence.

#### 1.1.4. Virulence

Virulence is defined as the ability of a parasitic organism to harm its host, hence to cause disease [61]. It is well known that virulence in *P. aeruginosa* is regulated by QS



**Figure 1.4.:** Schematic of the alkylquinolone biosynthesis in *Pseudomonas aeruginosa*. Oval shapes represent biosynthetic enzymes with orange shapes for Pqs enzymes and green shapes for other enzymes. Solid arrows indicate main biochemical reactions and dashed arrows biochemical side reactions. Square brackets signify unstable intermediates.

[44] and it has also been shown in the clinical setting that a correlation between QS activity and virulence exists in patients suffering from *Pseudomonas pneumonia* [62].

Typical pseudomonal virulence factors are redox-active phenazines that are responsible for the fluorescence of the bacteria [63]. Pyocyanin, the most extensively studied phenazine in *P. aeruginosa*, can be detected in large quantities in the sputum of infected CF patients and, along with other virulence factors, plays an important role in pathogenesis by generating reactive oxygen species that subject the host cells to oxidative stress [64].

Rhamnolipids are surface-active glycolipids that are accountable for the establishment of VAP in intensive care patients by disturbing the integrity of the lung epithelium and hence facilitating paracellular invasion of the bacteria [62, 65]. In addition, they enable *P. aeruginosa* to eliminate polymorphnuclear neutrophil leukocytes, consequently impairing the most important immune response in CF lungs [66, 67].

Furthermore, *P. aeruginosa* produces a range of peptidic virulence factors: The elastase LasB is involved in the destruction of connective tissue and the invasion into the epithelium, while exotoxin A inhibits eukaryotic protein biosynthesis and hence induces apoptosis. Alkaline protease, katalase and superoxiddismutase as well as further exoenzymes also contribute to *P. aeruginosa* virulence [68].

Besides, *P. aeruginosa* can produce hydrogen cyanide, which inhibits the host's cytochrome-C-oxidase und thereby interrupts the cellular respiratory chain. Interestingly, HCN can be detected in exhaled breath of infected patients and thus serve as a diagnostic biomarker [69].

Biofilm formation, i. e. the encapsulation of bacterial cells into an extracellular matrix that adheres to a surface, is also considered part of the *P. aeruginosa* virulence strategy since it enables the bacterium to persist in different infection sites and evade most defense mechanisms of the immune system [70]. *P. aeruginosa* biofilms are formed by secretion and maturation of polysaccharides, proteins and DNA [71]. The transition from planktonic to biofilm lifestyle follows a complex regulation that includes switches in metabolic profiles [72].

The fact that many pseudomonal virulence regulators and effectors are small-molecule secondary metabolites makes *P. aeruginosa* an interesting pathogen to study the interplay of metabolism and virulence by means of metabolomics.



## 1.2. Metabolomics

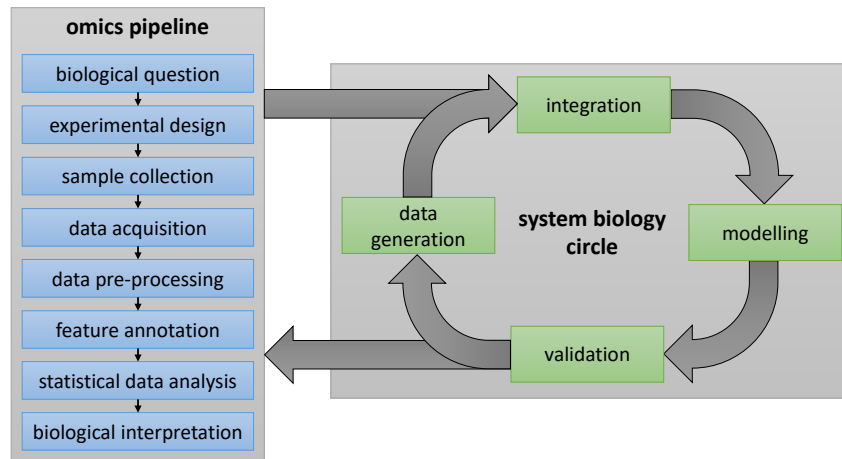
The term *metabolomics* refers to an interdisciplinary field of study that combines aspects of analytical chemistry, mass spectrometry, nuclear magnetic resonance spectroscopy and data analysis to investigate the entirety, a certain subset or characteristic profiles of low molecular weight compounds (<1500 Da) in a biological system under defined conditions [73].

The field is rooted in analytical chemistry, biochemistry, clinical chemistry, natural product chemistry as well as in the other omics disciplines, especially functional genomics. An often-cited historical ancestor of metabolomics experiments is the Urine Wheel by Ullrich Pinder from 1506 [74] that classifies urine qualities and links them to disease states, thus enabling diagnosis by analysis of a biofluid. This principle is still at the core of many modern metabolomics experiments, especially in the medical sciences [75]. Further milestones in the evolution of metabolomics were the introduction of techniques for the labelling of metabolites; first by means of chemical labelling of fatty acid precursors by Knoop in 1904 [76] and later by integration of stable isotopes by Shemin and Rittenberg in 1945 [77]. These pivotal experiments enabled scientists to study metabolism in chemical detail and paved the way for modern analytical biochemistry. The first study that followed a contemporary metabolomics approach—still *avant la lettre*—was published by Nobel laureate Linus Pauling and collaborators in 1971. They used gas chromatography—mass spectrometry to quantify metabolites in urine and breath and achieved an analytical coverage of 250 and 280 metabolites, respectively [78]. In 1998, the term ‘metabolome’ was coined by Oliver *et al.* [79] and one year later, Nicholson *et al.* described ‘metabonomics’ as the study of metabolite compositions in perturbed systems [80]. Historically, *metabolomics* was contrasted with *metabonomics* and described only the study of biological systems in their natural state, but is now mostly used as the general term for the systems biology investigation of metabolites [81].

### 1.2.1. Systems biology from an analytical chemistry perspective

Systems biology tries to understand functions and processes in biological entities by comprehensively examining their molecular components and their interaction networks [82]. While the bottom-up approach does so by deducing from information about parts or subsystem, top-down systems biology integrates global data from biological experiments and uses this information to construct models that explain biological phenomena

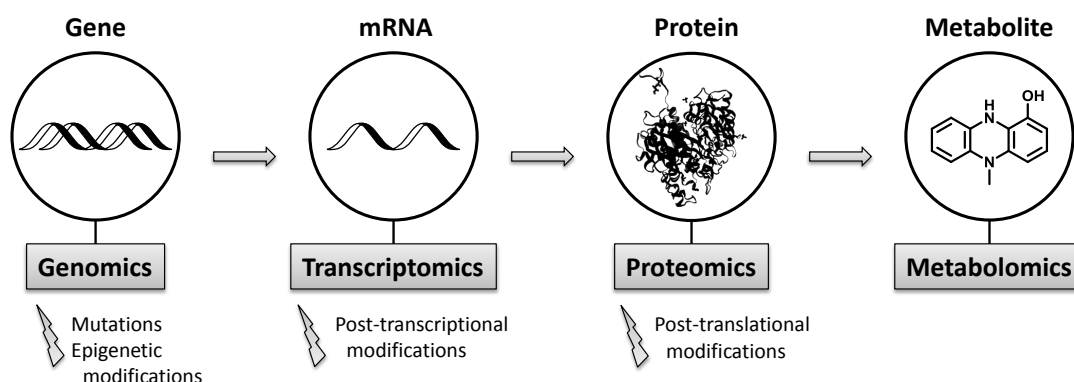
[82, 83]. This data is mostly generated by so-called omics technologies, i. e. experiments that characterise that complete pool of a class of biological molecules [84]. Genomics [85], transcriptomics [86], proteomics [87] and metabolomics are generally regarded as the core omics disciplines, while more specific fields like lipidomics [88] or metallomics [89] exist as well. The interrelation of omics experiments and systems biology is shown in Figure 1.5.



**Figure 1.5.:** Interrelation of omics experiments and systems biology. The systems biology circle consists of data generation, integration, modelling and validation. Omics data is integrated and used for model building. The validation of the model requires new data that is again integrated and used to improve the model. The figure is adapted from [90].

Metabolomics as a scientific discipline is complementary to the other omics technologies genomics, transcriptomics and proteomics (Figure 1.6). The identification and quantitation of metabolites produces a functional snap-shot of a biological system's biochemistry and offers the best correlation to the molecular phenotype, since metabolome data—unlike genome or the proteome data—directly represent substrates and products of biochemical processes in the cell, thus disclosing the final outcome of gene expression and enzyme activity [91]. Furthermore, metabolomics is not limited to the detection of endogenous metabolites but also covers exogenous substances and thereby enables the study of the interplay of genome and environment [73].

In contrast to genes, transcripts and proteins, metabolites are a chemically extremely diverse group of analytes. The metabolome comprises monomers, intermediates, cofactors, signalling molecules, structural elements and many more [92]. They are commonly classified as either primary or secondary metabolites, the latter being dispensable for basic vital functions [93]. Metabolites vastly differ in polarity, size and reactivity, as



**Figure 1.6.:** The Omics Cascade according to [91]. Genomics studies biological variation at the most basic level and can reveal fundamental differences between individuals. Yet, it cannot fully account for somatic mutations and misses out on all factors that influence the expression of genes, e. g. epigenetic modifications. Transcriptomics measures gene expression and thus covers the effects of gene regulation while it does not provide data on how many of the transcripts are translated into functional proteins (e. g. by the effects of posttranscriptional modifications). Similarly, proteomics analyses abundances of proteins, hence describing biological systems on a phenotypic level, but is limited by its inability to gauge the activity of proteins, which is affected by posttranslational modifications, regulatory mechanisms and environmental factors such as substrate and product concentrations. Metabolomics accounts for these as it quantifies products, intermediates and substrates of enzymatic reactions along with exogenous molecules.

illustrated by the *Staphylococcus aureus* metabolome that comprises, i. a. amino acids, peptides, carbohydrates, lipids, carboxylic acids, nucleosides, aromatic compounds and many more [94]. Moreover, metabolite concentration ranges can run the gamut from pg/ml to mg/ml [95]. The large variety of analytes requires a correspondingly large variety of analytical methods and study designs, which makes metabolomics a heterogeneous field of research [96].

### 1.2.2. Targeted and untargeted metabolomics

Metabolomics approaches are generally classified as *targeted* or *untargeted*. Targeted metabolomics refers to the analysis of a previously defined set of metabolites and often includes absolute quantification [97]. This approach is generally suited for classic hypothesis-driven research and requires previous knowledge about chemical properties of the metabolites to be analysed. If this information is known, modern mass spectrometry technologies enable the distinction of very similar compounds and thus provide detailed insights into metabolic processes in biological systems [73].

Untargeted or non-targeted metabolomics approaches aim to analyse the entire metabolome in the most comprehensive way possible. To that end, both NMR and MS based

methods can be used; however, liquid chromatography coupled to high resolution mass spectrometry is the most eminent technology in the field [91]. Owing to the complexity and heterogeneity of the metabolome, there is no standardised analytical methodology to capture the complete metabolome [98]. Without previous knowledge about the exact chemical composition of samples, the identification of detected metabolites is another substantial problem in untargeted metabolomics [99]. On the other hand, untargeted studies can be conducted with minimal assumptions, hence preventing biases which can be introduced by the selection of analytes [81]; and after years of technical improvement in high resolution mass spectrometry and progress in community curated spectral data bases, hundreds to thousands of metabolites can potentially be identified in a biological sample [91, 99–101]. Untargeted metabolomics generates huge amounts of data that have to be handled with multivariate statistics and other advanced data analysis methods [102]. The data can be analysed on a functional level or used to generate classification models and thus aid in the generation of new hypotheses. These can, in turn, be tested by targeted metabolomics or other types of experiments [81].

### 1.2.3. Metabolomics technologies

Due to the high complexity and diversity of possible samples and analytes, modern metabolomics comprises a multitude of analytical and pre-analytical methods. Generally, they require more or less extensive sample preparation and most of the analytical methodologies rely on chromatographic separation [103]. While other detection principles exist, mass spectrometry and nuclear magnetic resonance are the predominant methods for the qualitative and quantitative analysis of metabolites [104].



**Figure 1.7.:** Schematic representation of a metabolomics workflow. Samples are collected or generated, then subjected to sample preparation procedures such as extraction or purification to yield samples suitable for chemical analysis by mass spectrometry or nuclear magnetic resonance spectroscopy. The resulting analytical data has to be preprocessed and processed before it can be analysed by means of chemometric, statistical and systems biology methods and finally interpreted in the biological context.

### Sample preparation

Sample preparation strategies depend on the type of sample and on the experimental design. Methodologies range from minimal preparation, e. g. direct infusion from cultivation flasks [105] or desorption electrospray ionisation–mass spectrometry of bacterial colonies [106], to sophisticated purification and enrichment strategies [107]. Compromises have to be made to balance potential improvements in selectivity, sensitivity and robustness on the one hand and throughput on the other hand [108].

In most experimental settings, the crucial basic steps of sample preparation are to halt metabolic reactions as many metabolites have very short half-lives [109] and to separate the relevant part of the sample from irrelevant parts, e. g. blood serum from cellular components or bacterial biomass from cultivation fluid. Stopping the cellular metabolism and preventing metabolite degradation by enzymatic or chemical reactions is commonly referred to as ‘quenching’. Quenching is indispensable to ascertain that metabolite concentrations measured in a metabolomics experiment actually represent the biochemistry of the live sample [110]. Quenching approaches rely on quick temperature changes and/or organic solvents that denature enzymes in the sample and can be combined with the separation step [110, 111]. Standard methods for separation include centrifugation [112] and filtration [113]. To enable extraction of metabolites, cellular or tissue samples have to be lysed, which is done chemically by solvent addition, mechanically, e. g. by bead beating, or via ultrasound [114].

Extraction, including protein removal, is usually performed on the samples to facilitate instrumental analysis by removing potentially interfering substances and irrelevant analytes and by enrichment of analytes of interest. Liquid extraction is most commonly used in metabolomics with either hydrophobic or hydrophilic solvents (e. g. chloroform or methanol–water mixtures, respectively), depending on the chemical properties of the target analytes [107]. If strong matrix effects or low analyte concentrations are observed, solid phase extraction has been shown to be a valuable alternative or complementary extraction method [115].

The larger the sample size and the more complex the study design, the more important are sample preparation, handling and storage. Therefore, community guidelines for pre-analytical processing and biobanking of metabolomics sample have been developed [116].

**Separation: LC, GC, CE**

Metabolomics experiments that use NMR spectroscopy for metabolite identification and quantification as well as some specialised mass spectrometry techniques do not rely on the separation of metabolites from extracted samples. These exceptions include direct injection mass spectrometry [105] and (imaging) mass spectrometry by desorption electrospray ionisation, matrix-assisted laser desorption ionisation or related technologies [117]. For most mass spectrometry-based experimental setups, however, separating the individual components of metabolite mixtures is of high importance, e. g. to avoid signal distortion caused by ion suppression and to be able to distinguish isobaric analytes [118]. In the analysis of complex extracts, separation performance can be the decisive factor for analytical coverage and accuracy and hence determine the value of the information gained in the experiment [119].

Common separation techniques include liquid chromatography (LC), gas chromatography (GC) and capillary electrophoresis (CE) [120]. While the latter has so far remained a niche technology, LC and GC are fundamental for metabolomics.

**Liquid chromatography** LC can be used to separate a large variety of different analytes without pre-analytical derivatisation, which makes it a versatile separation method in the bioanalytical sciences. Ultra high performance liquid chromatography (UHPLC) is now the standard LC method used in both targeted and untargeted metabolomics [121]. Its main advantages are its wide range of analytes—also due to different possible column chemistries—, its easy interfaceability with different types of mass spectrometers via electrospray ionisation (ESI) and the large amount of available scientific literature on LC-MS metabolomics [121, 122].

Reversed phases, i. e. nonpolar stationary phases, are most often used in LC-MS metabolomics. Octadecylsilyl (RP18) columns are highly prevalent in all metabolomics applications and suited for many types of samples [123]. While the metabolome coverage of reversed phase LC can be improved by the correct choice of mobile phases [124], some metabolite classes are not easily separable on RP18 or comparable columns. If highly polar metabolites are of interest, RP separations can be complemented by hydrophilic lipophilic interaction chromatography (HILIC). HILIC columns have been designed for the analysis of polar compounds and the use of aqueous, low salt eluents renders it particularly suitable for ESI-MS [125].

**Gas chromatography** In contrast to LC, GC is suited for small compounds (<650 Da), especially for intermediates of the primary metabolism [126]. Analytes must either be volatile or derivatisable into volatile compounds, e. g. by silylation using *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamid (MSTFA) combined with methoxamination, and must not be thermolabile [126, 127]. These limitations render GC unsuitable for some research areas, such as bacterial secondary metabolism. On the other hand, LC methods often lack coverage of important primary metabolites, which makes GC-MS and LC-MS complementary techniques in metabolomics.

Unlike LC, GC is often coupled to mass spectrometers that use hard ionisation, which has various implications for data analysis and specifically metabolite identification. Various software and databases for annotation of GC-electron impact (EI)-mass spectra exist and GC-EI-MS fragmentation is considered highly reproducible [126].

### **Detection: Mass spectrometry and NMR spectroscopy**

Although some of the first metabolomics experiments were conducted with infrared spectroscopy [79], mass spectrometry and NMR spectroscopy are now the detection methods of choice.

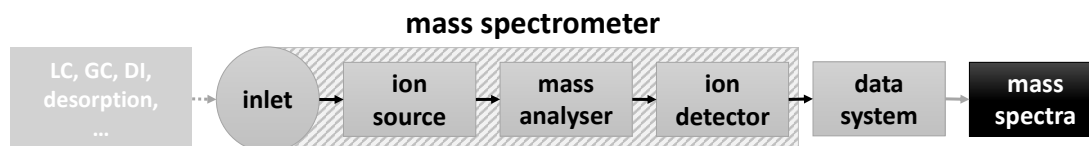
**NMR spectroscopy** As mentioned above, NMR spectroscopy can be directly applied to complex mixtures without prior separation and also sample preparation can be minimal which makes it a fast and experimentally simple method for metabolomics. Other advantages are that the detection principle is universal, i. e. not discriminating metabolite classes based on chemical properties, that measurements are non-destructive, enabling e. g. time series or *in vivo* metabolomics, and that absolute quantification is achievable without using multiple internal standards [128, 129].

The main disadvantage of the technology is its low sensitivity and the resulting poor metabolome coverage compared to MS-based approaches [129]. Along with the need for complex data analysis and expensive instrumentation, this limitation restricts the suitability of NMR spectroscopy for many metabolomics experiments, while it can play to its strength whenever speed of analysis and absolute quantification are the key requirements.

**Mass spectrometry** Complementary to NMR metabolomics, MS-based approaches excel due to broad metabolome coverage, high sensitivity, better flexibility and customisability and data analysis strategies that are more suitable for automation. On

the downside, absolute quantification is challenging, measurements are destructive, not all metabolites can be detected with the same sensitivity, and reproducibility between batches and platforms can be problematic [130].

Various MS methods are used in the field of metabolomics. The general setup is shown in Figure 1.8.



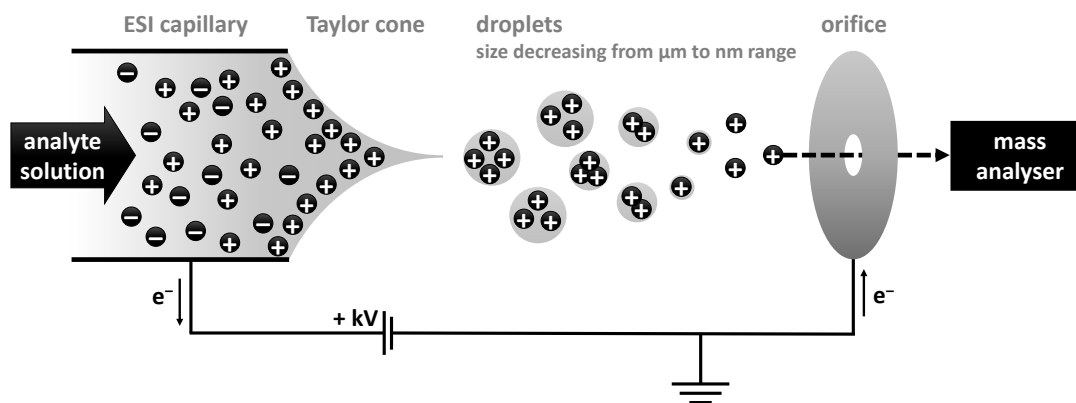
**Figure 1.8.:** Schematic of the general setup of a mass spectrometer. The analytes that have either been separated by chromatography or are directly infused or are desorbed from the sample are introduced into the mass spectrometer via the inlet. They are then ionised in the ion source, separated by their mass-to-charge ratio ( $m/z$ ) in the mass analyser and detected by the ion detector. The signals from the ion detector are transmitted to a data processing unit which returns MS data in the form of mass spectra. LC, liquid chromatography; GC, gas chromatography; DI, direct infusion.

**Ionisation** As mass spectrometry operates on charged ions in the gas phase, uncharged molecules have to be ionised prior to analysis and detection. The choice of the ionisation method for a metabolomics experiment depends on the sample type and the analytes, on the type of chromatography and mass spectrometry used and on the research question. In GC-MS, the analytes are already in the gas phase and can directly be ionised. The most common techniques are electron ionisation (EI, also called electron impact ionisation) and different types of atmospheric pressure ionisation (API) [131]. EI uses a beam of accelerated electrons (mostly at 70 eV) to ionise precursor molecules, which in most cases results in fragmentation of the parent molecule into charged and neutral fragments, hence it is considered a hard ionisation method [132]. The interpretation of EI fragmentation patterns is a long-established skill of analytical chemists by use of which various structural information about the analyte can be deduced [133]. Multiple spectral databases exist for the annotation of GC-EI-MS spectra [134].

API methods, in turn, are soft ionisation methods that mostly leave the parent ions intact and thus allow for the determination of the molecular mass of the analyte [131]. Ion sources that operate under atmospheric pressure include amongst others atmospheric pressure chemical ionisation (APCI) and electrospray ionisation (ESI). ESI, whose invention by Fenn and colleagues has been honoured with a Nobel Prize in



chemistry [135], is the standard ionisation method used in LC-MS [136]. Its functional principle is laid out in Figure 1.9.



**Figure 1.9.:** Schematic of the electrospray ionisation process in positive mode. The analyte solution is transported through a capillary with an electric potential, usually at rates in the  $\mu\text{l}/\text{min}$  range. Due to the electric potential, a so-called Taylor cone forms at the outlet of the capillary and emits a spray of droplets (positively charged in ESI-(+) mode). Once emitted, the droplets decrease in size as solvent evaporates, leading to increasing charge density. As soon as Coulombic repulsion exceeds surface tension, so-called Coulombic explosions generate smaller daughter droplets. This process recurs until analyte ions can transition to the gas phase from the nanodroplets. The ions are attracted by the oppositely charged orifice and fly towards the mass analyser through the orifice's aperture. In negative mode, the potential difference is reversed: Negatively charged droplets are emitted from the Taylor cone and anions pass through a positively charged orifice. Adapted from [137].

Chemically, ionisation happens predominantly by addition or removal of protons or by adduct formation with charged species from the matrix, e. g. sodium or potassium cations in positive mode and chloride or small organic anions in negative mode. Analytes can be multiply charged and combinations of (de-)protonation and adduct formation are possible. Despite the soft ionisation, various fragmentation reactions, often called in-source fragmentations, can take place during ESI depending on the chemical structure of the analytes [138]. Both adduct formation and in-source fragmentation provide valuable information for the annotation of unknown analytes.

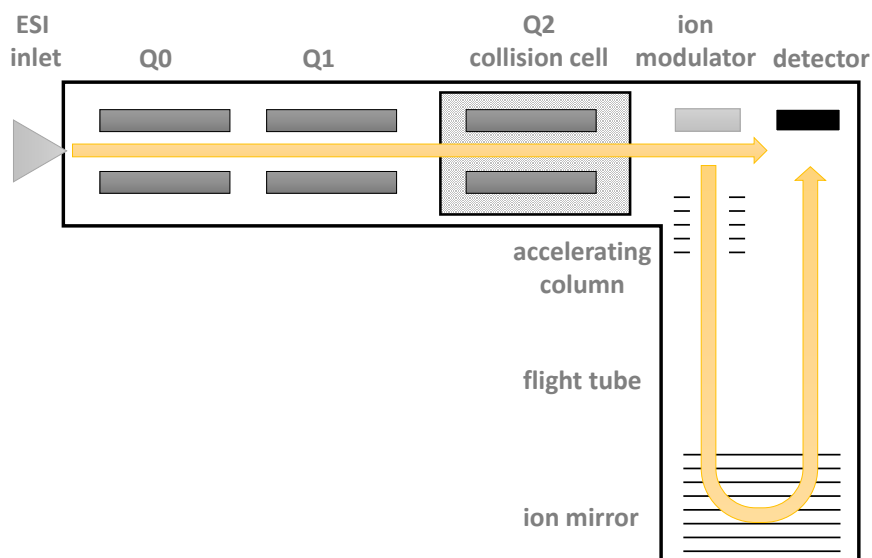
**Mass analysers** Once the analytes have been turned into gaseous ions, the mass analyser determines their mass-to-charge ratio ( $m/z$ ). Several basic principles for mass analysis exist and most mass analysers used in metabolomics rely on the ideas of quadrupole, ion trap, time-of-flight (TOF), ion cyclotron resonance (ICR) or Orbitrap technology [138]. Triple quadrupole (QqQ), quadrupole ion trap and quadrupole time-

of-flight (Q-TOF or QqTOF) mass analysers are popular variations or combinations of the aforementioned types that allow for tandem mass spectrometry. The main criteria for the choice of mass analyser in metabolomics experiments are resolution and sensitivity. Triple quadrupole and quadrupole ion trap instruments are able to detect very low amounts of analytes very specifically, which renders these technologies ideal for targeted metabolomics. Untargeted metabolomics, on the other hand, depends on metabolite identification and hence benefits from the powerful resolution and the resulting high or even ultra-high ( $< 5$  ppm or  $< 1$  ppm difference of measured and theoretical  $m/z$ , respectively) mass accuracy of Q-TOF, Orbitrap or Fourier transform ICR instruments [138, 139].

As EI *per se* results in characteristic fragmentation patterns, tandem mass spectrometry is usually not needed in GC-EI-MS. Respective instruments mostly use low resolution mass analysers such as quadrupoles or ion traps, both of which separate ions by an oscillating electric field [140].

In the liquid chromatography field, LC-QTOF-MS has demonstrated its qualities as a reliable and versatile technology for untargeted metabolomics and metabolite identification by widespread use in the community [141]. The functional principle of time-of-flight mass spectrometry is the acceleration of analyte ions by an electric field whose strength is known. As all ions of the same charge experience the same kinetic energy, the velocity of each ion depends on its mass—or its mass-to-charge ratio if ions of different charge are to be compared. The velocity in the electric field is measured indirectly by recording the time of flight to the detector in a flight tube, and high accuracy mass-to-charge ratios can be calculated from the results. Q-TOF instruments are hybrids of triple quadrupole and TOF instruments, where basically the third quadrupole of a QqQ is replaced by a TOF element. They offer the advantage of high resolution along with extended capacities for tandem mass spectrometry inherent to QqQ technology [142]. Design and function of a Q-TOF mass analyser are depicted in Figure 1.10.

**Tandem mass spectrometry and fragmentation** Tandem mass spectrometry ( $MS^2$ , MS/MS) is the coupling of two mass analysers or mass analysis steps with an additional reaction step between them [143]. In metabolomics,  $MS^2$  is used to generate fragmentation spectra of metabolites by a reaction called collision-induced dissociation (CID). Accelerated analyte ions are directed into a collision cell filled with an inert buffer gas, usually argon, and the release of kinetic energy upon collision leads to fragmentations of the ions similar to those observed in electron ionisation, but *via* even-electron species



**Figure 1.10.:** Design and function of an orthogonal reflectron quadrupole time-of-flight mass analyser. The yellow arrows depict the ions' trajectories from the ESI inlet to the detector. Starting at atmospheric pressure in the ESI unit, they fly through the first quadrupole (Q0) which is operated at low pressure and radiofrequency (r.f.) voltage. Q0 serves to focus the ion beam and for collisional damping. Q1 is held under vacuum and can be run in r.f. mode to let ions pass or as mass filter, which is relevant for MS<sup>2</sup> measurements. In Q2, the pressure can be several milliTorrs as in Q0 and it is also operated in r.f. mode to serve as a transmission element. It can be filled with a collision gas, e. g. argon, and thus serve as a collision cell for MS<sup>2</sup> fragmentation experiments. After Q2, ions are orthogonally diverted and accelerated into the flight tube by the ion modulator and the accelerating column. The flight tube replaces the third quadrupole (Q3) of a triple quadrupole mass analyser. Modern instruments use a reflectron or ion mirror that directs the ions back through the flight tube where they reach the detector. The time of flight is recorded and used for  $m/z$  calculation. Adapted from [142].

rather than radical ions [144]. In a Q-TOF mass spectrometer, CID happens in Q2 (cf. Figure 1.10) and the collision energy that influences the degree of fragmentation can be controlled. By using Q1 as a mass filter, ions of a distinct  $m/z$  can be selectively fragmented. Many MS instruments have an auto-MS/MS mode that selects the  $n$  ions with the highest intensity in each scan for fragmentation, a process called data-dependent acquisition (DDA). Data-independent acquisition (DIA) or all ion fragmentation (AIF) does not use a mass filter and results in various fragment ions that cannot be directly associated with an individual precursor  $m/z$  [145]. DDA—and DIA after additional processing—results in tandem mass spectra that can be matched against MS<sup>2</sup> libraries or used to deduce structural information from the fragmentation patterns [138].

#### 1.2.4. (Pre-)Processing and data analysis

##### Pre-processing

Mass spectrometers record (MS<sup>1</sup>) data in the two dimensions of  $m/z$  and signal intensity. If chromatography-coupled setups are used, retention time (RT) is an additional dimension, producing a three-dimensional data matrix. Two-dimensional chromatograms, i. e. plots of signal intensity over time, can be generated from this data by summation of the signals of all  $m/z$  traces (total ion chromatogram, TIC) or by filtering the  $m/z$  dimension for a specific value or range (extracted ion chromatogram, EIC).

As the aim of most metabolomics experiments is the comparison of signal differences between groups, experimental conditions etc., the MS data has to be pre-processed into so-called feature tables. A feature represents a signal defined by RT and  $m/z$ , to which a peak intensity or area in the individual samples can be assigned [130]. Several pre-processing software tools have been developed, and XCMS and MS-DIAL are the most widely used ones for DDA and DIA LC-MS, respectively [146, 147]. Popular GC-MS pre-processing tools include AMDIS and MetaboliteDetector [148, 149]. LC-MS and GC-MS data pre-processing consists of four basic steps: filtering, feature detection, alignment and normalisation: Filtering reduces noise and smoothes peak shapes [150]. Feature detection distills signals of defined  $m/z$ , RT boundaries and intensity from the chromatographic data; the most prominent feature detection algorithm is XCMS' centWave which uses continuous wavelet transformation to model peaks [151]. Inter-sample correspondence between features is established in the alignment step which often includes retention time correction to account for RT shifts between chromatographic runs. The ordered bijective interpolated warping (obiwarp) algorithm has become the

standard method for alignment of LC-MS data [152]. Normalisation, finally, is applied to level out inter-sample or inter-batch differences in feature intensities and thus remove systematic bias. The use of multiple internal standards is the normalisation method of choice for standard LC-MS and GC-MS experiments and it can be complemented by scaling approaches based on statistical models [150].

### Data analysis

Descriptive statistics is the easiest but also most important data analysis tool for metabolomics feature tables: mean intensity, standard deviation and—if groups or conditions are compared to each other or a reference—fold changes are routinely calculated for all features and intensity distribution is visualised by means of boxplots. Significance testing with Student's or Welch's *t*-test, Mann-Whitney *U*-test or ANOVA are standard methods from the arsenal of inferential statistics. Like in all omics disciplines, they should include correction for multiple testing in order to reduce false positive significance estimates [153].

In addition to these univariate methods, multivariate analysis is heavily used in the metabolomics community. As the multidimensional features  $\times$  samples matrix is not easily represented in two-dimensional plots for data exploration purposes, principal components analysis (PCA) can be used for dimension reduction. PCA is an unsupervised method (i. e. it does not use class information) that summarises information from multiple dimensions into principal components which best explain the variation between the data points [154]. It is used to gauge trends in the data set, e. g. whether groups of samples are separated from each other.

In contrast to unsupervised methods, supervised methods use the class label of samples to, for instance, generate discriminatory or prediction models. Partial least-squares discriminant analysis is a supervised method with resemblance to PCA which can be used for classification [155]. Classification models can also be build using machine learning algorithms like kernel support vector machines, artificial neural networks or random forests [156]. The latter is particularly suited for untargeted metabolomics data, because it can handle input data that is not on the same scale and the models generated are rather easily interpretable with regard to important metabolites [102, 157].

Another data analysis approach is the application of systems biology tools like metabolic network reconstruction or pathway analysis [90]. These methods enable data interpretation in a holistic biological context, but require extensive prior knowledge about the biological systems.

### 1.2.5. Metabolite identification

Metabolite identification is a key step for functional interpretation of metabolomics data but it also represents one of the major challenges in the field of MS-based metabolomics [158]. At the transition between data processing and metabolite identification is the annotation of adducts, in-source fragments, isotopes and multiply charged ions. These can make up a significant percentage of the detected features in an LC-MS data set, as any given metabolite may produce many different features beside the one of the actual (de-)protonated singly charged parent ion ( $[M+H]^+$  or  $[M-H]^-$ ) [159]. In the XCMS software family, CAMERA is the tool used to annotate these features and group them together [160].

While the feature's exact  $m/z$  and isotopic pattern can reveal its sum formula, definite identification necessitates the matching of at least two orthogonal analytical properties to an authentic standard measured on the same instrument [161, 162]. For GC-MS, that would be retention time or the standardised retention index and fragmentation pattern and for LC-MS at least exact  $m/z$  and retention time, preferably along with an MS<sup>2</sup> spectrum recorded with similar collision energy.

If authentic standards are not locally available, online databases are used to assign putative annotations to features based on exact  $m/z$  and MS<sup>2</sup> or EI fragmentation pattern [134]. METLIN [163], HMDB [164], MassBank [165] and the Golm Metabolome Database [166] are among the most comprehensive metabolite databases but more specialised databases like the *Pseudomonas aeruginosa* Metabolome Database (PAMDB) [167] exist as well. Search algorithms for metabolite databases compare spectral similarities which are continuous numerical values, not true-or-false statements, therefore suggested library matches have to be carefully examined [168, 169]. Moreover, mass spectra differ depending on multiple factors such as instrument setup or matrix composition which can complicate annotation by spectral matching [134].

Features that cannot be annotated with local or online metabolite databases are often referred to as 'unknowns'. While 'known unknowns' are features that are not annotated in the respective experiment but that are produced by compounds known in the literature, 'unknown unknowns' are previously unreported, novel molecules [158]. Various computational tools have been developed to aid the structure elucidation of both types of 'unknowns' [170]. Many of them rely on *in silico* generation of MS<sup>2</sup> (or MS<sup>*n*</sup>) spectra [171, 172], whereas others use substructure annotation [173] or spectral similarity [174, 175]. A more detailed description of computational annotation strategies can be found in the introduction to Chapter 3.

### 1.2.6. Microbial metabolomics

Since MS-based exploration of secondary metabolites has been an integral part of microbial natural product research for decades, it is logical to also use more global metabolomics techniques in this field [119]. Most analytical considerations described above are applicable to microbial metabolomics as well as to mammalian or plant metabolomics; however, some experimental and data analysis steps have to be tailored to the specific properties of microbial samples [176]. Particularly, fast sampling and quenching is of particular importance for bacterial samples and extraction protocols have been adjusted based on the expected metabolome composition [177]. Furthermore, the experimenter must decide if samples are directly collected, e. g. from the environment or the human host, or if cultivation-based methods are used. In the latter case, the type of cultivation—on agar, planktonic—and the specific parameters—growth conditions, medium, incubation time—have to be thoroughly planned as they can have vast effects on the metabolome [31]. The use of standardised conditions is often dictated by the need to integrate metabolomics data with other functional genomic or phenotypic data. Planktonic cultivation of bacteria offers the advantage that not only the cellular metabolome (the ‘endometabolome’) but also the excreted metabolites (the ‘exometabolome’) can be analysed [178].

According to Aldridge and Rhee, microbial metabolomics experiments can be classified into five general themes: enzyme annotation, *in situ* enzymology and pathway analysis, chemical inventorying, chemotaxonomy, and biochemical phenotyping [179]. Both the elucidation of unknown enzyme functions and the mechanistic or functional study of known enzymatic reactions can be achieved with *in vitro* and *in vivo* metabolomics methods [180]. Activity-based metabolomic profiling (ABMP) is used to observe consequences of presence, absence or excess of an enzyme of interest. Its *in vitro* variant compares the analytical profiles of metabolite extracts incubated with recombinantly produced enzyme with untreated extracts to find which metabolite levels are altered by *in vitro* turnover. *In vivo* ABMP, however, makes use of genetic engineering to produce knock-out and/or overexpressing strains and compares the metabolomes of these strains to analyse *in vivo* effects of the enzyme in question, e. g. differential abundance of metabolite in a certain metabolic pathway [181].

Chemical inventorying is where microbial metabolomics touches on classical natural product research. The revelation of microbial chemodiversity is of intrinsic scientific value and untargeted metabolomics is an invaluable addition to the natural product re-

searcher's toolbox [182], especially in conjunction with advanced computational methods [174].

Though MS-based methods have a long tradition in the identification of bacterial species, metabolomics enables chemotaxonomy on the strain level as demonstrated for, i. a., *Mycobacterium tuberculosis* [183], *Bacillus cereus* [184] or multiple *Streptomyces* species [185]. Similarly, phenotypic properties can be linked to metabolomic profiles, e. g. drug susceptibility [186]. In addition to classification, these biochemical phenotypes provide functional information in the form of differential abundances of metabolites [187]. Since phenotype characterisation of pathogens is one of the most important steps to understanding infections, metabolomics combined with microbiological expertise can be a powerful tool for infection research [188].

### 1.2.7. The *P. aeruginosa* metabolome

The metabolome of *P. aeruginosa* is a particularly interesting study object for two main reasons: First, the outstanding clinical importance of the pathogen and second, the richness of its secondary metabolome along with the fact that many of its secondary metabolites are associated with virulence or its regulation. Efforts to explore and catalogue the chemodiversity of pseudomonal secondary metabolites were enabled by advances in bioanalytical chemistry in the 2000s, when AQs [54], phenazines [63], rhamnolipids [189] and other groups of natural products were described in a more comprehensive manner and new congeners were discovered.

Around the same time, the first metabolomics experiments were conducted to study how *P. aeruginosa* adapts its metabolism to different growth conditions, highlighting the pathogen's metabolic versatility [190] and suggesting that the switch between planktonic and biofilm lifestyle includes an adaptation of cellular metabolism [191]. An early exometabolomics study found that the virulence-associated anti- $\sigma$  factor MucA modulates the excretion of osmoprotective metabolites, hence leaving a 'metabolic footprint' of its activity in the spent growth medium [192]. In combination with endometabolomics (or cellular metabolomics), Behrends *et al.* used NMR exometabolomics to systematically study a *P. aeruginosa* mutant library and revealed that not all metabolic phenotypes of knockout mutants can be easily linked to the known gene function of the respective deficient gene [193].

Experimental conditions that mimick the situation at specific infection sites in the human host enabled discoveries that were closer to the *in vivo* situation: In conditions resembling the urinary tract, microbial metabolomics showed that multiple metabolic



pathways, including amino acid utilisation and fatty acid biosynthesis, were altered compared to standard laboratory conditions [194]. The cultivation in artificial sputum medium, in turn, resulted in smaller differences in substrate utilisation but more pronounced variation in excreted metabolites [195].

The study of clinical *P. aeruginosa* isolates, most importantly those from CF lungs, led to the discovery that the bacterium strongly adapts its metabolism to the nutrient availability at the specific infection site over time and that this effect on the *P. aeruginosa* metabolome exceeds that of the clonal lineages, which demonstrates the usefulness of metabolomics as a complementary method to the sequence-based omics technologies [196–198]. Additionally, metabolomics has the potential to differentiate *P. aeruginosa* strains and aid in the discovery of biomarkers for clinical phenotypes. This was demonstrated for volatile metabolites measured by GC-MS by Bean *et al.* [199] and Bardin *et al.* used rapid evaporative ionisation-MS to discern metabolic phenotypes of CF isolates [200].

These examples of prior metabolomics studies on *P. aeruginosa* demonstrate the potential of targeted and untargeted metabolomics in the complex task of unravelling the interdependencies of metabolism and virulence in this important pathogen.

## References

- [1] M. Gomila, A. Peña, M. Mulet, J. Lalucat, and E. García-Valdés. “Phylogenomics and systematics in *Pseudomonas*”. In: *Frontiers in Microbiology* 6 (Mar. 2015). DOI: 10.3389/fmicb.2015.00214.
- [2] W. Migula. *Über ein neues System der Bakterien*. 1895.
- [3] W. Migula. *System der Bakterien: Handbuch der Morphologie, Entwicklungsgeschichte und Systematik der Bakterien. Zweiter Band. Spezielle Systematik der Bakterien*. Fischer, 1900.
- [4] N. J. Palleroni. “The *Pseudomonas* Story”. In: *Environmental Microbiology* 12.6 (June 2010), pp. 1377–1383. DOI: 10.1111/j.1462-2920.2009.02041.x.
- [5] N. Palleroni. *Genus 1. Pseudomonas Migula 1894*. In ‘*Bergey’s manual of systematic bacteriology. Vol 2: 2nd edn.* (Eds DJ Brenner, NR Krieg, JT Staley) pp. 323–379. 2005.
- [6] J. Schröter. “Über einige durch Bakterien gebildete Pigmente”. In: *Beiträge zur Biologie der Pflanzen* 1.2 (1872), pp. 109–126.
- [7] K. Botzenhart and G. Döring. “Ecology and epidemiology of *Pseudomonas aeruginosa*”. In: *Pseudomonas aeruginosa as an Opportunistic Pathogen*. Springer, 1993, pp. 1–18.
- [8] C. Bouchard. “Influence qu’exerce sur la maladie charbonneuse l’inoculation du bacille pyocyanique”. In: *CR Acad Sci* 108 (1889), pp. 713–714.
- [9] E. Fraenkel. “Über die Menschenpathogenität des *Bacillus pyocyaneus*”. In: *Zeitschrift für Hygiene und Infektionskrankheiten* 72.1 (1912), pp. 486–522.

- [10] A. Dalhoff and A. Bayer. “*Pseudomonas aeruginosa* als Erreger opportunistischer Infektionen”. In: *Infection* 15.2 (1987), S60–S63.
- [11] M. Finland. “Experiences with *Pseudomonas aeruginosa* at Boston City Hospital over the last half-century”. In: *Pseudomonas aeruginosa: the organism, diseases it causes, and their treatment*. Bern: Hans Huber (1980), pp. 244–64.
- [12] S. Mahajan-Miklos, L. G. Rahme, and F. M. Ausubel. “Elucidating the molecular mechanisms of bacterial virulence using non-mammalian hosts”. In: *Molecular Microbiology* 37.5 (Sept. 2000), pp. 981–988. DOI: 10.1046/j.1365-2958.2000.02056.x.
- [13] L. M. Weiner, A. K. Webb, B. Limbago, M. A. Dudeck, J. Patel, A. J. Kallen, J. R. Edwards, and D. M. Sievert. “Antimicrobial-Resistant Pathogens Associated With Healthcare-Associated Infections: Summary of Data Reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2011–2014”. In: *Infection Control & Hospital Epidemiology* 37.11 (Aug. 2016), pp. 1288–1301. DOI: 10.1017/ice.2016.174.
- [14] G. P. Bodey, R. Bolivar, V. Fainstein, and L. Jadeja. “Infections Caused by *Pseudomonas aeruginosa*”. In: *Clinical Infectious Diseases* 5.2 (Mar. 1983), pp. 279–313. DOI: 10.1093/clinids/5.2.279.
- [15] J. B. Lyczak, C. L. Cannon, and G. B. Pier. “Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist”. In: *Microbes and infection* 2.9 (2000), pp. 1051–1060.
- [16] S. Fujitani, H.-Y. Sun, V. L. Yu, and J. A. Weingarten. “Pneumonia Due to *Pseudomonas aeruginosa*”. In: *Chest* 139.4 (Apr. 2011), pp. 909–919. DOI: 10.1378/chest.10-0166.
- [17] S. L. Percival, L. Suleman, C. Vuotto, and G. Donelli. “Healthcare-associated infections, medical devices and biofilms: risk, tolerance and control”. In: *Journal of Medical Microbiology* 64.Pt\_4 (Feb. 2015), pp. 323–334. DOI: 10.1099/jmm.0.000032.
- [18] J. Chastre and J.-Y. Fagon. “Ventilator-associated Pneumonia”. In: *American Journal of Respiratory and Critical Care Medicine* 165.7 (Apr. 2002), pp. 867–903. DOI: 10.1164/ajrccm.165.7.2105078.
- [19] S. K. Green, M. N. Schroth, J. J. Cho, S. K. Kominos, and V. B. Vitanza-jack. “Agricultural plants and soil as a reservoir for *Pseudomonas aeruginosa*.” In: *Applied microbiology* 28 (6 Dec. 1974), pp. 987–991. ISSN: 0003-6919.
- [20] Y. K. Chitkara and T. C. Feierabend. “Endogenous and exogenous infection with *Pseudomonas aeruginosa* in a burns unit.” In: *International surgery* 66 (3 1981), pp. 237–240. ISSN: 0020-8868.
- [21] J. B. Lyczak, C. L. Cannon, and G. B. Pier. “Lung Infections Associated with Cystic Fibrosis”. In: *Clinical Microbiology Reviews* 15.2 (Apr. 2002), pp. 194–222. DOI: 10.1128/cmr.15.2.194-222.2002.
- [22] N. Mesaros, P. Nordmann, P. Plésiat, M. Roussel-Delvallez, J. V. Eldere, Y. Glupczynski, Y. V. Laethem, F. Jacobs, P. Lebecque, A. Malfroot, P. Tulkens, and F. V. Bambeke. “*Pseudomonas aeruginosa*: resistance and therapeutic options at the turn of the new millennium”. In: *Clinical Microbiology and Infection* 13.6 (June 2007), pp. 560–578. DOI: 10.1111/j.1469-0691.2007.01681.x.
- [23] E. Drenkard. “Antimicrobial resistance of *Pseudomonas aeruginosa* biofilms”. In: *Microbes and Infection* 5.13 (Nov. 2003), pp. 1213–1219. DOI: 10.1016/j.micinf.2003.08.009.

- [24] N. Høiby, T. Bjarnsholt, M. Givskov, S. Molin, and O. Ciofu. “Antibiotic resistance of bacterial biofilms”. In: *International Journal of Antimicrobial Agents* 35.4 (Apr. 2010), pp. 322–332. DOI: 10.1016/j.ijantimicag.2009.12.011.
- [25] G. Pappas, K. Saploura, and M. E. Falagas. “Current Treatment of Pseudomonal Infections in the Elderly”. In: *Drugs & Aging* 26.5 (May 2009), pp. 363–379. DOI: 10.2165/00002512-200926050-00001.
- [26] S. P. Diggle and M. Whiteley. “Microbe Profile: *Pseudomonas aeruginosa*: opportunistic pathogen and lab rat”. In: *Microbiology* 166.1 (Jan. 2020), pp. 30–33. DOI: 10.1099/mic.0.000860.
- [27] M.-V. Grosso-Becerra, C. Santos-Medellín, A. González-Valdez, J.-L. Méndez, G. Delgado, R. Morales-Espinosa, L. Servín-González, L.-D. Alcaraz, and G. Soberón-Chávez. “*Pseudomonas aeruginosa* clinical and environmental isolates constitute a single population with high phenotypic diversity”. In: *BMC Genomics* 15.1 (2014), p. 318. DOI: 10.1186/1471-2164-15-318.
- [28] M. F. Moradali, S. Ghods, and B. H. A. Rehm. “*Pseudomonas aeruginosa* Lifestyle: A Paradigm for Adaptation, Survival, and Persistence”. In: *Frontiers in Cellular and Infection Microbiology* 7 (Feb. 2017). DOI: 10.3389/fcimb.2017.00039.
- [29] L. den Dooren de Jong. “Bijdrage tot de kennis van het mineralisatieproces”. PhD thesis. TU Delft, Delft University of Technology, 1926.
- [30] M. Stephenson. *Bacterial metabolism*. Vol. 48. 4. LWW, 1939.
- [31] E. Frimmersdorf, S. Horatzek, A. Pelnikevich, L. Wiehlmann, and D. Schomburg. “How *Pseudomonas aeruginosa* adapts to various environments: a metabolomic approach”. In: *Environmental Microbiology* 12.6 (Feb. 2010), pp. 1734–1747. DOI: 10.1111/j.1462-2920.2010.02253.x.
- [32] S. K. Dolan, M. Kohlstedt, S. Trigg, P. V. Ramirez, C. F. Kaminski, C. Wittmann, and M. Welch. “Contextual Flexibility in *Pseudomonas aeruginosa* Central Carbon Metabolism during Growth in Single Carbon Sources”. In: *mBio* 11.2 (Mar. 2020). Ed. by J. Nogales and D. K. Newman. DOI: 10.1128/mbio.02684-19.
- [33] M. Beaume, T. Köhler, T. Fontana, M. Tognon, A. Renzoni, and C. van Delden. “Metabolic pathways of *Pseudomonas aeruginosa* involved in competition with respiratory bacterial pathogens”. In: *Frontiers in Microbiology* 6 (Apr. 2015). DOI: 10.3389/fmicb.2015.00321.
- [34] J. A. Bartell, A. S. Blazier, P. Yen, J. C. Thøgersen, L. Jelsbak, J. B. Goldberg, and J. A. Papin. “Reconstruction of the metabolic network of *Pseudomonas aeruginosa* to interrogate virulence factor synthesis”. In: *Nature Communications* 8.1 (Mar. 2017). DOI: 10.1038/ncomms14631.
- [35] P. Williams, K. Winzer, W. C. Chan, and M. Camara. “Look who’s talking: communication and quorum sensing in the bacterial world”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 362.1483 (July 2007), pp. 1119–1134. DOI: 10.1098/rstb.2007.2039.
- [36] W. C. Fuqua, S. C. Winans, and E. P. Greenberg. “Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators.” In: *Journal of bacteriology* 176 (2 Jan. 1994), pp. 269–275. ISSN: 0021-9193.
- [37] P. N. Jimenez, G. Koch, J. A. Thompson, K. B. Xavier, R. H. Cool, and W. J. Quax. “The Multiple Signaling Systems Regulating Virulence in *Pseudomonas aeruginosa*”. In: *Microbiology and Molecular Biology Reviews* 76.1 (Mar. 2012), pp. 46–65. DOI: 10.1128/mnbr.05007-11.

- [38] J. Lee, J. Wu, Y. Deng, J. Wang, C. Wang, J. Wang, C. Chang, Y. Dong, P. Williams, and L.-H. Zhang. “A cell-cell communication signal integrates quorum sensing and stress response”. In: *Nature Chemical Biology* 9.5 (Mar. 2013), pp. 339–343. DOI: 10.1038/nchembio.1225.
- [39] P. Cornelis. “Putting an end to the *Pseudomonas aeruginosa* IQS controversy”. In: *MicrobiologyOpen* 9.2 (Oct. 2019). DOI: 10.1002/mbo3.962.
- [40] M. Hentzer. “Attenuation of *Pseudomonas aeruginosa* virulence by quorum sensing inhibitors”. In: *The EMBO Journal* 22.15 (Aug. 2003), pp. 3803–3815. DOI: 10.1093/emboj/cdg366.
- [41] W. R. Galloway, J. T. Hodgkinson, S. Bowden, M. Welch, and D. R. Spring. “Applications of small molecule activators and inhibitors of quorum sensing in Gram-negative bacteria”. In: *Trends in Microbiology* 20.9 (Sept. 2012), pp. 449–458. DOI: 10.1016/j.tim.2012.06.003.
- [42] F. Soukarieh, P. Williams, M. J. Stocks, and M. Cámara. “*Pseudomonas aeruginosa* Quorum Sensing Systems as Drug Discovery Targets: Current Position and Future Perspectives”. In: *Journal of Medicinal Chemistry* (July 2018). DOI: 10.1021/acs.jmedchem.8b00540.
- [43] S. Mukherjee, D. A. Moustafa, V. Stergioula, C. D. Smith, J. B. Goldberg, and B. L. Bassler. “The PqsE and RhlR proteins are an autoinducer synthase–receptor pair that control virulence and biofilm development in *Pseudomonas aeruginosa*”. In: *Proceedings of the National Academy of Sciences* (Sept. 2018), p. 201814023. DOI: 10.1073/pnas.1814023115.
- [44] L. Passador, J. Cook, M. Gambello, L. Rust, and B. Iglewski. “Expression of *Pseudomonas aeruginosa* virulence genes requires cell-to-cell communication”. In: *Science* 260.5111 (May 1993), pp. 1127–1130. DOI: 10.1126/science.8493556.
- [45] M. J. Gambello, S. Kaye, and B. H. Iglewski. “LasR of *Pseudomonas aeruginosa* is a transcriptional activator of the alkaline protease gene (*apr*) and an enhancer of exotoxin A expression.” In: *Infection and immunity* 61 (4 Apr. 1993), pp. 1180–1184. ISSN: 0019-9567.
- [46] J. P. Pearson, E. C. Pesci, and B. H. Iglewski. “Roles of *Pseudomonas aeruginosa* *las* and *rhl* quorum-sensing systems in control of elastase and rhamnolipid biosynthesis genes.” In: *Journal of Bacteriology* 179.18 (Sept. 1997), pp. 5756–5767. DOI: 10.1128/jb.179.18.5756-5767.1997.
- [47] C. Fuqua. “The QscR Quorum-Sensing Regulon of *Pseudomonas aeruginosa*: an Orphan Claims Its Identity”. In: *Journal of Bacteriology* 188.9 (Apr. 2006), pp. 3169–3171. DOI: 10.1128/jb.188.9.3169-3171.2006.
- [48] S. Chugani and E. P. Greenberg. “LuxR homolog-independent gene regulation by acyl-homoserine lactones in *Pseudomonas aeruginosa*”. In: *Proceedings of the National Academy of Sciences* 107.23 (May 2010), pp. 10673–10678. DOI: 10.1073/pnas.1005909107.
- [49] S. P. Diggle, P. Cornelis, P. Williams, and M. Cámara. “4-Quinolone signalling in *Pseudomonas aeruginosa*: Old molecules, new perspectives”. In: *International Journal of Medical Microbiology* 296.2-3 (Apr. 2006), pp. 83–91. DOI: 10.1016/j.ijmm.2006.01.038.
- [50] F. Bredenbruch, M. Nimtz, V. Wray, M. Morr, R. Muller, and S. Haussler. “Biosynthetic Pathway of *Pseudomonas aeruginosa* 4-Hydroxy-2-Alkylquinolines”. In: *Journal of Bacteriology* 187.11 (May 2005), pp. 3630–3635. DOI: 10.1128/jb.187.11.3630-3635.2005.

- [51] D. Pistorius, A. Ullrich, S. Lucas, R. W. Hartmann, U. Kazmaier, and R. Müller. “Biosynthesis of 2-Alkyl-4(1*H*)-Quinolones in *Pseudomonas aeruginosa*: Potential for Therapeutic Interference with Pathogenicity”. In: *ChemBioChem* 12.6 (Mar. 2011), pp. 850–853. DOI: 10.1002/cbic.201100014.
- [52] C. E. Dulcey, V. Dekimpe, D.-A. Fauvelle, S. Milot, M.-C. Groleau, N. Doucet, L. G. Rahme, F. Lépine, and E. Déziel. “The End of an Old Hypothesis: The *Pseudomonas* Signaling Molecules 4-Hydroxy-2-Alkylquinolines Derive from Fatty Acids, Not 3-Ketofatty Acids”. In: *Chemistry & Biology* 20.12 (Dec. 2013), pp. 1481–1491. DOI: 10.1016/j.chembiol.2013.09.021.
- [53] F. Lépine, S. Milot, E. Déziel, J. He, and L. G. Rahme. “Electrospray/mass spectrometric identification and analysis of 4-hydroxy-2-alkylquinolines (HAQs) produced by *Pseudomonas aeruginosa*”. In: *Journal of the American Society for Mass Spectrometry* 15.6 (June 2004), pp. 862–869. DOI: 10.1016/j.jasms.2004.02.012.
- [54] C. A. Ortori, J.-F. Dubern, S. R. Chhabra, M. Cámara, K. Hardie, P. Williams, and D. A. Barrett. “Simultaneous quantitative profiling of *N*-acyl-L-homoserine lactone and 2-alkyl-4(1*H*)-quinolone families of quorum-sensing signaling molecules using LC-MS/MS”. In: *Analytical and Bioanalytical Chemistry* 399.2 (Oct. 2010), pp. 839–850. DOI: 10.1007/s00216-010-4341-0.
- [55] S. L. Drees and S. Fetzner. “PqsE of *Pseudomonas aeruginosa* Acts as Pathway-Specific Thioesterase in the Biosynthesis of Alkylquinolone Signaling Molecules”. In: *Chemistry & Biology* 22.5 (May 2015), pp. 611–618. DOI: 10.1016/j.chembiol.2015.04.012.
- [56] M.-C. Groleau, T. de Oliveira Pereira, V. Dekimpe, and E. Déziel. “PqsE Is Essential for RhlR-Dependent Quorum Sensing Regulation in *Pseudomonas aeruginosa*”. In: *mSystems* 5.3 (May 2020). Ed. by E. A. Shank. DOI: 10.1128/msystems.00194-20.
- [57] S. Heeb, M. P. Fletcher, S. R. Chhabra, S. P. Diggle, P. Williams, and M. Cámara. “Quinolones: from antibiotics to autoinducers”. In: *FEMS Microbiology Reviews* 35.2 (Mar. 2011), pp. 247–274. DOI: 10.1111/j.1574-6976.2010.00247.x.
- [58] D. Szamosvári and T. Böttcher. “An Unsaturated Quinolone *N*-Oxide of *Pseudomonas aeruginosa* Modulates Growth and Virulence of *Staphylococcus aureus*”. In: *Angewandte Chemie International Edition* 56.25 (May 2017), pp. 7271–7275. DOI: 10.1002/anie.201702944.
- [59] L. Mashburn-Warren, J. Howe, K. Brandenburg, and M. Whiteley. “Structural Requirements of the *Pseudomonas* Quinolone Signal for Membrane Vesicle Stimulation”. In: *Journal of Bacteriology* 191.10 (Mar. 2009), pp. 3411–3414. DOI: 10.1128/jb.00052-09.
- [60] S. P. Diggle, S. Matthijs, V. J. Wright, M. P. Fletcher, S. R. Chhabra, I. L. Lamont, X. Kong, R. C. Hider, P. Cornelis, M. Cámara, and P. Williams. “The *Pseudomonas aeruginosa* 4-Quinolone Signal Molecules HHQ and PQS Play Multifunctional Roles in Quorum Sensing and Iron Entrapment”. In: *Chemistry & Biology* 14.1 (Jan. 2007), pp. 87–96. DOI: 10.1016/j.chembiol.2006.11.014.
- [61] L.-a. Pirofski and A. Casadevall. “Q&A: What is a pathogen? A question that begs the point”. In: *BMC Biology* 10.1 (2012), p. 6. DOI: 10.1186/1741-7007-10-6.
- [62] T. Kohler, R. Guanella, J. Carlet, and C. van Delden. “Quorum sensing-dependent virulence during *Pseudomonas aeruginosa* colonisation and pneumonia in mechanically ventilated patients”. In: *Thorax* 65.8 (Aug. 2010), pp. 703–710. DOI: 10.1136/thx.2009.133082.

- [63] D. V. Mavrodi, W. Blankenfeldt, and L. S. Thomashow. “Phenazine Compounds in Fluorescent *Pseudomonas* Spp.: Biosynthesis and Regulation”. In: *Annual Review of Phytopathology* 44.1 (Sept. 2006), pp. 417–445. DOI: 10.1146/annurev.phyto.44.013106.145710.
- [64] G. W. Lau, D. J. Hassett, H. Ran, and F. Kong. “The role of pyocyanin in *Pseudomonas aeruginosa* infection”. In: *Trends in Molecular Medicine* 10.12 (Dec. 2004), pp. 599–606. DOI: 10.1016/j.molmed.2004.10.002.
- [65] L. Zulianello, C. Canard, T. Kohler, D. Caille, J.-S. Lacroix, and P. Meda. “Rhamnolipids Are Virulence Factors That Promote Early Infiltration of Primary Human Airway Epithelia by *Pseudomonas aeruginosa*”. In: *Infection and Immunity* 74.6 (May 2006), pp. 3134–3147. DOI: 10.1128/iai.01772-05.
- [66] P. O. Jensen, T. Bjarnsholt, R. Phipps, T. B. Rasmussen, H. Calum, L. Christoffersen, C. Moser, P. Williams, T. Pressler, M. Givskov, and N. Hoiby. “Rapid necrotic killing of polymorphonuclear leukocytes is caused by quorum-sensing-controlled production of rhamnolipid by *Pseudomonas aeruginosa*”. In: *Microbiology* 153.5 (May 2007), pp. 1329–1338. DOI: 10.1099/mic.0.2006/003863-0.
- [67] M. Alhede, T. Bjarnsholt, P. O. Jensen, R. K. Phipps, C. Moser, L. Christophersen, L. D. Christensen, M. van Gennip, M. Parsek, N. Hoiby, T. B. Rasmussen, and M. Givskov. “*Pseudomonas aeruginosa* recognizes and responds aggressively to the presence of polymorphonuclear leukocytes”. In: *Microbiology* 155.11 (July 2009), pp. 3500–3508. DOI: 10.1099/mic.0.031443-0.
- [68] T. Strateva and I. Mitov. “Contribution of an arsenal of virulence factors to pathogenesis of *Pseudomonas aeruginosa* infections”. In: *Annals of Microbiology* 61.4 (May 2011), pp. 717–732. DOI: 10.1007/s13213-011-0273-y.
- [69] B. Enderby, D. Smith, W. Carroll, and W. Lenney. “Hydrogen cyanide as a biomarker for *Pseudomonas aeruginosa* in the breath of children with cystic fibrosis”. In: *Pediatric Pulmonology* 44.2 (Feb. 2009), pp. 142–147. DOI: 10.1002/ppul.20963.
- [70] M. Rybtke, L. D. Hultqvist, M. Givskov, and T. Tolker-Nielsen. “*Pseudomonas aeruginosa* Biofilm Infections: Community Structure, Antimicrobial Tolerance and Immune Response”. In: *Journal of Molecular Biology* 427.23 (Nov. 2015), pp. 3628–3645. DOI: 10.1016/j.jmb.2015.08.016.
- [71] M. Harmsen, L. Yang, S. J. Pamp, and T. Tolker-Nielsen. “An update on *Pseudomonas aeruginosa* biofilm formation, tolerance, and dispersal”. In: *FEMS Immunology & Medical Microbiology* 59.3 (Aug. 2010), pp. 253–268. DOI: 10.1111/j.1574-695x.2010.00690.x.
- [72] P. Costaglioli, C. Barthe, S. Claverol, V. S. Brözel, M. Perrot, M. Crouzet, M. Bonneu, B. Garbay, and S. Vilain. “Evidence for the involvement of the anthranilate degradation pathway in *Pseudomonas aeruginosa* biofilm formation”. In: *MicrobiologyOpen* 1.3 (Sept. 2012), pp. 326–339. DOI: 10.1002/mbo3.33.
- [73] W. Griffiths, T. Koal, Y. Wang, M. Kohl, D. Enot, and H.-P. Deigner. ““Targeted Metabolomics” in der Biomarkerforschung”. In: *Angewandte Chemie* 122.32 (July 2010), pp. 5554–5575. DOI: 10.1002/ange.200905579.
- [74] U. Pinder. *Epiphanie medicorum*. 1506. URL: <https://books.google.de/books?id=D0S5YgEACAAJ>.
- [75] J. K. Nicholson and J. C. Lindon. “Metabonomics”. In: *Nature* 455.7216 (Oct. 2008), pp. 1054–1056. DOI: 10.1038/4551054a.



- [76] F. Knoop. *Der Abbau aromatischer Fettsäuren im Tierkörper*. Kuttruff, 1904. URL: <https://books.google.de/books?id=TvOUHAAACAAJ>.
- [77] D. Shemin and D. Rittenberg. “The Utilization of Glycine for the Synthesis of a Porphyrin”. In: *Journal of Biological Chemistry* 159.2 (1945), pp. 567–568. eprint: <http://www.jbc.org/content/159/2/567.full.pdf+html>. URL: <http://www.jbc.org/content/159/2/567.short>.
- [78] L. Pauling, A. B. Robinson, R. Teranishi, and P. Cary. “Quantitative Analysis of Urine Vapor and Breath by Gas-Liquid Partition Chromatography”. In: *Proceedings of the National Academy of Sciences* 68.10 (Oct. 1971), pp. 2374–2376. DOI: 10.1073/pnas.68.10.2374.
- [79] S. G. Oliver, M. K. Winson, D. B. Kell, and F. Baganz. “Systematic functional analysis of the yeast genome”. In: *Trends in Biotechnology* 16.9 (Sept. 1998), pp. 373–378. DOI: 10.1016/s0167-7799(98)01214-1.
- [80] J. K. Nicholson, J. C. Lindon, and E. Holmes. “‘Metabonomics’: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data”. In: *Xenobiotica* 29.11 (Jan. 1999), pp. 1181–1189. DOI: 10.1080/004982599238047.
- [81] D. C. Sévin, A. Kuehne, N. Zamboni, and U. Sauer. “Biological insights through non-targeted metabolomics”. In: *Current Opinion in Biotechnology* 34 (Aug. 2015), pp. 1–8. DOI: 10.1016/j.copbio.2014.10.001.
- [82] F. J. Bruggeman and H. V. Westerhoff. “The nature of systems biology”. In: *Trends in Microbiology* 15.1 (Jan. 2007), pp. 45–50. DOI: 10.1016/j.tim.2006.11.003.
- [83] T. Ideker, T. Galitski, and L. Hood. “A New Approach to Decoding Life: Systems Biology”. In: *Annual Review of Genomics and Human Genetics* 2.1 (Sept. 2001), pp. 343–372. DOI: 10.1146/annurev.genom.2.1.343.
- [84] S. P. Yadav. “The wholeness in suffix -omics, -omes, and the word om.” In: *Journal of biomolecular techniques : JBT* 18 (5 Dec. 2007), p. 277. ISSN: 1524-0215. ppublish.
- [85] F. S. Collins, E. D. Green, A. E. Guttmacher, and M. S. Guyer. “A vision for the future of genomics research”. In: *Nature* 422.6934 (Apr. 2003), pp. 835–847. DOI: 10.1038/nature01626.
- [86] Z. Wang, M. Gerstein, and M. Snyder. “RNA-Seq: a revolutionary tool for transcriptomics”. In: *Nature Reviews Genetics* 10.1 (Jan. 2009), pp. 57–63. DOI: 10.1038/nrg2484.
- [87] R. Aebersold and M. Mann. “Mass spectrometry-based proteomics”. In: *Nature* 422.6928 (Mar. 2003), pp. 198–207. DOI: 10.1038/nature01511.
- [88] A. Shevchenko and K. Simons. “Lipidomics: coming to grips with lipid diversity”. In: *Nature Reviews Molecular Cell Biology* 11.8 (July 2010), pp. 593–598. DOI: 10.1038/nrm2934.
- [89] H. Haraguchi. “Metallomics as integrated biometal science”. In: *Journal of Analytical Atomic Spectrometry* 19.1 (2004), p. 5. DOI: 10.1039/b308213j.
- [90] A. Rosato, L. Tenori, M. Cascante, P. R. D. A. Carulla, V. A. P. M. dos Santos, and E. Saccenti. “From correlation to causation: analysis of metabolomics data using systems biology approaches”. In: *Metabolomics* 14.4 (Feb. 2018). DOI: 10.1007/s11306-018-1335-y.
- [91] G. J. Patti, O. Yanes, and G. Siuzdak. “Metabolomics: the apogee of the omics trilogy”. In: *Nature Reviews Molecular Cell Biology* 13.4 (Apr. 2012), pp. 263–269. DOI: 10.1038/nrm3314.

- [92] S. B. Milne, T. P. Mathews, D. S. Myers, P. T. Ivanova, and H. A. Brown. “Sum of the Parts: Mass Spectrometry-Based Metabolomics”. In: *Biochemistry* 52.22 (Mar. 2013), pp. 3829–3840. DOI: 10.1021/bi400060e.
- [93] C. M. Grim, G. T. Luu, and L. M. Sanchez. “Staring into the void: demystifying microbial metabolomics”. In: *FEMS Microbiology Letters* 366.11 (June 2019). DOI: 10.1093/femsle/fnz135.
- [94] S. Aros-Calt, B. H. Muller, S. Boudah, C. Ducruix, G. Gervasi, C. Junot, and F. Fenaille. “Annotation of the *Staphylococcus aureus* Metabolome Using Liquid Chromatography Coupled to High-Resolution Mass Spectrometry and Application to the Study of Methicillin Resistance”. In: *Journal of Proteome Research* 14.11 (Oct. 2015), pp. 4863–4875. DOI: 10.1021/acs.jproteome.5b00697.
- [95] Y.-J. Xu, C. Wang, W. E. Ho, and C. N. Ong. “Recent developments and applications of metabolomics in microbiological investigations”. In: *TrAC Trends in Analytical Chemistry* 56 (Apr. 2014), pp. 37–48. DOI: 10.1016/j.trac.2013.12.009.
- [96] S. Moco, J. Vervoort, S. Moco, R. J. Bino, R. C. D. Vos, and R. Bino. “Metabolomics technologies and metabolite identification”. In: *TrAC Trends in Analytical Chemistry* 26.9 (Oct. 2007), pp. 855–866. DOI: 10.1016/j.trac.2007.08.003.
- [97] O. Begou, H. G. Gika, I. D. Wilson, and G. Theodoridis. “Hyphenated MS-based targeted approaches in metabolomics”. In: *The Analyst* 142.17 (2017), pp. 3079–3100. DOI: 10.1039/c7an00812k.
- [98] M. Baker. “Metabolomics: from small molecules to big ideas”. In: *Nature Methods* 8.2 (Feb. 2011), pp. 117–121. DOI: 10.1038/nmeth0211-117.
- [99] M. R. Viant, I. J. Kurland, M. R. Jones, and W. B. Dunn. “How close are we to complete annotation of metabolomes?” In: *Current Opinion in Chemical Biology* 36 (Feb. 2017), pp. 64–69. DOI: 10.1016/j.cbpa.2017.01.001.
- [100] W. B. Dunn, A. Erban, R. J. M. Weber, D. J. Creek, M. Brown, R. Breitling, T. Hankemeier, R. Goodacre, S. Neumann, J. Kopka, and M. R. Viant. “Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics”. In: *Metabolomics* 9.S1 (May 2012), pp. 44–66. DOI: 10.1007/s11306-012-0434-4.
- [101] Z. Lai, H. Tsugawa, G. Wohlgemuth, S. Mehta, M. Mueller, Y. Zheng, A. Ogiwara, J. Meissen, M. Showalter, K. Takeuchi, T. Kind, P. Beal, M. Arita, and O. Fiehn. “Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics”. In: *Nature Methods* 15.1 (Nov. 2017), pp. 53–56. DOI: 10.1038/nmeth.4512.
- [102] L. Yi, N. Dong, Y. Yun, B. Deng, D. Ren, S. Liu, and Y. Liang. “Chemometric methods in data processing of mass spectrometry-based metabolomics: A review”. In: *Analytica Chimica Acta* 914 (Mar. 2016), pp. 17–34. DOI: 10.1016/j.aca.2016.02.001.
- [103] I. Kohler, A. Verhoeven, R. J. Derks, and M. Giera. “Analytical pitfalls and challenges in clinical metabolomics”. In: *Bioanalysis* 8.14 (July 2016), pp. 1509–1532. DOI: 10.4155/bio-2016-0090.
- [104] A. Amberg, B. Riefke, G. Schlotterbeck, A. Ross, H. Senn, F. Dieterle, and M. Keck. “NMR and MS Methods for Metabolomics”. In: *Methods in Molecular Biology*. Springer New York, 2017, pp. 229–258. DOI: 10.1007/978-1-4939-7172-5\_13.
- [105] H. Link, T. Fuhrer, L. Gerosa, N. Zamboni, and U. Sauer. “Real-time metabolome profiling of the metabolic switch between starvation and growth”. In: *Nature Methods* 12.11 (Sept. 2015), pp. 1091–1097. DOI: 10.1038/nmeth.3584.



- [106] C. M. Rath, J. Y. Yang, T. Alexandrov, and P. C. Dorrestein. “Data-Independent Microbial Metabolomics with Ambient Ionization Mass Spectrometry”. In: *Journal of The American Society for Mass Spectrometry* 24.8 (Apr. 2013), pp. 1167–1176. DOI: 10.1007/s13361-013-0608-y.
- [107] R.-J. Raterink, P. W. Lindenburg, R. J. Vreeken, R. Ramautar, and T. Hankemeier. “Recent developments in sample-pretreatment techniques for mass spectrometry-based metabolomics”. In: *TrAC Trends in Analytical Chemistry* 61 (Oct. 2014), pp. 157–167. DOI: 10.1016/j.trac.2014.06.003.
- [108] B. Habchi, S. Alves, A. Paris, D. N. Rutledge, and E. Rathahao-Paris. “How to really perform high throughput metabolomic analyses efficiently?”. In: *TrAC Trends in Analytical Chemistry* 85 (Dec. 2016), pp. 128–139. DOI: 10.1016/j.trac.2016.09.005.
- [109] W. de Koning and K. van Dam. “A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH”. In: *Analytical Biochemistry* 204.1 (July 1992), pp. 118–123. DOI: 10.1016/0003-2697(92)90149-2.
- [110] F. Pinu, S. Villas-Boas, and R. Aggio. “Analysis of Intracellular Metabolites from Microorganisms: Quenching and Extraction Protocols”. In: *Metabolites* 7.4 (Oct. 2017), p. 53. DOI: 10.3390/metabo7040053.
- [111] M. R. Mashego, K. Rumbold, M. D. Mey, E. Vandamme, W. Soetaert, and J. J. Heijnen. “Microbial metabolomics: past, present and future methodologies”. In: *Biotechnology Letters* 29.1 (Nov. 2006), pp. 1–16. DOI: 10.1007/s10529-006-9218-0.
- [112] J. Börner, S. Buchinger, and D. Schomburg. “A high-throughput method for microbial metabolome analysis using gas chromatography/mass spectrometry”. In: *Analytical Biochemistry* 367.2 (Aug. 2007), pp. 143–151. DOI: 10.1016/j.ab.2007.04.036.
- [113] C. J. Bolten, P. Kiefer, F. Letisse, J.-C. Portais, and C. Wittmann. “Sampling for Metabolome Analysis of Microorganisms”. In: *Analytical Chemistry* 79.10 (May 2007), pp. 3843–3849. DOI: 10.1021/ac0623888.
- [114] O. Yanes, R. Tautenhahn, G. J. Patti, and G. Siuzdak. “Expanding Coverage of the Metabolome for Global Metabolite Profiling”. In: *Analytical Chemistry* 83.6 (Mar. 2011), pp. 2152–2161. DOI: 10.1021/ac102981k.
- [115] D. G. Sitnikov, C. S. Monnin, and D. Vuckovic. “Systematic Assessment of Seven Solvent and Solid-Phase Extraction Methods for Metabolomics Analysis of Human Plasma by LC-MS”. In: *Scientific Reports* 6.1 (Dec. 2016). DOI: 10.1038/srep38885.
- [116] J. A. Kirwan, L. Brennan, D. Broadhurst, O. Fiehn, M. Cascante, W. B. Dunn, M. A. Schmidt, and V. Velagapudi. “Preanalytical Processing and Biobanking Procedures of Biological Samples for Metabolomics Research: A White Paper, Community Perspective (for “Precision Medicine and Pharmacometabolomics Task Group”—The Metabolomics Society Initiative)”. In: *Clinical Chemistry* 64.8 (Aug. 2018), pp. 1158–1182. DOI: 10.1373/clinchem.2018.287045.
- [117] N. M. Stasulli and E. A. Shank. “Profiling the metabolic signals involved in chemical communication between microbes using imaging mass spectrometry”. In: *FEMS Microbiology Reviews* 40.6 (Sept. 2016). Ed. by K. Gibbs, pp. 807–813. DOI: 10.1093/femsre/fuw032.
- [118] T. O. Metz, J. S. Page, E. S. Baker, K. Tang, J. Ding, Y. Shen, and R. D. Smith. “High-resolution separations and improved ion production and transmission in metabolomics”. In: *TrAC Trends in Analytical Chemistry* 27.3 (Mar. 2008), pp. 205–214. DOI: 10.1016/j.trac.2007.11.003.

- [119] J.-L. Wolfender, G. Marti, A. Thomas, and S. Bertrand. “Current approaches and challenges for the metabolite profiling of complex natural extracts”. In: *Journal of Chromatography A* 1382 (Feb. 2015), pp. 136–164. DOI: 10.1016/j.chroma.2014.10.091.
- [120] N. L. Kuehnbaum and P. Britz-McKibbin. “New Advances in Separation Science for Metabolomics: Resolving Chemical Diversity in a Post-Genomic Era”. In: *Chemical Reviews* 113.4 (Mar. 2013), pp. 2437–2468. DOI: 10.1021/cr300484s.
- [121] R. López-Ruiz, R. Romero-González, and A. G. Frenich. “Ultrahigh-pressure liquid chromatography-mass spectrometry: An overview of the last decade”. In: *TrAC Trends in Analytical Chemistry* 118 (Sept. 2019), pp. 170–181. DOI: 10.1016/j.trac.2019.05.044.
- [122] G. A. Theodoridis, H. G. Gika, E. J. Want, and I. D. Wilson. “Liquid chromatography-mass spectrometry based global metabolite profiling: A review”. In: *Analytica Chimica Acta* 711 (Jan. 2012), pp. 7–16. DOI: 10.1016/j.aca.2011.09.042.
- [123] L. Cui, H. Lu, and Y. H. Lee. “Challenges and emergent solutions for LC-MS/MS based untargeted metabolomics in diseases”. In: *Mass Spectrometry Reviews* 37.6 (Feb. 2018), pp. 772–792. DOI: 10.1002/mas.21562.
- [124] A. Lindahl, S. Sääf, J. Lehtiö, and A. Nordström. “Tuning Metabolome Coverage in Reversed Phase LC-MS Metabolomics of MeOH Extracted Samples Using the Reconstitution Solvent Composition”. In: *Analytical Chemistry* 89.14 (June 2017), pp. 7356–7364. DOI: 10.1021/acs.analchem.7b00475.
- [125] P. Hemström and K. Irgum. “Hydrophilic interaction chromatography”. In: *Journal of Separation Science* 29.12 (Aug. 2006), pp. 1784–1821. DOI: 10.1002/jssc.200600199.
- [126] O. Fiehn. “Metabolomics by Gas Chromatography-Mass Spectrometry: Combined Targeted and Untargeted Profiling.” In: *Current protocols in molecular biology* 114 (Apr. 2016), pp. 30.4.1–30.4.32. ISSN: 1934-3647. DOI: 10.1002/0471142727.mb3004s114.epublish.
- [127] A. Ruiz-Matute, O. Hernández-Hernández, S. Rodríguez-Sánchez, M. Sanz, and I. Martínez-Castro. “Derivatization of carbohydrates for GC and GC-MS analyses”. In: *Journal of Chromatography B* 879.17-18 (May 2011), pp. 1226–1240. DOI: 10.1016/j.jchromb.2010.11.013.
- [128] A.-H. Emwas, R. Roy, R. T. McKay, L. Tenori, E. Saccenti, G. A. N. Gowda, D. Raftery, F. Alahmari, L. Jaremko, M. Jaremko, and D. S. Wishart. “NMR Spectroscopy for Metabolomics Research”. In: *Metabolites* 9.7 (June 2019), p. 123. DOI: 10.3390/metabo9070123.
- [129] J. L. Markley, R. Brüschweiler, A. S. Edison, H. R. Eghbalnia, R. Powers, D. Raftery, and D. S. Wishart. “The future of NMR-based metabolomics”. In: *Current Opinion in Biotechnology* 43 (Feb. 2017), pp. 34–40. DOI: 10.1016/j.copbio.2016.08.001.
- [130] X. Liu and J. W. Locasale. “Metabolomics: A Primer”. In: *Trends in Biochemical Sciences* 42.4 (Apr. 2017), pp. 274–284. DOI: 10.1016/j.tibs.2017.01.004.
- [131] D.-X. Li, L. Gan, A. Bronja, and O. J. Schmitz. “Gas chromatography coupled to atmospheric pressure ionization mass spectrometry (GC-API-MS): Review”. In: *Analytica Chimica Acta* 891 (Sept. 2015), pp. 43–61. DOI: 10.1016/j.aca.2015.08.002.
- [132] “electron ionization in mass spectrometry”. In: *IUPAC Compendium of Chemical Terminology*. IUPAC. DOI: 10.1351/goldbook.e01999.
- [133] F. McLafferty. *Interpretation of mass spectra*. Mill Valley, Calif: University Science Books, 1993. ISBN: 0935702253.

- [134] M. Vinaixa, E. L. Schymanski, S. Neumann, M. Navarro, R. M. Salek, and O. Yanes. “Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects”. In: *TrAC Trends in Analytical Chemistry* 78 (Apr. 2016), pp. 23–35. DOI: 10.1016/j.trac.2015.09.005.
- [135] J. Fenn, M. Mann, C. Meng, S. Wong, and C. Whitehouse. “Electrospray ionization for mass spectrometry of large biomolecules”. In: *Science* 246.4926 (Oct. 1989), pp. 64–71. DOI: 10.1126/science.2675315.
- [136] M. Wilm. “Principles of Electrospray Ionization”. In: *Molecular & Cellular Proteomics* 10.7 (July 2011), p. M111.009407. DOI: 10.1074/mcp.m111.009407.
- [137] L. Konermann, E. Ahadi, A. D. Rodriguez, and S. Vahidi. “Unraveling the Mechanism of Electrospray Ionization”. In: *Analytical Chemistry* 85.1 (Nov. 2012), pp. 2–9. DOI: 10.1021/ac302789c.
- [138] M. Holčápek, R. Jirásko, and M. Lísa. “Basic rules for the interpretation of atmospheric pressure ionization mass spectra of small molecules”. In: *Journal of Chromatography A* 1217.25 (June 2010), pp. 3908–3921. DOI: 10.1016/j.chroma.2010.02.049.
- [139] G. A. N. Gowda and D. Djukovic. “Overview of Mass Spectrometry-Based Metabolomics: Opportunities and Challenges”. In: *Methods in Molecular Biology*. Springer New York, 2014, pp. 3–12. DOI: 10.1007/978-1-4939-1258-2\_1.
- [140] D. J. Beale, F. R. Pinu, K. A. Kouremenos, M. M. Poojary, V. K. Narayana, B. A. Boughton, K. Kanojia, S. Dayalan, O. A. H. Jones, and D. A. Dias. “Review of recent developments in GC–MS approaches to metabolomics-based research”. In: *Metabolomics* 14.11 (Nov. 2018). DOI: 10.1007/s11306-018-1449-2.
- [141] Z.-J. Zhu, A. W. Schultz, J. Wang, C. H. Johnson, S. M. Yannone, G. J. Patti, and G. Siuzdak. “Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of metabolites guided by the METLIN database”. In: *Nature Protocols* 8.3 (Feb. 2013), pp. 451–460. DOI: 10.1038/nprot.2013.004.
- [142] I. V. Chernushevich, A. V. Loboda, and B. A. Thomson. “An introduction to quadrupole-time-of-flight mass spectrometry”. In: *Journal of Mass Spectrometry* 36.8 (2001), pp. 849–865. DOI: 10.1002/jms.207.
- [143] “tandem mass spectrometer”. In: *IUPAC Compendium of Chemical Terminology*. IUPAC. DOI: 10.1351/goldbook.t06250.
- [144] A. R. Johnson and E. E. Carlson. “Collision-Induced Dissociation Mass Spectrometry: A Powerful Tool for Natural Product Structure Elucidation”. In: *Analytical Chemistry* 87.21 (July 2015), pp. 10668–10678. DOI: 10.1021/acs.analchem.5b01543.
- [145] A. Doerr. “DIA mass spectrometry”. In: *Nature Methods* 12.1 (Dec. 2014), pp. 35–35. DOI: 10.1038/nmeth.3234.
- [146] C. A. Smith, E. J. Want, G. O’Maille, R. Abagyan, and G. Siuzdak. “XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification”. In: *Analytical Chemistry* 78.3 (Feb. 2006), pp. 779–787. DOI: 10.1021/ac051437y.
- [147] H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. VanderGheynst, O. Fiehn, and M. Arita. “MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis”. In: *Nature Methods* 12.6 (May 2015), pp. 523–526. DOI: 10.1038/nmeth.3393.
- [148] S. E. Stein. “An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data”. In: *Journal of the American Society for Mass Spectrometry* 10.8 (1999), pp. 770–781.

- [149] K. Hiller, J. Hangebrauk, C. Jäger, J. Spura, K. Schreiber, and D. Schomburg. “MetaboliteDetector: Comprehensive Analysis Tool for Targeted and Nontargeted GC/MS Based Metabolome Analysis”. In: *Analytical Chemistry* 81.9 (May 2009), pp. 3429–3439. DOI: 10.1021/ac802689c.
- [150] M. Katajamaa and M. Orešič. “Data processing for mass spectrometry-based metabolomics”. In: *Journal of Chromatography A* 1158.1-2 (July 2007), pp. 318–328. DOI: 10.1016/j.chroma.2007.04.021.
- [151] R. Tautenhahn, C. Böttcher, and S. Neumann. “Highly sensitive feature detection for high resolution LC/MS”. In: *BMC Bioinformatics* 9.1 (Nov. 2008). DOI: 10.1186/1471-2105-9-504.
- [152] J. T. Prince and E. M. Marcotte. “Chromatographic Alignment of ESI-LC-MS Proteomics Data Sets by Ordered Bijective Interpolated Warping”. In: *Analytical Chemistry* 78.17 (Sept. 2006), pp. 6140–6152. DOI: 10.1021/ac0605344.
- [153] M. Vinaixa, S. Samino, I. Saez, J. Duran, J. J. Guinovart, and O. Yanes. “A Guideline to Univariate Statistical Analysis for LC/MS-Based Untargeted Metabolomics-Derived Data”. In: *Metabolites* 2.4 (Oct. 2012), pp. 775–795. DOI: 10.3390/metabo2040775.
- [154] J. Bartel, J. Krumsiek, and F. J. Theis. “Statistical Methods for the Analysis of High-Throughput Metabolomics Data”. In: *Computational and Structural Biotechnology Journal* 4.5 (Jan. 2013), e201301009. DOI: 10.5936/csbj.201301009.
- [155] J. Trygg, E. Holmes, and T. Lundstedt. “Chemometrics in Metabonomics”. In: *Journal of Proteome Research* 6.2 (Feb. 2007), pp. 469–479. DOI: 10.1021/pr060594q.
- [156] K. M. Mendez, S. N. Reinke, and D. I. Broadhurst. “A comparative evaluation of the generalised predictive ability of eight machine learning algorithms across ten clinical metabolomics data sets for binary classification”. In: *Metabolomics* 15.12 (Nov. 2019). DOI: 10.1007/s11306-019-1612-4.
- [157] T. Chen, Y. Cao, Y. Zhang, J. Liu, Y. Bao, C. Wang, W. Jia, and A. Zhao. “Random Forest in Clinical Metabolomics for Phenotypic Discrimination and Biomarker Selection”. In: *Evidence-Based Complementary and Alternative Medicine* 2013 (2013), pp. 1–11. DOI: 10.1155/2013/298183.
- [158] D. S. Wishart. “Computational strategies for metabolite identification in metabolomics”. In: *Bioanalysis* 1.9 (Dec. 2009), pp. 1579–1596. DOI: 10.4155/bio.09.138.
- [159] X. Domingo-Almenara, J. R. Montenegro-Burke, H. P. Benton, and G. Siuzdak. “Annotation: A Computational Solution for Streamlining Metabolomics Analysis”. In: *Analytical Chemistry* 90.1 (Nov. 2017), pp. 480–489. DOI: 10.1021/acs.analchem.7b03929.
- [160] C. Kuhl, R. Tautenhahn, C. Böttcher, T. R. Larson, and S. Neumann. “CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets”. In: *Analytical Chemistry* 84.1 (Dec. 2011), pp. 283–289. DOI: 10.1021/ac202450g.
- [161] L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W.-M. Fan, O. Fiehn, R. Goodacre, J. L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A. N. Lane, J. C. Lindon, P. Marriott, A. W. Nicholls, M. D. Reily, J. J. Thaden, and M. R. Viant. “Proposed minimum reporting standards for chemical analysis”. In: *Metabolomics* 3.3 (Sept. 2007), pp. 211–221. DOI: 10.1007/s11306-007-0082-2.
- [162] R. M. Salek, C. Steinbeck, M. R. Viant, R. Goodacre, and W. B. Dunn. “The role of reporting standards for metabolite annotation and identification in metabolomic studies”. In: *GigaScience* 2.1 (Oct. 2013). DOI: 10.1186/2047-217x-2-13.

- [163] C. A. Smith, G. O. Maille, E. J. Want, C. Qin, S. A. Trauger, T. R. Brandon, D. E. Custodio, R. Abagyan, and G. Siuzdak. “METLIN”. In: *Therapeutic Drug Monitoring* 27.6 (Dec. 2005), pp. 747–751. DOI: 10.1097/01.ftd.0000179845.53213.39.
- [164] D. S. Wishart, Y. D. Feunang, A. Marcu, A. C. Guo, K. Liang, R. Vázquez-Fresno, T. Sajed, D. Johnson, C. Li, N. Karu, Z. Sayeeda, E. Lo, N. Assempour, M. Berjanskii, S. Singhal, D. Arndt, Y. Liang, H. Badran, J. Grant, A. Serra-Cayuela, Y. Liu, R. Mandal, V. Neveu, A. Pon, C. Knox, M. Wilson, C. Manach, and A. Scalbert. “HMDB 4.0: the human metabolome database for 2018”. In: *Nucleic Acids Research* 46.D1 (Nov. 2017), pp. D608–D617. DOI: 10.1093/nar/gkx1089.
- [165] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, and T. Nishioka. “MassBank: a public repository for sharing mass spectral data for life sciences”. In: *Journal of Mass Spectrometry* 45.7 (July 2010), pp. 703–714. DOI: 10.1002/jms.1777.
- [166] J. Kopka, N. Schauer, S. Krueger, C. Birkemeyer, B. Usadel, E. Bergmuller, P. Dornmann, W. Weckwerth, Y. Gibon, M. Stitt, L. Willmitzer, A. R. Fernie, and D. Steinhäuser. “GMD@CSB.DB: the Golm Metabolome Database”. In: *Bioinformatics* 21.8 (Dec. 2004), pp. 1635–1638. DOI: 10.1093/bioinformatics/bti236.
- [167] W. Huang, L. K. Brewer, J. W. Jones, A. T. Nguyen, A. Marcu, D. S. Wishart, A. G. Oglesby-Sherrouse, M. A. Kane, and A. Wilks. “PAMDB: a comprehensive *Pseudomonas aeruginosa* metabolome database”. In: *Nucleic Acids Research* 46.D1 (Nov. 2017), pp. D575–D580. DOI: 10.1093/nar/gkx1061.
- [168] K. X. Wan, I. Vidavsky, and M. L. Gross. “Comparing similar spectra: From similarity index to spectral contrast angle”. In: *Journal of the American Society for Mass Spectrometry* 13.1 (Jan. 2002), pp. 85–88. DOI: 10.1016/s1044-0305(01)00327-0.
- [169] T. Kind, H. Tsugawa, T. Cajka, Y. Ma, Z. Lai, S. S. Mehta, G. Wohlgemuth, D. K. Barupal, M. R. Showalter, M. Arita, and O. Fiehn. “Identification of small molecules using accurate mass MS/MS search”. In: *Mass Spectrometry Reviews* 37.4 (Apr. 2017), pp. 513–532. DOI: 10.1002/mas.21535.
- [170] I. Blaženović, T. Kind, J. Ji, and O. Fiehn. “Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics”. In: *Metabolites* 8.2 (May 2018), p. 31. DOI: 10.3390/metabo8020031.
- [171] K. Dührkop, H. Shen, M. Meusel, J. Rousu, and S. Böcker. “Searching molecular structure databases with tandem mass spectra using CSI:FingerID”. In: *Proceedings of the National Academy of Sciences* 112.41 (Sept. 2015), pp. 12580–12585. DOI: 10.1073/pnas.1509788112.
- [172] C. Ruttkies, E. L. Schymanski, S. Wolf, J. Hollender, and S. Neumann. “MetFrag relaunched: incorporating strategies beyond in silico fragmentation”. In: *Journal of Cheminformatics* 8.1 (Jan. 2016). DOI: 10.1186/s13321-016-0115-9.
- [173] J. J. J. van der Hooft, J. Wandy, M. P. Barrett, K. E. V. Burgess, and S. Rogers. “Topic modeling for untargeted substructure exploration in metabolomics”. In: *Proceedings of the National Academy of Sciences* 113.48 (Nov. 2016), pp. 13738–13743. DOI: 10.1073/pnas.1608041113.



- [174] J. Watrous, P. Roach, T. Alexandrov, B. S. Heath, J. Y. Yang, R. D. Kersten, M. van der Voort, K. Pogliano, H. Gross, J. M. Raaijmakers, B. S. Moore, J. Laskin, N. Bandeira, and P. C. Dorrestein. “Mass spectral molecular networking of living microbial colonies”. In: *Proceedings of the National Academy of Sciences* 109.26 (May 2012), E1743–E1752. DOI: 10.1073/pnas.1203689109.
- [175] H. Treutler, H. Tsugawa, A. Porzel, K. Gorzolka, A. Tissier, S. Neumann, and G. U. Balcke. “Discovering Regulated Metabolite Families in Untargeted Metabolomics Studies”. In: *Analytical Chemistry* 88.16 (Aug. 2016), pp. 8082–8090. DOI: 10.1021/acs.analchem.6b01569.
- [176] D. E. Garcia, E. E. Baidoo, P. I. Benke, F. Pingitore, Y. J. Tang, S. Villa, and J. D. Keasling. “Separation and mass spectrometry in microbial metabolomics”. In: *Current Opinion in Microbiology* 11.3 (June 2008), pp. 233–239. DOI: 10.1016/j.mib.2008.04.002.
- [177] W. M. van Gulik, A. B. Canelas, R. M. Seifar, and J. J. Heijnen. “The Sampling and Sample Preparation Problem in Microbial Metabolomics”. In: *Metabolomics in Practice*. Wiley-VCH Verlag GmbH & Co. KGaA, Feb. 2013, pp. 1–19. DOI: 10.1002/9783527655861.ch1.
- [178] F. Pinu and S. Villas-Boas. “Extracellular Microbial Metabolomics: The State of the Art”. In: *Metabolites* 7.3 (Aug. 2017), p. 43. DOI: 10.3390/metabo7030043.
- [179] B. B. Aldridge and K. Y. Rhee. “Microbial metabolomics: innovation, application, insight”. In: *Current Opinion in Microbiology* 19 (June 2014), pp. 90–96. DOI: 10.1016/j.mib.2014.06.009.
- [180] N. Saito, Y. Ohashi, T. Soga, and M. Tomita. “Unveiling cellular biochemical reactions via metabolomics-driven approaches”. In: *Current Opinion in Microbiology* 13.3 (June 2010), pp. 358–362. DOI: 10.1016/j.mib.2010.04.006.
- [181] G. A. Prosser, G. Larrouy-Maumus, and L. P. S. Carvalho. “Metabolomic strategies for the identification of new enzyme functions and metabolic pathways”. In: *EMBO reports* 15.6 (May 2014), pp. 657–669. DOI: 10.15252/embr.201338283.
- [182] T. Luzzatto-Knaan, A. V. Melnik, and P. C. Dorrestein. “Mass spectrometry tools and workflows for revealing microbial chemistry”. In: *The Analyst* 140.15 (2015), pp. 4949–4966. DOI: 10.1039/c5an00171d.
- [183] E. Layre, L. Sweet, S. Hong, C. A. Madigan, D. Desjardins, D. C. Young, T.-Y. Cheng, J. W. Annand, K. Kim, I. C. Shamputa, M. J. McConnell, C. A. Debono, S. M. Behar, A. J. Minnaard, M. Murray, C. E. Barry, I. Matsunaga, and D. B. Moody. “A Comparative Lipidomics Platform for Chemotaxonomic Analysis of *Mycobacterium tuberculosis*”. In: *Chemistry & Biology* 18.12 (Dec. 2011), pp. 1537–1549. DOI: 10.1016/j.chembiol.2011.10.013.
- [184] J. G. Bundy, T. L. Willey, R. S. Castell, D. J. Ellar, and K. M. Brindle. “Discrimination of pathogenic clinical isolates and laboratory strains of *Bacillus cereus* by NMR-based metabolomic profiling”. In: *FEMS Microbiology Letters* 242.1 (Jan. 2005), pp. 127–136. DOI: 10.1016/j.femsle.2004.10.048.
- [185] M. Y. Lee, H. Y. Kim, S. Lee, J.-G. Kim, J.-W. Suh, and C. H. Lee. “Metabolomics-Based Chemotaxonomic Classification of *Streptomyces* spp. and Its Correlation with Antibacterial Activity”. In: *Journal of Microbiology and Biotechnology* 25.8 (Aug. 2015), pp. 1265–1274. DOI: 10.4014/jmb.1503.03005.

- [186] M. H. M. Maifiah, S.-E. Cheah, M. D. Johnson, M.-L. Han, J. D. Boyce, V. Thamlikitkul, A. Forrest, K. S. Kaye, P. Hertzog, A. W. Purcell, J. Song, T. Velkov, D. J. Creek, and J. Li. “Global metabolic analyses identify key differences in metabolite levels between polymyxin-susceptible and polymyxin-resistant *Acinetobacter baumannii*”. In: *Scientific Reports* 6.1 (Feb. 2016). DOI: 10.1038/srep22287.
- [187] T. E. Hartman and K. Y. Rhee. “Microbial Metabolomics: Fifty Shades of Metabolism”. In: *ACS Infectious Diseases* 1.1 (Jan. 2015), pp. 73–75. DOI: 10.1021/id500041w.
- [188] S. Kreibich and W.-D. Hardt. “Experimental approaches to phenotypic diversity in infection”. In: *Current Opinion in Microbiology* 27 (Oct. 2015), pp. 25–36. DOI: 10.1016/j.mib.2015.06.007.
- [189] A. M. Abdel-Mawgoud, F. Lépine, and E. Déziel. “Rhamnolipids: diversity of structures, microbial origins and roles”. In: *Applied Microbiology and Biotechnology* 86.5 (Mar. 2010), pp. 1323–1336. DOI: 10.1007/s00253-010-2498-2.
- [190] M. J. van der Werf, K. M. Overkamp, B. Muilwijk, M. M. Koek, B. J. C. van der Werff-van der Vat, R. H. Jellema, L. Coulier, and T. Hankemeier. “Comprehensive analysis of the metabolome of *Pseudomonas putida* S12 grown on different carbon sources”. In: *Molecular BioSystems* 4.4 (2008), p. 315. DOI: 10.1039/b717340g.
- [191] E. L. Gjersing, J. L. Herberg, J. Horn, C. M. Schaldach, and R. S. Maxwell. “NMR Metabolomics of Planktonic and Biofilm Modes of Growth in *Pseudomonas aeruginosa*”. In: *Analytical Chemistry* 79.21 (Nov. 2007), pp. 8037–8045. DOI: 10.1021/ac070800t.
- [192] V. Behrends, B. Ryall, X. Wang, J. G. Bundy, and H. D. Williams. “Metabolic profiling of *Pseudomonas aeruginosa* demonstrates that the anti-sigma factor MucA modulates osmotic stress tolerance”. In: *Molecular BioSystems* 6.3 (2010), p. 562. DOI: 10.1039/b918710c.
- [193] V. Behrends, T. J. Bell, M. Liebeke, A. Cordes-Blauert, S. N. Ashraf, C. Nair, J. E. A. Zlosnik, H. D. Williams, and J. G. Bundy. “Metabolite Profiling to Characterize Disease-related Bacteria”. In: *Journal of Biological Chemistry* 288.21 (Apr. 2013), pp. 15098–15109. DOI: 10.1074/jbc.m112.442814.
- [194] P. Tielen, N. Rosin, A.-K. Meyer, K. Dohnt, I. Haddad, L. Jänsch, J. Klein, M. Narten, C. Pommerenke, M. Scheer, M. Schobert, D. Schomburg, B. Thielen, and D. Jahn. “Regulatory and Metabolic Networks for the Adaptation of *Pseudomonas aeruginosa* Biofilms to Urinary Tract-Like Conditions”. In: *PLoS ONE* 8.8 (Aug. 2013). Ed. by M. Hensel, e71845. DOI: 10.1371/journal.pone.0071845.
- [195] V. Behrends, B. Geier, H. D. Williams, and J. G. Bundy. “Direct Assessment of Metabolite Utilization by *Pseudomonas aeruginosa* during Growth on Artificial Sputum Medium”. In: *Applied and Environmental Microbiology* 79.7 (Jan. 2013), pp. 2467–2470. DOI: 10.1128/aem.03609-12.
- [196] V. Behrends, B. Ryall, J. E. A. Zlosnik, D. P. Speert, J. G. Bundy, and H. D. Williams. “Metabolic adaptations of *Pseudomonas aeruginosa* during cystic fibrosis chronic lung infections”. In: *Environmental Microbiology* 15.2 (Aug. 2012), pp. 398–408. DOI: 10.1111/j.1462-2920.2012.02840.x.
- [197] K. M. Jørgensen, O. Ciofu, S. Molin, T. Wassermann, H. K. Johansen, N. Høiby, and L. E. Christiansen. “Diversity of metabolic profiles of cystic fibrosis *Pseudomonas aeruginosa* during the early stages of lung infection”. In: *Microbiology* 161.7 (July 2015), pp. 1447–1462. DOI: 10.1099/mic.0.000093.

- [198] R. A. Quinn, V. V. Phelan, K. L. Whiteson, N. Garg, B. A. Bailey, Y. W. Lim, D. J. Conrad, P. C. Dorrestein, and F. L. Rohwer. “Microbial, host and xenobiotic diversity in the cystic fibrosis sputum metabolome”. In: *The ISME Journal* 10.6 (Dec. 2015), pp. 1483–1498. DOI: 10.1038/ismej.2015.207.
- [199] H. D. Bean, C. A. Rees, and J. E. Hill. “Comparative analysis of the volatile metabolomes of *Pseudomonas aeruginosa* clinical isolates”. In: *Journal of Breath Research* 10.4 (Nov. 2016), p. 047102. DOI: 10.1088/1752-7155/10/4/047102.
- [200] E. E. Bardin, S. J. S. Cameron, A. Perdones-Montero, K. Hardiman, F. Bolt, E. W. F. W. Alton, A. Bush, J. C. Davies, and Z. Takáts. “Metabolic Phenotyping and Strain Characterisation of *Pseudomonas aeruginosa* Isolates from Cystic Fibrosis Patients Using Rapid Evaporative Ionisation Mass Spectrometry”. In: *Scientific Reports* 8.1 (July 2018). DOI: 10.1038/s41598-018-28665-7.



## 2 | Aims of the Thesis

The principal aim of this Thesis is to explore how targeted and untargeted metabolomics can be used to gain novel insights into virulence-associated processes in *Pseudomonas aeruginosa*.

Since metabolomics is still a young and developing discipline, its usefulness in infection research has not been fully exploited. While there are some examples of interesting and relevant findings based on metabolomics data, a lot is left to discover and there is a plethora of research questions that could potentially be answered by using targeted and/or untargeted metabolomics as main or complimentary methodological approach. In this Thesis, metabolomics technologies were used to address highly relevant questions on the biology and pathogenicity of *P. aeruginosa* that span from the molecular to the cellular and population level. These are illustrated in the following paragraphs:

**Exploring the chemical inventory of *P. aeruginosa* and extending the analytical coverage of the *P. aeruginosa* metabolome (Publication 1)** Metabolite identification and analytical coverage of the metabolome is generally regarded as one of the main challenges in untargeted metabolomics. That is particularly true for prolific producers of secondary metabolites. The generation of useful biological knowledge from untargeted metabolomics data requires as much annotation as possible. At the starting point of this dissertation, the percentage of features in a *P. aeruginosa* cell extract metabolomics data set was very low—both in the CBIO lab and in the literature. This fact necessitated efforts to increase the annotation rate, and the first aim of the Thesis was to enable identification and annotation of more *P. aeruginosa* metabolites by means of chemometric and bioinformatic processing combined with advanced statistical methods. The second aim was to extend the existing knowledge on *P. aeruginosa* secondary metabolites by identifying novel congeners of important virulence-associated metabolites by the aforementioned tools and thus explore and describe the chemodiversity of pseudomonal small molecule virulence factors.

**Turning the methodological advances into a robust, versatile and shareable software and demonstrating its usefulness on different data types (Publication 2)**

Metabolomics, like any systems biology discipline, is a community-driven science and relies on methodologies developed by the community for the community. Thus, it is a scientifically worthwhile endeavour to create novel processing and analysis tools that facilitate the generation of knowledge from data and share them with peers. The third aim of the Thesis is to do so by developing an R/Bioconductor package as a shareable form of the CluMSID software on the basis of Publication 1. This software package is introduced in the main body of Publication 2. Additional aims are to show that CluMSID can help to extract useful information for feature annotation and identification not only from high-resolution LC-MS/MS data but also from data generated by other types of mass spectrometry. These are presented in the extensive Supplementary Information of Publication 2.

**Defining the chemodiversity of alkylquinolones and investigating their biosynthesis by means of targeted metabolomics (Publication 3)**

Alkylquinolones are key regulators of various virulence-associated processes in *P. aeruginosa* and hence an interesting object of study for researchers interested in the interplay of the metabolome and microbial pathogenicity. Building on the extended descriptive knowledge on the chemodiversity of alkylquinolones gained in Publication 1, the Thesis aimed to transfer this knowledge onto a functional level by investigating how this diversity is brought about by the biosynthetic machinery. By integrating targeted metabolomics with structural biology, proteomics and classical biochemical assays, a mechanistic model of alkylquinolone biosynthesis by PqsBC is to be developed. Moreover, the biosynthesis of an alkylquinolone subclass, mono-unsaturated alkylquinolones, should be studied mechanistically for the first time.

**Applying comprehensive untargeted metabolomics to study PrmC, a virulence-associated enzyme of partly unknown function (Publication 4)**

The elucidation of (partly) unknown enzyme functions is a demanding task and metabolomics can serve as one building block in the solution. In the Thesis' subproject described Publication 4, the aim is to examine the influence of PrmC, an enzyme with a known role in protein translation and an underexplored function in the regulation of virulence, on the cellular and extracellular metabolome of *P. aeruginosa* with a specific focus on virulence-associated metabolites.

---

**Discovering metabolic variation between *P. aeruginosa* clinical strains and generating a predictive model to gauge virulence from untargeted metabolomics data (Publication 5)**

While in Publication 4, a defined genetically modified strain is compared to wildtype *P. aeruginosa*, the aim of the subproject described in Publication 5 was to analyse virulence-associated metabolome differences between clinical isolates. These strains have been collected in various hospitals and private practices at different locations in Europe. Besides the description of metabolome differences between strains with different virulence phenotypes, the identification of putative biomarkers for pseudomonal virulence and the development of statistical models to gauge virulence based on untargeted metabolomics data represent major objectives. To summarise, this subproject will transfer existing and generate new knowledge about the interplay of virulence and the metabolome on the level of clinical *P. aeruginosa* strains.



### 3 | Clustering of MS<sup>2</sup> Spectra Using Unsupervised Methods to Aid the Identification of Secondary Metabolites from *Pseudomonas aeruginosa*

This Chapter has been published as peer-reviewed article in a scientific journal:

T. Depke, R. Franke, and M. Brönstrup. "Clustering of MS<sup>2</sup> spectra using unsupervised methods to aid the identification of secondary metabolites from *Pseudomonas aeruginosa*". In: *Journal of Chromatography B* 1071 (Dec. 2017), pp. 19–28. DOI: 10.1016/j.jchromb.2017.06.002

#### Abstract

*Pseudomonas aeruginosa* is an important opportunistic pathogen that produces a large arsenal of small molecule virulence factors and quorum sensing signal molecules. The annotation of these secondary metabolites in untargeted, mass spectrometry-based metabolomics is difficult, as many of them cannot be found in common metabolite databases, and as manual annotation is tedious. We therefore developed an algorithm named CluMSID that uses cosine similarities of product ion spectra and neutral loss patterns in combination with unsupervised clustering methods such as multidimensional scaling, density based clustering and hierarchical clustering to group structurally similar compounds and hence facilitate their annotation. The use of this tool allowed us to find clusters for several classes of primary and secondary metabolites, and helped identifying spectral similarities that would have gone unnoticed in standard untargeted metabolomics data analysis workflows. CluMSID enabled the annotation of 27 previously undescribed members of the canonical classes of alkyl quinolone quorum sensing signal molecules and provided evidence for the postulation of a new putative alkyl quinolone class. The CluMSID script written in R is open source and can be used by anyone in the metabolomics and natural product research community.

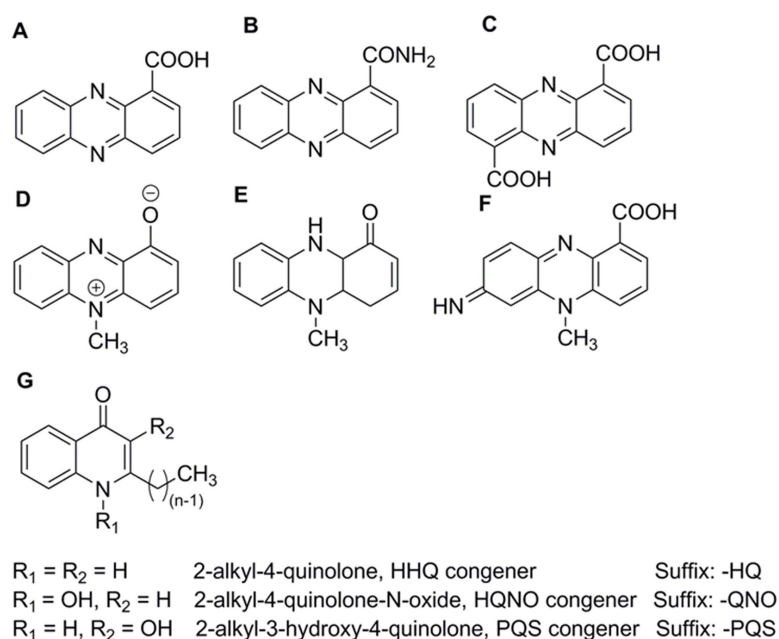
### 3.1. Introduction

The Gram-negative bacterium *Pseudomonas aeruginosa* is an important opportunistic pathogen that can infect various human tissues. It is particularly threatening for patients suffering from cystic fibrosis (CF), pneumonia, or wound infections [1, 2]. For example, *P. aeruginosa* accounts for potentially fatal respiratory failure in 80–95% of CF patients [3]. The intrinsic antibiotic resistance of *P. aeruginosa* [4] and its ability to evade the host immune system by forming biofilms [5, 6] redound further to the high clinical importance of this pathogen.

Like other bacteria, *P. aeruginosa* uses peptides and proteins as virulence factors [7], but it especially relies on small molecule secondary metabolites to exert and regulate its virulence, the most important classes being phenazines, rhamnolipids, homoserine lactones and alkyl quinolones [8]. Phenazines (Fig. 3.1) are redox-active, coloured heterocyclic compounds that are also responsible for the green fluorescence of the bacteria [9]. Pyocyanin, the best studied phenazine of *P. aeruginosa*, can be detected in the sputum of infected CF patients. It is an important contributor to pathogenesis by generating reactive oxygen species that subject host cells to harmful oxidative stress [10]. Rhamnolipids are glycolipids with surfactant-like properties that disturb the integrity of the lung epithelium and hence facilitate paracellular invasion of bacteria [11]. Furthermore, they enable the bacterium to eliminate polymorphonuclear leucocytes and thus evade the most important immune mechanism in CF patients' lungs [12, 13].

*P. aeruginosa* regulates its virulence to a large extent by quorum sensing (QS) [8]. Like most Gram-negative bacteria, it possesses a QS system based on homoserine lactones (HSLs), but it also features an additional system that consists of alkyl quinolones like the *Pseudomonas Quinolone Signal* (PQS) or its biosynthetic precursor 2-heptyl-[1*H*]-4-quinolone (HHQ). Although these two species and 2-heptyl-[1*H*]-4-quinolone-*N*-oxide (HQNO) are usually referred to as the most prominent and most important quorum sensing signal molecules, *P. aeruginosa* produces a large variety of alkyl quinolone derivatives (Fig. 3.1G) [14].

The fact that a large portion of both the regulators and the effectors of pseudomonal virulence are small molecules renders mass spectrometry-based metabolomics a suitable method to study the complex interplay of virulence and metabolism in this pathogen. Microbial metabolomics is a field of growing importance in basic microbiology, biotechnology, synthetic biology as well as medical microbiology and infection research [15]. Metabolomics, the analysis of the (near-)entirety of small molecules in a biological system, is an integral part of systems biology and from all omics disciplines, it is as-



**Figure 3.1.:** Chemical structures of selected secondary metabolites produced by *Pseudomonas aeruginosa*. **A-F:** Selected phenazines produced by *Pseudomonas aeruginosa*. **A:** Phenazine-1-carboxylic acid. **B:** Phenazine-1-carboxamide. **C:** Phenazine-1,6-dicarboxylic acid. **D:** Pyocyanin. **E:** Tetrahydropyocyanin (putative structure) **F:** Aeruginosin A. **G:** General structure of the three alkyl quinolone classes and the nomenclature used in this article. The side chain composition is indicated by the prefix  $C_n$  or, in the case of  $m$ -fold unsaturated side chains,  $C_n:m$ .

sumed to be the one closest to the phenotype [16]. Therefore, it can help elucidating metabolism as such, enzyme functions, or the interplay of metabolism and virulence or other phenotypic traits [17–20].

The most eminent problem encountered by researchers in MS-based metabolomics is metabolite identification [21]. Multiple tools have been developed to help the analyst deal with this issue [22–27] including data base querying [28], computational prediction of mass spectra [29, 30], substructure determination [31] and the analysis of fragmentation trees [32]. As all of them solve only parts of the problem, metabolite identification remains the major obstacle on the way to biological interpretations of metabolomics data. Whereas the metabolomes of some microbes such as *Staphylococcus aureus* have been extensively (albeit still incompletely) annotated [33], the *P. aeruginosa* metabolome measured by reversed phase LC–MS/MS contains in large part compounds that cannot be found in the established metabolite data bases like METLIN [34], MassBank [35] or HMDB/ECMDB [36, 37].

Similarity comparison of MS<sup>2</sup> spectra has long been used in the field of proteomics to group spectra derived from the same molecule [38]. Recent adaptations of this principle in combination with networking algorithms have been applied to dereplication problems in natural product research, metabolomics and molecular ecology [39–42]. The basic idea is that compounds which produce similar MS<sup>2</sup> spectra are likely to have a similar structure or at least share structural elements. The similarity of MS<sup>2</sup> spectra can be calculated in different ways [43], the most widely used being the cosine similarity between  $n$ -dimensional vectors constructed from two spectra containing peaks with  $n$  unique mass-to-charge ratios [44]. Treutler *et al.* have recently combined clustering of MS<sup>2</sup> spectra from data-independent measurements with information on feature regulation gained on the MS1 level in their web application “MetFamily” in order to study differences in abundance of metabolite families [45]. Another web application that uses MS<sup>2</sup> similarity is the “Molecular Networking” tool developed by the Doerrestein group [42]. This approach uses spectral correlation and visualisation to group spectra and allows for data exchange with other researchers but the possibilities for customisation of the data analysis methods are limited.

In this article, we present an algorithm called CluMSID, short for *Clustering of MS<sup>2</sup> Spectra for Metabolite Identification*. It applies cosine similarity in combination with clustering by unsupervised methods to group features by their MS<sup>2</sup> fragmentation in order to facilitate metabolite identification. The algorithm is encoded in an open source, user friendly program written in the programming language R and handles



LC–MS/MS data from data-dependent measurements which makes it usable for most experimentators and data analysts in the field of metabolomics. In contrast to the aforementioned web applications, CluMSID is a highly customisable and accessible tool that is dedicated to assist in the metabolite identification process. We explored the strengths and the limitations of CluMSID by clustering and assigning metabolite spectra from a *P. aeruginosa* cell extract, and achieved extended metabolome coverage through the identification of hitherto novel analogues of virulence regulators.

## 3.2. Materials and methods

### 3.2.1. Chemicals and analytical standards

LC eluents and extraction solvents were Baker Analyzed™ Ultra LC/MS grade (Fisher Scientific, Schwerte, Germany). D-(+)-Glucose monohydrate was from Merck Millipore (Darmstadt, Germany). LB broth powder, magnesium sulfate and casamino acids were purchased from Roth (Karlsruhe, Germany). All other chemicals were obtained from Sigma-Aldrich (Taufkirchen, Germany).

The MSMLS—Mass Spectrometry Metabolite Library of Standards (IROA Technologies, Bolton, MA) as well as a number of individually bought substances from Sigma-Aldrich (Taufkirchen, Germany) were used as analytical standards for our in-house library.

### 3.2.2. Bacterial strains and growth conditions

*Pseudomonas aeruginosa* PA14 from plate culture was used to inoculate four precultures of 4 mL volume each in Luria-Bertani broth [46]. The precultures were incubated overnight at 37 °C in 15 mL glass tubes under constant agitation (160 rotations per minute) in a shaking incubator, and subsequently united and centrifuged at 4 °C and 9000 *g* for 5 min. The supernatant was discarded, and the pellet was resuspended in 10 mL BM2 medium (2 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 40 mM K<sub>2</sub>HPO<sub>4</sub>, 22 mM KH<sub>2</sub>PO<sub>4</sub>, 2 mM MgSO<sub>4</sub>, 10 μM FeSO<sub>4</sub>, 0.4% (w/v) glucose, 0.01% (w/v) casamino acids) by vigorous shaking. Three 50 mL cultures in BM2 medium were inoculated with 290 μL of this suspension, resulting in a starting OD<sub>600</sub> of 0.05. The cultures were incubated at 37 °C and 160 rotations per minute in a shake incubator for 7.5 h until they reached an OD<sub>600</sub> of 1.3 ± 0.1. 10 mL from each culture flask were transferred to 50 mL plastic tubes and centrifuged at 4 °C and 9000 *g* for 5 min. The supernatants were discarded, while the pellet was shock frozen in liquid nitrogen and stored at –80 °C until further processing.

### 3.2.3. Metabolite extraction

The thawed cell pellets were resuspended in 1 mL 80% (v/v) methanol containing 0.1 mg/L trimethoprim, 0.1 mg/L nortriptyline and 0.3 mg/L glipizide as internal standards. Complete resuspension and cell disruption was achieved by vigorous shaking for 1 min, followed by sonication in an ice cold sonic bath (Sonorex Digiplus, BANDELIN electronic, Berlin, Germany) for 15 min. Cell debris was separated by centrifugation at 4 °C and 11000 *g* for 20 min. 900 µL of the supernatant were transferred to a 2 mL plastic tube and dried in a centrifugal evaporator (Refrigerated CentriVap® Concentrator with -50 °C CentriVap® Cold Trap, Labconco, Kansas City, MO) at 20 °C and full vacuum until complete dryness. Samples were reconstituted in 90 µL 80% (v/v) methanol containing 1 mg/L caffeine and 1 mg/L naproxen as internal standards which were added as part of a routine untargeted metabolomics workflow but were not used in the analysis presented here. Equal volumes of each of the three samples were combined to yield a pooled sample, which was used for further analysis.

### 3.2.4. Liquid chromatography—tandem mass spectrometry

The pooled metabolite extract was separated by ultra-high performance liquid chromatography on a Dionex Ultimate 3000 UPLC (Thermo Fisher Scientific, Waltham, MA) using a 150 mm Kinetex C18 reversed phase column with 1.7 µm particle size and 2.1 mm inner diameter (Phenomenex, Aschaffenburg, Germany) with a flow rate of 300 µL/min. Gradient elution with water with 0.1% (v/v) formic acid as eluent A and acetonitrile with 0.1% (v/v) formic acid as eluent B was run as follows: 1% B for  $t = 0$  min to  $t = 2$  min, linear gradient from 1% B to 100% B from  $t = 2$  min to  $t = 20$  min, hold 100% B until  $t = 25$  min and linear gradient from 100% B to 1% B from  $t = 25$  min to  $t = 30$  min.

The sample was analysed by positive mode electrospray ionization quadrupole time-of-flight mass spectrometry on a maXis™ HD QTOF (Bruker, Bremen, Germany) in full scan mode (50–1500 Da). Data dependent MS/MS was performed by collision-induced dissociation of the three most abundant ions in each scan, making use of Bruker's "Smart Exclusion" functionality to minimize multiple fragmentation of the same ion. The collision energy was ramped from 80% to 200% of the default auto-MS/MS collision energy in order to get more information rich spectra.

The sample was rerun under the same chromatographic and mass spectrometry settings except for a higher injection volume of 3 µL (instead of 1 µL) and the use of

a preferred mass list. The list contained the  $m/z$  of all theoretically possible alkyl quinolones with chain lengths of 0–21 carbon atoms and 0–2 double bonds (Supplementary Table 3.S1). The preferred mass list assured that alkyl quinolones contained in the sample in sufficient concentrations were fragmented on all accounts.

### 3.2.5. Data processing

#### 3.2.5.1. Data conversion and import

LC–MS/MS data were exported to mzXML files using Bruker DataAnalysis and Bruker Compass Xport. All subsequent data analysis steps were carried out utilizing R (version 3.3.1) [47] in the RStudio environment (version 0.99.896) [48]. The code can be accessed via GitHub: <https://github.com/tdepke/CluMSID> and can also be found in the Supplementary Material (Supplementary Computer Code 1). As can be seen from the code, all steps except for the addition of manually evaluated annotations are done automatically without the need for direct user intervention. The mzR package (version 2.4.1) [49–53] was used to import mzXML files into R. The overall number of MS<sup>2</sup> spectra was counted and checked for plausibility as a quality control measure. All MS<sup>2</sup> spectra were extracted and stored as a list. Each list entry corresponded to a spectrum that was represented by a two column matrix containing the  $m/z$  of all fragments and the respective signal intensities. The list was filtered to contain only spectra with more than one fragment peak. Precursor  $m/z$  and retention time were collected for each spectrum.

#### 3.2.5.2. Correction of precursor $m/z$

As already reported by Garg *et al.* [40], mzXML files generated by Bruker Compass Xport retain mass calibration from the \*.baf file for peaks in MS<sup>1</sup> and MS<sup>2</sup> scans, but contain uncalibrated masses for the precursor ions of MS<sup>2</sup> spectra—a circumstance that causes multiple problems in the handling of MS<sup>2</sup> data. To correct this error, we matched the indicated precursor  $m/z$  with the preceding MS<sup>1</sup> scan and took the peak  $m/z$  from the respective spectrum that was closest to it on the condition that it did not deviate more than 100 ppm. *In praxi*, deviations as high as approximately 80 ppm were observed.

### 3.2.5.3. Generation of consensus spectra

In order to reduce artefact masses, species eluting after  $t = 25$  min were excluded. As many ions are fragmented several times despite optimized acquisition parameters, multiple spectra from the same feature were merged employing the following procedure: Spectra were grouped if their precursor  $m/z$  differed less than 10 ppm. For the individual groups, precursor retention times were sorted in ascending order, and each difference between neighbouring retention times was calculated. If neighbouring spectra were recorded within an interval not bigger than 30 s, they were considered to belong to one feature and were merged. A retention time difference of more than 30 s indicated an isobaric species eluting separately and therefore, spectra were assigned to a new feature. For each feature, median  $m/z$  and median retention time of the contributing spectra were calculated. The respective spectra were merged by calculating the mean intensities for all present fragment ions. Fragments were considered the same if their  $m/z$  in the individual spectra had a difference of not more than 10 ppm. The mean of the  $m/z$  values in the individual spectra was given as  $m/z$  in the merged spectrum. If a fragment was not present in a spectrum, an intensity of 0 was used. Feature IDs for all merged spectra were created from precursor ion  $m/z$  and retention times using the same naming formula as XCMS [54].

### 3.2.5.4. Generation of neutral loss patterns

A second list that contained the neutral losses from each spectrum instead of the  $m/z$  of the charged fragment ions was derived from the original spectra. To that end, all fragment peak  $m/z$  were subtracted from the corrected precursor ion  $m/z$ , and the intensity of each fragment peak  $m/z$  was adopted for the respective neutral loss.

### 3.2.5.5. Assignment of manual annotations

A list of feature IDs along with exact precursor ion  $m/z$  and retention time was printed in a human-readable format. Level 1 identifications [55, 56], i. e. annotations based on comparison of exact mass, MS<sup>2</sup> spectrum and retention time to those of an authentic standard measured on the same machine, were assigned to features if they could be matched to authentic standards of our in-house library regarding their exact mass, retention time and MS/MS fragmentation using Bruker Compass DataAnalysis and LibraryEditor. Putative annotations were assigned based on exact mass and MS/MS fragmentation searches in METLIN [34]. In case of alkyl quinolones and rhamnolipids,

exact masses and fragmentation patterns were compared to literature data [14, 57, 58]. Metabolite names were appended to feature IDs, putative annotations in brackets. Putative annotations, including those resulting from the cluster analysis, were additionally validated regarding exact mass and isotope pattern conformity using Bruker's SmartFormula functionality in DataAnalysis software (cf. section 2.5.1) [59]. As recommended by Thiele *et al.* [59],  $50\text{ m}\sigma$  was chosen as upper threshold for isotope pattern conformity. The respective data is listed in Supplementary Table 3.S7.

### 3.2.5.6. Calculation of spectral similarity and cluster analysis

Spectral similarity between two merged spectra  $a$  and  $b$  was calculated using the well-established cosine similarity derived from the spectral contrast angle  $\theta$  with square root-transformed intensities  $a_i$  and  $b_i$  [43, 44, 60]:

$$\cos \theta = \frac{\sum_i a_i \times b_i}{\sqrt{\sum_i a_i^2 \times \sum_i b_i^2}} \quad (3.1)$$

All merged spectra, i. e. the consensus spectra generated as described in 2.5.3, were subjected to pairwise comparison, and a distance matrix was constructed using  $1 - \cos \theta$  as distance. The same procedure was applied for neutral loss patterns generated as described in 2.5.4. For data visualization, metric multidimensional scaling was applied to plot the distances between the spectra in two-dimensional space. The distance matrices were further examined by density based clustering employing the Ordering Points To Identify the Clustering Structure (OPTICS) algorithm [61] from the dbscan R package (version 0.9-8) [62] with an upper limit of the size of the epsilon neighbourhood of 10000 and three as minimum number of points in the epsilon region. The threshold to identify clusters was adjusted to 0.5 for original spectra and 0.7 for neutral loss patterns. Thresholds for cutting dendrograms are by definition to a certain degree arbitrary. The values we chose for this study proved to be useful to get appropriately separated clusters. Cluster IDs were assigned to spectra and reachability plots were created. Furthermore, hierarchical clustering with average linkage as agglomeration method was performed on these distance matrices, and dendrograms were plotted in traditional and circular layout, the latter by the use of the ape R package (version 3.5) [63]. To identify the main clusters, the dendrogram was cut in groups at a height of 0.95, and cluster IDs were assigned to spectra.

### 3.2.5.7. Analysis of samples measured with preferred mass list

The same data analysis workflow was used to analyse the samples measured with a preferred mass list. Putative annotations for alkyl quinolones were assigned based on exact mass and printed in square brackets in the respective figures. These annotations were confirmed or dismissed based on spectral and neutral loss similarity to other alkyl quinolones.

## 3.3. Results and discussion

### 3.3.1. Data pre-processing and manual assignment

The LC-MS/MS results file from the untargeted analysis of a *P. aeruginosa* PA14 cell extract featured 6844 spectra in total, thereof 4202 MS<sup>2</sup> spectra. 2765 MS<sup>2</sup> spectra contained more than one product ion peak. Dismissing spectra whose precursor ions eluted during column washing left 2290 MS<sup>2</sup> spectra that were considered in the analysis. By merging spectra from the same precursor, this number was further reduced to 518 consensus spectra that are assumed to represent distinct molecular features contained in the sample. The spectra held an average of 12 product ion peaks, with 90% of the spectra ranging between 2 and 25 peaks. 61 spectra could be identified by matching exact mass, MS<sup>2</sup> spectrum and retention time to our in-house library, including internal standards and in-source fragments. Another 89 spectra could be putatively identified by MS/MS and exact mass comparison to the METLIN database and relevant literature on *P. aeruginosa* secondary metabolites (Supplementary Table 3.S2). Thus, 368 spectra remained unassigned after matching to the in-house library and the manual assignment. We therefore employed data mining tools to reveal MS<sup>2</sup> spectra similarities that enable putative annotations of various features.

### 3.3.2. Data exploration by multidimensional scaling

Multidimensional scaling (MDS) is a versatile method to visualise distances between data points in multidimensional space and hence was considered adequate to provide an explorative overview of the MS<sup>2</sup> similarity data generated by CluMSID. An MDS plot of MS<sup>2</sup> spectra similarities is shown in Supplementary Fig. 3.S1. Although spatial separation is generally poor, three point clouds can be identified that comprise putatively identified alkyl quinolones, peptides and amino acids as well as several glutamate containing metabolites. Thus, this method could potentially be used to quickly

assess whether a feature belongs to one of those three major clusters of spectra, but other methodologies are to be preferred for a more detailed view that also covers other substance classes.

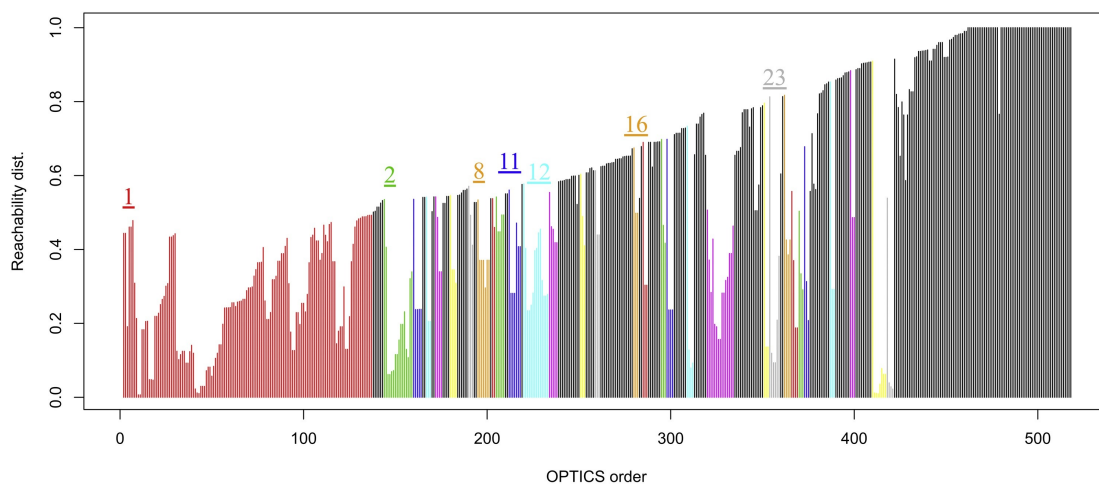
### 3.3.3. Density-based clustering of MS<sup>2</sup> spectra

Based on the structural diversity of metabolites in bacteria in general [64] and *Pseudomonas* species in particular [65], one assumes that an LC–MS/MS dataset of a bacterial cell extract contains sets of spectra displaying high similarity to each other, because they stem from secondary metabolites from the same chemical class or biosynthetic pathway, along with various spectra from structurally different metabolites that can only be weakly related to others. This fits well with the concept of density-based clustering that aims to identify clusters whose members show close proximity to each other, and to classify data points that do not integrate in clusters as noise. As the intra-group similarity of different classes of secondary metabolites is expected to be varying, we chose the OPTICS algorithm [61] as a density-based clustering tool to be included in CluMSID.

Fig. 3.2 provides a graphical overview of the OPTICS clustering results in form of a plot that depicts the reachability distance of each pair of subsequent spectra in the order derived from the density-based clustering. The clusters are colour-coded, while black lines represent noise.

For some groups that are expected to form clusters, density-based clustering yields very good results: Cluster 2 (green in Fig. 3.2) includes various adenine containing metabolites. Phenazine-1-carboxylic acid, phenazine-carboxamide and their in-source fragments are conglomerated in cluster 23 (grey), whereas pyocyanin is located remote from the other phenazines in cluster 16 (orange). The alkyl quinolone spectra are not gathered in one density-based cluster, but rather separately grouped by their chain length: While the C9-congeners are part of the large cluster 1 (red), species with C7, C13 and C11 are grouped in the individual clusters 8, 11 and 12, respectively, comprising both saturated and unsaturated alkyl quinolones of all three classes. Most other alkyl quinolones as well as the large part of rhamnolipids are classified as noise.

Though the density-based clustering functionality of CluMSID succeeded in grouping of some sets of similar metabolites and thus provides useful hints for feature annotation, its overall performance regarding the *P. aeruginosa* cell extract used in this study is not entirely convincing. Nonetheless, OPTICS clustering is implemented in CluMSID as one option of unsupervised methods for the clustering of MS<sup>2</sup> spectra.



**Figure 3.2.:** Reachability plot of OPTICS-clustered MS<sup>2</sup> spectra. The 518 consensus spectra are ordered by the OPTICS algorithm and each bar in the reachability plot represents the OPTICS distance of two consecutive spectra, i.e. low bars signify high similarity and “valleys” in the plot correspond to clusters of similar spectra. Clusters were identified with a reachability distance threshold of 0.5 and colour coded. Clusters discussed in the text are annotated in the plot using their cluster ID: 1: various compounds including C9 alkyl quinolones. 2: adenine containing metabolites. 8: C7 alkyl quinolones. 11: C13 alkyl quinolones. 12: C11 alkyl quinolones. 16: pyocyanin and its derivatives. 23: phenazine-1-carboxylic acid and related compounds. Spectra classified as not belonging to a cluster (“noise”) are coloured in black. A full list of cluster assignments including the colour-coding can be found in Supplementary Table 3.S3.

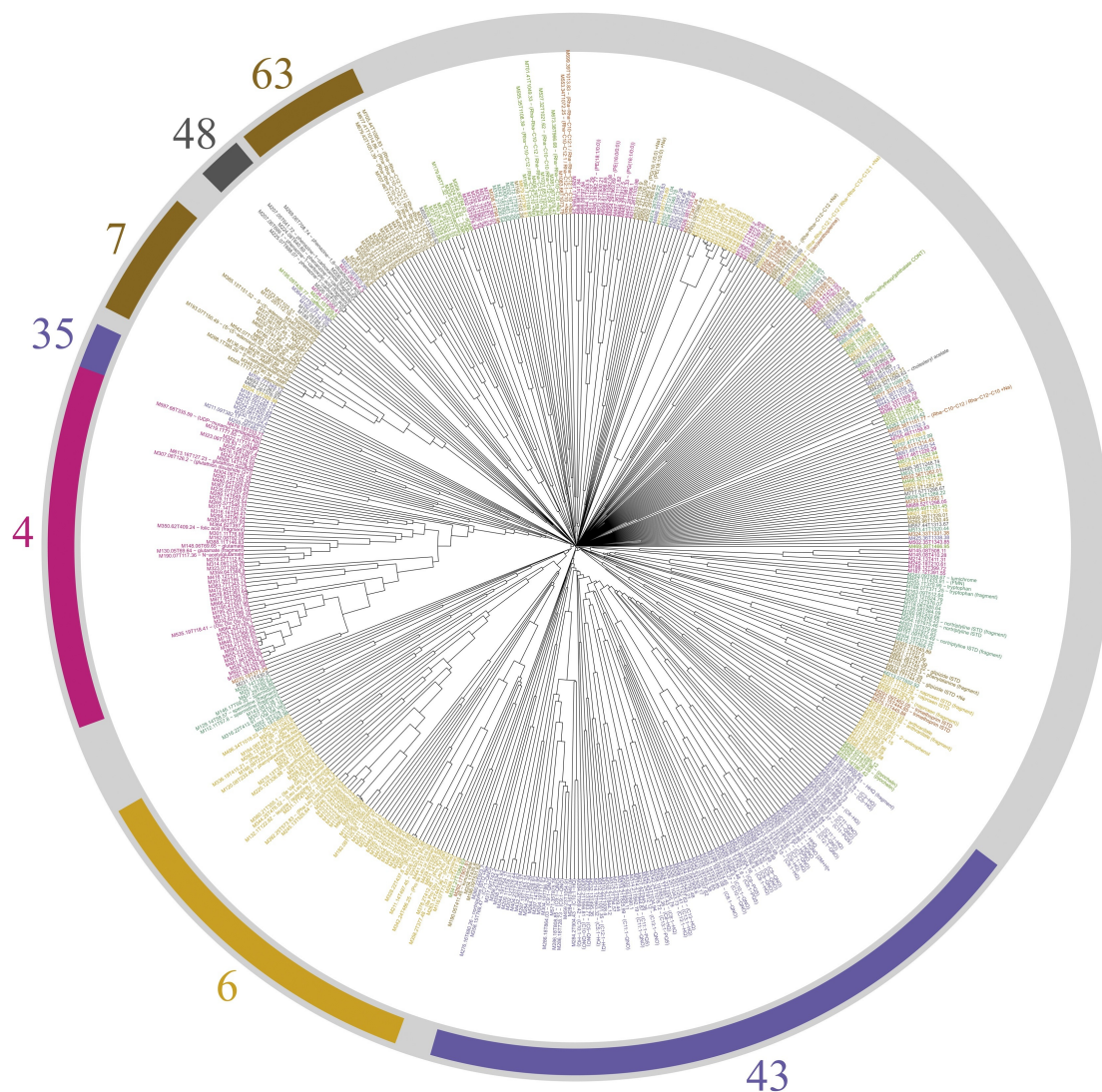


### 3.3.4. Hierarchical clustering of MS<sup>2</sup> spectra

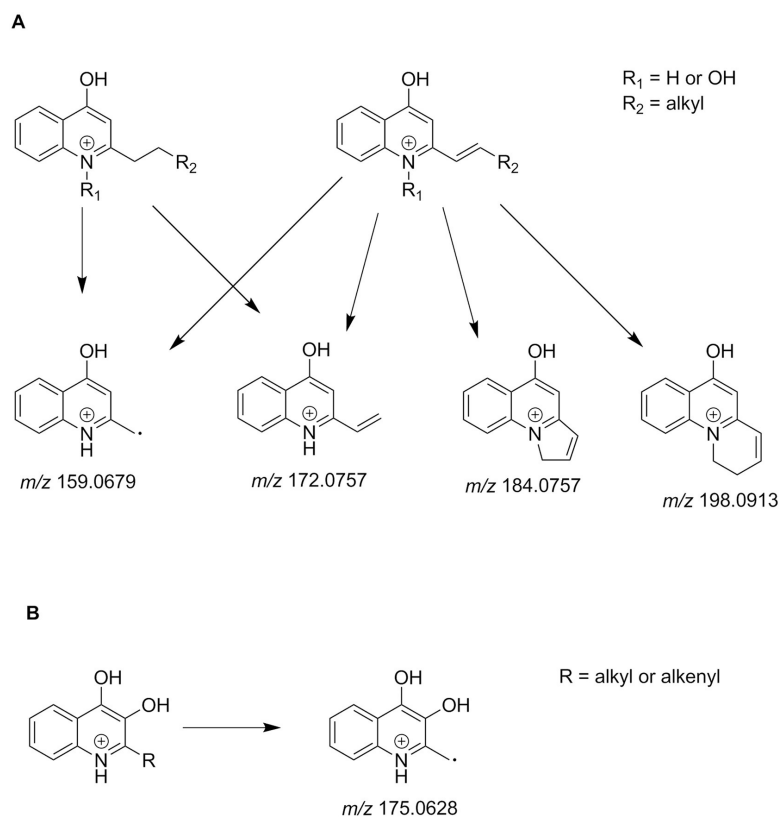
Agglomerative hierarchical clustering, one of the oldest clustering methods, is among the most frequently used data mining tools. It uses pairwise distances—that have in this case been calculated in advance by cosine similarity of MS<sup>2</sup> spectra—to perform a stepwise grouping of the data points into clusters [66]. The results of the hierarchical clustering of all 518 MS<sup>2</sup> spectra by cosine similarity of square root-transformed intensities have been visualized in form of a heat map (Supplementary Fig. 3.S2). Hierarchical clustering with average linkage agglomeration led to the formation of several clusters with different sizes and intermediate to high intra-cluster spectral similarities, whereas only very few features showed considerable similarity to spectra outside their own cluster. More than 90% of all pairwise comparisons (245090 of 268324) resulted in a spectral similarity of zero, i. e. the respective spectra had no fragment ion  $m/z$  in common. Cutting the dendrogram at a height of 0.95 units yielded 125 clusters (Fig. 3.3). Only 29 of them contain more than two features, indicating that many analytes produced MS<sup>2</sup> spectra that cannot be assigned to a group of structurally comparable features. A full list of cluster assignments is provided in Supplementary Table 3.S4.

Cluster 43 contains most features, with 98 spectra assigned to it (Fig. 3.3). According to the manual identification procedure (section 3.1), it includes the known alkyl quinolones HHQ, HQNO and PQS along with multiple putatively identified alkyl quinolones. The only annotated cluster member not belonging to the alkyl quinolone group is a sodium adduct of the internal standard nortriptyline. HHQ and HQNO, whose retention times overlapped, cluster very closely together as they share amongst other the typical alkyl quinolone fragments 159.068 and 172.076 (Fig. 3.4A). In even closer proximity to HQNO, a feature of 519.323 Da can be found, which corresponds to a proton-bound dimer of HQNO. A spectrum with a precursor mass of 503.327 Da clustered next to HHQ with a high cosine similarity of 0.73 also contained the aforementioned typical alkyl quinolone peaks, along with the parent masses of both HHQ and HQNO (Supplementary Fig. 3.S3). The ion probably represents a proton-bound mixed dimer of the two co-eluting alkyl quinolones formed in the ionization source. The phenomenon of closely clustered spectra from  $[M+H]^+$  and  $[2M+H]^+$  ions can also be observed for C9:1-QNO ( $m/z$  286.180 and 571.353), C11:1-PQS ( $m/z$  314.211 and 627.416) and C11:1-HQ ( $m/z$  298.217 and 595.426), amongst others. This proves that spectral similarity can also help to find features that derive from the same analyte.

Apart from the known alkyl quinolones, the cluster contains metabolite spectra assigned to alkyl quinolones that have not been reported in the literature (although



**Figure 3.3.:** Results of the hierarchical clustering with average linkage as agglomeration method of 518 consensus MS<sup>2</sup> spectra visualised as circular dendrogram. Clusters were obtained by cutting at 0.95 are colour coded in the dendrogram. Cluster IDs of clusters discussed in the main text as well as their colour coding is indicated in the outer circle: 4: predominantly glutamate containing metabolites. 6: various peptides and peptide-like metabolites. 7: adenine containing metabolites. 35: pyocyanin and related compounds. 43: alkyl quinolones. 48: non-pyocyanin phenazines. 63: among others rhamnolipid [M+H]<sup>+</sup> features.



**Figure 3.4.:** Most important alkyl quinolone fragmentation reactions according to [14]. A: Fragmentation reactions for HHQ and HQNO congeners with 159.068 as main fragment  $m/z$  for both classes and 184.076 as well as 198.091 as signature fragment  $m/z$  for species with an  $\alpha$ - $\beta$  double bond. B: Main fragmentation reaction for PQS congeners with 175.063 as signature fragment  $m/z$  that is not observed for the other classes.

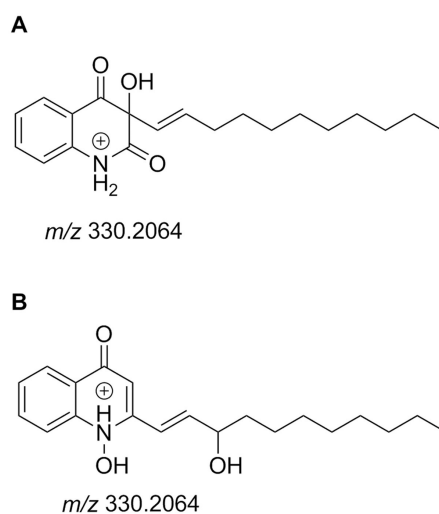
Taylor *et al.* have postulated the existence of short-chain alkyl quinolones based on low resolution GC-electron capture MS data [67]). The components show the typical alkyl quinolone fragmentation and conform to the homologous series of saturated or mono-unsaturated 2-alkyl-4-quinolones (HHQ-like) or 2-alkyl-4-hydroxyquinoline-*N*-oxides (HQNO-like): C4-HQ and C4:1-HQ with  $m/z$  202.122 and 200.107 elute in low concentration at similar retention times. They do not display the most typical fragment ion of  $m/z$  159.068, but another characteristic fragment of  $m/z$  172.076 (Supplementary Fig. 3.S4). A spectrum derived from a precursor of  $m/z$  176.071, clustering close to C5-HQ, can be assigned to C1-QNO based on exact mass and fragmentation pattern. In the same region of the dendrogram, a feature can be putatively identified as C6-HQ build on the same reasoning. The feature M232.13T495.36, whose exact mass matches the known compound C5-QNO, has not been identified by library search, as it lacked the abovementioned main typical fragments associated with alkyl quinolones, presumably because it is present in a too low concentration. However, the fact that it is grouped with the other alkyl quinolones based on spectral similarity supports the assignment.

In addition, the cluster contains a set of spectra that display the characteristic fragmentation pattern, but whose precursor masses do not correspond to members of reported series of alkyl quinolones. Most notably, several features show an  $m/z$  that is 2.02 Da below that of a mono-unsaturated alkyl quinolone, e.g. M312.2T794.57 with an exact mass of 312.1951 Da compared to 314.2121 Da for C11:1-QNO. Using low resolution data, this feature could have easily been mistaken as C12:1-HQ with a theoretical exact mass of 312.2322 Da, an analogue that has also been detected in the sample with a median  $m/z$  of 312.2327. The high resolution data, however, revealed a sum formula of C<sub>20</sub>H<sub>25</sub>NO<sub>2</sub> that is in accordance with a doubly unsaturated C11 hydroxyquinoline-*N*-oxide, C11:2-QNO. The spectrum shows fragment peaks of  $m/z$  184.075 and 198.092, which have been deemed characteristic for (mono-)unsaturated alkyl quinolones by Lépine *et al.* [14] and could be generated in the same manner from doubly unsaturated species (Supplementary Fig. 3.S5). Similarly, features with  $m/z$  of 268.1697, 296.2015, 310.2166, 312.1960, 324.2327 and 340.2276 can be putatively annotated as C9:2-HQ, C11:2-HQ, C12:2-HQ, C11:2-QNO (different isomer), C13:2-HQ and C13:2-QNO, respectively.

Another set of features from the alkyl quinolone cluster exhibits masses that conform to sum formulae of alkyl quinolones possessing an additional oxygen atom, e. g. M330.21T844.12 with an  $m/z$  of 330.2068 that is consistent with the sum formula

$C_{20}H_{27}NO_3$ . Lépine *et al.* have postulated a series of 3-alkyl-2,3-dihydroxy-4-quinolones that are presumably derived from the respective *N*-oxides and produce signature fragments at  $m/z$  178, 162, 160, 146, 144 and 132, the latter being the most intensive [14]. Although the aforementioned feature from this analysis would match the precursor mass of 3-undecenyl-2,3-dihydroxy-4-quinolone (Fig. 3.S5A), the MS<sup>2</sup> spectrum differs considerably from the fragment pattern reported by Lépine *et al.*, containing the characteristic alkyl quinolone fragment at  $m/z$  159.068 along with 184.076, that indicate an unsaturated side chain (Supplementary Fig. 3.S6). The most intensive peak at  $m/z$  312.1964, the mass of the putative C11:2-QNO, is connected to the precursor mass by a neutral loss of H<sub>2</sub>O. The same pattern is apparent in several other features: M304.19T717.91 is contained in the sample in very low concentrations, therefore, typical alkyl quinolone fragment ions cannot be observed. But 286.1797, corresponding to a loss of water and to the mass of C9:1-QNO, is the most prominent peaks in the MS<sup>2</sup> spectrum (Supplementary Fig. 3.S7). This is also confirmed in the isobaric feature M304.19T786.75 that in turn possesses the 159.068 and 184.075 peaks (Supplementary Fig. 3.S8). The same patterns appear in M332.22T890.1, containing a peak whose  $m/z$  corresponds to the mass of C11:1-QNO, and in M358.24T929.17 with a peak matching C13:2-QNO (Supplementary Fig. 3.S9). In a few more features included in the respective cluster, the exact masses conform to this scheme, whereas the fragmentation patterns diverge for unknown reasons. From the present MS/MS data, it cannot be excluded that the abovementioned series of spectra derive from 3-alkyl-2,3-dihydroxy-4-quinolones that undergo differing fragmentation reactions under the conditions applied in this study. Still, the facile loss of water in these compounds suggests that a side chain oxidation accounts for the additional oxygen. This assumption is backed by the absence of a hydroxylated quinolone fragment as produced by PQS congeners (Fig. 3.4B). Therefore, we propose that M330.21T844.12 are products of an allylic oxidation of the side chain (Fig. 3.5B). However, the hierarchical clustering based on spectral similarity as enabled by CluMSID has pointed to the fact that these analytes are related to alkyl quinolones which has been supported by examining the individual fragmentation patterns.

The fact that there are so few 2-alkyl-3-hydroxy-4-quinolones, i. e. PQS analogues, in the cluster seems surprising, but a closer inspection revealed the absence of such spectra outside of cluster 43. It is rather a separation problem that these compounds are only infrequently fragmented in data dependent MS/MS scans: As reported in the literature [68–70], we observed that PQS was prone to significant peak broadening (Supplementary



**Figure 3.5.:** Possible structures of an alkyl quinolone species with the sum formula C<sub>20</sub>H<sub>27</sub>NO<sub>3</sub>. A: 3-Alkyl-2,3-dihydroxy-4-quinolone as postulated by [14]. B: Side chain oxidation product as proposed in this study. Both structures have the same exact mass but the fragmentation pattern with a pronounced loss of water is better explained by structure B.

Fig. 3.S10). Thus, the analysis of PQS analogs by liquid chromatography requires measures like chelating additives in the mobile phase [68] or preanalytical derivatisation [70], which are not suitable for untargeted approaches. A detailed and easy to inspect representation of cluster 43 can be found in Supplementary Fig. 3.S11.

Cluster 4, the second largest cluster, contains glutamate with its signature fragments at  $m/z$  148.060 (precursor), 130.049, 102.055 and 84.044 as well as a number of glutamate-containing substances. For instance, a glutamate in-source fragment and *N*-acetylglutamate clustered very closely to glutamate, and also an in-source fragment of folic acid featuring all the characteristic glutamate fragments can be found in this region of the dendrogram. Furthermore, glutathione disulfide, UDP-muramyl-pentapeptide (doubly charged) and some peptides containing glutamate form part of this cluster. The third largest cluster is cluster 6 that comprises amino acids, peptides and peptide-related compounds such as panthotenate. As peptidic metabolites are not in the focus of this study, clusters 4 and 6 are not discussed in detail—however, it is worth noting that the algorithm is capable of grouping peptides together and can help in their annotation. Another example of the broader utility of the algorithm is cluster 7, which represents a dense group of metabolites containing adenine and some related compounds, namely adenosine, adenosine monophosphate, adenosine diphosphate, flavin adenine dinucleotide, nicotinamide adenine dinucleotide, nicotinamide



adenine dinucleotide phosphate, *S*-adenosyl-homocysteine, deoxyadenosine monophosphate, methylthioadenosine, hypoxanthine and nicotinamide.

While the phenazine virulence factors phenazine-1-carboxamide, phenazine-1-carboxylic acid and phenazine-1,6-dicarboxylic acid along with their in source fragments constitute cluster 48 (Supplementary Fig. 3.S12), the most prominent member of the family is located quite far away from its congeners in cluster 35 (Supplementary Fig. 3.S13). In fact, pyocyanin's tandem mass spectrum displayed a very low cosine similarity of 0.02 to that of phenazine-1-carboxylic acid and no similarity at all to phenazine-1-carboxamide and phenazine-1,6-dicarboxylic acid spectra. This can be attributed to the fact that although pyocyanin has a phenazine scaffold, its different hydroxylation state apparently affects MS/MS fragmentation substantially. Although the known species 1-hydroxyphenazine and 2-hydroxyphenazine were not found in this analysis, a number of other spectra exhibited high similarity to that of pyocyanin. None of them could be identified by library or database searches, but the clustering in combination with biochemical reasoning gave some hints on the compounds' identities. M215.12T626.24 produced a spectrum whose fragment peaks correspond to the most prominent fragments in the spectrum of pyocyanin, and as its precursor mass is 4.03 Da higher, we suggest an annotation as the reduced form of pyocyanin, tetrahydropyocyanin (Supplementary Fig. 3.S14). M254.09T400.89 could, once its structural connection to pyocyanin had been revealed by our algorithm, be identified as aeruginosin A (Supplementary Fig. 3.S15) as originally reported by Holliman in 1969 [71] (not to be confused with a group of octahydroindoles also called aeruginosins [72]). This assignment is supported by comparison of the MS<sup>2</sup> spectrum to one published by Abu *et al.* [73]. Another interesting spectrum with a precursor mass of 255.0761 Da is located directly next to pyocyanin with a high cosine similarity of 0.65. It contains the signature fragments of pyocyanin at  $m/z$  168.068 and 196.063 as well as a peak with the  $m/z$  of pyocyanin itself, 211.087, implying major structural similarity of the two compounds (Supplementary Fig. 3.S16). As the retention time of the two features differs by about 100 s, it can be excluded that the feature in question is an ESI adduct of pyocyanin. The mass difference respectively the neutral loss from 255.076 to 211.087 commonly occurs through the loss of CO<sub>2</sub> from a carboxylic acid. To the best of our knowledge, a pyocyanin carboxylic acid has not been described before, but as pyocyanin is biosynthesised from phenazine-1-carboxylic acid by oxidative decarboxylation, it is conceivable that the equivalent reaction of phenazine-1,6-dicarboxylic acid could produce such a

species. Finally, M181.08T724.14 could be identified as phenazine that is generated spontaneously during the pyocyanin biosynthesis [74].

The third important group of small molecule metabolites associated with *P. aeruginosa* virulence are rhamnolipids. In contrast to alkyl quinolones and phenazines, they are not as nicely grouped in one or two clusters. Most noteworthy is the fact that the [M+H]<sup>+</sup> and the [M+Na]<sup>+</sup> adduct of the same compound do not cluster together, reflecting different fragmentation pathways despite an identical structure of the analyte. Whereas all [M+H]<sup>+</sup> spectra are grouped in cluster 63, the respective sodium adducts are more scattered. At least spectra from some rhamnolipid sodium adducts were closely similar to those derived from [2M+Na]<sup>+</sup> adducts.

### 3.3.5. Hierarchical clustering of neutral loss patterns

Structural similarity can not only be revealed by shared charged fragments, but also by common neutral losses. Therefore, clustering based on the similarity of neutral loss patterns can provide useful complimentary information to the data generated by comparing product ion spectra. We therefore implemented functionalities in CluMSID that allow for the generation of neutral loss patterns and the respective distance matrix as well as the mining of this information by MDS, density based clustering and hierarchical clustering. In contrast to the approach employed by Li *et al.* [41] that relies on matching neutral losses to a library of commonly observed uncharged fragments, CluMSID preserves all neutral loss information by generating neutral loss patterns and making use of cosine similarity for the comparison.

A list of cluster assignments and the respective dendrograms produced by these methods can be found in Supplementary Table 3.S5 and Fig. 3.S17. An interesting observation of the utility of neutral loss based clustering can be made in this study by comparing the surrounding environment of nucleotides in the MS<sup>2</sup> spectra and the neutral loss patterns dendrograms. As mentioned in Section 3.4, the product ion spectrum of AMP clusters alongside those of other adenine-containing metabolites, whereas its highest neutral loss similarities are to GMP and CMP (cluster 9). That is expected, as all of these molecules lose phosphate and ribose.

Neutral loss similarity can also be misleading, as exemplified by cluster 11, which includes alkyl quinolones, two phenazines and the internal standard nortriptyline along with glutamyl-alanine, nicotinamide, hypoxanthine and anthranilate. This is a rather heterogeneous group from a structural point of view. On the other hand, the clusters



comprising the putatively identified rhamnolipids are much denser than those obtained by spectral clustering of product ions.

Density based clustering and multidimensional scaling of neutral loss pattern data yielded only little informative grouping of spectra (data not shown) and shall therefore not be discussed in this article; it is, however, possible with CluMSID and might prove useful in other contexts.

### 3.3.6. Semi-targeted analysis with preferred mass lists

The hierarchical cluster analysis of the data from the semi-targeted MS/MS method could nicely reproduce the grouping of alkyl quinolones in one cluster, as almost all of them are assigned to cluster 29 (see Supplementary Table 3.S6 for a full list and Supplementary Fig. 3.S18 for the respective dendrogram), underlining the utility of CluMSID.

Table 3.1 lists all alkyl quinolones from the three canonical classes that have been identified by semi-targeted analysis of a *P. aeruginosa* cell extract, in total 59 different species, not counting *cis-trans* isomers. Especially the description of doubly unsaturated congeners represents an important extension of the *P. aeruginosa* “quinolinome”. The detection of alkyl quinolones with C15 and even one with a C17 side chain also represent minor advances in defining the chemodiversity of quorum sensing signal molecules.

The authors are aware that the experimental setup of this study, i. e. the use of a pooled sample of cultures grown in minimal medium under limited oxygenisation, does not allow to generalise on the multitude of environments *P. aeruginosa* is able to inhabit. It should also be stressed that the CluMSID workflow may only be used to gain qualitative information on metabolite identities and no quantitative data on production levels *et cetera*. Still, in summary, CluMSID led, in combination with the semi-targeted analysis with preferred mass lists, to the MS<sup>2</sup>-based identification of 27 alkyl quinolone species that have, to the best of our knowledge, not been reported before. By enhancing the coverage of low-abundant congeners in a systematic manner, CluMSID helped defining the chemodiversity of quorum sensing signal molecules.

## 3.4. Conclusion

We addressed a major challenge in mass spectrometry-based untargeted metabolomics, the feature identification, by developing a tool that clusters MS<sup>2</sup> product ion spectra and neutral loss patterns based on their cosine similarity using different unsupervised

**Table 3.1.:** Alkyl quinolones identified by semi-targeted analysis.

<b>2-alkyl-4-quinolones</b>			<b>2-alkyl-4-hydroxyquinoline-<i>N</i>-oxides</b>		
side chain	<i>m/z</i>	previously reported:	side chain	<i>m/z</i>	previously reported:
C1	160.0757		C1	176.0706	
C2	174.0914		C2	190.0863	
C3	188.1070	[58]	C3	204.1019	
C4	202.1227		C3:1	202.0863	
C5	216.1383	[14, 58]	C5	232.1332	[14]
C5:1	214.1227	[14]	C5:1	230.1176	
C6	230.1540	[14, 58]	C6	246.1489	[14]
C6:1	228.1383	[14]	C7	260.1645	[14]
C7	244.1696	[14]	C7:1	258.1489	[14]
C7:1	242.1540	[14, 58]	C8	274.1802	[14]
C8	258.1853	[14, 58]	C8:1	272.1645	[14]
C8:1	256.1696	[14]	C9	288.195	[14]
C9	272.2009	[14, 58]	C9:1	286.1802	[14]
C9:1	270.1853	[14, 58]	C9:2	284.1645	
C9:2	268.1696		C10	302.2115	[14]
C10	286.2166	[14]	C10:1	300.1958	[14]
C10:1	284.2009	[14]	C11	316.2271	[14]
C11	300.2322	[14, 58]	C11:1	314.2115	[14]
C11:1	298.2166	[14, 58]	C11:2	312.1958	
C11:2	296.2009		C12:1	328.2271	
C12	314.2479		C13	344.2584	
C12:1	312.2322	[14, 58]	C13:1	342.2428	
C12:2	310.2166		C13:2	340.2271	
C13	328.2635	[14, 58]			
C13:1	326.2479	[14, 58]			
C13:2	324.2322				
C15	356.2948				
C15:1	354.2792				
C15:2	352.2635				
C17:1	382.3105				
C10:1	300.1958				
C11:1	314.2115				
			<b>2-alkyl-3-hydroxy-4-quinolones</b>		
side chain	<i>m/z</i>	previously reported:	side chain	<i>m/z</i>	previously reported:
C7	260.1645	[14]	C7	260.1645	[14]
C7:1	258.1489		C7:1	258.1489	
C9	288.1958	[14]	C9	288.1958	[14]
C9:1	286.1802		C9:1	286.1802	

methods. The program called CluMSID is provided as an open source tool usable by anyone in the metabolomics community.

The utility of CluMSID was demonstrated by the identification of 27 novel members of the canonical alkyl quinolones, including a series of AQs with doubly unsaturated side chain. Also pyocyanin congeners that were either unknown or could not be found in common metabolite databases were annotated. In addition, CluMSID correctly grouped metabolites with common functional elements (e. g. peptides, nucleotides) in clusters, thereby facilitating a structural annotation. Therefore, we are convinced that CluMSID is a beneficial tool for the study of metabolomes of diverse provenance.

## Acknowledgements

The authors would like to thank Prof. Susanne Häußler, HZI Braunschweig, for helpful guidance as well as for providing us with the *Pseudomonas aeruginosa* PA14 strain and Ulrike Beutling, HZI Braunschweig, for assistance and guidance. Many thanks also to Prof. Frank Klawonn, HZI Braunschweig and Ostfalia Hochschule für angewandte Wissenschaften, Wolfenbüttel, for fruitful discussions about data mining methods. T.D. would like to thank the Studienstiftung des deutschen Volkes for financial and non-material support. This research project has been supported by the President's Initiative and Networking Funds of the Helmholtz Association of German Research Centres (HGF) under contract number VH-GS-202, and the EU-funded European Marine Biological Research Infrastructure Cluster (EMBRIC, code 654008).

## References

- [1] G. P. Bodey, R. Bolivar, V. Fainstein, and L. Jadeja. "Infections Caused by *Pseudomonas aeruginosa*". In: *Clinical Infectious Diseases* 5.2 (Mar. 1983), pp. 279–313. DOI: 10.1093/clinids/5.2.279.
- [2] J. B. Lyczak, C. L. Cannon, and G. B. Pier. "Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist". In: *Microbes and infection* 2.9 (2000), pp. 1051–1060.
- [3] J. B. Lyczak, C. L. Cannon, and G. B. Pier. "Lung Infections Associated with Cystic Fibrosis". In: *Clinical Microbiology Reviews* 15.2 (Apr. 2002), pp. 194–222. DOI: 10.1128/cmr.15.2.194-222.2002.
- [4] N. Mesaros, P. Nordmann, P. Plésiat, M. Roussel-Delvallez, J. V. Eldere, Y. Glupczynski, Y. V. Laethem, F. Jacobs, P. Lebecque, A. Malfroot, P. Tulkens, and F. V. Bambeke. "*Pseudomonas aeruginosa*: resistance and therapeutic options at the turn of the new millennium". In: *Clinical Microbiology and Infection* 13.6 (June 2007), pp. 560–578. DOI: 10.1111/j.1469-0691.2007.01681.x.
- [5] N. Høiby, T. Bjarnsholt, M. Givskov, S. Molin, and O. Ciofu. "Antibiotic resistance of bacterial biofilms". In: *International Journal of Antimicrobial Agents* 35.4 (Apr. 2010), pp. 322–332. DOI: 10.1016/j.ijantimicag.2009.12.011.

- [6] E. Drenkard. “Antimicrobial resistance of *Pseudomonas aeruginosa* biofilms”. In: *Microbes and Infection* 5.13 (Nov. 2003), pp. 1213–1219. DOI: 10.1016/j.micinf.2003.08.009.
- [7] T. Strateva and I. Mitov. “Contribution of an arsenal of virulence factors to pathogenesis of *Pseudomonas aeruginosa* infections”. In: *Annals of Microbiology* 61.4 (May 2011), pp. 717–732. DOI: 10.1007/s13213-011-0273-y.
- [8] P. N. Jimenez, G. Koch, J. A. Thompson, K. B. Xavier, R. H. Cool, and W. J. Quax. “The Multiple Signaling Systems Regulating Virulence in *Pseudomonas aeruginosa*”. In: *Microbiology and Molecular Biology Reviews* 76.1 (Mar. 2012), pp. 46–65. DOI: 10.1128/mnbr.05007-11.
- [9] D. V. Mavrodi, W. Blankenfeldt, and L. S. Thomashow. “Phenazine Compounds in Fluorescent *Pseudomonas* Spp.: Biosynthesis and Regulation”. In: *Annual Review of Phytopathology* 44.1 (Sept. 2006), pp. 417–445. DOI: 10.1146/annurev.phyto.44.013106.145710.
- [10] G. W. Lau, D. J. Hassett, H. Ran, and F. Kong. “The role of pyocyanin in *Pseudomonas aeruginosa* infection”. In: *Trends in Molecular Medicine* 10.12 (Dec. 2004), pp. 599–606. DOI: 10.1016/j.molmed.2004.10.002.
- [11] L. Zulianello, C. Canard, T. Kohler, D. Caille, J.-S. Lacroix, and P. Meda. “Rhamnolipids Are Virulence Factors That Promote Early Infiltration of Primary Human Airway Epithelia by *Pseudomonas aeruginosa*”. In: *Infection and Immunity* 74.6 (May 2006), pp. 3134–3147. DOI: 10.1128/iai.01772-05.
- [12] P. O. Jensen, T. Bjarnsholt, R. Phipps, T. B. Rasmussen, H. Calum, L. Christoffersen, C. Moser, P. Williams, T. Pressler, M. Givskov, and N. Hoiby. “Rapid necrotic killing of polymorphonuclear leukocytes is caused by quorum-sensing-controlled production of rhamnolipid by *Pseudomonas aeruginosa*”. In: *Microbiology* 153.5 (May 2007), pp. 1329–1338. DOI: 10.1099/mic.0.2006/003863-0.
- [13] M. Alhede, T. Bjarnsholt, P. O. Jensen, R. K. Phipps, C. Moser, L. Christophersen, L. D. Christensen, M. van Gennip, M. Parsek, N. Hoiby, T. B. Rasmussen, and M. Givskov. “*Pseudomonas aeruginosa* recognizes and responds aggressively to the presence of polymorphonuclear leukocytes”. In: *Microbiology* 155.11 (July 2009), pp. 3500–3508. DOI: 10.1099/mic.0.031443-0.
- [14] F. Lépine, S. Milot, E. Déziel, J. He, and L. G. Rahme. “Electrospray/mass spectrometric identification and analysis of 4-hydroxy-2-alkylquinolines (HAQs) produced by *Pseudomonas aeruginosa*”. In: *Journal of the American Society for Mass Spectrometry* 15.6 (June 2004), pp. 862–869. DOI: 10.1016/j.jasms.2004.02.012.
- [15] B. B. Aldridge and K. Y. Rhee. “Microbial metabolomics: innovation, application, insight”. In: *Current Opinion in Microbiology* 19 (June 2014), pp. 90–96. DOI: 10.1016/j.mib.2014.06.009.
- [16] G. J. Patti, O. Yanes, and G. Siuzdak. “Metabolomics: the apogee of the omics trilogy”. In: *Nature Reviews Molecular Cell Biology* 13.4 (Apr. 2012), pp. 263–269. DOI: 10.1038/nrm3314.
- [17] E. Frimmersdorf, S. Horatzek, A. Pelnikevich, L. Wiehlmann, and D. Schomburg. “How *Pseudomonas aeruginosa* adapts to various environments: a metabolomic approach”. In: *Environmental Microbiology* 12.6 (Feb. 2010), pp. 1734–1747. DOI: 10.1111/j.1462-2920.2010.02253.x.
- [18] N. Saito, Y. Ohashi, T. Soga, and M. Tomita. “Unveiling cellular biochemical reactions via metabolomics-driven approaches”. In: *Current Opinion in Microbiology* 13.3 (June 2010), pp. 358–362. DOI: 10.1016/j.mib.2010.04.006.

- [19] V. Behrends, T. J. Bell, M. Liebeke, A. Cordes-Blauert, S. N. Ashraf, C. Nair, J. E. A. Zlosnik, H. D. Williams, and J. G. Bundy. “Metabolite Profiling to Characterize Disease-related Bacteria”. In: *Journal of Biological Chemistry* 288.21 (Apr. 2013), pp. 15098–15109. DOI: 10.1074/jbc.m112.442814.
- [20] R. t’Kindt, R. A. Scheltema, A. Jankevics, K. Brunker, S. Rijal, J.-C. Dujardin, R. Breitling, D. G. Watson, G. H. Coombs, and S. Decuypere. “Metabolomics to Unveil and Understand Phenotypic Diversity between Pathogen Populations”. In: *PLoS Neglected Tropical Diseases* 4.11 (Nov. 2010). Ed. by T. G. Geary, e904. DOI: 10.1371/journal.pntd.0000904.
- [21] D. S. Wishart. “Computational strategies for metabolite identification in metabolomics”. In: *Bioanalysis* 1.9 (Dec. 2009), pp. 1579–1596. DOI: 10.4155/bio.09.138.
- [22] M. Brown, W. B. Dunn, P. Dobson, Y. Patel, C. L. Winder, S. Francis-McIntyre, P. Begley, K. Carroll, D. Broadhurst, A. Tseng, N. Swainston, I. Spasic, R. Goodacre, and D. B. Kell. “Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics”. In: *The Analyst* 134.7 (2009), p. 1322. DOI: 10.1039/b901179j.
- [23] W. B. Dunn, A. Erban, R. J. M. Weber, D. J. Creek, M. Brown, R. Breitling, T. Hankemeier, R. Goodacre, S. Neumann, J. Kopka, and M. R. Viant. “Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics”. In: *Metabolomics* 9.S1 (May 2012), pp. 44–66. DOI: 10.1007/s11306-012-0434-4.
- [24] T. Kind and O. Fiehn. “Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry”. In: *BMC Bioinformatics* 8.1 (2007), p. 105. DOI: 10.1186/1471-2105-8-105.
- [25] F. Matsuda. “Regular expressions of MS/MS spectra for partial annotation of metabolite features”. In: *Metabolomics* 12.7 (June 2016). DOI: 10.1007/s11306-016-1052-3.
- [26] J. E. Peironcely, M. Rojas-Chertó, A. Tas, R. Vreeken, T. Reijmers, L. Coulier, and T. Hankemeier. “Automated Pipeline for De Novo Metabolite Identification Using Mass-Spectrometry-Based Metabolomics”. In: *Analytical Chemistry* 85.7 (Mar. 2013), pp. 3576–3583. DOI: 10.1021/ac303218u.
- [27] E. Rathahao-Paris, S. Alves, C. Junot, and J.-C. Tabet. “High resolution mass spectrometry for structural identification of metabolites in metabolomics”. In: *Metabolomics* 12.1 (Nov. 2015). DOI: 10.1007/s11306-015-0882-8.
- [28] M. Vinaixa, E. L. Schymanski, S. Neumann, M. Navarro, R. M. Salek, and O. Yanes. “Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects”. In: *TrAC Trends in Analytical Chemistry* 78 (Apr. 2016), pp. 23–35. DOI: 10.1016/j.trac.2015.09.005.
- [29] K. Dührkop, H. Shen, M. Meusel, J. Rousu, and S. Böcker. “Searching molecular structure databases with tandem mass spectra using CSI:FingerID”. In: *Proceedings of the National Academy of Sciences* 112.41 (Sept. 2015), pp. 12580–12585. DOI: 10.1073/pnas.1509788112.
- [30] S. Wolf, S. Schmidt, M. Müller-Hannemann, and S. Neumann. “In silico fragmentation for computer assisted identification of metabolite mass spectra”. In: *BMC Bioinformatics* 11.1 (2010), p. 148. DOI: 10.1186/1471-2105-11-148.
- [31] M. Kotera, A. G. McDonald, S. Boyce, and K. F. Tipton. “Functional Group and Substructure Searching as a Tool in Metabolomics”. In: *PLoS ONE* 3.2 (Feb. 2008). Ed. by J. Zhu, e1537. DOI: 10.1371/journal.pone.0001537.

- [32] S. Böcker and K. Dührkop. “Fragmentation trees reloaded”. In: *Journal of Cheminformatics* 8.1 (Feb. 2016). DOI: 10.1186/s13321-016-0116-8.
- [33] S. Aros-Calt, B. H. Muller, S. Boudah, C. Ducruix, G. Gervasi, C. Junot, and F. Fenaille. “Annotation of the *Staphylococcus aureus* Metabolome Using Liquid Chromatography Coupled to High-Resolution Mass Spectrometry and Application to the Study of Methicillin Resistance”. In: *Journal of Proteome Research* 14.11 (Oct. 2015), pp. 4863–4875. DOI: 10.1021/acs.jproteome.5b00697.
- [34] C. A. Smith, G. O. Maille, E. J. Want, C. Qin, S. A. Trauger, T. R. Brandon, D. E. Custodio, R. Abagyan, and G. Siuzdak. “METLIN”. In: *Therapeutic Drug Monitoring* 27.6 (Dec. 2005), pp. 747–751. DOI: 10.1097/01.ftd.0000179845.53213.39.
- [35] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, and T. Nishioka. “MassBank: a public repository for sharing mass spectral data for life sciences”. In: *Journal of Mass Spectrometry* 45.7 (July 2010), pp. 703–714. DOI: 10.1002/jms.1777.
- [36] D. S. Wishart, T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorn Dahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner, and A. Scalbert. “HMDB 3.0—The Human Metabolome Database in 2013”. In: *Nucleic Acids Research* 41.D1 (Nov. 2012), pp. D801–D807. DOI: 10.1093/nar/gks1065.
- [37] T. Sajed, A. Marcu, M. Ramirez, A. Pon, A. C. Guo, C. Knox, M. Wilson, J. R. Grant, Y. Djoumbou, and D. S. Wishart. “ECMDB 2.0: A richer resource for understanding the biochemistry of *E. coli*”. In: *Nucleic Acids Research* 44.D1 (Oct. 2015), pp. D495–D501. DOI: 10.1093/nar/gkv1060.
- [38] D. L. Tabb, M. J. MacCoss, C. C. Wu, S. D. Anderson, and J. R. Yates. “Similarity among Tandem Mass Spectra from Proteomic Experiments: Detection, Significance, and Utility”. In: *Analytical Chemistry* 75.10 (May 2003), pp. 2470–2477. DOI: 10.1021/ac026424o.
- [39] J. Watrous, P. Roach, T. Alexandrov, B. S. Heath, J. Y. Yang, R. D. Kersten, M. van der Voort, K. Pogliano, H. Gross, J. M. Raaijmakers, B. S. Moore, J. Laskin, N. Bandeira, and P. C. Dorrestein. “Mass spectral molecular networking of living microbial colonies”. In: *Proceedings of the National Academy of Sciences* 109.26 (May 2012), E1743–E1752. DOI: 10.1073/pnas.1203689109.
- [40] N. Garg, C. A. Kapon, Y. W. Lim, N. Koyama, M. J. Vermeij, D. Conrad, F. Rohwer, and P. C. Dorrestein. “Mass spectral similarity for untargeted metabolomics data analysis of complex mixtures”. In: *International Journal of Mass Spectrometry* 377 (Feb. 2015), pp. 719–727. DOI: 10.1016/j.ijms.2014.06.005.
- [41] D. Li, I. T. Baldwin, and E. Gaquerel. “Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis”. In: *Proceedings of the National Academy of Sciences* 112.30 (July 2015), E4147–E4155. DOI: 10.1073/pnas.1503106112.
- [42] M. Wang et al. “Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking”. In: *Nature Biotechnology* 34.8 (Aug. 2016), pp. 828–837. DOI: 10.1038/nbt.3597.



- [43] S. E. Stein and D. R. Scott. “Optimization and testing of mass spectral library search algorithms for compound identification”. In: *Journal of the American Society for Mass Spectrometry* 5.9 (Sept. 1994), pp. 859–866. DOI: 10.1016/1044-0305(94)87009-8.
- [44] K. X. Wan, I. Vidavsky, and M. L. Gross. “Comparing similar spectra: From similarity index to spectral contrast angle”. In: *Journal of the American Society for Mass Spectrometry* 13.1 (Jan. 2002), pp. 85–88. DOI: 10.1016/S1044-0305(01)00327-0.
- [45] H. Treutler, H. Tsugawa, A. Porzel, K. Gorzolka, A. Tissier, S. Neumann, and G. U. Balcke. “Discovering Regulated Metabolite Families in Untargeted Metabolomics Studies”. In: *Analytical Chemistry* 88.16 (Aug. 2016), pp. 8082–8090. DOI: 10.1021/acs.analchem.6b01569.
- [46] G. Bertani. “Studies on Lysogenesis I.: The Mode of Phage Liberation by Lysogenic *Escherichia coli*”. In: *Journal of bacteriology* 62.3 (1951), p. 293.
- [47] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2017. URL: <https://www.R-project.org/>.
- [48] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc. Boston, MA, 2015. URL: <http://www.rstudio.com/>.
- [49] M. C. Chambers, B. Maclean, R. Burke, D. Amodei, D. L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T. A. Baker, M.-Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S. L. Seymour, L. M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E. W. Deutsch, R. L. Moritz, J. E. Katz, D. B. Agus, M. MacCoss, D. L. Tabb, and P. Mallick. “A cross-platform toolkit for mass spectrometry and proteomics”. In: *Nature Biotechnology* 30.10 (Oct. 2012), pp. 918–920. DOI: 10.1038/nbt.2377.
- [50] L. Martens, M. Chambers, M. Sturm, D. Kessner, F. Levander, J. Shofstahl, W. H. Tang, A. Römpp, S. Neumann, A. D. Pizarro, L. Montecchi-Palazzi, N. Tasman, M. Coleman, F. Reisinger, P. Souda, H. Hermjakob, P.-A. Binz, and E. W. Deutsch. “mzML—a Community Standard for Mass Spectrometry Data”. In: *Molecular & Cellular Proteomics* 10.1 (Aug. 2010), R110.000133. DOI: 10.1074/mcp.r110.000133.
- [51] P. G. A. Pedrioli, J. K. Eng, R. Hubley, M. Vogelzang, E. W. Deutsch, B. Raught, B. Pratt, E. Nilsson, R. H. Angeletti, R. Apweiler, K. Cheung, C. E. Costello, H. Hermjakob, S. Huang, R. K. Julian, E. Kapp, M. E. McComb, S. G. Oliver, G. Omenn, N. W. Paton, R. Simpson, R. Smith, C. F. Taylor, W. Zhu, and R. Aebersold. “A common open representation of mass spectrometry data and its application to proteomics research”. In: *Nature Biotechnology* 22.11 (Nov. 2004), pp. 1459–1466. DOI: 10.1038/nbt1031.
- [52] A. Keller, J. Eng, N. Zhang, X.-j. Li, and R. Aebersold. “A uniform proteomics MS/MS analysis platform utilizing open XML file formats”. In: *Molecular Systems Biology* 1.1 (Aug. 2005), E1–E8. DOI: 10.1038/msb4100024.
- [53] D. Kessner, M. Chambers, R. Burke, D. Agus, and P. Mallick. “ProteoWizard: open source software for rapid proteomics tools development”. In: *Bioinformatics* 24.21 (July 2008), pp. 2534–2536. DOI: 10.1093/bioinformatics/btn323.
- [54] C. A. Smith, E. J. Want, G. O’Maille, R. Abagyan, and G. Siuzdak. “XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification”. In: *Analytical Chemistry* 78.3 (Feb. 2006), pp. 779–787. DOI: 10.1021/ac051437y.
- [55] R. M. Salek, C. Steinbeck, M. R. Viant, R. Goodacre, and W. B. Dunn. “The role of reporting standards for metabolite annotation and identification in metabolomic studies”. In: *GigaScience* 2.1 (Oct. 2013). DOI: 10.1186/2047-217x-2-13.

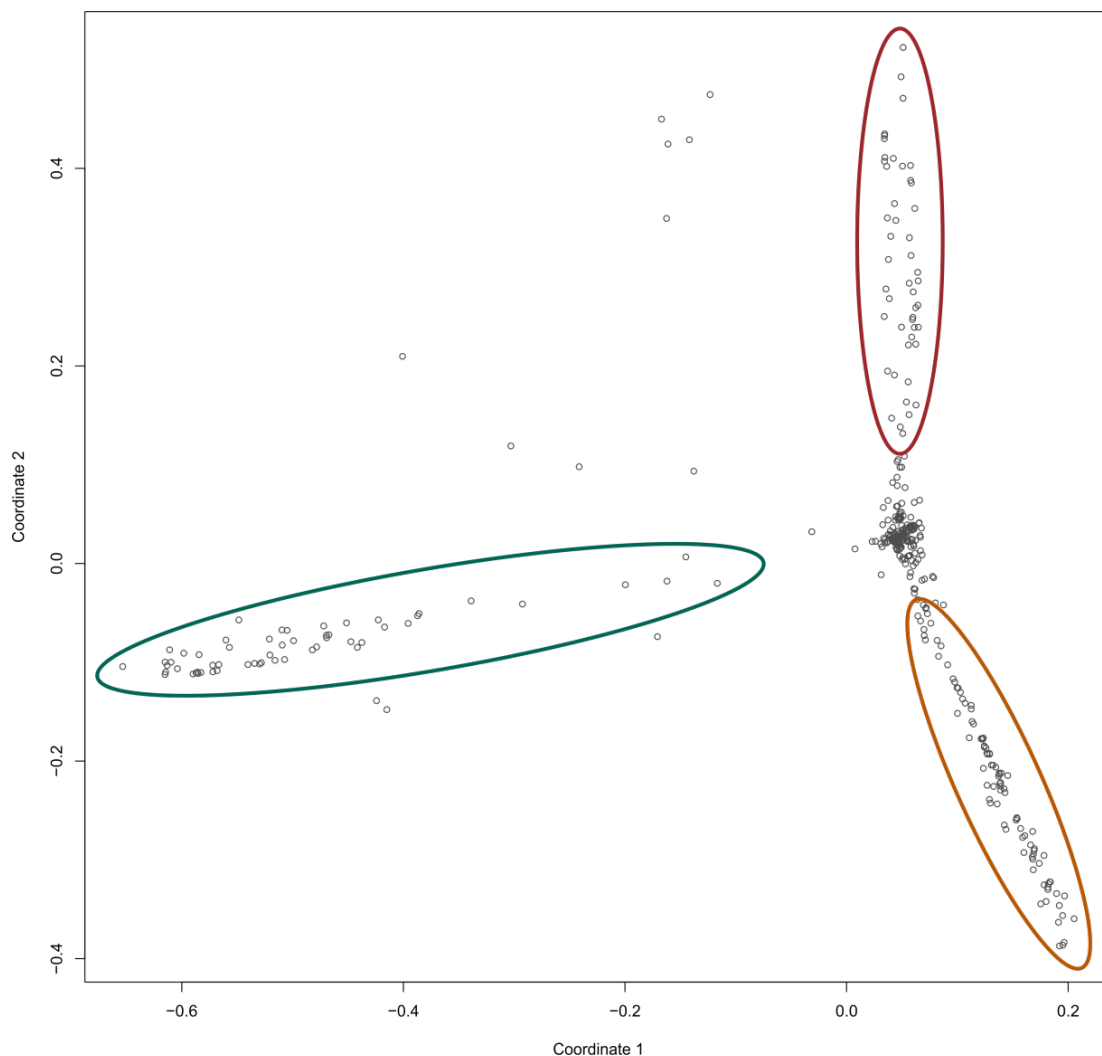
- [56] L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W.-M. Fan, O. Fiehn, R. Goodacre, J. L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A. N. Lane, J. C. Lindon, P. Marriott, A. W. Nicholls, M. D. Reily, J. J. Thaden, and M. R. Viant. “Proposed minimum reporting standards for chemical analysis”. In: *Metabolomics* 3.3 (Sept. 2007), pp. 211–221. DOI: 10.1007/s11306-007-0082-2.
- [57] A. M. Abdel-Mawgoud, F. Lépine, and E. Déziel. “Rhamnolipids: diversity of structures, microbial origins and roles”. In: *Applied Microbiology and Biotechnology* 86.5 (Mar. 2010), pp. 1323–1336. DOI: 10.1007/s00253-010-2498-2.
- [58] H. Niewerth, K. Bergander, S. R. Chhabra, P. Williams, and S. Fetzner. “Synthesis and biotransformation of 2-alkyl-4(1*H*)-quinolones by recombinant *Pseudomonas putida* KT2440”. In: *Applied Microbiology and Biotechnology* 91.5 (June 2011), pp. 1399–1408. DOI: 10.1007/s00253-011-3378-0.
- [59] H. Thiele, G. McLeod, M. Niemitz, and T. Kühn. “Structure verification of small molecules using mass spectrometry and NMR spectroscopy”. In: *Monatshefte für Chemie - Chemical Monthly* 142.7 (Apr. 2011), pp. 717–730. DOI: 10.1007/s00706-011-0486-6.
- [60] J. Liu and V. M. Karbhari. “Performance and design of fibre-reinforced polymer composites at cold temperatures current status and future needs”. In: *International Journal of Materials and Product Technology* 28.1/2 (2007), p. 1. DOI: 10.1504/ijmpt.2007.011506.
- [61] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander. “OPTICS”. In: *ACM SIGMOD Record* 28.2 (June 1999), pp. 49–60. DOI: 10.1145/304181.304187.
- [62] M. Hahsler and M. Piekenbrock. *dbscan: Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms*. R package version 1.1-1. 2017. URL: <https://CRAN.R-project.org/package=dbscan>.
- [63] E. Paradis, J. Claude, and K. Strimmer. “APE: Analyses of Phylogenetics and Evolution in R language”. In: *Bioinformatics* 20.2 (Jan. 2004), pp. 289–290. DOI: 10.1093/bioinformatics/btg412.
- [64] V. Malik. “Microbial secondary metabolism”. In: *Trends in Biochemical Sciences* 5.3 (Mar. 1980), pp. 68–72. DOI: 10.1016/0968-0004(80)90071-7.
- [65] H. Gross and J. E. Loper. “Genomics of secondary metabolite production by *Pseudomonas* spp.” In: *Natural Product Reports* 26.11 (2009), p. 1408. DOI: 10.1039/b817075b.
- [66] M. R. Berthold, C. Borgelt, F. Höppner, and F. Klawonn. *Guide to intelligent data analysis: how to intelligently make sense of real data*. Springer Science & Business Media, 2010.
- [67] G. W. Taylor, Z. A. Machan, S. Mehmet, P. J. Cole, and R. Wilson. “Rapid identification of 4-hydroxy-2-alkylquinolines produced by *Pseudomonas aeruginosa* using gas chromatography—electron-capture mass spectrometry”. In: *Journal of Chromatography B: Biomedical Sciences and Applications* 664.2 (Feb. 1995), pp. 458–462. DOI: 10.1016/0378-4347(94)00494-p.
- [68] C. A. Ortori, J.-F. Dubern, S. R. Chhabra, M. Cámara, K. Hardie, P. Williams, and D. A. Barrett. “Simultaneous quantitative profiling of *N*-acyl-L-homoserine lactone and 2-alkyl-4(1*H*)-quinolone families of quorum-sensing signaling molecules using LC-MS/MS”. In: *Analytical and Bioanalytical Chemistry* 399.2 (Oct. 2010), pp. 839–850. DOI: 10.1007/s00216-010-4341-0.



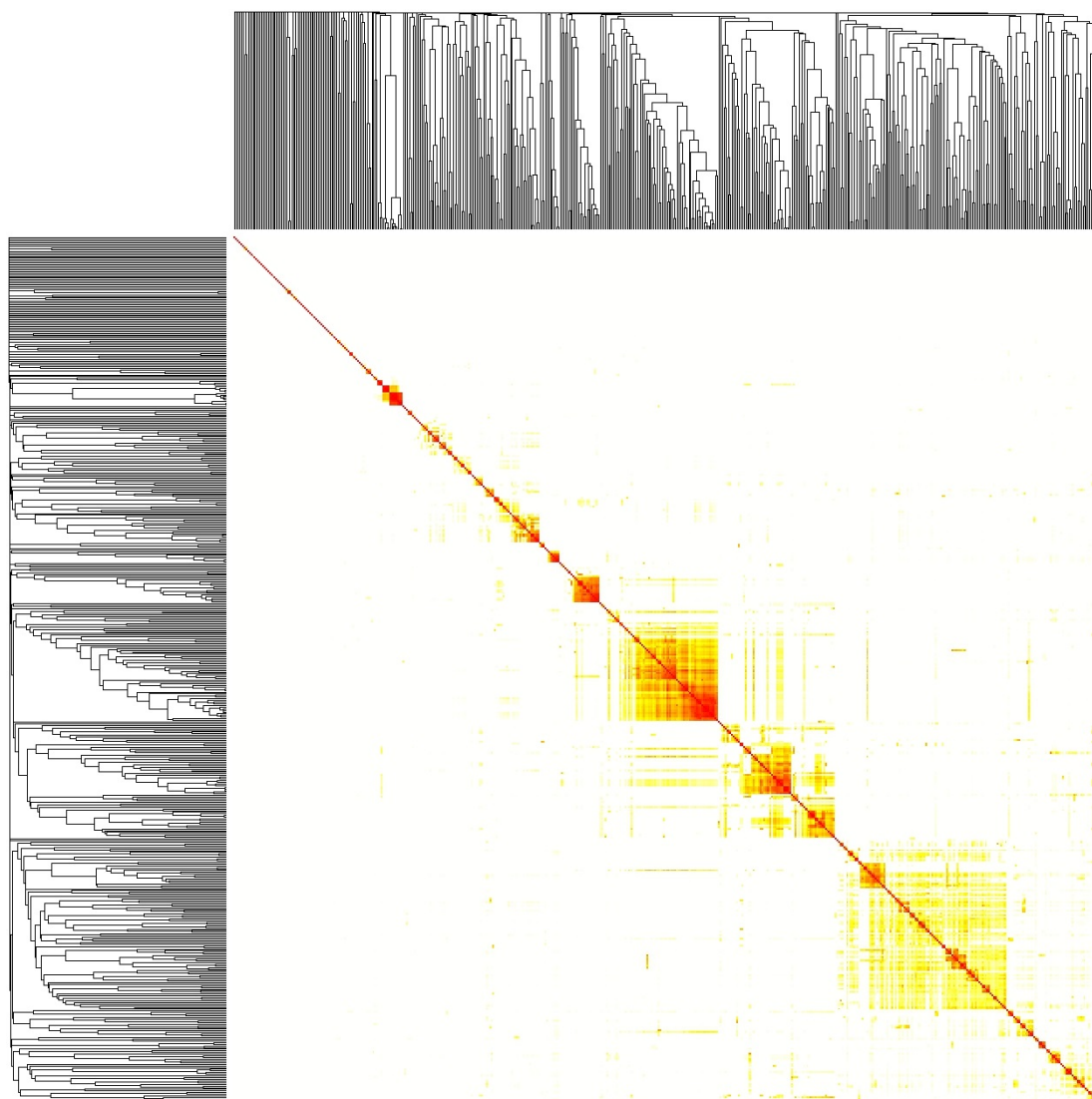
- [69] C. A. Ortori, N. Halliday, M. Cámara, P. Williams, and D. A. Barrett. “LC-MS/MS Quantitative Analysis of Quorum Sensing Signal Molecules”. In: *Methods in Molecular Biology*. Springer New York, 2014, pp. 255–270. DOI: 10.1007/978-1-4939-0473-0\_21.
- [70] C. K. Maurer, A. Steinbach, and R. W. Hartmann. “Development and validation of a UHPLC–MS/MS procedure for quantification of the *Pseudomonas* Quinolone Signal in bacterial culture after acetylation for characterization of new quorum sensing inhibitors”. In: *Journal of Pharmaceutical and Biomedical Analysis* 86 (Dec. 2013), pp. 127–134. DOI: 10.1016/j.jpba.2013.07.047.
- [71] F. G. Holliman. “Pigments of *Pseudomonas* species. Part I. Structure and synthesis of aeruginosin A”. In: *Journal of the Chemical Society C: Organic* 19 (1969), p. 2514. DOI: 10.1039/j39690002514.
- [72] K. Ersmark, J. Del Valle, and S. Hanessian. “Chemistry and Biology of the Aeruginosin Family of Serine Protease Inhibitors”. In: *Angewandte Chemie International Edition* 47.7 (Feb. 2008), pp. 1202–1223. DOI: 10.1002/anie.200605219.
- [73] E. A. Abu, S. Su, L. Sallans, R. E. Boissy, A. Greatens, W. R. Heineman, and D. J. Hassett. “Cyclic voltammetric, fluorescence and biological analysis of purified aeruginosin A, a secreted red pigment of *Pseudomonas aeruginosa* PAO1”. In: *Microbiology* 159.Pt\_8 (June 2013), pp. 1736–1747. DOI: 10.1099/mic.0.065235-0.
- [74] E. G. Ahuja, P. Janning, M. Mentel, A. Graebisch, R. Breinbauer, W. Hiller, B. Costisella, L. S. Thomashow, D. V. Mavrodi, and W. Blankenfeldt. “PhzA/B Catalyzes the Formation of the Tricycle in Phenazine Biosynthesis”. In: *Journal of the American Chemical Society* 130.50 (Dec. 2008), pp. 17053–17061. DOI: 10.1021/ja806325k.

## Supplementary Material

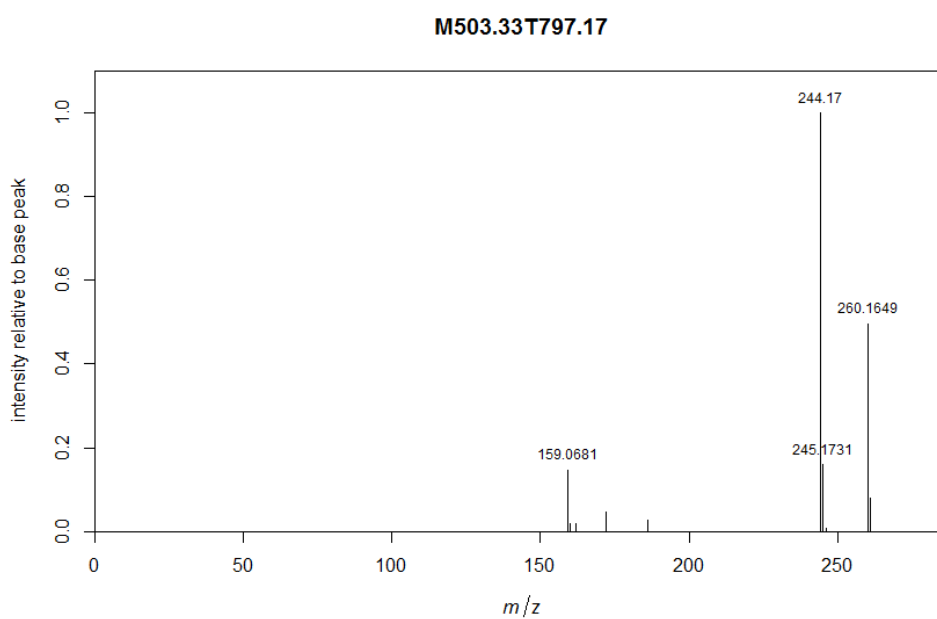
### Supplementary Figures



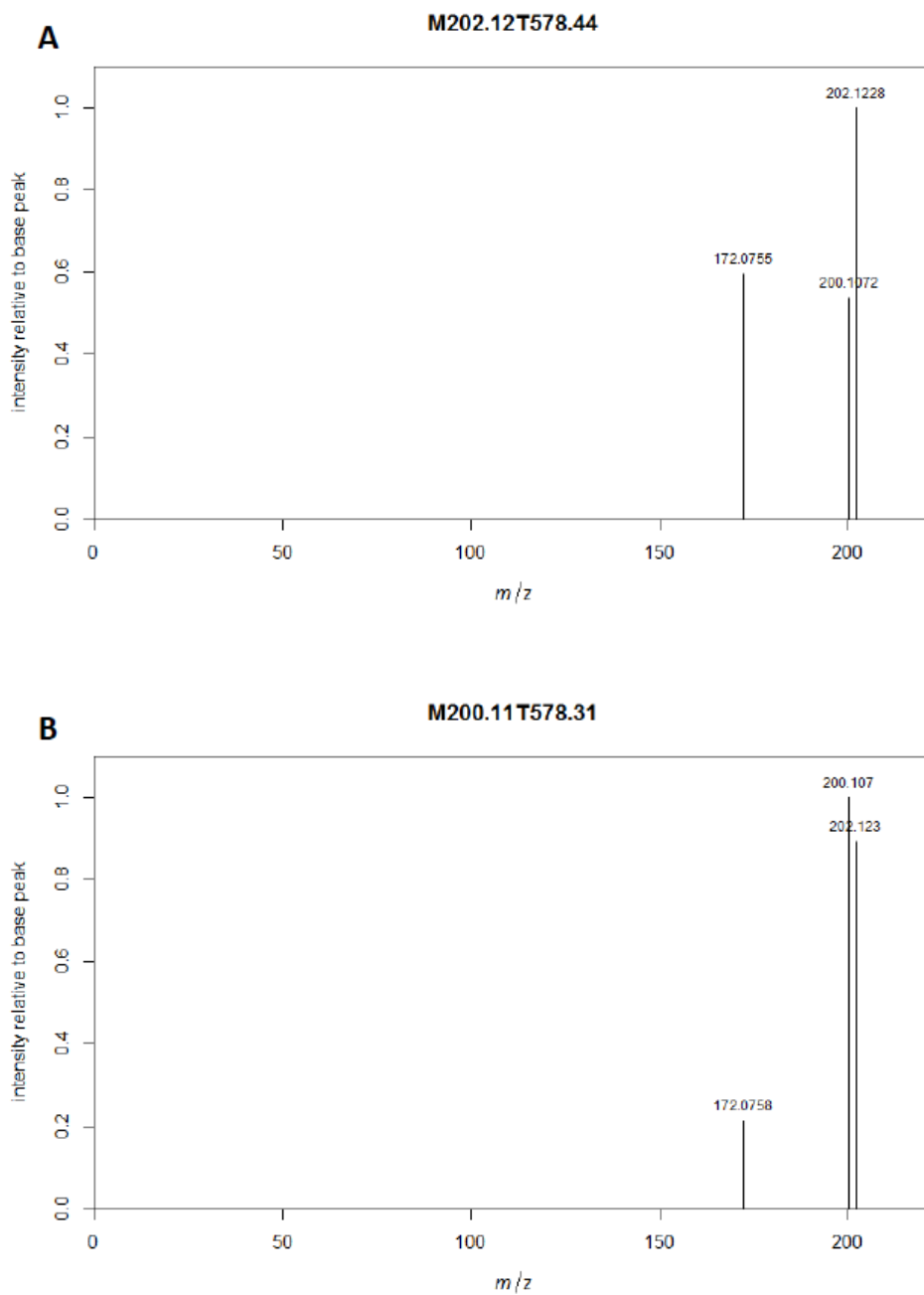
**Figure 3.S1.:** MDS plot displaying cosine differences of 518 MS<sup>2</sup> spectra. Major groups are marked by coloured circles. Orange: predominantly alkyl quinolones. Green: predominantly glutamate-containing metabolites. Red: predominantly peptides and amino acids.



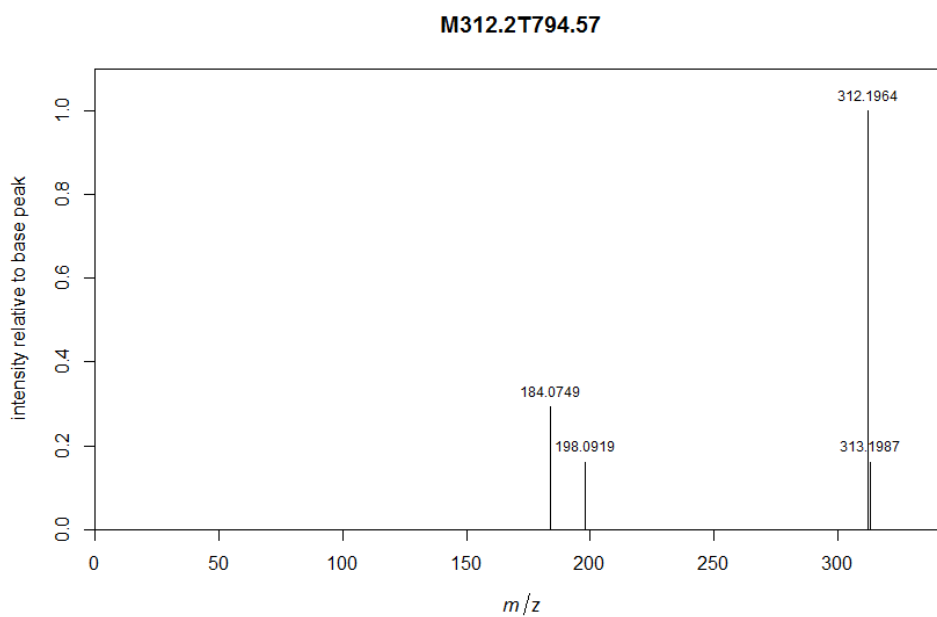
**Figure 3.S2.:** Heatmap displaying cosine similarity between 518 consensus spectra and dendrogram as result of hierarchical clustering with average linkage. Red = high similarity, yellow = intermediate similarity, white = no similarity.



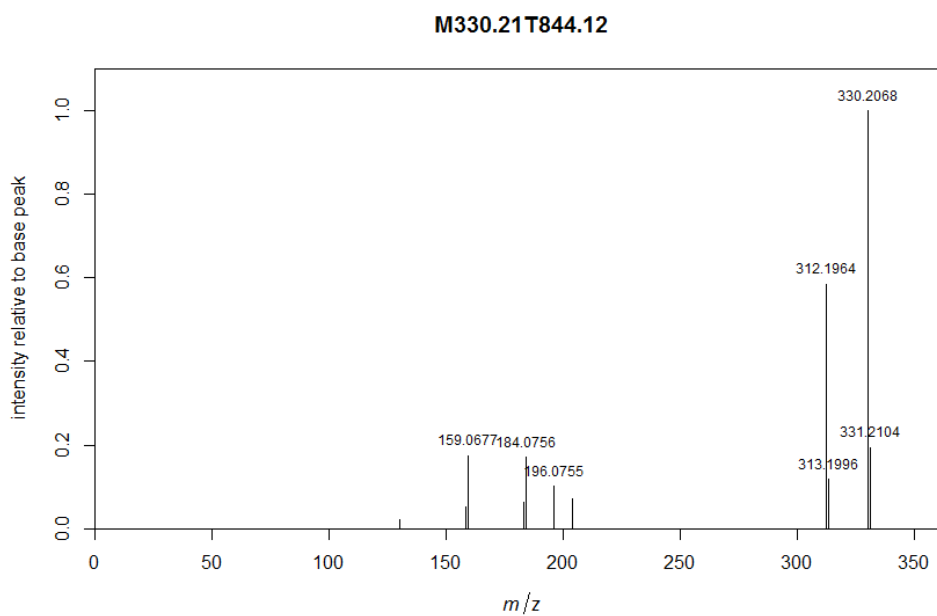
**Figure 3.S3.:** MS<sup>2</sup> spectrum of M503.33T797.17, a putative proton-bound mixed dimer of HHQ and HQNO.



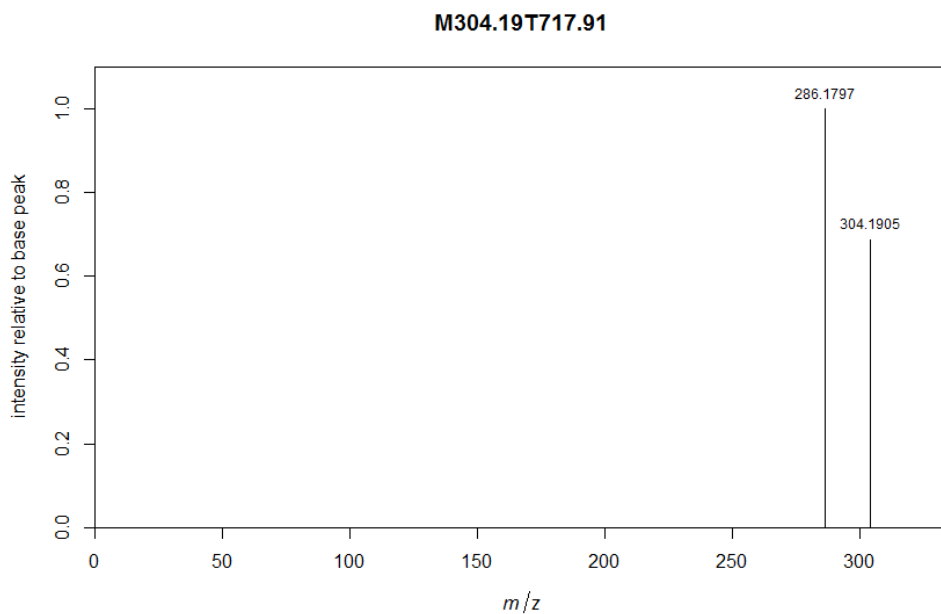
**Figure 3.S4.:** MS<sup>2</sup> spectra of M202.12T578.44 (A) and M200.11T578.31 (B) that have been putatively annotated as C4-HQ and C4:1-HQ.



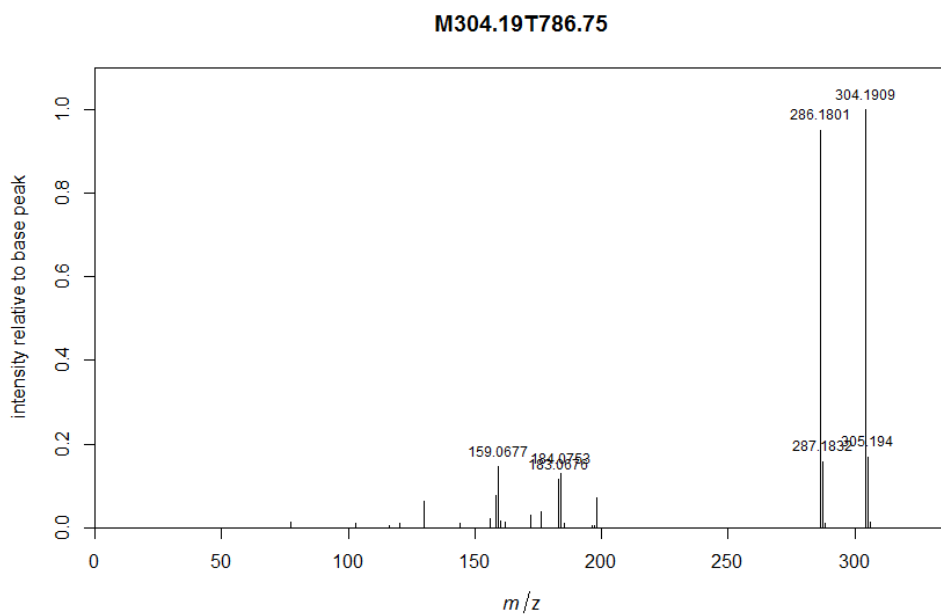
**Figure 3.S5.:** MS<sup>2</sup> spectrum of M312.2T794.57, putative C11:2-QNO.



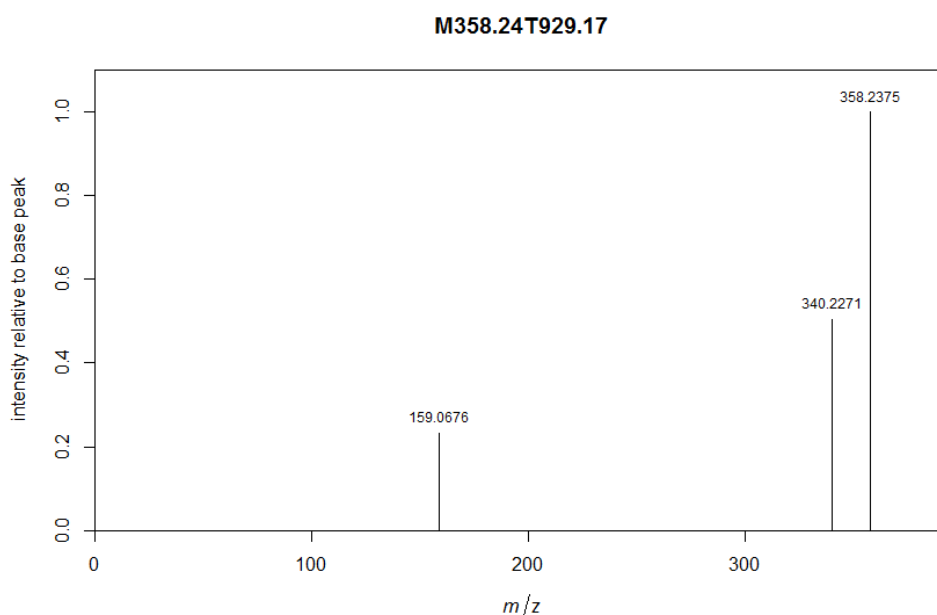
**Figure 3.S6.:** MS<sup>2</sup> spectrum of M330.21T844.12, a putative side chain-oxidation product of C11:1-QNO.



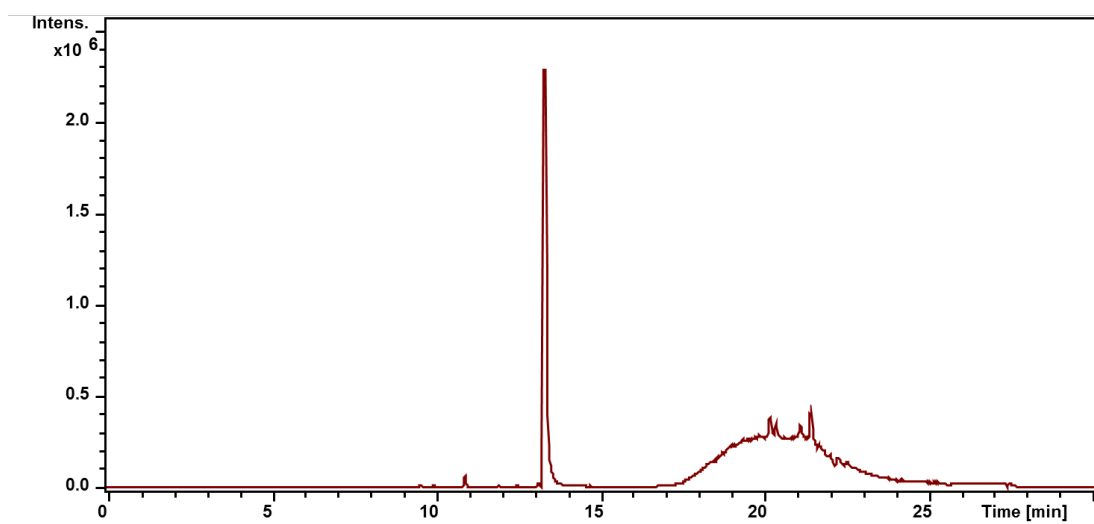
**Figure 3.S7.:** MS<sup>2</sup> spectrum of M304.19T717.91, a putative side chain-oxidation product of C9-QNO.



**Figure 3.S8.:** MS<sup>2</sup> spectrum of M304.19T786.75, another putative side chain-oxidation product of C9-QNO.



**Figure 3.S9.:** MS<sup>2</sup> spectrum of M358.24T929.17, a putative side chain-oxidation product of C13:1-QNO.



**Figure 3.S10.:** Extracted ion chromatogram of  $m/z$  260.164, the exact mass of the  $[M+H]^+$  of both HQNO and PQS. HQNO elutes first and in a good peak shape, while PQS elutes later and fails to produce a defined peak.



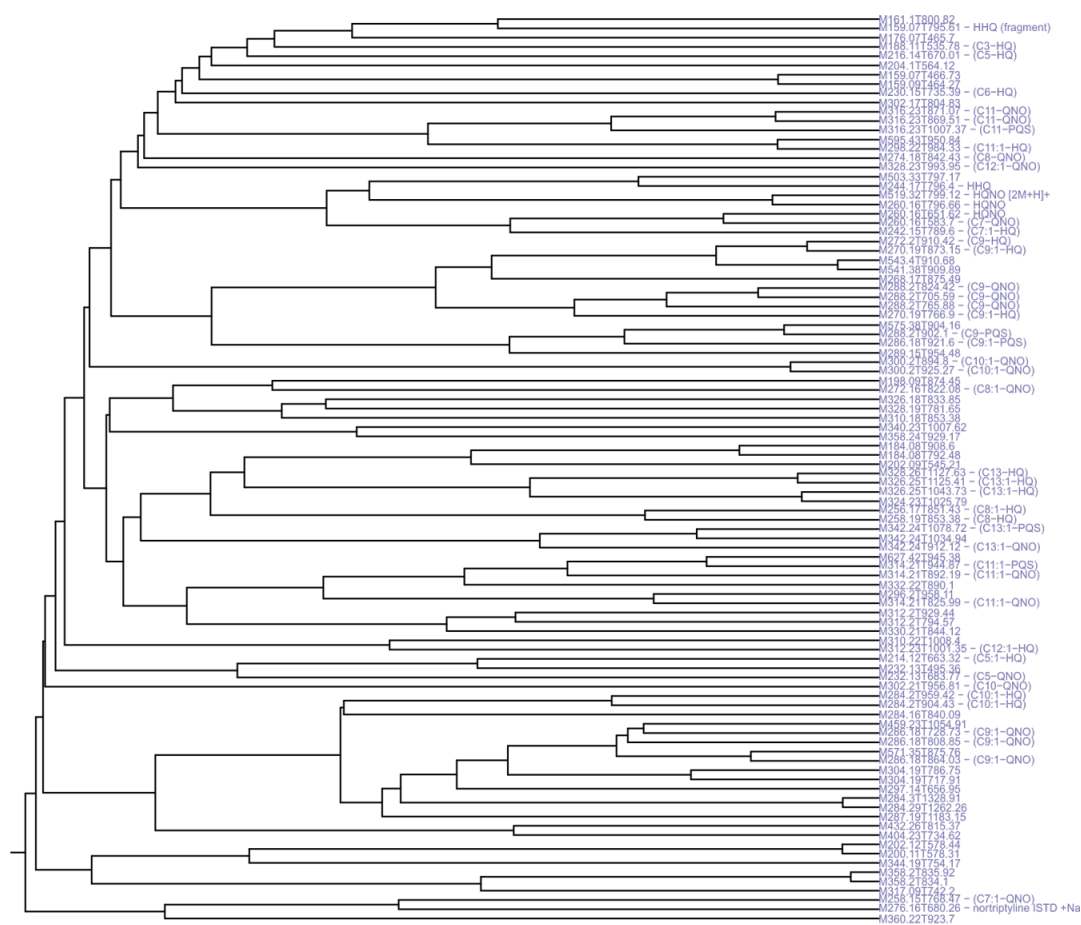


Figure 3.S11.: Cluster 43 containing various alkyl quinolone species.

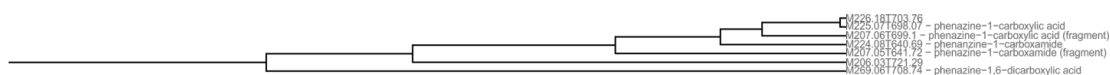


Figure 3.S12.: Cluster 48 from the dendrogram obtained by hierarchical clustering of MS<sup>2</sup> spectra similarities. The Cluster contains the phenazine derivatives that are not *N*-methylated.

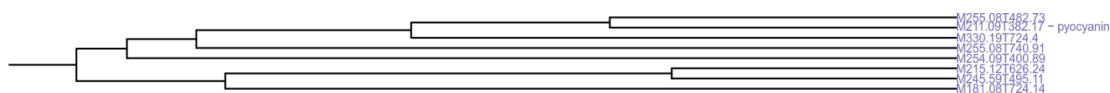
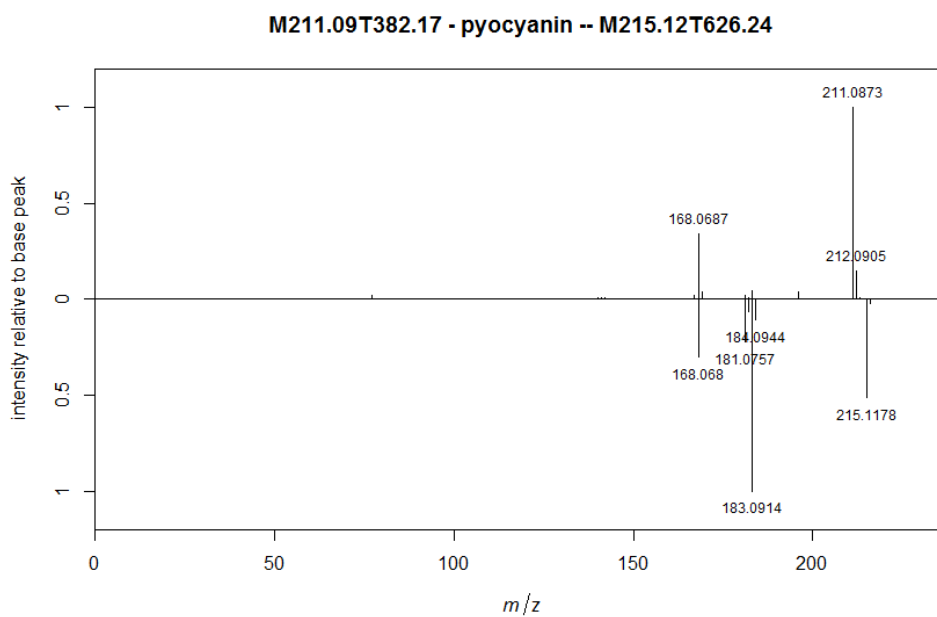
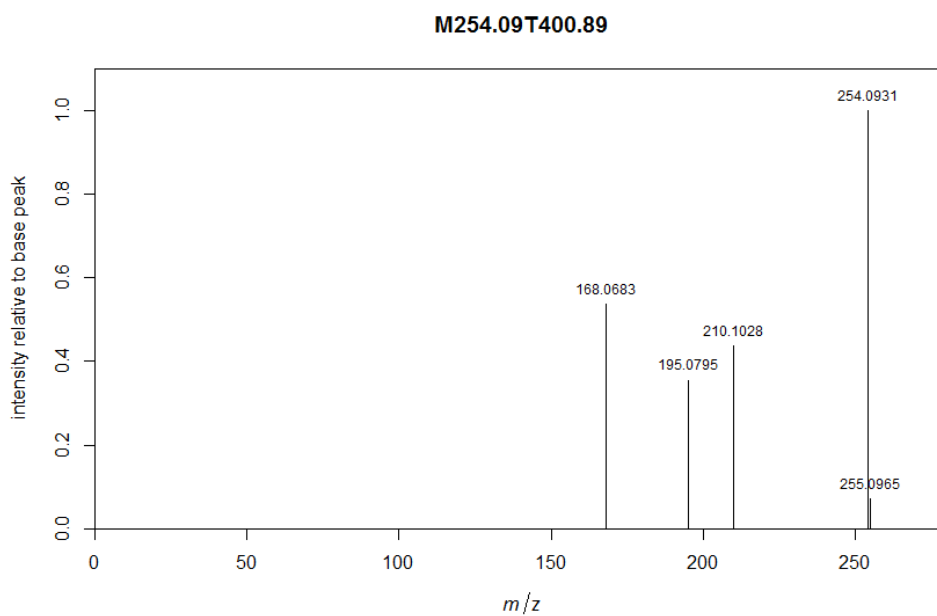


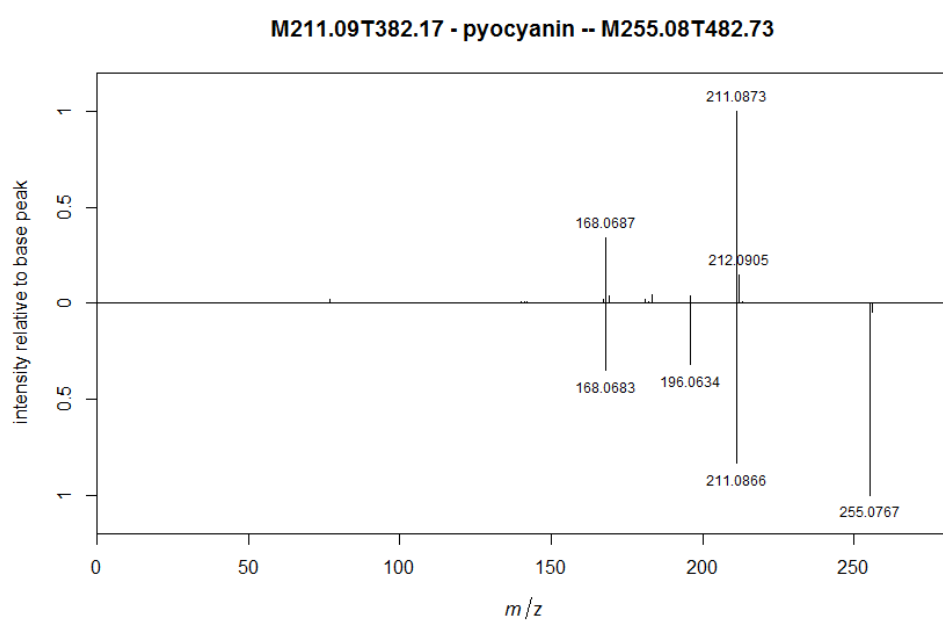
Figure 3.S13.: Cluster 35 from the dendrogram obtained by hierarchical clustering of MS<sup>2</sup> spectra similarities. The Cluster contains the pycocyanin and related phenazine species.



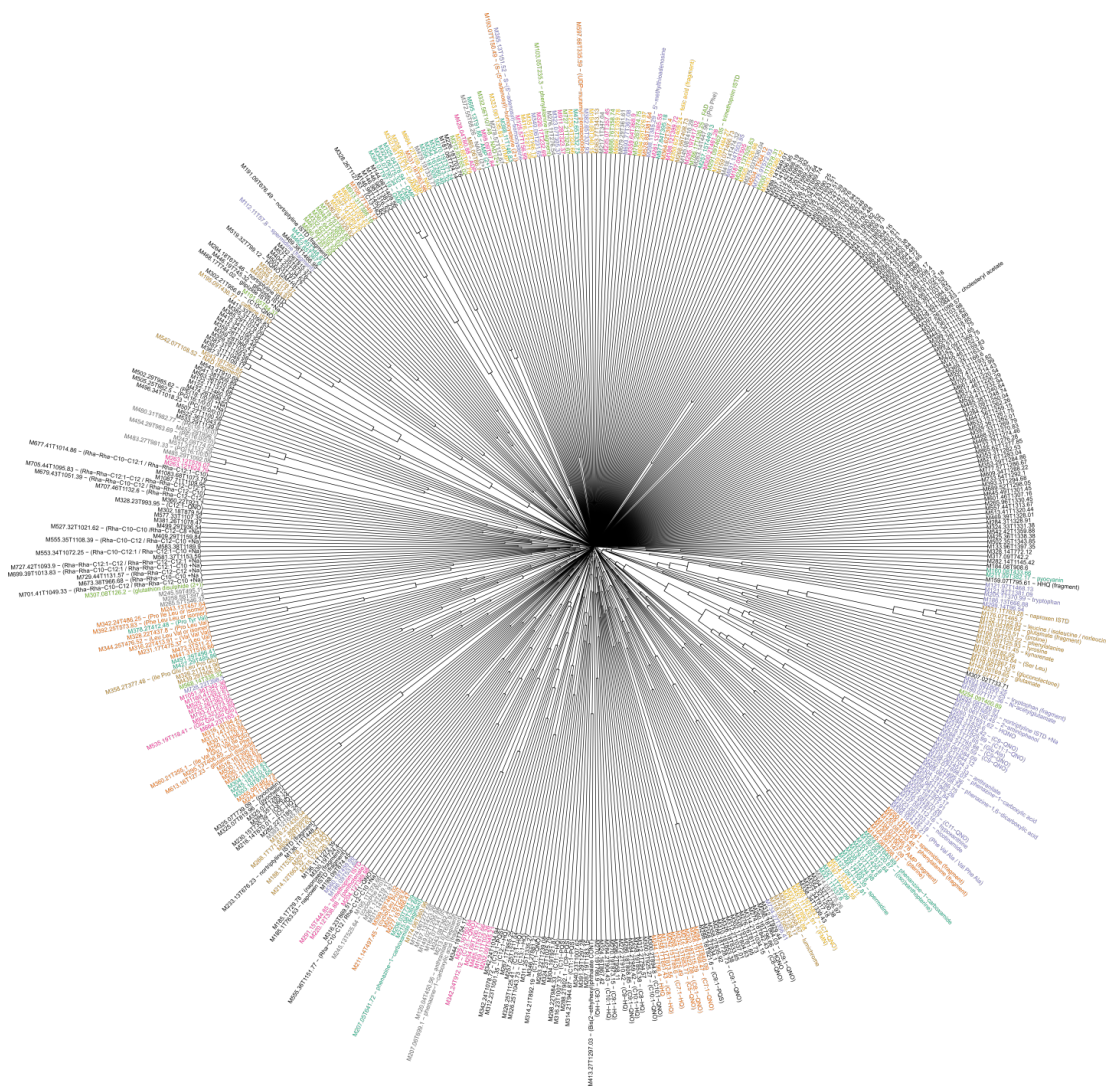
**Figure 3.S14.:** MS<sup>2</sup> spectra of pyocyanin (upper pane) and M215.12T626.24, putatively annotated as tetrahydropyocyanin (lower pane).



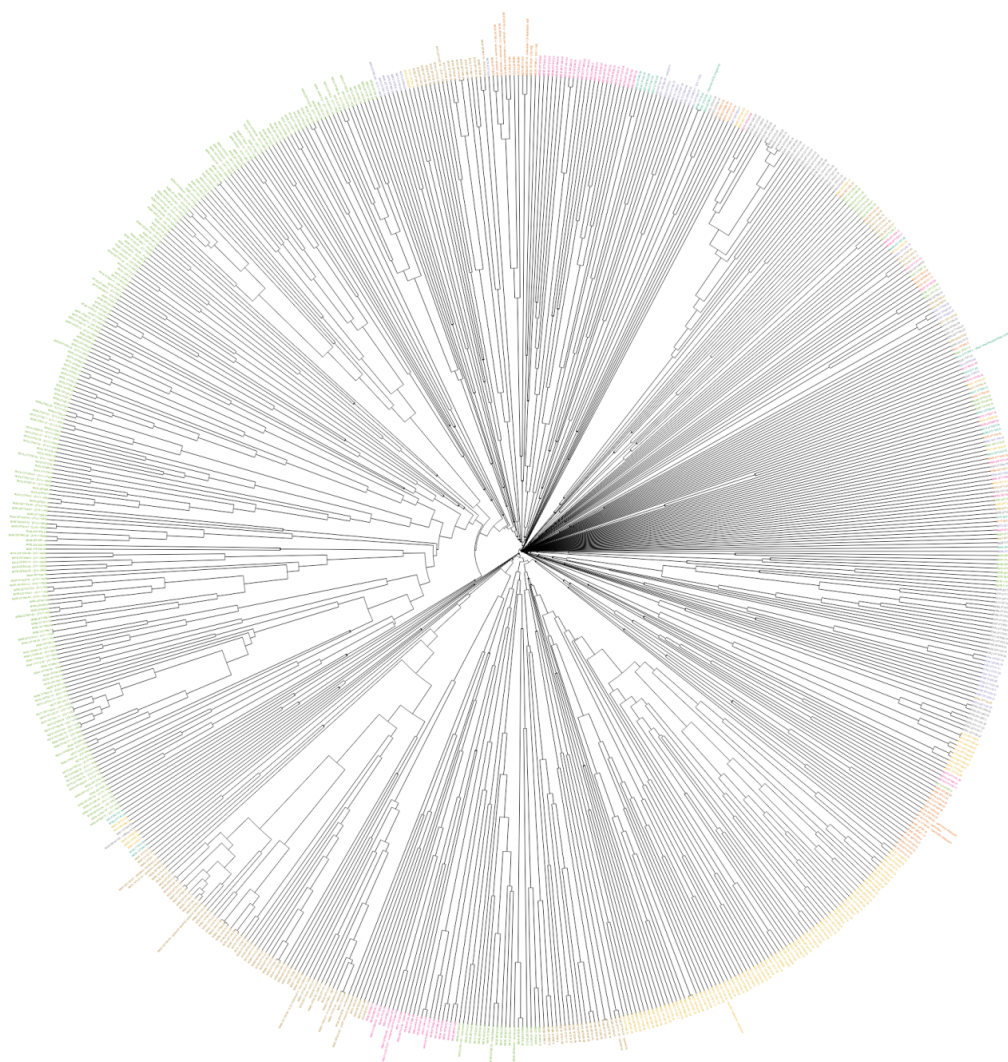
**Figure 3.S15.:** MS<sup>2</sup> spectrum of M254.09T400.89, annotated as aeruginosin A.



**Figure 3.S16.:** MS<sup>2</sup> spectra of pyocyanin (upper pane) and M255.08T482.73, putatively annotated as pyocyanin carboxylic acid (lower pane).



**Figure 3.S17.:** Dendrogram obtained by hierarchical clustering of neutral loss pattern similarities. Feature IDs are colour coded according to the cluster ID they have been assigned to (cf. Supplementary Table 5).



**Figure 3.S18.:** Dendrogram obtained by hierarchical clustering of product ion spectra similarities of the spectra acquired in the alkyl quinolone-biased semi-targeted analysis. Feature IDs are colour coded according to the cluster ID they have been assigned to (cf. Supplementary Table 6).

## Supplementary Tables

For the sake of readability, Supplementary Tables 1.S1 to 1.S7 have been moved to Appendix A.

## Supplementary Computer Code

```
1 # CluMSID 0.1
2
3 library(mzR)
4 library(iontree)
5 library(ape)
6 library(dbscan)
7 library(RColorBrewer)
8
9 ## Open connection to mzXML file
10 aa <- openMSfile("PoolA_R_SE.mzXML")
11
12 ## Count MS2 spectra (QC)
13 mslvl <- c()
14 for (z in 1:length(aa)) {
15   mslvl[z] <- header(aa, z)$msLevel
16 }
17 length(mslvl[mslvl == 2])
18
19 ## Extract MS2 spectra
20 spectra <- list()
21 for (z in 1:length(aa)) {
22   spectra[[z]] <- peaks(aa, z)
23 }
24 ms2log <- mslvl == 2
25 ms2spectra <- spectra[ms2log]
26
27 ## Create list with spectra containing 2 or more peaks
28 vec <- c()
29 for (k in 1:length(ms2spectra)) {
30   vec[k] <- (nrow(ms2spectra[[k]]) >= 2)
31 }
32 ms2spectra2 <- ms2spectra[vec]
33
34 ## Correct uncalibrated precursor masses
35 pmz <- header(aa)$precursorMZ
36 new.pmz <- 0
37 for (i in 2:length(pmz)) {
38   if (pmz[i] == 0) {
39     x <-
40     0
41   } else {
42     if (pmz[(i - 1)] == 0) {
43       x <-
44       peaks(aa, (i - 1))[which.min(abs(pmz[i] - peaks(aa, (i - 1))[, 1])), 1]
45     } else {
46       if (pmz[(i - 2)] == 0) {
47         x <-
```



```

48     peaks(aa, (i - 2))[which.min(abs(pmz[i] - peaks(aa, (i - 2))[, 1])), 1]
49   } else {
50     if (pmz[(i - 3)] == 0) {
51       x <-
52         peaks(aa, (i - 3))[which.min(abs(pmz[i] - peaks(aa, (i - 3))[, 1])), 1]
53     } else {
54       x <- NA
55     }
56   }
57 }
58 }
59 if (x == 0 ||
60     ((abs(x - pmz[i]) / pmz[i]) * 1e06) <= 100) {
61   new.pmz[i] <- x
62 } else {
63   new.pmz[i] <- NA
64 }
65 }
66
67 ## Create a matrix with precursor m/z and retention time for all spectra in ms2list
68 precursor <- cbind(new.pmz, header(aa)$retentionTime)
69 precursor2 <- precursor[ms2log,][vec,]
70
71 ## Exclude everything with RT >25min
72 cutend <- precursor2[, 2] < 25 * 60
73 precursormzrt <- precursor2[cutend,]
74 ms2list <- ms2spectra2[cutend]
75
76 ## Get median m/z and median RT for precursor masses that differ less than 10ppm
77 flist <- list()
78 mz1 <- precursormzrt
79 while (nrow(mz1) >= 1) {
80   l1 <- abs(mz1[1, 1] - mz1[, 1]) <= mz1[1, 1] * 1E-5
81   l2 <- matrix(mz1[c(l1, l1)], ncol = 2)
82   l3 <- diff(l2[, 2])
83   l4 <- c(0, which(l3 > 30), nrow(l2))
84   l5 <- list()
85   for (i in 1:(length(l4) - 1)) {
86     l5[[i]] <- l2[(l4[i] + 1):(l4[i + 1]),]
87   }
88   flist <- append(flist, l5)
89   mz1 <- matrix(mz1[c(!l1, !l1)], ncol = 2)
90 }
91 ### Calculate median m/z and median RT for all 'features'
92 for (i in 1:length(flist)) {
93   if (is.matrix(flist[[i]])) {
94     flist[[i]] <- cbind(flist[[i]],
95                       rep(median(flist[[i]][, 1]), times = nrow(flist[[i]])),
96                       rep(median(flist[[i]][, 2]), times = nrow(flist[[i]])))
97   } else {
98     flist[[i]] <- c(flist[[i]], flist[[i]])
99   }
100 }
101 medmzrt <- c()
102 for (i in 1:length(flist)) {
103   medmzrt <- rbind(medmzrt, flist[[i]])

```

```
104 }
105 medmzrt <-
106   as.data.frame(medmzrt)
107 colnames(medmzrt) <- c("mz", "rt", "med.mz", "med.rt")
108 ##### Create IDs
109 medmzrt$id <- paste("M", round((medmzrt$med.mz), 2),
110                   "T", round((medmzrt$med.rt), 2), sep = "")
111 ### Put back in original order
112 medmzrt <- medmzrt[order(medmzrt$rt),]
113
114 ## Define merging functions
115 mergeTolerance <- function(x, y, tolerance = 1e-5) {
116   mrg <- merge(x, y, by = "V1", all = T)
117   mrg[is.na(mrg)] <- 0
118   i <- 1
119   while (!is.na(mrg[(i + 1), 1])) {
120     if (abs(mrg[i, 1] - mrg[(i + 1), 1]) <= mrg[i, 1] * tolerance) {
121       mrg[i, 1] <- (mrg[i, 1] + mrg[(i + 1), 1]) / 2
122       mrg[i,-1] <- mrg[i,-1] + mrg[(i + 1),-1]
123       mrg <- mrg[-(i + 1),]
124       i <- i + 1
125       colnames(mrg) <-
126         c("V1", 2:ncol(mrg)) #suppresses error warning 'duplicate column names'
127     } else {
128       i <- i + 1
129     }
130   }
131   mrg
132 }
133
134 mergeSpecList <- function(specList, mzmed) {
135   mrgls <- list()
136   for (z in 1:length(specList)) {
137     z0 <- c()
138     for (j in 1:(z - 1)) {
139       z0[j] <- mzmed[z] == mzmed[j]
140     }
141     if (z != 1 & any(z0)) {
142       mrgls[[z]] <- mrgls[[which(z0 == T)[1]]]
143     } else {
144       if (sum(mzmed == mzmed[z]) > 1) {
145         z1 <- as.matrix(Reduce(mergeTolerance,
146                             specList[mzmed == mzmed[z]]))
147         z1[is.na(z1)] <- 0
148         z2 <-
149           cbind(z1[, 1], round((rowSums(z1) - z1[, 1]) / ncol(z1[, -1])))
150         mrgls[[z]] <- z2
151       } else {
152         mrgls[[z]] <- specList[[z]]
153       }
154     }
155   }
156   names(mrgls) <- names(specList)
157   mrgls
158 }
159
```



```

160 ## Merge spectra in the list that fulfill identity criteria
161 names(ms2list) <- medmzrt$id
162 mergedlist <- mergeSpecList(ms2list, medmzrt$id)
163 shortlist <- mergedlist[!duplicated(mergedlist)]
164 shortmzrt <- medmzrt[!duplicated(mergedlist), 3:4]
165
166 ## Make list with neutral loss spectra
167 nllist <- list()
168 for (i in 1:length(shortlist)) {
169   nl <-
170     cbind((shortmzrt[i, 1] - shortlist[[i]][, 1]), shortlist[[i]][, 2])
171   nl <- subset(nl, nl[, 1] >= -(shortmzrt[i, 1] * 1e-5))
172   nllist[[i]] <- nl
173 }
174 names(nllist) <- names(shortlist)
175
176 ## Print precursor m/z and RT from all merged spectra
177 ## and identify in Bruker DataAnalysis
178 write.table(
179   cbind(names(shortlist), shortmzrt),
180   file = "161019metaboident_SE_pre.csv",
181   sep = ",",
182   row.names = F
183 )
184
185 ## Read in manual annotations
186 ident <-
187   read.csv(file = "161022metaboident_SE_post.csv", stringsAsFactors = F)
188 metabonames <- c()
189 for (n in 1:nrow(ident)) {
190   if (is.na(ident[n, 4]) && is.na(ident[n, 5])) {
191     metabonames[n] <- ident[n, 1]
192   }
193   if (!is.na(ident[n, 4])) {
194     metabonames[n] <- paste(ident[n, 1], "_", ident[n, 4], sep = "")
195   }
196   if (is.na(ident[n, 4]) && !is.na(ident[n, 5])) {
197     metabonames[n] <-
198       paste(ident[n, 1], "_", "(", ident[n, 5], ")", sep = "")
199   }
200 }
201 names(shortlist) <- metabonames
202 names(nllist) <- metabonames
203
204 ## Define similarity score
205 cossim <- function(x, y) {
206   colnames(x) <- NULL
207   colnames(y) <- NULL
208   mm <- mergeTolerance(x, y)
209   sum(sqrt(mm[, 2]) * sqrt(mm[, 3])) / (sqrt(sum(mm[, 2])) * sqrt(sum(mm[, 3])))
210 }
211
212 ## Create distance matrix for MS2 spectra
213 distmat <-
214   matrix(nrow = length(shortlist), ncol = length(shortlist))
215 for (m in 1:length(shortlist)) {

```

```
216 for (l in m:length(shortlist)) {
217   if (is.na(distmat[m, l])) {
218     distmat[m, l] <- 1 - cossim(shortlist[[m]], shortlist[[l]])
219     distmat[l, m] <- distmat[m, l]
220   }
221 }
222 }
223 colnames(distmat) <- names(shortlist)
224 row.names(distmat) <- names(shortlist)
225
226 ## Create distance matrix for neutral loss spectra
227 distmat.nl <-
228   matrix(nrow = length(nlalist), ncol = length(nlalist))
229 for (m in 1:length(nlalist)) {
230   for (l in m:length(nlalist)) {
231     if (is.na(distmat.nl[m, l])) {
232       distmat.nl[m, l] <- 1 - cossim(nlalist[[m]], nlalist[[l]])
233       distmat.nl[l, m] <- distmat.nl[m, l]
234     }
235   }
236 }
237 distmat.nl[is.na(distmat.nl)] <- 1
238 colnames(distmat.nl) <- names(nlalist)
239 row.names(distmat.nl) <- names(nlalist)
240
241 ## Multidimensional scaling
242
243 fit <- cmdscale(as.dist(distmat), eig = TRUE, k = 2)
244 x <- fit$points[, 1]
245 y <- fit$points[, 2]
246 pdf(file = "figure_mds.pdf",
247     height = 12,
248     width = 12)
249 plot(
250   x,
251   y,
252   xlab = "Coordinate_1",
253   ylab = "Coordinate_2",
254   type = "p",
255   col = grey(0.3)
256 )
257 dev.off()
258
259 ## Density based clustering using OPTICS
260 opt <-
261   optics(as.dist(distmat),
262         eps = 10000,
263         minPts = 3,
264         search = "dist")
265 opt.nl <-
266   optics(
267     as.dist(distmat.nl),
268     eps = 10000,
269     minPts = 3,
270     search = "dist"
271   )
```

```

272
273 ### Identify clusters by cutting the reachability plot (black is noise)
274 res <- optics_cut(opt, eps_cl = 0.5)
275
276 pdf(file = "20161025_optics_SE.pdf",
277     height = 6,
278     width = 12)
279 plot(res)
280 dev.off()
281
282 clustmat <- NULL
283 for (i in c(1:max(res$cluster), 0)) {
284   x <-
285     cbind(colnames(distmat)[res$cluster == i], rep(i, length(colnames(distmat)[res$
286       cluster == i])))
287   clustmat <- rbind(clustmat, x)
288 }
289 write.csv(clustmat, file = "20161024_dbclust.csv")
290 res.nl <- optics_cut(opt.nl, eps_cl = 0.7)
291
292 ### Create plot
293 opal <- palette()
294 palette(c(opal, rep(c("orange", opal[-1]),10)))
295
296 pdf(file = "figure_optics.pdf",
297     height = 6,
298     width = 12)
299 plot(res)
300 dev.off()
301 palette(opal)
302
303 clustmat.nl <- NULL
304 for (i in c(1:max(res.nl$cluster), 0)) {
305   x <-
306     cbind(colnames(distmat)[res.nl$cluster == i], rep(i, length(colnames(distmat.nl)[
307       res.nl$cluster == i])))
308   clustmat.nl <- rbind(clustmat.nl, x)
309 }
310 write.csv(clustmat.nl, file = "20161024_dbclustnl.csv")
311
312 ### Hierarchical clustering
313 clust <- hclust(as.dist(distmat), method = "average")
314 hclusttree <- cutree(clust, h = 0.95)
315 hclustmat <- cbind(names(hclusttree), hclusttree)
316
317 ### Plot heatmap
318 hm <- heatmap(distmat, Rowv = as.dendrogram(clust), Colv = "Rowv", distfun = NULL, symm
319   = T)
320
321 ### Plot dendrogram (2 different layouts)
322 clr <- brewer.pal(n = 8, name = "Dark2")
323 pdf(file = "figure_hclust_dendrogram.pdf",
324     width = 30,
325     height = 30)
326 plot(
327   as.phylo(clust),

```

```
325 type = "fan",
326 cex = 0.7,
327 tip.color = rep(clr, 16)[hclusttree]
328 )
329 dev.off()
330 pdf(file = "figure_clusters.pdf",
331     width = 10,
332     height = 35)
333 plot(
334   as.phylo(clust),
335   type = "phylo",
336   cex = 0.4,
337   tip.color = rep(clr, 16)[hclusttree]
338 )
339 dev.off()
340
341 write.csv(hclusttree, file = "20161024_hclust.csv")
342
343 clust.nl <- hclust(as.dist(distmat.nl), method = "average")
344 hclusttree.nl <- cutree(clust.nl, h = 0.95)
345 write.csv(hclusttree.nl, file = "20161024_hclustnl.csv")
346
347 ### Plot dendrogram
348 pdf(file = "figure_hclust_dendrogram_nl.pdf",
349     width = 30,
350     height = 30)
351 plot(
352   as.phylo(clust.nl),
353   type = "fan",
354   cex = 0.7,
355   tip.color = rep(clr, 16)[hclusttree.nl]
356 )
357 dev.off()
358
359 close(aa)
360
361 #####
362
363 ## Tools for analysis & interpretation
364
365 ### Print spectrum as table from 'shortlist'
366 print.matrix <- function(m){
367   write.table(format(m, justify="right"),
368             row.names=F, col.names=F, quote=F)
369 }
370 ms2 <- function(x){print.matrix(shortlist[[x]])}
371
372 ### Plot spectrum from 'shortlist'
373 specplot <- function(n, list = shortlist) {
374   plot(x = list[[n]][,1],
375        y = list[[n]][,2] / max(list[[n]][,2]),
376        type = "h",
377        xlim = c(0, (max(list[[n]][, 1]) * 1.1)),
378        xaxs = "i",
379        xlab = expression(italic(m/z)),
380        ylim = c(0, 1.1),
```

```

381     yaxs = "i",
382     ylab = "intensity_relative_to_base_peak",
383     main = names(list[n])
384     text(x = (list[[n]][,1])[(list[[n]][,2] / max(list[[n]][,2])) > 0.1],
385         y = (list[[n]][,2] / max(list[[n]][,2]))[(list[[n]][,2] / max(list[[n]][,2])) >
386           0.1],
387         labels = round((list[[n]][, 1])[(list[[n]][,2] / max(list[[n]][,2])) > 0.1], 4),
388         pos = 3,
389         cex = 0.75)
390 }
391 ### Create mirror plot of two spectra from 'shortlist'
392 specplot2 <- function(n, o, list = shortlist) {
393   plot(
394     x = list[[n]][, 1],
395     y = list[[n]][, 2] / max(list[[n]][, 2]),
396     type = "h",
397     xlim = c(0, (max(c(
398       list[[n]][, 1], list[[o]][, 1]
399     )) * 1.1)),
400     xaxs = "i",
401     xlab = expression(italic(m / z)),
402     ylim = c(-1.2, 1.2),
403     yaxs = "i",
404     yaxt = "n",
405     ylab = "intensity_relative_to_base_peak",
406     main = paste(names(list[n]), "--", names(list[o]))
407   )
408   points(x = list[[o]][, 1],
409         y = -(list[[o]][, 2] / max(list[[o]][, 2])),
410         type = "h")
411   abline(a = 0, b = 0)
412   axis(2,
413        at = seq(-1, 1, 0.5),
414        labels = c(1.0, 0.5, 0.0, 0.5, 1.0))
415   text(
416     x = (list[[n]][, 1])[(list[[n]][, 2] / max(list[[n]][, 2])) > 0.1],
417     y = (list[[n]][, 2] / max(list[[n]][, 2]))[(list[[n]][, 2] / max(list[[n]][, 2])) >
418       0.1],
419     labels = round((list[[n]][, 1])[(list[[n]][, 2] / max(list[[n]][, 2])) > 0.1], 4),
420     pos = 3,
421     cex = 0.75
422   )
423   text(
424     x = (list[[o]][, 1])[(list[[o]][, 2] / max(list[[o]][, 2])) > 0.1],
425     y = -(list[[o]][, 2] / max(list[[o]][, 2]))[(list[[o]][, 2] / max(list[[o]][, 2])) >
426       0.1],
427     labels = round((list[[o]][, 1])[(list[[o]][, 2] / max(list[[o]][, 2])) > 0.1], 4),
428     pos = 1,
429     cex = 0.75
430   )
431 }

```



## 4 | CluMSID: an R package for similarity-based clustering of tandem mass spectra to aid feature annotation in metabolomics

This Chapter has been published as peer-reviewed article in a scientific journal:

T. Depke, R. Franke, and M. Brönstrup. “CluMSID: an R package for similarity-based clustering of tandem mass spectra to aid feature annotation in metabolomics”. In: *Bioinformatics* (Jan. 2019). ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btz005

4

### Abstract

**Summary:** Compound identification is one of the most eminent challenges in the untargeted analysis of complex mixtures of small molecules by mass spectrometry. Similarity of tandem mass spectra can provide valuable information on putative structural similarities between known and unknown analytes and hence aids feature identification in the bioanalytical sciences. We have developed CluMSID (Clustering of MS<sup>2</sup> spectra for metabolite identification), an R package that enables researchers to make use of tandem mass spectra and neutral loss pattern similarities as a part of their metabolite annotation workflow. CluMSID offers functions for all analysis steps from import of raw data to data mining by unsupervised multivariate methods along with respective (interactive) visualisations. A detailed tutorial with example data is provided as supplementary information.

**Availability:** CluMSID is available as R package from <https://github.com/tdepke/CluMSID/>

**Contact:** tobias.depke@helmholtz-hzi.de or mark.broenstrup@helmholtz-hzi.de

**Supplementary information:** Supplementary data are available at *Bioinformatics* online

## 4.1. Introduction

The untargeted analysis of complex mixtures of small molecules using liquid chromatography coupled to (tandem) mass spectrometry (LC-MS(/MS)) has developed into an important technology to study biological systems, from the various applications of metabolomics to natural product research, drug discovery, environmental and forensic sciences. Still, the technique faces various challenges, the most important being metabolite identification [1]. To address this issue, multiple tools have been developed that aid identification by different approaches, mostly relying on computational mass spectrometry [2].

In proteomics and metabolomics, similarity of tandem mass spectra is routinely used to gauge the match score of experimental and library spectra [3]. It is now also established that similarities in tandem mass spectrometry ( $MS^2$ ) fragmentation can hint towards structural relations between analytes [4]. In this paper, we present a customisable open access tool for similarity-based clustering of LC-MS/MS data from data-dependent acquisitions. It enables reproducible analyses and is fully integratable into R pipelines that use e. g. the popular packages from the ‘xcms’ family. The tool is also applicable to flow injection or GC-EI-MS data. However, we will refer to the most frequent experimental type, i. e. LC-MS/MS, in the following. The tool can handle mass spectra following positive and negative ionization from low- and high-resolution mass analyzers. With these assets, CluMSID has the potential to become a valuable extension of the metabolomics data analyst’s toolbox.

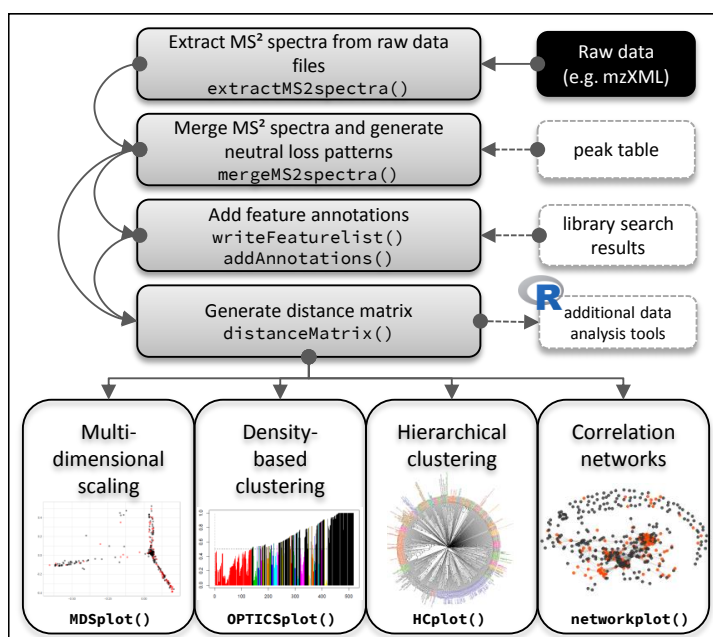
## 4.2. Implementation and main functions

CluMSID is a highly flexible open source tool written in R that combines  $MS^2$  spectral similarity comparisons with several unsupervised data mining methods. The data can be accessed at all stages and custom data analysis steps can easily be integrated. CluMSID is available as R package that can be downloaded from GitHub and used on all platforms.

### 4.2.1. Data import, pre-processing and similarity calculation

For its main functions (Fig. 4.1), CluMSID requires LC-MS/MS data in a standard non-proprietary formats that can be parsed by mzR (<http://github.com/sneumann/mzR>), e. g. mzXML.  $MS^2$  spectra are extracted from the raw file and redundant  $MS^2$  spectra





**Figure 4.1.:** Schematic of a CluMSID workflow. The main functions of CluMSID and their logical order are illustrated. Grey boxes signify data (pre-)processing steps that serve as the basis for multidimensional scaling, clustering, correlation networks or other additional analysis tools. White boxes on the right symbolise optional in- or output, black boxes required input.

are merged into consensus spectra. This process can either operate stand-alone by grouping spectra from the same precursor ion within a retention time interval defined by the user, or spectra can be assigned to peaks picked by a different method or software by means of an external peak table, e. g. generated by ‘xcms’ [5].

During the merging process, neutral loss patterns are generated for all features by subtraction of each fragment  $m/z$  from the precursor  $m/z$ . Neutral losses represent uncharged fragments that also convey structural information comparable to charged fragments. The neutral loss patterns can be processed like MS<sup>2</sup> spectra. Annotations can be added to features that have been identified by the user, e. g. by using online or in-house spectral libraries.

A distance matrix is produced from the list of (annotated) consensus spectra by calculating spectral similarities for every pairwise combination. The similarity measure used by CluMSID is the spectral contrast angle, also known as cosine similarity or cosine score [6]. This distance matrix is the basis for the clustering functions and can be used with non-CluMSID functions to customise the analysis workflow.

### 4.2.2. Mining and visualisation of similarity data

Four data mining methods along with respective (interactive) visualisations are included in CluMSID: multidimensional scaling (MDS), density-based clustering with the OPTICS (Ordering points to identify the clustering structure) algorithm (<https://cran.r-project.org/web/packages/dbscan/>), hierarchical clustering and the generation of correlation networks using the ‘network’ package [7]. In the first study that employed a CluMSID prototype, hierarchical clustering proved the most suitable method to putatively annotate *P. aeruginosa* secondary metabolites [8].

CluMSID’s correlation network functions are particularly useful for researchers who have worked with ‘Molecular Networking’ before but would like to use a more interactive and flexible tool. The MDS plot and the correlation network visualisation can also be generated in an interactive and zoomable version based on the ‘plotly’ package (<https://cran.r-project.org/web/packages/plotly/>).

### 4.2.3. Additional functionalities

CluMSID harbours a set of accessory functions, e. g. to find spectra that contain a specific fragment or neutral loss or to match single spectra against a set of (library) spectra.

Furthermore, MS<sup>1</sup> pseudospectra, i. e. groups of MS<sup>1</sup> peaks derived from the same analyte like adducts, fragments etc., as produced by the ‘CAMERA’ package [9] can be extracted from raw data and analysed like MS<sup>2</sup> spectra.

### 4.2.4. Comparison to existing tools

Similarity-based clustering of LC-MS/MS data from data-independent acquisition can also be performed using the “MetFamily” tool [10] or the more specialised MS/MS data mining tool “MetCirc” [11] while the best known method to study MS<sup>2</sup> spectra from data-dependent acquisition (auto-MS/MS) is “Molecular Networking”, a widely used web tool that however offers relatively little possibilities for customisation [12]. Whereas “Molecular Networking” and “MetFamily” are web applications that can be accessed via a webserver with the need to upload data, we provide a package that does not rely on the public upload of data and offers greater transparency and customisability by enabling access to the data at every step of the analysis. The incorporation of CluMSID in an existing R pipeline that uses for instance ‘xcms’ for peak picking is a usecase which we expect to be very common and useful for the community working

with untargeted metabolomics data. “MetCirc” is an interactive visualisation tool to compare MS<sup>2</sup> experimental data across organisms or tissues, thus it has a different focus than CluMSID. While the distance matrix is calculated in a similar manner, CluMSID features more options for data preprocessing, analysis and visualisation, like hierarchical and density-based clustering. The analysis of pseudospectra and neutral loss patterns are additional, valuable features.

### 4.3. Conclusion

CluMSID is a tool that aids the identification of features in untargeted LC-MS/MS analysis by the use of MS<sup>2</sup> spectral similarity and unsupervised learning methods. It offers functions for a complete and customisable workflow from raw data to visualisations in the form of a freely accessible R package. We are convinced that CluMSID will benefit both the wider metabolomics community and scientists from other bioanalytical fields applying untargeted LC-MS/MS analysis by enabling researchers to integrate MS<sup>2</sup> spectra and neutral loss patterns similarity data into their feature annotation workflow.

### Funding

This work was supported by the President’s Initiative and Networking Funds of the Helmholtz Association of German Research Centres [VH-GS-202]; the EU-funded European Marine Biological Research Infrastructure Cluster [654008]; and by a PhD scholarship of the *Studienstiftung des deutschen Volkes* to T.D.

### References

- [1] C. H. Johnson, J. Ivanisevic, and G. Siuzdak. “Metabolomics: beyond biomarkers and towards mechanisms”. In: *Nature Reviews Molecular Cell Biology* 17.7 (Mar. 2016), pp. 451–459. DOI: 10.1038/nrm.2016.25.
- [2] H. Tsugawa. “Advances in computational metabolomics and databases deepen the understanding of metabolisms”. In: *Current Opinion in Biotechnology* 54 (Dec. 2018), pp. 10–17. DOI: 10.1016/j.copbio.2018.01.008.
- [3] D. L. Tabb, M. J. MacCoss, C. C. Wu, S. D. Anderson, and J. R. Yates. “Similarity among Tandem Mass Spectra from Proteomic Experiments: Detection, Significance, and Utility”. In: *Analytical Chemistry* 75.10 (May 2003), pp. 2470–2477. DOI: 10.1021/ac026424o.
- [4] N. Garg, C. A. Kapon, Y. W. Lim, N. Koyama, M. J. Vermeij, D. Conrad, F. Rohwer, and P. C. Dorrestein. “Mass spectral similarity for untargeted metabolomics data analysis of complex mixtures”. In: *International Journal of Mass Spectrometry* 377 (Feb. 2015), pp. 719–727. DOI: 10.1016/j.ijms.2014.06.005.

- [5] C. A. Smith, E. J. Want, G. O'Maille, R. Abagyan, and G. Siuzdak. "XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification". In: *Analytical Chemistry* 78.3 (Feb. 2006), pp. 779–787. DOI: 10.1021/ac051437y.
- [6] K. X. Wan, I. Vidavsky, and M. L. Gross. "Comparing similar spectra: From similarity index to spectral contrast angle". In: *Journal of the American Society for Mass Spectrometry* 13.1 (Jan. 2002), pp. 85–88. DOI: 10.1016/s1044-0305(01)00327-0.
- [7] C. T. Butts. "network: A Package for Managing Relational Data in R". In: *Journal of Statistical Software* 24.2 (2008). DOI: 10.18637/jss.v024.i02.
- [8] T. Depke, R. Franke, and M. Brönstrup. "Clustering of MS<sup>2</sup> spectra using unsupervised methods to aid the identification of secondary metabolites from *Pseudomonas aeruginosa*". In: *Journal of Chromatography B* 1071 (Dec. 2017), pp. 19–28. DOI: 10.1016/j.jchromb.2017.06.002.
- [9] C. Kuhl, R. Tautenhahn, C. Böttcher, T. R. Larson, and S. Neumann. "CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets". In: *Analytical Chemistry* 84.1 (Dec. 2011), pp. 283–289. DOI: 10.1021/ac202450g.
- [10] H. Treutler, H. Tsugawa, A. Porzel, K. Gorzolka, A. Tissier, S. Neumann, and G. U. Balcke. "Discovering Regulated Metabolite Families in Untargeted Metabolomics Studies". In: *Analytical Chemistry* 88.16 (Aug. 2016), pp. 8082–8090. DOI: 10.1021/acs.analchem.6b01569.
- [11] T. Naake and E. Gaquerel. "MetCirc: navigating mass spectral similarity in high-resolution MS/MS metabolomics data". In: *Bioinformatics* 33.15 (Apr. 2017), pp. 2419–2420. DOI: 10.1093/bioinformatics/btx159.
- [12] M. Wang et al. "Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking". In: *Nature Biotechnology* 34.8 (Aug. 2016), pp. 828–837. DOI: 10.1038/nbt.3597.

## Supplementary information

Detailed tutorial (PDF file).

# CluMSID — Clustering of MS<sup>2</sup> Spectra for Metabolite Identification

A General Tutorial.

*Tobias Depke*

*December 31, 2018*

## Contents

<b>Introduction</b>	<b>A-2</b>
<b>MS2spectrum and pseudospectrum classes</b>	<b>A-2</b>
<b>Extract MS<sup>2</sup> spectra from *.mzXML file</b>	<b>A-3</b>
<b>Merge MS<sup>2</sup> spectra that derive from the same peak/feature</b>	<b>A-5</b>
Merge spectra without external peaktable . . . . .	A-5
Merge spectra with external peaktable, e.g. from XCMS . . . . .	A-6
<b>Add annotations</b>	<b>A-8</b>
Manual procedure . . . . .	A-8
Alternative procedures . . . . .	A-8
<b>Generate distance matrices</b>	<b>A-9</b>
Distance matrix for product ion spectra . . . . .	A-9
Distance matrix for neutral loss patterns . . . . .	A-9
<b>Visualise distance/similarity data using multidimensional scaling (MDS)</b>	<b>A-10</b>
<b>Perform density-based clustering using the OPTICS algorithm</b>	<b>A-12</b>
<b>Perform hierarchical clustering</b>	<b>A-14</b>
Create a heatmap . . . . .	A-14
Create a dendrogram . . . . .	A-16
<b>Generate a correlation network</b>	<b>A-18</b>
<b>Additional functionalities</b>	<b>A-22</b>
Access individual spectra from a list of spectra by various slot entries . . . . .	A-22
Find spectra that contain a specific fragment or neutral loss . . . . .	A-24
Match one spectrum against a set of spectra . . . . .	A-25
Convert MSnbase objects to class MS2spectrum . . . . .	A-27
Split polarities from polarity-switching runs . . . . .	A-28
<b>Use MS<sup>1</sup> pseudospectra instead of or in addition to MS<sup>2</sup> data</b>	<b>A-28</b>
Extract pseudospectra . . . . .	A-28
Create distance matrix for pseudospectra . . . . .	A-29
Generate a correlation network for pseudospectra . . . . .	A-29

## Introduction

This tutorial shows how to use the CluMSID package to help annotate MS<sup>2</sup> spectra from untargeted LC-MS/MS data. CluMSID works with MS<sup>2</sup> data generated by data-dependent acquisition and requires an mzXML file (like in this example) or any other file that can be parsed by mzR, like mzML, mzTab or netCDF, as input. It can be used both stand-alone and together with the XCMS suite of preprocessing tools.

CluMSID extracts and merges MS<sup>2</sup> spectra and generates neutral loss patterns for each feature. Additionally, it can make use of information from the CAMERA package to generate pseudospectra from MS<sup>1</sup> level data. The tool uses cosine similarity to generate distance matrices from MS<sup>2</sup> spectra, neutral loss patterns and pseudospectra.

These distance matrices are the basis for multivariate statistics methods such as multidimensional scaling, density-based clustering, hierarchical clustering and correlation networks. The CluMSID package provides functions for these methods including (interactive) visualisation but the distance/similarity data can also be analysed with other R functions.

For the demonstrations in this tutorial, we will mainly use data from pooled *Pseudomonas aeruginosa* cell extracts, measured in ESI-(+) mode with auto-MS/MS on a Bruker maxis<sup>HD</sup> qTOF after reversed phase separation by UPLC. For details, please refer to the Depke *et al.* 2017 publication (doi: 10.1016/j.jchromb.2017.06.002.).

To be able to access the example data, we also need the related package CluMSIDdata. Both packages are available from Bioconductor, starting from version 3.9, and can be installed as follows:

```
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install(c("CluMSIDdata", "CluMSID"))
```

Before the release of R 3.6 in April/May 2019, the installation from Bioconductor requires the user to install the development versions of R and Bioconductor. For those who wish to avoid working with devel version, R 3.5 compatible versions of CluMSID and CluMSIDdata are available from GitHub and can be installed as follows:

```
if (!requireNamespace("devtools", quietly = TRUE)) install.packages("devtools")
devtools::install_github("tdepke/CluMSIDdata", ref = "pkg")
devtools::install_github("tdepke/CluMSID", ref = "pkg")
```

Once installed, both packages are loaded, along with tidyverse which we will use later.

```
library(ClusMSID)
library(ClusMSIDdata)
library(tidyverse)
```

## MS2spectrum and pseudospectrum classes

CluMSID uses a custom S4 class named MS2spectrum to store spectral information in the following slots:

- **id**: a character string similar to the ID used by XCMSonline or the ID given in a predefined peak list
- **annotation**: a character string containing a user-defined annotation, defaults to empty
- **precursor**: (median)  $m/z$  of the spectrum's precursor ion
- **rt**: (median) retention time of the spectrum's precursor ion
- **polarity**: the polarity with which the spectrum was recorded, either positive or negative
- **spectrum**: the actual MS<sup>2</sup> spectrum as two-column matrix (column 1 is (median)  $m/z$ , column 2 is (median) intensity of the product ions)
- **neutral\_losses**: a neutral loss pattern generated by subtracting the product ion mass-to-charge ratios from the precursor  $m/z$  in a matrix format analogous to the spectrum slot

The pseudospectrum class is very similar but it contains no information on precursor  $m/z$  and therefore no neutral loss pattern, either. By default, the **id** slot contains the "pcgroup" number assigned by CAMERA.

The individual slots of `MS2spectrum` and `pseudospectrum` objects can be accessed via the standard S4 way using `object@slot`, e.g. `object@annotation` or by using an accessor function. These exist for all slots and are called `accessFoo()`, where `Foo` is the slot name (not exactly, though, because Bioconductor does not allow to mix snake\_case and camelCase in function names):

- `accessID(object)`
- `accessAnnotation(object)`
- `accessPrecursor(object)`
- `accessRT(object)`
- `accessPolarity(object)`
- `accessSpectrum(object)`
- `accessNeutralLosses(object)`.

## Extract MS<sup>2</sup> spectra from \*.mzXML file

The first step in the CluMSID workflow is to extract MS<sup>2</sup> spectra from the raw data file (in mzXML format). This is done by the `extractMS2spectra` function which internally uses several functions from the `mzR` package. The function offers the possibility to filter spectra that contain less a defined number of peaks and/or do not fall in a defined retention time window. Setting the `recalibrate_precursor` argument to `TRUE` activates a correction process for uncalibrated precursor *m/z* data that existed in older version of Bruker's Compass Xport (cf. Depke *et al.* 2017). It is not necessary to use it with files generated by other software but does not corrupt the data, either.

*Please be aware that mzR often throws warnings concerning the Rcpp version that can usually be ignored.*

```
ms2list <- extractMS2spectra(system.file("extdata",
                                         "PoolA_R_SE.mzXML",
                                         package = "CluMSIDdata"),
                             min_peaks = 2,
                             recalibrate_precursor = TRUE,
                             RTlims = c(0,25))
```

This operation has now extracted all the MS<sup>2</sup> spectra from the raw data file and stored them in a list. Each list entry is an object of class `MS2spectrum`. The list is quite long because it still contains a lot of spectra that derive from the same chromatographic peak.

```
length(ms2list)
#> [1] 2290
```

In our example, the first two spectra in the list derive from the same peak and thus have the same precursor ion and almost the same retention time.

```
head(ms2list, 4)
#> [[1]]
#> An object of class "MS2spectrum"
#> id:
#> annotation:
#> precursor: 146.1652
#> retention time: 56.266
#> polarity: positive
#> MS2 spectrum with 2 fragment peaks
#> neutral loss pattern with 0 neutral losses
#> [[2]]
#> An object of class "MS2spectrum"
#> id:
#> annotation:
```

```
#> precursor: 146.1653
#> retention time: 57.292
#> polarity: positive
#> MS2 spectrum with 3 fragment peaks
#> neutral loss pattern with 0 neutral losses
#> [[3]]
#> An object of class "MS2spectrum"
#> id:
#> annotation:
#> precursor: 129.1387
#> retention time: 57.545
#> polarity: positive
#> MS2 spectrum with 2 fragment peaks
#> neutral loss pattern with 0 neutral losses
#> [[4]]
#> An object of class "MS2spectrum"
#> id:
#> annotation:
#> precursor: 112.1119
#> retention time: 57.797
#> polarity: positive
#> MS2 spectrum with 2 fragment peaks
#> neutral loss pattern with 0 neutral losses
```

From the output above, you also see that the `MS2spectrum` class has a `show()` generic that summarises the MS<sup>2</sup> spectrum and neutral loss pattern data. To show the default output, use `showDefault()`. Be aware that neutral loss patterns have not been calculated in this step.

```
showDefault(ms2list[[2]])
#> An object of class "MS2spectrum"
#> Slot "id":
#> character(0)
#>
#> Slot "annotation":
#> character(0)
#>
#> Slot "precursor":
#> [1] 146.1653
#>
#> Slot "rt":
#> [1] 57.292
#>
#> Slot "polarity":
#> [1] "positive"
#>
#> Slot "spectrum":
#>      [,1] [,2]
#> [1,] 72.08064 2448
#> [2,] 84.08077 328
#> [3,] 112.11228 843
#>
#> Slot "neutral_losses":
#> <0 x 0 matrix>
```



## Merge MS<sup>2</sup> spectra that derive from the same peak/feature

To reduce the amount of redundant MS<sup>2</sup> spectra, the `mergeMS2spectra()` function is used to generate consensus spectra from the MS<sup>2</sup> spectra that derive from the same precursor. `CluMSID` offers two possibilities to do so:

### Merge spectra without external peaktable

This possibility is the standard method for stand-alone use of `CluMSID` and is equivalent to what has been described in Depke *et al.* 2017. It does not need additional input and summarises consecutive spectra that have the same precursor *m/z* if their retention time fall within a defined threshold (`rt_tolerance`, defaults to 30s). A retention time difference between consecutive spectra larger than `rt_tolerance` is interpreted as chromatographic separation and respective spectra will be assigned to a new feature. The `mz_tolerance` argument should be set according to your instruments *m/z* precision, the default is  $1 * 10^{-5}$  (10ppm, equivalent to  $\pm 5$ ppm instrument precision). The `peaktable` and `exclude_unmatched` arguments are not used in this method and are to be left at their default.

```
featlist <- mergeMS2spectra(ms2list)

length(featlist)
#> [1] 518

head(featlist, 4)
#> [[1]]
#> An object of class "MS2spectrum"
#> id: M146.17T59.35
#> annotation:
#> precursor: 146.1653
#> retention time: 59.35
#> polarity: positive
#> MS2 spectrum with 8 fragment peaks
#> neutral loss pattern with 7 neutral losses
#> [[2]]
#> An object of class "MS2spectrum"
#> id: M129.14T58.57
#> annotation:
#> precursor: 129.1387
#> retention time: 58.57
#> polarity: positive
#> MS2 spectrum with 4 fragment peaks
#> neutral loss pattern with 3 neutral losses
#> [[3]]
#> An object of class "MS2spectrum"
#> id: M112.11T57.8
#> annotation:
#> precursor: 112.1119
#> retention time: 57.8
#> polarity: positive
#> MS2 spectrum with 2 fragment peaks
#> neutral loss pattern with 1 neutral losses
#> [[4]]
#> An object of class "MS2spectrum"
#> id: M251.16T60.64
#> annotation:
```

```
#> precursor: 251.1603
#> retention time: 60.64
#> polarity: positive
#> MS2 spectrum with 9 fragment peaks
#> neutral loss pattern with 8 neutral losses
```

The total amount of spectra was reduced from 2290 to 518 and as many other, the redundant spectra #1 and #2 in the raw list are now merged to one consensus spectrum (#1 in the merged list).

In this step, neutral loss patterns have been generated that look like this:

```
accessNeutralLosses(featlist[[1]])
#>      [,1] [,2]
#> [1,] 74.08475 6429
#> [2,] 73.08163 262
#> [3,] 71.07394 239
#> [4,] 62.08476 1044
#> [5,] 34.05341 2363
#> [6,] 33.05024 144
#> [7,] 17.02688 852
```

## Merge spectra with external peaktable, e.g. from XCMS

The second possibility is to supply a peaktable, i.e. a list of picked peaks with their mass-to-charge ratios and retention times. This is particularly useful if you want to annotate a complete metabolomics data set. In our example, we have a metabolomics dataset called "TD035" in which we have measured a range of samples in MS<sup>1</sup> mode for relative quantification. Additionally, we have measured a pooled QC sample in MS<sup>2</sup> mode for annotation. The MS<sup>1</sup> data were analysed using XCMSonline and we want to group the MS<sup>2</sup> spectra so that they match the XCMSonline peak picking.

The spectra are extracted as shown above:

```
ms2list2 <- extractMS2spectra(system.file("extdata",
                                         "TD035-PoolMSMS2.mzXML",
                                         package = "CluMSIDdata"),
                             min_peaks = 2,
                             recalibrate_precursor = TRUE,
                             RTlims = c(0,25))
```

The peaklist is imported from the XCMSonline output. The list has to contain at least 3 columns:

- column 1: name/identifier of the feature
- column 2: *m/z*
- column 3: retention time

Shown below is an easy way of getting from an XCMSonline annotated diffreport to a suitable peaktable using tidyverse functions. Of course, you can achieve the same goal with base R functions or even in Excel. Depending on the retention time format in your \*.mzXML file, you might have to convert from minutes to seconds or vice versa. Here, we have minutes in the XCMSonline output but seconds in the MS<sup>2</sup> file, so we multiply by 60.

```
phtable <- read_delim(file = system.file("extdata",
                                         "TD035_XCMS.annotated.diffreport.tsv",
                                         package = "CluMSIDdata"),
                     delim = "\t") %>%
  select(c(name, mzmed, rtmed)) %>%
  mutate(rtmed = rtmed * 60)
```

```
head(ptable)
#> # A tibble: 6 x 3
#>   name      mzmed  rtmed
#>   <chr>    <dbl> <dbl>
#> 1 M245T2    245.   100.
#> 2 M440T2_1 440.   107.
#> 3 M578T2    578.   104.
#> 4 M85T1     85.0   60.8
#> 5 M126T1_1 126.    61.0
#> 6 M688T24   688.  1468.
```

We can now use this peaktable as an argument for `mergeMS2spectra()`. You can choose whether you want to keep or exclude MS<sup>2</sup> spectra that do not match any peak in the peaktable. These can occur in regions of the chromatogram where there are no clear peaks but the auto-MS/MS still fragments the most abundant ions. These unmatched spectra are merged following the same rules as described above (method without peaktable). In this example, we keep the unmatched spectra. We use the default values for *m/z* and retention time tolerance and thus do not need to specify them.

```
featlist2 <- mergeMS2spectra(ms2list2,
                             peaktable = ptable,
                             exclude_unmatched = FALSE)
```

```
head(featlist2, 4)
#> [[1]]
#> An object of class "MS2spectrum"
#> id: M213T0
#> annotation:
#> precursor: 213.1462
#> retention time: 6.04
#> polarity: positive
#> MS2 spectrum with 5 fragment peaks
#> neutral loss pattern with 3 neutral losses
#> [[2]]
#> An object of class "MS2spectrum"
#> id: xM158T31.17
#> annotation:
#> precursor: 158.0027
#> retention time: 31.17
#> polarity: positive
#> MS2 spectrum with 3 fragment peaks
#> neutral loss pattern with 3 neutral losses
#> [[3]]
#> An object of class "MS2spectrum"
#> id: M146T1_3
#> annotation:
#> precursor: 146.1650
#> retention time: 61.15
#> polarity: positive
#> MS2 spectrum with 7 fragment peaks
#> neutral loss pattern with 6 neutral losses
#> [[4]]
#> An object of class "MS2spectrum"
#> id: M129T1_4
#> annotation:
```

```
#> precursor: 129.1384  
#> retention time: 60.74  
#> polarity: positive  
#> MS2 spectrum with 2 fragment peaks  
#> neutral loss pattern with 2 neutral losses
```

Note that the 2<sup>nd</sup> entry in `featlist2` is marked with an 'x' which means that it could not be assigned to a feature in the `peaktable`.

For the sake of simplicity, only the data generated from the stand-alone procedure will be used for the following examples. Be assured that all of them would also work with the data generated with the help of an external `peaktable` (`featlist2`).

## Add annotations

The next step is to add (external) annotations to the list of features, e.g. from a spectral library that you curate in-house or one that has been supplied by your instrument manufacturer. If you do not (want to) annotate your features at all, this step can be skipped completely, leaving the annotation slot of the `MS2spectrum` objects empty.

### Manual procedure

CluMSID offers several possibilities to add annotations to your feature list. The most basic one first generates a list of features and saves it as \*.csv file. For that you use the `writeFeaturelist()` function and only have to specify your list of spectra and a file name for the output file (here: `pre_anno.csv`). You can then manually fill in your annotations in a new column in the table, save it (in this example under the name `post_anno.csv`) and reload it to R:

```
writeFeaturelist(featlist, "pre_anno.csv")  
  
annotatedSpeclist <- addAnnotations(featlist,  
                                   system.file("extdata",  
                                               "post_anno.csv",  
                                               package = "CluMSIDdata"))
```

`annotatedSpeclist` will then be equivalent to `featlist` with annotations added to the annotation slot of the list entries.

### Alternative procedures

You can add annotations without leaving the R environment, too. `addAnnotations()` also accepts objects of class `data.frame` as `annolist` argument. Be aware that `addAnnotations()` assigns the annotation based on the position in the feature list. I.e., if the order of the features in your list of features (`featlist`) and your list of annotations (`annolist`) is different, you will get nonsense results.

The savest ways to `addAnnotations()` with a `data.frame` is to use `Featurelist()` to generate a `data.frame` that is formatted in the same way as the file output from `writeFeaturelist()` and then match your identifications against this `data.frame` and use the result as argument for `addAnnotations()`.

Say you have an object called `annos` that contains feature IDs (the same as in `featlist`) and annotations in a two-column `data.frame` with "id" and "annotation" as column names. It could look like this:

```
str(annos)
#> 'data.frame': 154 obs. of 2 variables:
#> $ id : chr "M146.17T59.35" "M129.14T58.57" "M112.11T57.8" "M148.06T69.65" ...
#> $ annotation: chr "spermidine" "spermidine (fragment)" "spermidine (fragment)" "glutamate"
head(annos)
#>      id      annotation
#> 1 M146.17T59.35 spermidine
#> 2 M129.14T58.57 spermidine (fragment)
#> 3 M112.11T57.8 spermidine (fragment)
#> 4 M148.06T69.65 glutamate
#> 5 M130.05T69.64 glutamate (fragment)
#> 6 M179.06T71.32 gluconolactone
```

`addAnnotations(featlist, annos, annotationColumn = 2)` will throw an error because `featlist` and `annos` are of different length. Instead, you need to do the following:

```
fl <- featureList(featlist)

fl_annos <- dplyr::left_join(fl, annos, by = "id")
```

Now, you can annotate your list of spectra using `addAnnotations(featlist, fl_annos, annotationColumn = 4)`.

An analogous procedure works if you have your annotations stored in a peaktable that you have used for `mergeMSspectra()`. As the order of spectra in the list will not be same as the order of features in your peaktable, you need to do a matching with the output of `featureList()` as well.

## Generate distance matrices

Once we have a list of `MS2spectrum` objects containing all the required information with or without annotation, we can generate distance matrices from (product ion)  $MS^2$  spectra as well as from neutral loss patterns. These distance matrices serve as the basis for further analysis of the data. Both for  $MS^2$  spectra and neutral loss patterns, cosine similarity is used as similarity metric:

$$\cos(\theta) = \frac{\sum_i a_i \cdot b_i}{\sqrt{\sum_i a_i^2 \cdot \sum_i b_i^2}}$$

### Distance matrix for product ion spectra

For most applications, analysing the similarity of product ion  $MS^2$  spectra will be most useful. The generation of the distance matrix is done by just one simple command but it can take some time to calculate.

```
distmat <- distanceMatrix(annotatedSpecList)
```

### Distance matrix for neutral loss patterns

Common neutral losses and neutral loss patterns can convey information about structural similarity, as well, e.g. with nucleotides or glycosylated secondary metabolites. `CluMSID` offers the possibility to study neutral loss patterns independently from product ion spectra. The generation of a distance matrix is analogous, you just need to set the `type` argument to `"neutral_losses"`:

```
nlmat <- distanceMatrix(annotatedSpeclist, type = "neutral_losses")
```

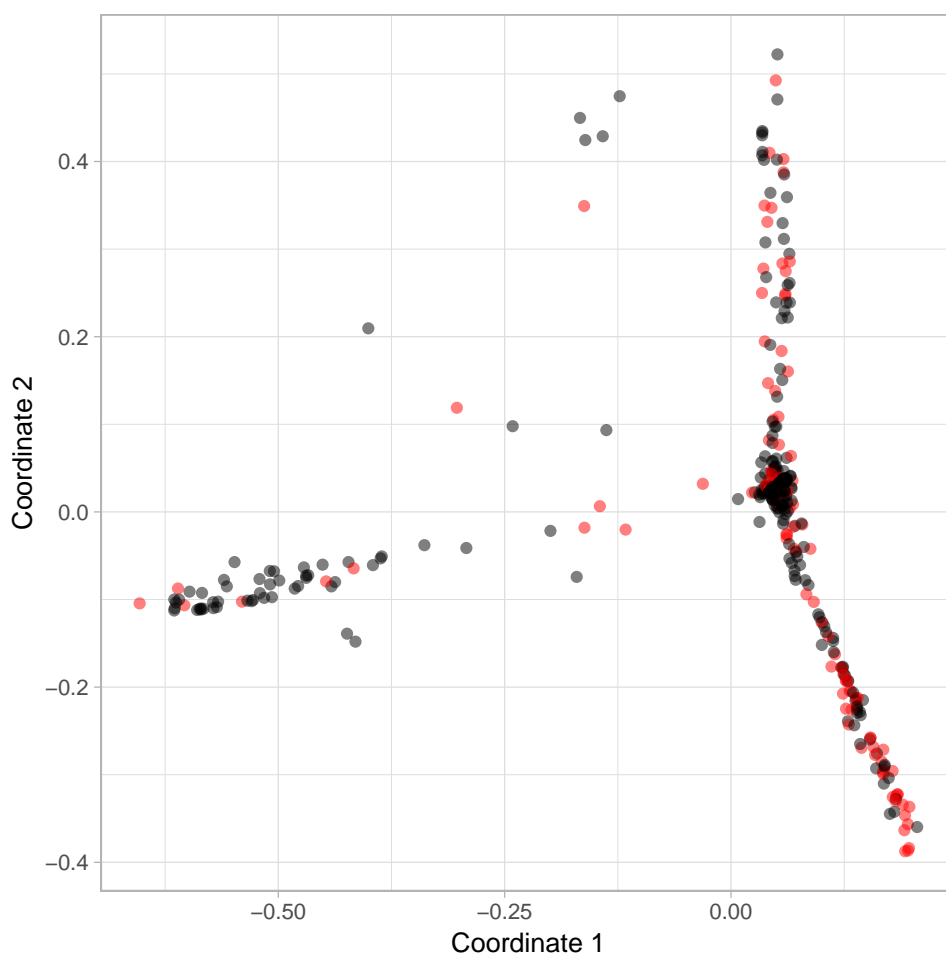
## Visualise distance/similarity data using multidimensional scaling (MDS)

One rather simple possibility to visually analyse the spectral similarity data is multidimensional scaling, a dimension reduction method that simplifies distances in  $n$ -dimensional space to those in two-dimensional space ( $n$  in this case being the number of consensus spectra or neutral loss patterns that were used to generate the distance matrix in the previous step). CluMSID offers a simple function to produce an MDS plot from the distance matrix with the option to highlight annotated metabolites and the possibility to generate an interactive plot using `plotly`.

Standard MDS plots are generated as follows:

For MS<sup>2</sup> spectra:

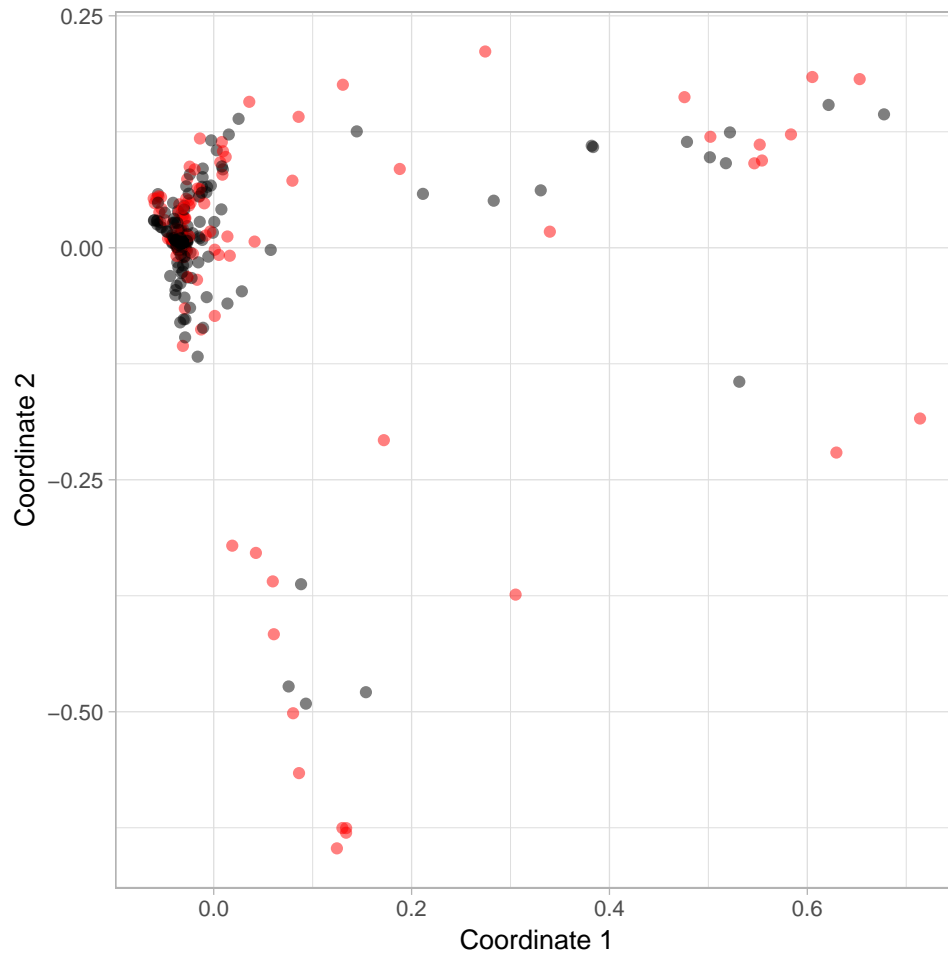
```
MDSplot(distmat, highlight_annotated = TRUE)
```



**Figure A-1:** Multidimensional scaling plot as a visualisation of MS<sup>2</sup> spectra similarities of the example data set. Red dots signify annotated spectra, black dots spectra from unknown metabolites.

For neutral loss patterns:

```
MDSplot(nlmat, highlight_annotated = TRUE)
```

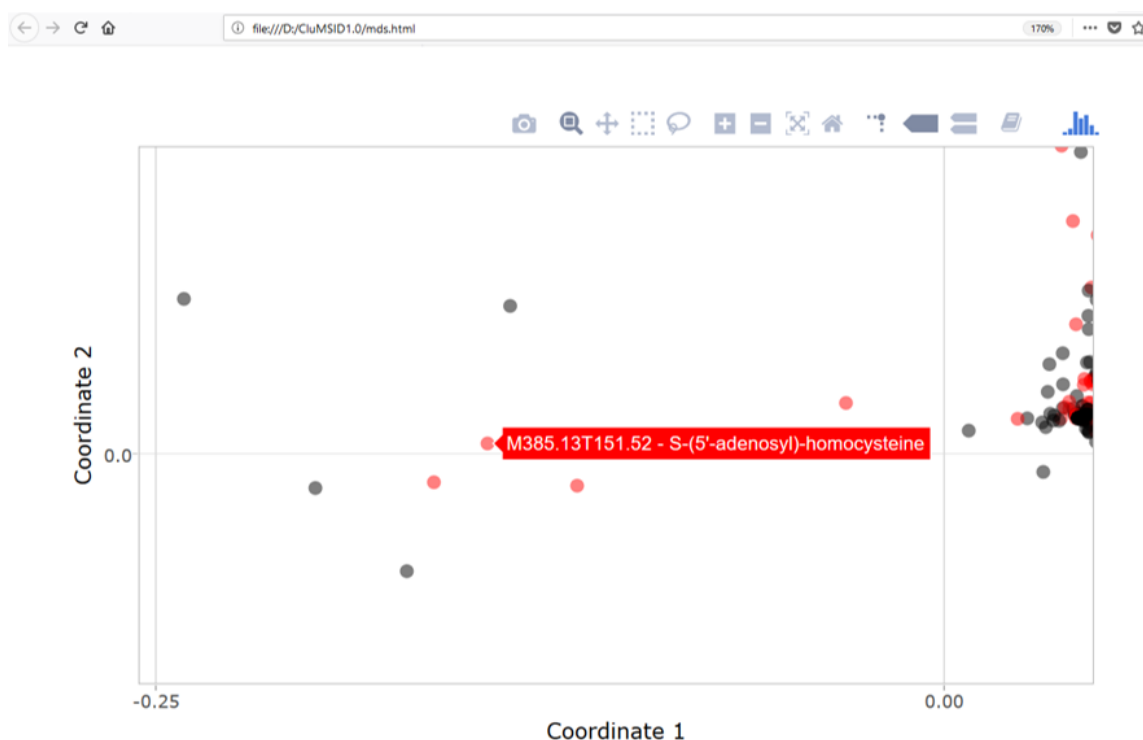


**Figure A-2:** Multidimensional scaling plot as a visualisation of neutral loss similarities of the example data set. Red dots signify annotated spectra, black dots spectra from unknown metabolites.

Interactive plots are zoomable and show feature names upon mouse-over. They are generated like normal MDS plots and can be viewed within RStudio or—after saving as html file using `htmlwidgets`—displayed in a normal web browser.

```
my_mds <- MDSplot(distmat, interactive = TRUE, highlight_annotated = TRUE)  
htmlwidgets::saveWidget(my_mds, "mds.html")
```

This is how it looks like if you open the html file in Firefox and mouse over a feature:



**Figure A-3:** Screenshot of the interactive version of the Multidimensional scaling plot visualising  $MS^2$  spectra similarities of the example data set (cf Figure 1). Zoomed image section with tooltip displaying feature information upon mouse-over.

## Perform density-based clustering using the OPTICS algorithm

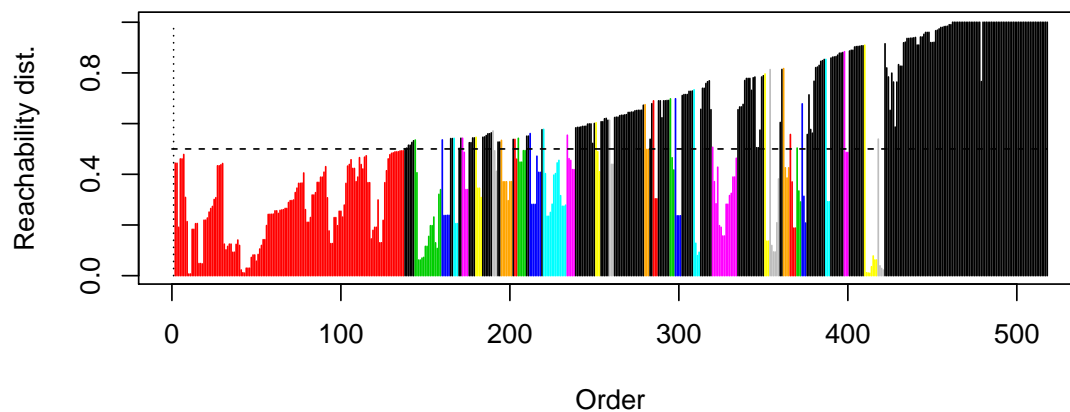
For density-based clustering with CluMSID, the 'OPTICS' algorithm and its implementation in the `dbscan` package is used. Density-based clustering is a useful clustering method that often yields different results than hierarchical clustering and can thus provide additional insight into the data. CluMSID has two functions to perform density-based clustering, one for the reachability plot which is the most useful visualisation of OPTICS results and one that outputs a `data.frame` containing the cluster assignments for every feature.

Both functions require as arguments a distance matrix as well as three parameters for the underlying functions `dbscan::optics` and `dbscan::extractDBSCAN`: `eps`, `minPts` and `eps_c1`. Lowering the `eps` parameter (default is 10000) limits the size of the epsilon neighbourhood which from experience has very little effect on the results. `minPts` defaults to 3 in CluMSID. It defines how many points are considered for reachability distance calculation between clusters. The `dbscan::optics` default for `minPts` is 5. Users are encouraged to experiment with this parameter. `eps_c1` is the reachability threshold to identify clusters and can be varied based on your data. Lowering `eps_c1` leads to a larger number of smaller clusters and vice versa for raising the value. In general, it is advisable to choose a higher `eps_c1` for  $MS^2$  spectra than for neutral loss patterns, since the latter tend to show less similarity to each other. For details, please refer to the `dbscan` help for the `dbscan::optics` and `dbscan::extractDBSCAN` functions.

If the default parameters are used, the generation of an OPTICS reachability plots is very simple, shown here for  $MS^2$  spectra and neutral loss patterns:

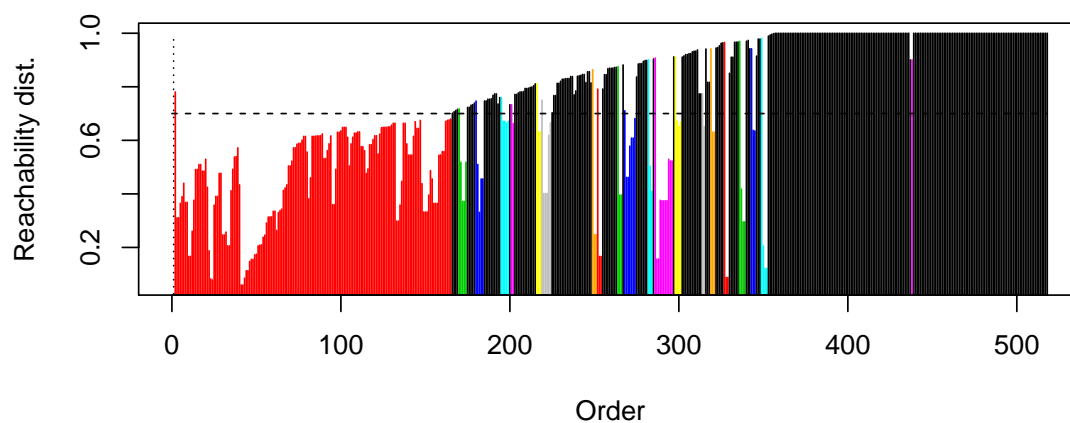


```
OPTICSplot(distmat)
```



**Figure A-4:** Reachability distance plot resulting from OPTICS density based clustering of the MS<sup>2</sup> spectra similarities of the example data set. Bars represent features in OPTICS order with heights corresponding to the reachability distance to the next feature. The dashed horizontal line marks the reachability threshold that separates clusters. The resulting clusters are colour-coded with black representing noise, i.e. features not assigned to any cluster.

```
OPTICSplot(nlmat, eps_cl = 0.7)
```



**Figure A-5:** Reachability distance plot resulting from OPTICS density based clustering of the neutral loss similarities of the example data set (cf Figure 4).

In the reachability plots, every line represents a feature and the height of the line is the reachability distance to the next feature in the OPTICS order. Thus, valleys represent groups of similar spectra or neutral loss patterns. The order and the cluster assignment can be studied using the `OPTICStbl` function that outputs a three-column `data.frame` with feature id, cluster assignment and OPTICS order. The order of features in the `data.frame` corresponds to the original order in the input distance matrix. Features that were not assigned to a cluster are black in the reachability plot and have the cluster ID 0. `OPTICStbl` takes the same arguments as `OPTICSplot`. The two functions have to be run with exactly the same parameters to assure compatibility of results.

```
OPTICStable <- OPTICStbl(distmat)

head(OPTICStable)
#>           feature cluster_ID OPTICS_order
#> 1      M146.17T59.35 - spermidine           1           1
#> 2 M129.14T58.57 - spermidine (fragment)       1           3
#> 3 M112.11T57.8 - spermidine (fragment)       1           4
#> 4                        M251.16T60.64           0          185
#> 5                        M212.85T65.02           0          518
#> 6                        M290.85T64.76           0          517
```

## Perform hierarchical clustering

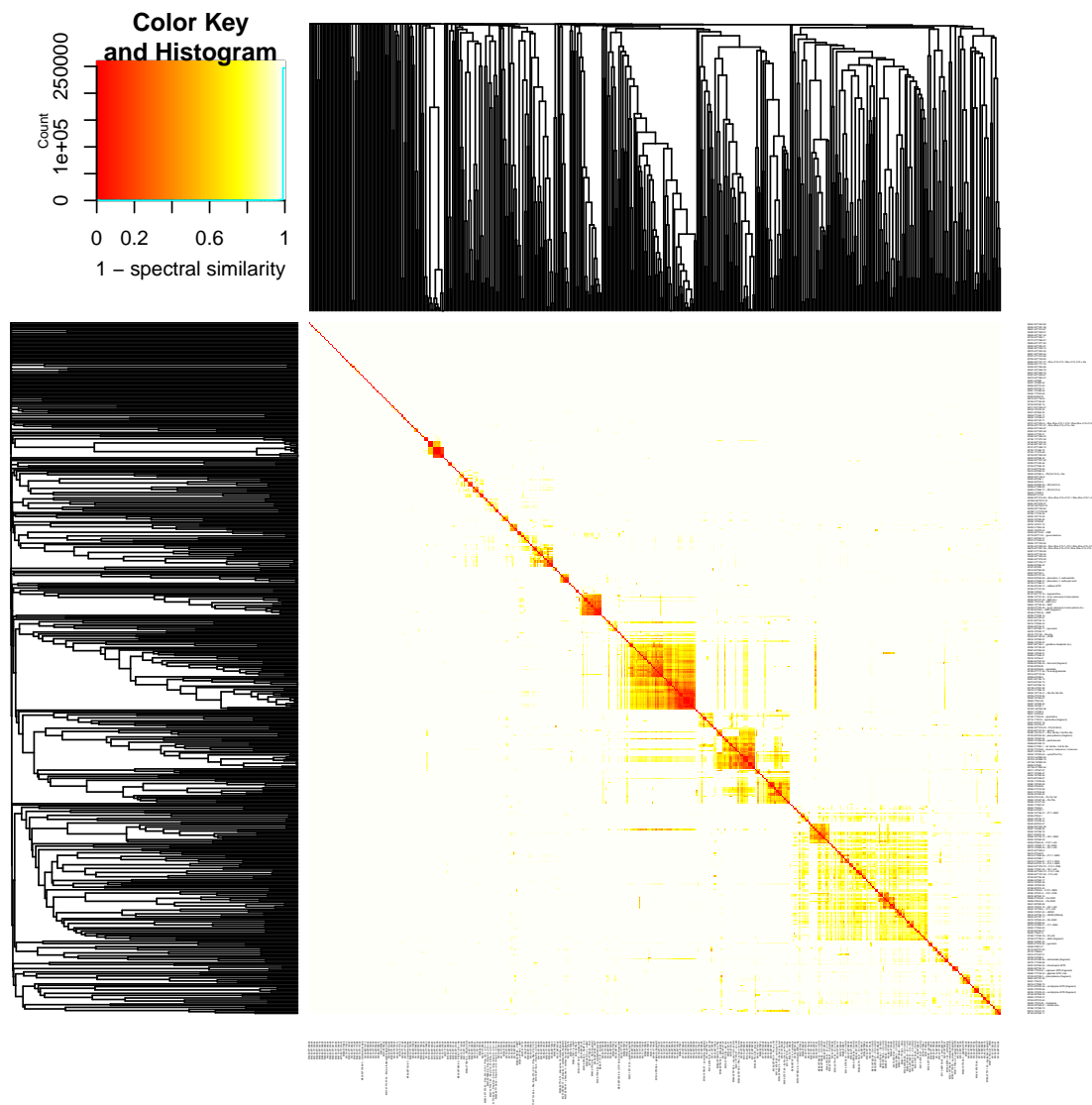
In Depke *et al.* 2017, hierarchical clustering proved the most useful method to unveil structural similarities between features. analogous to density-based clustering, CluMSID offers two functions, one for plots and one for a `data.frame` with cluster assignments, both taking a distance matrix as the only compulsory argument. The other two parameters are `h` (defaults to 0.95), the height where the tree should be cut (see `stats::cutree` for details) and `type` that determines the type visualisation:

- heatmap: a heatmap displaying pairwise similarities/distances along with cluster dendrograms
- dendrogram (default): a circular dendrogram with colour code for cluster assignment

## Create a heatmap

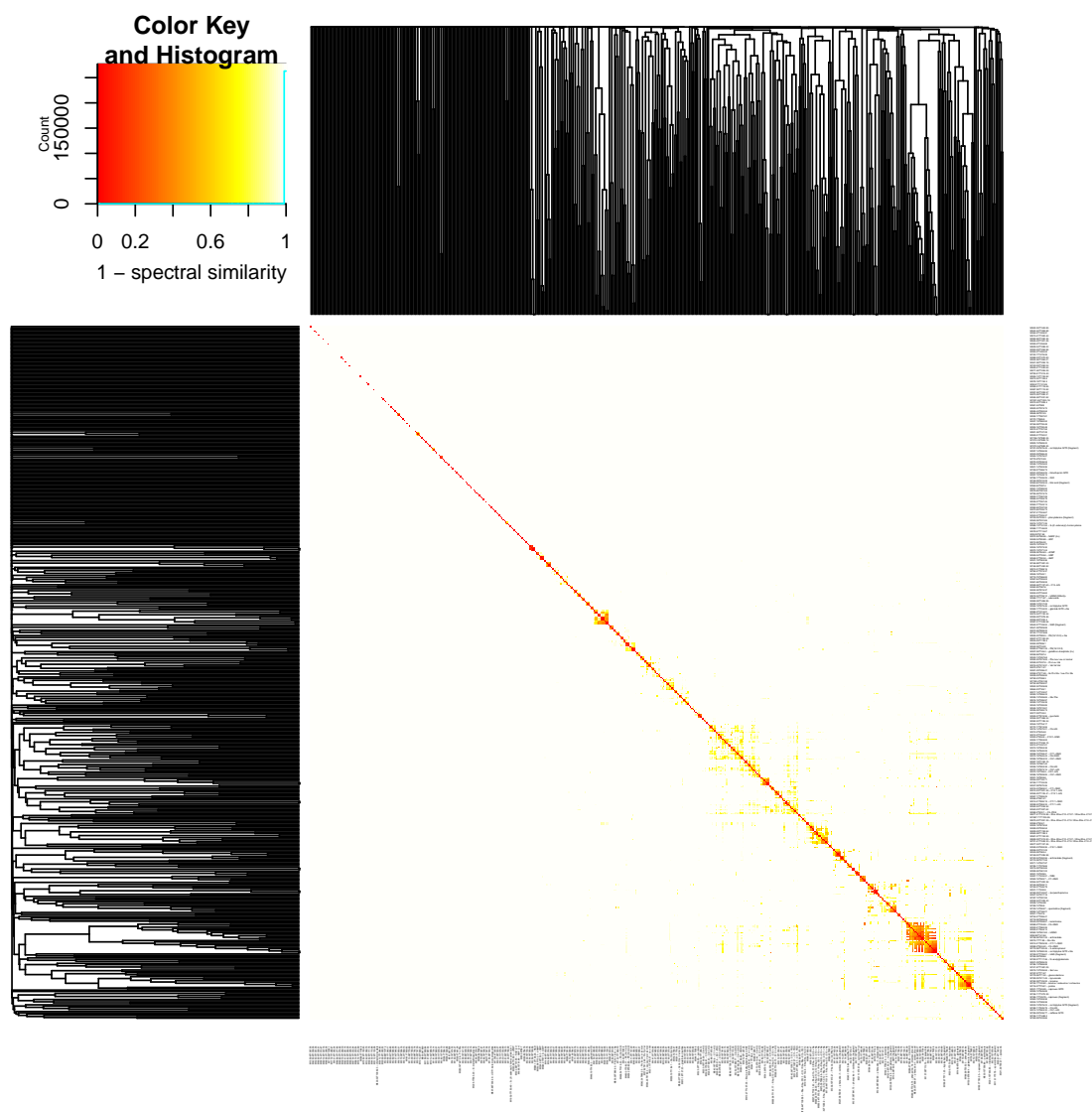
Heatmaps of our example data for MS<sup>2</sup> and neutral loss pattern similarity are created as follows (with reduced label font size by changing `cexRow` and `cexCol` as well as `margins` of the underlying `heatmap.2` function):

```
HCplot(distmat, type = "heatmap",
       cexRow = 0.06, cexCol = 0.06,
       margins = c(5,5))
```



**Figure A-6:** Symmetric heat map of the distance matrix displaying  $MS^2$  spectra similarities of the example data set along with dendrograms resulting from hierarchical clustering based on the distance matrix. The colour encoding is shown in the top-left insert.

```
HCplot(nlmat, type = "heatmap",
       cexRow = 0.06, cexCol = 0.06,
       margins = c(5,5))
```



**Figure A-7:** Symmetric heat map of the distance matrix displaying neutral loss similarities of the example data set along with dendrograms resulting from hierarchical clustering based on the distance matrix. The colour encoding is shown in the top-left insert.

Obviously, it makes sense to export the plots to larger pdf or png files (e.g.  $2000 \times 2000$  pixels) to examine them closely. If exported to pdf, the feature names remain searchable (Ctrl+F in Windows).

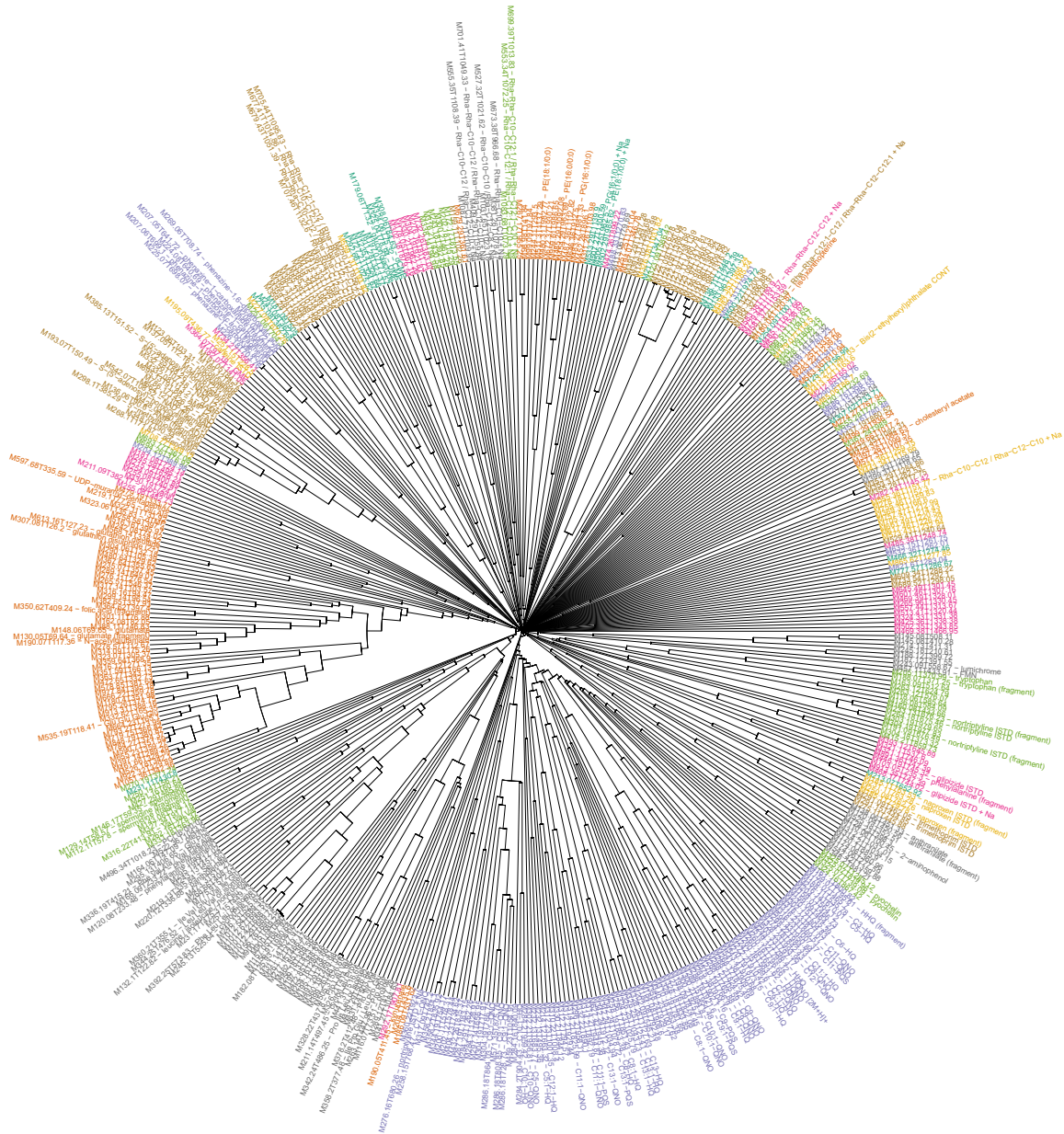
### Create a dendrogram

With the dendrogram, too, it is advisable to export it to pdf in a large format, e.g. as follows:

```
pdf(file = "CluMSID_dendro.pdf", width = 20, height = 20)  
HCplot(distmat)
```

```
dev.off()
```

The plot from our example data looks like this:



**Figure A-8:** Circularised dendrogram as a result of agglomerative hierarchical clustering with average linkage as agglomeration criterion based on MS<sup>2</sup> spectra similarities of the example data set. Each leaf represents one feature and colours encode cluster affiliation of the features. Leaf labels display feature IDs, along with feature annotations, if existent. Distance from the central point is indicative of the height of the dendrogram.

The clusters are colour-coded and if exported to pdf, the tip labels containing feature ID and annotation are searchable. The height of the dendrogram's branching points serves as another piece of information when interpreting the clustered data as it signifies similarity of features.

For a detailed example of how to interpret, please refer to Depke *et al.* 2017, where CluMSID helped to identify new members of several classes of secondary metabolites in *Pseudomonas aeruginosa*.

Like with density-based clustering, it is also possible to generate a list of features with respective cluster assignments using HCtbl. As mentioned above for OPTISplot and OPTICStbl, it is crucial to run HCplot and HCtbl using the same parameters.

```
HCTable <- HCtbl(distmat)

head(HCTable)
#>               feature cluster_ID
#> 1      M146.17T59.35 - spermidine      1
#> 2 M129.14T58.57 - spermidine (fragment)  1
#> 3 M112.11T57.8 - spermidine (fragment)  1
#> 4      M251.16T60.64                  1
#> 5      M212.85T65.02                  2
#> 6      M290.85T64.76                  3
```

## Generate a correlation network

As a new functionality, CluMSID offers the possibility to analyse the similarity data using weighted correlation networks. These networks offer some advantages with respect to standard clustering methods, most notably that they do not strictly assign every feature to a distinct cluster but also represent similarities between features that would fall into different clusters in hierarchical or density-based clustering. Thus, correlation networks potentially contain more useful information for data interpretation. On the downside, the interpretation is also complicated by this lack of concrete cluster assignments. E.g., we cannot simply look up which features belong to the same cluster in order to examine their spectra closely but we have to go back to the correlation network visualisation and search for connected features manually.

networkplot requires some arguments:

- `distmat`: *matrix*; a distance matrix like for all other functions described above
- `interactive`: *logical*; Similar to MDSplot, correlation network can be generate as interactive plots that are zoomable and display feature IDs on mouse-over. If that is desired, set `interactive` to `TRUE` (default is `FALSE`).
- `show_labels`: *logical*; whether to display feature IDs in the (non-interactive) plot (default is `FALSE`, ignored if `interactive = TRUE`)
- `label_size`: *numeric*; font size of feature ID labels (default is 1.5, which is way smaller than the default in `GGally::ggnet2`, 4.5)
- `highlight_annotated`: *logical*; whether to plot dots for features with annotation in a different colour (same as in `MDSplot`, default is `FALSE`)
- `min_similarity`: *numeric*; the minimum similarity (1 – distance) threshold (similarities below this threshold will be ignored, default is 0.1)
- `exclude_singletons`: *logical*; whether to exclude features from the plot that do not have connections to other features, particularly useful with data sets containing very dissimilar spectra, e.g. neutral loss patterns or  $MS^1$  pseudospectra (default is `FALSE`)

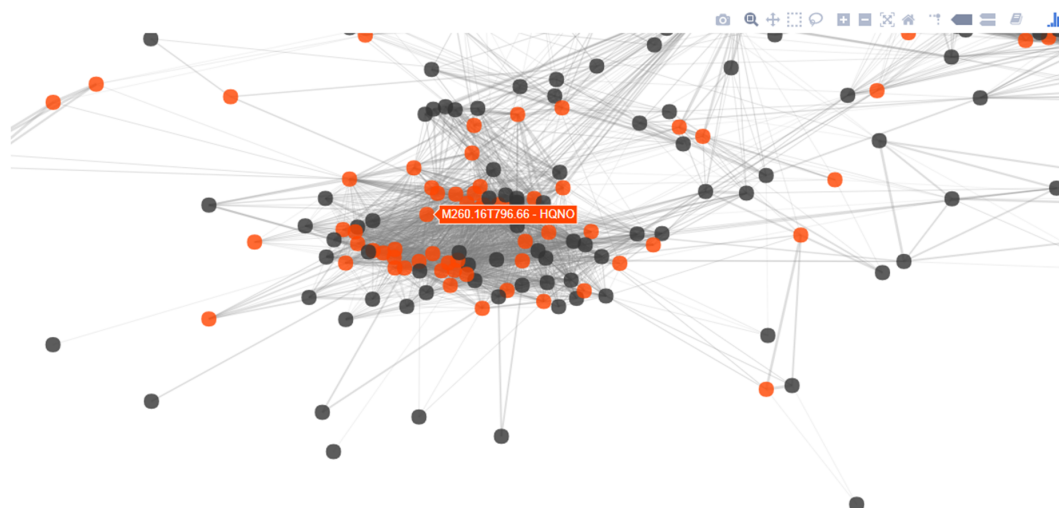




As you can guess from this busy plot, it makes sense to use the interactive visualisation. Just like with `MDSplot`, you can view the interactive plot within RStudio or save it as html and view it in web browser.

```
my_net <- networkplot(distmat, interactive = TRUE,  
                      highlight_annotated = TRUE)  
  
htmlwidgets::saveWidget(my_net, "net.html")
```

This is how it looks like if you open the html file in Firefox, zoom in on a cluster and mouse over a feature:



**Figure A-10:** Screenshot of the interactive version of the Correlation network plot based on  $MS^2$  spectra similarities of the example data set (cf Figure 9). Zoomed image section with tooltip displaying feature information upon mouse-over.

Please be aware that the spatial arrangement of the data points in the plot has a random component, i.e. while the relative position of the points (the distance to each other) is always the same, the absolute position varies and will not be the same even if the same command is executed twice.

The pairwise similarity of spectra or neutral loss patterns of features expressed by the cosine score is signified by the width of the line connecting the two features. All pairwise similarities greater than `min_similarity` result in a connecting line in the plot. The spatial proximity in which the features are mapped onto the plot is determined by the multivariate method underlying the network generation.

As we have already noticed after inspection of the heatmaps on p.13–14, the neutral loss patterns show much less similarity to each other than the  $MS^2$  spectra data. Thus, we expect quite a few neutral loss patterns that do not show any similarity to another neutral loss pattern. This expectation justifies the exclusion of these 'singletons' from the correlation network analysis. To do so, just set `exclude_singletons` to `TRUE`:





## Additional functionalities

Multidimensional scaling, density-based clustering, hierarchical clustering and correlation network analysis are the main CluMSID tools to analyse MS<sup>2</sup> spectra or neutral loss pattern similarity data, however, the package contains some additional functionalities that may facilitate data analysis in some cases and can also be used in other contexts with or without the above-mentioned unsupervised methods.

### Access individual spectra from a list of spectra by various slot entries

Accessing S4 objects within lists is not trivial. Therefore, CluMSID offers a function to access individual or several MS2spectrum objects by their slot entries. `getSpectrum()` requires the following arguments:

- `featlist`: a list that contains only objects of class MS2spectrum
- `slot`: the slot to be searched (invalid slot arguments will produce errors):
  - `id`
  - `annotation`
  - `precursor` ( $m/z$  of precursor ion)
  - `rt` (retention time of precursor)
- `what`: the search term or number, must be *character* for `id` and `annotation` and *numeric* for `precursor` and `rt`
- `mz.tol`: the tolerance used for precursor ion  $m/z$  searches, defaults to 1E-05 (10ppm)
- `rt.tol`: the tolerance used for precursor ion retention time searches, defaults to 30s; high values can be used to specify retention time ranges (see example)

Some examples will demonstrate the use of `getSpectrum()`:

**1. Accessing a spectrum by its ID.** For this, the exact feature ID must be known:

```
getSpectrum(annotatedSpeclist, "id", "M244.17T796.4")
#> An object of class "MS2spectrum"
#> id: M244.17T796.4
#> annotation: HHQ
#> precursor: 244.1700
#> retention time: 796.4
#> polarity: positive
#> MS2 spectrum with 98 fragment peaks
#> neutral loss pattern with 81 neutral losses
```

**2. Accessing a spectrum by its annotation.** For this, the exact annotation has to be known as well, other annotations will produce a message:

```
getSpectrum(annotatedSpeclist, "annotation", "HHQ")
#> An object of class "MS2spectrum"
#> id: M244.17T796.4
#> annotation: HHQ
#> precursor: 244.1700
#> retention time: 796.4
#> polarity: positive
#> MS2 spectrum with 98 fragment peaks
#> neutral loss pattern with 81 neutral losses
```

```
getSpectrum(annotatedSpeclist, "annotation", "C7-HQ")
#> No spectrum with that annotation.
```

**3. Accessing spectra by their precursor ion  $m/z$ .** If the list contains more than one spectrum with a precursor ion  $m/z$  within the tolerance, the output is again a list of MS2spectrum objects that meet the specified criterion:

```
getSpectrum(annotatedSpeclist, "precursor", 286.18, mz.tol = 1E-03)
#> [[1]]
#> An object of class "MS2spectrum"
#> id: M286.18T728.73
#> annotation: C9:1-QNO
#> precursor: 286.1799
#> retention time: 728.73
#> polarity: positive
#> MS2 spectrum with 4 fragment peaks
#> neutral loss pattern with 2 neutral losses
#> [[2]]
#> An object of class "MS2spectrum"
#> id: M286.18T808.85
#> annotation: C9:1-QNO
#> precursor: 286.1804
#> retention time: 808.85
#> polarity: positive
#> MS2 spectrum with 7 fragment peaks
#> neutral loss pattern with 5 neutral losses
#> [[3]]
#> An object of class "MS2spectrum"
#> id: M286.18T864.03
#> annotation: C9:1-QNO
#> precursor: 286.1808
#> retention time: 864.03
#> polarity: positive
#> MS2 spectrum with 183 fragment peaks
#> neutral loss pattern with 167 neutral losses
#> [[4]]
#> An object of class "MS2spectrum"
#> id: M286.18T921.6
#> annotation: C9:1-PQS
#> precursor: 286.1808
#> retention time: 921.6
#> polarity: positive
#> MS2 spectrum with 3 fragment peaks
#> neutral loss pattern with 1 neutral losses
```

**4. Accessing spectra by their precursor retention time.** Here, too, we can extract several MS2spectrum objects by setting a larger retention time tolerance. If we want to extract the spectra of all compounds that elute from 6min (360s) to 8min (480s), we proceed as follows:

```
six_eight <- getSpectrum(annotatedSpeclist, "rt", 420, rt.tol = 60)
length(six_eight)
#> [1] 75
```

## Find spectra that contain a specific fragment or neutral loss

Another pair of accessory functions is `findFragment()` and `findNL()` which are used to find spectra that contain a specific fragment ion or neutral loss. Analogous to `getSpectrum()`, they need as arguments a list of `MS2spectrum` objects, the  $m/z$  of the fragment or neutral loss of interest and the respective  $m/z$  tolerance in ppm (default is 10ppm). The two functions can be useful in many situation, e.g. when working with lipid data where head groups and fatty acids often give characteristic fragments or neutral losses. In the world of *P. aeruginosa* secondary metabolites, alkylquinolones (AQs) play an important role and most of the AQ  $MS^2$  spectra contain a signature fragment with an  $m/z$  of 159.068. Based on this fragment  $m/z$ , we can create a list of putative AQs:

```
putativeAQs <- findFragment(annotatedSpeclist, 159.068)
#> 70 spectra were found that contain a fragment of m/z 159.068 +/- 10 ppm.
```

An example for common neutral losses are nucleoside monophosphates that all loose ribose-5'-monophosphate, resulting in a neutral loss of 212.009 in ESI(+). Using `findNL()` we find CMP, UMP, AMP and GMP.

```
findNL(annotatedSpeclist, 212.009)
#> 4 neutral loss patterns were found that contain a neutral loss of m/z 212.009 +/- 10 ppm.
#> [[1]]
#> An object of class "MS2spectrum"
#> id: M324.06T75.32
#> annotation: CMP
#> precursor: 324.0591
#> retention time: 75.32
#> polarity: positive
#> MS2 spectrum with 8 fragment peaks
#> neutral loss pattern with 8 neutral losses
#> [[2]]
#> An object of class "MS2spectrum"
#> id: M325.04T78.94
#> annotation: UMP
#> precursor: 325.0429
#> retention time: 78.94
#> polarity: positive
#> MS2 spectrum with 5 fragment peaks
#> neutral loss pattern with 5 neutral losses
#> [[3]]
#> An object of class "MS2spectrum"
#> id: M348.07T90.34
#> annotation: AMP
#> precursor: 348.0707
#> retention time: 90.34
#> polarity: positive
#> MS2 spectrum with 21 fragment peaks
#> neutral loss pattern with 19 neutral losses
#> [[4]]
#> An object of class "MS2spectrum"
#> id: M364.07T97.19
#> annotation: GMP
#> precursor: 364.0659
#> retention time: 97.19
#> polarity: positive
#> MS2 spectrum with 6 fragment peaks
#> neutral loss pattern with 6 neutral losses
```

## Match one spectrum against a set of spectra

If you are mainly interested in one or a few number of spectra or neutral loss patterns, it may be sufficient to match one feature at a time against a larger set of spectra. This set of spectra can be all spectra contained in one mzXML file like in all the examples in this tutorial or they could be a spectral library, as long as its format in R is a list of MS2spectrum objects.

The `getSimilarities()` function requires several arguments:

- `spec`: The spectrum to be compared to other spectra. Can be either an object of class `MS2spectrum` or a two-column numerical matrix that contains fragment mass-to-charge ratios in the first and intensities in the second column.
- `speclist`: The set of spectra to which `spec` is to be compared. Must be a list where every entry is an object of class `MS2spectrum`. Can be generated from an mzXML file as shown above or constructed using `new("MS2spectrum", ...)` for every list entry (see example).
- `type`: Specifies whether MS<sup>2</sup> spectra or neutral loss patterns are to be compared. Must be either 'spectrum' (default) or 'neutral\_losses'.
- `hits_only`: Logical that indicates whether the result should contain only similarities greater than zero (see example).

In the first example, we want to find all MS<sup>2</sup> spectra in our example data set that are similar to the spectrum of pyocyanin, an important secondary metabolite from *Pseudomonas aeruginosa* and therefore match the pyocyanin spectrum against our annotatedSpeclist. Because we have already identified pyocyanin in the data set, we can use `getSpectrum` to extract the MS2spectrum object from annotatedSpeclist. We do not want to search all 518 elements of the result vector, so we set `hits_only` to `TRUE` to exclude spectra that have 0 similarity to the pyocyanin spectrum.

```
pyo <- getSpectrum(annotatedSpeclist, "annotation", "pyocyanin")

sim_pyo <- getSimilarities(pyo, annotatedSpeclist, hits_only = TRUE)
sim_pyo
#> M110.06T100.45 M123.06T103.31 M332.56T107.48 M332.08T113.21
#> 0.0235166588 0.0071763662 0.0032575891 0.0035153018
#> M182.08T125.93 M166.09T233.22 M120.08T233.48 M103.05T235.3
#> 0.0414005385 0.0394723541 0.0492390806 0.0826780036
#> M174.06T277.59 M220.12T336.88 M525.18T352.51 M243.08T362.4
#> 0.0391004892 0.0205482303 0.0060019991 0.0145904545
#> M188.07T371.25 M205.1T370.99 M211.09T382.17 M187.12T391.55
#> 0.0176900909 0.0179895663 1.0000000000 0.0210280136
#> M188.12T399.72 M254.09T400.89 M160.08T433.66 M291.15T444.85
#> 0.0105392131 0.2071528536 0.0489638040 0.0106479317
#> M120.04T450.56 M138.06T451.33 M176.07T465.7 M491.29T496.41
#> 0.0287432023 0.0202198052 0.0275059908 0.0610208210
#> M255.08T482.73 M245.59T495.11 M145.08T508.11 M163.09T512.64
#> 0.6451546287 0.2583432230 0.0473127795 0.0034167239
#> M188.11T535.78 M321.1T537.6 M243.09T558.67 M136.08T584.09
#> 0.0057005179 0.0293635312 0.0116275804 0.0132716679
#> M118.06T585.64 M215.12T626.24 M224.08T640.69 M213.07T652.92
#> 0.0203921366 0.3252546561 0.0325490977 0.0083842257
#> M216.14T670.01 M227.08T670.26 M264.18T675.46 M233.13T676.23
#> 0.0009299928 0.0034818309 0.0172023762 0.0143332573
#> M225.07T698.07 M207.06T699.1 M257.06T704.3 M226.18T703.76
#> 0.0253940205 0.0230298767 0.0028192053 0.0255571995
#> M325.07T739.09 M181.08T724.14 M330.19T724.4 M255.08T740.91
#> 0.0010572974 0.1283755709 0.5019236568 0.2030839014
#> M446.19T745.32 M321.1T746.09 M891.36T747.39 M231.1T763.28
```

```
#> 0.0750793676 0.1017149711 0.0714108632 0.0070294838
#> M185.1T763.53 M288.2T765.88 M258.15T768.47 M328.14T772.12
#> 0.0094225379 0.0018840838 0.0168976240 0.0009052790
#> M242.15T789.6 M304.19T786.75 M260.16T796.66 M244.17T796.4
#> 0.0071606643 0.0066966285 0.0141834395 0.0106077162
#> M159.07T795.61 M314.21T825.99 M326.18T833.85 M286.18T864.03
#> 0.0124577125 0.0006167122 0.0023434969 0.0205024121
#> M270.19T873.15 M268.17T875.49 M178.05T867.16 M198.09T874.45
#> 0.0089744139 0.0011687551 0.0413446455 0.0064665757
#> M170.1T885.9 M288.2T902.1 M184.08T908.6 M272.2T910.42
#> 0.0062571323 0.0100270721 0.0036517997 0.0086043375
#> M312.2T929.44 M314.21T944.87 M296.2T958.11 M298.22T984.33
#> 0.0085388744 0.0128215188 0.0054678320 0.0065812089
#> M500.22T974.72 M304.19T977.83 M303.19T979.65 M340.23T1007.62
#> 0.0396920510 0.0059045590 0.0049045093 0.0052002486
#> M324.23T1025.79 M314.21T1034.15 M326.25T1043.73 M679.43T1051.39
#> 0.0005826366 0.0030626495 0.0005424581 0.0008290325
```

We get 84 spectra that have a non-zero similarity to the pyocyanin spectrum, including pyocyanin itself with a similarity of 1. Of course, we can further filter the data by subsetting the result vector in order to exclude spectra that have only minimal similarity, e.g. M679.43T1051.39 with a cosine similarity of only 0.0008 (the last element in the vector).

In the second example, we generate a new `speclist`, e.g. from a spectral library. We look at the unknown feature that has most similarity to pyocyanin. As pyocyanin is contained in `annotatedSpeclist` itself, we have to look at the second highest similarity. Again, we use `getSpectrum()` to extract the object from `annotatedSpeclist`:

```
highest_sim <- sort(sim_pyoc, decreasing = TRUE)[2]

sim_spec <- getSpectrum(annotatedSpeclist, "id", names(highest_sim))
sim_spec
#> An object of class "MS2spectrum"
#> id: M255.08T482.73
#> annotation:
#> precursor: 255.0761
#> retention time: 482.73
#> polarity: positive
#> MS2 spectrum with 5 fragment peaks
#> neutral loss pattern with 3 neutral losses
```

We see that the feature is not annotated. We are interested whether this feature also shows similarity to other members of the phenazine family of *P. aeruginosa* secondary metabolites. Some phenazines are contained in `annotatedSpeclist` but some are not, so we make a new `speclist` called `phenazines` and add the missing spectra manually from an in-house library:

```
phenazines <- list()
phenazines[[1]] <- getSpectrum(annotatedSpeclist,
                              "annotation", "pyocyanin")
phenazines[[2]] <- getSpectrum(annotatedSpeclist,
                              "annotation", "phenazine-1-carboxamide")
phenazines[[3]] <- getSpectrum(annotatedSpeclist,
                              "annotation", "phenazine-1-carboxylic acid")
phenazines[[4]] <- getSpectrum(annotatedSpeclist,
                              "annotation", "phenazine-1,6-dicarboxylic acid")
phenazines[[5]] <- new("MS2spectrum", id = "lib_entry_1",
```

```

        annotation = "1-hydroxyphenazine",
        spectrum = matrix(c(168.0632, 14,
                           169.0711, 288,
                           170.0743, 33,
                           179.0551, 62,
                           197.0653, 999),
                          byrow = TRUE,
                          ncol = 2))
phenazines[[6]] <- new("MS2spectrum", id = "lib_entry_2",
                     annotation = "2-hydroxy-phenazine-1-carboxylic acid",
                     spectrum = matrix(c(167.0621, 43,
                                          179.0619, 93,
                                          180.0650, 12,
                                          195.0564, 40,
                                          223.0509, 999,
                                          224.0541, 142,
                                          241.0611, 60),
                                        byrow = TRUE,
                                        ncol = 2))
phenazines[[7]] <- new("MS2spectrum", id = "lib_entry_3",
                     annotation = "pyocyanin (library spectrum)",
                     spectrum = matrix(c(168.0690, 58,
                                          183.0927, 152,
                                          184.0958, 19,
                                          196.0640, 118,
                                          197.0674, 15,
                                          211.0873, 999,
                                          212.0905, 145),
                                        byrow = TRUE,
                                        ncol = 2))

getSimilarities(sim_spec, phenazines, hits_only = FALSE)
#> M211.09T382.17 M224.08T640.69 M225.07T698.07 M269.06T708.74 lib_entry_1
#> 0.6451546 0.0000000 0.0000000 0.0000000 0.0000000
#> lib_entry_2 lib_entry_3
#> 0.0000000 0.6375061

```

As a result, we get the interesting information that the MS<sup>2</sup> spectra similarity of our unknown feature seems to be specific to pyocyanin (both the experimental and the library spectrum).

### Convert MSnbase objects to class MS2spectrum

The MSnbase package—which is commonly used for proteomics applications and is also associated with XCMS3—has two classes for (MS<sup>2</sup>) spectra, `Spectrum` and `Spectrum2` which contain spectra along with meta-information. These meta-information differ from those contained in `MS2spectrum` objects and are not very well suited for metabolomics applications. Still, it is possible to use `CluMSID` functions with objects of those two classes by converting them to `MS2spectrum` objects using `as.MS2spectrum()`:

```

CluMSID_object <- as.MS2spectrum(MSnbase_object)
# or alternatively
CluMSID_object <- as(MSnbase_object, "MS2spectrum")

```



## Split polarities from polarity-switching runs

As polarity-switching and similar methods are gaining importance in LC-MS/MS metabolomics, CluMSID offers the possibility to process LC-MS/MS data containing spectra of different polarities. As spectra from positive and negative ionisation show different fragmentation mechanisms and patterns, it does not appear to be useful to compare spectra of different polarity to each other. Therefore, CluMSID provides a function to separate positive and negative spectra from each other. This has to be done in the very beginning of the analysis to not interfere with spectral merging. Positive and negative spectra can then be processed independently from each other as shown above.

A schematic workflow would look like this:

```
raw_list_mixedpolarities <- extractMS2spectra("raw_file_mixedpolarities.mzXML")

raw_list_positive <- splitPolarities(raw_list_mixedpolarities, "positive")
raw_list_negative <- splitPolarities(raw_list_mixedpolarities, "negative")

speclist_positive <- mergeMS2spectra(raw_list_positive)
speclist_negative <- mergeMS2spectra(raw_list_negative)
```

... and so on as described in this tutorial.

## Use MS<sup>1</sup> pseudospectra instead of or in addition to MS<sup>2</sup> data

MS<sup>1</sup> pseudospectra are groups of peaks/ions that derive or are assumed to derive from the same compound. They consist of peaks for in-source fragment, adducts etc. Pseudospectra can contain structural information about analytes, e.g. about moieties that easily fragment even in MS<sup>1</sup> mode without CID. Thus, it might sometimes be useful to study similarities between pseudospectra analogously to those between MS<sup>2</sup> spectra. CluMSID makes use of the CAMERA package to assign peaks to pseudospectra. A custom S4 class named `pseudospectrum` is used which is very similar to the `MS2spectrum` class. For obvious reasons, it does not contain a precursor ion *m/z* slot and thus no neutral loss pattern, either. The `pcgroup` defined by CAMERA is used as ID, an annotation can be added if desired.

### Extract pseudospectra

To extract pseudospectra, you first have to process your data using the CAMERA package, either in R or via XCMSonline, where this is done automatically. There are two possibilities to use the `extractPseudospectra()` function in CluMSID: either with an `xsAnnotate` object which you generate with CAMERA in R or with a `data.frame` that contains data on *m/z*, retention time, intensity and `pcgroup`, e.g. the results table from XCMSonline.

The latter is demonstrated with the XCMSonline results table already used to generate a peak table. If the column names are not changed, the `data.frame` can be supplied as-is and `intensity_columns` does not have to be specified. We want to exclude pseudospectra that have only one peak, so we set `min_peaks = 2`.

```
pstable <-
  read_delim(file = system.file("extdata",
                                "TD035_XCMS.annotated.diffreport.tsv",
                                package = "CluMSIDdata"),
            delim = "\t")

pseudospeclist <- extractPseudospectra(pstable, min_peaks = 2)
```

As a result, we get a list with 198 pseudospectra that we can now process further.



## Create distance matrix for pseudospectra

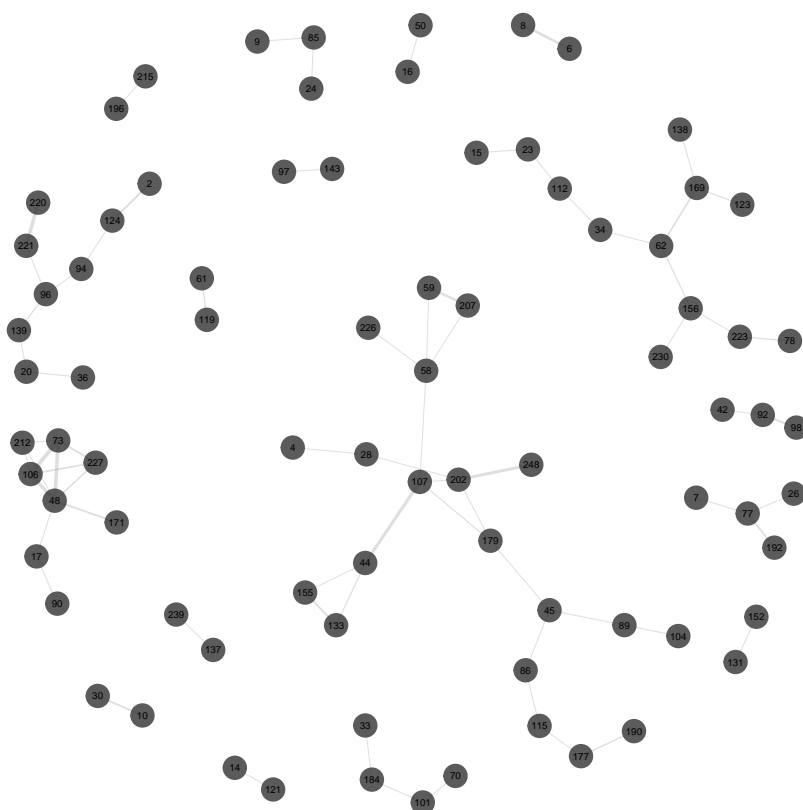
The creation of a distance matrix is analogous to the procedure for MS<sup>2</sup> spectra:

```
pseudodistmat <- distanceMatrix(pseudospeclist)
```

## Generate a correlation network for pseudospectra

The distance matrix can now be used for MDS, clustering and correlation networks just like described above. For demonstration, we generate a correlation network:

```
networkplot(pseudodistmat, show_labels = TRUE, exclude_singletons = TRUE)
```



**Figure A-12:** Correlation network plot based on similarities of pseudospectra of the example data set. Grey dots indicate non-identified features, orange dots identified ones. Labels display CAMERA's pseudospectra IDs. Edge widths are proportional to spectral similarity of the connected features.

With the exclusion of singletons, we get a much less busy plot than for MS<sup>2</sup> data but we still find quite a few connections that may prove informative.

# Clustering Mass Spectra from Low Resolution GC-EI-MS Data Using CluMSID

Tobias Depke

December 31, 2018

## Contents

Introduction . . . . .	B-1
Data import and preprocessing . . . . .	B-1
Extraction and annotation of spectra . . . . .	B-2
Generation of distance matrix . . . . .	B-3
Data exploration . . . . .	B-3
Conclusion . . . . .	B-8

## Introduction

Although originally developed for high resolution LC-MS/MS data, CluMSID can also be used to find similarities in GC-EI-MS data, i.e. data from hard ionisation mass spectrometry.

As the peak picking and spectral merging differs considerably from data dependent ESI-MS/MS, we cannot use the standard CluMSID functions `extractMS2spectra()` and `mergeMS2spectra()`. In fact, the analysis of mass spectra from hard ionisation mass spectrometry resembles the one of MS<sup>1</sup> pseudospectra in ESI-MS. Thus, we can use the CluMSID function `extractPseudospectra()` in conjunction with pseudospectra generated by the CAMERA package.

Since `xcms` and CAMERA sometimes have difficulties in handling GC-EI-MS data, we use the `metaMS` package that enables workflows specialised to the analysis of such data. We also require the `metaMSdata` package from which we import the `FEMSettings` object that contains `xcms` and CAMERA settings for GC-EI-MS data.

```
library(CluMSID)
library(CluMSIDdata)
library(metaMS)
library(metaMSdata)
data(FEMSettings)
```

## Data import and preprocessing

As example data, we use GC-EI-MS metabolomics data from pooled cell extracts of *Pseudomonas aeruginosa* measured on a Thermo Scientific ITQ linear ion trap that has been converted to netCDF using Thermo Xcalibur. A netCDF file is available in the CluMSIDdata package:

```
pool <- system.file("extdata",
                    "1800802_TD_pool_total_1.cdf",
                    package = "CluMSIDdata")
```

To generate a list of (pseudo)spectra, we first need an `xsAnnotate` object as generated by CAMERA. In the case of GC-MS data, it is more convenient to use the `metaMS` function `runCAMERA()` than actual CAMERA functions. `metaMS::runCAMERA` requires an `xcmsSet` object which we generate by using `xcms::xcmsSet` on our netCDF file (we can do that in one go). We used standard GC-MS settings for `runCAMERA()` as they are proposed in the `metaMS` vignette.

```
xA <- runCAMERA(xcmsSet(pool),
  chrom = "GC",
  settings = metaSetting(TSQXLS.GC, "CAMERA"))
```

## Extraction and annotation of spectra

From the `xsAnnotate` object, we can now extract the (pseudo)spectra using the `CluMSID` function `extractPseudospectra()` function as we would do for  $MS^1$  pseudospectra from LC-ESI-MS data.

```
pslist <- extractPseudospectra(xA, min_peaks = 0)
```

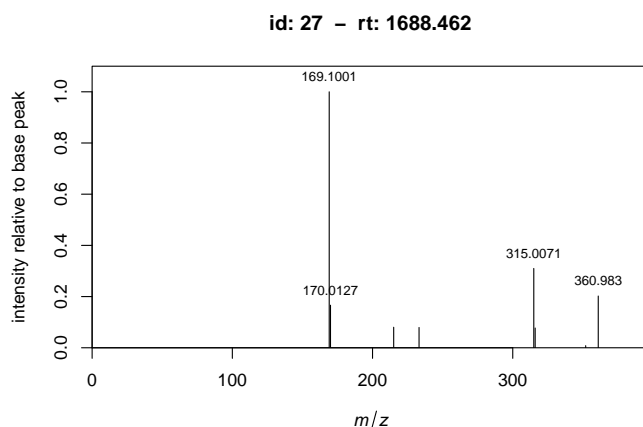
Adding annotations is not as easy as with LC-(DDA-)MS/MS data, because only the retention time and the spectrum itself describe the feature and no precursor  $m/z$  is available. Thus, feature annotations/identifications made in a different programme, in this case `MetaboliteDetector`, have to be compared to the spectra in the `pslist` object.

Like with LC-(DDA-)MS/MS data, we can use `writeFeaturelist()` and `addAnnotations()` to add external annotations. The table output from `writeFeaturelist()` will give NA for all precursor  $m/z$ .

```
writeFeaturelist(pslist, "GC_pre.csv")
```

To facilitate manual annotation, it helps to plot the spectra along with the relevant information for every feature/pseudospectrum. That can be done by `CluMSID`'s `specplot` function:

```
specplot(pslist[[27]])
```



**Figure B-1:** Barplot for pseudospectrum 27, displaying fragment  $m/z$  on the x-axis and intensity normalised to the maximum intensity on the y-axis.

In this example, we load the list of feature annotations from CluMSIDdata:

```
apslist <- addAnnotations(fealist = pslist,  
                          annolist = system.file("extdata",  
                                                  "GC_post.csv",  
                                                  package = "CluMSIDdata"))
```

### Generation of distance matrix

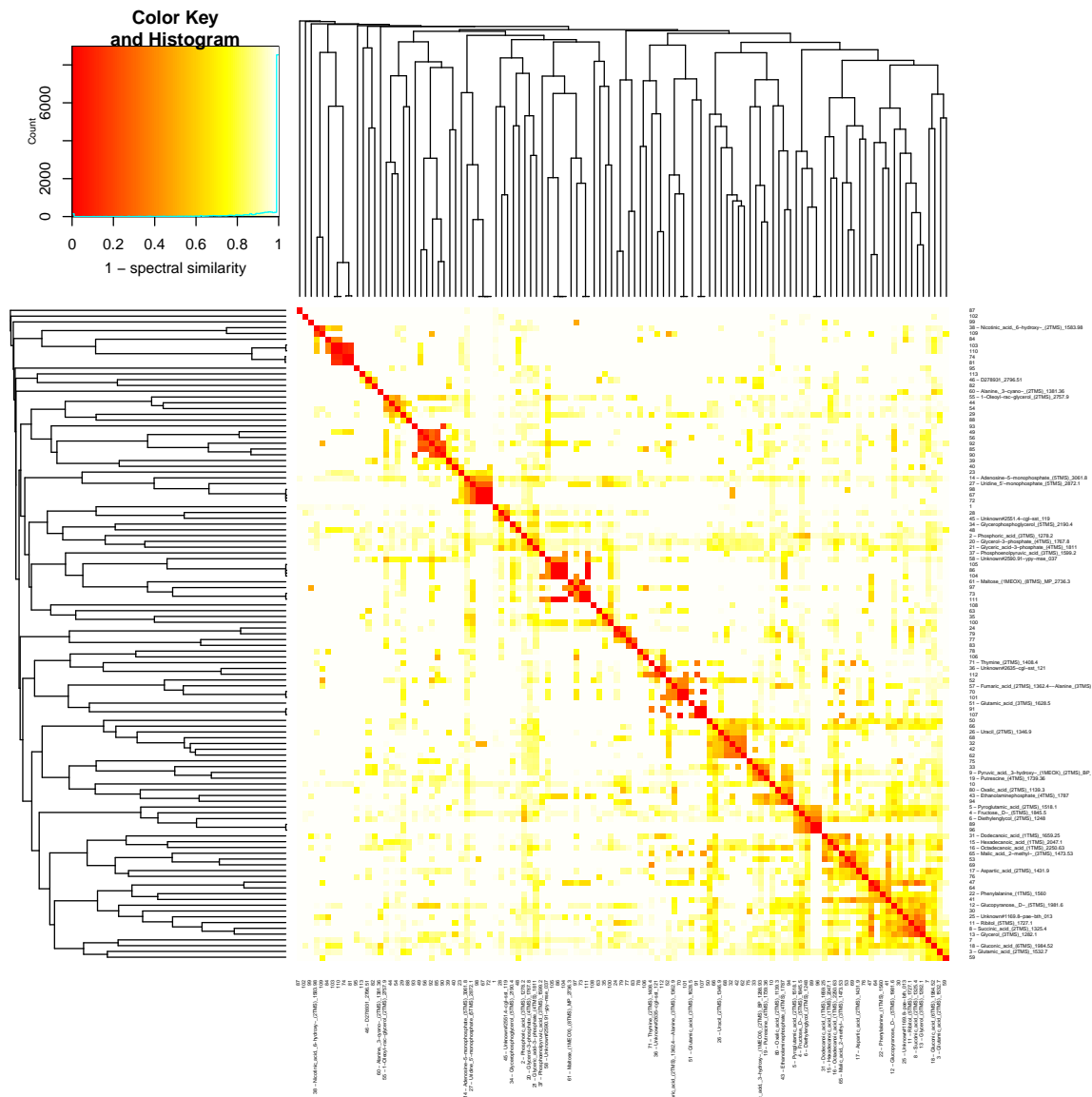
This list of spectra in turn serves as an input for `distanceMatrix()`. As we are dealing with low resolution data, we have to adjust the  $m/z$  tolerance. The default value, 10ppm, is suitable for time-of-flight mass spectrometers while linear ion traps or single quadrupoles which are commonly used in GC-EI-MS only have unit mass resolution, equivalent to a relative mass error of 0.02 to 0.001 depending on the  $m/z$  of the analyte. We chose 0.02 to be tolerant enough for low molecular weight analytes:

```
pseudodistmat <- distanceMatrix(apslist, mZ_tolerance = 0.02)
```

### Data exploration

Starting from this distance matrix, we can use all the data exploration functions that CluMSID offers. In this example workflow, we look at a cluster dendrogram:

```
HClplot(pseudodistmat, type = "heatmap",  
        cexRow = 0.3, cexCol = 0.3,  
        margins = c(7,7))
```



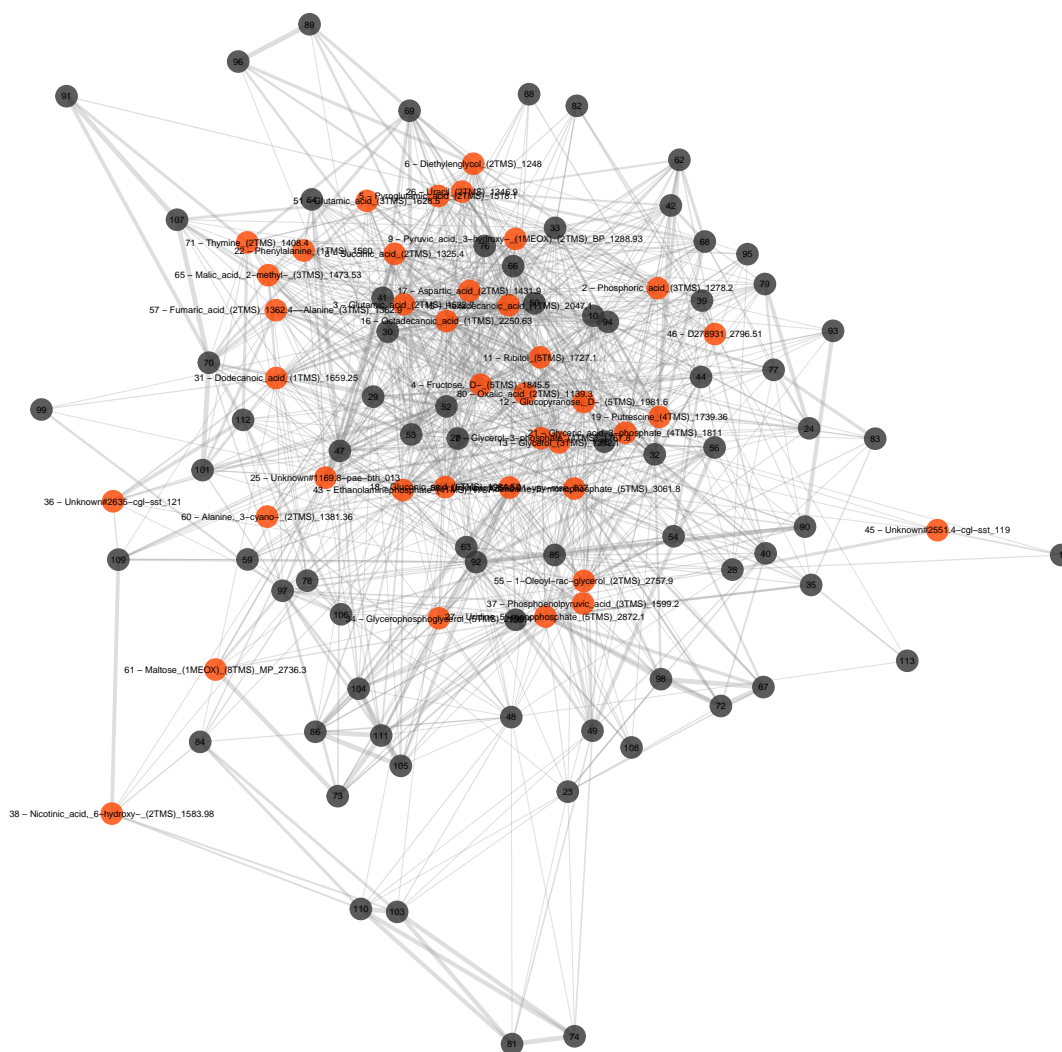
**Figure B-2:** Symmetric heat map of the distance matrix displaying pseudospectra similarities of the GC-EI-MS example data set along with dendrograms resulting from hierarchical clustering based on the distance matrix. The colour encoding is shown in the top-left insert.

It is directly visible that the resulting clusters are not as dense as with the LC-MS/MS example data. In turn, there are more between-cluster similarities.

B-4

This also shows in the correlation network, resulting in a chaotic plot when used with the default minimal similarity of 0.1:

```
networkplot(pseudodistmat, highlight_annotated = TRUE,
            show_labels = TRUE, exclude_singletons = TRUE)
```

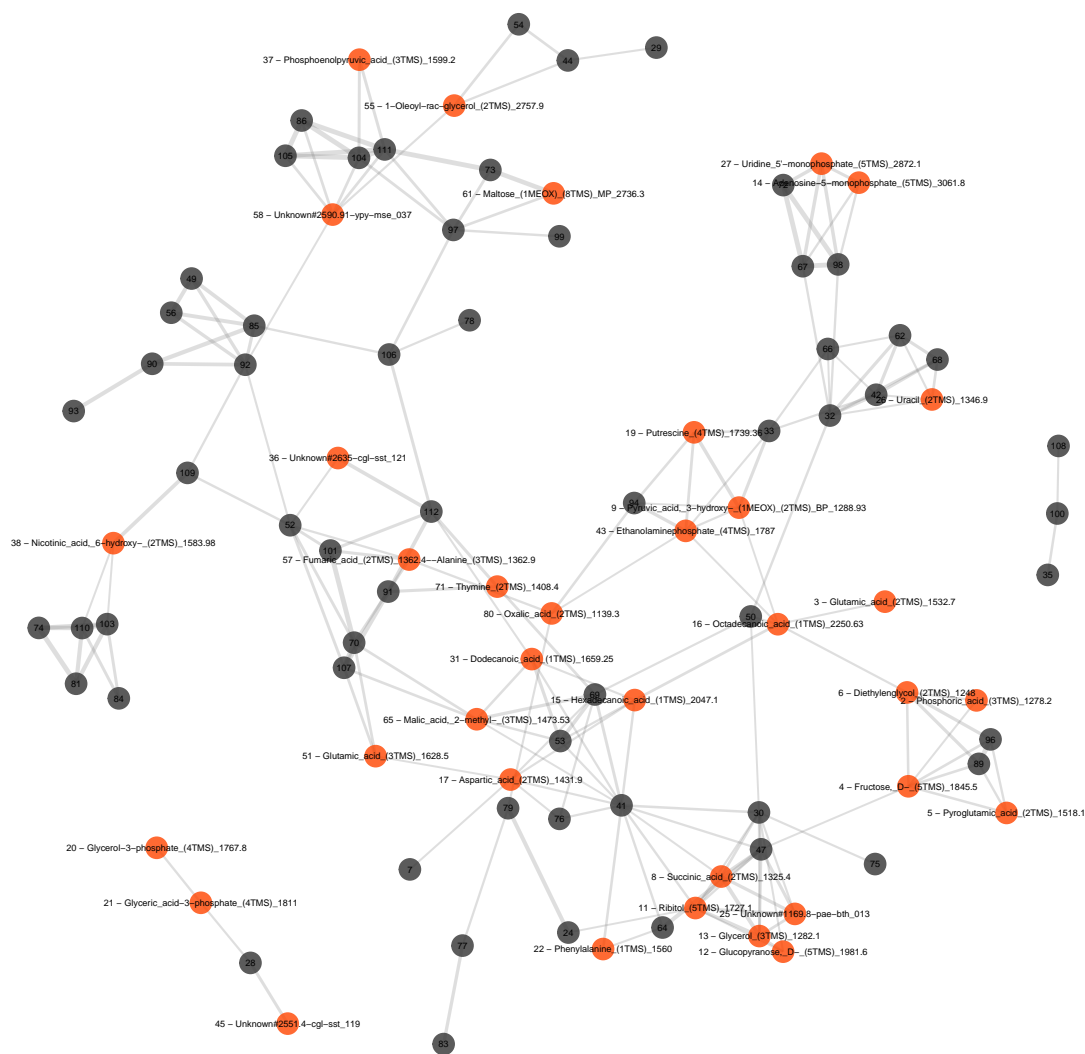


**Figure B-3:** Correlation network plot based on pseudospectra similarities of the GC-El-MS example data set, generated with the default similarity threshold of 0.1. Grey dots indicate non-identified features, orange dots identified ones. Labels display feature IDs, along with feature annotations, if existent. Edge widths are proportional to spectral similarity of the connected features.

B-5

By choosing a higher similarity threshold of e.g. 0.4, it is far easier to identify clusters:

```
networkplot(pseudodistmat, highlight_annotated = TRUE,
            show_labels = TRUE, exclude_singletons = TRUE,
            min_similarity = 0.4)
```

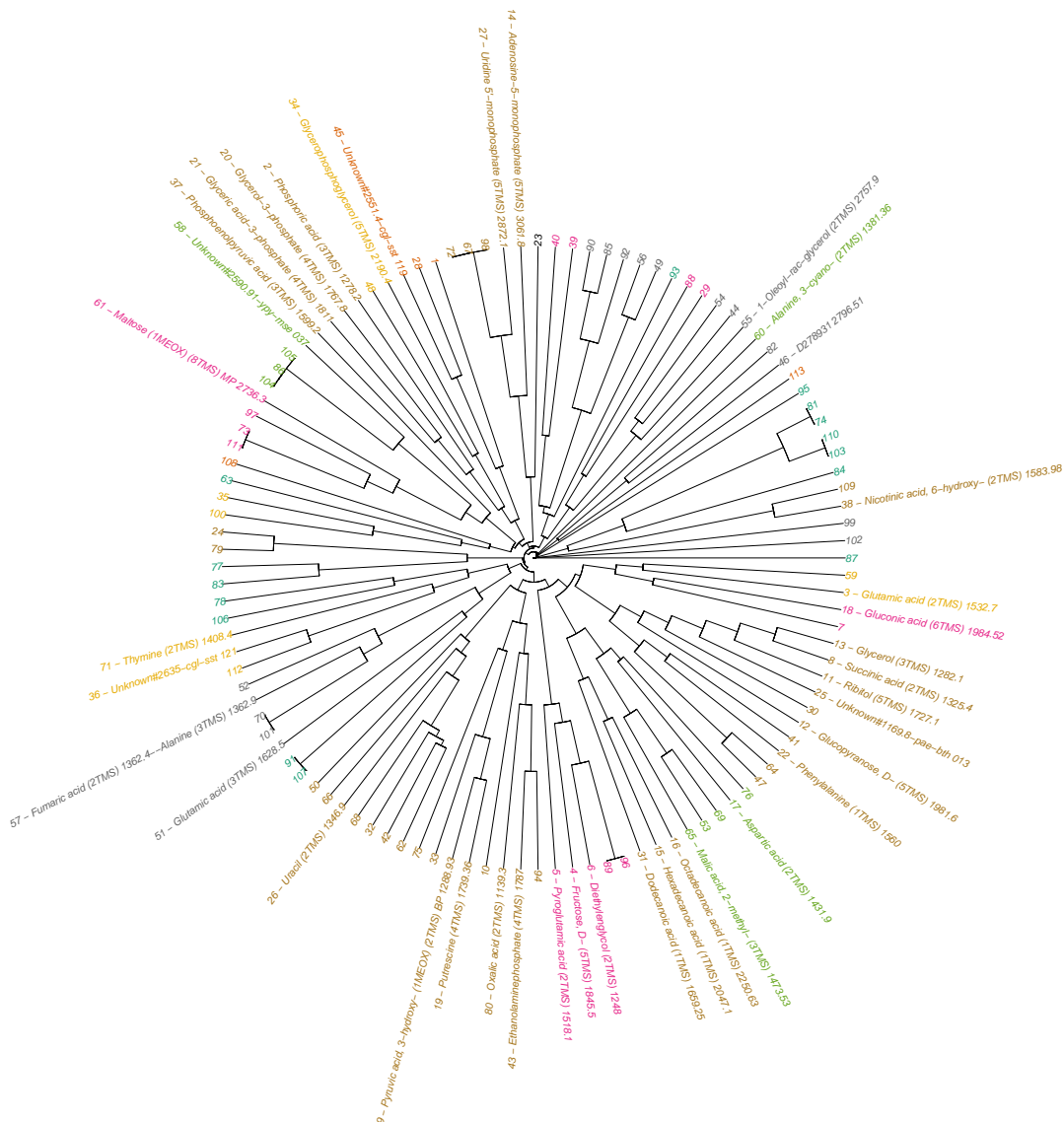


**Figure B-4:** Correlation network plot based on pseudospectra similarities of the GC-EI-MS example data set, generated with the custom similarity threshold of 0.4. Grey dots indicate non-identified features, orange dots identified ones. Labels display feature IDs, along with feature annotations, if existent. Edge widths are proportional to spectral similarity of the connected features.

Presumably, the high between-cluster similarities are due to the low resolution data and the resulting fact, that fragment with different chemical composition but same unit resolution mass cannot be distinguished.

We can also use hierarchical clustering to identify clusters of similar (pseudo-)spectra. Here, too, we have to adjust  $h$  to account for higher between-cluster similarities:

```
HCplot(pseudodistmat, h = 0.7)
```



**Figure B-5:** Circularised dendrogram as a result of agglomerative hierarchical clustering with average linkage as agglomeration criterion based on pseudospectra similarities of the GC-EL-MS example data set. Each leaf represents one feature and colours encode cluster affiliation of the features. Leaf labels display feature IDs, along with feature annotations, if existent. Distance from the central point is indicative of the height of the dendrogram.

We see that e.g. octadecanoic acid, hexadecanoic acid and dodecanoic acid form a nice cluster as well as the phosphate containing metabolites phosphoenolpyruvic acid, glyceric acid-3-phosphate, glycerol-3-phosphate and phosphoric acid itself.



It is also apparent that some features have a similarity of 1 and could therefore represent the same compound, like e.g. the features 98, 67 and 72. Those three features cluster together with AMP and UMP, suggesting that they could be nucleotides as well.

To illustrate the use of CluMSID's accessory function with this type of data, we take another look at nucleotides: A signature fragment for nucleotides in GC-EI-MS is  $m/z$  315 that derives from pentose-5-phosphates. We see this fragment in Figure 1, the spectrum of UMP (derivatised with 5 TMS groups). We can use `findFragment` to see if there are more spectra outside the cluster that feature this fragment. As we deal with unit masses, we would like to find  $m/z$  of  $315 \pm 0.5$  which we can do by setting `tolerance = 0.5/315`:

```
fragmentlist <- findFragment(apslist, mz = 315, tolerance = 0.5/315)
#> 6 spectra were found that contain a fragment of m/z 315 +/- 1587.30158730159 ppm.

vapply(X = fragmentlist, FUN = accessID, FUN.VALUE = integer(1))
#> [1] 2 14 20 21 27 35
```

We find four more spectra that contain a 315 fragment that could be investigated closer.

## Conclusion

In conclusion, every annotation method is extremely limited if only low resolution data is available and so is CluMSID. Still, we see that the tool works independently of chromatography and mass spectrometry method and even has the potential to give some good hints for feature annotation in GC-EI-MS metabolomics.

# Clustering Mass Spectra from High Resolution DI-MS/MS Data Using CluMSID

Tobias Depke

December 31, 2018

## Contents

Introduction . . . . .	C-1
Data import . . . . .	C-1
Data preprocessing . . . . .	C-1
Generation of distance matrix . . . . .	C-2
Data exploration . . . . .	C-2
Conclusion . . . . .	C-4

## Introduction

Although originally developed for liquid chromatography-tandem mass spectrometry (LC-MS/MS) data, CluMSID can also be used with direct infusion-tandem mass spectrometry (DI-MS/MS) data.

Generally, the missing retention time dimension makes feature annotation in metabolomics harder but if only direct infusion data is at hand, CluMSID can help to get an overview of the chemodiversity of a sample measured by DI-MS/MS.

In this example, we will use a similar sample (1 $\mu$ L *Pseudomonas aeruginosa* PA14 cell extract) as in the General Tutorial, measured on the same machine, a Bruker maxis<sup>HD</sup> qTOF operated in ESI-(+) mode with auto-MS/MS but without chromatographic separation.

## Data import

We load the file from the CluMSIDdata package:

```
library(CluMSID)
library(CluMSIDdata)

DIfile <- system.file("extdata",
                      "PA14_maxis_DI.mzXML",
                      package = "CluMSIDdata")
```

## Data preprocessing

The extraction of spectra works the same way as with LC-MS/MS data:

```
ms2list <- extractMS2spectra(DIfile)
length(ms2list)
#> [1] 373
```

Merging of redundant spectra is less straightforward when retention time is not available. Depending on the MS/MS method it can be next to impossible to decide whether two spectra with the same precursor  $m/z$  and similar fragmentation patterns derive from the same analyte or from two different but structurally similar ones.

In this example, we would like to merge spectra with identical precursor ions only if they were recorded one right after another. We can do so by setting `rt_tolerance` to 1 second:

```
featlist <- mergeMS2spectra(ms2list, rt_tolerance = 1)
length(featlist)
#> [1] 349
```

We see that we have hardly reduced the number of spectra in the list. If we would decide to merge all spectra with identical precursor  $m/z$  from the entire run, we could do so by setting `rt_tolerance` to the duration of the run, in this case approx. 250 seconds:

```
testlist <- mergeMS2spectra(ms2list, rt_tolerance = 250)
length(testlist)
#> [1] 75
```

The resulting number of spectra is drastically lower but the danger of merging spectra that do not actually derive from the same analyte is also very big.

### Generation of distance matrix

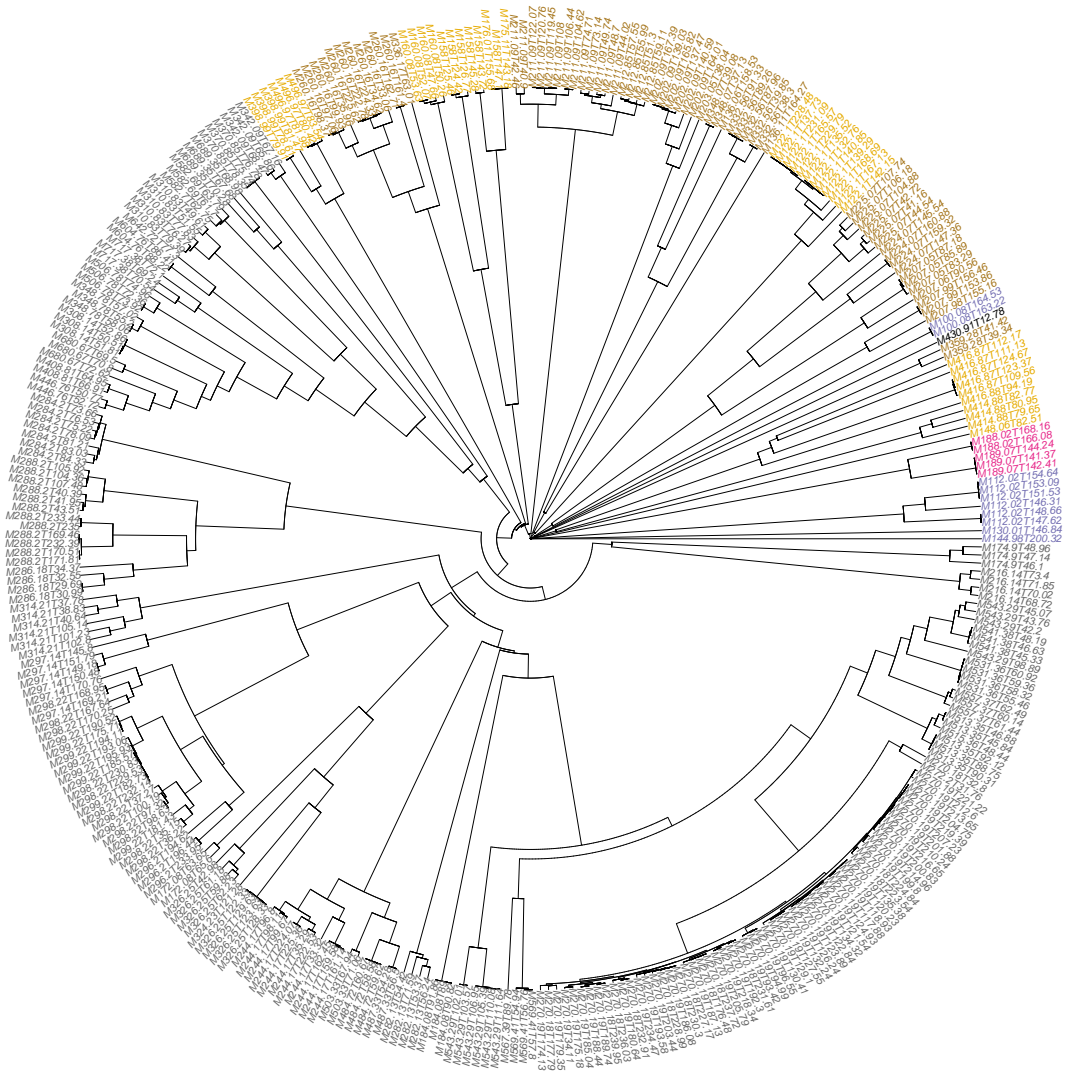
In this very explorative example, we skip the integration of previous knowledge on feature identities and generate a distance matrix right away:

```
distmat <- distanceMatrix(featlist)
```

### Data exploration

Starting from this distance matrix, we can use all the data exploration functions that `CluMSID` offers. In this example workflow, we look at a cluster dendrogram:

```
HCplot(distmat)
```



**Figure C-1:** Circularised dendrogram as a result of agglomerative hierarchical clustering with average linkage as agglomeration criterion based on  $MS^2$  spectra similarities of the DI-MS/MS example data set. Each leaf represents one feature and colours encode cluster affiliation of the features. Leaf labels display feature IDs, along with feature annotations, if existent. Distance from the central point is indicative of the height of the dendrogram.

It is directly obvious that we have some spectra that are nearly identical and thus most likely derive from the same analyte, e.g. the many spectra with a precursor  $m/z$  of 270.19. But we still see nice clustering of similar spectra with different precursor  $m/z$ , e.g. the huge gray cluster that contains a lot of different alkylquinolone type metabolites (see General Tutorial).

## Conclusion

In conclusion, CluMSID is very useful to provide an overview of spectral similarities within DI-MS/MS runs but wherever annotation is in the focus, one should not do without the additional layer of information created by chromatographic separation.

# Clustering Mass Spectra from Low Resolution LC-MS/MS Data Using CluMSID

Tobias Depke

December 31, 2018

## Contents

Introduction . . . . .	D-1
Data import . . . . .	D-1
Data preprocessing . . . . .	D-1
Generation of distance matrix . . . . .	D-2
Data exploration . . . . .	D-2
Conclusion . . . . .	D-5

## Introduction

As described in the GC-EI-MS tutorial, CluMSID can also be used to analyse low resolution data—although using low resolution data comes at a cost.

In this example, we will use a similar sample (1 $\mu$ L *Pseudomonas aeruginosa* PA14 cell extract) as in the General Tutorial, measured with similar chromatography but on a different mass spectrometer, a Bruker amaZon ion trap instrument operated in ESI-(+) mode with auto-MS/MS. In addition to introducing a workflow for low resolution LC-MS/MS data, this example also demonstrates that CluMSID can work with data from different types of mass spectrometers.

## Data import

We load the file from the CluMSIDdata package:

```
library(CluMSID)
library(CluMSIDdata)

lowresfile <- system.file("extdata",
                          "PA14_amazon_lowres.mzXML",
                          package = "CluMSIDdata")
```

## Data preprocessing

The extraction of spectra works the same way as with high resolution LC-MS/MS data:

```
ms2list <- extractMS2spectra(lowresfile)
length(ms2list)
#> [1] 1989
```

Like in the GC-EI-MS example, we have to adjust `mz_tolerance` to a much higher value compared to high resolution data, while the retention time tolerance can remain unchanged.

```
featlist <- mergeMS2spectra(ms2list, mz_tolerance = 0.02)
```

```
length(featslist)
#> [1] 525
```

We see that we have similar numbers of spectra as in the General Tutorial, because we tried to keep all parameters except for the mass spectrometer type comparable.

## Generation of distance matrix

As we do not have low resolution spectral libraries at hand, we skip the integration of previous knowledge on feature identities in this example and generate a distance matrix right away:

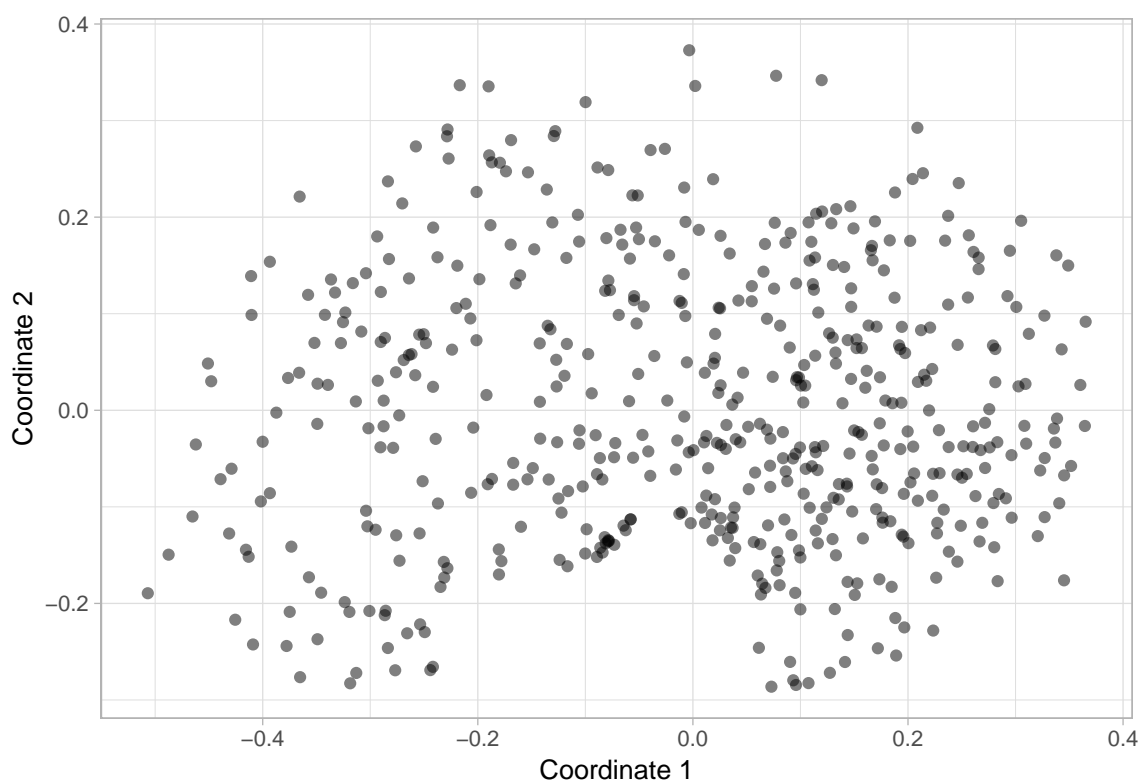
```
distmat <- distanceMatrix(featslist)
```

## Data exploration

Starting from this distance matrix, we can use all the data exploration functions that CLuMSID offers.

When we now make an MDS plot, we learn that the similarity data is very different from the comparable high resolution data:

```
MDSplot(distmat)
```

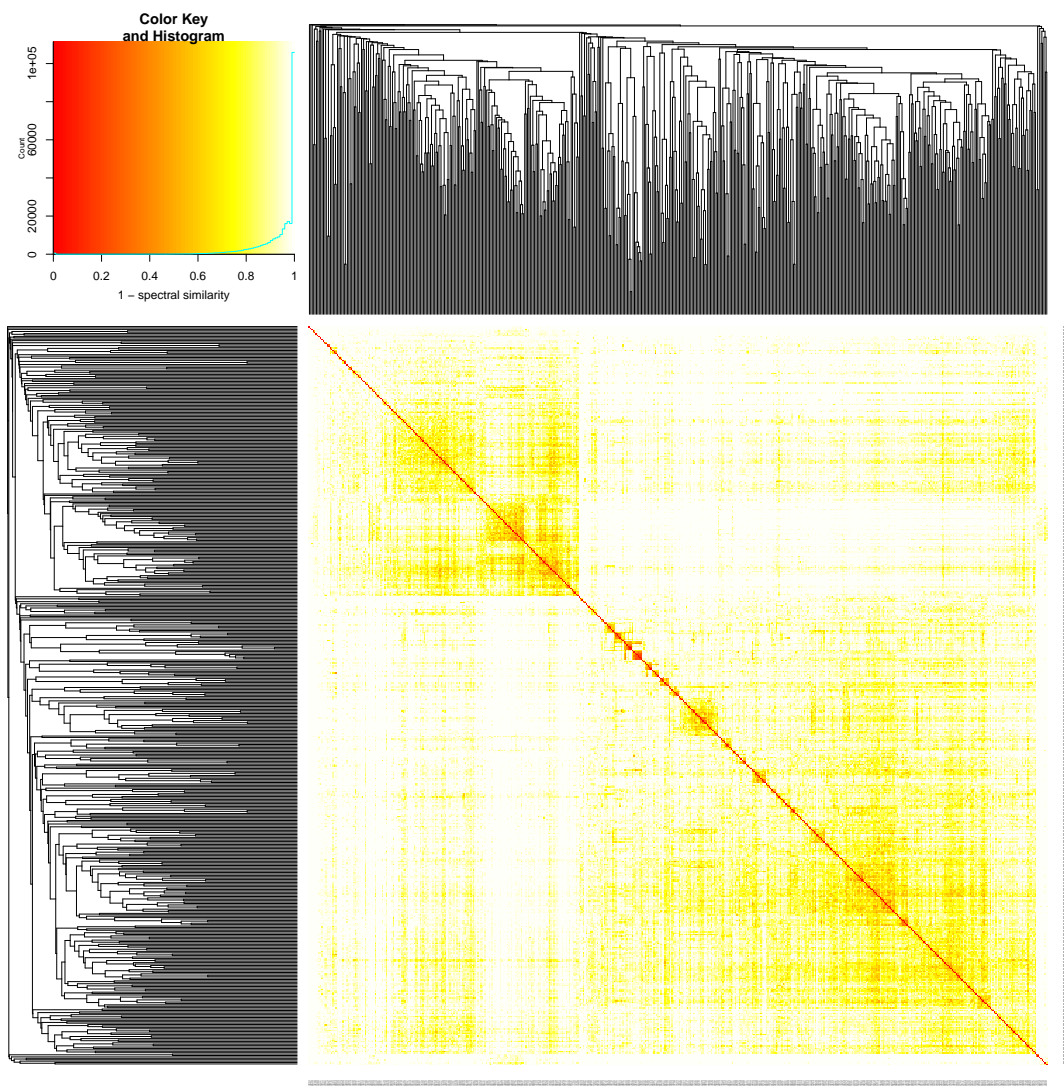


**Figure D-1:** Multidimensional scaling plot as a visualisation of MS<sup>2</sup> spectra similarities of the low resolution LC-MS/MS example data set. Black dots signify spectra from unknown metabolites.

D-2

To get a better overview of the data and the general similarity behaviour, we create a heat map of the distance matrix:

```
HCplot(distmat, type = "heatmap",  
       cexRow = 0.1, cexCol = 0.1,  
       margins = c(6,6))
```



**Figure D-2:** Symmetric heat map of the distance matrix displaying  $MS^2$  spectra similarities of the low resolution LC-MS/MS example data set, along with dendrograms resulting from hierarchical clustering based on the distance matrix. The colour encoding is shown in the top-left insert.

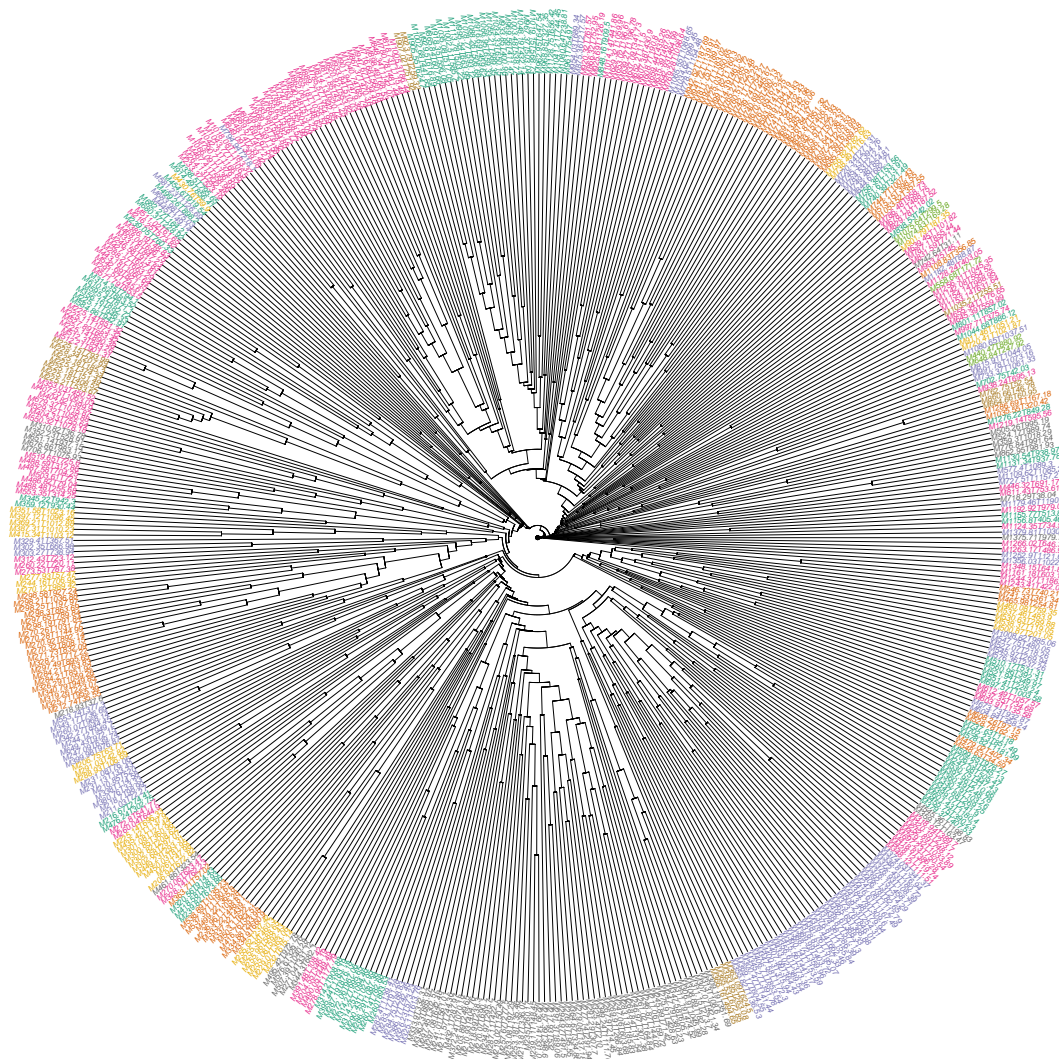
We clearly see that the heat map is generally a lot “warmer” than in the General Tutorial (an intuition that is supported by the histogram in the top-left corner), i.e. we have a higher general degree of similarity between spectra. That is not surprising as the  $m/z$  information has much fewer levels than in high resolution data and



fragments of different sum formula are more likely to have indistinguishable mass-to-charge ratios.

We also see that some more or less compact clusters can be identified. This is easier to inspect in the dendrogram visualisation:

```
HCplot(distmat, h = 0.8, cex = 0.5)
```



**Figure D-3:** Circularised dendrogram as a result of agglomerative hierarchical clustering with average linkage as agglomeration criterion based on  $MS^2$  spectra similarities of the low resolution LC-MS/MS example data set. Each leaf represents one feature and colours encode cluster affiliation of the features. Leaf labels display feature IDs. Distance from the central point is indicative of the height of the dendrogram.

## Conclusion

In conclusion, CluMSID is capable of processing low resolution LC-MS/MS data and if high resolution data is not available, it can be very useful to provide an overview of spectral similarities in low resolution data, thereby helping metabolite annotation if some individual metabolites can be identified by comparison to authentic standards. However, concerning feature annotation, high resolution methods should always be favoured for the many benefits they provide.

# Using CluMSID with a Publicly Available MetaboLights Data Set

Tobias Depke

December 31, 2018

## Contents

Introduction . . . . .	E-1
Extract MS <sup>2</sup> spectra from multiple *.mzML files . . . . .	E-1
Merge spectra with external peak list . . . . .	E-2
Add annotations . . . . .	E-3
Generate distance matrix . . . . .	E-3
Explore data . . . . .	E-3
Conclusion . . . . .	E-6

## Introduction

In this tutorial, we would like to demonstrate the use of CluMSID with a publicly available LC-MS/MS data set deposited on MetaboLights. We chose data set MTBLS433 that can be accessed on the MetaboLights web page (<https://www.ebi.ac.uk/metabolights/MTBLS433>) and which has been published in the following article:

Kalogiouri, N. P., Alygizakis, N. A., Aalizadeh, R., & Thomaidis, N. S. (2016). Olive oil authenticity studies by target and nontarget LC-QTOF-MS combined with advanced chemometric techniques. *Analytical and bioanalytical chemistry*, 408(28), 7955-7970.

The authors analysed olive oil of various provenance using reversed-phase ultra high performance liquid chromatography-electrospray ionisation quadrupole time of flight tandem mass spectrometry in negative mode with auto-MS/MS fragmentation.

As a representative pooled sample is not provided, we will combine MS<sup>2</sup> data from several runs and use the peak picking done by the authors of the study for the merging of MS<sup>2</sup> spectra. Some metabolite annotations are also included in the MTBLS433 data set which we will integrate into our analysis.

```
library(CluMSID)
library(CluMSIDdata)
library(tidyverse)
```

## Extract MS<sup>2</sup> spectra from multiple \*.mzML files

For demonstration, not all files from the analysis will be included into the analysis. Four data files from the data set have been chosen that represent olive oil samples from different regions in Greece:

- YH1\_GA7\_01\_10463.mzML: YH1, from Komi
- AX1\_GB5\_01\_10470.mzML: AX1, from Megaloxori
- LP1\_GB3\_01\_10467.mzML: LP1, from Moria
- BR1\_GB6\_01\_10471.mzML: BR1, from Agia Paraskevi

Note that these are mzML files that can be processed the exact same way as mzXML files.

Furthermore, we would like to use the peak picking and annotation data from the original authors which we can read from the file `m_mtbls433_metabolite_profiling_mass_spectrometry_v2_maf.tsv`.

First, we extract MS<sup>2</sup> spectra from the respective files separately by using `extractMS2spectra()`. Then, we just combine the resulting lists into one list using base R functionality:

```
YH1 <- system.file("extdata", "YH1_GA7_01_10463.mzML",
  package = "CluMSIDdata")
AX1 <- system.file("extdata", "AX1_GB5_01_10470.mzML",
  package = "CluMSIDdata")
LP1 <- system.file("extdata", "LP1_GB3_01_10467.mzML",
  package = "CluMSIDdata")
BR1 <- system.file("extdata", "BR1_GB6_01_10471.mzML",
  package = "CluMSIDdata")

YH1list <- extractMS2spectra(YH1)
AX1list <- extractMS2spectra(AX1)
LP1list <- extractMS2spectra(LP1)
BR1list <- extractMS2spectra(BR1)

raw_oillist <- c(YH1list, AX1list, LP1list, BR1list)
```

### Merge spectra with external peak list

First, we import the peak list by reading the respective table and filtering for the relevant information. We only need the columns `metabolite_identification`, `mass_to_charge` and `retention_time` and we would like to replace "unknown" with an empty field in the `metabolite_identification` column. Plus, the features do not have a unique identifier in the table but we can easily generate that from *m/z* and RT. Note that the retention time in the raw data is given in seconds and in the data table it is in minutes, so we have to convert. For the sake of consistency, we also change the column names. We use tidyverse syntax but users can do as they prefer.

```
raw_mtbls_df <- system.file("extdata",
  "m_mtbls433_metabolite_profiling_mass_spectrometry_v2_maf.tsv",
  package = "CluMSIDdata")

mtbls_df <- readr::read_delim(raw_mtbls_df, "\t") %>%
  mutate(metabolite_identification =
    str_replace(metabolite_identification, "unknown", "")) %>%
  mutate(id = paste0("M", mass_to_charge, "T", retention_time)) %>%
  mutate(retention_time = retention_time * 60) %>%
  select(id,
    mass_to_charge,
    retention_time,
    metabolite_identification) %>%
  rename(mz = mass_to_charge,
    rt = retention_time,
    annotation = metabolite_identification)
```

This peak list, or its first three columns, can now be used to merge spectra. We exclude spectra that do not match to any of the peaks in the peak list. As we are not very familiar with instrumental setup, we set the limits for retention time and *m/z* deviation a little wider. To make an educated guess on mass accuracy, we take a look at an identified metabolite, its measured *m/z* and its theoretical *m/z*. We use arachidic acid  $[M-H]^-$ , whose theoretical *m/z* is 311.2956:

```
## Define theoretical m/z
th <- 311.2956

## Get measured m/z for arachidic acid data from mtbls_df
ac <- mtbls_df %>%
  filter(annotation == "Arachidic acid") %>%
  select(mz) %>%
  as.numeric()

## Calculate relative m/z difference in ppm
abs(th - ac)/th * 1e6
#> [1] 28.91143
```

So, we will work with an an  $m/z$  tolerance of  $\pm 30$ ppm (which seems rather high for a high resolution mass spectrometer).

```
oillist <- mergeMS2spectra(raw_oillist,
  peaktable = mtbls_df[,1:3],
  exclude_unmatched = TRUE,
  rt_tolerance = 60,
  mz_tolerance = 3e-5)
```

### Add annotations

To add annotations, we use `mtbls_df` as well, as described in the General Tutorial:

```
f1 <- featureList(oillist)
fl_annos <- dplyr::left_join(f1, mtbls_df, by = "id")

annolist <- addAnnotations(oillist, fl_annos, annotationColumn = 6)
```

### Generate distance matrix

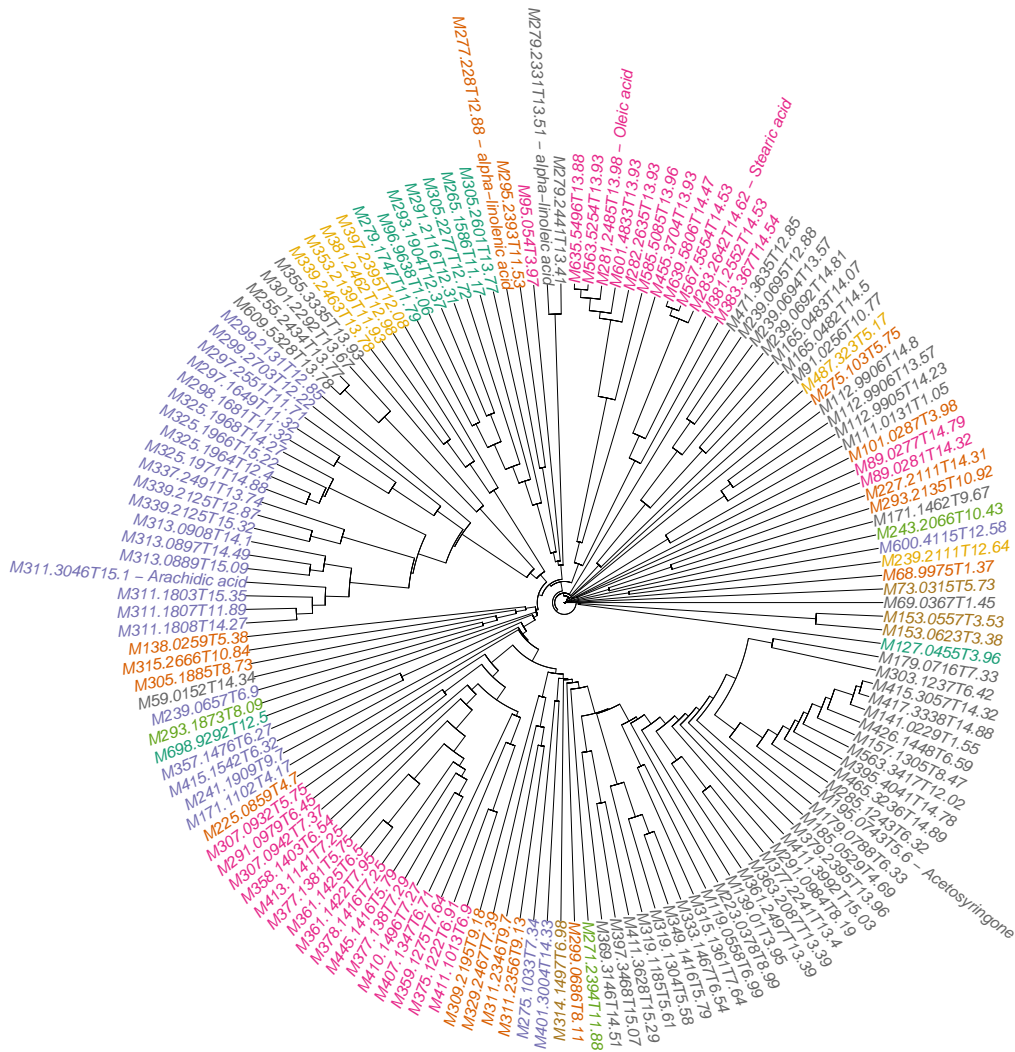
For the generation of the distance matrix, too, we use an  $m/z$  tolerance of  $\pm 30$ ppm:

```
distmat <- distanceMatrix(annolist, mz_tolerance = 3e-5)
```

### Explore data

To explore the data, we have a look at a cluster dendrogram:

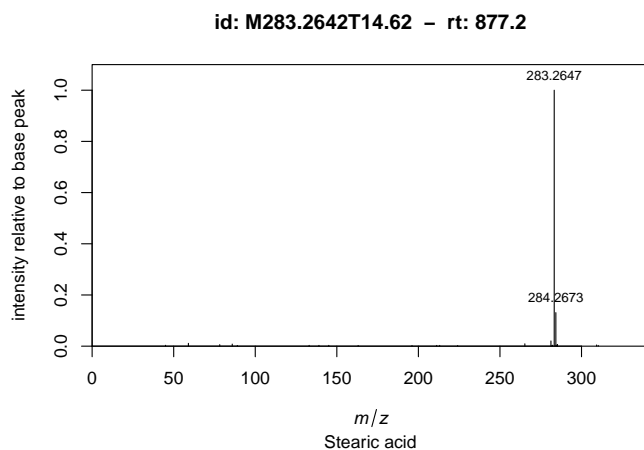
```
HCplot(distmat, h = 0.7, cex = 1)
```



**Figure E-1:** Circularised dendrogram as a result of agglomerative hierarchical clustering with average linkage as agglomeration criterion based on MS<sup>2</sup> spectra similarities of the MTBLS433 LC-MS/MS example data set. Each leaf represents one feature and colours encode cluster affiliation of the features. Leaf labels display feature IDs. Distance from the central point is indicative of the height of the dendrogram.

Since it was not in the focus of their study, the authors identified only a few metabolites. If we look at the positions of these metabolites in the cluster dendrogram, we see that the poly-unsaturated fatty acids alpha-linolenic acid and alpha-linolenic acid are nicely separated from the saturated fatty acids stearic acid and arachidic acid. We would expect the latter to cluster together but a look at the spectra reveals that stearic acid barely produces any fragment ions and mainly contains the unfragmented  $[M-H]^-$  parent ion:

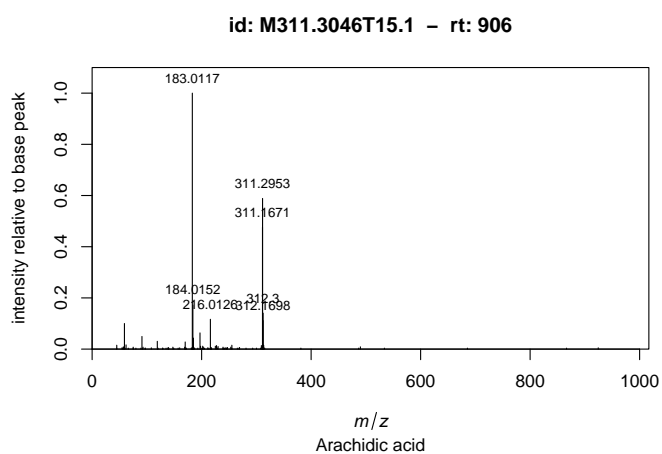
```
specplot(getSpectrum(annolist, "annotation", "Stearic acid"))
```



**Figure E-2:** Barplot for the feature M283.2642T14.62, identified as stearic acid, displaying fragment  $m/z$  on the x-axis and intensity normalised to the maximum intensity on the y-axis.

In contrast, arachidic acid produces a much richer spectrum:

```
specplot(getSpectrum(annolist, "annotation", "Arachidic acid"))
```



**Figure E-3:** Barplot for the feature M311.3046T15.1, identified as arachidic acid, displaying fragment  $m/z$  on the x-axis and intensity normalised to the maximum intensity on the y-axis.

Inspecting the features that cluster close to arachidic acid shows that many of them have an exact  $m/z$  that conforms with other fatty acids of different chain length or saturation (within the  $m/z$  tolerance), e.g. the neighbouring feature M339.2125T15.32 that could be arachidonic acid  $[M+C]^-$ .

Looking at oleic acid  $[M-H]^-$ , we see that it clusters very closely to M563.5254T13.93, whose  $m/z$  is consistent with oleic acid  $[2M-H]^-$  and some other possible adducts.

As a last example, the only identified metabolite that does not belong to the class of fatty acids is acetosyringone, a phenolic secondary plant metabolite. It forms part of a rather dense cluster in the dendrogram, suggesting high spectral similarities to the other members of the cluster. It would be interesting to try to annotate more of these metabolite to find out if they are also phenolic compounds.

### **Conclusion**

In conclusion, we demonstrated how to use CluMSID with a publicly available data set from the MetaboLights repository and how to include external information such as peak lists or feature annotations into a CluMSID workflow. In doing so, we had a look on a few example findings that could help to annotate more of the features in the data set and thereby showed the usefulness of CluMSID for samples very different from the ones in the other tutorials.



## 5 | The Alkylquinolone Repertoire of *Pseudomonas aeruginosa* is Linked to Structural Flexibility of the FabH-like PQS Biosynthesis Enzyme PqsBC

This Chapter has been published as peer-reviewed article in a scientific journal:

F. Witzgall, T. Depke, M. Hoffmann, M. Empting, M. Brönstrup, R. Müller, and W. Blankenfeldt. "The Alkylquinolone Repertoire of *Pseudomonas aeruginosa* is Linked to Structural Flexibility of the FabH-like 2-Heptyl-3-hydroxy-4(1*H*)-quinolone (PQS) Biosynthesis Enzyme PqsBC". in: *ChemBioChem* 19.14 (May 2018), pp. 1531–1544. DOI: 10.1002/cbic.201800153

### Abstract

*Pseudomonas aeruginosa* is a bacterial pathogen that causes life-threatening infections in immunocompromised patients. It produces a large armory of saturated and mono-unsaturated 2-alkyl-4(1*H*)-quinolones (AQs) or AQ *N*-oxides (AQNOs) that serve as signaling molecules to control the production of virulence factors, are involved in membrane vesicle formation and iron chelation and also have e. g. antibiotic properties. It has been shown that the FabH-like heterodimeric enzyme PqsBC catalyzes the last step in the biosynthesis of the most abundant AQ congener 2-heptyl-4(1*H*)-quinolone (HHQ) by condensing octanoyl-coenzyme A (CoA) with 2-aminobenzoylacetate (2-ABA), but the basis for the large number of other AQs/AQNOs produced by *P. aeruginosa* is not known. Here, we demonstrate that PqsBC uses different medium-chain acyl-CoAs to produce various saturated AQs/AQNOs and also biosynthesizes mono-unsaturated congeners. Further, we have determined structures of PqsBC in four different crystal forms at 1.5 Å to 2.7 Å resolution. Together with a previous report, this reveals that PqsBC adopts open, intermediate and closed conformations that alter the shape of the acyl-binding cavity and explain the promiscuity of PqsBC. The different conformations also allow

**us to propose a model for structural transitions that accompany the catalytic cycle of PqsBC, which may have broader implications for other FabH-enzymes, where such structural transitions have been postulated but never been observed.**

**Keywords:**

enzymes; protein structures, structure-activity relationships, transferases, enzyme catalysis, quorum sensing, FabH, conformational change

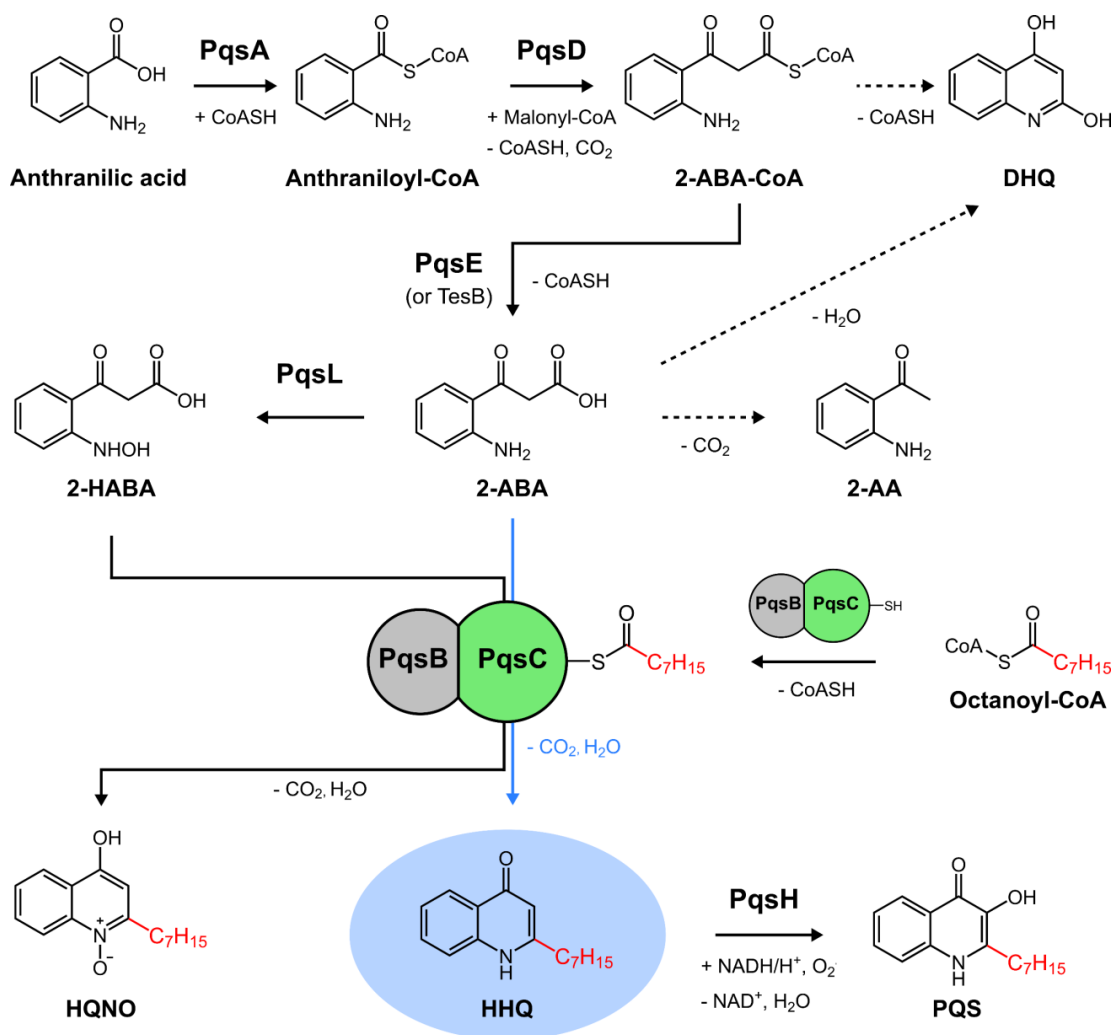
## 5.1. Introduction

The Gram-negative bacterium *P. aeruginosa* is a serious burden for public health. It is a major causative agent of hospital-acquired lung infections often leading to death in cystic fibrosis patients, of catheter-associated urinary tract infections, bacterial keratitis due to contaminated contact lenses and of wound infections [1–3]. The ability of *P. aeruginosa* to adapt to diverse environments and to cause different types of infections requires a time- and habitat-dependent coordinated expression of target genes. This is achieved by a complex intercellular communication network that controls up to 10% of the *P. aeruginosa* genome and is therefore considered as a potential therapeutic target [4–6]. Bacterial cell-to-cell communication is often referred to as “quorum sensing (QS)”, as its underlying mechanisms involve the synthesis and the secretion of small signaling molecules (autoinducers, AIs) that are sensed by specific transcriptional receptors, which are activated as soon as a certain AI level (quorum) is reached [7, 8]. Besides the production of a large arsenal of virulence factors needed for survival and persistence in the host, QS-regulated processes in *P. aeruginosa* include biofilm formation, motility and the activation of the CRISPR-Cas adaptive immune system [9, 10]. *P. aeruginosa* has three major intercellular communication circuits, which are also interlinked with each other [9, 11]. Two of these pathways, the *las* [12–14] and *rhl* [15–17] systems, rely on *N*-acyl-L-homoserine lactones, which are common autoinducers in Gram-negative bacteria [7, 8]. The third QS circuit, the *pqs* system, responds to 2-heptyl-3-hydroxy-4(1*H*)-quinolone (PQS) or its biosynthetic precursor 2-heptyl-4(1*H*)-quinolone (HHQ), which both activate the transcriptional multiple virulence factor regulator MvfR (also known as PqsR) and belong to the class of 2-alkyl-4(1*H*)-quinolones (AQs) [18–21]. HHQ- or PQS-bound PqsR complexes induce the expression of the *pqsABCDE* operon, which encodes HHQ biosynthetic proteins (Scheme 5.1) [20–23]. The aryl-coenzyme A (CoA) ligase PqsA activates anthranilate to anthraniloyl-CoA, which is then condensed with malonyl-CoA to 2-aminobenzoylacetyl-CoA (2-ABA-CoA) by PqsD [24,

25]. The thioesterase PqsE and other thioesterases of *P. aeruginosa* catalyze the hydrolysis of 2-ABA-CoA to 2-aminobenzoylacetate (2-ABA), which is the building block for the 4-quinolone ring structure of AQs [26, 27]. 2-ABA is a branch point metabolite in HHQ biosynthesis, as it is shuttled into different reaction pathways. Due to its instability, it can spontaneously decompose into 2-aminoacetophenone (2-AA) by the loss of CO<sub>2</sub> or by undergoing intramolecular cyclization into 2,4-dihydroxyquinoline (DHQ) [27, 28]. In the main biosynthetic route to HHQ, however, the heterodimeric PqsBC complex (EC 2.3.1.180) transfers 2-ABA onto octanoate to synthesize HHQ in a decarboxylative Claisen condensation after an acyl-enzyme intermediate between the active site cysteine (C129) of PqsC and octanoyl-CoA has been formed [28, 29]. In a final step, HHQ gets hydroxylated at carbon position C3 by the monooxygenase PqsH [30]. Additionally, a very recent study demonstrated that PqsBC can also use 2-hydroxylaminobenzoylacetate (2-HABA) as a second substrate, generated from 2-ABA by the monooxygenase PqsL, to yield 2-heptyl-4-hydroxyquinoline *N*-oxide (HQNO) [31], which is one of the most abundant 2-alkyl-4-hydroxyquinoline *N*-oxides (AQNOs) in *P. aeruginosa* [21, 32]. PqsB and PqsC are homologous to  $\beta$ -ketoacyl-acyl-carrier protein synthases III (FabH) of the thiolase superfamily, but unlike other FabH homologs that have a conserved His-Asn-Cys catalytic triad, PqsB lacks all of these three residues and PqsC has only a catalytic dyad composed of His-Cys (H269-C129) [28, 29, 33]. It has recently been shown by Drees *et al.* [29] that the missing asparagine side chain in PqsC is mimicked by the NH<sub>2</sub> group of 2-ABA in the reaction cycle.

In addition to HHQ and PQS, which are the most relevant AQs in terms of cell-to-cell communication, *P. aeruginosa* produces more than 50 distinct AQs/AQNOs with saturated or unsaturated alkyl side chains of different lengths [34]. The repertoire of AQs in *P. aeruginosa* is highly diverse and so are their functions. Besides their role as signaling molecules in cell-to-cell communication, AQs and AQNOs are involved in iron chelation, membrane vesicle formation, show antimicrobial activities and can manipulate the immune system of an infected mammalian host [35]. All these diverse QS-dependent and QS-independent functions of AQs and AQNOs help *P. aeruginosa* to outcompete other microorganisms in the environment or to escape cellular immune responses of the host in order to create a favorable growth habitat [35].

Although it is well known that *P. aeruginosa* mainly produces AQ and AQNO congeners with alkyl chain lengths of 7 or 9 carbon atoms [21, 34, 36, 37], the molecular mechanisms leading to this selectivity are unknown. Until now, it has only been demonstrated that the alkyl chain of HHQ/PQS originates from octanoate introduced by



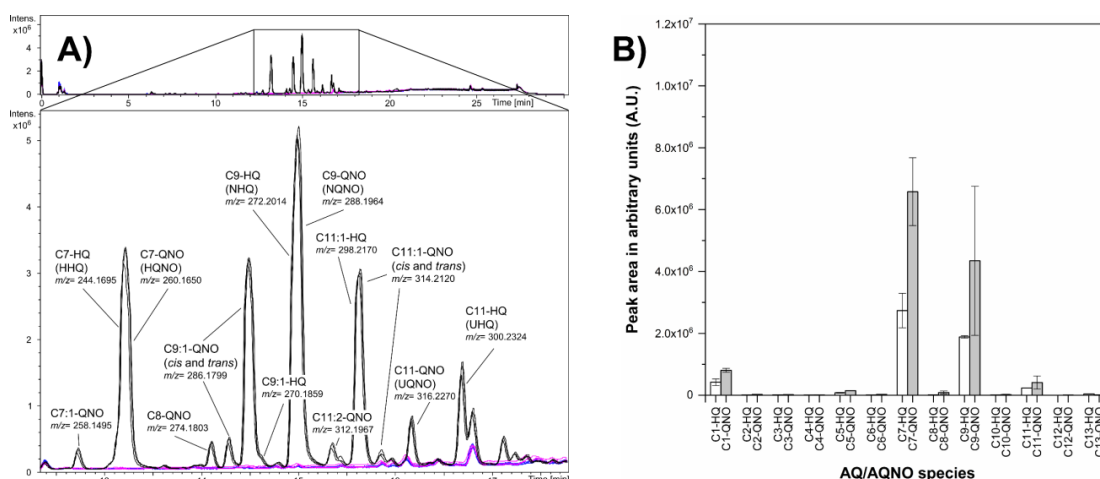
**Scheme 5.1.:** Current understanding of AQ and AQNO biosynthesis in *Pseudomonas aeruginosa*, highlighting the role of PqsBC. Dashed arrows indicate breakdown reactions. TesB is a broad-specific thioesterase that can replace PqsE in this pathway.

PqsBC [28]. However, since *P. aeruginosa* also synthesizes AQ and AQNO derivatives with shorter, longer and mono-unsaturated alkyl chains, it seems likely that PqsBC can also utilize the respective acyl-CoA primers for condensation. We therefore investigated the acyl-CoA substrate specificity of PqsBC, using enzymatic assays and feeding experiments as well as by analyzing PqsBC crystal structures from five different crystal forms. This shows that PqsBC prefers medium-chain acyl-CoAs and is the key factor of the *pqs* system that drives the AQ/AQNO distribution found in *P. aeruginosa*. Our data also provide evidence that PqsBC is directly involved in the biosynthesis of mono-unsaturated AQs/AQNOs. Crystal structures demonstrate that PqsBC exists in open, closed and intermediate conformations. Such conformations have previously been postulated for other FabH enzymes, but have never been observed. Our data therefore provide new insight into the acyl-CoA binding mechanism of FabH-like proteins.

## 5.2. Results

### 5.2.1. The diversity of AQs/AQNOs produced by *P. aeruginosa* depends strictly on PqsBC

Recently, it has been shown that the heptyl side chain of HHQ derives from octanoyl-CoA, which is incorporated by the heterodimeric FabH-like enzyme PqsBC [28]. To investigate the importance of PqsBC for the production of the whole spectrum of AQs and AQNOs, we analyzed their production in *P. aeruginosa* PA14 wildtype and in nonpolar *pqsB*<sup>-</sup> and *pqsC*<sup>-</sup> mutant strains grown in minimal medium using LC-MS. Hydroxylated species (PQS analogs) were excluded from analysis, as AQ/AQNO hydroxylation occurs downstream of PqsBC (Scheme 1) [20, 21, 30]. Towards this, it has been described previously that PQS analogues and AQNOs of the same side chain length can be distinguished both by their MS/MS fragmentation and by their chromatographic behavior despite identical mass-to-charge ratios [34, 37, 38]. In addition, PQS congeners did not give defined peaks and had different retention times than AQs and AQNOs under the chromatographic conditions used in this study. While AQs/AQNOs were completely absent in the *pqsB*<sup>-</sup> and *pqsC*<sup>-</sup> mutants, we observed a large variety of different AQs/AQNOs in the wildtype strain (Figure 5.1A). We mainly detected saturated and mono-unsaturated (*cis* and *trans*) AQs and AQNOs with odd-numbered aliphatic side chains containing 7 to 11 carbon atoms, while only traces of AQ congeners with even-numbered alkyl chains were found (Figures 5.1A, B; the nomenclature for AQs/AQNOs used here was adapted from Depke *et al.* [38]). The two most abundant

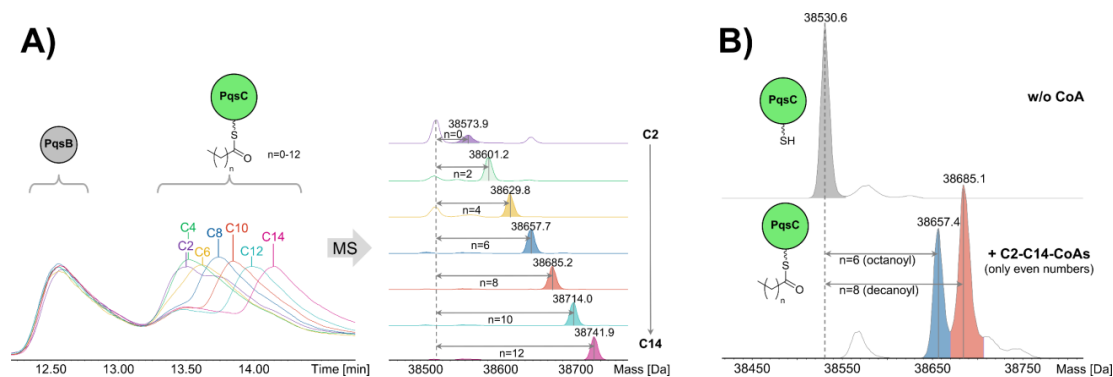


**Figure 5.1.:** AQ/AQNO profile of *Pseudomonas aeruginosa* PA14. **A)** Total ion chromatograms of bacterial cell extracts of wildtype *P. aeruginosa* PA14 (black), ED117 (*pqsB*<sup>-</sup>, blue) and ED218 (*pqsC*<sup>-</sup>, magenta). The magnified insert shows the relevant AQ/AQNO section. Chromatograms from three replicates are shown for each strain. CX:n-HQ are AQs (HHQ series) and CX:n-QNO are AQNOs (HQNO series) with X and n representing the number of carbon atoms and the number of double bonds in the aliphatic chain, respectively (e.g. C7-HQ is equivalent to HHQ). The most prominent AQ/AQNO peaks in the chromatograms of wildtype *P. aeruginosa* PA14 are annotated with the *m/z* of their [M+H]<sup>+</sup> ions. None of the annotated AQs/AQNOs is present in the mutant strains. **B)** Mean peak areas of saturated AQs (white bars) or AQNOs (gray bars) in combined cell and supernatant extracts of exponentially growing wildtype *P. aeruginosa* PA14 in arbitrary units (A.U.). Differences in AQ/AQNO levels shown in panels A) and B) are due to the fact that A) shows only cellular AQs/AQNOs, while both cellular and extracellular AQs/AQNOs were quantified for B). Since the annotation of AQs and AQNOs is based on their MS<sup>2</sup> spectra containing the characteristic radical cation with *m/z* = 159, which cannot be formed from C1-HQ/C1-QNO, the identity of C1-HQ/C1-QNO cannot be verified. The error bars are the standard deviations from three independent measurements. For comparability, the scaling is identical to Figure 5.7A.

saturated species produced by *P. aeruginosa* PA14 were C7-HQ (HHQ) or C7-QNO (HQNO) and C9-HQ (NHQ) or C9-QNO (NQNO) (Figure 5.1B), which is line with previous reports [21, 34, 36, 37]. These results suggest that PqsBC synthesizes not only HHQ but also other AQs/AQNOs by using shorter or longer acyl-CoAs.

### 5.2.2. PqsBC accepts a broad spectrum of acyl-CoAs

To corroborate the hypothesis that PqsBC synthesizes other AQ/AQNO derivatives directly, we performed proteomic assays to test whether PqsBC can also accept other saturated acyl-CoA substrates besides octanoyl-CoA. Towards this, PqsBC was incubated with different even-numbered saturated acyl-CoAs ranging from acetyl- to tetradecanoyl-CoA (C2- to C14-CoA) and protein acylation was analyzed with ESI-



**Figure 5.2.:** Acylation of PqsBC. **A)** Overlaid total ion chromatograms of PqsBC loaded with saturated acyl-CoAs, ranging from C2- to C14-CoA ( $n = 0-12$ ; only even numbers), and the respective deconvoluted mass spectra confirming acylation. **B)** Deconvoluted mass spectra of a competitive acyl-CoA loading experiment with PqsBC treated with C2- to C14-CoA ( $n = 0-12$ ; only even numbers) in equimolar ratio reveals only octanoylated and decanoylated PqsC species. The peak at 38530.6 Da corresponds to free PqsC (theoretical mass including the residual residues GPH of the affinity tag: 38530.0 Da).

MS. In all cases, PqsC was loaded with the respective acyl chain, while PqsB remained unmodified as expected (Figure 5.2A). However, incubation with acetyl-CoA, butyryl-CoA and hexanoyl-CoA resulted in incomplete PqsC modification.

In a second analysis, PqsBC was incubated with a mixture containing all acyl-CoAs at equal concentrations. Notably, PqsC was only loaded with octanoate and decanoate, revealing a clear preference for the corresponding acyl-CoAs (Figure 5.2B). To evaluate if substrate utilization can be shifted by increasing the concentration of one acyl-CoA over the other, we incubated PqsBC with a constant amount of octanoyl-CoA and equimolar or 10-fold higher concentrations of hexanoyl-, decanoyl-, dodecanoyl- or tetradecanoyl-CoA in 1:1 competition assays (Figure 5.S1 in the Supporting Information). Even with 10-fold excess of hexanoyl- or tetradecanoyl-CoA, PqsBC was only loaded with octanoate (Figures 5.S1A, D). In the case of a 10-fold excess of decanoyl-CoA, however, the loading shifted towards the decanoylated PqsBC species and only a minor fraction modified with the octanoyl moiety remained (Figure 5.S1B). Similar observations were made with dodecanoyl-CoA (Figure 5.S1C). Incubation of PqsBC with equimolar concentrations of octanoyl- and dodecanoyl-CoA, on the other hand, resulted in almost complete modification with octanoate. Together, these acyl-CoA loading experiments indicate that octanoyl- and decanoyl-CoA are the preferred acyl-CoA substrates for PqsBC and that transacylation of PqsC can be influenced to some extent by increasing the concentration of the acyl-CoA primers for C8- to C12-CoA,

**Table 5.1.:** Acyl-CoA substrate specificity of PqsBC.

Acyl-CoA	Acyl chain length	[CoA] ( $\mu\text{M}$ ) (PqsBC)	[CoA] ( $\mu\text{M}$ ) (PqsBC <sup>C129A</sup> )
Acetyl-CoA	2	n. d.	-
Butyryl-CoA	4	n. d.	-
Hexanoyl-CoA	6	205 $\pm$ 8	-
Octanoyl-CoA	8	459 $\pm$ 10	n. d.
Decanoyl-CoA	10	261 $\pm$ 19	-
Dodecanoyl-CoA	12	n. d.	-
Tetradecanoyl-CoA	14	n. d.	-
Hexadecanoyl-CoA	16	n. d.	-

PqsBC (1  $\mu\text{M}$ ) was mixed with acyl-CoA (500  $\mu\text{M}$ ) and 2-ABA (1 mM). The reaction was stopped after 20 min. The absorbance of the sample was measured at 412 nm after the addition of DTNB (2 mM) to determine the concentration of free coenzyme A (CoA). The conversion of octanoyl-CoA by PqsBC<sup>C129A</sup> was tested as a negative control. The errors are the standard deviations from three independent measurements. n. d.: no detectable CoA release. -: not tested.

suggesting that the AQ/AQNO spectrum produced by *P. aeruginosa* is also influenced by acyl-CoA availability.

We also investigated the acyl-CoA substrate specificity of PqsBC by endpoint measurements under turnover conditions in the presence of the second substrate 2-ABA, since transacylation only reflects the first reaction step and the substrate specificity could also be influenced by the overall kinetics of the enzyme. Acyl-CoA turnover was monitored by the release of CoA upon enzyme-acyl intermediate formation, which was detected spectrophotometrically with 5,5'-dithiobis-2-nitrobenzoic acid (DTNB, Ellman's reagent) at 412 nm. PqsBC displayed its highest activity with octanoyl-CoA (459  $\mu\text{M}$  CoA), followed by decanoyl-CoA (261  $\mu\text{M}$  CoA) and hexanoyl-CoA (205  $\mu\text{M}$  CoA) (Table 5.1). No CoA was detected from the short-chain acetyl- or butyryl-CoA or the long-chain dodecanoyl-, tetradecanoyl- or hexadecanoyl-CoA. These data show that PqsBC has a clear preference for medium-chain acyl-CoAs with aliphatic tails of 6 to 10 carbon atoms, exhibiting the highest activity with octanoyl-CoA.

### 5.2.3. Structure determination of PqsBC in four different crystal forms

In order to correlate the observed preference for medium-chain acyl-CoAs with structural features of PqsBC, we aimed at obtaining a crystal structure of octanoylated wildtype PqsBC and of the active site mutants PqsBC<sup>C129A</sup> or PqsBC<sup>C129S</sup> to capture an enzyme-octanoyl-CoA complex, which is expected to form before the acyl



group is transferred onto active site C129 of PqsC. Despite extensive efforts in soaking and cocrystallization experiments with wildtype PqsBC and with the mutants PqsBC<sup>C129A</sup>/PqsBC<sup>C129S</sup>, however, this was not successful. Nevertheless, we determined the structures of ligand-free PqsBC wildtype and PqsBC<sup>C129A</sup>/PqsBC<sup>C129S</sup> mutants in four different crystal forms (crystal form 1 to 4) at 1.53 Å to 2.7 Å resolution. The crystals contained two, four or eight PqsBC heterodimers in their asymmetric units (Figures 5.S2 and 5.S3; Tables 5.S1 and 5.S2). Interestingly, crystals belonging to crystal forms 1, 2 and 3 could only be obtained with protein still carrying the purification tag at the *N*-terminus of PqsC. Crystal packing analysis revealed that the purification tag mediated crucial crystal contacts to neighboring symmetry-related PqsBC molecules in the respective crystal lattices (Figure 5.S4).

Molecules from the precipitant were observed in the active site of PqsC in crystal forms 1 and 3 (Figures 5.S5A, B), and in the wildtype PqsBC structure of crystal form 2, we identified additional elongated but ambiguous electron density around the active site cysteine C129 (Figure 5.S5C). This electron density probably originated from a covalently bound co-purified ligand.

#### 5.2.4. Analysis of PqsBC in five different crystal forms reveals flexibility around the active site of PqsC

We used the Protein Structural Statistics Web Server (PSSweb) [39, 40] to compare all PqsBC heterodimers in the asymmetric units of the different crystal forms. Together with the crystal structure determined by Drees *et al.* [29] (PDB ID: 5DWZ; designated as crystal form 5), this analysis included 22 independent copies of PqsBC. PqsB shows only minor structural variations except for the *C*-terminus, some surface exposed residues and a loop region (residues 186-188) at the upper area of the dimeric interface (Figures 5.3A, C). PqsC, in contrast, is highly variable in several regions surrounding the putative acyl-CoA binding channel near active site residues C129 and H269 (Figures 5.3B, C). The variable structural elements include helix  $\alpha$ 1 (residues 33-39) with W35 and a 310 helix (residues 168-171) with R168. W35 and R168 are highly conserved in FabH enzymes (Figure 5.S6) and are known to be involved in CoA binding by intercalating the adenine moiety of CoA between the guanidine (R168) and the indole (W35) groups [41–44]. The largest structural variations in PqsC are found in a region between  $\beta$ 10 and helix  $\alpha$ 5 (residues 212-243), which is part of an area termed as “flap” in FabH from *Mycobacterium tuberculosis* (*mtFabH*), where it has been predicted to undergo large structural rearrangements during the binding of the acyl-CoA primer

(residues 201-234 in PqsC correspond to the flap region in *mtFabH*; Figure 5.S6) [45]. In contrast, amino acids in the immediate vicinity of the catalytic residues C129 and H269 are highly similar between the analyzed structures, indicating rigidity (Figures 5.3B, C).

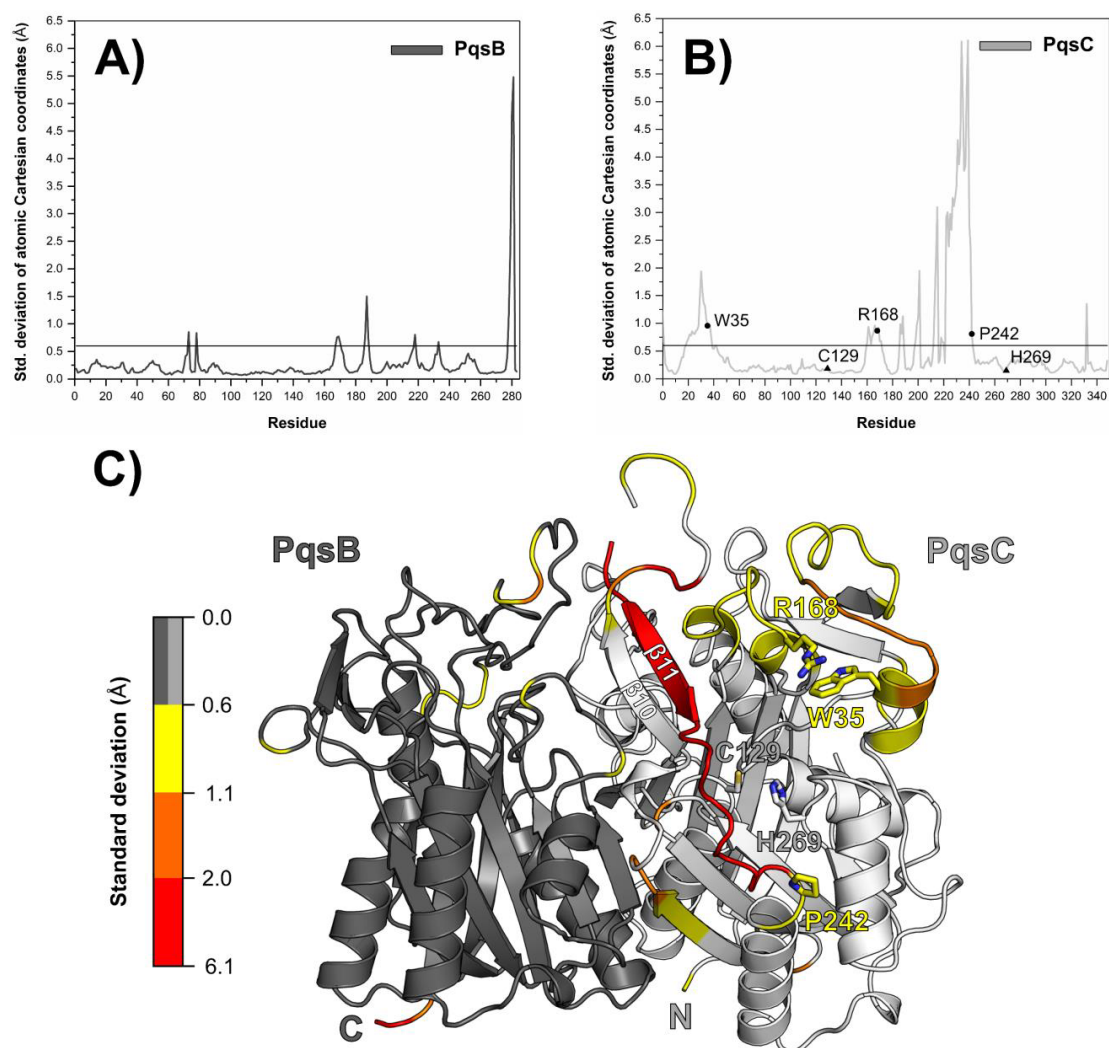
### 5.2.5. PqsC adopts open, intermediate and closed conformations

Superimposition of the 22 copies of PqsBC contained in the five different crystals forms revealed that these heterodimers can be assigned to “open”, “intermediate” and “closed” states, depending on the positions of structural elements surrounding the active site (Figure 5.4A, Table 5.S3). They show different conformations of the flap and helix  $\alpha 1$ , whose flexibility is also reflected in high B-factors (Figures 5.4A and S7). The open form is only found in crystal form 5, while crystal forms 1 to 4 belong to the intermediate or closed states (Table 5.S3).

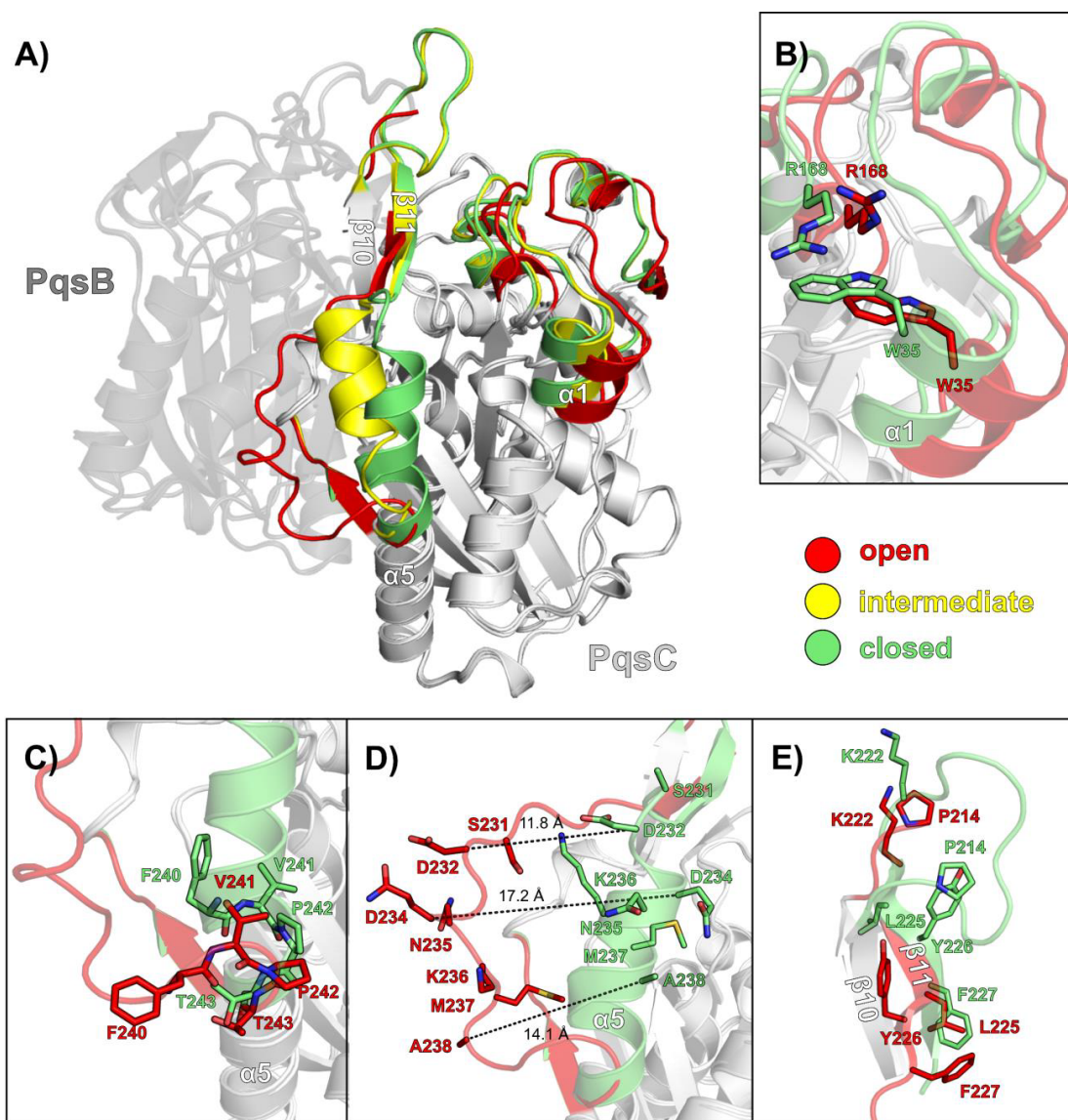
In the open form, secondary structure elements containing W35 and R168 are pushed outward relative to the closed state (Figure 5.4B). An interesting observation is that the peptide bond between V241 and P242 adopts a *cis* conformation in open PqsBC whereas it is in the *trans* conformation in the closed and intermediate structures (Figures 5.4C and 5.S8), suggesting that P242 could serve as a hinge during the catalytic cycle of PqsBC. The shape of the active site changes dramatically between the open and closed form: while residues 231 to 240 adopt a loop structure in the open state that enlarges the catalytic cavity (termed “loop A” by Drees *et al.* [29]; Figure 5.S6), they form an imperfect, kinked  $\alpha$ -helix acting as an *N*-terminal extension of helix  $\alpha 5$  in the closed form, leading to reduction in size of the substrate-binding pocket. This transition moves residues 231 to 238 closer to the active site by up to 17.2 Å (Figure 5.4D). At the same time, strand  $\beta 11$  of the mobile flap is raised, which also contributes to closing and shrinking of the binding pocket (Figure 5.4E). The distance between C $\alpha$ -atoms of Q234 (loop A/helix  $\alpha 5$ ) and W35 (helix  $\alpha 1$ ) decreases from 29 Å in the open to 11 Å in the closed conformation, highlighting the long-range structural rearrangements that accompany the interconversion of the two states.

### 5.2.6. The acyl-binding site acts as a molecular ruler that determines the acyl-CoA specificity of PqsBC

It is evident from other FabH homologs that the shape and the size of the acyl-binding channel dictate the acyl-CoA substrate specificity of these enzymes [44, 46, 47]. To investigate if the geometry of the acyl-binding site of PqsBC also reflects its speci-



**Figure 5.3.:** Structural comparison of 22 individual PqsBC heterodimers in crystal forms 1 to 5 with the Protein Structural Statistics Web Server (PSSweb) [39, 40]. **A/B)** Average standard deviations of backbone coordinates (Å) of all PqsB (A) and PqsC (B) chains by residue. The catalytic dyad (C129, H269) and conserved residues for acyl-CoA binding (W35, R168) as well as the potential hinge residue P242 are indicated as triangles and dots, respectively. **C)** Average PqsBC heterodimer in cartoon representation colored by the standard deviation (Å) for the backbone coordinates. Backbone atoms of PqsB or PqsC with an average standard deviation smaller than 0.6 Å are colored in dark and light gray, respectively. Important residues and secondary structure elements of PqsC discussed in the main text are highlighted.



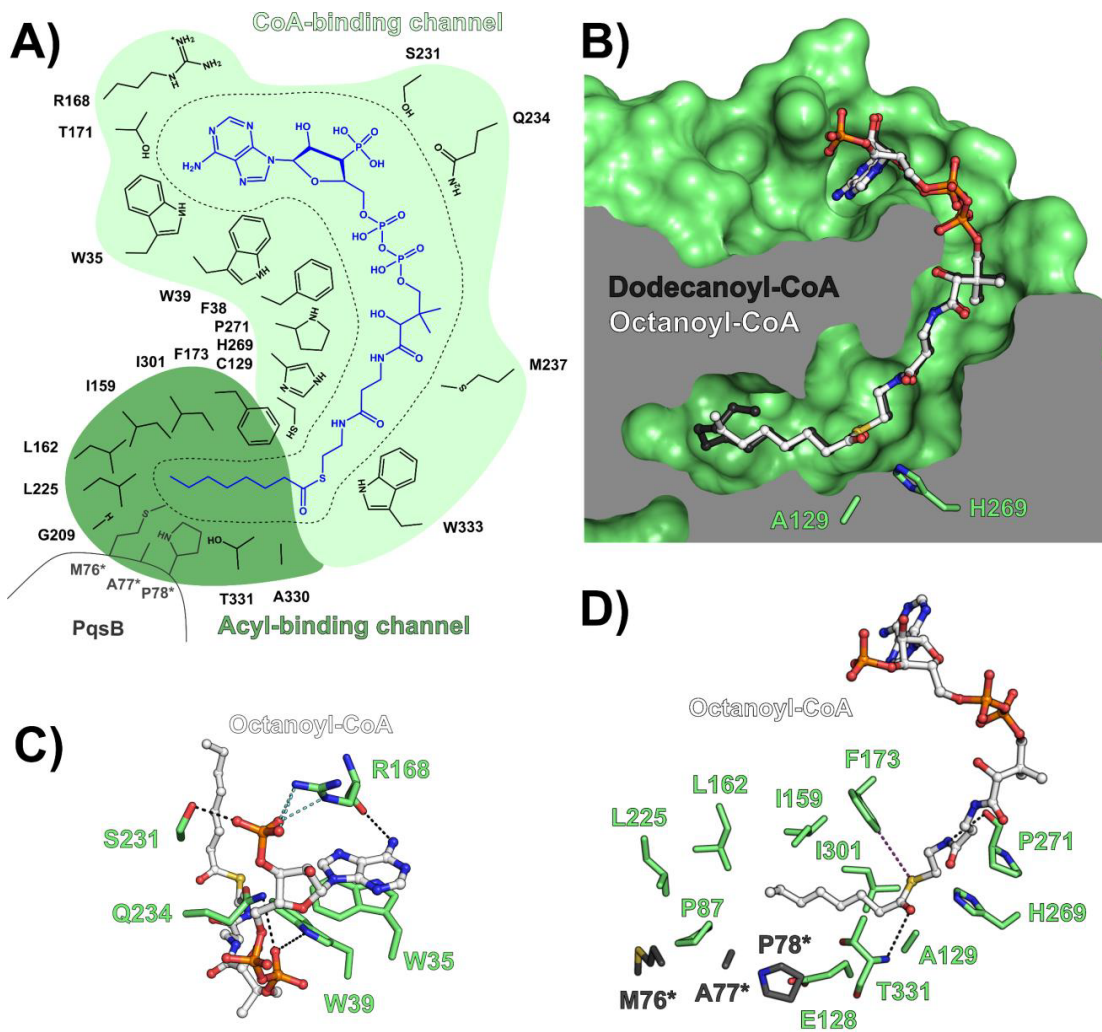
**Figure 5.4.:** Conformational changes in PqsBC. **A)** Open (red; PDB ID: 5DWZ [29]), intermediate (yellow; PDB ID: 6ET1, this study) and closed conformations (green; PDB ID: 6ET2, this study) of PqsC. PqsB is shown in dark gray and residues with only small or no changes in PqsC are shown in light gray. **B)** R168 and W35, which clamp the adenine ring of the acyl-CoA substrate in other FabH enzymes, move inwards upon closing of the active site. **C)** Residue P242 in helix  $\alpha 5$  adopts a *cis* configuration in the open form (red) and a *trans* configuration in the closed state (green). **D)** Residues S231-A238 of loop A in the open conformation of PqsC (red) are part of helix  $\alpha 5$  in the closed form (green). Dashed lines indicate distances between the same residue in the open and closed state. **E)** Amino acids of the mobile flap including  $\beta 11$  of the open conformation (green) move upwards relative to those in the closed conformation (red). Intermediate conformations are omitted in panels B-E) for clarity.

ficity for octanoyl-CoA (Table 5.1), we aligned the closed form of PqsBC<sup>C129A</sup> with *mtFabH* bound to dodecanoyl-CoA (PDB ID: 1U6S [44]) to model a complex between PqsBC<sup>C129A</sup> and octanoyl-CoA (Figure 5.5). According to this, octanoyl-CoA occupies an L-shaped pocket in PqsC, which can be subdivided into a long pantetheinate or CoA-binding tunnel and a shorter, buried acyl-binding channel (Figures 5.5A, B) that is similar to those found in crystal structures of *mtFabH*-dodecanoyl-CoA [44] or of FabH from *Micrococcus luteus* (*mlFabH*) [48]. The substrate-binding pocket is only accessible via the pantetheinate L-arm, and the two subchannels are separated by the catalytic dyad of PqsC (C129, H269) at the bottom of the binding cleft (Figures 5.5A, B).

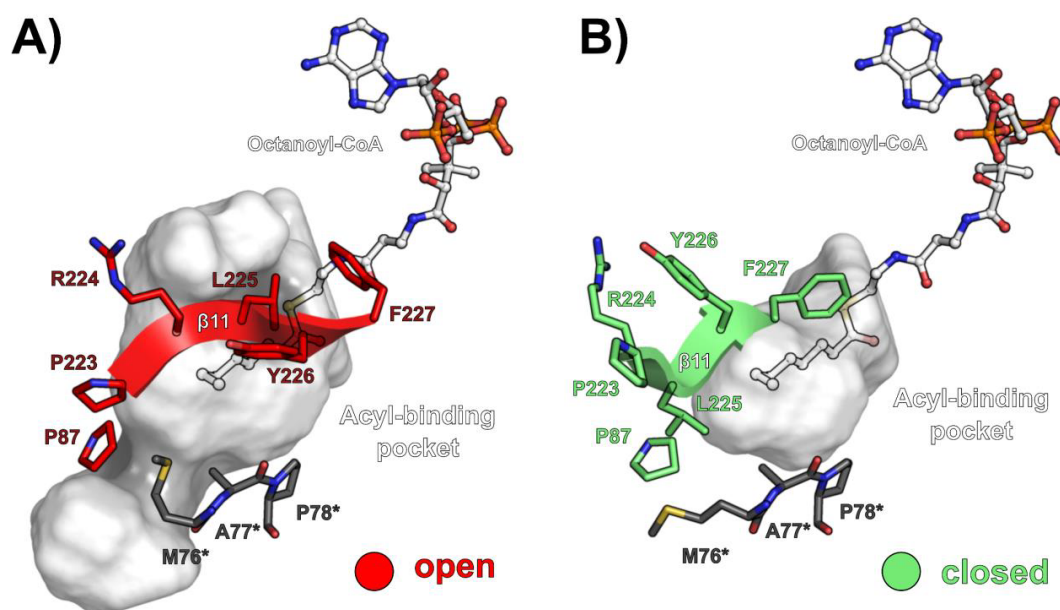
Our model suggests that the side chains of W35 and R168 sandwich the adenine ring via  $\pi$ -alkyl and  $\pi$ - $\pi$  interactions, respectively, thereby anchoring the CoA portion of octanoyl-CoA like a clamp at the protein surface of PqsC (Figure 5.5C). This stacking recognition motif is fully conserved in FabH enzymes (Figure 5.S6). The main chain carbonyl oxygen of R168 additionally fixes the adenine ring by forming a hydrogen bond with the amino group. The ribose phosphate forms a salt bridge with R168 and is also hydrogen bonded to S231. Additional hydrogen bonds are formed between the diphosphate moiety and indole nitrogen atom of W39 as well as the side chain of Q234 (Figure 5.5C). S231 and Q234 are both located on loop A in the open state (Figure 5.4D), which is one of the most divergent regions between the open and the closed conformations of PqsBC. While the pantothenate moiety is stabilized by hydrophobic interactions, the  $\beta$ -mercaptoethylamine unit is hydrogen-bonded to the backbone carbonyl oxygen of P271 (Figure 5.5D). The aromatic ring of F173 mediates a sulfur- $\pi$  interaction with the sulfur atom of the thioester bond and the main chain amino group of T331 forms a hydrogen bond with the thioester's carbonyl oxygen. The octanoyl chain is embedded in the hydrophobic acyl-binding pocket (Figures 5.5A, B, D). The distal end of this tunnel is closed by M76\*/A77\* of PqsB and by P87/L162/L225 of PqsC (Figure 5.5D).

Side views of the acyl-binding sites of PqsBC<sup>C129A</sup> models bound to octanoyl- and dodecanoyl-CoA show that the 11.5 Å long alkyl-binding channel provides enough space for medium-chain acyl-CoAs (hexanoyl- to decanoyl-CoA) and has an optimal length for octanoyl-CoA (Figure 5.5B). Alkyl chains with more than 10 carbon atoms are expected to have to fold their aliphatic tails back to fit in the hydrophobic pocket (Figure 5.5B), which would lead to unfavorably strained alkyl conformations and to disadvantageous contacts with backbone atoms, whereas shorter chains will establish





**Figure 5.5.:** Modeling of PqsBC-acyl-CoA complexes. **A)** Schematic representation of the L-shaped acyl-CoA binding site of PqsBC, which can be subdivided into a channel for the CoA moiety (CoA-binding channel) and a tunnel for the hydrophobic fatty acid part (acyl-binding channel). Octanoyl-CoA (blue) and important residues (dark gray for PqsB, indicated by asterisks; black for PqsC) lining the binding pocket are indicated. **B)** Surface cutaway side view of the L-shaped acyl-CoA binding pocket of PqsBC in complex with octanoyl-CoA (white) superposed with dodecanoyl-CoA (black). **C)** Predicted interactions between residues of the CoA-binding channel of PqsC (green) and octanoyl-CoA (white). Hydrogen bonds and salt bridges are indicated with dashed black and light blue lines, respectively. The adenine moiety of octanoyl-CoA is sandwiched by the side chains of R168 and W35. **D)** Alkyl-binding channel of PqsBC in complex with octanoyl-CoA (white). Hydrogen bonds (black dashed lines) and sulfur- $\pi$  interactions (magenta dashed lines) between PqsC and the cysteamine moiety or the thioester of octanoyl-CoA are shown. The octanoyl chain is embedded in a hydrophobic pocket formed by residues of both PqsB (dark gray) and PqsC (green).



**Figure 5.6.:** Transition from the open **A)** to the closed state **B)** reduces the size of the acyl-binding pocket from 708 to 256 Å<sup>3</sup> (white surfaces, calculated with KVFinder [49]) as a consequence of changes in the indicated residues. The model of octanoyl-CoA (white, ball-and-stick) is shown to highlight the tight fit to the acyl-CoA binding site in the closed state. PqsB and PqsC are shown in dark gray and red/green, respectively. Residues of PqsB are indicated by an asterisk.

fewer hydrophobic interactions. The modeled PqsBC-acyl-CoA complexes thus explain the substrate specificity of PqsBC towards medium-chain acyl-CoAs.

Next, we compared the alkyl-binding site of the open and closed conformations of PqsBC (Figure 5.6) to obtain insight into changes at the active site that are expected to accompany substrate binding. The acyl cavity of the open state (708 Å<sup>3</sup>) is calculated to be about three times larger than that of the closed conformation (256 Å<sup>3</sup>). As mentioned above, strand  $\beta$ 11 of the flap region also adopts a completely different conformation between the two structures: residues F227 and L225 of PqsC are located at the side of the acyl cavity in the open state (Figure 5.6A) but cover the acyl cavity from top and front in the closed conformation (Figure 5.6B). Other residues with significantly different orientations are M76\* of PqsB and P87 of PqsC (Figures 5.6A, B). In the open form, M76\* and P87 face each other and are oriented upwards to provide space for the alkyl chain (Figure 5.6A), while they are turned downwards in the PqsBC<sup>C129A</sup>-octanoyl-CoA model, thereby closing the cavity at the distal end of the channel (Figure 5.6B).

### 5.2.7. The AQ/AQNO spectrum of *P. aeruginosa* depends on acyl-CoA availability to PqsBC

The analysis of the structure (Figure 5.5) and the results of the enzymatic *in vitro* assays outlined above (Figure 5.2 and Table 5.1) explain PqsBC's selectivity for medium-chain acyl-CoAs but also suggest that the enzyme should not be able to distinguish between odd- and even-numbered acyl-CoA substrates, provided they fall into the correct size regime. At equal substrate concentrations, this should lead to a bell-shaped profile for AQ/AQNO derivatives of different alkyl size in *P. aeruginosa*, suggesting that the observed deviations from this distribution in cell cultures (Figure 5.1B) originate from acyl-CoA availability as a consequence of the fatty acid metabolism in the bacterium. To test this, we performed feeding experiments with exogenously supplied individual fatty acid precursors from acetic acid to tetradecanoic acid in wild-type *P. aeruginosa* PA14 cultures at a fixed concentration of 2 mM and quantified the resulting relative AQ/AQNO levels (Figure 5.7A). In contrast to non-supplemented cultures (Figure 5.1B), the AQ/AQNO profile was indeed bell-shaped with a maximum at C6-HQ/C7-QNO, indicative of a preference of heptanoyl- and octanoyl-CoA. Further, the profile is characterized by a sharp increase from C3-HQ/QNO to C4-HQ/QNO and a steep drop after C9-HQ/C9-QNO. These results are in line with the preference of PqsBC for medium-chain acyl-CoAs and show that the AQ/AQNO distribution in unfed *P. aeruginosa* reflects the availability of octanoyl- and decanoyl-CoA as the C2-fragment nature of fatty acid metabolism.

Unexpectedly, we observed that addition of fatty acids longer than decanoic acid did not result in an increase of the respective AQ/AQNO but in elevated levels of AQs/AQNOs with shorter alkyl chains (Figure 5.S9). Apparently, long-chain fatty acids (> C10) run through the  $\beta$ -oxidation cycle until they reach an alkyl chain length that falls into the substrate range of PqsBC, and are then channeled into the AQ/AQNO biosynthetic pathway.

### 5.2.8. PqsBC also produces mono-unsaturated AQs/AQNOs

In addition to saturated AQs/AQNOs, *P. aeruginosa* also produces mono-unsaturated congeners (Figure 5.S10; C $n$ :1-HQ/QNO, where  $n$  is the number of carbon atoms of the aliphatic side chain) [21, 34, 36, 38]. It has been shown that the double bond is located between the  $\alpha$  and  $\beta$  carbons and that the respective *cis* and *trans* isomers have different chromatographic retention times [34]. Recently, it was found that the *trans*

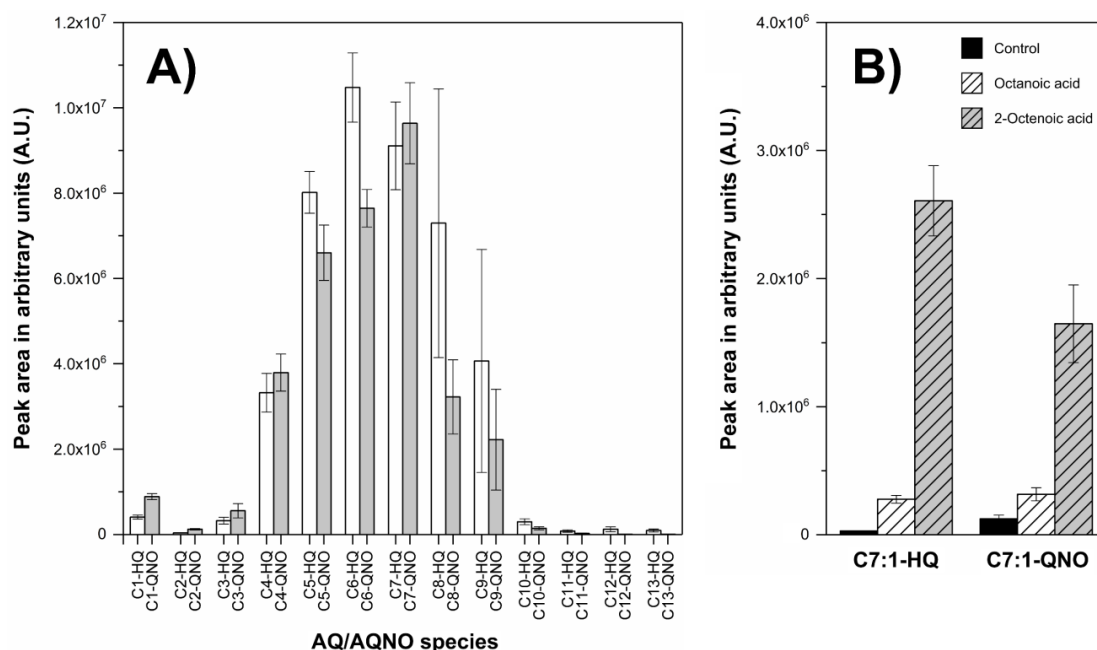


isomer of mono-unsaturated C9:1-QNO exhibits high bacteriostatic activity against *S. aureus* MRSA strains, while the respective *cis* isomer and its saturated AQ congener C9-HQ were inactive [50]. Although mono-unsaturated AQs (C9:1-HQ) have already been described by Wells in 1952 [51], there is currently no information whether the double bond is already introduced by PqsBC-mediated incorporation of an unsaturated acyl-CoA precursor or whether it is the consequence of an unidentified desaturase that acts downstream of AQ/AQNO biosynthesis.

In the course of our feeding experiments, we found that addition of octanoic acid to the growth medium not only increased the level of saturated C7-HQ but also that of its mono-unsaturated congener C7:1-HQ (Figure 5.S11). To shed further light onto this, we conducted an analogous experiment by adding fully deuterated octanoic-d15 acid as fatty acid precursor to the culture medium of wildtype *P. aeruginosa* PA14. We detected mass shifts of 15 Da for both C7-HQ and C7-QNO as expected, but also observed mass shifts of 13 Da for the respective mono-unsaturated C7 congeners, demonstrating that octanoic-d15 acid was also incorporated into C7:1-HQ or C7:1-QNO either prior or after unsaturation (Figure 5.S12). To discriminate between these possibilities, cultures were supplemented with a mixture of *cis/trans*-2-octenoic acid (2 mM). As shown in Figure 5.7B, C7:1-HQ/QNO levels strongly increase after addition of this compound, indicating that PqsBC can also use this unsaturated precursor. It therefore seems likely that C7:1-HQ and C7:1-QNO observed in unfed cultures derive from octenoyl-CoA that arises as an intermediate of the  $\beta$ -oxidation pathway.

### 5.3. Discussion

Alkylquinolones and their *N*-oxides (AQs/AQNOs) play an important role in *P. aeruginosa* and it has long been known that abundant amounts and a diverse spectrum of these compounds are produced by the bacterium [21, 34–37]. Here, we have used targeted metabolomics, feeding experiments, proteomics, biochemical assays and structural biology to investigate the role of the FabH-like heterodimeric PqsBC in the production of AQ/AQNO derivatives. Our study shows that PqsBC is a promiscuous enzyme that generates many of these compounds directly by utilizing acyl-CoA substrates with saturated and unsaturated alkyl chains of different length. PqsBC's preference for medium-chain acyl-CoAs provides a rationale why the most abundant AQs/AQNOs found in *P. aeruginosa* culture contain a C7- or a C9-alkyl chain, but it fails to explain why derivatives of uneven-chain acyl-CoAs and of smaller substrates with significant



**Figure 5.7.:** The AQ/AQNO spectrum of *P. aeruginosa* depends on the availability of fatty acids and 2-octenoic acid is the precursor of C7:1-HQ and C7:1-QNO. **A)** Mean peak area in A.U. of Aqs (white bars) and AQNOs (gray bars) in combined cell and supernatant extracts generated from cultures of exponentially growing wildtype *P. aeruginosa* PA14 fed with 2 mM of the respective precursor fatty acid (e.g. octanoic acid for C7-HQ). Error bars show the standard deviation from three independent measurements. Note that the AQ/AQNO levels of C4-HQ/C4-QNO to C9-HQ/C9-QNO after precursor feeding are higher than those of unfed cultures (Figure 5.1B) and show a bell-shaped profile, indicative of incorporation of the supplied fatty acids into Aqs/AQNOs. **B)** Mean peak area of *cis/trans* C7:1-HQ and C7:1-QNO in cell extracts of exponentially growing wildtype *P. aeruginosa* PA14 fed with EtOH (control, black bars), with octanoic acid (2 mM, white cross-striped bars) or with 2-octenoic acid (2 mM, grey cross-striped bars). The error bars represent the standard deviation from three independent measurements. The large increase in C7:1-HQ/C7:1-QNO upon the addition of 2-octenoic acid suggests that the mono-unsaturated C7-alkyl chain originates from 2-octenoic acid.

turnover in *in vitro* assays, such as hexanoyl-CoA, are underrepresented *in vivo*. Feeding experiments with C2- to C14-fatty acids suggest that these AQs/AQNOs are absent because the respective fatty acid/acyl-CoA precursors are not available for AQ/AQNO biosynthesis.

The level and composition of acyl-CoAs is dynamic and influenced by several enzymes involved in fatty acid degradation (Fad) and biosynthesis (Fab) [52], some of them still uncharacterized in *P. aeruginosa*. The initial acyl-CoA pool is defined by the substrate specificity of multiple fatty acyl-CoA ligases (FadDs), which prime fatty acids with CoA for the  $\beta$ -oxidation cycle after uptake [53, 54]. Acyl-CoA dehydrogenases (FadE) introduce the  $\alpha$ - $\beta$  double bond to produce the respective unsaturated enoyl-CoAs in the first step of  $\beta$ -oxidation, before additional enzymes (FadB, FadA) catalyze their further breakdown. A recent study revealed that the transcriptional regulator PsrA counteracts the conversion of acyl-CoAs to enoyl-CoAs by repressing the transcription of the FadE-homolog PA0506, thereby increasing the level of saturated acyl-CoAs that are then available for AQ/AQNO biosynthesis [55]. Interestingly, octanoyl-CoA, which is the preferred substrate of PqsBC as demonstrated here, has a special role in the fatty acid metabolism of *P. aeruginosa*: in addition to degradation via  $\beta$ -oxidation, it can be condensed with malonyl-ACP to produce the rhamnolipid precursor  $\beta$ -ketodecanoyl-ACP by the enzyme PA3286, which directly links  $\beta$ -oxidation with *de novo* fatty acid and rhamnolipid biosynthesis [56, 57]. In addition to Fad or Fab enzymes, there are also acyl carrier proteins (ACPs) and enzymes involved in the biosynthesis of cell wall components or of rhamnolipids that also make use of the fatty acid pool. This highlights that PqsBC competes with many other fatty acid-metabolizing enzymes to branch medium-chain acyl-CoAs into AQ/AQNO biosynthesis. In this context, it may be worth noting that the *N*-acyl-L-homoserine lactone synthases RhlI of the *rhl* system and LasI of the *las* circuit prefer ACP-loaded C4- and 3-oxo-C12 acyl chains, respectively [58, 59]. As PqsBC selectively uses medium-chain acyl-CoAs, it is probably ensured that the *rhl*, *las* and *pqs* quorum sensing circuits access different fatty acid subpools and do not steal acyl-substrates from each other. Further, AQs with C7- or C9-alkyl chains are the most potent co-inducers of PqsR and the best substrates of PqsH [30, 60, 61]. This shows that PqsBC, PqsR and PqsH are perfectly attuned to one another, suggesting co-evolution not only of the *rhl*, *las* and *pqs* systems but also within the *pqs* system itself.

The finding that the AQNO profile of *P. aeruginosa* cultures in both the presence and the absence of exogenous fatty acids follows the same trend as that of AQs (Figures 5.1B

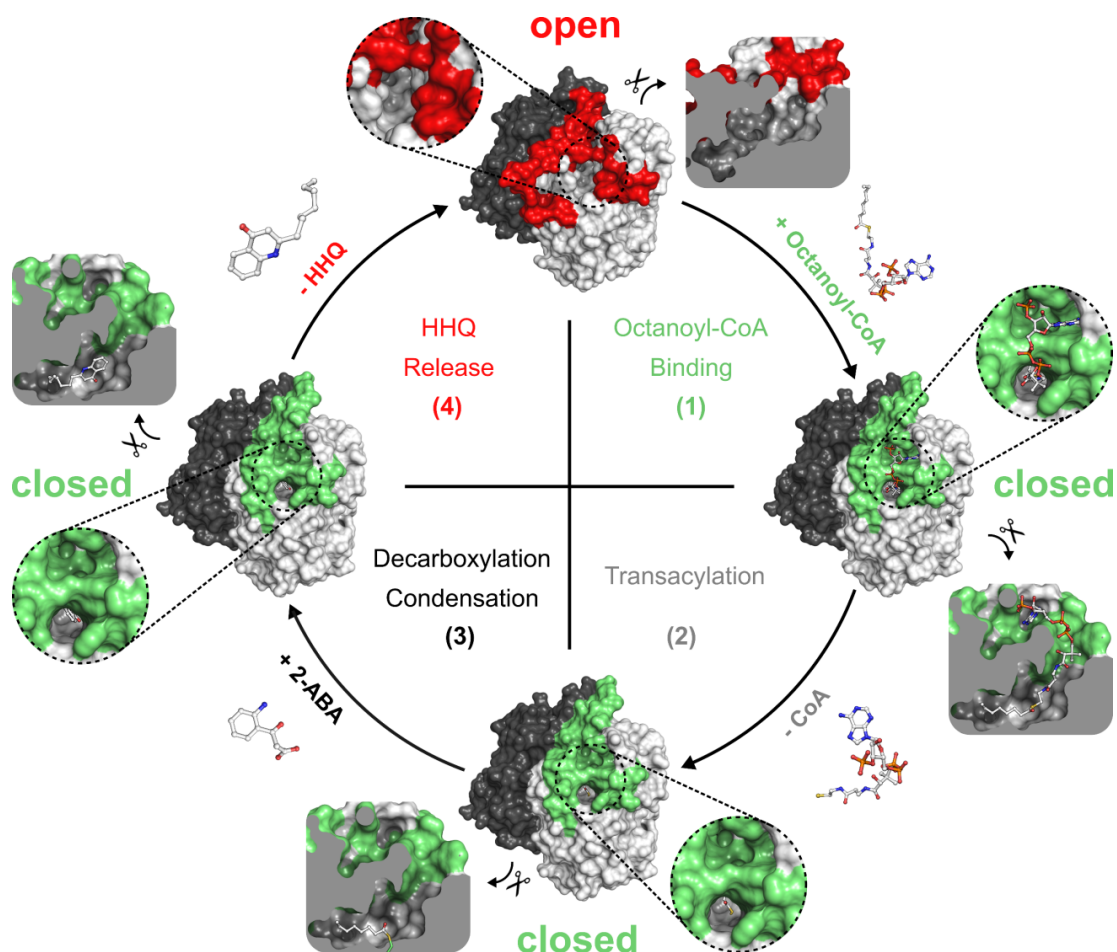
and 5.7A) supports the findings of a recent study [31] that PqsBC also catalyzes the synthesis of AQNOs from 2-HABA (produced by PqsL) and from acyl-CoAs. Our study also provides evidence that in addition to saturated acyl-CoAs, PqsBC also accepts  $\alpha,\beta$ -unsaturated acyl-CoAs, which are most probably formed in the first step of the  $\beta$ -oxidation cycle, to produce the respective mono-unsaturated AQs/AQNOs directly.

The substrate preferences of PqsBC are rooted in its three-dimensional structure, and we found that PqsBC in its closed conformation contains a hydrophobic alkyl-binding chamber that provides sufficient space to accommodate octanoyl- and decanoyl-moieties in addition to shorter alkyl chains. These shorter substrates will establish fewer interactions with the active site whereas larger molecules will have to adopt unfavorable geometries, explaining the bell-shaped distribution of AQs/AQNOs when cultures of *P. aeruginosa* were supplied with large amounts of C2- to C14-fatty acids. Apparently, the acyl-binding tunnel of PqsBC serves as a “molecular ruler” that restricts the alkyl chain length of the substrate to 10 or fewer carbon atoms. While catalytic residues are only found in PqsC, the acyl-binding tunnel is in part also built by residues from PqsB, showing that this otherwise inactive monomer not only acts as a chaperone for PqsC but also plays a role in shaping the acyl-CoA substrate selectivity of PqsBC.

Importantly, our extensive sampling of crystallization conditions yielded four new crystal forms, of which two have large asymmetric units containing four and eight copies of the protein complex. Together with the structure recently published by Drees *et al.* [29], this allowed us to compare 22 PqsBC heterodimers in different crystallographic environments. Structural analysis showed that PqsB is very similar in all heterodimers, at the same time revealing a structural plasticity of PqsC that manifests itself in open, intermediate and closed conformations of PqsC. With this, our study provides the first direct experimental corroboration for the existence of distinct conformational states in a FabH-like enzyme. Earlier studies have reported only closed conformations, suggesting that the closed state is in general thermodynamically favored [41, 42, 44, 46–48, 62–64]. Evidence for the dynamic nature of these enzymes and for the existence of open conformations has been rather indirect until now [43, 45, 65, 66]. For example, an apo structure of *Escherichia coli* FabH (*ecFabH*, PDB ID: 1HNK) was found to be unstructured in regions implicated in ligand binding [43], and kinetic as well as crystallographic analyses with alkyl-CoA disulfide inhibitors suggested that this unstructured *ecFabH* resembles an open form that closes upon binding of the CoA-ligand [65]. This view was later refined to the “open state model” by studies on *mtFabH* showing that a mutant with a blocked acyl-binding pocket could still bind an inhibitor, but this was

only possible if the enzyme opened by movement of the flap region and exposed its binding site [45, 67]. The open state model is further supported by molecular dynamics (MD) simulations of the FabH-like PQS biosynthesis enzyme PqsD and of *ecFabH*, which demonstrated large motions of the flap region and in the areas with conserved residues for adenine stacking [66, 68]. These MD profiles are strikingly similar to the structural fluctuations observed in the ensemble of PqsC crystal structures shown in Figure 5.3B. We are therefore confident that the open, intermediate and closed conformations of PqsBC identified here are not crystallographic artifacts but resemble trapped “conformational snapshots” that occur during acyl-CoA binding of PqsBC.

Comparison of the three conformational states of PqsBC enables us to speculate about a structural mechanism for the opening and closing of PqsBC during catalysis, which is summarized in Figure 5.8. At the outset of the catalytic cycle, substrate-free PqsBC is expected to be in the open form with P242 in the *cis*-conformation. The adenine moiety of the acyl-CoA substrate would then bind between the conserved W35 and R168 (Figure 5.5C) and *cis-trans* isomerization at V241-P242 would accompany the transition of loop A in the open state to the elongated, kinked helix  $\alpha 5$  observed in the closed conformation (Figures 5.4C, D), which concomitantly pushes strand  $\beta 11$  upward (Figure 5.4E). This allows residues S231 and Q234 of the extended helix  $\alpha 5$  to establish favorable interaction with the CoA moiety of the substrate (Figure 5.5C), pulling the structural elements containing the W35/R168 clamp towards the active site (Figure 5.4B). The active site will close further by movement of strand  $\beta 11$  from PqsC, which slides over the acyl-binding pocket like a zipper. Due to this displacement, F227 (PqsC) can then shield the pocket from the top like a lid and L225 (PqsC) moves to the far end of the tunnel, thereby pushing P87 (PqsC) and M76\* (PqsB) downwards to establish the distal walls of the channel (Figures 5.6A, B). As a consequence, the bound substrate would be locked in the closed conformation to undergo the transacylation reaction with C129. Interestingly, the finding that PqsBC was also loaded with long-chain dodecanoyl or tetradecanoyl units (Figure 5.2A), which are expected not to fit into the acyl-binding cavity of PqsC, seems to indicate that full closing is not required for this reaction. However, competition experiments (Figure 5.S1) and enzymatic reactions including the second substrate 2-ABA (Table 5.1) suggest that transacylation in the open state of these long-chain primers is less efficient and blocks structural changes required for the second half-reaction of the catalytic cycle. This second half-reaction likely is initiated by dissociation of CoA, which would not require large structural perturbations, since CoA is bound mostly at the surface of PqsC. The release of CoA



**Figure 5.8.:** A structural, hypothetical model of the catalytic cycle of PqsBC. (1) After octanoyl-CoA binds to the enzyme, PqsBC (PqsB: dark gray; PqsC: light gray) transitions from the open (red) into the closed state (green). (2) The octanoyl chain is transferred to active site C129 of PqsC, followed by the release of CoA. (3) 2-ABA enters the active site and undergoes decarboxylative condensation with the octanoyl-PqsC intermediate to produce HHQ. (4) The re-opening of PqsBC leads to the release of HHQ (or of the Claisen condensation intermediate that finally undergoes cyclization to HHQ). A surface cutaway side view of the L-shaped acyl-CoA binding channel is shown at each reaction step.

is expected to open a channel through which the second substrate 2-ABA can enter the active site while the enzyme would remain in the closed state. The AQ/AQNO product forms by decarboxylative condensation followed by intramolecular cyclization of the condensation intermediate. While it is at present unclear if this final cyclization proceeds within the active site and if it requires catalysis [29], the decarboxylative condensation releases the acyl group from C129. As a consequence, forces that hold the enzyme in the closed state would be interrupted, allowing re-opening and dissociation of the product.

A similar model has already been suggested for *mtFabH* [45, 67]. It has to be noted, however, that P242 of PqsC is not conserved in other FabH enzymes such as *mtFabH* (Figure 5.S6) and therefore, due to the potential importance of its *cis/trans* isomerization, details of the hypothetical model purported here may not apply to all FabH proteins. This is not surprising because of the large diversity of this family and the special role that PqsBC takes within these proteins. Most FabH enzymes are homodimers involved in fatty acid biosynthesis, have a catalytic triad, and utilize acyl-CoA primers and malonyl-ACP to produce 3-ketoacyl-ACPs. In contrast, PqsBC is heterodimeric, requires only a catalytic dyad, and produces a bicyclic AQ product from acyl-CoA and 2-ABA. The flexibility of the flap region in most other FabH enzymes may therefore be encoded in other residues, and Sachdeva *et al.* [45] have previously suggested that conserved glycine residues that flank the flap region in other FabH enzymes but not in PqsBC (Figure 5.S6) play a role in opening and closing the active site. Although it is obvious from the crystal structures that the orientation of P242 has large structural consequences for the active site of PqsBC, further in-depth investigation will be necessary to evaluate its potential role in the catalytic mechanism of PqsBC.

Due to its importance in controlling virulence and because of its unique occurrence in *P. aeruginosa*, the *pqs* system is currently evaluated as a drug target in novel anti-virulence strategies [69–71]. The finding that PqsBC adopts distinct conformations that have been resolved here to high resolution and could be addressed selectively may offer new opportunities for these drug discovery programs.

## 5.4. Experimental Section

Full experimental procedures can be found in the supporting online information.



### Chemicals and bacterial strains

UCBPP-PA14 was used as the wildtype strain for all experiments. The nonpolar mutant strains ED117 (*pqsB::TnphoA*) and ED218 (*pqsC::Kan*) were kindly provided by Eric Déziel and co-workers [28]. 2-ABA was kindly provided by Steffen L. Drees and Susanne Fetzner [27]. All acyl-CoAs were purchased from Larodan with the exception of acetyl-CoA, which was acquired from Sigma-Aldrich. DTNB was obtained from Thermo Fisher Scientific and CoA was purchased from AppliChem.

### Metabolomic analysis of *Pseudomonas aeruginosa*

AQ/AQNO analysis in *P. aeruginosa* has been performed as described elsewhere [38]. Briefly, bacteria were grown in BM2 minimal medium with or without exogenous fatty acids (2 mM). Methanol extracts of cells and culture supernatants were analyzed by HPLC-coupled mass spectrometry, using a 150 mm Kinetex C18 reversed-phase column with 1.7  $\mu\text{m}$  particle size and 2.1 mm inner diameter (Phenomenex) and a quadrupole time-of-flight mass spectrometer (maXis HD QTOF, Bruker) with positive mode electrospray ionization. Signal quantification was based on the calculation of peak areas in extracted ion chromatograms of the respective analytes.

### Acyl-CoA loading assays and LC-MS measurements of intact proteins

Acyl-CoA loading assays were performed with PqsBC (5  $\mu\text{M}$ ) and acyl-CoAs (50  $\mu\text{M}$ ) in Tris/HCl (50 mM, pH 7.6) at 30 °C. After incubation for 30 min, the samples (20  $\mu\text{L}$ ) were directly submitted for intact protein analyses. All ESI-MS measurements were performed on a Dionex Ultimate 3000 RSLC system using a ProSwift RP-4H (monolithic PS-DVB), 250  $\times$  1 mm column (Thermo Fisher Scientific) and a maXis 4G hr-ToF mass spectrometer (Bruker Daltonics) equipped with the standard Bruker ESI source.

### Cloning and site-directed mutagenesis

The *pqsB* (PA0997) and *pqsC* (PA0998) genes from *Pseudomonas aeruginosa* PAO1 were amplified from chromosomal DNA by PCR (primers listed in Table 5.S4). *pqsB* was cloned into pET26b (Merck Millipore) and *pqsC* was ligated into pET19m or p10\$ [72], both based on the pET19b vector (Merck Millipore). The resulting plasmid pET19m-*pqsC* produces PqsC with an *N*-terminal His<sub>6</sub> tag followed by a recognition motif for TEV (tobacco etch virus) protease, while p10\$-*pqsC* encodes PqsC with an *N*-terminal His<sub>6</sub>-tagged T7 lysozyme removable by human rhinovirus 3C protease. The active site



cysteine C129 of PqsC was mutated to alanine (PqsC<sup>C129A</sup>) or serine (PqsC<sup>C129S</sup>) in pET19mod-*pqsC* and p10\$-*pqsC* by PCR-based mutagenesis (Table 5.S4).

#### **Expression and purification of (His<sub>6</sub>-)PqsBC, (His<sub>6</sub>-)PqsBC<sup>C129A</sup> and PqsBC<sup>C129S</sup>**

Recombinant proteins were produced in *E. coli* BL21(DE3)pLysS (Promega) or BL21-CodonPlus(DE3)-RIL (Agilent Technologies) co-transformed with pET26b-*pqsB* and pET19m-*pqsC* or p10\$-*pqsC*. Purification involved nickel affinity and size exclusion chromatography with or without an intermittent protease cleavage and chromatography step to remove the His<sub>6</sub>-affinity tag (Table 5.S5). The purified proteins were concentrated to 20–35 mg mL<sup>-1</sup>, flash-cooled in liquid nitrogen and stored at -80 °C.

#### **PqsBC activity assay with 2-ABA and acyl-CoAs of different carbon chain lengths**

Enzymatic activities were measured in a spectrophotometric assay by using 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB, Ellman's reagent) to monitor the formation of free CoA at 412 nm [73, 74]. Acyl-CoA or DTNB stock solutions were freshly prepared in MilliQ water or DMSO, respectively. The concentrations of acyl-CoA stocks were measured at 259 nm [75] and the concentration of free CoA already present in the respective acyl-CoA stock was calculated by adding DTNB and comparing to a standard curve.

For the enzyme activity assay, PqsBC (1 μM) was mixed with acyl-CoA (500 μM) and 2-ABA (1 mM) and the samples were incubated at room temperature for 20 min. The reaction was stopped with SDS before measuring released CoA with DTNB.

#### **Crystallization, data-collection, phasing, refinement and structural analysis**

Initial crystallization conditions were identified with commercial screens using the sitting drop vapor diffusion method. After optimization by random and grid screening, crystals were harvested in cryo-protectant (Table 5.S1, Figure 5.S2) and diffraction data were collected on beamlines X06DA/X10SA at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) and on PETRA III beamline P11 at DESY (Hamburg, Germany). Diffraction data were indexed and integrated with XDS [76] or XDSAPP [77], and scaled with Aimless [78]. Since no other PqsBC structure was available at the outset of these studies, initial phases were derived by a combination of molecular replacement and heavy atom derivatization, using data from crystal form 2. The model was finalized by alternating steps of manual model building in COOT [79] and maximum-likelihood refinement in phenix.refine [80]. The final model was used to de-

termine the structures of PqsBC in other space groups by molecular replacement using Phaser [81]. Model qualities were evaluated with MolProbity [82]. Data collection and refinement statistics are listed in Table 5.S2. Atomic coordinates and structure factor amplitudes have been deposited in the Protein Data Bank ([www.rcsb.org](http://www.rcsb.org)) [83] with accession codes 6ESZ, 6ET0, 6ET1, 6ET2 and 6ET3.

Structural representation were rendered with PyMOL [84], secondary structure elements were assigned with DSSP [85] and protein cavities were calculated with KVFinder [49].

### Modeling of PqsBC<sup>C129A</sup>-acyl-CoA complexes

PqsBC<sup>C129A</sup>-acyl-CoA complexes were modeled with the structure of His<sub>6</sub>-PqsBC<sup>C129A</sup> determined in this study (chains IJ of crystal form 3, PDB ID: 6ET2) and the X-ray coordinates of *mtFabH* in complex with dodecanoyl-CoA (PDB ID: 1U6S [44]) using Molecular Operating Environment (MOE, Chemical Computing Group) [86].

## 5

### Acknowledgements

We thank SLS (Paul Scherrer Institute, Villigen, Switzerland) and PETRA III (DESY, Hamburg, Germany) for beamline access and the beamline staff for technical assistance. We are grateful to the X-ray community at the Helmholtz Centre for Infection Research (HZI, Braunschweig, Germany) for their help with data collection. We also thank Steffen L. Drees and Susanne Fetzner (University of Münster, Germany) for providing 2-ABA for PqsBC activity experiments and Eric Déziel for the kind provision of the *Pseudomonas aeruginosa* mutant strains used in this study. F. Witzgall and T. Depke are supported by the HZI Graduate School for Infection Research. T. Depke acknowledges financial and non-material support by the Studienstiftung des deutschen Volkes.

## Abbreviations

2-AA,	2-aminoacetophenone;
2-ABA,	2-aminobenzoylacetate;
2-ABA-CoA,	2-aminobenzoyl-acetyl-CoA;
2-HABA,	2-hydroxylaminobenzoylacetate;
AI,	autoinducer;
AQ,	2-alkyl-4(1 <i>H</i> )-quinolone;
AQNO,	2-alkyl-4-hydroxyquinoline <i>N</i> -oxide;
ASU,	asymmetric unit;
DHQ,	2,4-dihydroxyquinoline;
DTNB,	5,5'-dithiobis-2-nitrobenzoic acid;
<i>ec</i> FabH,	<i>Escherichia coli</i> FabH;
Fab,	fatty acid biosynthesis;
Fad,	fatty acid degradation;
HHQ,	2-heptyl-4(1 <i>H</i> )-quinolone;
HQNO,	2-heptyl-4-hydroxyquinoline <i>N</i> -oxide;
MD,	molecular dynamics;
<i>ml</i> FabH,	<i>Micrococcus luteus</i> FabH;
<i>mt</i> FabH,	<i>Mycobacterium tuberculosis</i> FabH;
MvfR,	multiple virulence factor regulator;
PDB,	Protein Data Bank;
PQS,	2-heptyl-3-hydroxy-4(1 <i>H</i> )-quinolone;
PSSweb,	Protein Structural Statistics Web Server;
QS,	quorum sensing;
TEV,	tobacco etch virus.

## References

- [1] J. B. Lyczak, C. L. Cannon, and G. B. Pier. “Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist”. In: *Microbes and infection* 2.9 (2000), pp. 1051–1060.
- [2] G. P. Bodey, R. Bolivar, V. Fainstein, and L. Jadeja. “Infections Caused by *Pseudomonas aeruginosa*”. In: *Clinical Infectious Diseases* 5.2 (Mar. 1983), pp. 279–313. DOI: 10.1093/clinids/5.2.279.
- [3] S. Stefani, S. Campana, L. Cariani, V. Carnovale, C. Colombo, M. Lleo, V. Iula, L. Minicucci, P. Morelli, G. Pizzamiglio, and G. Taccetti. “Relevance of multidrug-resistant *Pseudomonas aeruginosa* infections in cystic fibrosis”. In: *International Journal of Medical Microbiology* 307.6 (Sept. 2017), pp. 353–362. DOI: 10.1016/j.ijmm.2017.07.004.

- [4] M. Schuster, C. P. Lostroh, T. Ogi, and E. P. Greenberg. “Identification, Timing, and Signal Specificity of *Pseudomonas aeruginosa* Quorum-Controlled Genes: a Transcriptome Analysis”. In: *Journal of Bacteriology* 185.7 (Apr. 2003), pp. 2066–2079. DOI: 10.1128/jb.185.7.2066-2079.2003.
- [5] R. S. Smith and B. H. Iglewski. “*Pseudomonas aeruginosa* quorum sensing as a potential antimicrobial target”. In: *Journal of Clinical Investigation* 112.10 (Nov. 2003), pp. 1460–1465. DOI: 10.1172/jci200320364.
- [6] V. E. Wagner, D. Bushnell, L. Passador, A. I. Brooks, and B. H. Iglewski. “Microarray Analysis of *Pseudomonas aeruginosa* Quorum-Sensing Regulons: Effects of Growth Phase and Environment”. In: *Journal of Bacteriology* 185.7 (Apr. 2003), pp. 2080–2095. DOI: 10.1128/jb.185.7.2080-2095.2003.
- [7] N. A. Whitehead, A. M. Barnard, H. Slater, N. J. Simpson, and G. P. Salmond. “Quorum-sensing in Gram-negative bacteria”. In: *FEMS Microbiology Reviews* 25.4 (Aug. 2001), pp. 365–404. DOI: 10.1111/j.1574-6976.2001.tb00583.x.
- [8] K. Papenfort and B. L. Bassler. “Quorum sensing signal-response systems in Gram-negative bacteria”. In: *Nature Reviews Microbiology* 14.9 (Sept. 2016), pp. 576–588. DOI: 10.1038/nrmicro.2016.89.
- [9] J. Lee and L. Zhang. “The hierarchy quorum sensing network in *Pseudomonas aeruginosa*”. In: *Protein & Cell* 6.1 (Sept. 2014), pp. 26–41. DOI: 10.1007/s13238-014-0100-x.
- [10] N. M. Høyland-Kroghsbo, J. Paczkowski, S. Mukherjee, J. Broniewski, E. Westra, J. Bondy-Denomy, and B. L. Bassler. “Quorum sensing controls the *Pseudomonas aeruginosa* CRISPR-Cas adaptive immune system”. In: *Proceedings of the National Academy of Sciences* 114.1 (Nov. 2016), pp. 131–135. DOI: 10.1073/pnas.1617415113.
- [11] S. L. McKnight, B. H. Iglewski, and E. C. Pesci. “The *Pseudomonas* Quinolone Signal Regulates *rhl* Quorum Sensing in *Pseudomonas aeruginosa*”. In: *Journal of Bacteriology* 182.10 (May 2000), pp. 2702–2708. DOI: 10.1128/jb.182.10.2702-2708.2000.
- [12] M. J. Gambello and B. H. Iglewski. “Cloning and characterization of the *Pseudomonas aeruginosa lasR* gene, a transcriptional activator of elastase expression.” In: *Journal of Bacteriology* 173.9 (May 1991), pp. 3000–3009. DOI: 10.1128/jb.173.9.3000-3009.1991.
- [13] L. Passador, J. Cook, M. Gambello, L. Rust, and B. Iglewski. “Expression of *Pseudomonas aeruginosa* virulence genes requires cell-to-cell communication”. In: *Science* 260.5111 (May 1993), pp. 1127–1130. DOI: 10.1126/science.8493556.
- [14] J. P. Pearson, K. M. Gray, L. Passador, K. D. Tucker, A. Eberhard, B. H. Iglewski, and E. P. Greenberg. “Structure of the autoinducer required for expression of *Pseudomonas aeruginosa* virulence genes.” In: *Proceedings of the National Academy of Sciences* 91.1 (Jan. 1994), pp. 197–201. DOI: 10.1073/pnas.91.1.197.
- [15] U. A. Ochsner, A. K. Koch, A. Fiechter, and J. Reiser. “Isolation and characterization of a regulatory gene affecting rhamnolipid biosurfactant synthesis in *Pseudomonas aeruginosa*.” In: *Journal of Bacteriology* 176.7 (Apr. 1994), pp. 2044–2054. DOI: 10.1128/jb.176.7.2044-2054.1994.
- [16] U. A. Ochsner and J. Reiser. “Autoinducer-mediated regulation of rhamnolipid biosurfactant synthesis in *Pseudomonas aeruginosa*.” In: *Proceedings of the National Academy of Sciences* 92.14 (July 1995), pp. 6424–6428. DOI: 10.1073/pnas.92.14.6424.

- [17] J. P. Pearson, L. Passador, B. H. Iglewski, and E. P. Greenberg. “A second *N*-acyl-homoserine lactone signal produced by *Pseudomonas aeruginosa*.” In: *Proceedings of the National Academy of Sciences* 92.5 (Feb. 1995), pp. 1490–1494. DOI: 10.1073/pnas.92.5.1490.
- [18] E. C. Pesci, J. B. J. Milbank, J. P. Pearson, S. McKnight, A. S. Kende, E. P. Greenberg, and B. H. Iglewski. “Quinolone signaling in the cell-to-cell communication system of *Pseudomonas aeruginosa*.” In: *Proceedings of the National Academy of Sciences* 96.20 (Sept. 1999), pp. 11229–11234. DOI: 10.1073/pnas.96.20.11229.
- [19] H. Cao, G. Krishnan, B. Goumnerov, J. Tsongalis, R. Tompkins, and L. G. Rahme. “A quorum sensing-associated virulence gene of *Pseudomonas aeruginosa* encodes a LysR-like transcription regulator with a unique self-regulatory mechanism.” In: *Proceedings of the National Academy of Sciences* 98.25 (Nov. 2001), pp. 14613–14618. DOI: 10.1073/pnas.251465298.
- [20] L. A. Gallagher, S. L. McKnight, M. S. Kuznetsova, E. C. Pesci, and C. Manoil. “Functions Required for Extracellular Quinolone Signaling by *Pseudomonas aeruginosa*.” In: *Journal of Bacteriology* 184.23 (Dec. 2002), pp. 6472–6480. DOI: 10.1128/jb.184.23.6472-6480.2002.
- [21] E. Deziel, F. Lepine, S. Milot, J. He, M. N. Mindrinos, R. G. Tompkins, and L. G. Rahme. “Analysis of *Pseudomonas aeruginosa* 4-hydroxy-2-alkylquinolines (HAQs) reveals a role for 4-hydroxy-2-heptylquinoline in cell-to-cell communication.” In: *Proceedings of the National Academy of Sciences* 101.5 (Jan. 2004), pp. 1339–1344. DOI: 10.1073/pnas.0307694100.
- [22] D. S. Wade, M. W. Calfee, E. R. Rocha, E. A. Ling, E. Engstrom, J. P. Coleman, and E. C. Pesci. “Regulation of *Pseudomonas* Quinolone Signal Synthesis in *Pseudomonas aeruginosa*.” In: *Journal of Bacteriology* 187.13 (June 2005), pp. 4372–4380. DOI: 10.1128/jb.187.13.4372-4380.2005.
- [23] G. Xiao, E. Déziel, J. He, F. Lépine, B. Lesic, M.-H. Castonguay, S. Milot, A. P. Tampakaki, S. E. Stachel, and L. G. Rahme. “MvfR, a key *Pseudomonas aeruginosa* pathogenicity LTTR-class regulatory protein, has dual ligands.” In: *Molecular Microbiology* 62.6 (Dec. 2006), pp. 1689–1699. DOI: 10.1111/j.1365-2958.2006.05462.x.
- [24] J. P. Coleman, L. L. Hudson, S. L. McKnight, J. M. Farrow, M. W. Calfee, C. A. Lindsey, and E. C. Pesci. “*Pseudomonas aeruginosa* PqsA Is an Anthranilate-Coenzyme A Ligase.” In: *Journal of Bacteriology* 190.4 (Dec. 2007), pp. 1247–1255. DOI: 10.1128/jb.01140-07.
- [25] Y.-M. Zhang, M. W. Frank, K. Zhu, A. Mayasundari, and C. O. Rock. “PqsD Is Responsible for the Synthesis of 2,4-Dihydroxyquinoline, an Extracellular Metabolite Produced by *Pseudomonas aeruginosa*.” In: *Journal of Biological Chemistry* 283.43 (Aug. 2008), pp. 28788–28794. DOI: 10.1074/jbc.m804555200.
- [26] S. Yu, V. Jensen, J. Seeliger, I. Feldmann, S. Weber, E. Schleicher, S. Häussler, and W. Blankenfeldt. “Structure Elucidation and Preliminary Assessment of Hydrolase Activity of PqsE, the *Pseudomonas* Quinolone Signal (PQS) Response Protein.” In: *Biochemistry* 48.43 (Nov. 2009), pp. 10298–10307. DOI: 10.1021/bi900123j.
- [27] S. L. Drees and S. Fetzner. “PqsE of *Pseudomonas aeruginosa* Acts as Pathway-Specific Thioesterase in the Biosynthesis of Alkylquinolone Signaling Molecules.” In: *Chemistry & Biology* 22.5 (May 2015), pp. 611–618. DOI: 10.1016/j.chembio.2015.04.012.

- [28] C. E. Dulcey, V. Dekimpe, D.-A. Fauvelle, S. Milot, M.-C. Groleau, N. Doucet, L. G. Rahme, F. Lépine, and E. Déziel. “The End of an Old Hypothesis: The *Pseudomonas* Signaling Molecules 4-Hydroxy-2-Alkylquinolines Derive from Fatty Acids, Not 3-Ketofatty Acids”. In: *Chemistry & Biology* 20.12 (Dec. 2013), pp. 1481–1491. DOI: 10.1016/j.chembiol.2013.09.021.
- [29] S. L. Drees, C. Li, F. Prasetya, M. Saleem, I. Dreveny, P. Williams, U. Hennecke, J. Emsley, and S. Fetzner. “PqsBC, a Condensing Enzyme in the Biosynthesis of the *Pseudomonas aeruginosa* Quinolone Signal”. In: *Journal of Biological Chemistry* 291.13 (Jan. 2016), pp. 6610–6624. DOI: 10.1074/jbc.m115.708453.
- [30] J. W. Schertzer, S. A. Brown, and M. Whiteley. “Oxygen levels rapidly modulate *Pseudomonas aeruginosa* social behaviours via substrate limitation of PqsH”. In: *Molecular Microbiology* 77.6 (Sept. 2010), pp. 1527–1538. DOI: 10.1111/j.1365-2958.2010.07303.x.
- [31] S. L. Drees, S. Ernst, B. D. Belviso, N. Jagmann, U. Hennecke, and S. Fetzner. “PqsL uses reduced flavin to produce 2-hydroxylaminobenzoylacetate, a preferred PqsBC substrate in alkyl quinolone biosynthesis in *Pseudomonas aeruginosa*”. In: *Journal of Biological Chemistry* (Apr. 2018), jbc.RA117.000789. DOI: 10.1074/jbc.ra117.000789.
- [32] J. W. Cornforth and A. T. James. “Structure of a naturally occurring antagonist of dihydrostreptomycin”. In: *Biochemical Journal* 63.1 (May 1956), pp. 124–130. DOI: 10.1042/bj0630124.
- [33] A. M. Haapalainen, G. Meriläinen, and R. K. Wierenga. “The thiolase superfamily: condensing enzymes with diverse reaction specificities”. In: *Trends in Biochemical Sciences* 31.1 (Jan. 2006), pp. 64–71. DOI: 10.1016/j.tibs.2005.11.011.
- [34] F. Lépine, S. Milot, E. Déziel, J. He, and L. G. Rahme. “Electrospray/mass spectrometric identification and analysis of 4-hydroxy-2-alkylquinolines (HAQs) produced by *Pseudomonas aeruginosa*”. In: *Journal of the American Society for Mass Spectrometry* 15.6 (June 2004), pp. 862–869. DOI: 10.1016/j.jasms.2004.02.012.
- [35] S. Heeb, M. P. Fletcher, S. R. Chhabra, S. P. Diggle, P. Williams, and M. Cámara. “Quinolones: from antibiotics to autoinducers”. In: *FEMS Microbiology Reviews* 35.2 (Mar. 2011), pp. 247–274. DOI: 10.1111/j.1574-6976.2010.00247.x.
- [36] G. W. Taylor, Z. A. Machan, S. Mehmet, P. J. Cole, and R. Wilson. “Rapid identification of 4-hydroxy-2-alkylquinolines produced by *Pseudomonas aeruginosa* using gas chromatography—electron-capture mass spectrometry”. In: *Journal of Chromatography B: Biomedical Sciences and Applications* 664.2 (Feb. 1995), pp. 458–462. DOI: 10.1016/0378-4347(94)00494-p.
- [37] C. A. Ortori, J.-F. Dubern, S. R. Chhabra, M. Cámara, K. Hardie, P. Williams, and D. A. Barrett. “Simultaneous quantitative profiling of *N*-acyl-L-homoserine lactone and 2-alkyl-4(1*H*)-quinolone families of quorum-sensing signaling molecules using LC-MS/MS”. In: *Analytical and Bioanalytical Chemistry* 399.2 (Oct. 2010), pp. 839–850. DOI: 10.1007/s00216-010-4341-0.
- [38] T. Depke, R. Franke, and M. Brönstrup. “Clustering of MS<sup>2</sup> spectra using unsupervised methods to aid the identification of secondary metabolites from *Pseudomonas aeruginosa*”. In: *Journal of Chromatography B* 1071 (Dec. 2017), pp. 19–28. DOI: 10.1016/j.jchromb.2017.06.002.
- [39] T. Gaillard, B. B. L. Schwarz, Y. Chebaro, R. H. Stote, and A. Dejaegere. “Protein Structural Statistics with PSS”. In: *Journal of Chemical Information and Modeling* 53.9 (Sept. 2013), pp. 2471–2482. DOI: 10.1021/ci400233j.



- [40] T. Gaillard, R. H. Stote, and A. Dejaegere. “PSSweb: protein structural statistics web server”. In: *Nucleic Acids Research* 44.W1 (May 2016), W401–W405. DOI: 10.1093/nar/gkw332.
- [41] A. K. Bera, V. Atanasova, H. Robinson, E. Eisenstein, J. P. Coleman, E. C. Pesci, and J. F. Parsons. “Structure of PqsD, a *Pseudomonas* Quinolone Signal Biosynthetic Enzyme, in Complex with Anthranilate”. In: *Biochemistry* 48.36 (Sept. 2009), pp. 8644–8655. DOI: 10.1021/bi9009055.
- [42] C. Davies, R. J. Heath, S. W. White, and C. O. Rock. “The 1.8 Å crystal structure and active-site architecture of  $\beta$ -ketoacyl-acyl carrier protein synthase III (FabH) from *Escherichia coli*”. In: *Structure* 8.2 (2000), pp. 185–195.
- [43] X. Qiu, C. A. Janson, W. W. Smith, M. Head, J. Lonsdale, and A. K. Konstantinidis. “Refined structures of beta-ketoacyl-acyl carrier protein synthase III.” In: *Journal of molecular biology* 307 (1 Mar. 2001), pp. 341–356. ISSN: 0022-2836. DOI: 10.1006/jmbi.2000.4457.
- [44] F. Musayev, S. Sachdeva, J. N. Scarsdale, K. Reynolds, and H. Wright. “Crystal Structure of a Substrate Complex of *Mycobacterium tuberculosis*  $\beta$ -Ketoacyl-acyl Carrier Protein Synthase III (FabH) with Lauroyl-coenzyme A”. In: *Journal of Molecular Biology* 346.5 (Mar. 2005), pp. 1313–1321. DOI: 10.1016/j.jmb.2004.12.044.
- [45] S. Sachdeva, F. N. Musayev, M. M. Alhamadsheh, J. N. Scarsdale, H. T. Wright, and K. A. Reynolds. “Separate Entrance and Exit Portals for Ligand Traffic in *Mycobacterium tuberculosis* FabH”. In: *Chemistry & Biology* 15.4 (Apr. 2008), pp. 402–412. DOI: 10.1016/j.chembiol.2008.03.007.
- [46] X. Qiu, C. A. Janson, A. K. Konstantinidis, S. Nwagwu, C. Silverman, W. W. Smith, S. Khandekar, J. Lonsdale, and S. S. Abdel-Meguid. “Crystal Structure of  $\beta$ -Ketoacyl-Acyl Carrier Protein Synthase III”. In: *Journal of Biological Chemistry* 274.51 (Dec. 1999), pp. 36465–36471. DOI: 10.1074/jbc.274.51.36465.
- [47] X. Qiu, A. E. Choudhry, C. A. Janson, M. Grooms, R. A. Daines, J. T. Lonsdale, and S. S. Khandekar. “Crystal structure and substrate specificity of the  $\beta$ -ketoacyl-acyl carrier protein synthase III (FabH) from *Staphylococcus aureus*”. In: *Protein Science* 14.8 (Aug. 2005), pp. 2087–2094. DOI: 10.1110/ps.051501605.
- [48] J. H. Pereira, E.-B. Goh, J. D. Keasling, H. R. Beller, and P. D. Adams. “Structure of FabH and factors affecting the distribution of branched fatty acids in *Micrococcus luteus*”. In: *Acta Crystallographica Section D Biological Crystallography* 68.10 (Sept. 2012), pp. 1320–1328. DOI: 10.1107/s0907444912028351.
- [49] S. H. Oliveira, F. A. Ferraz, R. V. Honorato, J. Xavier-Neto, T. J. Sobreira, and P. S. de Oliveira. “KVFinder: steered identification of protein cavities as a PyMOL plugin”. In: *BMC Bioinformatics* 15.1 (2014), p. 197. DOI: 10.1186/1471-2105-15-197.
- [50] D. Szamosvári and T. Böttcher. “An Unsaturated Quinolone *N*-Oxide of *Pseudomonas aeruginosa* Modulates Growth and Virulence of *Staphylococcus aureus*”. In: *Angewandte Chemie International Edition* 56.25 (May 2017), pp. 7271–7275. DOI: 10.1002/anie.201702944.
- [51] I. C. Wells. “Antibiotic substances produced by *Pseudomonas aeruginosa*; syntheses of Pyo Ib, Pyo Ic, and Pyo III.” In: *The Journal of biological chemistry* 196 (1 May 1952), pp. 331–340. ISSN: 0021-9258.
- [52] Y. Fujita, H. Matsuoka, and K. Hirooka. “Regulation of fatty acid metabolism in bacteria”. In: *Molecular Microbiology* 66.4 (Nov. 2007), pp. 829–839. DOI: 10.1111/j.1365-2958.2007.05947.x.

- [53] Y. Kang, J. Zarzycki-Siek, C. B. Walton, M. H. Norris, and T. T. Hoang. “Multiple FadD Acyl-CoA Synthetases Contribute to Differential Fatty Acid Degradation and Virulence in *Pseudomonas aeruginosa*”. In: *PLoS ONE* 5.10 (Oct. 2010). Ed. by M. Otto, e13557. DOI: 10.1371/journal.pone.0013557.
- [54] J. Zarzycki-Siek, M. H. Norris, Y. Kang, Z. Sun, A. P. Bluhm, I. A. McMillan, and T. T. Hoang. “Elucidating the *Pseudomonas aeruginosa* Fatty Acid Degradation Pathway: Identification of Additional Fatty Acyl-CoA Synthetase Homologues”. In: *PLoS ONE* 8.5 (May 2013). Ed. by M. Skurnik, e64554. DOI: 10.1371/journal.pone.0064554.
- [55] G. Wells, S. Palethorpe, and E. C. Pesci. “PsrA controls the synthesis of the *Pseudomonas aeruginosa* quinolone signal via repression of the FadE homolog, PA0506”. In: *PLOS ONE* 12.12 (Dec. 2017). Ed. by M.-J. Viroille, e0189331. DOI: 10.1371/journal.pone.0189331.
- [56] Y. Yuan, J. A. Leeds, and T. C. Meredith. “*Pseudomonas aeruginosa* Directly Shunts  $\beta$ -Oxidation Degradation Intermediates into De Novo Fatty Acid Biosynthesis”. In: *Journal of Bacteriology* 194.19 (June 2012), pp. 5185–5196. DOI: 10.1128/jb.00860-12.
- [57] L. Zhang, T. A. Veres-Schalnat, A. Somogyi, J. E. Pemberton, and R. M. Maier. “Fatty Acid Cosubstrates Provide  $\beta$ -Oxidation Precursors for Rhamnolipid Biosynthesis in *Pseudomonas aeruginosa*, as Evidenced by Isotope Tracing and Gene Expression Assays”. In: *Applied and Environmental Microbiology* 78.24 (Oct. 2012), pp. 8611–8622. DOI: 10.1128/aem.02111-12.
- [58] M. R. Parsek, D. L. Val, B. L. Hanzelka, J. E. Cronan, and E. P. Greenberg. “Acyl homoserine-lactone quorum-sensing signal generation.” In: *Proceedings of the National Academy of Sciences of the United States of America* 96 (8 Apr. 1999), pp. 4360–4365. ISSN: 0027-8424.
- [59] T. A. Gould, H. P. Schweizer, and M. E. A. Churchill. “Structure of the *Pseudomonas aeruginosa* acyl-homoserinelactone synthase LasI”. In: *Molecular Microbiology* 53.4 (July 2004), pp. 1135–1146. DOI: 10.1111/j.1365-2958.2004.04211.x.
- [60] M. P. Fletcher, S. P. Diggle, S. A. Crusz, S. R. Chhabra, M. Cámara, and P. Williams. “A dual biosensor for 2-alkyl-4-quinolone quorum-sensing signal molecules”. In: *Environmental Microbiology* 9.11 (Nov. 2007), pp. 2683–2693. DOI: 10.1111/j.1462-2920.2007.01380.x.
- [61] A. Ilangovan, M. Fletcher, G. Rampioni, C. Pustelny, K. Rumbaugh, S. Heeb, M. Cámara, A. Truman, S. R. Chhabra, J. Emsley, and P. Williams. “Structural Basis for Native Agonist and Synthetic Inhibitor Recognition by the *Pseudomonas aeruginosa* Quorum Sensing Regulator PqsR (MvfR)”. In: *PLoS Pathogens* 9.7 (July 2013). Ed. by F. M. Ausubel, e1003508. DOI: 10.1371/journal.ppat.1003508.
- [62] J. N. Scarsdale, G. Kazanina, X. He, K. A. Reynolds, and H. T. Wright. “Crystal Structure of the Mycobacterium tuberculosis  $\beta$ -Ketoacyl-Acyl Carrier Protein Synthase III”. In: *Journal of Biological Chemistry* 276.23 (Mar. 2001), pp. 20516–20522. DOI: 10.1074/jbc.m010762200.
- [63] K. S. Gajiwala, S. Margosiak, J. Lu, J. Cortez, Y. Su, Z. Nie, and K. Appelt. “Crystal structures of bacterial FabH suggest a molecular basis for the substrate specificity of the enzyme”. In: *FEBS Letters* 583.17 (Aug. 2009), pp. 2939–2946. DOI: 10.1016/j.febslet.2009.08.001.
- [64] J. D. Nanson, Z. Himiari, C. M. D. Swarbrick, and J. K. Forwood. “Structural Characterisation of the Beta-Ketoacyl-Acyl Carrier Protein Synthases, FabF and FabH, of *Yersinia pestis*”. In: *Scientific Reports* 5 (Oct. 2015), p. 14797. DOI: 10.1038/srep14797.



- [65] M. M. Alhamadsheh, F. Musayev, A. A. Komissarov, S. Sachdeva, H. T. Wright, N. Scarsdale, G. Florova, and K. A. Reynolds. “Alkyl-CoA Disulfides as Inhibitors and Mechanistic Probes for FabH Enzymes”. In: *Chemistry & Biology* 14.5 (May 2007), pp. 513–524. DOI: 10.1016/j.chembiol.2007.03.013.
- [66] A. Steinbach, C. K. Maurer, E. Weidel, C. Henn, C. Brengel, R. W. Hartmann, and M. Negri. “Molecular basis of HHQ biosynthesis: molecular dynamics simulations, enzyme kinetic and surface plasmon resonance studies”. In: *BMC Biophysics* 6.1 (2013), p. 10. DOI: 10.1186/2046-1682-6-10.
- [67] J. Z. Lu and S. T. Prigge. “The Tail of Mycolic Acids”. In: *Chemistry & Biology* 15.4 (Apr. 2008), pp. 309–310. DOI: 10.1016/j.chembiol.2008.04.001.
- [68] D. Ramamoorthy, E. Turos, and W. C. Guida. “Identification of a New Binding Site in *E. coli* FabH using Molecular Dynamics Simulations: Validation by Computational Alanine Mutagenesis and Docking Studies”. In: *Journal of Chemical Information and Modeling* 53.5 (Apr. 2013), pp. 1138–1156. DOI: 10.1021/ci3003528.
- [69] D. Maura, S. L. Drees, A. Bandyopadhyaya, T. Kitao, M. Negri, M. Starkey, B. Lesic, S. Milot, E. Déziel, R. Zahler, M. Pucci, A. Felici, S. Fetzner, F. Lépine, and L. G. Rahme. “Polypharmacology Approaches against the *Pseudomonas aeruginosa* MvfR Regulon and Their Application in Blocking Virulence and Antibiotic Tolerance”. In: *ACS Chemical Biology* 12.5 (Apr. 2017), pp. 1435–1443. DOI: 10.1021/acscchembio.6b01139.
- [70] G. Allegretta, C. K. Maurer, J. Eberhard, D. Maura, R. W. Hartmann, L. Rahme, and M. Empting. “In-depth Profiling of MvfR-Regulated Small Molecules in *Pseudomonas aeruginosa* after Quorum Sensing Inhibitor Treatment”. In: *Frontiers in Microbiology* 8 (May 2017). DOI: 10.3389/fmicb.2017.00924.
- [71] Z. Zhou and S. Ma. “Recent Advances in the Discovery of PqsD Inhibitors as Antimicrobial Agents”. In: *ChemMedChem* 12.6 (Mar. 2017), pp. 420–425. DOI: 10.1002/cmdc.201700015.
- [72] T. Bock, E. Luxenburger, J. Hoffmann, V. Schütza, C. Feiler, R. Müller, and W. Blankenfeldt. “AibA/AibB Induces an Intramolecular Decarboxylation in Isovalerate Biosynthesis by *Myxococcus xanthus*”. In: *Angewandte Chemie International Edition* 56.33 (May 2017), pp. 9986–9989. DOI: 10.1002/anie.201701992.
- [73] G. L. Ellman. “Tissue sulfhydryl groups”. In: *Archives of Biochemistry and Biophysics* 82.1 (May 1959), pp. 70–77. DOI: 10.1016/0003-9861(59)90090-6.
- [74] P. W. Riddles, R. L. Blakeley, and B. Zerner. “Reassessment of Ellman’s reagent.” In: *Methods in enzymology* 91 (1983), pp. 49–60. ISSN: 0076-6879.
- [75] P. G. Killenberg and D. F. Dukes. “Coenzyme A derivatives of bile acids-chemical synthesis, purification, and utilization in enzymic preparation of taurine conjugates.” In: *Journal of lipid research* 17 (5 Sept. 1976), pp. 451–455. ISSN: 0022-2275.
- [76] W. Kabsch. “XDS”. In: *Acta Crystallographica Section D Biological Crystallography* 66.2 (Jan. 2010), pp. 125–132. DOI: 10.1107/s0907444909047337.
- [77] M. Krug, M. S. Weiss, U. Heinemann, and U. Mueller. “XDSAPP: a graphical user interface for the convenient processing of diffraction data using XDS”. In: *Journal of Applied Crystallography* 45.3 (Apr. 2012), pp. 568–572. DOI: 10.1107/s0021889812011715.
- [78] P. R. Evans and G. N. Murshudov. “How good are my data and what is the resolution?” In: *Acta Crystallographica Section D Biological Crystallography* 69.7 (June 2013), pp. 1204–1214. DOI: 10.1107/s0907444913000061.

- [79] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan. “Features and development of Coot”. In: *Acta Crystallographica Section D Biological Crystallography* 66.4 (Mar. 2010), pp. 486–501. DOI: 10.1107/s0907444910007493.
- [80] P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, and P. D. Adams. “Towards automated crystallographic structure refinement with phenix.refine”. In: *Acta Crystallographica Section D Biological Crystallography* 68.4 (Mar. 2012), pp. 352–367. DOI: 10.1107/s0907444912001308.
- [81] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. “Phaser crystallographic software”. In: *Journal of Applied Crystallography* 40.4 (July 2007), pp. 658–674. DOI: 10.1107/s0021889807021206.
- [82] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson. “MolProbity: all-atom structure validation for macromolecular crystallography”. In: *Acta Crystallographica Section D Biological Crystallography* 66.1 (Dec. 2009), pp. 12–21. DOI: 10.1107/s0907444909042073.
- [83] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. “The Protein Data Bank.” In: *Nucleic acids research* 28 (1 Jan. 2000), pp. 235–242. ISSN: 0305-1048.
- [84] *The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.*
- [85] W. Kabsch and C. Sander. “Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features”. In: *Biopolymers* 22.12 (Dec. 1983), pp. 2577–2637. DOI: 10.1002/bip.360221211.
- [86] *Molecular Operating Environment (MOE), 2013.08; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite No.910, Montreal, QC, Canada, H3A 2R7, 2018.*

## Supporting Information

### Supporting Experimental Section

#### Metabolomic analysis of *Pseudomonas aeruginosa*

Bacteria were pre-cultured overnight in lysogeny broth (LB) at 37 °C and 160 rpm in a shaking incubator. The cells were pelleted by centrifugation (4 °C, 9000 *g*, 5 min) and used to inoculate BM2 minimal medium consisting of (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (2 mM), K<sub>2</sub>HPO<sub>4</sub> (40 mM), KH<sub>2</sub>PO<sub>4</sub> (22 mM), MgSO<sub>4</sub> (2 mM), FeSO<sub>4</sub> (10 μM), glucose (0.4%, *w/v*) and casein hydrolysate (0.01%, *w/v*). The starting OD<sub>600</sub> was adjusted to 0.05 and the cultures were harvested after 8 h at an OD<sub>600</sub> of 1.5 ± 0.1. For the analysis of cellular AQs/AQNOs, the cells were pelleted by centrifugation (4 °C, 9000 *g*, 5 min) and the pellet was snap frozen in liquid nitrogen while the supernatant was discarded. For the combined analysis of cellular and extracellular AQs/AQNOs, the collected volume of culture broth was dried overnight in a centrifugal evaporator (Refrigerated CentriVap Concentrator with –50 °C CentriVap Cold Trap, Labconco) at 20 °C.

Fatty acid feeding experiments were conducted by supplementing the minimal medium with final concentrations of 2 mM fatty acid. As the fatty acids were dissolved in ethanol (100%, *v/v*), the respective amount of ethanol was added to the control samples. For the fatty acid feeding experiments, all members of the homologous series from acetic acid to tetradecanoic acid as well as 2-octenoic acid (mixture of *cis* and *trans*) and the perdeuterated octanoic-d15 acid were tested. All experiments were performed in triplicates.

Cell pellets and dried culture broth samples were reconstituted to the original concentration in methanol (100%, *v/v*) by two rounds of vigorous shaking for 1 min and subsequent 10 min sonication in an ice cold ultrasonic bath (Sonorex Digiplus, BANDLIN electronic). After centrifugation (4 °C, 9000 *g*, 5 min), a defined volume of the supernatant was collected and dried in a centrifugal evaporator overnight at 20 °C. The sample was then reconstituted in half the volume of acetonitrile (50%, *v/v*) with formic acid (0.1%, *v/v*) by shaking for 5 s, followed by sonication for 5 min and again shaking for 5 s.

The analytical procedure was essentially conducted as described by Depke *et al.* [1]. In short, the extracts were separated on a 150 mm Kinetex C18 reversed-phase column with 1.7 μm particle size and 2.1 mm inner diameter (Phenomenex) using a gradient of water with formic acid (0.1%, *v/v*) as eluent A and acetonitrile with formic acid (0.1%, *v/v*) as eluent B. The gradient elution was performed as follows: hold 1% B for

2 min, proceed with a linear gradient from 1% B to 100% B over 18 min, remain at 100% B for 5 min and return to 1% B with a linear gradient over 5 min. The samples were analysed on a quadrupole time-of-flight mass spectrometer (maXis HD QTOF, Bruker) using positive mode electrospray ionization. Full scan data (50–1500 Da) were recorded and collision-induced dissociation was used to collect data-dependent MS/MS data for feature annotation. Signal quantification was based on the calculation of peak areas in extracted ion chromatograms of the respective analytes. AQs and AQNOs were annotated as described by Depke *et al.* [1]. The nomenclature used in this study is  $CX:n$ -HQ for 2-alkyl-4(1*H*)-quinolones (HHQ series) and  $CX:n$ -QNO for 2-alkyl-4-hydroxyquinoline *N*-oxides (HQNO series) with *X* representing the number of carbons in the aliphatic chain and *n* the number of double bonds in the chain, i.e. C7-HQ is equivalent to HHQ (2-heptyl-4(1*H*)-quinolone).

#### Acyl-CoA loading assays and LC-MS measurements of intact proteins

Acyl-CoA loading assays were performed with PqsBC (5  $\mu$ M) and acyl-CoAs (50  $\mu$ M) in Tris/HCl (50 mM, pH 7.6) at 30 °C. After incubation for 30 min, the samples (20  $\mu$ L) were directly submitted for intact protein analyses. All ESI-MS measurements of intact protein samples were performed on a Dionex Ultimate 3000 RSLC system using a ProSwift RP-4H (monolithic PS-DVB), 250  $\times$  1 mm column (Thermo Fisher Scientific). Separation of the samples (1.0  $\mu$ L) was achieved by a multistep gradient from (A) H<sub>2</sub>O + formic acid (0.1%, *v/v*) + dimethyl sulfoxide (DMSO, 1%, *v/v*) to (B) acetonitrile + formic acid (0.1%, *v/v*) + DMSO (1%, *v/v*) at a flow rate of 200  $\mu$ L min<sup>-1</sup> at 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5% (B), followed by an increase to 65% (B) in 18 min, followed by an increase to 98% in 0.5 min to end up with a 3 min step at 98% (B) before reequilibration to the initial conditions. UV spectra were recorded with a diode array detector in the range from 200 to 600 nm. The LC flow was split to 75  $\mu$ L min<sup>-1</sup> before entering a maXis 4G hr-ToF mass spectrometer (Bruker Daltonics) using the standard Bruker ESI source. In the source region, the temperature was set to 180 °C, the capillary voltage was 4000 V, the dry-gas flow was 6.0 L min<sup>-1</sup> and the nebulizer was set to 1.1 bar. Mass spectra were acquired in positive ionization mode ranging from 600–1800 *m/z* at 2.5 Hz scan rate. Protein masses were deconvoluted by using the Maximum Entropy algorithm (Spectrum Square Associates, Inc.).

### Cloning and site-directed mutagenesis

The *pqsB* (PA0997) and *pqsC* (PA0998) genes from *Pseudomonas aeruginosa* PAO1 were amplified from chromosomal DNA by PCR with sequence specific primers (Table 5.S44). To generate an untagged construct of PqsB, the PCR product of *pqsB* was cloned into pET26b (Merck Millipore) using the restriction enzymes *NdeI* (NEB) and *BamHI* (NEB). The amplicon coding for PqsC was cloned into *NdeI/XhoI*-digested pET19m and p10\$ [2], which are engineered plasmids based on the pET19b vector (Merck Millipore). The resulting plasmid pET19m-*pqsC* produces PqsC with a *N*-terminal His<sub>6</sub> tag followed by a recognition motif for TEV (tobacco etch virus) protease (MGHHHHHHA-ENLYFQGH-PqsC; recognition sequence of TEV protease is underlined), while p10\$-*pqsC* encodes PqsC with a *N*-terminal His<sub>6</sub>-tagged T7 lysozyme (T7L) (MGHHHHHHAENLYFQGH-T7L-*LEVLFQGH*-PqsC; recognition sequences for TEV protease and 3C protease are indicated by underlined and italics characters, respectively), which can be removed by human rhinovirus 3C protease. The active site cysteine C129 of PqsC was mutated to alanine (PqsC<sup>C129A</sup>) or serine (PqsC<sup>C129S</sup>) in pET19mod-*pqsC* and p10\$-*pqsC* by site-directed mutagenesis via PCR using KAPA HiFi DNA Polymerase (Kapa Biosystems) with primers listed in Table 5.S4. Parental template DNA was digested with *DpnI* (NEB).

### Expression and purification of (His<sub>6</sub>-)PqsBC, (His<sub>6</sub>-)PqsBC<sup>C129A</sup> and PqsBC<sup>C129S</sup>

The expression plasmids pET26b-*pqsB*/pET19m-*pqsC* or pET26b-*pqsB*/p10\$-*pqsC* were co-transformed into chemically competent *E. coli* BL21(DE3)pLysS (Promega) or BL21-CodonPlus(DE3)-RIL cells (Agilent Technologies), respectively. An overnight pre-culture grown from a single colony in lysogeny broth (LB) medium containing ampicillin (100 mg L<sup>-1</sup>), kanamycin (50 mg L<sup>-1</sup>) and chloramphenicol (34 mg L<sup>-1</sup>) at 37 °C was used for inoculation of Terrific Broth (TB) medium supplemented with the same antibiotics to an optical density (OD<sub>600</sub>) of 0.05. When this culture reached an OD<sub>600</sub> of 0.8–1.0 at 37 °C, the temperature was shifted to 20 °C and protein expression was induced with isopropyl-β-D-thiogalactopyranoside (IPTG, 0.5 mM) for 20 h. Cells were centrifuged at 6.800 *g* for 15 min at 4 °C, the pellet was shock frozen in liquid nitrogen and stored at -80 °C. In the following, the general purification procedures for His<sub>6</sub>-tagged and untagged PqsBC are described. For the purification of distinct PqsBC variants, slightly different buffers were used (summarized in Table 5.S5).

The thawed cell pellet was resuspended in buffer A supplemented with MgCl<sub>2</sub> (2 mM), DNase I (0.6 mg L<sup>-1</sup>, Roche Life Science) and a cComplete mini EDTA-free protease in-

hibitor cocktail tablet (Roche Life Science). Resuspended cells were disrupted with an Emulsiflex-C3 homogenizer (Avestin) and the insoluble fraction was removed by centrifugation (37,000 *g* for 45 min at 4 °C). Recombinant His<sub>6</sub>-PqsC/PqsB (His<sub>6</sub>-PqsBC) or His<sub>6</sub>-T7L-PqsC/PqsB (His<sub>6</sub>-T7L-PqsBC) was isolated from the supernatant by nickel affinity chromatography using a 5 mL HisTrap HP column (GE Healthcare) charged with NiSO<sub>4</sub> (100 mM) and equilibrated with buffer A. The complex was eluted with an imidazole gradient of 2–40% buffer B [buffer A with imidazole (500 mM)]. Fractions containing the target proteins were pooled and 1 mg His<sub>6</sub>-tagged TEV protease per 25 mg His<sub>6</sub>-PqsBC or 1 mg His<sub>6</sub>-tagged 3C protease per 40 mg His<sub>6</sub>-T7L-PqsBC was added to cut off the His<sub>6</sub> tag or the His<sub>6</sub>-T7L tag from the *N*-terminus of PqsC. The mixture was dialyzed against buffer C containing β-mercaptoethanol (2 mM) in a dialysis bag (Thermo Scientific Snakeskin, 3.5 kDa MWCO) overnight at 4 °C. A second nickel affinity chromatography was performed using a 5 mL HisTrap HP column (GE Healthcare) with buffer C and buffer D [buffer C with imidazole (500 mM)] to separate cleaved and uncleaved protein complex and to remove the protease. Because in the case of His<sub>6</sub>-PqsBC TEV protease cleavage efficiency was low, the majority of PqsC still exhibited the *N*-terminal hexahistidine tag. Therefore, uncleaved His<sub>6</sub>-PqsBC and cleaved PqsBC fractions were pooled separately. In contrast, 3C protease removed the His<sub>6</sub>-T7L tag from PqsC with high efficiency. Protein fractions were concentrated (Sartorius Vivaspin, 30 kDa MWCO), loaded onto a HiLoad 26/60 Superdex 200 column (GE Healthcare) and eluted with buffer C. Fractions containing pure His<sub>6</sub>-PqsBC or PqsBC were concentrated (Sartorius Vivaspin, 30 kDa MWCO) to 20–35 mg mL<sup>-1</sup>, flash-cooled in liquid nitrogen and stored at –80 °C.

Another batch of His<sub>6</sub>-tagged PqsBC was purified according to the same protocol, but instead of a second nickel affinity chromatography, His<sub>6</sub>-PqsBC complex was purified by anion exchange chromatography (Table 5.S5). For this, His<sub>6</sub>-PqsBC was applied onto a 5 mL HiTrap Q FF column (GE Healthcare) in buffer C after dialysis and eluted with a NaCl gradient of 17–22% buffer D [buffer C with NaCl (1 M)]. For size-exclusion chromatography buffer E was used. Untagged or His<sub>6</sub>-tagged PqsBC variants (His<sub>6</sub>-PqsBC<sup>C129A</sup>, PqsBC<sup>C129A</sup>, PqsBC<sup>C129S</sup>) were purified in the same manner as described for wildtype His<sub>6</sub>-PqsBC (pET26b-*pqsB*/pET19m-*pqsC*) or untagged PqsBC (pET26b-*pqsB*/p10\$-*pqsC*), respectively.

### PqsBC activity assay with 2-ABA and acyl-CoAs with different carbon chain lengths

The enzymatic activity of PqsBC with 2-ABA and different acyl-CoAs was measured in a spectrophotometric assay with 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB, Ellman's reagent) by monitoring the formation of free coenzyme A (CoASH) at 412 nm [3, 4]. For the quantification of CoASH product, a CoASH standard curve was prepared by mixing DTNB (200 mM, 1  $\mu$ L) with different concentrations of CoASH in the range of 0.01–0.2 mM in a total sample volume of 100  $\mu$ L in inactivation buffer [HEPES (20 mM), NaCl (150 mM), SDS (0.5%, *w/v*), pH 7.5].

Acyl-CoA or DTNB stock solutions were freshly prepared in MilliQ water (adjusted to pH 5.0 with HCl) or in DMSO, respectively. The concentration of each acyl-CoA stock ( $[\text{acyl-CoA}]_{\text{total}}$ ) was measured at 259 nm ( $\epsilon_{259} = 15000 \text{ M}^{-1} \text{ cm}^{-1}$ ) [5] and the concentration of free CoASH ( $[\text{CoASH}]_{\text{free}}$ ) already present in the respective acyl-CoA stock was calculated from the CoASH standard curve in the presence of DTNB. The concentration of intact, non-hydrolyzed acyl-CoA ( $[\text{acyl-CoA}]_{\text{intact}}$ ) was obtained by subtracting  $[\text{CoASH}]_{\text{free}}$  from  $[\text{acyl-CoA}]_{\text{total}}$ .

For the enzyme activity assay, PqsBC (1  $\mu$ M) was mixed with acyl-CoA (500  $\mu$ M) and 2-ABA (1 mM) in assay buffer [HEPES (20 mM), NaCl (150 mM), pH 7.5] and the samples were incubated at room temperature for 20 min in a total volume of 25  $\mu$ L. The reaction was stopped by adding 24.5  $\mu$ L of termination buffer [HEPES (20 mM), NaCl (150 mM), SDS (1%, *w/v*), pH 7.5]. Free thiols of released CoASH were reacted with DTNB (200 mM, 0.5  $\mu$ L) to the yellow 2-nitro-5-thiobenzoate ( $\text{TNB}^{2-}$ ) and its absorbance at 412 nm ( $A_{412}$ ) was measured with a NanoDrop 2000 UV-Vis spectrophotometer (Thermo Fisher Scientific). As a negative control, the C129A active site mutant PqsBC<sup>C129A</sup> was incubated with octanoyl-CoA and 2-ABA. The background signal at 412 nm, determined from blank samples in which PqsBC was inactivated with inactivation buffer [HEPES (20 mM), NaCl (150 mM), SDS (0.5%, *w/v*), pH 7.5] prior to the addition of 2-ABA and acyl-CoA substrates, was subtracted from the  $A_{412}$  value of the respective sample. Blank samples were treated in the same manner described above, but the reaction was terminated with inactivation buffer. The concentration of free CoASH produced in the enzymatic reaction was calculated from the CoASH standard curve. All acyl-CoAs were purchased from Larodan with the exception of acetyl-CoA, which was acquired from Sigma-Aldrich. DTNB was obtained from Thermo Fisher Scientific and CoASH was purchased from AppliChem.



### Crystallization, data collection, phasing and refinement

Initial crystallization conditions were identified by mixing His<sub>6</sub>-PqsBC, His<sub>6</sub>-PqsBC<sup>C129A</sup>, PqsBC, PqsBC<sup>C129A</sup> or PqsBC<sup>C129S</sup> solution (0.2 μL, 2.5 to 25 mg mL<sup>-1</sup>) with precipitant (0.2 μL), using a Honeybee 961 dispensing robot (Digilab Genomic Solutions) at 20 °C or an OryxNano robot (Douglas Instruments) at 4 °C in 96 well sitting drop INTELLI crystallization plates (Art Robbins Instruments) charged with precipitant reservoirs (60 μL). Several hits for His<sub>6</sub>-PqsBC, His<sub>6</sub>-PqsBC<sup>C129A</sup> and PqsBC<sup>C129S</sup> were identified at 20 °C within 3–13 days. Crystals were optimized with random and grid screens dispensed by the Formulatrix liquid handling system (Formulatrix), either with the same setup as described for initial crystallization screening, or in 24 well hanging drop VDX plates (Hampton Research) by mixing protein solution (1 μL) with precipitant (1 μL) equilibrated against precipitant reservoir (500 μL) at 20 °C. Information about cryoprotectants, final crystallization conditions and unit cell dimensions are summarized in Table 5.S1. Crystal morphologies are shown in Figure 5.S2. X-ray diffraction data were collected at 100 K on beamlines X06DA/X10SA at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) and on PETRA III beamline P11 at DESY (Hamburg, Germany). Diffraction data were indexed and integrated with XDS [6] or XDSAPP [7], and scaled with Aimless [8] of the CCP4 program suite [9]. Data quality and correct space group assignment of the processed data was assessed with Xtriage [10] of the Phenix software suite [11].

Since the PqsBC crystal structure by Drees *et al.* [12] was not available at the outset of these studies, initial phases were derived by a combination of molecular replacement and heavy atom derivatization. First, the molecular replacement pipeline BALBES [13] selected a crystal structure of the homodimeric FabH from *Aquifex aeolicus* VF5 (PDB ID: 2EBD; 18% and 25% sequence identity to PqsB and PqsC, respectively) to locate three out of four expected PqsBC heterodimers in a native dataset of crystal form 2 (C121,  $a = 151.5 \text{ \AA}$ ,  $b = 230.9 \text{ \AA}$ ,  $c = 134.3 \text{ \AA}$ ,  $\alpha = 90.0^\circ$ ,  $\beta = 113.9^\circ$  and  $\gamma = 90^\circ$ ). These initial phases were then used to calculate a single-wavelength anomalous diffraction (SAD) difference density map from a dataset obtained at 1.0077 Å from a crystal in the same space group but soaked with HgCl<sub>2</sub> (1 mM) for 16 h. The SAD difference map revealed the presence of four anomalous scatterers, three located close to the active site cysteine residues of the dimers positioned by BALBES and one lying in unoccupied space, hinting at the position of the missing PqsBC molecule. By feeding these four positions as mercury atoms into the phenix.autosol routine of the Phenix software suite [11, 14], it was possible to obtain protein fragments that allowed rough



pre-positioning of a fourth *A. aeolicus* VF5 FabH homodimer, which was then correctly oriented by phased molecular replacement in MOLREP [15] from the CCP4 software package [9]. After several rounds of density modification with DM [16], automated iterative model building and refinement with the ARP/wARP web service (<http://arpwarp.embl-hamburg.de>) [17] and with Buccaneer [18], the model was finalized by alternating steps of manual model building in COOT [19] and maximum-likelihood refinement cycles in phenix.refine [20]. In the last few refinement cycles, water molecules and ligands were included, and TLS (translation, liberation, screw) refinement [21] was enabled after automatically dividing the model into TLS groups in phenix.refine [20]. The final model was used to determine the structures of PqsBC variants belonging to other space groups by molecular replacement using Phaser [22] or by rigid body refinement implemented in phenix.refine [20]. The overall quality of model geometries was evaluated with MolProbity (<http://molprobity.biochem.duke.edu/>) [23]. Data collection and refinement statistics are listed in Table 5.S2. Final atomic coordinates and structure factor amplitudes have been deposited in the Protein Data Bank ([www.rscb.org](http://www.rscb.org)) [24] with accession codes 6ESZ, 6ET0, 6ET1, 6ET2 and 6ET3.

All figures showing PqsBC crystal structures were prepared with PyMOL (Schrödinger) [25], secondary structure elements were assigned with DSSP [26], interactions between PqsBC and ligands were analyzed with Discovery Studio Visualizer 4.5 (Biovia) [27] and protein cavities were calculated with the KVFinder PyMOL plugin (default settings) [28].

### Structural comparison of PqsBC in five different crystal forms

The crystal structures of the PqsBC variants His<sub>6</sub>-PqsBC, His<sub>6</sub>-PqsBC<sup>C129A</sup>, PqsBC<sup>C129S</sup> in crystal forms 1–4 determined in this study and of crystal form 5 determined by Drees *et al.* [12] contain 2, 4 or 8 copies of the PqsBC heterodimer in their asymmetric units (22 copies in total). For structural comparison, the single PqsBC chains were extracted from the PDB files, ligands and water molecules were removed. The individual heterodimers were analyzed with the Protein Structural Statistics Web Server (PSSweb; <http://pssweb.org>) [29, 30]. The *N*-terminal purification tag of PqsC was excluded from analysis. Prior to the calculation of Cartesian coordinate statistics from the ensemble of PDB files, a multiple sequence alignment was performed with Clustal Omega [31] and structures were superposed with Theseus [32] implemented in the PSSweb workflow [29, 30]. Finally, the individual PqsBC heterodimers were manually clustered

into open, intermediate and closed states based on the locations of helix  $\alpha 1$  and loop A/elongated helix  $\alpha 5$  (Table 5.S3).

### Modeling of PqsBC<sup>C129A</sup>-acyl-CoA complexes

PqsBC<sup>C129A</sup>-acyl-CoA complexes were modeled with the structure of His<sub>6</sub>-PqsBC<sup>C129A</sup> determined in this study (chains IJ of crystal form 3, PDB ID: 6ET2) and the X-ray coordinates of the related enzyme FabH from *Mycobacterium tuberculosis* (*mtFabH*) in complex with dodecanoyl-CoA (PDB ID: 1U6S [33]) using Molecular Operating Environment (MOE, Chemical Computing Group) [34]. In detail, protein and ligand coordinates of chain A of PDB ID 1U6S were superimposed on His<sub>6</sub>-PqsC<sup>C129A</sup> to reveal structural similarities between both proteins and to obtain a putative binding pose of acyl-CoA ligands in PqsC. Starting from this, the His<sub>6</sub>-PqsC<sup>C129A</sup>-acyl-CoA complexes were generated and geometrically optimized after removal of the co-crystallized MOPS buffer component. Initially, pronounced steric clashes between His<sub>6</sub>-PqsC<sup>C129A</sup> and the adenosyl-diphosphate moiety as well as the terminal hydrocarbon units of the dodecanoyl moiety were observed. Hence, the dodecanoyl-CoA ligand was truncated to octanoyl-CoA and the conformation of the diphosphate linker was adjusted to minimize unfavourable contacts. Then, the built-in QuickPrep function was applied using the AMBER10:EHT force field and default parameters. Subsequently, the restraints were removed and the structure was energy-minimized with the same force field. The dodecanoyl-CoA complexes was modeled based on this structure via addition of terminal hydrocarbon units followed by the same QuickPrep and energy minimization steps.

## Supporting Tables

Table 5.S1.: Final crystallization conditions and cryoprotectants for different PqsBC variants.

PqsBC variant	[Protein] (mg/ml)	Crystal form	Space group	Cell dimensions	Crystallization condition	Cryoprotectant
His <sub>6</sub> -PqsBC (batch 2)	10	1	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	a, b, c (Å): 62, 142, 172	0.1 M Bis-Tris (pH 6.6),	15% (v/v) MPD
				α, β, γ (°): 90, 90, 90	0.2 M Lithium sulfate, 16% (w/v) PEG 3350, 5% (v/v) MPD	
His <sub>6</sub> -PqsBC <sup>C129A</sup>	15	1	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	a, b, c (Å): 62, 141, 171	0.1 M Bis-Tris (pH 6.7),	5% (v/v) (2R,3R)- (-)-2,3-Butanediol
				α, β, γ (°): 90, 90, 90	0.2 M Lithium sulfate, 14% (w/v) PEG 3350, 5% (v/v) MPD	
His <sub>6</sub> -PqsBC (batch 1)	20	2	C121	a, b, c (Å): 152, 231, 134	0.1 M HEPES (pH 8.0),	–
				α, β, γ (°): 90, 114, 90	0.2 M Ammonium sulfate, 30% (w/v) Poly(acrylic acid sodium salt) 2100	
His <sub>6</sub> -PqsBC <sup>C129A</sup>	10	3	P12 <sub>1</sub> 1	a, b, c (Å): 121, 170, 148 α, β, γ (°): 90, 101, 90	0.1 M MOPS (pH 7.7), 0.1 M Magnesium acetate, 15% (w/v) PEG 6000	20% (v/v) Glycerol
PqsBC <sup>C129S</sup>	5	4	C121	a, b, c (Å): 205, 64, 119 α, β, γ (°): 90, 121, 90	0.1 M Bis-Tris (pH 6.5), 2 M Ammonium sulfate	10% (v/v) (2R,3R)- (-)-2,3-Butanediol

Crystal form 1 was optimized in a 24 well hanging drop setup, while all other crystals (crystal form 2–4) were improved in 96 well crystallization plates. MPD: 2-methyl-2,4-pentanediol; MOPS: 3-morpholinopropane-1-sulfonic acid; Bis-Tris: Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane. Different preparations (batch 1/batch 2) of His<sub>6</sub>-PqsBC were used for crystallization setup (purification details can be found in Table 5.S5).

**Table 5.S2.:** Crystallographic data collection and refinement statistics for different PqsBC variants.

Data set	His <sub>6</sub> -WT	His <sub>6</sub> -C129A	His <sub>6</sub> -WT	His <sub>6</sub> -WT (HgCl <sub>2</sub> )	His <sub>6</sub> -C129A	C129S
Crystal form	1	1	2	2	3	4
<b>Data collection</b>						
X-ray source <sup>[a]</sup>	SLS	SLS	SLS	SLS	PETRA III	SLS
Beamline	X06DA	X06DA	X06DA	X06DA	P11	X10SA
Detector	Pilatus 2MF	Pilatus 2MF	Pilatus 2MF	Pilatus 2MF	Pilatus 6MF	Pilatus 6MF
Wavelength (Å)	1.0000	1.0000	1.0000	1.0077	1.0332	1.0000
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	C121	C121	P12 <sub>1</sub> 1	C121
Resolution range (Å)	47.46-1.84 (1.87-1.84)	47.12-1.53 (1.56-1.53)	48.11-2.65 (2.70-2.65)	48.18-3.30 (3.39-3.30)	49.61-2.60 (2.64-2.60)	48.21-2.25 (2.31-2.25)
Unit-cell parameters						
a (Å)	62.0	61.7	151.5	146.3	121.4	205.2
b (Å)	142.4	141.4	230.9	231.9	170.2	63.8
c (Å)	171.1	170.7	134.3	134.2	148.4	119.1
α (°)	90.0	90.0	90.0	90.0	90.0	90.0
β (°)	90.0	90.0	113.9	114.1	101.0	120.8
γ (°)	90.0	90.0	90.0	90.0	90.0	90.0
Mosaicity (°)	0.10	0.10	0.37	0.19	0.08	0.11
Total No. of reflections	1087443 (51635)	1655372 (83877)	470948 (23877)	2616517 (182592)	1255674 (63696)	430035 (28204)
Unique reflections	132059 (6539)	224695 (11031)	121019 (5927)	61995 (4570)	181283 (8998)	63167 (4410)
Multiplicity	8.2 (7.9)	7.4 (7.6)	3.9 (4.0)	42.2 (40.0)	6.9 (7.1)	6.8 (6.4)
Completeness (%)	100.0 (100.0)	100.0 (99.8)	99.4 (99.0)	100.0 (100.0)	99.9 (99.9)	100.0 (100.0)
Mean I/σ(I)	14.1 (2.0)	14.1 (2.0)	10.7 (1.6)	13.0 (1.6)	12.0 (2.0)	11.9 (1.7)
R <sub>meas</sub> (%) <sup>[b]</sup>	12.8 (114.2)	8.1 (121.9)	13.1 (111.4)	36.4 (315.4)	14.4 (109.4)	12.7 (113.4)
R <sub>p.i.m.</sub> (%) <sup>[c]</sup>	4.4 (40.4)	2.9 (43.6)	6.5 (54.9)	5.6 (50.1)	5.4 (40.9)	4.8 (44.5)
CC <sub>1/2</sub> (%) <sup>[d]</sup>	99.8 (80.6)	99.0 (74.1)	99.5 (60.3)	99.8 (79.1)	99.6 (68.6)	99.8 (70.0)
Solvent content (%)	53	52	67	67	52	48
<b>Refinement</b>						
Resolution range (Å)	47.34-1.84 (1.86-1.84)	47.12-1.53 (1.55-1.53)	48.11-2.65 (2.68-2.65)		48.98-2.60 (2.63-2.60)	48.21-2.25 (2.29-2.25)
R <sub>work</sub> (%) <sup>[e]</sup>	15.0 (26.4)	14.0 (24.6)	18.1 (31.2)		18.4 (28.4)	16.9 (26.0)
R <sub>free</sub> (%) <sup>[e]</sup>	17.9 (32.3)	15.8 (28.0)	20.3 (34.0)		22.0 (32.8)	20.1 (28.8)
R.m.s. deviations						
Bonds (Å)	0.007	0.005	0.002		0.003	0.002
Angles (°)	0.829	0.902	0.547		0.596	0.546
No. of non H-atoms						
Protein	10970	11184	21000		42197	10176
Ligands	102	142	-		104	43
Water	1212	1518	719		825	512
Average B factors (Å <sup>2</sup> )						
Protein	30	26	64		53	51
Ligands	61	50	-		67	85
Water	40	40	48		36	46
Ramachandran plot (%)						
Favored regions	97.6	97.6	97		97.6	97.8
Outliers	0.2	0.2	0.2		0.2	0.2
MolProbity score <sup>[f]</sup>	1.17	1.27	1.08		0.99	0.96
PDB ID	6ESZ	6ET0	6ET1		6ET2	6ET3

Values in parentheses are for the highest resolution shell.

<sup>[a]</sup> SLS: Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland); PETRA III: Positron-Electron Tandem Ring Accelerator (DESY, Hamburg, Germany).

<sup>[b]</sup>  $R_{meas} = \frac{\sum_{hkl} [N(hkl)/(N(hkl) - 1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$ , where  $N(hkl)$  is the number of observations of the reflection with the index  $hkl$ ,  $I_i(hkl)$  is the intensity of the  $i$ th measurement of the reflection  $hkl$  and  $\langle I(hkl) \rangle$  is the mean intensity of multiple observations of the reflection  $hkl$  [35].

<sup>[c]</sup>  $R_{p.i.m.} = \frac{\sum_{hkl} [1/(N(hkl) - 1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$  [36].

<sup>[d]</sup>  $CC_{1/2}$  is the Pearson correlation coefficient between the intensities of two random half-data sets [37].

<sup>[e]</sup>  $R_{work/free} = \frac{\sum_{hkl} |F_{obs}(hkl) - F_{calc}(hkl)|}{\sum_{hkl} F_{obs}(hkl)}$ , where  $F_{obs}(hkl)$  and  $F_{calc}(hkl)$  are the observed and the calculated structure factor amplitudes, respectively.  $R_{free}$  is calculated for 5% of reflections that are excluded from structure refinement [38].

<sup>[f]</sup> MolProbity score: quality criterion including the clash score, the Ramachandran and rotamer statistics [23].

**Table 5.S3.:** Clustering of PqsBC heterodimers into open, closed and intermediate states.

Number	PqsBC variant	Space group	Crystal form	Chains	Conformation
1	Wildtype	P <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	1	AB	intermediate
2	Wildtype	P <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	1	CD	intermediate
3	Wildtype	C2	2	AB	intermediate
4	Wildtype	C2	2	CD	intermediate
5	Wildtype	C2	2	EF	intermediate
6	Wildtype	C2	2	GH	intermediate
7	C129A	P <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	1	AB	intermediate
8	C129A	P <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	1	CD	intermediate
9	C129A	P <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	5	AD	open
10	C129A	P <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	5	BC	open
11	C129A	P <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	5	EF	open
12	C129A	P <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	5	GH	open
13	C129A	P <sub>2</sub> <sub>1</sub>	3	AB	intermediate
14	C129A	P <sub>2</sub> <sub>1</sub>	3	CD	intermediate
15	C129A	P <sub>2</sub> <sub>1</sub>	3	EF	intermediate
16	C129A	P <sub>2</sub> <sub>1</sub>	3	GH	intermediate
17	C129A	P <sub>2</sub> <sub>1</sub>	3	IJ	closed
18	C129A	P <sub>2</sub> <sub>1</sub>	3	KL	intermediate
19	C129A	P <sub>2</sub> <sub>1</sub>	3	MN	intermediate
20	C129A	P <sub>2</sub> <sub>1</sub>	3	OP	intermediate
21	C129S	C2	4	AB	disordered
22	C129S	C2	4	CD	closed

Structures of PqsBC variants belonging to crystal form 1-4 were solved in this study, while the structure of PqsBC<sup>C129A</sup> in crystal form 5 was determined by Drees *et al.* [12] (PDB ID: 5DWZ). Classification of chains AB of PqsBC<sup>C129S</sup> was not possible because the mobile flap of PqsC<sup>C129S</sup> (chain A) was completely disordered in this heterodimer.

**Table 5.S4.:** Oligonucleotides for amplification of *pqsB* and *pqsC* and for site-directed mutagenesis of *pqsC*.

Primer	Sequence (5'→3')	Final plasmid(s)
<i>pqsB</i> _NdeI_for	GTA TTA CAT ATG TTG ATT CAG GCT GTG GGG	pET26b- <i>pqsB</i>
<i>pqsB</i> _BamHI_rev	TTA TTC GGA TCC TTA TGC ATG AGC TTC TCC	
<i>pqsC</i> _NdeI_for	GTG TCG CAT ATG CAT AAG GTC AAA CTG GC	pET19m- <i>pqsC</i> , p10\$- <i>pqsC</i>
<i>pqsC</i> _XhoI_rev	TAC TTA CTC GAG TCA GCA CAC CAG CAC CTC	
<i>pqsC</i> _C129A_for	CCA TTG GAT TCG CAG ATG GAG <b>GCG</b> GCC AGC TTC CTG CTC AAC CTG	pET19m- <i>pqsC</i> <sup>C129A</sup> , p10\$- <i>pqsC</i> <sup>C129A</sup>
<i>pqsC</i> _C129A_rev	CAG GTT GAG CAG GAA GCT GGC <b>CGC</b> CTC CAT CTG CGA ATC CAA TGG	
<i>pqsC</i> _C129S_for	CCA TTG GAT TCG CAG ATG GAG <b>TCT</b> GCC AGC TTC CTG CTC AAC CTG	pET19m- <i>pqsC</i> <sup>C129S</sup> , p10\$- <i>pqsC</i> <sup>C129S</sup>
<i>pqsC</i> _C129S_rev	CAG GTT GAG CAG GAA GCT GGC <b>AGA</b> CTC CAT CTG CGA ATC CAA TGG	

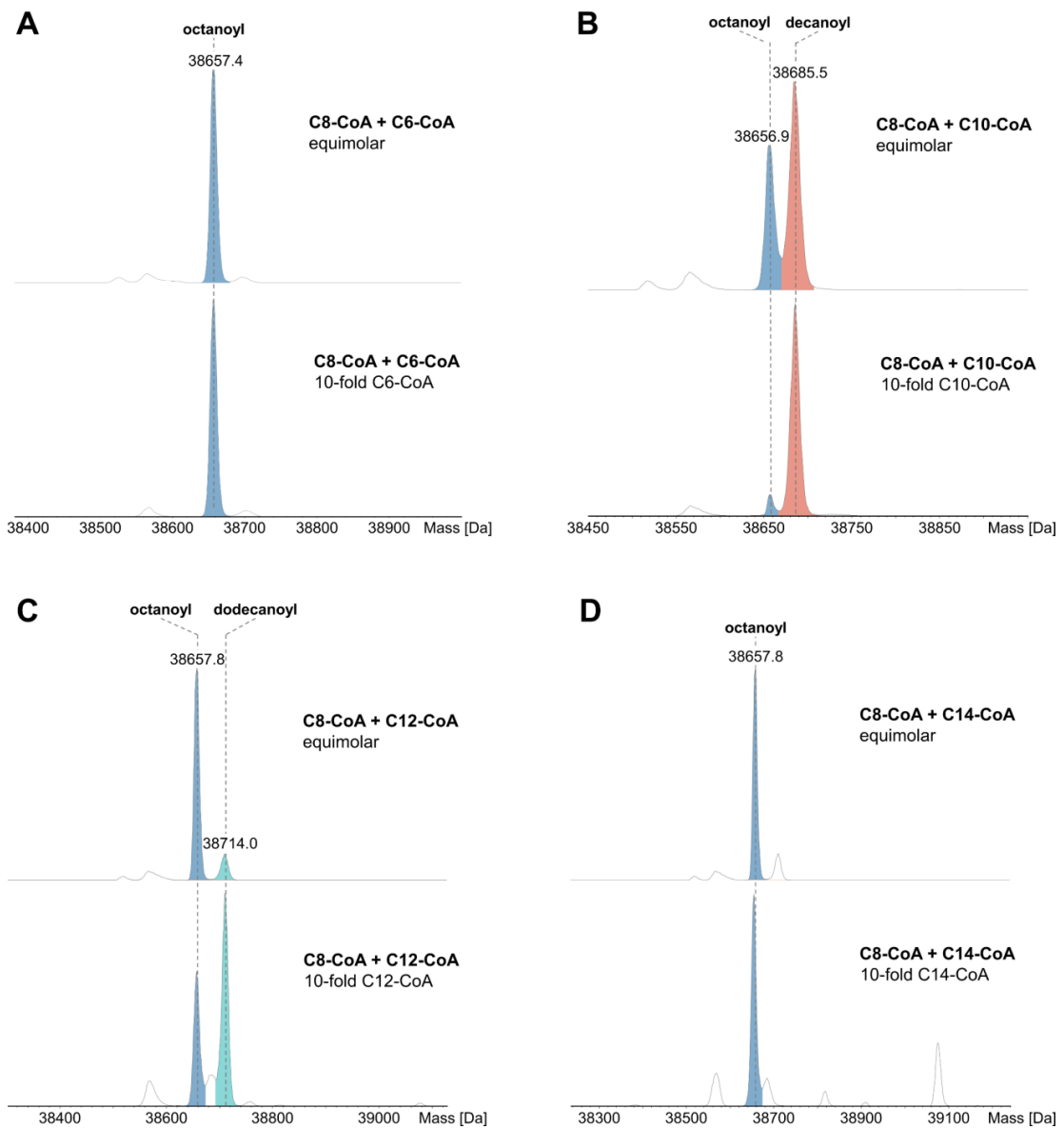
Restriction sites (*NdeI*, *BamHI*, *XhoI*) are written as bold characters and stop codons are underlined. The mutations (C129A or C129S) that were introduced by site-directed mutagenesis are highlighted in yellow.

**Table 5.S5.:** PqsBC variants purified in this study.

Plasmids	Expressed protein	Purification steps	Purification buffers	Purified protein
pET26b- <i>pqsB</i> pET19m- <i>pqsC</i>	His <sub>6</sub> -PqsBC	1st Ni <sup>2+</sup> -IMAC, TEV cleavage, 2nd Ni <sup>2+</sup> -IMAC, SEC	Buffer A (150 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> , 300 mM NaCl; pH 7.5) Buffer B (buffer A with 500 mM imidazole) Buffer C (20 mM HEPES, 150 mM NaCl; pH 7.5) Buffer D (buffer C with 500 mM imidazole)	His <sub>6</sub> -PqsBC and PqsBC (batch 1)
		1st Ni <sup>2+</sup> -IMAC, anion exchange, SEC	Buffer A (150 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> , 300 mM NaCl; 2 mM TCEP; pH 7.8) Buffer B (buffer A with 500 mM imidazole) Buffer C (15 mM Tris-HCl, 20 mM NaCl, 2 mM TCEP; pH 8.0) Buffer D (buffer C with 1 M NaCl) Buffer E (20 mM HEPES, 150 mM NaCl, 1 mM TCEP; pH 7.8)	His <sub>6</sub> -PqsBC (batch 2)
pET26b- <i>pqsB</i> p10 $\delta$ - <i>pqsC</i>	His <sub>6</sub> -T7L-PqsBC	1st Ni <sup>2+</sup> -IMAC, 3C cleavage, 2nd Ni <sup>2+</sup> -IMAC, SEC	Buffer A (150 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> , 300 mM NaCl, 2 mM TCEP; pH 7.8) Buffer B (buffer A with 500 mM imidazole) Buffer C (20 mM HEPES, 150 mM NaCl, 1 mM TCEP; pH 7.8) Buffer D (buffer C with 500 mM imidazole)	PqsBC (batch 2)
pET26b- <i>pqsB</i> pET19m- <i>pqsCC129A</i>	His <sub>6</sub> -PqsBC <sup>C129A</sup>	1st Ni <sup>2+</sup> -IMAC, TEV cleavage, 2nd Ni <sup>2+</sup> -IMAC, SEC	Buffer A (150 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> , 300 mM NaCl, 1 mM DTT; pH 7.5) Buffer B (buffer A with 500 mM imidazole) Buffer C (20 mM HEPES, 150 mM NaCl, 1 mM DTT; pH 7.5) Buffer D (buffer C with 500 mM imidazole)	His <sub>6</sub> -PqsBC <sup>C129A</sup> and PqsBC <sup>C129A</sup>
pET26b- <i>pqsB</i> p10 $\delta$ - <i>pqsCC129S</i>	His <sub>6</sub> -T7L-PqsBC <sup>C129S</sup>	1st Ni <sup>2+</sup> -IMAC, 3C cleavage, 2nd Ni <sup>2+</sup> -IMAC, SEC	Buffer A (150 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> , 300 mM NaCl, 2 mM TCEP; pH 7.8) Buffer B (buffer A with 500 mM imidazole) Buffer C (20 mM HEPES, 150 mM NaCl, 0.5 mM TCEP; pH 7.8) Buffer D (buffer C with 500 mM imidazole)	PqsBC <sup>C129S</sup>

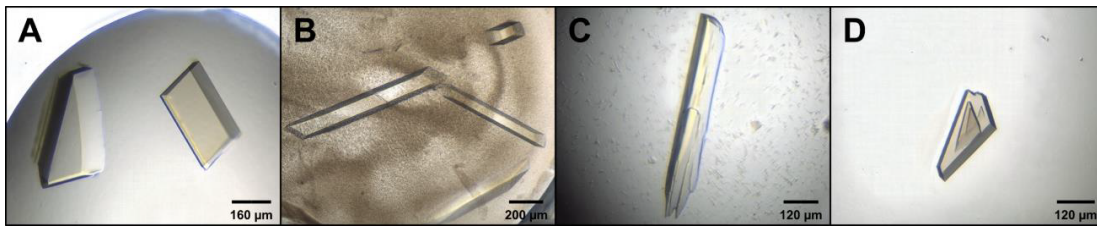
IMAC: immobilized metal ion affinity chromatography; SEC: size exclusion chromatography; DTT: dithiothreitol; TCEP: tris(2-carboxyethyl)phosphine. Different preparations (batch 1/batch 2) of His<sub>6</sub>-PqsBC were used for crystallization setup (Table 5.S1).

## Supporting Figures

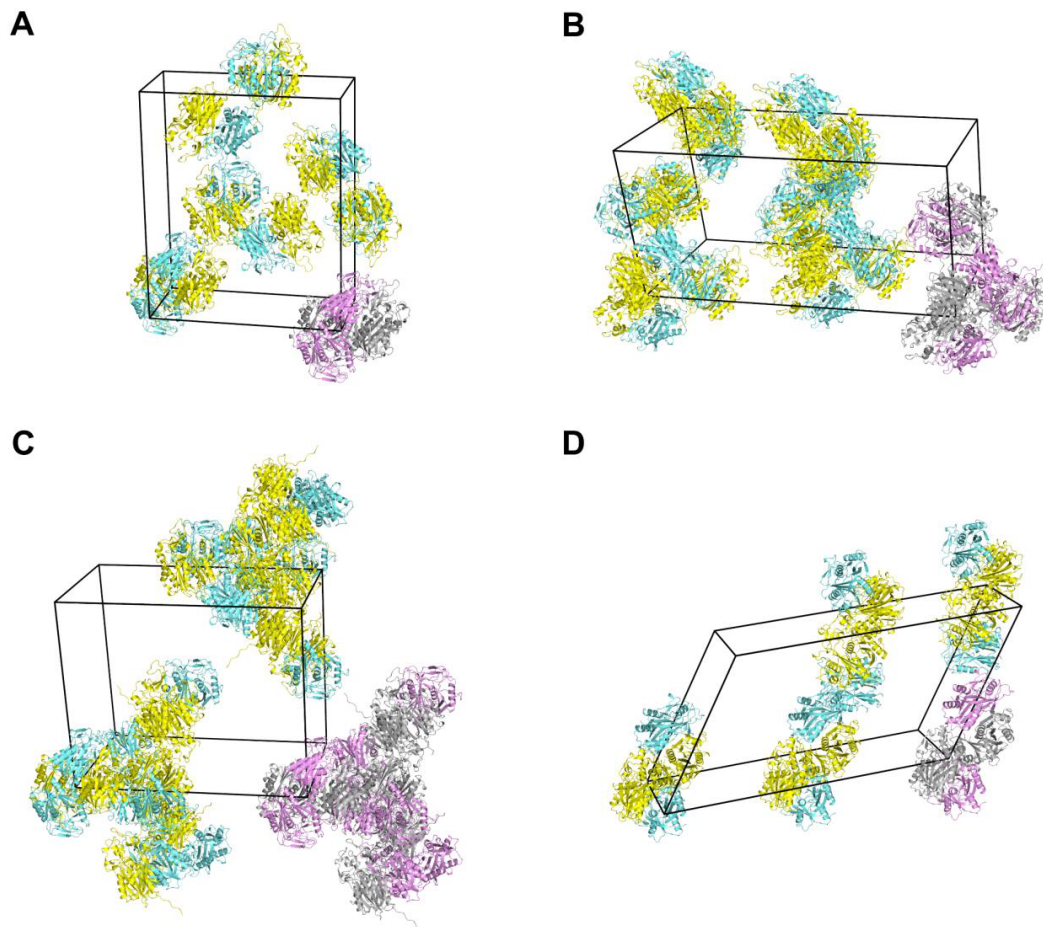


**Figure 5.S1.:** Deconvoluted mass spectra of PqsBC competitive acyl-CoA loading experiments. **(A)** C8- vs. C6-CoA (octanoyl- vs. hexanoyl-CoA). **(B)** C8- vs. C10-CoA (octanoyl- vs. decanoyl-CoA). **(C)** C8- vs. C12-CoA (octanoyl- vs. dodecanoyl-CoA). **(D)** C8- vs. C14-CoA (octanoyl- vs. tetradecanoyl-CoA). For each combination, an experiment with equimolar quantities of the two acyl-CoAs was performed as well as an experiment with 10-fold higher concentrations of hexanoyl-, decanoyl-, dodecanoyl- or tetradecanoyl-CoA over octanoyl-CoA.

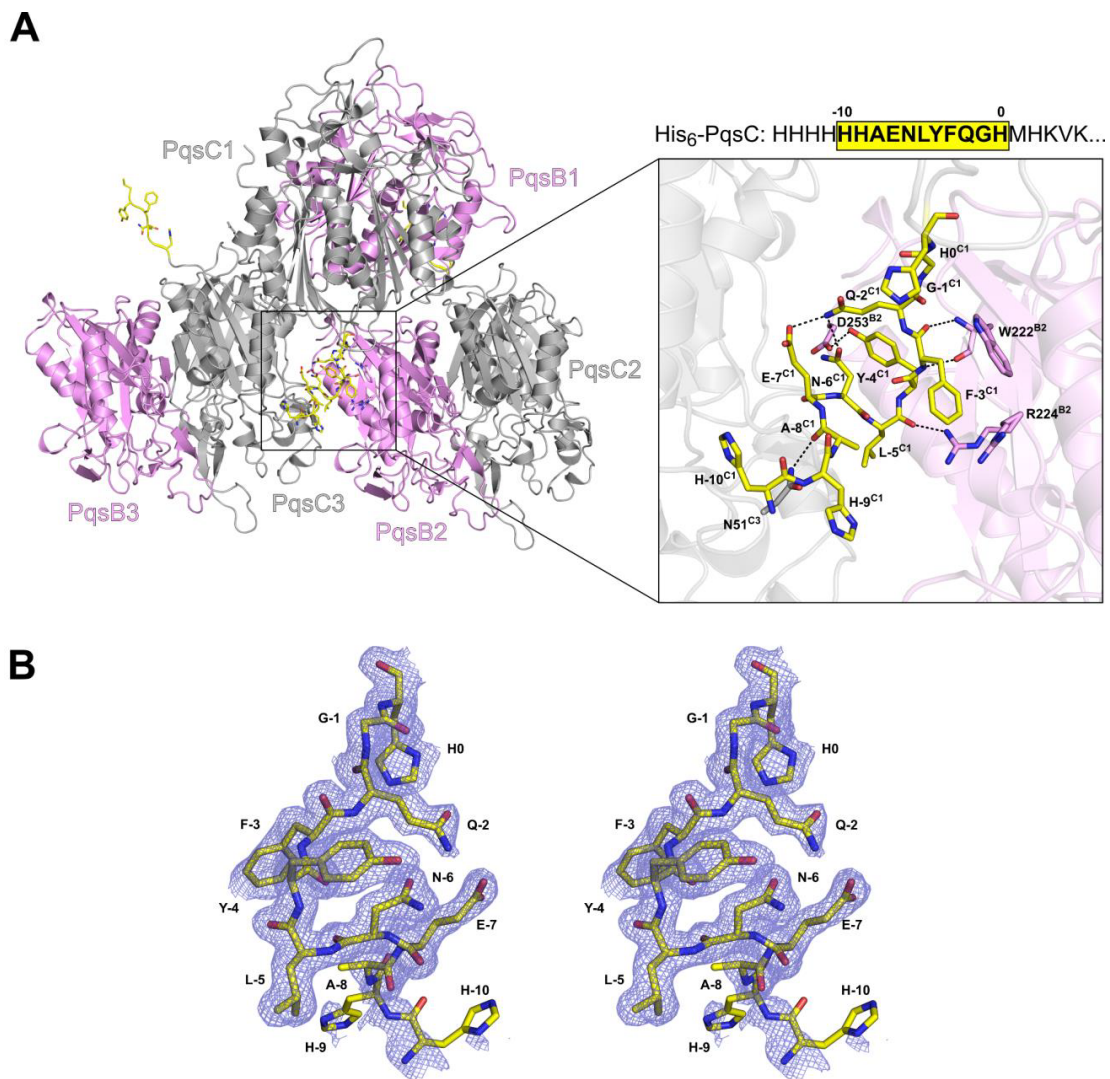




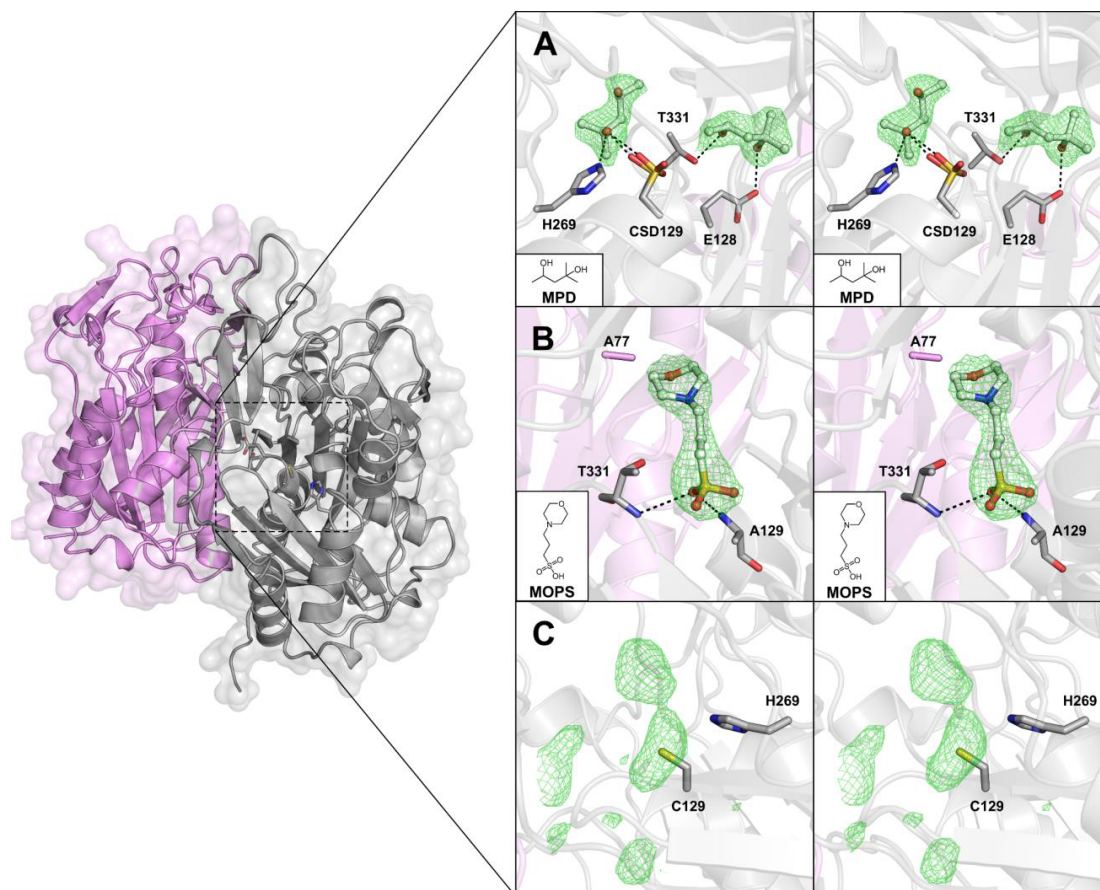
**Figure 5.S2.:** Crystal morphologies of different PqsBC variants. **(A)** His<sub>6</sub>-PqsBC<sup>C129A</sup> crystals (crystal form 1). **(B)** His<sub>6</sub>-PqsBC crystals (crystal form 2). **(C)** His<sub>6</sub>-PqsBC<sup>C129A</sup> crystal (crystal form 3). **(D)** PqsBC<sup>C129S</sup> crystal (crystal form 4). Crystals in A–D grew within 3 to 13 days. Crystallization conditions and cell dimensions of the different crystal forms are reported in Table 5.S1.



**Figure 5.S3.:** Crystal packing of different PqsBC variants. Unit cell representations of **(A)** crystal form 1 with two His<sub>6</sub>-PqsBC heterodimers in the asymmetric unit (ASU), **(B)** crystal form 2 with four His<sub>6</sub>-PqsBC/ASU, **(C)** crystal form 3 with eight His<sub>6</sub>-PqsBC<sup>C129A</sup>/ASU and of **(D)** crystal form 4 with two PqsBC<sup>C129S</sup>/ASU. PqsB (light magenta/cyan) and PqsC (silver/yellow) are shown in cartoon representation. PqsBC heterodimers in the ASUs are colored in light magenta and silver. Details about the unit cell dimensions of crystal forms 1–4 are reported in Table 5.S1.



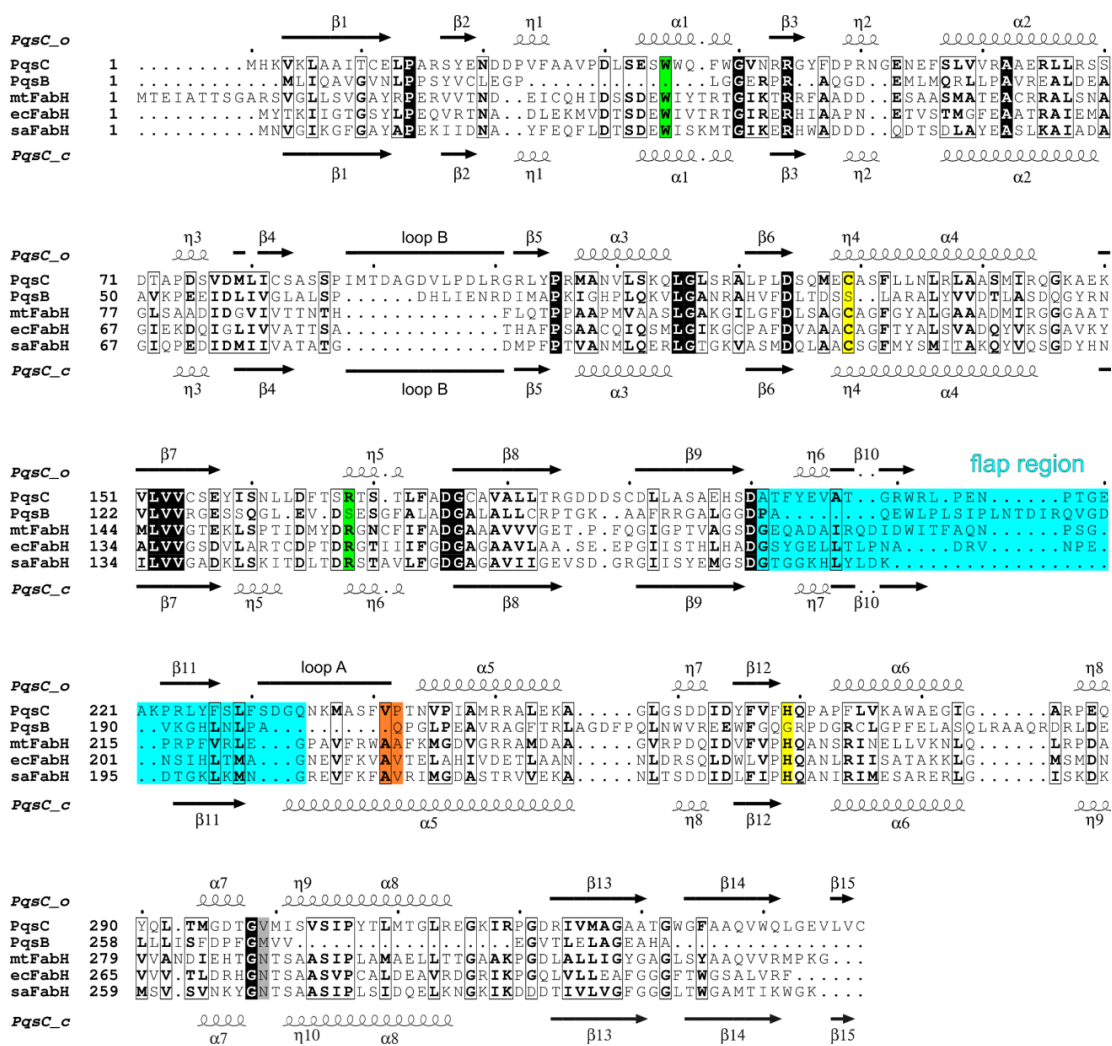
**Figure 5.S4.:** Crystal contacts of the TEV-protease cleavable His<sub>6</sub>-tag at the *N*-terminus of PqsC with symmetry related molecules in crystal form 1. **(A)** The overall structures of three PqsBC heterodimers (PqsBC1, PqsBC2 and PqsBC3) in the crystal lattice are shown. PqsC and PqsB monomers are colored in gray and light magenta, respectively. The His<sub>6</sub>-tags followed by TEV-protease recognition sites at the *N*-termini of the PqsC monomers are highlighted in yellow. The close-up view (right) depicts the purification tag of PqsC1 that interacts with neighboring PqsB2 and PqsC3 molecules in the crystal lattice. Hydrogen bonds (dashed lines) are formed between the backbone and side-chains of A-8<sup>C1</sup> ↔ N51<sup>C3</sup> and L-5<sup>C1</sup> ↔ R224<sup>B2</sup>, between side-chains of Y-4<sup>C1</sup> ↔ D253<sup>B2</sup> and between main-chain carbonyl and amide groups of F-3<sup>C1</sup> ↔ W222<sup>B2</sup>. In addition, the purification tag forms a cation-π contact with the guanidinium group of R224<sup>B2</sup> via F-3<sup>C1</sup>. Further hydrogen bonds between the side-chains of E-7<sup>C1</sup> ↔ Q-2<sup>C1</sup> and N-6<sup>C1</sup> ↔ Q-2<sup>C1</sup> stabilize the tag. **(B)** Stereo view of the 2mF<sub>o</sub>-DF<sub>c</sub> electron density map (blue mesh) of the purification tag at the *N*-terminus of PqsC (yellow) contoured at 0.7σ.



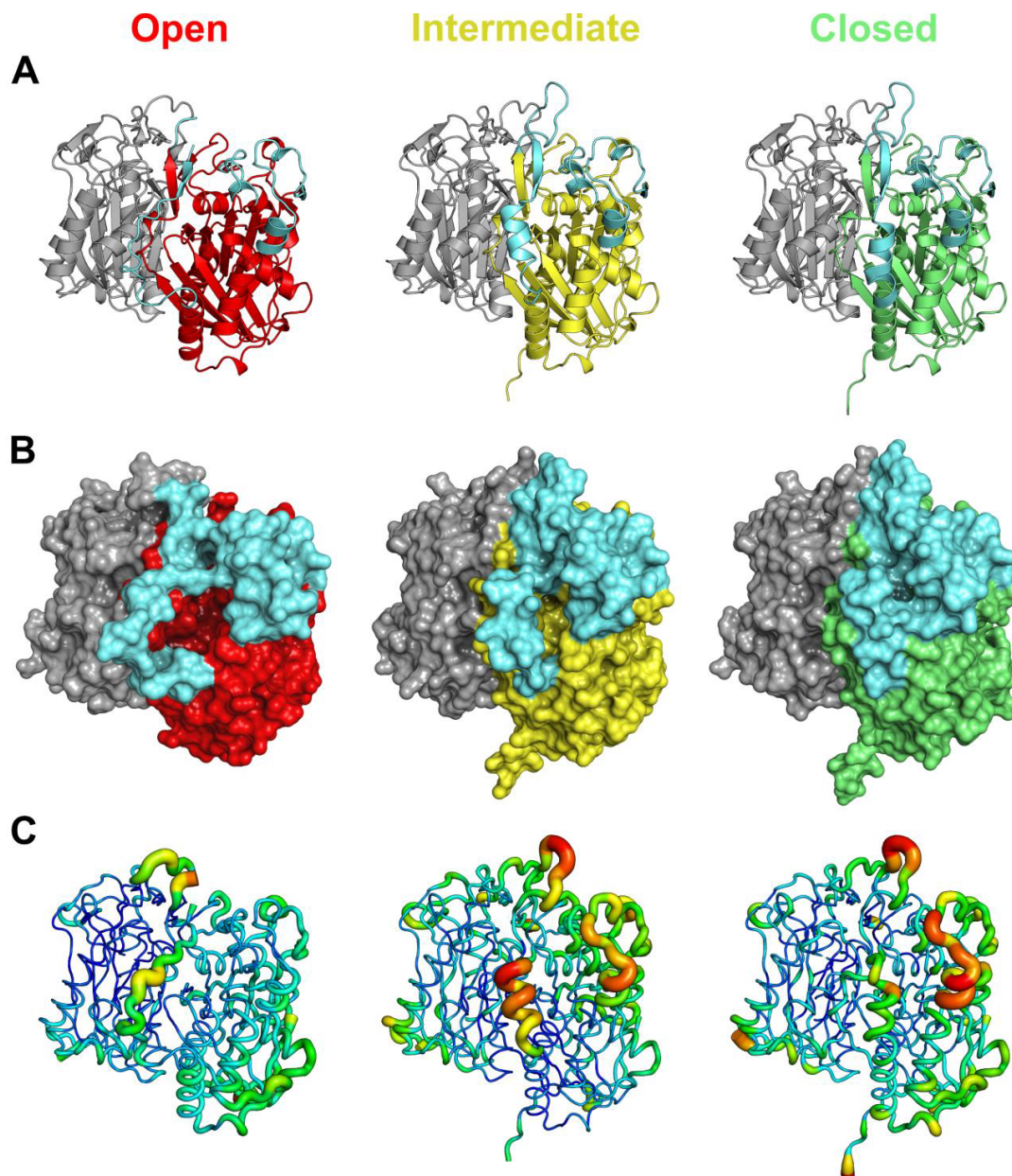
**Figure 5.S5.:** Ligands bound to the active site of PqsC. Stereo view of the active sites of **(A)** His<sub>6</sub>-PqsBC (crystal form 1) in complex with (4S)-2-methylpentane-2,4-diol (MPD) and of **(B)** His<sub>6</sub>-PqsBC<sup>C129A</sup> (crystal form 3) bound to MOPS. MPD and MOPS are components of the respective precipitant. Hydrogen bonds are depicted with dashed black lines. CSD129 in (A) is a sulfinic acid resulting from oxidation of C129. **(C)** Stereo plot of an unknown modification at C129 in the active site of His<sub>6</sub>-PqsBC (crystal form 2). PqsB (light magenta) and PqsC (silver) are shown in cartoon representations. Unbiased difference electron density  $mF_o - DF_c$  maps (green meshes) of bound ligands are displayed at  $3\sigma$ .



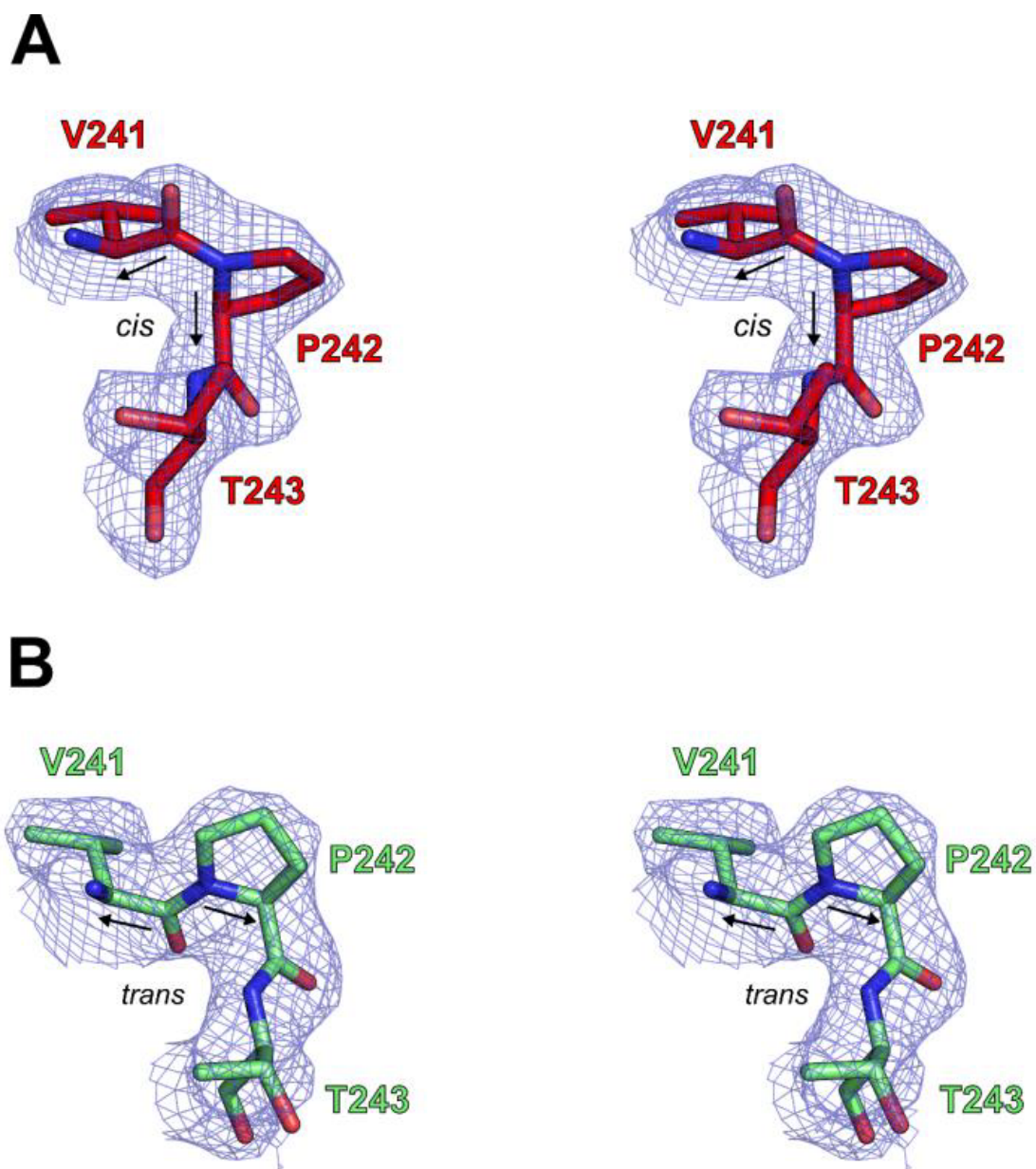
Chapter 5: Publication 3:  
*PqsBC* determines the AQ repertoire of *P. aeruginosa*



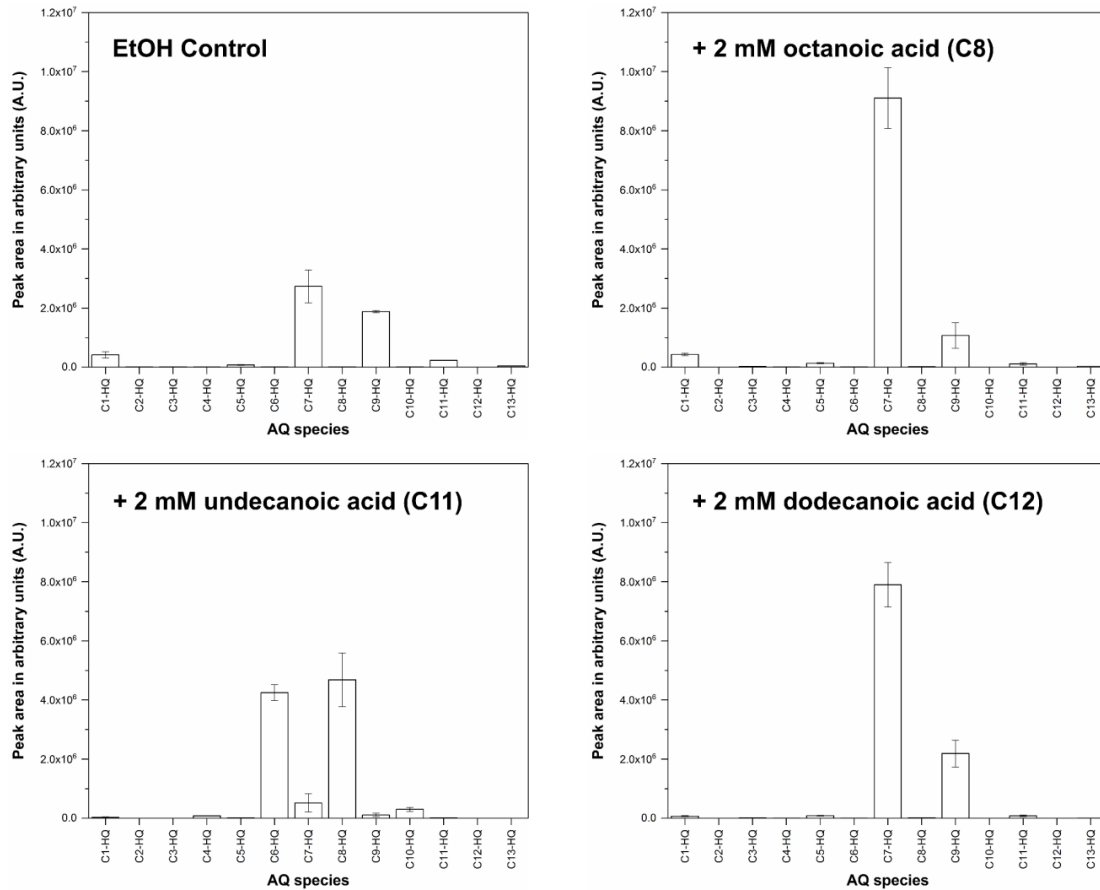
**Figure 5.S6.:** Multiple sequence alignment of PqsC (Uniprot ID: Q9I4X1), PqsB (Uniprot ID: Q9I4X2) and FabH enzymes from *Mycobacterium tuberculosis* (*mtFabH*; Uniprot ID: P9WNG3), *Escherichia coli* (*ecFabH*; Uniprot ID: P0A6R0) and from *Staphylococcus aureus* (*saFabH*; Uniprot ID: Q8NXE2). The alignment was performed with Clustal Omega [31] and processed using ESript3.0 (<http://esript.ibcp.fr>) [39]. Secondary structure elements ( $\alpha$ :  $\alpha$ -helix;  $\beta$ :  $\beta$ -strand;  $\eta$ :  $3_{10}$  helix) of PqsC crystal structures in the open (*PqsC\_o*) and in the closed form (*PqsC\_c*) are drawn at the top and at the bottom of the alignment, respectively. Strictly conserved residues (arginine, tryptophan) involved in acyl-CoA binding are highlighted in green, the catalytic triad is displayed in yellow (cysteine and histidine) and gray (asparagine; absent in PqsC). The mobile flap suggested to undergo large conformational rearrangements during the catalytic cycle of *mtFabH* is indicated in cyan [40]. The V241-P242 prolyl peptide bond, which adopts cis and trans configurations in *PqsC\_o* and *PqsC\_c*, respectively, is colored in orange. In *PqsC\_o*, helix  $\alpha 5$  is notably shorter than in *PqsC\_c*. PqsB lacks all conserved residues of FabH enzymes. Identical residues are in white letters boxed in black and similar residues are in bold black letters.



**Figure 5.S7.:** Conformational states of PqsBC. **(A)** Cartoon, **(B)** surface and **(C)** B-factor putty representations of open (red; PDB ID: 5DWZ [12]), intermediate (yellow; PDB ID: 6ET1; this article) and closed (green; PDB ID: 6ET2; this article) conformations of PqsBC. PqsB in (A) and (B) is colored in gray. Structural elements of PqsC undergoing large conformational changes between the open and the closed forms are highlighted in light blue. (C) The backbone atoms of these mobile regions have also higher B-factors (thick red tubes) compared to rigid areas (thin dark blue tubes).

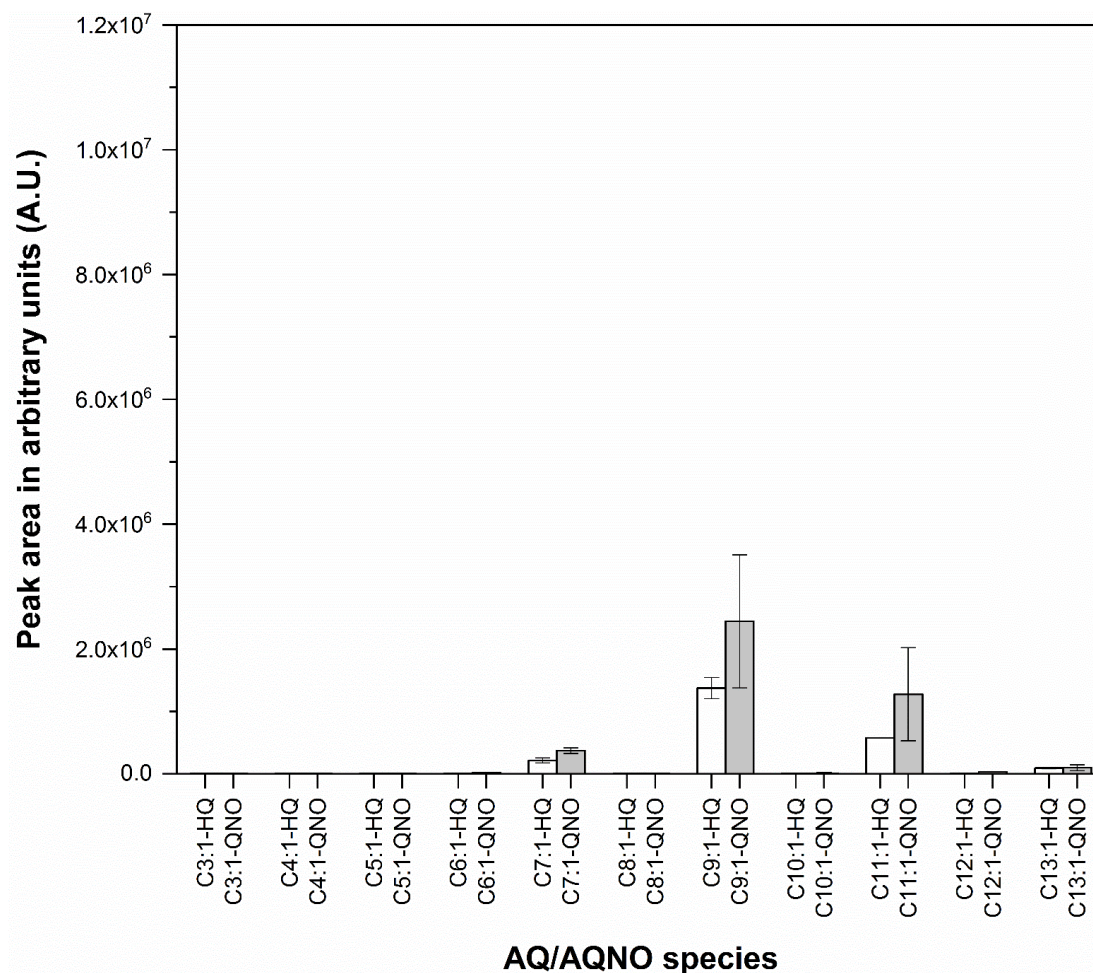


**Figure 5.S8.:** Stereo plots of the V241-P242 peptide bond in *cis* (A, open PqsBC) and *trans* configuration (B, closed PqsBC) with  $2mF_o-DF_c$  electron density maps (blue mesh) contoured at  $1\sigma$ .



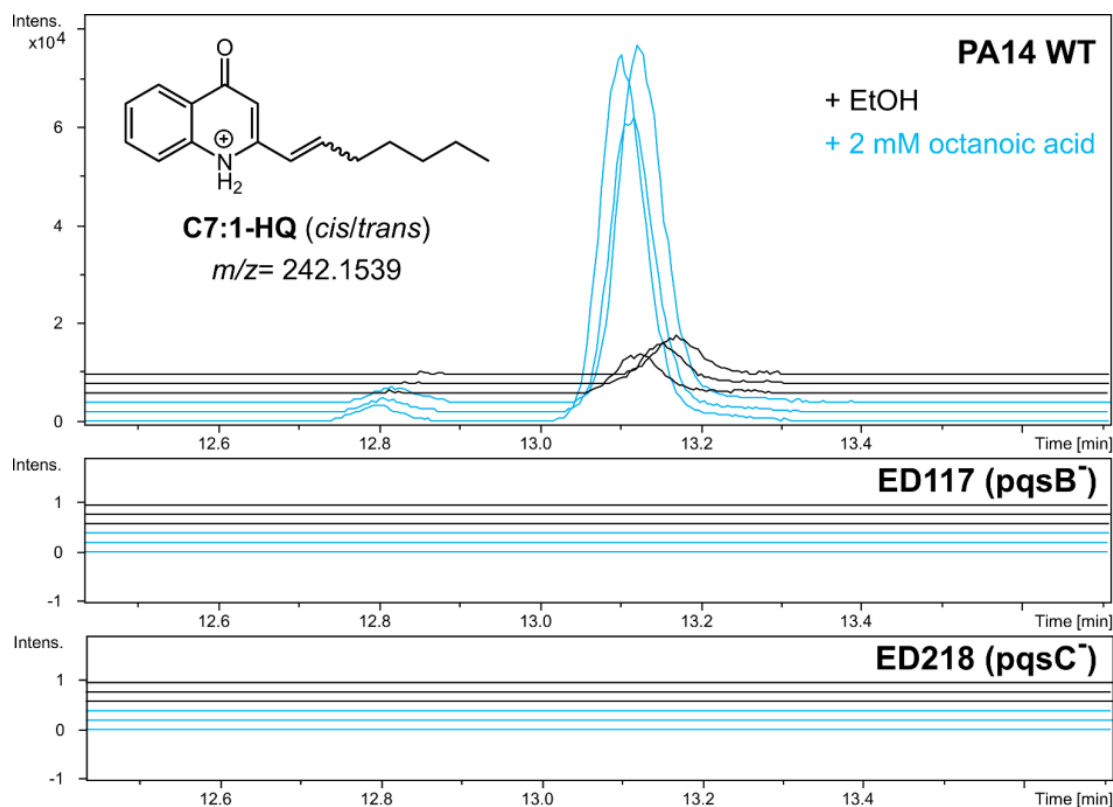
**Figure 5.S9.:** Effect of exogenous fatty acid addition on the AQ profile of *P. aeruginosa*. Mean peak areas of Aqs are in arbitrary units (A.U.). Exponentially growing wildtype *P. aeruginosa* PA14 were fed with EtOH (control), octanoic, undecanoic or dodecanoic acid (2 mM) and both cellular and extracellular metabolites were extracted. The error bars are the standard deviations from three independent measurements. The scaling of the y-axis is the same as in Figure 5.7A of the main text. In contrast to medium-chain fatty acids such as octanoic acid (C8), fatty acids longer than decanoic acid (C10) are not directly incorporated into the respective one carbon shorter AQ, but they are successively degraded in C2 units (most probably via  $\beta$ -oxidation) until they reach a length that can be used by PqsBC.



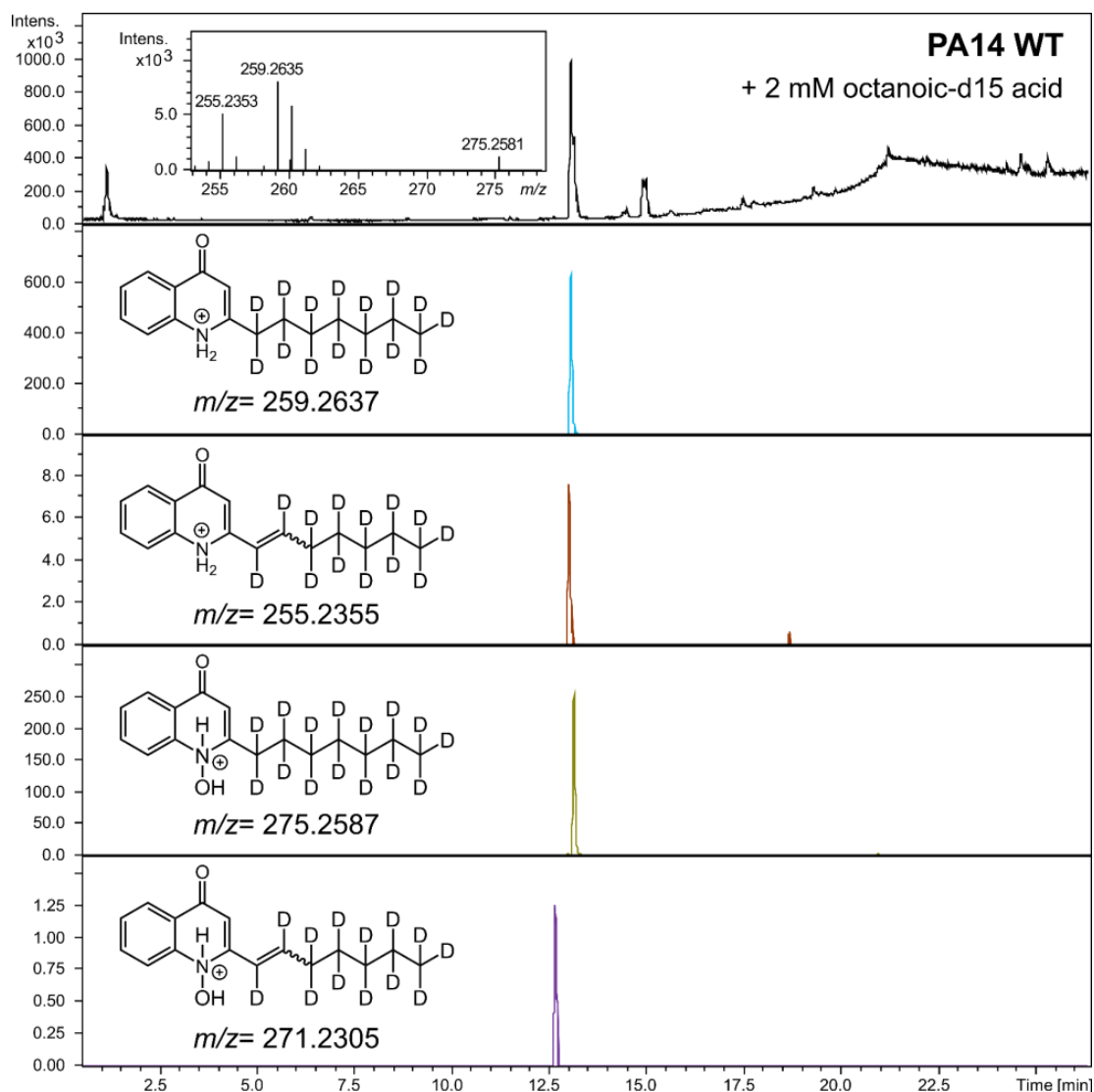


**Figure 5.S10.:** Mono-unsaturated Aqs/AQNOs in combined cell and supernatant extracts of exponentially growing wildtype *P. aeruginosa* PA14. Mean peak areas of cis/trans mono-unsaturated Aqs (white bars) or AQNOs (gray bars) are in arbitrary units (A.U.). The error bars are the standard deviations from three independent measurements. The scaling of the y-axis is the same as shown in main text Figure 5.1B for the respective saturated Aqs/AQNOs.





**Figure 5.S11.:** Addition of octanoic acid increased the levels of C7:1-HQ in *P. aeruginosa* PA14. Extracted ion chromatograms of  $m/z = 242.1539$  (mass-to-charge ratio for C7:1-HQ) of cell extracts of exponentially growing wildtype *P. aeruginosa* PA14 (first panel, PA14 WT), ED117 (second panel,  $pqsB^-$ ) and ED218 (third panel,  $pqsC^-$ ) mutant strains supplemented with EtOH (control, black) or octanoic acid (2 mM, blue). Each condition was measured in triplicates. The two peaks in the extracted ion diagram of PA14 WT derive from the *cis* and *trans* isomers of C7:1-HQ. PA14 WT cultures supplemented with octanoic acid exhibit higher levels of C7:1-HQ compared to the EtOH controls, while C7:1-HQ is not produced in the  $pqsB^-$  and  $pqsC^-$  mutant strains.



**Figure 5.S12.:** Addition of fully deuterated octanoic-d15 acid led to deuterated C7-HQ/C7-QNO and C7:1-HQ/C7:1-QNO species. Total ion chromatogram with full scan mass spectrum of the chromatogram region where C7-HQ, C7-QNO and C7:1-HQ co-elute (first panel) and extracted ion chromatograms of deuterated C7-HQ (second panel), C7:1-HQ (third panel), C7-QNO (fourth panel) and C7:1-QNO (fifth panel) of cell extracts of exponentially growing wildtype *P. aeruginosa* PA14 (PA14 WT) supplemented with octanoic-d15 acid (2 mM). The three most abundant ions in the full scan mass spectrum (first panel) represent C7-HQ ( $m/z = 259.2635$ ), C7-QNO ( $m/z = 275.2581$ ) and C7:1-HQ ( $m/z = 255.2353$ ) exhibit mass shifts of 15 Da (saturated congeners) or 13 Da (mono-unsaturated congeners), showing that octanoic-d15 acid was incorporated. Theoretical  $m/z$  values are indicated in the respective panel for each AQ/AQNO species; extracted ion chromatograms were constructed with  $\pm 0.002$  Da tolerance from the theoretical  $m/z$ .

## Supplementary References

- [1] T. Depke, R. Franke, and M. Brönstrup. “Clustering of MS<sup>2</sup> spectra using unsupervised methods to aid the identification of secondary metabolites from *Pseudomonas aeruginosa*”. In: *Journal of Chromatography B* 1071 (Dec. 2017), pp. 19–28. DOI: 10.1016/j.jchromb.2017.06.002.
- [2] T. Bock, E. Luxenburger, J. Hoffmann, V. Schütza, C. Feiler, R. Müller, and W. Blankenfeldt. “AibA/AibB Induces an Intramolecular Decarboxylation in Isovalerate Biosynthesis by *Myxococcus xanthus*”. In: *Angewandte Chemie International Edition* 56.33 (May 2017), pp. 9986–9989. DOI: 10.1002/anie.201701992.
- [3] G. L. Ellman. “Tissue sulfhydryl groups”. In: *Archives of Biochemistry and Biophysics* 82.1 (May 1959), pp. 70–77. DOI: 10.1016/0003-9861(59)90090-6.
- [4] P. W. Riddles, R. L. Blakeley, and B. Zerner. “Reassessment of Ellman’s reagent.” In: *Methods in enzymology* 91 (1983), pp. 49–60. ISSN: 0076-6879.
- [5] P. G. Killenberg and D. F. Dukes. “Coenzyme A derivatives of bile acids-chemical synthesis, purification, and utilization in enzymic preparation of taurine conjugates.” In: *Journal of lipid research* 17 (5 Sept. 1976), pp. 451–455. ISSN: 0022-2275.
- [6] W. Kabsch. “XDS”. In: *Acta Crystallographica Section D Biological Crystallography* 66.2 (Jan. 2010), pp. 125–132. DOI: 10.1107/s0907444909047337.
- [7] M. Krug, M. S. Weiss, U. Heinemann, and U. Mueller. “XDSAPP: a graphical user interface for the convenient processing of diffraction data using XDS”. In: *Journal of Applied Crystallography* 45.3 (Apr. 2012), pp. 568–572. DOI: 10.1107/s0021889812011715.
- [8] P. R. Evans and G. N. Murshudov. “How good are my data and what is the resolution?” In: *Acta Crystallographica Section D Biological Crystallography* 69.7 (June 2013), pp. 1204–1214. DOI: 10.1107/s0907444913000061.
- [9] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson. “Overview of the CCP4 suite and current developments”. In: *Acta Crystallographica Section D Biological Crystallography* 67.4 (Mar. 2011), pp. 235–242. DOI: 10.1107/s0907444910045749.
- [10] P. H. Zwart, R. W. Grosse-Kunstleve, P. D. Adams, *CCP4 newsletter*. 2005, Contribution 7.
- [11] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. “PHENIX: a comprehensive Python-based system for macromolecular structure solution”. In: *Acta Crystallographica Section D Biological Crystallography* 66.2 (Jan. 2010), pp. 213–221. DOI: 10.1107/s0907444909052925.
- [12] S. L. Drees, C. Li, F. Prasetya, M. Saleem, I. Dreveny, P. Williams, U. Hennecke, J. Emsley, and S. Fetzner. “PqsBC, a Condensing Enzyme in the Biosynthesis of the *Pseudomonas aeruginosa* Quinolone Signal”. In: *Journal of Biological Chemistry* 291.13 (Jan. 2016), pp. 6610–6624. DOI: 10.1074/jbc.m115.708453.
- [13] F. Long, A. A. Vagin, P. Young, and G. N. Murshudov. “BALBES: a molecular-replacement pipeline”. In: *Acta Crystallographica Section D Biological Crystallography* 64.1 (Dec. 2007), pp. 125–132. DOI: 10.1107/s0907444907050172.

- [14] T. C. Terwilliger, P. D. Adams, R. J. Read, A. J. McCoy, N. W. Moriarty, R. W. Grosse-Kunstleve, P. V. Afonine, P. H. Zwart, and L.-W. Hung. “Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard”. In: *Acta Crystallographica Section D Biological Crystallography* 65.6 (May 2009), pp. 582–601. DOI: 10.1107/s0907444909012098.
- [15] A. Vagin and A. Teplyakov. “Molecular replacement with MOLREP”. In: *Acta Crystallographica Section D Biological Crystallography* 66.1 (Dec. 2009), pp. 22–25. DOI: 10.1107/s0907444909042589.
- [16] K. Cowtan. “Error estimation and bias correction in phase-improvement calculations”. In: *Acta Crystallographica Section D Biological Crystallography* 55.9 (Sept. 1999), pp. 1555–1567. DOI: 10.1107/s0907444999007416.
- [17] G. Langer, S. X. Cohen, V. S. Lamzin, and A. Perrakis. “Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7”. In: *Nature Protocols* 3.7 (July 2008), pp. 1171–1179. DOI: 10.1038/nprot.2008.91.
- [18] K. Cowtan. “The Buccaneer software for automated model building. 1. Tracing protein chains”. In: *Acta Crystallographica Section D Biological Crystallography* 62.9 (Aug. 2006), pp. 1002–1011. DOI: 10.1107/s0907444906022116.
- [19] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan. “Features and development of Coot”. In: *Acta Crystallographica Section D Biological Crystallography* 66.4 (Mar. 2010), pp. 486–501. DOI: 10.1107/s0907444910007493.
- [20] P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, and P. D. Adams. “Towards automated crystallographic structure refinement with phenix.refine”. In: *Acta Crystallographica Section D Biological Crystallography* 68.4 (Mar. 2012), pp. 352–367. DOI: 10.1107/s0907444912001308.
- [21] V. Schomaker and K. N. Trueblood. “On the rigid-body motion of molecules in crystals”. In: *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry* 24.1 (Jan. 1968), pp. 63–76. DOI: 10.1107/s0567740868001718.
- [22] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. “Phaser crystallographic software”. In: *Journal of Applied Crystallography* 40.4 (July 2007), pp. 658–674. DOI: 10.1107/s0021889807021206.
- [23] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson. “MolProbity: all-atom structure validation for macromolecular crystallography”. In: *Acta Crystallographica Section D Biological Crystallography* 66.1 (Dec. 2009), pp. 12–21. DOI: 10.1107/s0907444909042073.
- [24] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. “The Protein Data Bank.” In: *Nucleic acids research* 28 (1 Jan. 2000), pp. 235–242. ISSN: 0305-1048.
- [25] *The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.*
- [26] W. Kabsch and C. Sander. “Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features”. In: *Biopolymers* 22.12 (Dec. 1983), pp. 2577–2637. DOI: 10.1002/bip.360221211.
- [27] *Discovery Studio Visualizer Version 4.5, Dassault Systèmes BIOVIA, San Diego: Dassault Systèmes.*
- [28] S. H. Oliveira, F. A. Ferraz, R. V. Honorato, J. Xavier-Neto, T. J. Sobreira, and P. S. de Oliveira. “KVFinder: steered identification of protein cavities as a PyMOL plugin”. In: *BMC Bioinformatics* 15.1 (2014), p. 197. DOI: 10.1186/1471-2105-15-197.

- [29] T. Gaillard, B. B. L. Schwarz, Y. Chebaro, R. H. Stote, and A. Dejaegere. “Protein Structural Statistics with PSS”. In: *Journal of Chemical Information and Modeling* 53.9 (Sept. 2013), pp. 2471–2482. DOI: 10.1021/ci400233j.
- [30] T. Gaillard, R. H. Stote, and A. Dejaegere. “PSSweb: protein structural statistics web server”. In: *Nucleic Acids Research* 44.W1 (May 2016), W401–W405. DOI: 10.1093/nar/gkw332.
- [31] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins. “Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega”. In: *Molecular Systems Biology* 7.1 (Apr. 2011), pp. 539–539. DOI: 10.1038/msb.2011.75.
- [32] D. L. Theobald and D. S. Wuttke. “THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures”. In: *Bioinformatics* 22.17 (June 2006), pp. 2171–2172. DOI: 10.1093/bioinformatics/bt1332.
- [33] F. Musayev, S. Sachdeva, J. N. Scarsdale, K. Reynolds, and H. Wright. “Crystal Structure of a Substrate Complex of *Mycobacterium tuberculosis*  $\beta$ -Ketoacyl-acyl Carrier Protein Synthase III (FabH) with Lauroyl-coenzyme A”. In: *Journal of Molecular Biology* 346.5 (Mar. 2005), pp. 1313–1321. DOI: 10.1016/j.jmb.2004.12.044.
- [34] *Molecular Operating Environment (MOE), 2013.08; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite No.910, Montreal, QC, Canada, H3A 2R7, 2018.*
- [35] K. Diederichs and P. A. Karplus. “Improved R-factors for diffraction data analysis in macromolecular crystallography”. In: *Nature Structural Biology* 4.4 (Apr. 1997), pp. 269–275. DOI: 10.1038/nsb0497-269.
- [36] M. S. Weiss. “Global indicators of X-ray data quality”. In: *Journal of Applied Crystallography* 34.2 (Apr. 2001), pp. 130–135. DOI: 10.1107/s0021889800018227.
- [37] P. A. Karplus and K. Diederichs. “Linking Crystallographic Model and Data Quality”. In: *Science* 336.6084 (May 2012), pp. 1030–1033. DOI: 10.1126/science.1218231.
- [38] A. T. Brünger. “Free R value: a novel statistical quantity for assessing the accuracy of crystal structures”. In: *Nature* 355.6359 (Jan. 1992), pp. 472–475. DOI: 10.1038/355472a0.
- [39] X. Robert and P. Gouet. “Deciphering key features in protein structures with the new ENDscript server”. In: *Nucleic Acids Research* 42.W1 (Apr. 2014), W320–W324. DOI: 10.1093/nar/gku316.
- [40] S. Sachdeva, F. N. Musayev, M. M. Alhamadsheh, J. N. Scarsdale, H. T. Wright, and K. A. Reynolds. “Separate Entrance and Exit Portals for Ligand Traffic in *Mycobacterium tuberculosis* FabH”. In: *Chemistry & Biology* 15.4 (Apr. 2008), pp. 402–412. DOI: 10.1016/j.chembiol.2008.03.007.



## 6 | The peptide chain release factor methyltransferase PrmC influences the *Pseudomonas aeruginosa* PA14 endo- and exometabolome

This Chapter has been published in a peer-reviewed scientific journal:

T. Depke, S. Häussler, and M. Brönstrup. “The Peptide Chain Release Factor Methyltransferase PrmC Influences the *Pseudomonas aeruginosa* PA14 Endo- and Exometabolome”. In: *Metabolites* 10.10 (Oct. 2020), p. 417. DOI: 10.3390/metabo10100417

### Abstract

*Pseudomonas aeruginosa* is one of the most important nosocomial pathogens and understanding its virulence is the key to effective control of *P. aeruginosa* infections. The regulatory network governing virulence factor production in *P. aeruginosa* is exceptionally complex. Previous studies have shown that the peptide chain release factor methyltransferase PrmC plays an important role in bacterial pathogenicity. Yet, the underlying molecular mechanism is incompletely understood. In this study, we used untargeted liquid and gas chromatography coupled to mass spectrometry to characterise the metabolome of a *prmC* defective *P. aeruginosa* PA14 strain in comparison with the corresponding strain complemented with *prmC* *in trans*. The comprehensive metabolomics data provided new insight into the influence of *prmC* on virulence and metabolism. *prmC* deficiency had broad effects on the endo- and exometabolome of *P. aeruginosa* PA14 with a marked decrease of the levels of aromatic compounds accompanied by reduced precursor supply from the shikimate pathway. Furthermore, a pronounced decrease of phenazine production was observed as well as lower abundance of alkylquinolones. Unexpectedly, the metabolomics data showed no *prmC* dependent effect on rhamnolipid production and an increase in pyochelin levels. A putative virulence biomarker identified in a previous study was significantly less abundant in the *prmC* deficient strain.

## Keywords

*Pseudomonas aeruginosa*; PrmC; metabolomics; virulence

### 6.1. Introduction

The ubiquitous  $\gamma$ -proteobacterium *Pseudomonas aeruginosa* is an important causative agent of infections in predisposed individuals such as those suffering from cystic fibrosis or burn wounds [1]. Deemed a priority antibiotic resistant bacterium by the WHO [2], it poses a considerable threat to public health, and causes a high number of nosocomial infections [3, 4]. *P. aeruginosa* possesses an arsenal of virulence factors that is intricately regulated by quorum sensing systems and activated in response to environmental factors via multiple signalling pathways [5]. While a large number of proteins and pathways have known roles in the regulation of virulence, there are enzymes that contribute to a virulent phenotype by an unknown mechanism. The S-adenosyl-L-methionine dependent peptide chain release factor methyltransferase PrmC (synonym: HemK) belongs to the latter group. Its absence results in a reduction of *in vivo* virulence—specifically the inability to produce pyocyanin—and an impaired adaptation to challenging environments [6]. Its original role is to posttranslationally methylate the release factors RF-1 and RF-2 (PrfA and PrfB in *P. aeruginosa*), thereby facilitating translation termination and decreasing the rate of stop-codon readthrough [7, 8]. Previous transcriptomic and proteomic analyses gave new insight and suggested that PrmC activity alters mRNA-protein ratios depending on which stop codon is used [6, 9]. In this study, we have applied metabolomics, the characterisation of the set of small-molecule primary and secondary metabolites, as a complementary technology to capture the effects of *prmC* deficiency in *Pseudomonas aeruginosa* PA14. The exometabolome, i. e., the metabolites secreted into and consumed from the growth media, was recorded along with the cellular metabolome, i. e., the endometabolome). Furthermore, liquid chromatography–mass spectrometry (LC-MS) and gas chromatography–mass spectrometry (GC-MS) were combined for advanced analytical coverage.

### 6.2. Materials and methods

To ensure comparability to previous research, the same bacterial strains as in Krueger et al. were used [9]: The *P. aeruginosa* PA14 *prmC* transposon mutant from the Harvard PA14 mutant library [10] carrying an empty pUCP20 vector, abbreviated PA14



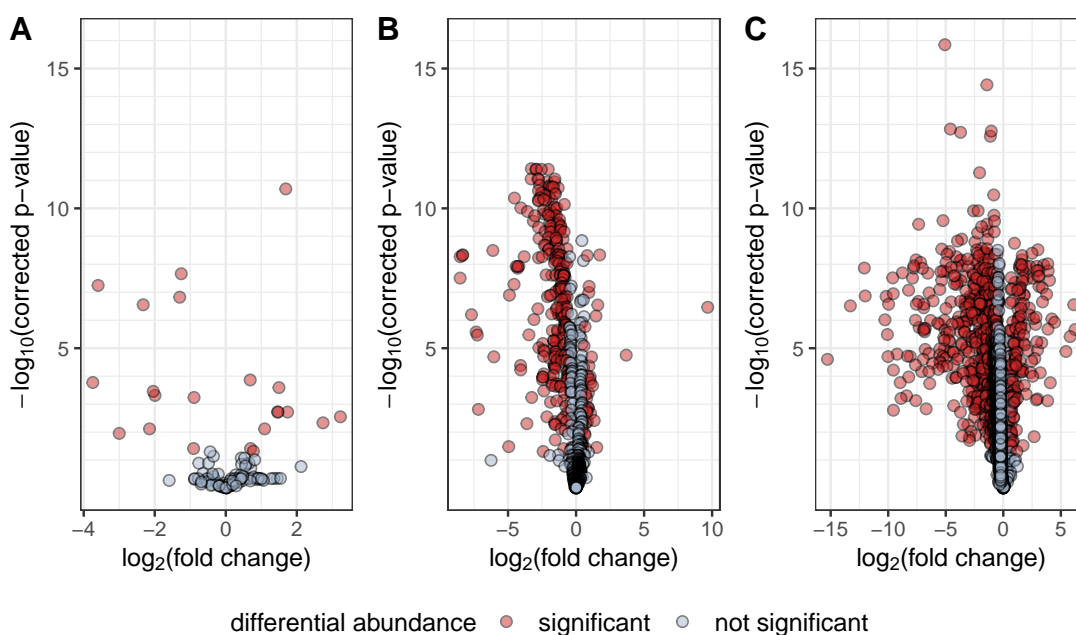
*tnprmC* and the same strain complemented with the *prmC* gene on the vector, PA14 *tnprmC::prmC*. In order to assure that all phenotypic effects are due to the absence or presence of the *prmC* gene, and to exclude that effects are caused by secondary mutations that might occur in addition to the inactivation of *prmC* in the transposon mutant, the transposon mutant was not compared to the corresponding wildtype, but to an equivalent, *prmC*-complemented transposon mutant. Five biological replicates of each strain were cultivated in glass flasks in 60 mL modified BM2 medium supplemented with additional amino acids (7 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 40 mM K<sub>2</sub>HPO<sub>4</sub>, 22 mM KH<sub>2</sub>PO<sub>4</sub>, 2 mM MgSO<sub>4</sub>, 10 μM FeSO<sub>4</sub> with 0.01% casamino acids) in a shaking incubator (approx. 150 rpm) at 37 °C. Blank samples (pure media without inoculation) were prepared accordingly and processed like the bacterial samples throughout the experiment. Bacteria were grown until an OD<sub>600</sub> of 1.6–1.7 before harvesting. Each biological replicate was split into two technical replicates for GC-MS analysis and two for LC-MS analysis by transferring 20 mL and 5 mL, respectively, to pre-cooled tubes, which were then immediately centrifuged at 9000 × *g* for 10 min at 4 °C. 1.5 mL supernatant was transferred from the LC-MS samples to fresh tubes and dried overnight in a centrifugal evaporator at 20 °C and full vacuum until complete dryness as described before [11]. The cell pellets for both LC-MS and GC-MS samples were snap frozen in liquid N<sub>2</sub>. Cell pellets for GC-MS analysis were resuspended in 1.5 mL 75% (v/v) methanol with 2 mg/L ribitol as an internal standard and supplemented with 1.5 mL deionised water and chloroform each, followed by vigorous shaking. The organic phase was discarded, and 1 mL of the aqueous phase was evaporated to dryness. Sample preparation, GC-MS analysis, data pre-processing and metabolite identification were performed as described by Berndt et al. [12]: Metabolite extracts were derivatised by methoxyamine/pyridine and silylated using MSTFA to volatilise the analytes. The samples were separated on a 30 m column with a stationary phase of 5% phenyl arylen and 95% dimethylpolysiloxane material (ZB-5MS<sup>®</sup>, Phenomenex, Torrance, CA, USA) using a temperature gradient from 70 °C to 330 °C over 32.5 min. MS data was recorded in full scan mode by electron impact ionisation in positive mode on an iontrap mass spectrometer (Thermo Scientific ITQ 900<sup>™</sup>, Thermo Fisher Scientific, Waltham, MA, USA). A mix of alkanes (chain length 10 to 36) was used to calibrate the retention index of the analytes. Spectral deconvolution, alignment and annotation were achieved using the MetaboliteDetector software [13] and an in-house spectral library. Directional fold changes were calculated by comparison of the mean feature intensities of the two strains and *p*-values were calculated using Welch's *t*-test with Benjamini-Hochberg correction for multiple testing.

Four replicates (two of *tnprmC* samples and two of *tnprmC::prmC* samples) were lost for data analysis due to a technical problem with the GC-MS instrument. Cell pellets and dried supernatants for LC-MS analysis were extracted using 1 mL and 500  $\mu$ L 100% methanol containing 0.1 mg/L trimethoprim, 0.1 mg/L nortriptyline and 0.3 mg/L glipizide as internal standards, respectively, by vigorous shaking, followed by sonication in a cooled ultrasonic bath for 15 min. After centrifugation, 800  $\mu$ L and 400  $\mu$ L, respectively, were transferred to fresh tubes, evaporated to dryness and reconstituted in 80  $\mu$ L / 40  $\mu$ L 50% acetonitrile containing 1 mg/L caffeine and 8 mg/L naproxen as internal standards, respectively, sonicated and centrifuged again and then used as LC-MS samples. Pooled samples containing aliquots of all samples were generated for quality control purposes. LC-MS analysis, pre-processing using XCMSonline [14], metabolite identification and statistical analysis were performed as described previously [11, 15]. In brief, the metabolite extracts were separated by ultra-high performance liquid chromatography on a C18 reversed phase column by gradient elution with H<sub>2</sub>O plus 0.1% formic acid and acetonitrile plus 0.1% formic acid. The MS data was recorded by a quadrupole time-of-flight mass spectrometer (maXis<sup>TM</sup>HD QTOF, Bruker, Bremen, Germany) after electrospray ionisation in positive mode. Pooled samples were additionally subjected to collision-induced dissociation by Bruker's data-dependent auto-MS/MS functionality to record tandem mass spectra for metabolite identification by comparison to authentic chemical standards and/or metabolite databases. Data pre-processing by XCMS used the same parameters as an earlier study [15]. Further processing involved exclusion of features eluting in the first 0.8 min or after more than 20 min, features with a standard deviation below 20% over all samples and features with an intensity below 10000 counts. The data was then normalised on the levels of the internal standards and differences in the amount of biomaterial was accounted for by normalising on the OD<sub>600</sub> of the respective samples. Isotopes and adducts identified by CAMERA [16] were also excluded. Annotations were added and the resulting feature tables were used for all analyses. Directional fold changes were and *p*-values were calculated as described for the GC-MS data. Throughout data processing, intra- and inter-group variability were assessed with the help of principal component analysis and by monitoring the levels of internal standards.

## 6.3. Results and discussion

*P. aeruginosa* PA14 with and without a functional *prmC* gene showed various differences in their metabolic profiles (Figure 6.1). Differences were less pronounced in the part of the metabolome accessible to GC-MS analysis. 24 of 116 features were significantly differentially abundant with a directional fold change  $\geq 1.5$  or  $\leq -1.5$  and a Benjamini-Hochberg corrected *p*-value  $\leq 0.05$ . However, a substantial proportion of the endo- and exometabolome measured by LC-MS was affected by the absence of *prmC*. 291 of 763 features and 893 of 1780 features were significantly differentially abundant, respectively. In both LC-MS data sets, decreased abundance upon loss of *prmC* dominated over increased abundance (Figure 6.1), which is in concordance with previous research showing that more proteins and transcripts were down-regulated than up-regulated in *prmC* deficient strains [9].

The GC-MS metabolomics data, that generally represent primary and intermediary metabolism, showed few consistent trends for metabolic pathways affected by *prmC* deficiency. Nevertheless, a consistent finding was that aromatic compounds were depleted in strains lacking *prmC*. For example, phenylalanine and tyrosine were significantly less abundant, with directional fold changes of  $-2.5$  and  $-4.4$ , respectively (Supplementary Table 1). Strikingly, shikimate and shikimate-3-phosphate, biogenic precursors of most phenylic compounds, could not be detected in any of the *tnprmC* samples, whereas they are present in *tnprmC::prmC*. This could point towards a general down-regulation of the shikimate pathway with downstream effects on secondary metabolite production. Specifically, this seems to affect alkyl quinolones and phenazines whose production relies on precursors derived from shikimate [17, 18]. Although other intermediates from the pathway were not detected, 3-deoxyarabinoheptulosonate was found to be 1.8 times less abundant in *tnprmC*. This feature could be a fragment or degradation product of 3-deoxy-arabino-heptulosonate 7-phosphate (DAHP), a precursor of shikimate. In the proteomic data of Krueger et al., the only enzyme involved in the shikimate pathway with a significant difference between *tnprmC* and *tnprmC::prmC* was 3-phosphoshikimate-1-carboxyvinyl-transferase (PA14\_23310) [9]. This enzyme acts downstream of shikimate-3-phosphate and thus its downregulation cannot directly explain shikimate and shikimate-3-phosphate depletion. On the mRNA level, the transcripts for *phzC1* and *phzC2* were significantly downregulated in *tnprmC* [9]. The gene product PhzC is a DAHP synthase that complements other DAHP synthases such as AroF and is believed to control flux of precursors into phenazine biosynthesis [18].



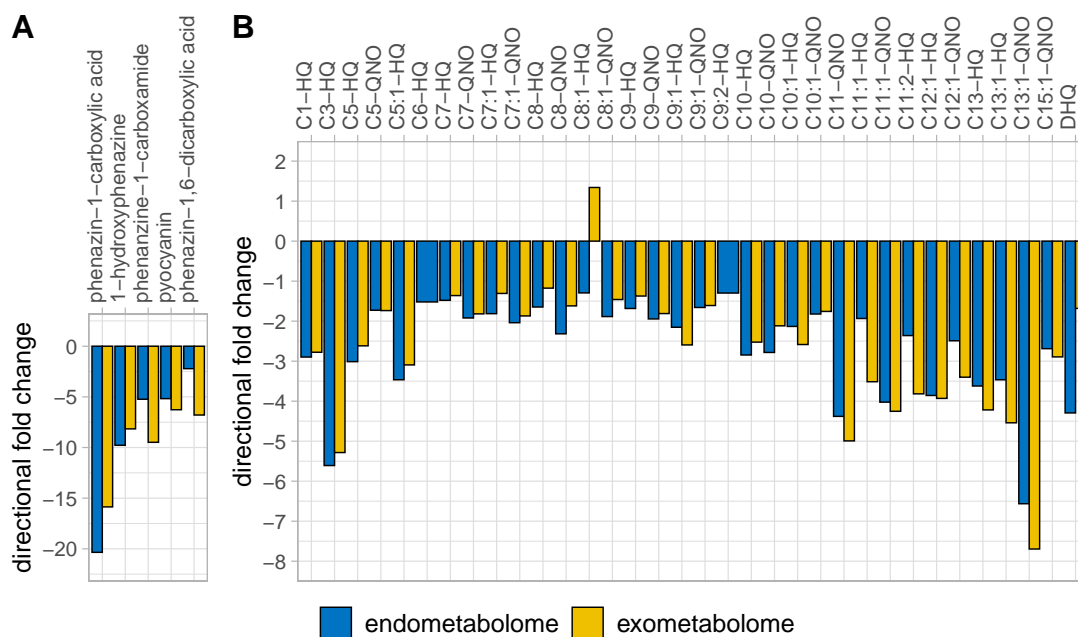
**Figure 6.1.:** Volcano plots of GC-MS metabolomics data (A), LC-MS endometabolomics data (B) and LC-MS exometabolomics data (C). All features of the respective data sets are plotted by their  $\log_2(\text{fold change})$  and  $-\log_{10}(\text{corrected } p\text{-value})$ . Features are considered significantly differentially abundant if their fold change exceeds 1.5 or  $-1.5$  and their corrected  $p$ -value is at most 0.05. While similar numbers of features displayed higher and lower abundance, respectively, in the GC-MS data, both endo- and exometabolomics LC-MS data showed a higher proportion of significantly decreased feature intensities, highlighting the dampening effect of *PrmC* deficiency on the production of secondary metabolites. For the GC-MS data, only identified features and reproducible unknowns were analysed, explaining the difference in overall feature number to the LC-MS data. Two features which were not detected in  $\Delta prmC$ , shikimate and shikimate-3-phosphate, are missing in the GC-MS volcano plot since no fold changes could be calculated.

Depletion of aromatic compounds (i. a. phenylalanine, tyrosine, anthranilate) upon absence of *prmC* was also observed in the LC-MS endo- and exometabolomics data (Supplementary Tables 2 and 3). From their lower abundance in the spent media, it might be concluded that reduced activity of the shikimate pathway is counteracted by an increased uptake of aromatic amino acids from the medium, thereby ensuring a near-normal growth phenotype.

The metabolomic effects of *prmC* deficiency were strongest on phenazine and alkyl-quinolone (AQ) levels, which was expected from previous studies [6, 9]. It was found that not only pyocyanin, but all phenazines were consistently less abundant both in the cellular metabolome and in the exometabolome, if *prmC* was non-functional (Figure 6.2A). The same was true for AQs but to a lesser degree (Figure 6.2B). Consistent with this phenomenon and the generally lower levels of aromatic metabolites, LC-MS endometabolomics also showed decreased levels of anthranilate in *tnprmC* (directional fold change  $-2.0$ , corrected  $p$ -value  $3.4 \times 10^{-8}$ ), a direct precursor of alkyl quinolones. The levels of the short or especially long alkyl chain AQs appeared to be more strongly affected by *prmC* deficiency, whereas the effect on the medium chain lengths (C6 – C9) was less pronounced. This is in accordance with recent findings on the substrate specificity of the biosynthetic enzyme complex PqsBC, which prefers medium chain activated fatty acids over long and short chain ones [19]. Krueger et al. demonstrated lower levels of PqsB and PqsC in *tnprmC* [9]. Thus, the reduced abundance of PqsBC might lead to higher relative consumption of the preferred substrates, while the less favoured long and short chain substrates are neglected.

Rhamnolipids are other important components of pseudomonal small molecule mediated virulence. Members of this class of metabolites were not consistently affected by *prmC* deficiency. In the extracellular metabolome, none of the features that were annotated as a rhamnolipid met the significance criteria (directional fold change  $\geq 1.5$  or  $\leq -1.5$  and corrected  $p$ -value  $\leq 0.05$ ). In the endometabolome, only one of the rhamnolipids was significantly differentially abundant. This feature was more abundant in *tnprmC*. These findings contradict those of Pustelny *et al.*, who found reduced rhamnolipid levels in *tnprmC* [6]. The discrepancy might be explained by the later growth phases at the time of measurement: Pustelny et al. quantified rhamnolipids after 48h of growth. The effect of *prmC* on rhamnolipid production might be time and/or growth phase dependent and might not manifest before stationary phase.

Interestingly, a significantly higher abundance of the two pyochelin isomers was detected in *tnprmC* (fold changes  $+2.8$  and  $+1.9$  in the endometabolome and  $+2.6$  and



**Figure 6.2.:** Directional fold changes of phenazines (**A**) and AQs (**B**) in the LC-MS endo- and exometabolomics data. All detected phenazines were much less abundant in *tnprnC* with no clear difference between cellular and extracellular levels. AQs were consistently less abundant in *tnprnC*. This trend was more pronounced for AQs with very long or very short alkyl chains. Panel A is sorted by magnitude of directional fold change, panel B by the length of the alkyl chain of the respective AQ. Panel B uses the nomenclature of Depke et al. [11]. While technically not an AQ, 2,4-dihydroxyquinoline (DHQ) has been added as a by-product of AQ biosynthesis. Only significantly differentially abundant features (corrected  $p$ -value  $\leq 0.05$ ) are shown. C6-HQ and C9:2-HQ were not significantly differentially abundant in the exometabolome.

+1.8 in the exometabolome; all corrected  $p$ -values  $< 10^{-4}$ ). Pyochelins are low affinity siderophores of *P. aeruginosa*, whose production is coordinated with pyoverdine, a high affinity siderophore [20]. Unfortunately, pyoverdine was not detected by our analytical methods, which complicates the interpretation of elevated pyochelin levels. Earlier research has associated elevated pyochelin levels with a virulent phenotype [15, 21, 22], and proteome and transcriptome data show the *pchD* transcript as being upregulated in *tnprmC* [9].

Among the non-annotated features, it is particularly noteworthy that M187T7, a metabolite with an  $m/z$  of 187.1230 and the putative sum formula  $C_{12}H_{15}N_2$ , displays a significantly differential abundance with approximately 5-fold lower levels in *tnprmC* and a corrected  $p$ -value of 0.0001 in the cellular metabolomics data (fold change  $-2.8$  and corrected  $p$ -value 0.016 in the exometabolome). The ‘unknown’ metabolite has been identified as a putative biomarker for virulent phenotypes in clinical *P. aeruginosa* strains in a previous study [15]. Though its identity and function are unassigned, this finding emphasises the potential importance of the metabolite for pseudomonal virulence.

There is a substantial number of additional non-annotated features that show highly significant differences in abundances between *tnprmC* and *tnprmC::prmC*, suggesting an influence of *prmC* on other parts of pseudomonal metabolism. While the annotated significantly affected metabolites in the endometabolome measured by LC-MS are almost exclusively directly or indirectly associated with virulence as described above, the respective list for the exometabolome also comprises glutathion disulphide with a directional fold change of  $+7.5$  and a corrected  $p$ -value of  $1.2 \times 10^{-8}$ . This indicates further consequences of *prmC* deficiency on the redox regulation of *P. aeruginosa*, which could potentially be linked to the distorted phenazine production. In the GC-MS metabolomics data, some features stand out by their low  $p$ -value and/or prominent fold change, but their connections to *prmC* cannot be easily established. For instance, succinate is depleted in *prmC* deficient bacteria (fold change  $-2.4$ , corrected  $p$ -value  $2.2 \times 10^{-8}$ ), whereas acetylserine is significantly more abundant in *tnprmC* (fold change  $+9.4$ , corrected  $p$ -value 0.0028). These findings show that the complexity of the metabolic consequences of *prmC* deficiency is considerable and far from being completely understood.

While this study consists of a comprehensive metabolomic profiling experiment, it also comes with several limitations. Like many untargeted metabolomics investigations, the data presented here suffers from incomplete metabolite annotation that leaves sev-

eral interesting pathways underexplored. Furthermore, metabolic fluxes have not been determined which hinders quantitative correlation with transcriptome data. Along with the novel findings outlined above, most results obtained by means of untargeted metabolomics are consistent with previous transcriptomic and proteomic studies [6, 9] and support and complement their findings.

## 6.4. Conclusion

In conclusion, we present the first metabolomics study of the role of *prmC* in virulence and metabolism of *P. aeruginosa*. The results support findings from earlier and complementary -omics studies [6, 9]. In addition, they shed light on the importance of the shikimate pathway on important AQ and phenazine virulence-mediating metabolites at high molecular resolution and highlight the potential of the putative novel virulence marker M187T7.

## Author contributions

Conceptualization, T.D., S.H. and M.B.; methodology, T.D.; validation, T.D.; formal analysis, T.D.; investigation, T.D.; resources, S.H. and M.B.; data curation, T.D.; writing—original draft preparation, T.D.; writing—review and editing, S.H. and M.B.; visualization, T.D.; supervision, S.H. and M.B.; project administration, T.D., S.H. and M.B.; funding acquisition, T.D., S.H. and M.B. All authors have read and agreed to the published version of the manuscript.

## Funding

T.D. received financial and non-material support through a PhD scholarship of the *Stiftung des deutschen Volkes*. The research was funded by the President's Initiative and Networking Funds of the Helmholtz Association of German Research Centres [VH-GS-202] and by EMBRIC, the EU-funded European Marine Biological Research Infrastructure Cluster [654008].

## Acknowledgements

The authors express gratitude for technical advice from Raimo Franke, Heike Overwin, Ulrike Beutling and Michael Hensler. Tabea Linde is acknowledged for excellent work on a related project during her bachelor's thesis project that also helped this study.



## Conflicts of interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Abbreviations

WHO	World Health Organisation
RF	release factor
LC-MS	liquid chromatography-mass spectrometry
GC-MS	gas chromatography-mass spectrometry
MSTFA	<i>N</i> -methyl- <i>N</i> -(trimethylsilyl)trifluoroacetamide
DAHP	3-deoxy-arabino-heptulosonate 7-phosphate
AQ	alkylquinolone
DHQ	2,4-dihydroxyquinoline
<i>C<sub>n</sub>:m</i>	hydrocarbon of chain length <i>n</i> with <i>m</i> double bonds
HQ	hydroxyquinoline
QNO	quinolone- <i>N</i> -oxide

## References

- [1] J. B. Lyczak, C. L. Cannon, and G. B. Pier. “Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist”. In: *Microbes and infection* 2.9 (2000), pp. 1051–1060.
- [2] E. Tacconelli, E. Carrara, A. Savoldi, S. Harbarth, M. Mendelson, D. L. Monnet, C. Pulcini, G. Kahlmeter, J. Kluytmans, Y. Carmeli, M. Ouellette, K. Outterson, J. Patel, M. Cavaleri, E. M. Cox, C. R. Houchens, M. L. Grayson, P. Hansen, N. Singh, U. Theuretzbacher, N. Magrini, A. O. Aboderin, S. S. Al-Abri, N. A. Jalil, N. Benzoni, S. Bhattacharya, A. J. Brink, F. R. Burkert, O. Cars, G. Cornaglia, O. J. Dyar, A. W. Friedrich, A. C. Gales, S. Gandra, C. G. Giske, D. A. Goff, H. Goossens, T. Gottlieb, M. G. Blanco, W. Hryniewicz, D. Kattula, T. Jinks, S. S. Kanj, L. Kerr, M.-P. Kieny, Y. S. Kim, R. S. Kozlov, J. Labarca, R. Laxminarayan, K. Leder, L. Leibovici, G. Levy-Hara, J. Littman, S. Malhotra-Kumar, V. Manchanda, L. Moja, B. Ndoye, A. Pan, D. L. Paterson, M. Paul, H. Qiu, P. Ramon-Pardo, J. Rodríguez-Baño, M. Sanguinetti, S. Sengupta, M. Sharland, M. Si-Mehand, L. L. Silver, W. Song, M. Steinbakk, J. Thomsen, G. E. Thwaites, J. W. van der Meer, N. V. Kinh, S. Vega, M. V. Villegas, A. Wechsler-Fördös, H. F. L. Wertheim, E. Wesangula, N. Woodford, F. O. Yilmaz, and A. Zorzet. “Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis”. In: *The Lancet Infectious Diseases* 18.3 (Mar. 2018), pp. 318–327. DOI: 10.1016/s1473-3099(17)30753-3.

- [3] M. D. Obritsch, D. N. Fish, R. MacLaren, and R. Jung. “Nosocomial Infections Due to Multidrug-Resistant *Pseudomonas aeruginosa*: Epidemiology and Treatment Options”. In: *Pharmacotherapy* 25.10 (Oct. 2005), pp. 1353–1364. DOI: 10.1592/phco.2005.25.10.1353.
- [4] S. de Bentzmann and P. Plésiat. “The *Pseudomonas aeruginosa* opportunistic pathogen and human infections”. In: *Environmental Microbiology* 13.7 (Mar. 2011), pp. 1655–1665. DOI: 10.1111/j.1462-2920.2011.02469.x.
- [5] P. N. Jimenez, G. Koch, J. A. Thompson, K. B. Xavier, R. H. Cool, and W. J. Quax. “The Multiple Signaling Systems Regulating Virulence in *Pseudomonas aeruginosa*”. In: *Microbiology and Molecular Biology Reviews* 76.1 (Mar. 2012), pp. 46–65. DOI: 10.1128/mubr.05007-11.
- [6] C. Pustelny, S. Brouwer, M. Müsken, A. Bielecka, A. Dötsch, M. Nimtz, and S. Häussler. “The peptide chain release factor methyltransferase PrmC is essential for pathogenicity and environmental adaptation of *Pseudomonas aeruginosa* PA14”. In: *Environmental Microbiology* 15.2 (Dec. 2012), pp. 597–609. DOI: 10.1111/1462-2920.12040.
- [7] V. Heurgué-Hamard, S. Champ, Å. Engström, M. Ehrenberg, and R. H. Buckingham. “The hemK gene in *Escherichia coli* encodes the N5-glutamine methyltransferase that modifies peptide release factors”. In: *The EMBO Journal* 21.4 (2002), pp. 769–778. DOI: 10.1093/emboj/21.4.769.
- [8] K. Nakahigashi, N. Kubo, S.-i. Narita, T. Shimaoka, S. Goto, T. Oshima, H. Mori, M. Maeda, C. Wada, and H. Inokuchi. “HemK, a class of protein methyl transferase with similarity to DNA methyl transferases, methylates polypeptide chain release factors, and *hemK* knockout induces defects in translational termination”. In: *Proceedings of the National Academy of Sciences* 99.3 (Jan. 2002), pp. 1473–1478. DOI: 10.1073/pnas.032488499.
- [9] J. Krueger, S. Pohl, M. Preusse, A. Kordes, N. Rugen, M. Schniederjans, A. Pich, and S. Häussler. “Unravelling post-transcriptional PrmC-dependent regulatory mechanisms in *Pseudomonas aeruginosa*”. In: *Environmental Microbiology* 18.10 (July 2016), pp. 3583–3592. DOI: 10.1111/1462-2920.13435.
- [10] N. T. Liberati, J. M. Urbach, S. Miyata, D. G. Lee, E. Drenkard, G. Wu, J. Villanueva, T. Wei, and F. M. Ausubel. “An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants”. In: *Proceedings of the National Academy of Sciences* 103.8 (Feb. 2006), pp. 2833–2838. DOI: 10.1073/pnas.0511100103.
- [11] T. Depke, R. Franke, and M. Brönstrup. “Clustering of MS<sup>2</sup> spectra using unsupervised methods to aid the identification of secondary metabolites from *Pseudomonas aeruginosa*”. In: *Journal of Chromatography B* 1071 (Dec. 2017), pp. 19–28. DOI: 10.1016/j.jchromb.2017.06.002.
- [12] V. Berndt, M. Beckstette, M. Volk, P. Dersch, and M. Brönstrup. “Metabolome and transcriptome-wide effects of the carbon storage regulator A in enteropathogenic *Escherichia coli*”. In: *Scientific Reports* 9.1 (Jan. 2019). DOI: 10.1038/s41598-018-36932-w.
- [13] K. Hiller, J. Hangebrauk, C. Jäger, J. Spura, K. Schreiber, and D. Schomburg. “MetaboliteDetector: Comprehensive Analysis Tool for Targeted and Nontargeted GC/MS Based Metabolome Analysis”. In: *Analytical Chemistry* 81.9 (May 2009), pp. 3429–3439. DOI: 10.1021/ac802689c.
- [14] R. Tautenhahn, G. J. Patti, D. Rinehart, and G. Siuzdak. “XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data”. In: *Analytical Chemistry* 84.11 (June 2012), pp. 5035–5039. DOI: 10.1021/ac300698c.

- [15] T. Depke, J. G. Thöming, A. Kordes, S. Häussler, and M. Brönstrup. “Untargeted LC-MS Metabolomics Differentiates Between Virulent and Avirulent Clinical Strains of *Pseudomonas aeruginosa*”. In: *Biomolecules* 10.7 (July 2020), p. 1041. DOI: 10.3390/biom10071041.
- [16] C. Kuhl, R. Tautenhahn, C. Böttcher, T. R. Larson, and S. Neumann. “CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets”. In: *Analytical Chemistry* 84.1 (Dec. 2011), pp. 283–289. DOI: 10.1021/ac202450g.
- [17] F. Bredenbruch, M. Nimtz, V. Wray, M. Morr, R. Muller, and S. Haussler. “Biosynthetic Pathway of *Pseudomonas aeruginosa* 4-Hydroxy-2-Alkylquinolines”. In: *Journal of Bacteriology* 187.11 (May 2005), pp. 3630–3635. DOI: 10.1128/jb.187.11.3630-3635.2005.
- [18] M. Mentel, E. G. Ahuja, D. V. Mavrodi, R. Breinbauer, L. S. Thomashow, and W. Blankenfeldt. “Of Two Make One: The Biosynthesis of Phenazines”. In: *ChemBioChem* 10.14 (Sept. 2009), pp. 2295–2304. DOI: 10.1002/cbic.200900323.
- [19] F. Witzgall, T. Depke, M. Hoffmann, M. Empting, M. Brönstrup, R. Müller, and W. Blankenfeldt. “The Alkylquinolone Repertoire of *Pseudomonas aeruginosa* is Linked to Structural Flexibility of the FabH-like 2-Heptyl-3-hydroxy-4(1*H*)-quinolone (PQS) Biosynthesis Enzyme PqsBC”. In: *ChemBioChem* 19.14 (May 2018), pp. 1531–1544. DOI: 10.1002/cbic.201800153.
- [20] P. Cornelis and J. Dingemans. “*Pseudomonas aeruginosa* adapts its iron uptake strategies in function of the type of infections”. In: *Frontiers in Cellular and Infection Microbiology* 3 (2013). DOI: 10.3389/fcimb.2013.00075.
- [21] N. J. Hare, C. Z. Soe, B. Rose, C. Harbour, R. Codd, J. Manos, and S. J. Cordwell. “Proteomics of *Pseudomonas aeruginosa* Australian Epidemic Strain 1 (AES-1) Cultured under Conditions Mimicking the Cystic Fibrosis Lung Reveals Increased Iron Acquisition via the Siderophore Pyochelin”. In: *Journal of Proteome Research* 11.2 (Dec. 2011), pp. 776–795. DOI: 10.1021/pr200659h.
- [22] C. D. Cox. “Effect of pyochelin on the virulence of *Pseudomonas aeruginosa*.” In: *Infection and Immunity* 36.1 (1982), pp. 17–23. DOI: 10.1128/iai.36.1.17-23.1982.

## Supporting Information

6.S1	Feature table of the GC-MS data. . . . .	228
6.S2	Feature table of all annotated features of the LC-MS endometabolomics data. . . . .	233
6.S3	Feature table of all annotated features of the LC-MS exometabolomics data. . . . .	236

**Table 6.S1.:** Feature table of the GC-MS data.

metabolite	library hit	retention index	retention time (min)	directional fold change	corrected p-value
unknown	Unknown#1031.8-pae-bth_008	1052.4	5.3	-1.04	0.952
2-hydroxypyridine	Pyridine,_2-hydroxy-(1TMS)_1032.4	1059.4	5.4	-1.41	0.462
lactate	Lactic_acid_(2TMS)_1062.9	1078.4	5.7	6.67	$4.65 \times 10^{-3}$
glycolate	Glycolic_acid_(2TMS)_1083	1084.9	5.8	1.91	0.451
alanine	Alanine_(2TMS)_1106.8	1111.6	6.3	1.59	0.451
oxalate	Oxalic_acid_(2TMS)_1139.3	1140.1	6.7	-1.09	0.870
glycine	Glycine_(2TMS)_1126.6	1140.1	6.7	1.63	0.451
unknown	Unknown#1169.8-pae-bth_013	1168.0	7.2	-1.25	0.483
leucine	Leucine_(1TMS)_1162.15	1176.4	7.3	-1.45	0.563
monomethylphosphate	Phosphoric_acid_monomethyl_ester_(2TMS)_1177.5	1193.1	7.6	-1.60	0.554
valine	Valine_(2TMS)_1214.5	1223.0	8.1	1.41	0.533
diethylenglycol	Diethylenglycol_(2TMS)_1248	1253.3	8.6	1.01	0.959
benzoate	Benzoic_acid_(1TMS)_1249.7	1259.8	8.7	-3.02	0.533
unknown	Unknown#1248.7-cgl-sst_008	1263.7	8.7	2.02	0.483
ethanolamine	Ethanolamine_(3TMS)_1269.07	1267.2	8.8	1.11	0.714
glycerol	Glycerol_(3TMS)_1282.1	1275.8	8.9	1.85	0.505
phosphate	Phosphoric_acid_(3TMS)_1278.2	1282.3	9.0	1.08	0.828
nicotinate	Nicotinic_acid_(1TMS)_1300.27	1314.5	9.5	-5.01	$2.80 \times 10^{-7}$
succinate	Succinic_acid_(2TMS)_1325.4	1324.8	9.7	-2.38	$2.18 \times 10^{-8}$
glycerate	Glyceric_acid_(3TMS)_1343.4	1336.6	9.9	1.63	0.195
lumichrome	Lumichrome_(2MEOX)_1342.36	1341.0	10.0	-1.87	0.0386
uracil	Uracil_(2TMS)_1346.9	1344.8	10.0	1.21	0.337
fumarate	Fumaric_acid_(2TMS)_1362.4	1358.1	10.2	-1.27	0.390
nonanoate	Nonanoic_acid_(1TMS)_1371.8	1368.2	10.4	1.32	0.462
unknown	Unknown#1385.3-pae-bth_024	1385.9	10.7	1.09	0.834
thymine	Thymine_(2TMS)_1408.4	1403.3	11.0	-1.16	0.725

**Table 6.S1.: Continued:** Feature table of the GC-MS data.

metabolite	library hit	retention index	retention time (min)	directional fold change	corrected p-value
methionine	Methionine_(1TMS)_1413.7	1416.5	11.2	1.11	0.819
unknown	Unknown#1422-pae-bth_028	1422.1	11.3	1.72	0.403
aspartate	Aspartic_acid_(2TMS)_1431.9	1427.0	11.4	-1.07	0.663
unknown	NA102	1451.4	11.7	1.18	0.652
decanoate	Decanoic_acid_(1TMS)_1463.1	1459.7	11.9	1.85	0.451
2-methyl-malate	Malic_acid,_2-methyl-(3TMS)_1473.53	1468.9	12.0	2.82	$2.56 \times 10^{-4}$
<i>N</i> -acetylserine	Serine,_N-acetyl-(2TMS)_1514.29	1506.5	12.6	9.36	$2.83 \times 10^{-3}$
pyroglutamate	Pyroglutamic_acid_(1TMS)_1499.6	1509.9	12.7	1.57	0.129
aspartate	Aspartic_acid_(3TMS)_1520.9	1518.8	12.8	1.24	0.578
pyroglutamate	Pyroglutamic_acid_(2TMS)_1518.1	1522.0	12.8	1.04	0.868
cytosine	Cytosine_(2TMS)_1523.1	1523.2	12.8	-1.09	0.870
4-aminobutanoate	Butanoic_acid,_4-amino-(3TMS)_1530.28	1528.5	12.9	-1.56	0.301
glutamate	Glutamic_acid_(2TMS)_1532.7	1535.2	13.0	1.11	0.462
phenylalanine	Phenylalanine_(1TMS)_1560	1549.9	13.2	-2.46	$1.50 \times 10^{-7}$
3-hydroxybenzoate	Benzoic_acid,_3-hydroxy-(2TMS)_1566.5	1568.0	13.5	-13.40	$1.68 \times 10^{-4}$
6-hydroxynicotinate	Nicotinic_acid,_6-hydroxy-(2TMS)_1583.98	1575.6	13.6	-3.98	$4.88 \times 10^{-4}$
proline	Proline_[+CO2]_(2TMS)_1587.3	1583.6	13.7	-1.44	0.462
unknown	Unknown#1586.9-pae-bth_039	1590.2	13.8	-1.21	0.653
phosphoenolpyruvate	Phosphoenolpyruvic_acid_(3TMS)_1599.2	1600.0	13.9	-1.69	0.129
glutamate	Glutamic_acid_(3TMS)_1628.5	1623.0	14.2	-1.45	0.602
4-hydroxybenzoate	Benzoic_acid,_4-hydroxy-(2TMS)_1631.79	1631.9	14.4	1.89	0.451
unknown	Unknown#1638.3-cgl-sst_041	1633.4	14.4	-1.84	0.483
1,2-diaminopropane	Propane,_1,2-diamino-(4TMS)_1653.83	1644.3	14.5	-1.71	0.462
unknown	NA149	1649.8	14.6	-1.40	0.129
3-hydroxydecanoate	Decanoic_acid,_3-hydroxy-(2TMS)_1659.8	1656.2	14.7	1.49	0.136
dodecanoate	Dodecanoic_acid_(1TMS)_1659.25	1658.1	14.7	1.62	0.0386

**Table 6.S1.: Continued:** Feature table of the GC-MS data.

metabolite	library hit	retention index	retention time (min)	directional fold change	corrected p-value
xylose/ribose/arabinose/...	Xylose_(1MEOX)_(4TMS)_MP_1653.4 (ribose, arabinose, ...)	1660.8	14.8	-1.32	0.737
pyrophosphate	Pyrophosphate_(4TMS)_1669.4	1662.3	14.8	-1.84	0.462
xylulose/ribulose/...	Xylulose_(1MEOX)_(4TMS)_MP_1685.1 (ribulose, ...)	1678.2	15.0	-1.33	0.805
2-dodecenoate	2-Dodecenoic-acid_1TMS_1715.5	1713.9	15.5	1.12	0.779
glycerol-2-phosphate	Glycerol-2-phosphate_(4TMS)_1732.4	1727.9	15.7	1.18	0.462
putrescine	Putrescine_(4TMS)_1739.36	1737.4	15.8	1.02	0.955
dihydroxyacetone phosphate	Dihydroxyacetone_phosphate_(1MEOX)_(3TMS)_BP_1764.8	1753.1	16.0	1.62	0.554
glycerol-3-phosphate	Glycerol-3-phosphate_(4TMS)_1767.8	1766.5	16.2	1.17	0.625
glucose	Glucopyranose_-[H2O]_(4TMS)_1763.61	1769.4	16.3	1.34	0.338
glycerate-3-phosphate	Glyceric_acid-2-phosphate_(4TMS)_1793.2	1779.3	16.4	-1.03	0.870
ethanolaminophosphate	Ethanolaminophosphate_(4TMS)_1787	1783.2	16.5	-1.16	0.318
N-acetyl-glutamate	Glutamic_acid,_N-acetyl-_(2TMS)_1788.4	1786.3	16.5	1.01	0.981
azelate	Azelaic_Acid_(2TMS)_1798.42	1798.8	16.7	-1.62	0.722
glycerate-1-phosphate	Glyceric_acid-3-phosphate_(4TMS)_1811	1808.1	16.8	-1.29	0.318
shikimate	Shikimic_acid_(4TMS)_1815.3	1810.6	16.8	not detected in <i>tnprnC</i>	
citrate	Citric_acid_(4TMS)_1825.5	1817.5	16.9	1.35	0.0844
3,4-dihydroxybenzoate	Benzoic_acid,_3,4-dihydroxy-_(3TMS)_1821.5	1818.5	17.0	-8.00	0.0111
fructose	Fructose,_D-_(5TMS)_1845.5	1857.8	17.5	-1.06	0.790
tyrosine	Tyrosine_(2TMS)_1882.5	1881.6	17.8	-4.43	$7.66 \times 10^{-3}$
mannitol/sorbitol/galactitol	Mannitol_(6TMS)_1925.5 (sorbitol, galactitol, ...)	1925.5	18.4	3.23	$2.02 \times 10^{-11}$
3-deoxyarabinoheptulosonate	Arabinoheptulosonic_acid_enol,_3-deoxy-_(5TMS)_MP_1933.32	1935.3	18.5	-1.86	$5.78 \times 10^{-4}$
glucose	Glucopyranose,_D-_(5TMS)_1981.6	1975.7	19.0	1.72	0.048

**Table 6.S1.: Continued:** Feature table of the GC-MS data.

metabolite	library hit	retention index	retention time (min)	directional fold change	corrected p-value
unknown	Unknown#2018.7-pin-mhe_028	2016.0	19.5	1.36	0.143
unknown	Unknown#2013.24-ypy-mse_018	2017.1	19.5	1.27	0.451
unknown	NA321	2022.7	19.5	3.34	$1.92 \times 10^{-3}$
hexadecanoate	Hexadecanoic_acid_(1TMS)_2047.1	2046.5	19.8	1.05	0.797
unknown	Unknown#2051.8-cgl-sst_001	2049.6	19.8	2.75	$1.91 \times 10^{-3}$
<i>N</i> -acetyl-glucosamine	Glucosamine,_N-acetyl-_(1MEOX)_(4TMS)_2077	2071.3	20.1	1.19	0.663
xylulose-5-phosphate/ribose-5-phosphate/...	Xylulose-5-phosphate_(1MEOX)_(5TMS)_MP_2099.9 (ribose-5-phosphat, ...)	2100.1	20.4	1.30	0.625
unknown	NA374	2126.0	20.7	1.02	0.981
glycerophosphoglycerol	Glycerophosphoglycerol_(5TMS)_2190.4	2187.8	21.5	1.09	0.554
octadecenoate	Octadecenoic_acid,_9-(E)-_(1TMS)_2220.6	2221.5	21.8	1.01	0.981
shikimate-3-phosphate	Shikimic_acid-3-phosphate_(5TMS)_2216.26	2224.1	21.9	not detected in <i>tnpmC</i>	
octadecenoate	Octadecanoic_acid_(1TMS)_2250.63	2243.2	22.0	1.04	0.752
unknown	NA_2276.1	2271.2	22.3	-12.10	$5.69 \times 10^{-8}$
fructose-1-phosphate	Fructose-1-phosphate_(1MEOX)_(6TMS)_MP_2292.4	2289.4	22.5	-1.13	0.834
fructose-6-phosphate	Fructose-6-phosphate_(1MEOX)_(6TMS)_MP_2308.33	2303.4	22.6	2.73	0.451
hexose-6-phosphate	Mannose-6-phosphate_(1MEOX)_(6TMS)_MP_2324.53 (hexose-6-phosphate)	2317.5	22.8	-1.27	0.073
hexose-6-phosphate	Glucose-6-phosphate_(1MEOX)_(6TMS)_BP_2340.2 (hexose-6-phosphate)	2337.3	23.0	-1.36	0.0508
unknown	NA454	2364.7	23.2	2.56	0.451
gluconate-6-phosphate	Gluconic_acid-6-phosphate_(7TMS)_2427.8	2423.2	23.8	-1.03	0.952
unknown	Unknown#2454.7-pin-mhe_060	2454.2	24.1	2.13	$7.66 \times 10^{-3}$
myo-inositol-2-phosphate	Inositol-2-phosphate,_myo-_(7TMS)_2465.95	2470.3	24.3	-4.12	$3.49 \times 10^{-4}$

**Table 6.S1.: Continued:** Feature table of the GC-MS data.

metabolite	library hit	retention index	retention time (min)	directional fold change	corrected p-value
myo-inositol-1-phosphate	Inositol-1-phosphate,_myo-_(7TMS)_2508.56	2513.5	24.7	2.42	0.462
unknown	Unknown#2551.4-cgl-sst_119	2550.1	25.1	1.44	0.0844
sedoheptulose-7-phosphate	Sedoheptulose-7-phosphate_(xMeOX)_(yTMS)_2555.17	2562.2	25.2	1.47	0.162
unknown	NA_2600.1	2594.7	25.5	1.39	0.382
unknown	Unknown#2590.91-ypy-mse_037	2594.8	25.5	-1.61	0.602
unknown	D260482_2610.31	2600.9	25.6	1.20	0.625
unknown	Unknown#2635-cgl-sst_121	2631.7	25.9	1.61	$1.37 \times 10^{-4}$
unknown	NA540	2661.6	26.2	1.76	0.0981
1-oleoyl-rac-glycerol	1-Oleoyl-rac-glycerol_(2TMS)_2757.9	2759.9	27.1	-1.14	0.834
disaccharide	alpha-D-Galactopyranosyl-(1,4)-D-galactopyranoside_(1MEOX)_(8TMS)_BP_2790.07 (disaccharide)	2802.0	27.5	2.78	$1.92 \times 10^{-3}$
disaccharide	Melibiose_(1MEOX)_(8TMS)_MP_2865.91 (disaccharide)	2859.5	28.0	1.33	0.474
UMP	Uridine_5'-monophosphate_(5TMS)_2872.1	2871.2	28.1	1.01	0.983
disaccharide	Melibiose_8TMS_2904.74 (disaccharide)	2898.3	28.3	4.35	0.170
AMP	Adenosine-5-monophosphate_(4TMS)_(Derivate_not_found)_3047.9	3052.9	29.6	1.14	0.760
AMP	Adenosine-5-monophosphate_(5TMS)_3061.8	3064.1	29.7	1.15	0.554
GMP	Guanosine-5-monophosphate_(6TMS)_3121.8	3116.8	30.1	1.36	0.451
unknown	NA820	3522.3	33.2	1.23	0.834
unknown	Unknown#3522.7-cgl-sst_133	3529.3	33.2	2.91	0.451
unknown	NA824	3547.5	33.4	2.08	0.533
trisaccharide	Cellotriose_(1MEOX)_(11TMS)_3649.07 (trisaccharide)	3641.0	34.1	2.00	0.462



**Table 6.S2.:** Feature table of all annotated features of the LC-MS endometabolomics data.

identifier	median m/z	median retention time (min)	annotation	directional fold change	corrected p-value
M129T1	129.1385	1.0	spermidine (fragment)	1.07	0.760
M146T1	146.1651	1.0	spermidine	1.05	0.796
M179T1	179.0550	1.2	gluconolactone	-1.57	$9.75 \times 10^{-5}$
M116T1	116.0705	1.2	proline	-1.21	0.119
M130T1_1	130.0499	1.2	5-oxoproline	-1.16	0.285
M148T1	148.0606	1.2	glutamate	-1.10	0.446
M118T1	118.0863	1.3	betaine	1.50	$7.44 \times 10^{-7}$
M325T1_1	325.0430	1.3	UMP	1.25	0.113
M277T1_2	277.1031	1.3	Glu Glu	-1.33	0.134
M324T1	324.0592	1.3	CMP	1.20	0.147
M428T1	428.0362	1.3	ADP	1.09	0.452
M124T1	124.0394	1.4	nicotinate	-2.89	0.0190
M535T1	535.1883	1.4	Glu Glu Glu Glu	1.89	0.0523
M162T1	162.0760	1.4	N-methylglutamate	1.25	0.103
M307T1	307.0833	1.4	glutathione disulphide (2+)	-2.37	0.104
M235T1_2	235.1189	1.4	His Pro (II)	1.05	0.491
M364T1	364.0651	1.4	GMP	1.07	0.504
M348T1	348.0706	1.4	AMP	1.08	0.612
M373T1_1	372.5449	1.4	NADP (2+)	1.04	0.701
M136T1	136.0617	1.4	adenine (AMP fragment)	1.06	0.697
M664T1_1	664.1165	1.4	NAD	1.06	0.794
M333T1_1	332.5619	1.4	NAD (2+)	1.02	0.894
M542T1	542.0680	1.4	NAD (fragment)	1.02	0.904
M123T1_2	123.0553	1.4	nicotinamide (NAD fragment)	1.00	0.983
M332T1	332.0751	1.4	dAMP	1.01	0.953
M182T2	182.0809	2.0	tyrosine	-3.40	$2.14 \times 10^{-4}$
M132T2	132.1017	2.0	Leucine / Isoleucine / Norleucine	-1.12	0.447
M235T2	235.1187	2.1	His Pro (I)	1.11	0.0861
M193T2	193.0681	2.5	S-(5'-adenosyl)- homocysteine (2+)	-1.03	0.806
M166T3	166.0862	3.5	phenylalanine	-2.91	$5.11 \times 10^{-3}$
M120T3	120.0807	3.5	phenylethanolamine (Phe fragment)	-2.93	$5.39 \times 10^{-3}$
M254T4	254.1612	4.3	Pro Arg	1.11	0.469
M194T5	194.0788	5.2	C4-HSL	1.34	$1.80 \times 10^{-5}$
M220T5	220.1180	5.2	panthotenate	1.42	$5.19 \times 10^{-3}$
M188T6	188.0706	5.9	tryptophan (fragment)	-1.19	0.274
M205T6	205.0972	5.9	tryptophan	-1.17	0.365
M298T6	298.0969	6.3	5'-methylthioadenosine	-1.00	0.973
M211T6	211.0869	6.4	pyocyanin	-5.17	$1.59 \times 10^{-10}$
M187T7	187.1230	6.5	putative virulence marker	-5.44	$1.07 \times 10^{-4}$

**Table 6.S2.: Continued:** Feature table of all annotated features of the LC-MS endometabolomics data.

identifier	median m/z	median retention time (min)	annotation	directional fold change	corrected p-value
M348T7	348.0702	6.9	AMP (FMN fragment)	-1.09	0.0956
M786T7	786.1640	6.9	FAD	-1.10	0.151
M120T7	120.0444	7.0	anthranilate (fragment)	-1.95	$3.45 \times 10^{-8}$
M160T7	160.0760	7.1	C1-HQ	-2.90	$1.07 \times 10^{-10}$
M457T7	457.1117	7.1	FMN	-1.29	$3.32 \times 10^{-5}$
M328T7	328.2229	7.3	Pro Leu Val	-1.11	0.253
M342T8	342.2386	7.8	Pro Ile Leu or isomer	1.03	0.718
M162T8	162.0552	8.3	DHQ	-4.29	$1.44 \times 10^{-7}$
M245T9	245.1283	8.6	Phe Pro	1.09	0.173
M188T9	188.1071	8.8	C3-HQ	-5.61	$5.34 \times 10^{-9}$
M243T9	243.0877	9.2	lumichrome	1.17	0.127
M224T11	224.0823	10.6	phenazine-1-carboxamide	-5.27	$1.15 \times 10^{-5}$
M207T11_2	207.0555	10.6	phenazine-1-carboxamide (fragment)	-5.21	$1.95 \times 10^{-5}$
M197T11	197.0708	10.9	1-hydroxyphenazine	-9.78	$3.81 \times 10^{-12}$
M214T11	214.1228	10.9	C5:1-HQ	-3.46	$4.94 \times 10^{-11}$
M216T11	216.1386	11.0	C5-HQ	-3.01	$2.84 \times 10^{-11}$
M232T11	232.1332	11.3	C5-QNO	-1.73	$8.28 \times 10^{-9}$
M207T11_1	207.0554	11.4	phenazine-1-carboxylic acid (fragment)	-21.20	$1.19 \times 10^{-8}$
M225T11_1	225.0661	11.4	phenazin-1-carboxylic acid	-19.50	$1.19 \times 10^{-8}$
M269T12	269.0557	11.6	phenazin-1,6-dicarboxylic acid	-2.22	$5.36 \times 10^{-9}$
M325T12_1	325.0681	11.6	pyochelin (II)	1.93	$4.29 \times 10^{-5}$
M230T12	230.1541	12.0	C6-HQ	-1.52	$6.50 \times 10^{-5}$
M325T12_2	325.0682	12.3	pyochelin (I)	2.76	$7.08 \times 10^{-7}$
M270T13	270.1854	12.6	C9:1-HQ (I)	-3.04	$1.63 \times 10^{-10}$
M288T13	288.1963	12.6	C9-QNO (I)	-2.61	$6.75 \times 10^{-10}$
M258T13	258.1494	12.7	C7:1-QNO	-2.04	$4.88 \times 10^{-9}$
M242T13	242.1544	13.0	C7:1-HQ	-1.81	$2.1 \times 10^{-8}$
M159T13	159.0680	13.0	HHQ (fragment)	-1.74	$6.47 \times 10^{-8}$
M244T13_1	244.1704	13.0	HHQ	-1.22	$3.30 \times 10^{-6}$
M260T13	260.1651	13.1	C7-QNO	-1.58	$9.28 \times 10^{-9}$
M519T13	519.3221	13.1	HQNO [2M+H] <sup>+</sup>	-2.26	$1.81 \times 10^{-8}$
M272T14_1	272.1646	13.5	C8:1-QNO	-1.89	$2.48 \times 10^{-8}$
M314T14	314.2119	13.6	C11:1-QNO	-4.02	$3.05 \times 10^{-9}$
M286T14_2	286.1806	13.9	C9:1-QNO (I)	-1.77	$1.80 \times 10^{-7}$
M256T14_2	256.1700	13.9	C8:1-HQ	-1.29	$5.70 \times 10^{-3}$
M274T14	274.1805	14.0	C8-QNO	-2.32	$2.54 \times 10^{-10}$
M258T14	258.1857	14.0	C8-HQ	-1.65	$1.74 \times 10^{-5}$
M320T14	320.1835	14.2	C12-HSL	-1.53	$1.65 \times 10^{-5}$
M296T14	296.2010	14.3	C11:2-HQ (II)	-1.91	$2.58 \times 10^{-7}$
M268T14	268.1702	14.3	C9:2-HQ	-1.30	$3.18 \times 10^{-3}$

**Table 6.S2.: Continued:** Feature table of all annotated features of the LC-MS endometabolomics data.

identifier	median m/z	median retention time (min)	annotation	directional fold change	corrected p-value
M316T14_2	316.2277	14.4	C11-QNO	-4.38	$1.13 \times 10^{-11}$
M286T14_1	286.1808	14.4	C9:1-QNO (II)	-1.55	$5.54 \times 10^{-7}$
M284T15	284.2015	14.7	C10:1-HQ (II)	-2.10	$4.53 \times 10^{-8}$
M300T15_1	300.1962	14.7	C10:1-QNO	-1.82	$6.82 \times 10^{-7}$
M270T15_1	270.1864	14.8	C9:1-HQ (II)	-1.26	$1.51 \times 10^{-5}$
M288T15	288.1964	14.9	C9-QNO (II)	-1.28	$5.94 \times 10^{-8}$
M342T15	342.2430	15.0	C13:1-QNO	-6.56	$1.20 \times 10^{-9}$
M452T15	452.2779	15.0	LPE(16:1)	-1.09	0.418
M260T15	260.1652	15.2	PQS	-2.46	0.0650
M286T15	286.1806	15.3	C9:1-PQS	-1.15	0.516
M314T16_1	314.2121	15.5	C11:1-PQS (I)	-1.82	$7.23 \times 10^{-11}$
M298T16_1	298.2173	15.5	C11:1-HQ (II)	-1.77	$1.49 \times 10^{-6}$
M505T16	505.2532	15.6	LPE(16:1) +Na	-1.01	0.965
M286T16	286.2171	15.7	C10-HQ	-2.85	$3.04 \times 10^{-11}$
M296T16	296.2014	15.7	C11:2-HQ (I)	-2.81	$4.26 \times 10^{-9}$
M302T16_2	302.2118	15.7	C10-QNO	-2.78	$1.13 \times 10^{-9}$
M284T16	284.2014	15.7	C10:1-HQ (I)	-2.16	$1.48 \times 10^{-8}$
M272T16_2	272.2011	15.7	C9-HQ	-1.68	$2.32 \times 10^{-6}$
M673T16	673.3780	15.9	Rha-Rha-C10-C10+Na	1.09	0.114
M651T16	651.3960	15.9	Rha-Rha-C10-C10	1.13	0.217
M314T16_2	314.2120	16.0	C11:1-PQS (I)	-2.19	$4.82 \times 10^{-9}$
M298T16_2	298.2171	16.2	C11:1-HQ (I)	-2.41	$5.82 \times 10^{-9}$
M312T16_2	312.2326	16.3	C12:1-HQ (I)	-3.77	$1.84 \times 10^{-11}$
M328T16	328.2275	16.3	C12:1-QNO	-2.49	$4.31 \times 10^{-9}$
M454T16	454.2933	16.3	LPE(16:0)	1.08	0.186
M316T17	316.2278	16.5	C11-PQS	-1.84	$6.87 \times 10^{-10}$
M298T17_1	298.2173	16.6	C11:1-HQ (III)	-1.62	$2.73 \times 10^{-5}$
M527T17	527.3197	16.7	Rha-C10-C10+Na	1.26	$2.30 \times 10^{-3}$
M699T17_2	699.3933	16.7	Rha-Rha-C10-C12:1 / Rha-Rha-C12:1-C10 +Na	1.11	0.0414
M677T17	677.4110	16.7	Rha-Rha-C10-C12:1 / Rha-Rha-C12:1-C10	1.12	0.312
M326T17_1	326.2484	17.0	C13:1-HQ (I)	-4.12	$4.02 \times 10^{-12}$
M502T17	502.2909	17.1	LPE(18:1) +Na	1.24	$3.54 \times 10^{-4}$
M480T17_1	480.3092	17.1	LPE(18:1)	1.20	$5.39 \times 10^{-3}$
M701T17_2	701.4089	17.2	Rha-Rha-C10-C12+Na	1.08	0.177
M679T17	679.4270	17.2	Rha-Rha-C10-C12 / Rha-Rha-C12-C10	1.14	0.226
M312T17	312.2325	17.4	C12:1-HQ (II)	-3.94	$3.90 \times 10^{-11}$
M553T17	553.3356	17.5	Rha-C10-C12:1+Na	1.10	0.294
M342T18	342.2430	17.7	C13:1-PQS	-2.21	$7.55 \times 10^{-8}$
M727T18	727.4245	17.9	Rha-Rha-C12-C12:1+Na	-1.10	0.336

**Table 6.S2.:** Continued: Feature table of all annotated features of the LC-MS endometabolomics data.

identifier	median m/z	median retention time (min)	annotation	directional fold change	corrected p-value
M705T18	705.4415	17.9	Rha-Rha-C12:1-C12 / Rha-Rha-C12-C12:1	-1.06	0.644
M555T18_2	555.3511	18.1	Rha-C10-C12+Na (II)	1.01	0.877
M328T18	328.2641	18.3	C13-HQ	-3.62	$1.97 \times 10^{-10}$
M326T18	326.2485	18.3	C13:1-HQ (II)	-2.81	$2.47 \times 10^{-8}$
M370T19	370.2742	18.5	C15:1-QNO	-2.69	$3.8 \times 10^{-10}$
M255T18	255.2320	18.5	palmitoleate	1.02	0.924
M707T19	707.4579	18.6	Rha-Rha-C12-C12	-1.06	0.319
M555T19	555.3571	18.9	Rha-C10-C12+Na (I)	1.58	$1.69 \times 10^{-6}$
M282T19_2	282.2792	19.1	petroselinic acid	-1.00	0.996
M257T20	257.2476	19.6	palmitate	-1.05	0.746
M327T20	327.2271	19.8	oleate	1.06	0.765

**Table 6.S3.:** Feature table of all annotated features of the LC-MS exometabolomics data.

identifier	median m/z	median retention time (min)	annotation	directional fold change	corrected p-value
M112T1_2	112.1119	1.0	spermidine (fragment)	2.07	$4.06 \times 10^{-4}$
M146T1_3	146.1651	1.0	spermidine	1.66	$1.79 \times 10^{-3}$
M129T1_2	129.1384	1.0	spermidine (fragment)	1.87	$8.84 \times 10^{-4}$
M179T1	179.0548	1.2	gluconolactone	-2.26	$7.69 \times 10^{-6}$
M162T1	162.0758	1.2	N-methylglutamate	1.03	0.762
M130T1_1	130.0497	1.2	5-oxoproline	-1.02	0.786
M118T1	118.0862	1.3	betaine	1.28	$3.74 \times 10^{-3}$
M219T1_3	219.0972	1.3	Glu Ala	1.26	$2.40 \times 10^{-3}$
M307T1	307.0831	1.4	glutathion disulphide (2+)	7.54	$1.18 \times 10^{-8}$
M136T1	136.0616	1.4	adenine (AMP fragment)	1.29	$9.15 \times 10^{-3}$
M137T1	137.0455	1.4	hypoxanthine	-1.63	$3.69 \times 10^{-5}$
M110T1	110.0599	1.4	2-aminophenol	-1.92	$1.10 \times 10^{-3}$
M235T1_2	235.1190	1.4	His Pro (II)	-1.45	$1.71 \times 10^{-4}$
M333T1_1	332.5616	1.4	NAD (2+)	-1.58	$5.30 \times 10^{-3}$
M664T1	664.1152	1.4	NAD	-1.53	0.0248
M348T1_1	348.0700	1.4	AMP	-1.37	0.039
M182T2	182.0807	2.0	tyrosine	-9.42	$8.9 \times 10^{-8}$
M235T2	235.1189	2.1	His Pro (I)	-1.62	$9.42 \times 10^{-5}$
M166T3	166.0860	3.4	phenylalanine	-2.02	$4.77 \times 10^{-6}$
M120T3	120.0805	3.4	phenylethanolamine (Phe fragment)	-2.06	$1.39 \times 10^{-5}$
M254T4	254.1612	4.3	Pro Arg	-1.54	$3.80 \times 10^{-7}$
M194T5	194.0786	5.2	C4-HSL	1.12	$7.66 \times 10^{-3}$

**Table 6.S3.: Continued:** Feature table of all annotated features of the LC-MS exometabolomics data.

identifier	median m/z	median retention time (min)	annotation	directional fold change	corrected p-value
M220T5	220.1175	5.2	panthotenate	-1.30	0.0101
M298T6_1	298.0965	6.3	5'-methylthioadenosine	-1.05	0.398
M211T6	211.0867	6.4	pyocyanin (I)	-5.91	$5.04 \times 10^{-8}$
M187T7	187.1228	6.6	putative virulence marker	-2.81	0.0155
M378T7	378.2014	6.6	Pro Tyr Val	-1.18	0.0753
M190T7	190.0495	6.7	kynurenate	-1.32	$2.77 \times 10^{-3}$
M138T7	138.0547	6.9	anthranilate	-3.00	$1.36 \times 10^{-8}$
M211T7_1	211.0866	7.0	pyocyanin (II)	-5.77	$7.20 \times 10^{-8}$
M120T7	120.0443	7.0	anthranilate (fragment)	-3.34	$1.60 \times 10^{-8}$
M160T7	160.0758	7.1	C1-HQ	-2.78	$4.31 \times 10^{-9}$
M328T7_2	328.2228	7.3	Pro Leu Val	-1.77	$5.15 \times 10^{-6}$
M211T7_2	211.0866	7.4	pyocyanin (III)	-7.13	$1.72 \times 10^{-7}$
M342T8	342.2384	7.8	Pro Ile Leu or isomer	-1.62	$1.71 \times 10^{-6}$
M211T8_2	211.1441	8.1	cyclo(Leu Pro)	-1.20	$4.51 \times 10^{-3}$
M162T8	162.0550	8.4	DHQ	-1.68	$8.39 \times 10^{-4}$
M245T9	245.1286	8.6	Phe Pro	-1.21	$4.23 \times 10^{-3}$
M188T9	188.1070	8.8	C3-HQ	-5.28	$3.73 \times 10^{-8}$
M243T9_2	243.0873	9.2	lumichrome	2.41	$6.42 \times 10^{-6}$
M224T11_2	224.0820	10.6	phenazine-1-carboxamide	-9.49	$2.73 \times 10^{-8}$
M260T11	260.1646	10.8	HQNO	-2.36	$3.73 \times 10^{-8}$
M214T11	214.1227	10.9	C5:1-HQ	-3.09	$3.35 \times 10^{-8}$
M197T11	197.0709	10.9	1-hydroxyphenazine	-8.16	$3.12 \times 10^{-5}$
M216T11	216.1384	11.0	C5-HQ	-2.62	$2.02 \times 10^{-8}$
M232T11	232.1332	11.3	C5-QNO	-1.74	$1.85 \times 10^{-7}$
M225T11_1	225.0659	11.4	phenazin-1-carboxylic acid (I)	-12.70	$6.95 \times 10^{-9}$
M207T11	207.0553	11.4	phenazine-1-carboxylic acid (fragment) (I)	-15.70	$4.38 \times 10^{-8}$
M325T12_2	325.0678	11.6	pyochelin (II)	1.81	$2.30 \times 10^{-5}$
M269T12	269.0552	11.6	phenazin-1,6-dicarboxylic acid	-6.79	$5.32 \times 10^{-7}$
M225T12	225.0658	11.9	phenazin-1-carboxylic acid (II)	-16.70	$1.30 \times 10^{-8}$
M207T12_2	207.0552	11.9	phenazine-1-carboxylic acid (fragment) (II)	-18.40	$2.55 \times 10^{-8}$
M286T12	286.1800	12.0	C9:1-QNO (I)	-1.67	$4.37 \times 10^{-9}$
M230T12	230.1537	12.1	C6-HQ	1.11	0.0505
M325T12_1	325.0676	12.3	pyochelin (I)	2.63	$1.21 \times 10^{-8}$
M260T12	260.1645	12.4	C7-QNO (II)	1.50	0.086
M270T13	270.1852	12.7	C9:1-HQ (I)	-2.59	$6.57 \times 10^{-10}$
M288T13	288.1960	12.7	C9-QNO (I)	-2.41	$4.19 \times 10^{-9}$
M258T13	258.1490	12.7	C7:1-QNO	-1.87	$5.02 \times 10^{-8}$
M242T13	242.1540	13.0	C7:1-HQ	-1.31	$1.64 \times 10^{-5}$

**Table 6.S3.: Continued:** Feature table of all annotated features of the LC-MS exometabolomics data.

identifier	median m/z	median retention time (min)	annotation	directional fold change	corrected p-value
M244T13_1	244.1702	13.1	HHQ	-1.27	$2.23 \times 10^{-7}$
M159T13	159.0678	13.1	HHQ (fragment)	-1.45	$8.96 \times 10^{-8}$
M260T13_1	260.1648	13.2	C7-QNO (I)	-1.24	$2.30 \times 10^{-6}$
M519T13	519.3218	13.2	HQNO [2M+H] <sup>+</sup>	-1.28	0.0104
M296T14_2	296.1997	13.6	C11:2-HQ (I)	-6.00	$1.22 \times 10^{-10}$
M314T14_1	314.2115	13.6	C11:1-QNO	-4.25	$4.86 \times 10^{-9}$
M272T14	272.1642	13.6	C8:1-QNO	-1.46	$3.41 \times 10^{-5}$
M256T14_2	256.1694	14.0	C8:1-HQ	1.34	$9.02 \times 10^{-3}$
M258T14	258.1853	14.0	C8-HQ	-1.17	0.0297
M274T14	274.1802	14.1	C8-QNO	-1.62	$2.09 \times 10^{-3}$
M320T14	320.1832	14.2	C12-HSL	2.45	$1.12 \times 10^{-6}$
M296T14_1	296.2003	14.3	C11:2-HQ (III)	-1.56	$2.31 \times 10^{-5}$
M316T14_2	316.2270	14.4	C11-QNO	-4.99	$3.12 \times 10^{-8}$
M286T14	286.1802	14.4	C9:1-QNO (II)	-1.55	$2.51 \times 10^{-4}$
M268T14	268.1696	14.4	C9:2-HQ	-1.33	0.0616
M300T15	300.1955	14.7	C10:1-QNO	-1.76	$1.43 \times 10^{-4}$
M260T15_1	260.1645	14.7	PQS (II)	-1.13	0.208
M284T15	284.2009	14.8	C10:1-HQ	-2.58	$6.52 \times 10^{-6}$
M272T15_2	272.2010	14.9	C9-HQ	-1.37	$3.80 \times 10^{-5}$
M288T15_2	288.1959	14.9	C9-QNO (II)	-1.21	0.0205
M270T15_1	270.1855	14.9	C9:1-HQ (II)	-1.17	0.167
M342T15	342.2424	15.0	C13:1-QNO	-7.70	$2.69 \times 10^{-7}$
M260T15_2	260.1647	15.3	PQS (I)	-1.80	0.0105
M286T15	286.1799	15.3	C9:1-PQS	-1.72	0.0468
M314T16_1	314.2116	15.5	C11:1-PQS (II)	-2.41	$3.66 \times 10^{-5}$
M298T16	298.2167	15.6	C11:1-HQ (I)	-4.16	$1.93 \times 10^{-7}$
M296T16	296.2009	15.7	C11:2-HQ (II)	-3.89	$2.55 \times 10^{-7}$
M286T16	286.2163	15.8	C10-HQ	-2.52	$2.28 \times 10^{-5}$
M302T16	302.2110	15.8	C10-QNO	-2.12	$2.86 \times 10^{-3}$
M314T16_2	314.2114	16.1	C11:1-PQS (I)	-2.76	$1.35 \times 10^{-5}$
M328T16	328.2266	16.3	C12:1-QNO	-3.40	$1.21 \times 10^{-5}$
M312T16	312.2321	16.4	C12:1-HQ (I)	-5.67	$5.78 \times 10^{-7}$
M677T17	677.4105	16.6	Rha-Rha-C10-C12:1 / Rha-Rha-C12:1-C10	-1.34	$1.90 \times 10^{-7}$
M699T17_3	699.3926	16.6	Rha-Rha-C10-C12:1 / Rha-Rha-C12:1-C10 +Na	-1.31	$4.48 \times 10^{-7}$
M316T17	316.2270	16.6	C11-PQS	-2.22	$1.07 \times 10^{-3}$
M298T17_1	298.2166	16.7	C11:1-HQ (II)	-2.87	$1.07 \times 10^{-5}$
M342T17_1	342.2424	17.0	C13:1-PQS	-3.35	$4.17 \times 10^{-5}$
M326T17_1	326.2477	17.1	C13:1-HQ (I)	-5.58	$5.20 \times 10^{-7}$
M679T17	679.4265	17.2	Rha-Rha-C10-C12 / Rha-Rha-C12-C10	-1.15	$6.83 \times 10^{-3}$
M553T17	553.3348	17.4	Rha-C10-C12:1+Na	-1.24	$3.17 \times 10^{-3}$

**Table 6.S3.: Continued:** Feature table of all annotated features of the LC-MS exometabolomics data.

identifier	median m/z	median retention time (min)	annotation	directional fold change	corrected p-value
M312T18	312.2318	17.5	C12:1-HQ (II)	-2.19	$2.65 \times 10^{-4}$
M705T18	705.4404	17.9	Rha-Rha-C12:1-C12 / Rha-Rha-C12-C12:1	-1.32	0.0595
M555T18	555.3506	18.0	Rha-C10-C12+Na	-1.21	0.0119
M328T18	328.2635	18.4	C13-HQ	-4.22	$2.64 \times 10^{-6}$
M326T18	326.2479	18.4	C13:1-HQ (II)	-3.50	$4.71 \times 10^{-6}$
M707T19	707.4571	18.5	Rha-Rha-C12-C12	-1.05	0.719
M729T19	729.4394	18.5	Rha-Rha-C12-C12+Na	-1.11	0.379
M370T19	370.2735	18.6	C15:1-QNO	-2.89	$5.15 \times 10^{-4}$
M282T19_2	282.2788	19.1	petroselinate	-1.07	0.551
M257T20	257.2472	19.6	palmitate	-1.10	0.362





## 7 | Untargeted LC-MS Metabolomics Differentiates Between Virulent and Avirulent Clinical Strains of *Pseudomonas aeruginosa*

This Chapter has been published as peer-reviewed article in a scientific journal:

T. Depke, J. G. Thöming, A. Kordes, S. Häussler, and M. Brönstrup. “Untargeted LC-MS Metabolomics Differentiates Between Virulent and Avirulent Clinical Strains of *Pseudomonas aeruginosa*”. In: *Biomolecules* 10.7 (July 2020), p. 1041. DOI: 10.3390/biom10071041

### Abstract

*Pseudomonas aeruginosa* is a facultative pathogen that can cause, inter alia, acute or chronic pneumonia in predisposed individuals. The Gram-negative bacterium displays considerable genomic and phenotypic diversity that is also shaped by small molecule secondary metabolites. The discrimination of virulence phenotypes is highly relevant to the diagnosis and prognosis of *P. aeruginosa* infections. In order to discover small molecule metabolites that distinguish different virulence phenotypes of *P. aeruginosa*, 35 clinical strains were cultivated under standard conditions, characterized in terms of virulence and biofilm phenotype, and their metabolomes were investigated by untargeted liquid chromatography—mass spectrometry. The data was both mined for individual candidate markers as well as used to construct statistical models to infer the virulence phenotype from metabolomics data. We found that clinical strains that differed in their virulence and biofilm phenotype also had pronounced divergence in their metabolomes, as underlined by 332 features that were significantly differentially abundant with fold changes greater than 1.5 in both directions. Important virulence-associated secondary metabolites like rhamnolipids, alkyl quinolones or phenazines were found to be strongly upregulated in virulent strains. In contrast, we observed little change in primary metabolism. A hitherto novel cationic metabolite with a sum formula of  $C_{12}H_{15}N_2$  could be

identified as a candidate biomarker. A random forest model was able to classify strains according to their virulence and biofilm phenotype with an area under the Receiver Operation Characteristics curve of 0.84. These findings demonstrate that untargeted metabolomics is a valuable tool to characterize *P. aeruginosa* virulence, and to explore interrelations between clinically important phenotypic traits and the bacterial metabolome.

## Keywords

*Pseudomonas aeruginosa*; virulence; untargeted metabolomics; LC-MS; Random forest classification; biomarker; phenotyping

## 7.1. Introduction

The  $\gamma$ -proteobacterium *Pseudomonas aeruginosa* thrives in various aquatic and terrestrial habitats [1], as well as in multiple niches in the human host such as the lungs, eyes and ears [2–4]. Its affinity and ability to form biofilms on surfaces enables this bacterium to colonize burn wounds and the surfaces of invasive medical devices such as catheters and implants [5]. This wide niche variability, along with its ability to produce various virulence factors and evade the immune system through numerous mechanisms including biofilm formation renders it a clinically relevant, yet difficult-to-treat opportunistic pathogen [6]. *P. aeruginosa* belongs to the group of most dominant bacteria involved in hospital-acquired infections, comprising an estimated 10% of nosocomial infections in the EU [7]. In particular, *P. aeruginosa* is a major threat to artificially ventilated patients [8] and those with cystic fibrosis (CF), of which roughly 60% are chronically infected by *P. aeruginosa* [9].

*P. aeruginosa* displays high metabolic versatility, enabling it to infect and persist in different human tissues and organs [2]. As an example, it has been found that *P. aeruginosa* adapts its iron uptake strategies depending on the type of infection [10]. Nutrient availability in general differs between the various infection sites, and besides metabolism in the narrower sense, strategies of pathogenicity and persistence also need to be tailored to the specific infection situation. Many aspects of *P. aeruginosa*'s adaptability to different infection sites and types have been studied on the genomic level [11]. Genes coding for virulence factors are highly conserved among *P. aeruginosa* strains; however, there is little correlation between genomic features and the type of

infection [12]. Despite a high conservation of virulence factors between clinical and environment samples [13], clinical *P. aeruginosa* strains have been demonstrated to express variable metabolomic, transcriptomic and phenotypic signatures despite almost identical genomes [14, 15]. With respect to pathogenicity, clinical isolates can have vastly different phenotypes and elicit the full range of host responses [16, 17].

Due to the diversity of genotypic and phenotypic traits, it is of considerable clinical interest to describe and differentiate the various metabolic and virulence properties of *P. aeruginosa* clinical isolates. At present, our understanding of genomics is insufficient to fully elucidate the metabolic and phenotypic variation of this bacterium at a clinically relevant level. Transcriptome data can be indicative of virulence phenotypes, yet not always sufficient, either, if used in isolation [18]. While numerous different phenotypic traits, including, but not limited to, swimming motility have been proposed as virulence markers for *P. aeruginosa* clinical strains [19], several studies have suggested investigating metabolomes as functional signatures that might be closer to the actual phenotype [20–22]. In *P. aeruginosa*, many regulators and effectors of virulence are small-molecule secondary metabolites [23] that are generally amenable to liquid chromatography—mass spectrometry (LC-MS) [24–26]. Microbial metabolomics is becoming more prevalent in many areas of microbiology and infection research [27, 28] and has demonstrated itself to be a successful methodology to, e.g., discriminate between different *Bacillus cereus* strains [29], to describe and differentiate drug susceptibility phenotypes in the parasite *Leishmania donovani* [30] as well as in the  $\gamma$ -proteobacterium *Acinetobacter baumannii* [31], to identify volatile metabolites in different *P. aeruginosa* strains [32] and to describe the metabolic adaptations of *P. aeruginosa* strains colonizing different niches in CF lungs [33].

Hence, we tested whether the virulence properties of *P. aeruginosa* clinical strains can be differentiated by untargeted LC-MS metabolomics. Metabolomics data for 35 clinical isolates recovered from diverse infection sites was acquired, stratified according to *in vivo* virulence phenotypes in the *Galleria mellonella* infection model [34] and biofilm phenotypes [15] and analyzed for discriminating markers. Data on the metabolite level and metabolic profiles were investigated, and a statistical model was generated to differentiate virulent and avirulent strains based solely on LC-MS data.

## 7.2. Materials and methods

### 7.2.1. Bacterial strains

Bacterial strains were selected from a biobank of *P. aeruginosa* clinical isolates curated at the Helmholtz Centre for Infection Research in Braunschweig, Germany, which is documented in the ‘Bactome’ database [35]. Strains were collected in clinical microbiology laboratories, private practice laboratories, or provided by strain collection curators across Germany, Spain, Hungary and Romania. Clinical isolates used in this study were previously characterized with regard to clinically relevant phenotypes [35–37], including *in vitro* biofilm phenotypes [15] and an *in vivo* virulence infection model using *Galleria mellonella* [34]. A list of strains and their phenotypic properties (biofilm and virulence) can be found in Table 7.1.

### 7.2.2. Transcriptomics

Transcriptional profiles of all clinical isolates used in this study were produced for a previous study [36]. Briefly, planktonic bacteria were cultivated to early stationary phase ( $OD_{600} = 2$ ) in 10 ml LB under shaking conditions (37 °C, 180 rpm). Three independent cultures were pooled to obtain one transcriptional profile per strain. cDNA libraries were generated using the ScriptSeq™ v2 RNA-seq Library Preparation Kit (Illumina), and samples were sequenced in single end mode on an Illumina HiSeq 2500 device (1 × 50 bp reads). The reads were mapped to the UCBPP-PA14 reference genome (NC\_008463.1, available for download from the Pseudomonas Genome database: <http://v2.pseudomonas.com>) using the stampy pipeline [38]. RNA-Seq data of clinical isolates is uploaded at NCBI’s Gene Expression Omnibus (GSE123544). Differential gene expression analysis was performed using the R package DESeq2 (v.1.18.1) [39] with default settings to calculate the normalized reads per gene (nrpg). For the identification of differentially expressed genes between virulent and avirulent strains, a threshold of  $\log_2(\text{fold change}) \geq 1$  and  $\leq -1$  respectively with  $p_{\text{adj}} < 0.05$  was applied. Only genes assigned to the core genes (according to Mathee *et al.* [40]) were considered for the analysis to account for differences in strain backgrounds (PA14 vs. PAO1). DESeq2 was used to generate a principal component analysis (PCA) plot from the transcriptional profiles.

**Table 7.1.:** Strains used in this study. Strains are listed by their identifier in the Bactome data base [35]. The biofilm cluster corresponds to the biofilm morphology [15] and the survival rate in a *G. mellonella* infection model (according to [34]) is given in percent survival after 48 h. The infection or sampling site is indicated if available. nd – not determined.

Discovery data set			
Strain	Biofilm cluster	<i>Galleria</i> survival (48 h)	Infection/sampling site
CH2860	A	5	Respiratory tract
CH4528	A	0	Respiratory tract
ESP046	A	0	nd/other
ESP088	A	5	nd/other
F2030	A	0	Respiratory tract
MHH16798	A	20	Respiratory tract
ZG302383	A	0	nd/other
CH2682	B	95	Rectal swab
ESP027	B	100	nd/other
F1959	B	100	Respiratory tract
F2165	B	100	Respiratory tract
F2166	B	100	Respiratory tract
F2224	B	95	nd/other
MHH17767	B	100	Respiratory tract
Validation data set			
Strain	Biofilm cluster	<i>Galleria</i> survival (48 h)	Infection/sampling site
CH2690	A	0	Rectal swab
ESP058	A	0	nd/other
ESP067	A	5	nd/other
F1997	A	0	Rectal swab
MHH17704	A	5	nd/other
Psae1439	A	10	Respiratory tract
ZG8038581181	A	10	Respiratory tract
CH4681	B	90	Respiratory tract
F1764	B	95	Respiratory tract
F2020	B	95	Wound infection
MHH16050	B	60	nd/other
MHH16563	B	95	Respiratory tract
MHH17546	B	100	Respiratory tract
Psae1837	B	75	nd/other
Additional data set			
Strain	Biofilm cluster	<i>Galleria</i> survival (48 h)	Infection/sampling site
CH2706	C	0	Rectal swab
CH4591	C	0	Rectal swab
ESP083	C	0	nd/other
F1864	C	0	nd/other
F2059	C	0	Wound infection
ZG316717	C	5	Ear infection
ZG8510487	C	0	Urinary tract infection

### 7.2.3. Untargeted metabolomics

All chemicals and analytical standards used in the metabolomics experiments in this study correspond to those previously described [24]. Selected strains were cultivated and measured in two distinct and independent batches, a discovery and a validation batch (cf. Table 7.1). For data analysis, the validation batch was divided into two sub-batches: the actual validation set containing strains with phenotypes that were also present in the discovery batch, and an additional set of isolates with virulent phenotypes containing the cluster C biofilm phenotype that was not present in the discovery batch. With this setup, the actual validation set can gauge the performance and validity of the applied statistical classification models (see below), because it contained strains that should be classified into the same categories as the strains in the discovery data set. As the phenotypes in the discovery data set are defined by virulence as well as biofilm morphology, an additional set of strains with a different combination of these two properties was needed to assess whether the model was able to differentiate solely the virulence phenotype irrespective of biofilm morphology. Overnight precultures grown in 3 mL LB medium in glass tubes were inoculated from plate cultures for each strain and incubated for approximately 16 h at 37 °C and 140 rpm in a shaking incubator. Three independent biological replicates were inoculated with a starting OD<sub>600</sub> of 0.05 from each preculture. Cultures were subsequently grown to an OD<sub>600</sub> of approximately 2. Measured OD<sub>600</sub> values for each sample were later used for normalization and can be found in Tables 7.S1, 7.S2 and 7.S3. 2 mL of each sample was collected and immediately centrifuged at 9000×g, at 4 °C for 5 min. Cell pellets were snap-frozen in liquid nitrogen and subsequently stored at –20 °C until all of the batch samples were processed to this stage.

Metabolite extraction was performed as previously [24]. In brief, cell pellets were extracted in 500 µL methanol containing 0.1 mg/L trimethoprim, 0.1 mg/L nortriptylin and 0.3 mg/L glipizide as internal standards through the use of vigorous shaking and sonication. Extracts were separated from solid matter by centrifugation. 400 µL of each extract was concentrated to dryness and resuspended in 40 µL 50% (v/v) acetonitrile with 0.1% formic acid containing 1 mg/L caffeine and 8 mg/L naproxen as internal standards.

A 1 µL aliquot of each sample was analyzed by reversed phase ultra-high performance liquid chromatography coupled to quadrupole time-of-flight mass spectrometry as previously [24, 41]. Tandem mass spectra were recorded from pooled quality control samples and used for metabolite identification by comparison to authentic standards and/or

metabolite databases as described in a previous study [24] (cf. Table 7.S4). LC-MS data were exported to mzXML using Bruker Compass Xport, and preprocessed with XCMSonline [42] with the parameters listed in Table 7.S5. Our discovery, validation and additional data sets were all processed separately.

After preprocessing, the untargeted metabolomics data underwent further processing using R/RStudio with ‘tidyverse’ packages as statistical software [43–45]. First, features eluting at retention times  $\leq 0.8$  min and  $\geq 20$  min, those displaying a relative standard deviation of  $\leq 20\%$  over all samples, and those with an intensity of  $\leq 10000$  counts were removed. Subsequently, the data was consecutively normalized through the use of internal standards; first with those added upon reconstitution (caffeine and naproxen), and then with those added during extraction (trimethoprim and nortriptyline). The data for each sample was further normalized through the use of the respective OD<sub>600</sub> at harvest as a proxy for cell number (Tables 7.S1, 7.S2 and 7.S3). Annotations were added and isotope peaks were identified using ‘CAMERA’ [46] (as part of the XCMSonline workflow) and removed from the data sets. The resulting feature tables for the discovery and the validation data sets were used for data analysis and model building.

Feature credentialing by means of stable isotope enriched growth medium [47] was performed in a previous study [48] and used to verify the biological origin of a candidate biomarker.

The Mass Spectrometry Search Tool (MASST) on the Global Natural Products Social Molecular Networking (GNPS) repository was used to match unidentified spectra of particular interest to previously reported MS<sup>2</sup> data [49]. The standard parameters of the search were used: MS<sup>2</sup> fragment ions were excluded if their  $m/z$  difference to the precursor ion was less than 17 Da and spectra were filtered using an approach called window filtering that keeps the 6 most abundant fragment ions within a  $\pm 50$  Da window throughout the spectrum. The  $m/z$  tolerance of the search was 2 Da for the precursor ion and 0.5 Da for MS<sup>2</sup> fragment ions. To be considered a match, the queried spectrum and library spectra had to display a cosine similarity score of  $\geq 0.7$  and  $\geq 6$  matched peaks.

The raw data was uploaded to MetaboLights [50] and can be accessed via the study identifier MTBLS1749.

#### 7.2.4. Data analysis and model building

PCA was used with centering and rotating of the variables, and PCA scores were plotted for data exploration. Directional fold changes were calculated for all features

with positive values signifying higher abundances in the virulent group and negative values signifying higher abundances in the avirulent group. Statistical significance of between-group differences was assessed by performing a Benjamini-Hochberg corrected two-sided Welch's *t*-test for each feature. The Mann-Whitney *U*-test (Wilcoxon rank-sum test) was used to test for statistical significance for individual comparisons when non-normality was suspected. Using the 'vegan' R package [51], permutational multivariate analysis of variance (PERMANOVA) of a Bray-Curtis distance matrix of the metabolite abundance data was employed to test the correlation of metabolite profile to the phenotypic group. The same package was used to calculate the Shannon index to gauge differences in metabolome diversity between the samples of the different datasets.

Predictive models were built using random forest classification and the 'randomForest' R package [52, 53] with 1000 trees per forest, and 500 randomly sampled variables considered as candidates at each split. Feature importance was assessed by mean increase of the Gini coefficient and mean increase of variable importance (VIP).

Model validation was performed by matching features of the discovery and validation data set and subsequent prediction of the phenotypes for all samples of the validation data set in the form of probabilities. The features were matched by comparing *m/z* and retention time, using a tolerance of 5 ppm and 1 min, respectively. Only features present in the discovery data set that matched a feature in the validation data set were considered in model building and validation. The same procedure was applied for the additional data set. Model quality was assessed by calculation of the area under the receiver operating characteristics (ROC) curve (AUC), using the 'ROCR' R package [54], where 1 corresponds to a perfect model and 0.5 is equivalent to random prediction. As the additional cluster C data set contains only one group of isolates, it was not possible to construct a ROC curve. Instead, the frequency of correct predictions for the strains in 100 independent constructions and predictions by the random forest model was assessed. This was also done for the first validation set.

### 7.3. Results

Untargeted LC-MS metabolomics data were recorded and analysed for 35 *P. aeruginosa* clinical strains differing in their virulence as determined by a *G. mellonella* survival model and their biofilm phenotype which was categorized into the three main clusters A, B and C [15]. 14 strains, seven virulent strains with a cluster A biofilm phenotype and seven avirulent cluster B strains constituted the discovery data set,

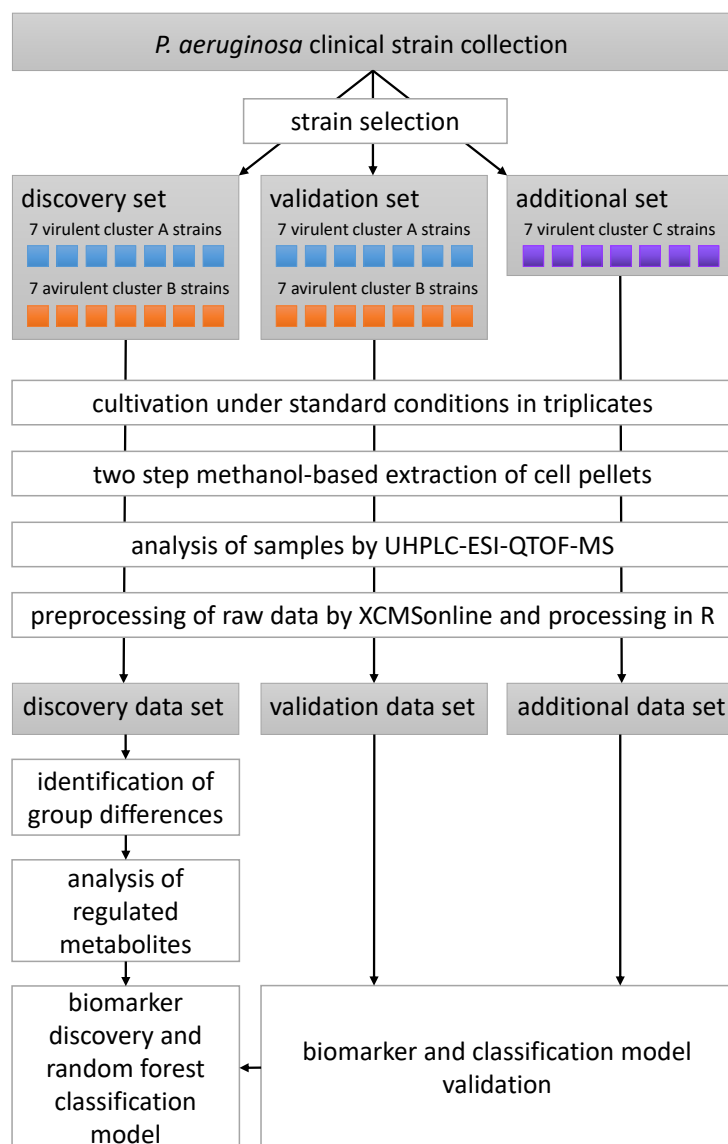


that was analysed in depth for metabolomic differences between the phenotypes and used to build a random forest classification model. Another 14 strains equally split between the two phenotypic groups represented in the discovery group served as the validation data set. These strains were used to test the model constructed from the discovery data set. An additional seven virulent strains corresponding to a tertiary cluster C biofilm phenotype—i.e. a phenotype which was not present in the discovery data set—were used to investigate whether our classification model is capable of predicting virulence in a biofilm phenotype independent manner. All strains were cultivated in rich medium under standard planktonic conditions, extracted using a methanol-based protocol, separated on a reversed-phase C18 column and detected using time-of-flight mass spectrometry following electrospray ionization in positive mode (ESI-QTOF-MS). The study design is visualized in Scheme 7.1.

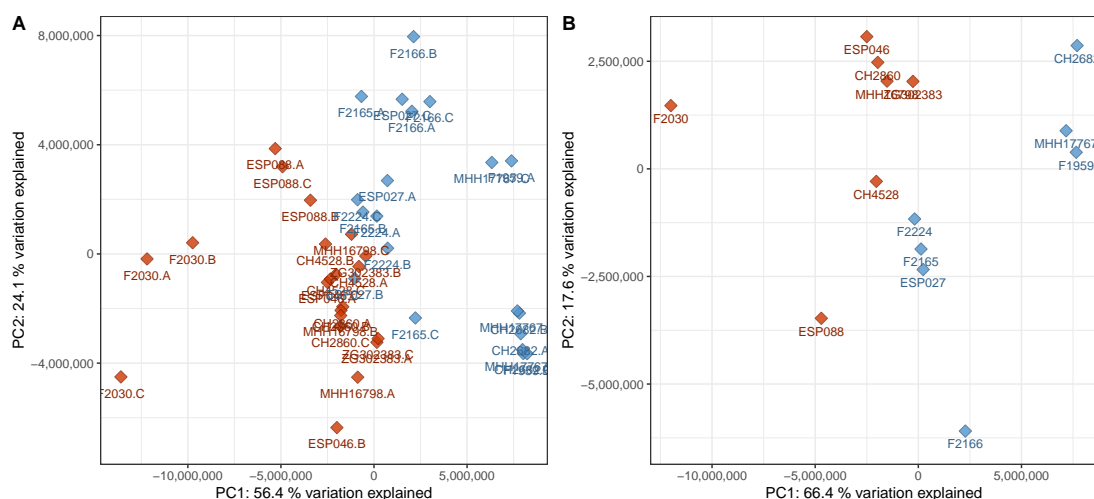
### 7.3.1. Virulent cluster A and avirulent cluster B strains have different metabolic profiles

Overall variation in the untargeted metabolomics dataset (differences of signal abundances between and within groups after normalization and filtering with respect to all detected signals) was assessed using PCA. As an unsupervised method, PCA does not use class information, thereby preventing potential bias when judging separation between sample groups. Upon visual inspection, the PCA scores plot of the discovery data set, i.e. the data from the set of *P. aeruginosa* strains used to generate the classification model, shows a good but not complete separation between virulent cluster A and avirulent cluster B strains (Figure 7.1). Although the two phenotypes do not form compact clusters, there is little overlap between virulent and avirulent strains. Clear separation in an unsupervised analysis suggests that there is potential for a supervised method to model the data in a superior manner. A PERMANOVA analysis further supports the notion that the metabolome differences are associated with the virulence and biofilm phenotype ( $F = 10.7$ ,  $p = 0.001$ ). Moreover, the strains do not cluster according to other parameters such as time to reach the specified OD<sub>600</sub>, or the hospital they were originally isolated in (Figure 7.S1). These findings suggest that, among our available metadata on the utilized *P. aeruginosa* isolates, virulence and biofilm phenotypes are the main drivers of variation between metabolomes.

The discovery data set contains 2359 features, whereof 135 were structurally annotated that corresponded to 96 unique metabolites. 332 features were significantly differentially abundant with fold changes greater than 1.5 in both directions as well



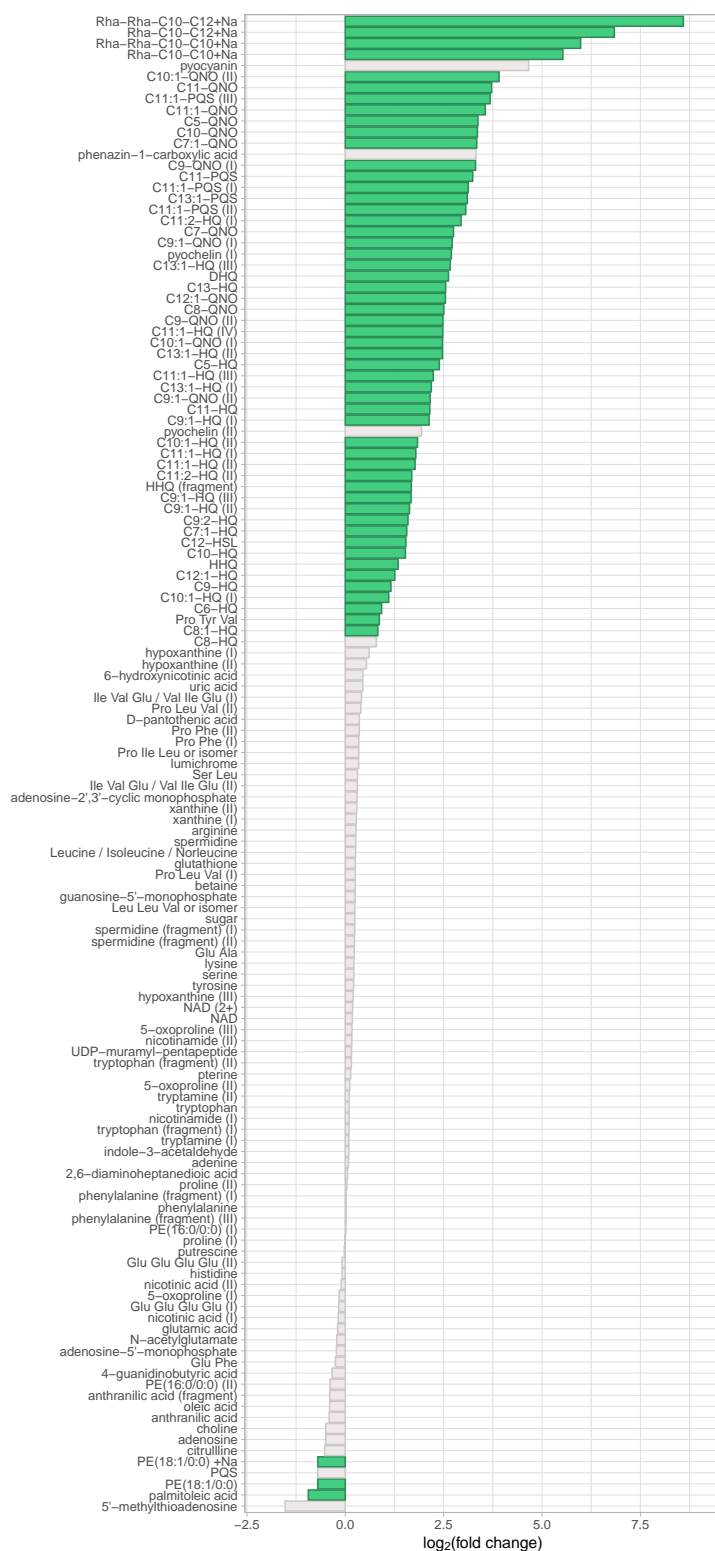
**Scheme 7.1.:** Schematic of the experimental and data analysis workflow of the metabolomics part of this study.



**Figure 7.1.:** PCA scores plot of the discovery data set. **A:** All samples in the discovery data set were subjected to a principal components analysis with centering and rotation of the variables. The samples were plotted using principal components 1 and 2 as coordinates. A good separation of the groups is visible, predominantly in principal component 1, which explains 56% of the variation in the data set. **B:** The data from the triplicate samples was averaged per strain and subjected to a PCA as described above. The separation is very similar to the data on the sample level. A clear separation can be achieved by considering both principal components 1 and 2. Red – virulent cluster A strains, blue – avirulent cluster B strains.

as with a Benjamini-Hochberg corrected  $p$ -value of less than 0.05. 299 of these features were more abundant in the virulent group and 56 of the 332 (17%) have been structurally annotated (cf. Figure 7.S2), corresponding to 40 unique metabolites.

Among the identified metabolites with differential abundance, secondary metabolites are found along with lipids. Secondary metabolites were discovered at higher levels, while lipids were less abundant, in the virulent strains (Figures 7.2 and 7.3). Virulent and avirulent *P. aeruginosa* strains did not differ in their relative abundances of primary metabolites. A PCA scores plot considering only identified metabolites provides good separation between the groups that were tested (Figure 7.S3). Additionally, the PCA loadings plot of the complete discovery set, to which degree features contribute to the principal components, provides evidence that most features with high loadings have been annotated (Figure 7.S4). These two observations demonstrate that most relevant metabolites, or at least members of the most relevant metabolite families, have been annotated. Interestingly, a PCA plot based on gene expression profiles did not show any clustering of the isolates according to their affiliation to a particular virulence phenotype (Figure 7.S5).



**Figure 7.2.:** Regulation of identified metabolites in the discovery data set. The binary logarithm of their (non-directional) fold change was plotted on the x-axis for all identified metabolites. Features exhibiting a fold change  $\geq 1.5$  and a corrected  $p$ -value  $\leq 0.05$  were coloured in green.



**Figure 7.3.:** Box plots of identified metabolites in the discovery data set. Data distribution for all identified metabolites with an absolute fold change of  $\geq 1.5$  is shown as box plots. For each metabolite, one boxplot shows the abundances in each group. Red – virulent cluster A strains, blue – avirulent cluster B strains.

### 7.3.2. Metabolic differences between virulent cluster A and avirulent cluster B strains manifest in differential abundance of virulence-associated secondary metabolites

Phenazines, prominent pseudomonal secondary metabolites with well-studied roles in pathogenesis and host cell damage, act as signaling molecules for the transcription factor SoxR which promotes the production of efflux pumps and the depletion of glutathione leading to redox instability of the host [55, 56]. In this study, two phenazines were identified: pyocyanin and its congener phenazine-1-carboxylic acid. When comparing our virulent strains to our avirulent strains, both metabolites exhibit large differences in abundance with fold changes of +25 and +10, respectively. However, the difference between our virulent and avirulent strains is not statistically significant if tested using a Benjamini-Hochberg corrected Welch's *t*-test due to the large variation within the virulent group containing two high producing strains, while all of the other strains produced much more modest phenazine levels (Figure 7.S6). A non-parametric significance assessment using the Mann-Whitney *U*-test yields *p*-values of  $2.6 \times 10^{-7}$  and  $9.5 \times 10^{-5}$  for pyocyanin and phenazine-1-carboxylic acid, respectively, thus suggesting significant differences in the abundance of phenazines in the two groups. Phenazine biosynthesis was also highly upregulated at the transcriptional level in virulent *P. aeruginosa* strains in comparison to the avirulent clinical isolates tested (Table 7.S6), supporting the notion that phenazine production is associated with a virulent phenotype.

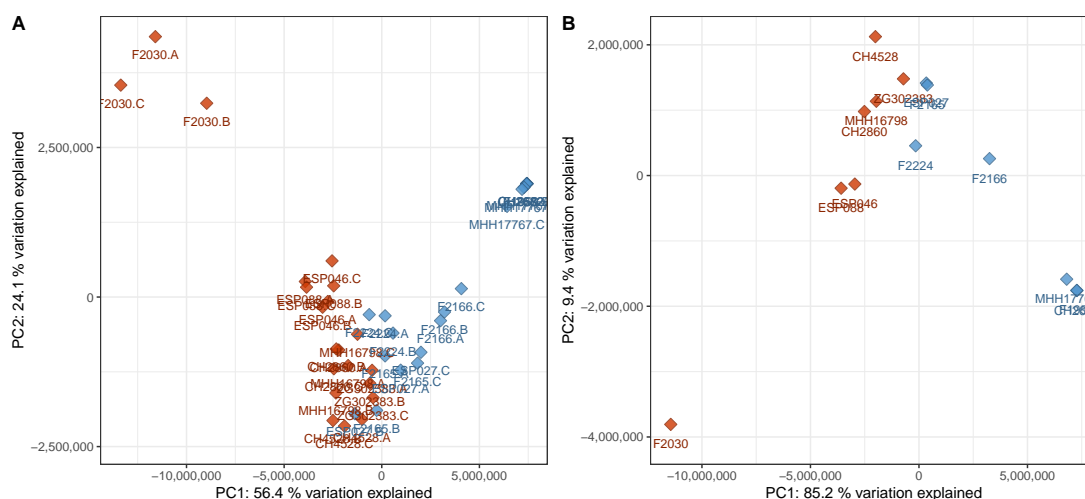
Alkyl quinolones (AQ), important quorum sensing signaling molecules unique to *P. aeruginosa* and closely related species, are involved in various virulence-associated processes [23]. Transcriptional profiles tend to provide evidence of the elevated expression of genes involved in the AQ biosynthesis in virulent isolates, however, the expression levels are not statistically significant (threshold  $\log_2(\text{fold change}) \geq 1$  and  $\leq -1$  with  $p_{\text{adj}} < 0.05$ ) between the two groups (Table 7.S7). Strikingly, the abundance of AQs in the metabolome of virulent *P. aeruginosa* strains is much greater than in the metabolome of the avirulent strains (Figure 7.2). For the highly abundant and important AQs HHQ (C7-HQ) and HQNO (C7-QNO), directional fold changes of +2.5 and +6.7 and corrected *p*-values of 0.0002 and 0.007, respectively, were observed. The most differentially abundant AQ is C10:1-QNO (directional fold change +15, corrected *p*-value 0.04), a metabolite with very low abundance (roughly 40× and 100× lower levels than C7-QNO in virulent cluster A and avirulent cluster B strains, respectively). The most significant difference was recorded for C9:1-HQ which was 3.2× more abundant in virulent cluster A strains with a corrected *p*-value of  $3.1 \times 10^{-6}$ . The various AQ

congeners detected in this study consistently showed significantly higher levels in the virulent group, had high loadings in the PCA and were good predictors in the random forest classification models, thereby emphasizing their importance in the regulation of virulence.

The largest fold changes between our virulent and avirulent strains were found in the rhamnolipids, another class of virulence-associated secondary metabolites (Figure 7.S7). These surface-active glycolipids play multiple roles in the establishment and maintenance of infection, including the transition between biofilm and planktonic lifestyle [57] and the impairment of the host airway epithelium [58]. Rhamnolipids enable the poorly water-soluble *Pseudomonas* quinolone signal (PQS) to diffuse in aqueous environments, as they enhance the solubility of PQS through their amphiphilic properties, thereby potentiating PQS-driven effects on virulence [59]. Four different rhamnolipid structures have been annotated; namely Rha-Rha-C10-C12, Rha-Rha-C10-C10, Rha-C10-C12 and Rha-C10-C10 (Figure 7.S8), and all of them were significantly more abundant in the virulent group (fold changes of +386, +63, +46, +114, respectively and corrected *p*-values of 0.001, 0.001, 0.01 and 0.02, respectively; always for Na adduct). In most avirulent strains, rhamnolipids were practically absent; likewise, some virulent strains barely produced any rhamnolipids, while others produced highly elevated levels of this secondary metabolite (Figure 7.S7): For Rha-Rha-C10-C12, all avirulent cluster B strains and the virulent cluster A strains F2030, ESP088 and CH4528 showed a peak area in arbitrary units below 100, whereas the other virulent cluster A strains featured peak areas ranging from approximately 3000 to 40,000.

The pseudomonal siderophore pyochelin exists in *trans* and *cis* isoforms [60], both of which have been annotated. Interestingly, only one of the species is significantly regulated, and more prevalent in the virulent strains (directional fold change +6.5, corrected *p*-value 0.006). Pyochelin, an important player in iron acquisition and homeostasis, has been linked to virulence, although it is not necessarily directly harmful to the host [10]. Despite the significant differences in the amount of rhamnolipids and pyochelin produced by virulent strains over avirulent strains, the corresponding genes for both rhamnolipid and pyochelin biosynthesis were not differentially expressed between these two groups (Table 7.S7).

Multiple primary and intermediate metabolites have been associated with pseudomonal virulence; however, no clear trends could be identified in the present study. For instance, tryptophan and phenylalanine, both of which share biosynthetic pathways with alkylquinolones and phenazines through the common precursor chorismate



**Figure 7.4.:** PCA scores plot of the discovery data set with only virulence-associated secondary metabolites included into the calculation. The plot has been created analogously to Figure 7.1 and shows a comparably good separation between the two phenotypes based on virulence-associated secondary metabolites only. **A:** Data on sample level, **B:** data on strain level (cf. Figure 7.1). Red – virulent cluster A strains, blue – avirulent cluster B strains.

and are known inducers of PQS production [61], are not differentially abundant in the two groups (directional fold changes  $+1.1$  and  $\pm 1.0$ , corrected  $p$ -values  $0.5$  and  $0.9$ , respectively). Anthranilic acid, which is closely connected to the biosynthesis of phenazines, was also not significantly differentially abundant between our virulent and avirulent strains (directional fold change  $-1.3$ , corrected  $p$ -value  $0.4$ ).

When only identified metabolites that are known to be virulence-associated—AQs, DHQ, homoserine lactones, pyochelin, phenazines and rhamnolipids—are considered, a good separation between cluster A and cluster B is still visible in the PCA scores plot (Figure 7.4).

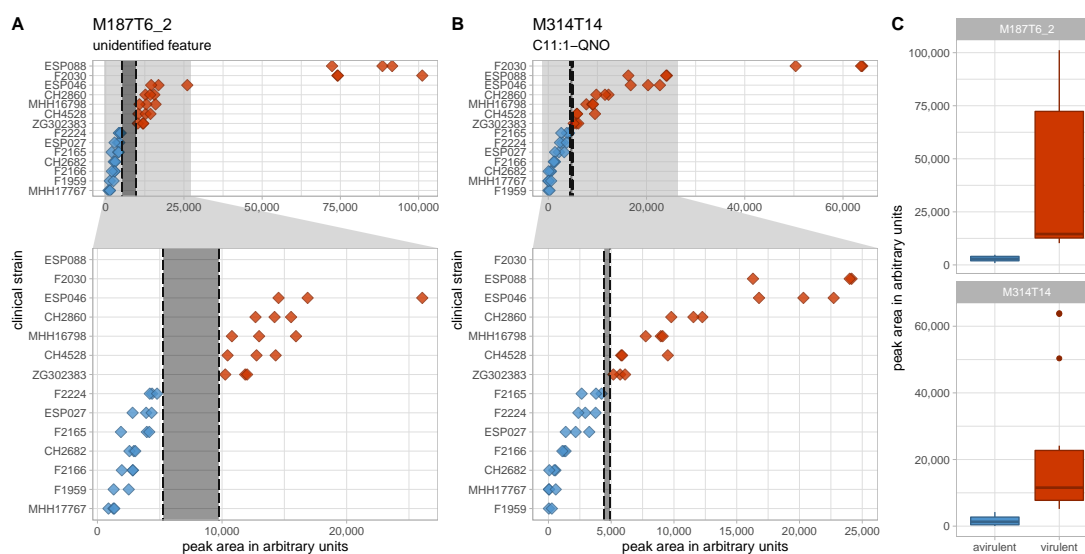
Figure 7.4 reveals that the metabolic profiles of one strain, F2030, differ from those of the other strains of the virulent cluster A group. This strain produces even higher AQ levels than the other virulent strains but displays lower levels of other virulence-associated metabolites. Compared to the other samples of the same phenotype, F2030 sample harbour  $2.1\times$  more HHQ,  $5.6\times$  more C11-QNO and  $3.6\times$  more DHQ, but  $2.2\times$  less C12-HSL,  $20\times$  less Rha-C10-C10 and  $27\times$  less pyocyanin. The trends for the respective congeners are consistent.



### 7.3.3. An unknown metabolite is a potential biomarker for virulent phenotypes

In the search for classification biomarkers, the most interesting features are those that enable a clear group separation. In the discovery data set, there were only two features whose maximum levels in one group were lower than the minimum levels in the other group, thus allowing for an intensity threshold for the respective features to separate the groups completely (Figure 7.5). One is M314T14, putatively identified as C11:1-QNO, a relatively low abundant AQ. As all other AQ congeners show overlapping levels, M314T14 is most probably not a robust separator. The second feature, M187T6\_2, however, has a larger non-overlapping intensity space between the two groups and thus appears to be a more promising separator. This feature exhibits a directional fold change of +11.7 and a Benjamini-Hochberg corrected  $p$ -value of 0.003. Statistical significance is also suggested by a  $p$ -value of  $7.4 \times 10^{-12}$  determined by the non-parametric Mann-Whitney  $U$ -test.

Despite considerable efforts, the identity of the feature could not be revealed by annotation strategies from the mixture, and efforts to purify the compound from raw extracts failed due to its very low abundance. Nonetheless, it could be demonstrated by means of feature credentialing [47] that the feature represents a metabolite produced by *P. aeruginosa* as it incorporated  $^{13}\text{C}$  from  $^{13}\text{C}_6$ -glucose if supplied to the growth medium (Figure 7.S9) [48]. Exact  $m/z$  and isotopic pattern analysis suggested the sum formula  $\text{C}_{12}\text{H}_{15}\text{N}_2$  for the positively charged ion. The  $\text{MS}^2$  spectrum of the feature is rather uninformative due to very weak fragmentation (Figure 7.S10). However, the most dominant fragment peaks (relative intensity compared to base peak > 5%) display  $m/z$  ratios that support the aforementioned sum formula, with an  $m/z$  of 145.076 corresponding to  $\text{C}_9\text{H}_9\text{N}_2$ , an  $m/z$  of 144.068 to  $\text{C}_9\text{H}_8\text{N}_2$ , and an  $m/z$  of 91.054 to  $\text{C}_7\text{H}_7$  – all possible fragments of  $\text{C}_{12}\text{H}_{15}\text{N}_2$ . This formula is consistent with a reduced phenazine structure, namely hexahydrophenazine. A MASST search for similar  $\text{MS}^2$  with the same precursor in the GNPS data base [49] revealed that the feature has been detected in three other mass spectrometry studies that examined *P. aeruginosa* samples or bacterial samples from patients infected with CF (data sets MSV000080397, MSV000080337, MSV000080251, and MSV000079680, accessible at <https://gnps.ucsd.edu/>). Although none of studies found a meaningful annotation for the feature in question, its presence in other *P. aeruginosa* related data sets further supports the notion that it is an actual pseudomonas metabolite rather than an artifact.



**Figure 7.5.:** Levels of the two non-overlapping features in the discovery data set. **A, B:** Peak area in arbitrary units is plotted on the x-axis for the two perfectly separating features M187T6\_2 (**A**) and M314T14 (**B**). Full range of the x-axis is shown in the top panels and an enlarged region of the lower x-axis range is depicted in the lower panels. Non-overlapping spaces are marked by dashed lines and grey shading. The difference between the lowest level in the virulent cluster A group and the highest level in the avirulent cluster B group is a lot more pronounced for M187T6\_2. **C:** Peak area in arbitrary units is plotted on the y-axis for the two perfectly separating features M187T6\_2 and M314T14 as box plots to visualize data distribution. Red – virulent cluster A strains, blue – avirulent cluster B strains.

The feature's intensity levels do not correlate with those of pyocyanin or phenazine-1-carboxylic acid and correlate only weakly or insignificantly with rhamnolipids, C12-HSL or aromatic amino acids like phenylalanine, tryptophan or anthranilic acid. However, they do display a strong and significant correlation with Aqs and pyochelin (Figure 7.S11). The strongest (positive) correlation to an annotated feature is with DHQ (Pearson's correlation coefficient of +0.93), suggesting a potential link to AQ biosynthesis. The strongest negative correlation to an annotated feature, in turn, was observed for adenosine, but appears uninformative with a weak Pearson's correlation coefficient of -0.47 despite its statistical significance.

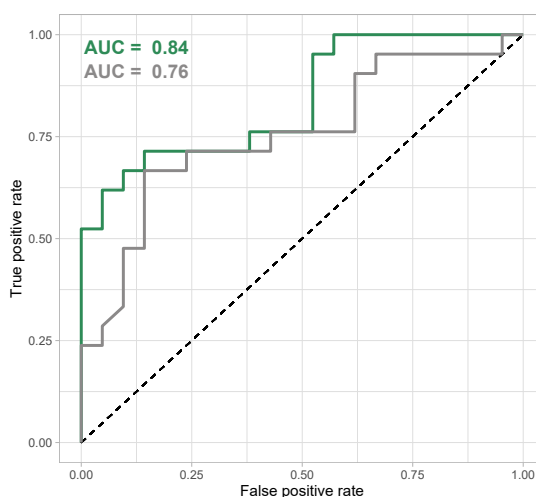
M187T6\_2 was tested as a potential marker for the differentiation of virulent and avirulent strains using a validation data set of another 14 clinical isolates—7 virulent and 7 avirulent, displaying the same biofilm phenotype as those in the discovery data set—that was processed and analyzed analogously to the discovery data set. The feature corresponding to M187T6\_2 in the validation set was significantly differentially abundant between the two virulence phenotype groups with a directional fold change

of +3.6 and a  $p$ -value of 0.005 (Welch's  $t$ -test) and 0.002 (Mann–Whitney  $U$ -test); however, it was unable to perfectly separate the groups (Figure 7.S12). This is in concordance with higher metabolome diversity of the clinical strains (mean Shannon index of the samples in the discovery and validation data set: 5.93 and 6.25, respectively;  $p < 10^{-17}$ ), as illustrated by a more diffuse distribution of the virulent and avirulent strains in the PCA of the validation data set (Figure 7.S13). A univariate model using the abundance of M187T6\_2 as a separator for virulent and avirulent phenotypes yields a fair area under the ROC curve of 0.75 (Figure 7.S14). The feature in question does not display an intra-group correlation with the surrogate parameter used for *in vivo* virulence in this study – the survival of infected *Galleria mellonella* larvae after 48h (Figure 7.S15).

#### 7.3.4. Virulent and avirulent strains with distinct biofilm phenotypes can be differentiated based on untargeted metabolomics data by machine learning

Since neither the single putative marker M187T6\_2, nor any other metabolite could achieve a perfect group separation in the validation set, we tested whether a multi-metabolite classification model is able to reliably discriminate virulent and avirulent strains with their respective biofilm phenotypes. Random forest classification was selected from the plethora of machine learning classification models [62] as it does not require data to be on the same scale, and allows for easy interpretation of the features' contribution to the model. Random forest classification has also been shown to be a powerful method for phenotype discrimination based on clinical metabolomics data [63].

A random forest model was trained using the discovery data set, regarding only features that have been found and integrated both in the validation and discovery data set. The model is able to discriminate the groups in the discovery data set very clearly (Figure 7.S16). Unsurprisingly, the M187T6\_2 feature described above is the most important feature in the model. Among the ten most important features were also C9-QNO and two isomers of C9:1-HQ, placing three AQ features in the top ten of separating markers in the model. The remaining six of the most important features could not be identified. These include features with low  $m/z$  (M85T1\_1 and M126T1\_1), features in the  $m/z$  range of AQs and other secondary metabolites (M231T7\_3, M246T3\_1 and M228T12) and a slightly larger one (M464T9\_3) (Figure 7.S17). In a model based only



**Figure 7.6.:** ROC curve for random forest classification model based on all features (green) and on identified features only (grey). The area under the Receiver Operation Characteristics curve is used as a quality metric for classification models and shows a good classification for a random forest classification model that uses all features in the data set, compared to a slightly weaker performance for a model that only considers identified features.

on identified features, AQ congeners make up eight of the ten most important features along with Rha-Rha-C10-C12 and pyocyanin.

The classification model was applied to the validation data set to gauge its capacity to correctly predict virulence phenotypes from new metabolomics data obtained from *P. aeruginosa* clinical strains. The model shows a good prediction performance—especially regarding the larger heterogeneity of the validation data set—as signified by an area under the ROC curve of 0.84 (Figure 7.6). If only identified features are regarded, the area under the ROC curve of 0.76 is still fair, but including unknown features improves the prediction performance of the model. Remarkably, some strains are systematically misclassified (their metabolomes do not correspond to their virulence and biofilm phenotype in the way the classification model connects these two types of data) (Figure 7.S18).

As the isolates of the discovery and validation data set differ in two phenotypes, virulence and biofilm phenotype, we analysed a third group consisting of isolates that have a virulent phenotype in the *G. mellonella* model but a different (cluster C) biofilm morphology (Figure 7.S19). Applying the random forest model to this group of isolates resulted in a true positive rate of only 47% (Figure 7.S20) suggesting that despite a similar virulence phenotype, cluster A and cluster C isolates differ significantly in their metabolome.

## 7.4. Discussion

The present study provides evidence of systematic differences in the virulence-associated metabolome of several clinical *P. aeruginosa* strains isolated from several different types of infections in different hospitals across Europe. The metabolic profiles of virulent and avirulent strains with cluster A and cluster B biofilm phenotypes, respectively, in the discovery data set were sufficiently different to identify a separation between the two groups, even with an unsupervised method such as PCA, while other strain or cultivation-related properties were unable to separate these two groups.

Very few primary metabolites were significantly differentially abundant between the virulent cluster A and the avirulent cluster B groups, corroborating data from previous studies demonstrating that the primary metabolome of *P. aeruginosa* strongly depends on growth conditions, and weakly on the strain or genetic background [64]. Of the 43 identified distinct primary metabolites (cf. Table 7.S8), only 11 had significantly different levels in the two groups in the discovery data set (with an adjusted  $p$ -value of  $\leq 0.05$ ), and only palmitoleic acid and lyso-PE-18:1 met the additional criterion of having a fold change  $\geq 1.5$  or  $\leq -1.5$ . Furthermore, the rich growth medium used in this study may have left some anabolic pathways inactivated, potentially excluding a group of metabolites that correlate with the virulence phenotype under different growth conditions.

In contrast to the primary metabolome, the secondary metabolome was substantially different between the two groups. The variance between the secondary metabolomes of the two groups enabled a group separation based solely on the abundances of identified virulence-associated secondary metabolites (Figure 7.4). The ability to detect all major groups of virulence-associated secondary metabolites in *P. aeruginosa* in a single LC-MS method underlines the usefulness of this technology in gauging pseudomonal virulence and its relation to metabolism. The strong upregulation of virulence-associated metabolites is responsible for the asymmetry in the volcano plot (Figure 7.S2), that exhibits a substantially larger number of upregulated features compared to downregulated features. The most important differentially abundant secondary metabolites discovered were AQs, which are known to regulate virulence in several ways [23]. In a recent study using Rapid Evaporative Ionisation Mass Spectrometry to differentiate CF and non-CF *P. aeruginosa* isolates, Bardin *et al.* identified AQs and rhamnolipids as important features for phenotype classification, with lower AQ levels in mucoid isolates [65]; thereby highlighting the notion that AQ profiles are an integral part of strain-specific metabolic profiles in *P. aeruginosa*. The abundance of quorum sensing signaling molecules in

clinical isolates have been investigated in depth. Despite their role in the regulation of virulence, low to non-existent levels have been observed in infectious clinical isolates [66], correlating with the low AQ levels measured in our avirulent clinical strains. Furthermore, Davenport *et al.* demonstrated that roughly 1/3 of the *P. aeruginosa* metabolome is linked to quorum sensing, with primary and secondary metabolite levels affected by lack of quorum sensing signaling molecules [67]. Accordingly, Aqs are among the most relevant features of our random forest classification model that successfully differentiates between virulent and avirulent strains.

Recent research on the CF sputum microbiome and metabolome highlights the importance of Aqs, rhamnolipids and phenazines in the *in vivo* virulence in the human host and suggests a correlation of the abundance of these secondary metabolites and clinical disease severity [68, 69]. In line with findings by Quinn *et al.* the prominent AQ PQS was not among the most important metabolites associated with infection and virulence [68], whereas a rhamnolipid (Rha-Rha-C10-C10 in their case) was [69]. The fact that the same molecules we identified as virulence-associated in our cultivation-based approach or closely related ones are also connected to virulence in a clinical human setting further supports the validity of our findings.

Furthermore, our search for discriminatory markers of virulence and biofilm phenotypes pointed towards several features that could not be identified. Our random forest classification model included seven ‘unknowns’ in the group of the ten most important features. These were within a broad  $m/z$  range from 85 Da to 464 Da, suggesting that they belong to different metabolite classes (Figure 7.S17).

The most important feature in our model is an ‘unknown’ feature that clearly separates the two groups in our discovery data set and performs acceptably in our validation data set. Its putative identity is speculative; a sum formula of  $C_{12}H_{15}N_2$  for the cation points towards a phenazine-like structure, and the correlation with other features suggests a link to AQ biosynthesis. However, caution needs to be taken in the interpretation of unknowns in LC-MS metabolomics, as they may actually derive from other metabolites or represent artifacts [70, 71] and AQ and phenazine levels are of course not independent of each other due to multiple direct and indirect regulatory effects [23]. The fact that an unknown *P. aeruginosa* metabolite is a putative virulence biomarker demonstrates the value of studying the secondary metabolites of pathogenic bacteria, including those produced by highly studied species. Further efforts are needed to elucidate the structure of the feature in question and study its role in pseudomonal virulence.

As with any *in vitro* model, the work presented here comes with limitations. Quinn *et al.* have shown that significant differences exist between cultured *P. aeruginosa* strains and those in the host environment [68]. Important factors such as interspecies relations [72], and the duration of infection before the sampling and isolation of the strains [73] were outside the scope of this study. The virulence assessment in *G. melonella* presents an additional limitation as it might not fully reflect virulence properties in the human host. Differences in metabolite diversity as measured by the Shannon index have been observed, but the current dataset is insufficient to conclude whether and how strain-specific metabolite diversity is related to the virulence phenotype; this aspect remains to be investigated in future studies. However, the combination of known metabolites and unannotated features in a random forest classification model achieves an area under the ROC curve of  $> 0.8$  for the validation data set, achieving a good discrimination of virulent and avirulent *P. aeruginosa* strains. The model is, at present, not suitable for virulence prediction of strains with different biofilm morphologies. The inference of virulence phenotypes of *P. aeruginosa* clinical strains cannot be achieved from genomics data alone [14], and are difficult to construe from transcriptomics data [18, 19]. Thus, the strength of LC-MS metabolomics in classifying *P. aeruginosa* strains is a logical reflection of the high relevance of secondary metabolites for infection processes in this pathogen.

## 7.5. Conclusions

*P. aeruginosa* clinical strains with different virulence and biofilm phenotypes have different metabolic profiles. A large portion of these differences can be attributed to known virulence-associated secondary metabolites; however, structurally unidentified features are important separators and putative virulence biomarkers. Using machine learning on the complete metabolome dataset, we obtained a classification model that differentiates virulent and avirulent *P. aeruginosa* clinical strains with good accuracy (area under the ROC curve of  $> 0.8$ ). We have demonstrated that metabolomics can play an important role in the discovery of reliable virulence biomarkers or biomarker panels that are applicable in the clinic to gauge virulence and provide invaluable information on the potential clinical course of an infection.

## Author contributions

Conceptualization, T.D., S.H. and M.B.; methodology, T.D.; validation, T.D.; formal analysis, T.D., J.G.T. and A.K.; investigation, T.D., J.G.T. and A.K.; resources, S.H. and M.B.; data curation, T.D.; writing—original draft preparation, T.D.; writing—review and editing, J.G.T., S.H. and M.B.; visualization, T.D.; supervision, S.H. and M.B.; project administration, T.D., S.H. and M.B.; funding acquisition, T.D. and M.B. All authors have read and agreed to the published version of the manuscript.

## Funding

This research was funded by the President’s Initiative and Networking Funds of the Helmholtz Association of German Research Centres [VH-GS-202] and by EMBRIC, the EU-funded European Marine Biological Research Infrastructure Cluster [654008]. T.D. has been supported by a PhD scholarship of the *Studienstiftung des deutschen Volkes*.

## Acknowledgements

The authors wish to thank Ariane Khaledi for assistance with strain handling, Raimo Franke for fruitful discussions throughout the project and Christopher Kesthely for proofreading the manuscript.

## Conflicts of interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.



## Abbreviations

AQ	alkyl quinolone
AUC	area under the curve
CF	cystic fibrosis
DHQ	2,4-dihydroxyquinoline
ESI-QTOF-MS	electrospray ionisation quadrupole time-of-flight mass spectrometry
GNPS	Global Natural Product Social Molecular Networking
HSL	homoserine lactone
LC-MS	liquid chromatography–mass spectrometry
MASST	Mass Spectrometry Search Tool
$m/z$	mass-to-charge ratio
nd	not determined
nrpg	normalized reads per gene
OD <sub>600</sub>	optical density at 600 nm
$p_{\text{adj}}$	adjusted $p$ -value
PCA	principal component analysis
PE	phosphatidylethanolamine
PERMANOVA	Permutational multivariate analysis of variation
PQS	<i>Pseudomonas</i> quinolone signal
QNO	quinoline- $N$ -oxide
Rha	rhamnose, rhamnosyl
ROC	receiver operating characteristics
VIP	variable importance in projection

## References

- [1] M.-V. Grosso-Becerra, C. Santos-Medellín, A. González-Valdez, J.-L. Méndez, G. Delgado, R. Morales-Espinosa, L. Servín-González, L.-D. Alcaraz, and G. Soberón-Chávez. “*Pseudomonas aeruginosa* clinical and environmental isolates constitute a single population with high phenotypic diversity”. In: *BMC Genomics* 15.1 (2014), p. 318. DOI: 10.1186/1471-2164-15-318.
- [2] J. Gross, I. J. Passmore, J. C. S. Chung, O. Rzhepishevskaya, M. Ramstedt, and M. Welch. “Universal soldier: *Pseudomonas aeruginosa* — an opportunistic generalist”. In: *Frontiers in Biology* 8.4 (July 2013), pp. 387–394. DOI: 10.1007/s11515-013-1267-x.
- [3] F. Stapleton, J. K. G. Dart, D. V. Seal, and M. Matheson. “Epidemiology of *Pseudomonas aeruginosa* keratitis in contact lens wearers”. In: *Epidemiology and Infection* 114.3 (June 1995), pp. 395–402. DOI: 10.1017/s0950268800052109.

- [4] E. W. Wang, J. Y. Jung, M. E. Pashia, R. Nason, S. Scholnick, and R. A. Chole. "Otopathogenic *Pseudomonas aeruginosa* Strains as Competent Biofilm Formers". In: *Archives of Otolaryngology-Head & Neck Surgery* 131.11 (Nov. 2005), p. 983. DOI: 10.1001/archoto1.131.11.983.
- [5] J. B. Lyczak, C. L. Cannon, and G. B. Pier. "Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist". In: *Microbes and infection* 2.9 (2000), pp. 1051–1060.
- [6] J. A. Driscoll, S. L. Brody, and M. H. Kollef. "The Epidemiology, Pathogenesis and Treatment of *Pseudomonas aeruginosa* Infections". In: *Drugs* 67.3 (2007), pp. 351–368. DOI: 10.2165/00003495-200767030-00003.
- [7] S. de Bentzmann and P. Plésiat. "The *Pseudomonas aeruginosa* opportunistic pathogen and human infections". In: *Environmental Microbiology* 13.7 (Mar. 2011), pp. 1655–1665. DOI: 10.1111/j.1462-2920.2011.02469.x.
- [8] J. Chastre and J.-Y. Fagon. "Ventilator-associated Pneumonia". In: *American Journal of Respiratory and Critical Care Medicine* 165.7 (Apr. 2002), pp. 867–903. DOI: 10.1164/ajrccm.165.7.2105078.
- [9] H. Lund-Palau, A. R. Turnbull, A. Bush, E. Bardin, L. Cameron, O. Soren, N. Wierre-Gore, E. W. F. W. Alton, J. G. Bundy, G. Connett, S. N. Faust, A. Filloux, P. Freemont, A. Jones, V. Khoo, S. Morales, R. Murphy, R. Pabary, A. Simbo, S. Schelenz, Z. Takats, J. Webb, H. D. Williams, and J. C. Davies. "*Pseudomonas aeruginosa* infection in cystic fibrosis: pathophysiological mechanisms and therapeutic approaches". In: *Expert Review of Respiratory Medicine* 10.6 (May 2016), pp. 685–697. DOI: 10.1080/17476348.2016.1177460.
- [10] P. Cornelis and J. Dingemans. "*Pseudomonas aeruginosa* adapts its iron uptake strategies in function of the type of infections". In: *Frontiers in Cellular and Infection Microbiology* 3 (2013). DOI: 10.3389/fcimb.2013.00075.
- [11] B. Tümmler, L. Wiehlmann, J. Klockgether, and N. Cramer. "Advances in understanding *Pseudomonas*". In: *F1000Prime Reports* 6 (Feb. 2014). DOI: 10.12703/p6-9.
- [12] M. C. Wolfgang, B. R. Kulasekara, X. Liang, D. Boyd, K. Wu, Q. Yang, C. G. Miyada, and S. Lory. "Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*". In: *Proceedings of the National Academy of Sciences* 100.14 (June 2003), pp. 8484–8489. DOI: 10.1073/pnas.0832438100.
- [13] R. S. Bradbury, L. F. Roddam, A. Merritt, D. W. Reid, and A. C. Champion. "Virulence gene distribution in clinical, nosocomial and environmental isolates of *Pseudomonas aeruginosa*". In: *Journal of Medical Microbiology* 59.8 (Aug. 2010), pp. 881–890. DOI: 10.1099/jmm.0.018283-0.
- [14] J. Klockgether, N. Miethke, P. Kubesch, Y.-S. Bohn, I. Brockhausen, N. Cramer, L. Eberl, J. Greipel, C. Herrmann, S. Herrmann, S. Horatzek, M. Lingner, L. Luciano, P. Salunkhe, D. Schomburg, M. Wehsling, L. Wiehlmann, C. F. Davenport, and B. Tümmler. "Intraclonal diversity of the *Pseudomonas aeruginosa* cystic fibrosis airway isolates TBCF10839 and TBCF121838: distinct signatures of transcriptome, proteome, metabolome, adherence and pathogenicity despite an almost identical genome sequence". In: *Environmental Microbiology* 15.1 (Aug. 2012), pp. 191–210. DOI: 10.1111/j.1462-2920.2012.02842.x.

- [15] J. G. Thöming, J. Tomasch, M. Preusse, M. Koska, N. Grahl, S. Pohl, S. D. Willger, V. Kaefer, M. Müsken, and S. Häussler. “Parallel evolutionary paths to produce more than one *Pseudomonas aeruginosa* biofilm phenotype”. In: *npj Biofilms and Microbiomes* 6.1 (Jan. 2020). DOI: 10.1038/s41522-019-0113-6.
- [16] J. Klockgether and B. Tümmler. “Recent advances in understanding *Pseudomonas aeruginosa* as a pathogen”. In: *F1000Research* 6 (July 2017), p. 1261. DOI: 10.12688/f1000research.10506.1.
- [17] R. Hilker, A. Munder, J. Klockgether, P. M. Losada, P. Chouvarine, N. Cramer, C. F. Davenport, S. Dethlefsen, S. Fischer, H. Peng, T. Schönfelder, O. Türk, L. Wiehlmann, F. Wöbeling, E. Gulbins, A. Goesmann, and B. Tümmler. “Interclonal gradient of virulence in the *Pseudomonas aeruginosa* pangenome from disease and environment”. In: *Environmental Microbiology* 17.1 (Sept. 2014), pp. 29–46. DOI: 10.1111/1462-2920.12606.
- [18] P. Bielecki, U. Komor, A. Bielecka, M. Müsken, J. Puchałka, M. W. Pletz, M. Ballmann, V. A. M. dos Santos, S. Weiss, and S. Häussler. “Ex vivo transcriptional profiling reveals a common set of genes important for the adaptation of *Pseudomonas aeruginosa* to chronically infected host sites”. In: *Environmental Microbiology* 15.2 (Nov. 2012), pp. 570–587. DOI: 10.1111/1462-2920.12024.
- [19] A. M. Sousa, R. Monteiro, and M. O. Pereira. “Unveiling the early events of *Pseudomonas aeruginosa* adaptation in cystic fibrosis airway environment using a long-term in vitro maintenance”. In: *International Journal of Medical Microbiology* 308.8 (Dec. 2018), pp. 1053–1064. DOI: 10.1016/j.ijmm.2018.10.003.
- [20] G. J. Patti, O. Yanes, and G. Siuzdak. “Metabolomics: the apogee of the omics trilogy”. In: *Nature Reviews Molecular Cell Biology* 13.4 (Apr. 2012), pp. 263–269. DOI: 10.1038/nrm3314.
- [21] O. Fiehn. “Metabolomics – the link between genotypes and phenotypes”. In: *Plant Molecular Biology* 48.1/2 (2002), pp. 155–171. DOI: 10.1023/a:1013713905833.
- [22] C. Guijas, J. R. Montenegro-Burke, B. Warth, M. E. Spilker, and G. Siuzdak. “Metabolomics activity screening for identifying metabolites that modulate phenotype”. In: *Nature Biotechnology* 36.4 (Apr. 2018), pp. 316–320. DOI: 10.1038/nbt.4101.
- [23] P. N. Jimenez, G. Koch, J. A. Thompson, K. B. Xavier, R. H. Cool, and W. J. Quax. “The Multiple Signaling Systems Regulating Virulence in *Pseudomonas aeruginosa*”. In: *Microbiology and Molecular Biology Reviews* 76.1 (Mar. 2012), pp. 46–65. DOI: 10.1128/mubr.05007-11.
- [24] T. Depke, R. Franke, and M. Brönstrup. “Clustering of MS<sup>2</sup> spectra using unsupervised methods to aid the identification of secondary metabolites from *Pseudomonas aeruginosa*”. In: *Journal of Chromatography B* 1071 (Dec. 2017), pp. 19–28. DOI: 10.1016/j.jchromb.2017.06.002.
- [25] C. A. Ortori, J.-F. Dubern, S. R. Chhabra, M. Cámara, K. Hardie, P. Williams, and D. A. Barrett. “Simultaneous quantitative profiling of *N*-acyl-L-homoserine lactone and 2-alkyl-4(1*H*)-quinolone families of quorum-sensing signaling molecules using LC-MS/MS”. In: *Analytical and Bioanalytical Chemistry* 399.2 (Oct. 2010), pp. 839–850. DOI: 10.1007/s00216-010-4341-0.
- [26] G. Allegretta, C. K. Maurer, J. Eberhard, D. Maura, R. W. Hartmann, L. Rahme, and M. Empting. “In-depth Profiling of MvfR-Regulated Small Molecules in *Pseudomonas aeruginosa* after Quorum Sensing Inhibitor Treatment”. In: *Frontiers in Microbiology* 8 (May 2017). DOI: 10.3389/fmicb.2017.00924.

- [27] P. Gao and G. Xu. “Mass-spectrometry-based microbial metabolomics: recent developments and applications”. In: *Analytical and Bioanalytical Chemistry* 407.3 (Sept. 2014), pp. 669–680. DOI: 10.1007/s00216-014-8127-7.
- [28] C. M. Grim, G. T. Luu, and L. M. Sanchez. “Staring into the void: demystifying microbial metabolomics”. In: *FEMS Microbiology Letters* 366.11 (June 2019). DOI: 10.1093/femsle/fnz135.
- [29] J. G. Bundy, T. L. Willey, R. S. Castell, D. J. Ellar, and K. M. Brindle. “Discrimination of pathogenic clinical isolates and laboratory strains of *Bacillus cereus* by NMR-based metabolomic profiling”. In: *FEMS Microbiology Letters* 242.1 (Jan. 2005), pp. 127–136. DOI: 10.1016/j.femsle.2004.10.048.
- [30] R. t’Kindt, R. A. Scheltema, A. Jankevics, K. Brunker, S. Rijal, J.-C. Dujardin, R. Breitling, D. G. Watson, G. H. Coombs, and S. Decuypere. “Metabolomics to Unveil and Understand Phenotypic Diversity between Pathogen Populations”. In: *PLoS Neglected Tropical Diseases* 4.11 (Nov. 2010). Ed. by T. G. Geary, e904. DOI: 10.1371/journal.pntd.0000904.
- [31] M. H. M. Maifiah, S.-E. Cheah, M. D. Johnson, M.-L. Han, J. D. Boyce, V. Thamlikitkul, A. Forrest, K. S. Kaye, P. Hertzog, A. W. Purcell, J. Song, T. Velkov, D. J. Creek, and J. Li. “Global metabolic analyses identify key differences in metabolite levels between polymyxin-susceptible and polymyxin-resistant *Acinetobacter baumannii*”. In: *Scientific Reports* 6.1 (Feb. 2016). DOI: 10.1038/srep22287.
- [32] H. D. Bean, C. A. Rees, and J. E. Hill. “Comparative analysis of the volatile metabolomes of *Pseudomonas aeruginosa* clinical isolates”. In: *Journal of Breath Research* 10.4 (Nov. 2016), p. 047102. DOI: 10.1088/1752-7155/10/4/047102.
- [33] K. M. Jørgensen, O. Ciofu, S. Molin, T. Wassermann, H. K. Johansen, N. Høiby, and L. E. Christiansen. “Diversity of metabolic profiles of cystic fibrosis *Pseudomonas aeruginosa* during the early stages of lung infection”. In: *Microbiology* 161.7 (July 2015), pp. 1447–1462. DOI: 10.1099/mic.0.000093.
- [34] A. Kordes, N. Grahl, M. Koska, M. Preusse, A. Arce-Rodriguez, W.-R. Abraham, V. Kaever, and S. Häussler. “Establishment of an induced memory response in *Pseudomonas aeruginosa* during infection of a eukaryotic host”. In: *The ISME Journal* 13.8 (Apr. 2019), pp. 2018–2030. DOI: 10.1038/s41396-019-0412-1.
- [35] K. Hornischer, A. Khaledi, S. Pohl, M. Schniederjans, L. Pezoldt, F. Casilag, U. Muthukumarasamy, S. Bruchmann, J. Thöming, A. Kordes, and S. Häussler. “BACTOME—a reference database to explore the sequence- and gene expression-variation landscape of *Pseudomonas aeruginosa* clinical isolates”. In: *Nucleic Acids Research* 47.D1 (Oct. 2018), pp. D716–D720. DOI: 10.1093/nar/gky895.
- [36] A. Khaledi, A. Weimann, M. Schniederjans, E. Asgari, T.-H. Kuo, A. Oliver, G. Cabot, A. Kola, P. Gastmeier, M. Hogardt, D. Jonas, M. R. Mofrad, A. Bremges, A. C. McHardy, and S. Häussler. “Predicting antimicrobial resistance in *Pseudomonas aeruginosa* with machine learning-enabled molecular diagnostics”. In: *EMBO Molecular Medicine* 12.3 (Feb. 2020). DOI: 10.15252/emmm.201910264.
- [37] J. Erdmann, J. G. Thöming, S. Pohl, A. Pich, C. Lenz, and S. Häussler. “The Core Proteome of Biofilm-Grown Clinical *Pseudomonas aeruginosa* Isolates”. In: *Cells* 8.10 (Sept. 2019), p. 1129. DOI: 10.3390/cells8101129.
- [38] G. Lunter and M. Goodson. “Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads”. In: *Genome Research* 21.6 (Oct. 2010), pp. 936–939. DOI: 10.1101/gr.111120.110.

- [39] M. I. Love, W. Huber, and S. Anders. “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2”. In: *Genome Biology* 15.12 (Dec. 2014). DOI: 10.1186/s13059-014-0550-8.
- [40] K. Mathee, G. Narasimhan, C. Valdes, X. Qiu, J. M. Matewish, M. Koehrsen, A. Rokas, C. N. Yandava, R. Engels, E. Zeng, R. Olavarietta, M. Doud, R. S. Smith, P. Montgomery, J. R. White, P. A. Godfrey, C. Kodira, B. Birren, J. E. Galagan, and S. Lory. “Dynamics of *Pseudomonas aeruginosa* genome evolution”. In: *Proceedings of the National Academy of Sciences* 105.8 (Feb. 2008), pp. 3100–3105. DOI: 10.1073/pnas.0711982105.
- [41] F. Witzgall, T. Depke, M. Hoffmann, M. Empting, M. Brönstrup, R. Müller, and W. Blankenfeldt. “The Alkylquinolone Repertoire of *Pseudomonas aeruginosa* is Linked to Structural Flexibility of the FabH-like 2-Heptyl-3-hydroxy-4(1*H*)-quinolone (PQS) Biosynthesis Enzyme PqsBC”. In: *ChemBioChem* 19.14 (May 2018), pp. 1531–1544. DOI: 10.1002/cbic.201800153.
- [42] R. Tautenhahn, G. J. Patti, D. Rinehart, and G. Siuzdak. “XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data”. In: *Analytical Chemistry* 84.11 (June 2012), pp. 5035–5039. DOI: 10.1021/ac300698c.
- [43] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL: <https://www.R-project.org/>.
- [44] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc. Boston, MA, 2015. URL: <http://www.rstudio.com/>.
- [45] H. Wickham, M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Golemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani. “Welcome to the tidyverse”. In: *Journal of Open Source Software* 4.43 (2019), p. 1686. DOI: 10.21105/joss.01686.
- [46] C. Kuhl, R. Tautenhahn, C. Böttcher, T. R. Larson, and S. Neumann. “CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets”. In: *Analytical Chemistry* 84.1 (Dec. 2011), pp. 283–289. DOI: 10.1021/ac202450g.
- [47] N. G. Mahieu, X. Huang, Y.-J. Chen, and G. J. Patti. “Credentialing Features: A Platform to Benchmark and Optimize Untargeted Metabolomic Methods”. In: *Analytical Chemistry* 86.19 (Sept. 2014), pp. 9583–9589. DOI: 10.1021/ac503092d.
- [48] W. Scherber. “Stable Isotope Labeling to Improve Metabolite Identification in Untargeted Metabolomics of Pathogenic Bacteria”. Master’s Thesis. Hochschule Aalen – Technik und Wirtschaft, 2020.
- [49] M. Wang, A. K. Jarmusch, F. Vargas, A. A. Aksenov, J. M. Gauglitz, K. Weldon, D. Petras, R. da Silva, R. Quinn, A. V. Melnik, J. J. J. van der Hooft, A. M. Caraballo-Rodríguez, L. F. Nothias, C. M. Aceves, M. Panitchpakdi, E. Brown, F. D. Ottavio, N. Sikora, E. O. Elijah, L. Labarta-Bajo, E. C. Gentry, S. Shalapur, K. E. Kyle, S. P. Puckett, J. D. Watrous, C. S. Carpenter, A. Bouslimani, M. Ernst, A. D. Swafford, E. I. Zú niga, M. J. Balunas, J. L. Klassen, R. Loomba, R. Knight, N. Bandeira, and P. C. Dorrestein. “Mass spectrometry searches using MASST”. In: *Nature Biotechnology* 38.1 (Jan. 2020), pp. 23–26. DOI: 10.1038/s41587-019-0375-9.
- [50] K. Haug, K. Cochrane, V. C. Nainala, M. Williams, J. Chang, K. V. Jayaseelan, and C. O’Donovan. “MetaboLights: a resource evolving in response to the needs of its scientific community”. In: *Nucleic Acids Research* (Nov. 2019). DOI: 10.1093/nar/gkz1019.

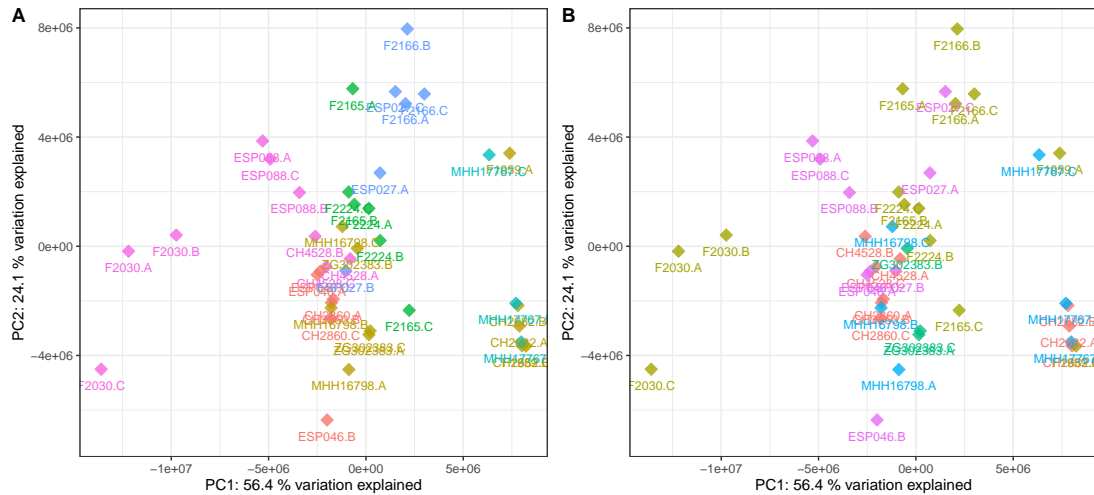
- [51] J. Oksanen, F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. *vegan: Community Ecology Package*. R package version 2.5-6. 2019. URL: <https://CRAN.R-project.org/package=vegan>.
- [52] A. Liaw and M. Wiener. "Classification and Regression by randomForest". In: *R News* 2.3 (2002), pp. 18–22. URL: <https://CRAN.R-project.org/doc/Rnews/>.
- [53] L. Breiman. "Random Forests". In: *Machine Learning* 45.1 (2001), pp. 5–32. DOI: 10.1023/a:1010933404324.
- [54] T. Sing, O. Sander, N. Beerenwinkel, and T. Lengauer. "ROCR: visualizing classifier performance in R". In: *Bioinformatics* 21.20 (Aug. 2005), pp. 3940–3941. DOI: 10.1093/bioinformatics/bti623.
- [55] D. V. Mavrodi, W. Blankenfeldt, and L. S. Thomashow. "Phenazine Compounds in Fluorescent *Pseudomonas* Spp.: Biosynthesis and Regulation". In: *Annual Review of Phytopathology* 44.1 (Sept. 2006), pp. 417–445. DOI: 10.1146/annurev.phyto.44.013106.145710.
- [56] N. Guttenberger, W. Blankenfeldt, and R. Breinbauer. "Recent developments in the isolation, biological function, biosynthesis, and synthesis of phenazine natural products". In: *Bioorganic & Medicinal Chemistry* 25.22 (Nov. 2017), pp. 6149–6166. DOI: 10.1016/j.bmc.2017.01.002.
- [57] B. R. Boles, M. Thoendel, and P. K. Singh. "Rhamnolipids mediate detachment of *Pseudomonas aeruginosa* from biofilms". In: *Molecular Microbiology* 57.5 (July 2005), pp. 1210–1223. DOI: 10.1111/j.1365-2958.2005.04743.x.
- [58] L. Zulianello, C. Canard, T. Kohler, D. Caille, J.-S. Lacroix, and P. Meda. "Rhamnolipids Are Virulence Factors That Promote Early Infiltration of Primary Human Airway Epithelia by *Pseudomonas aeruginosa*". In: *Infection and Immunity* 74.6 (May 2006), pp. 3134–3147. DOI: 10.1128/iai.01772-05.
- [59] M. W. Calfee, J. G. Shelton, J. A. McCubrey, and E. C. Pesci. "Solubility and Bioactivity of the *Pseudomonas* Quinolone Signal Are Increased by a *Pseudomonas aeruginosa*-Produced Surfactant". In: *Infection and Immunity* 73.2 (Feb. 2005), pp. 878–882. DOI: 10.1128/iai.73.2.878-882.2005.
- [60] K. Schlegel, K. Taraz, and H. Budzikiewicz. "The stereoisomers of pyochelin, a siderophore of *Pseudomonas aeruginosa*". In: *BioMetals* 17.4 (Aug. 2004), pp. 409–414. DOI: 10.1023/b:biom.0000029437.42633.73.
- [61] G. C. Palmer, K. L. Palmer, P. A. Jorth, and M. Whiteley. "Characterization of the *Pseudomonas aeruginosa* Transcriptional Response to Phenylalanine and Tyrosine". In: *Journal of Bacteriology* 192.11 (Mar. 2010), pp. 2722–2728. DOI: 10.1128/jb.00112-10.
- [62] L. Yi, N. Dong, Y. Yun, B. Deng, D. Ren, S. Liu, and Y. Liang. "Chemometric methods in data processing of mass spectrometry-based metabolomics: A review". In: *Analytica Chimica Acta* 914 (Mar. 2016), pp. 17–34. DOI: 10.1016/j.aca.2016.02.001.
- [63] T. Chen, Y. Cao, Y. Zhang, J. Liu, Y. Bao, C. Wang, W. Jia, and A. Zhao. "Random Forest in Clinical Metabolomics for Phenotypic Discrimination and Biomarker Selection". In: *Evidence-Based Complementary and Alternative Medicine* 2013 (2013), pp. 1–11. DOI: 10.1155/2013/298183.
- [64] E. Frimmersdorf, S. Horatzek, A. Pelnikevich, L. Wiehlmann, and D. Schomburg. "How *Pseudomonas aeruginosa* adapts to various environments: a metabolomic approach". In: *Environmental Microbiology* 12.6 (Feb. 2010), pp. 1734–1747. DOI: 10.1111/j.1462-2920.2010.02253.x.



- [65] E. E. Bardin, S. J. S. Cameron, A. Perdones-Montero, K. Hardiman, F. Bolt, E. W. F. W. Alton, A. Bush, J. C. Davies, and Z. Takáts. “Metabolic Phenotyping and Strain Characterisation of *Pseudomonas aeruginosa* Isolates from Cystic Fibrosis Patients Using Rapid Evaporative Ionisation Mass Spectrometry”. In: *Scientific Reports* 8.1 (July 2018). DOI: 10.1038/s41598-018-28665-7.
- [66] J. A. Schaber, N. L. Carty, N. A. McDonald, E. D. Graham, R. Cheluvappa, J. A. Griswold, and A. N. Hamood. “Analysis of quorum sensing-deficient clinical isolates of *Pseudomonas aeruginosa*”. In: *Journal of Medical Microbiology* 53.9 (Sept. 2004), pp. 841–853. DOI: 10.1099/jmm.0.45617-0.
- [67] P. W. Davenport, J. L. Griffin, and M. Welch. “Quorum Sensing Is Accompanied by Global Metabolic Changes in the Opportunistic Human Pathogen *Pseudomonas aeruginosa*”. In: *Journal of Bacteriology* 197.12 (Apr. 2015). Ed. by G. A. O’Toole, pp. 2072–2082. DOI: 10.1128/jb.02557-14.
- [68] R. A. Quinn, V. V. Phelan, K. L. Whiteson, N. Garg, B. A. Bailey, Y. W. Lim, D. J. Conrad, P. C. Dorrestein, and F. L. Rohwer. “Microbial, host and xenobiotic diversity in the cystic fibrosis sputum metabolome”. In: *The ISME Journal* 10.6 (Dec. 2015), pp. 1483–1498. DOI: 10.1038/ismej.2015.207.
- [69] R. A. Quinn, S. Adem, R. H. Mills, W. Comstock, L. D. Goldasich, G. Humphrey, A. A. Aksenov, A. V. Melnik, R. da Silva, G. Ackermann, N. Bandeira, D. J. Gonzalez, D. Conrad, A. J. O’Donoghue, R. Knight, and P. C. Dorrestein. “Neutrophilic proteolysis in the cystic fibrosis lung correlates with a pathogenic microbiome”. In: *Microbiome* 7.1 (Feb. 2019). DOI: 10.1186/s40168-019-0636-3.
- [70] R. Chaleckis, I. Meister, P. Zhang, and C. E. Wheelock. “Challenges, progress and promises of metabolite annotation for LC–MS-based metabolomics”. In: *Current Opinion in Biotechnology* 55 (Feb. 2019), pp. 44–50. DOI: 10.1016/j.copbio.2018.07.010.
- [71] Y.-F. Xu, W. Lu, and J. D. Rabinowitz. “Avoiding Misannotation of In-Source Fragmentation Products as Cellular Metabolites in Liquid Chromatography–Mass Spectrometry–Based Metabolomics”. In: *Analytical Chemistry* 87.4 (Jan. 2015), pp. 2273–2281. DOI: 10.1021/ac504118y.
- [72] B. Gao, T. Gallagher, Y. Zhang, M. Elbadawi-Sidhu, Z. Lai, O. Fiehn, and K. L. Whiteson. “Tracking Polymicrobial Metabolism in Cystic Fibrosis Airways: *Pseudomonas aeruginosa* Metabolism and Physiology Are Influenced by *Rothia mucilaginosa*-Derived Metabolites”. In: *mSphere* 3.2 (Apr. 2018). Ed. by K. McMahon. DOI: 10.1128/mSphere.00151-18.
- [73] A. Sousa and M. Pereira. “*Pseudomonas aeruginosa* Diversification during Infection Development in Cystic Fibrosis Lungs—A Review”. In: *Pathogens* 3.3 (Aug. 2014), pp. 680–703. DOI: 10.3390/pathogens3030680.

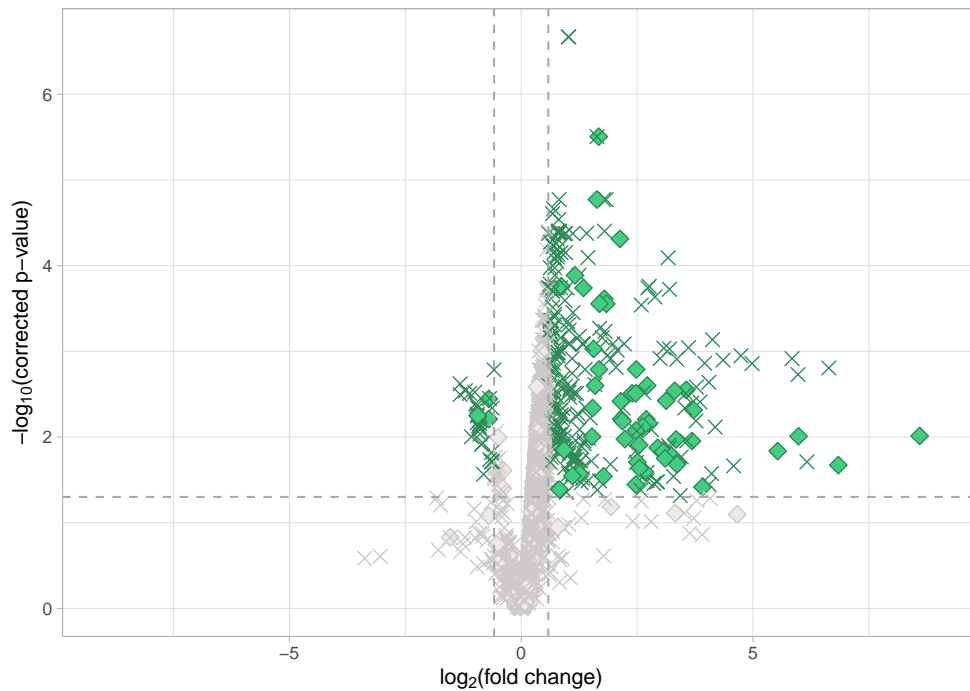
## Supporting Information

### Supplementary Figures

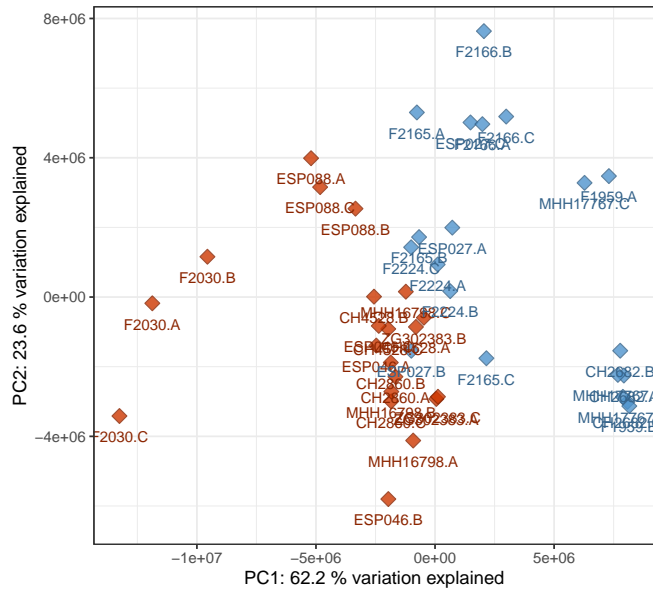


**Figure 7.S1.:** PCA scores plot of the discovery data set. Data points are coloured according to A) timepoint of harvest, i.e. duration of cultivation (red, 4 h; brown, 4.5 h; green, 5 h; cyan, 5.5 h; blue, 6 h; pink, 6.5 h), and B) geographical origin of the sample (red, Berlin (Germany); brown, Frankfurt am Main (Germany); green, Görlitz (Germany); blue, Hannover (Germany); pink, Palma de Mallorca (Spain)). No separation or grouping according to timepoint of harvest or geographical origin of the sample can be observed.

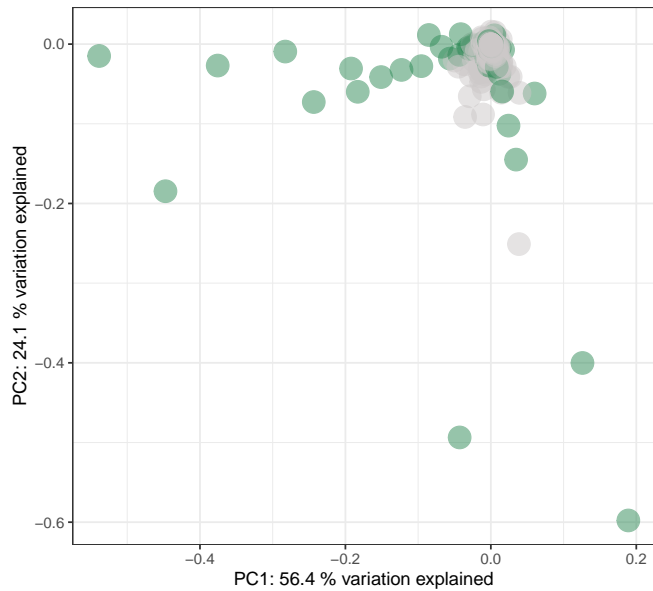




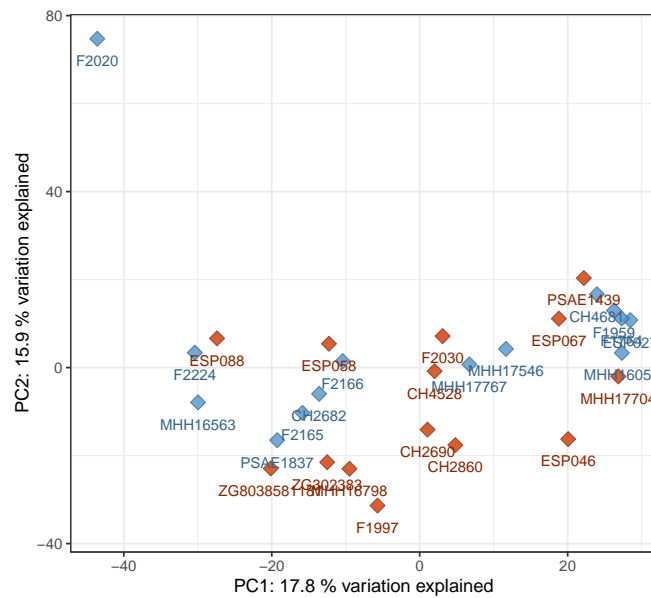
**Figure 7.S2.:** Volcano plot of the discovery data set. All features were plotted with the binary logarithm of their (non-directional) fold change on the x-axis and the negative decadic logarithm of their corrected  $p$ -value on the y-axis. Thresholds for significantly differentially abundant features are indicated by dashed lines (fold change  $\geq 1.5$ , corrected  $p$ -value  $\leq 0.05$ ) and data points were colour coded according to these thresholds (green – significantly differentially abundant features, grey – other features). Diamonds signify identified features whereas unknowns are indicated by crosses. It is apparent that the majority of significantly differentially abundant features have higher levels in the virulent cluster A group which is consistent with the high number of virulence-associated secondary metabolites in the data set.



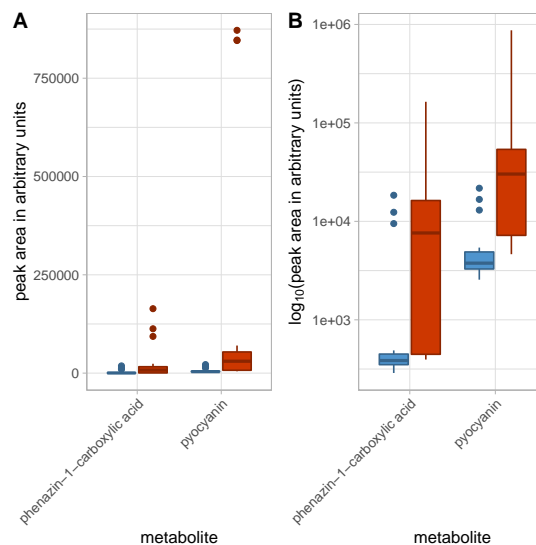
**Figure 7.S3.:** PCA scores plot with only annotated features considered in the analysis. The plot was generated analogously to Figure 1 in the main text. It suggests that the overall group separation is maintained if unknown features are ignored, indicating that the main drivers of separation or features correlated to them have been annotated. Red – virulent cluster A strains, blue – avirulent cluster B strains.



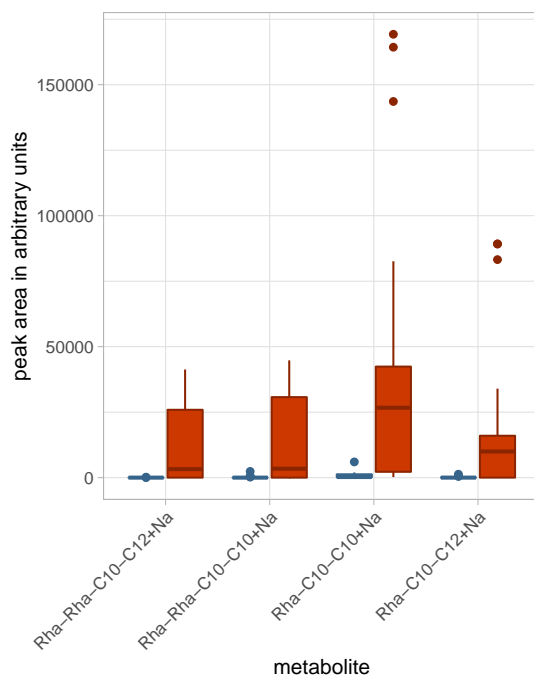
**Figure 7.S4.:** PCA loadings plot of the discovery data set. Green points symbolize annotated features and grey points features that could not be annotated. Most features with high loadings, i.e. a strong contribution to the first two principal components, have been annotated.



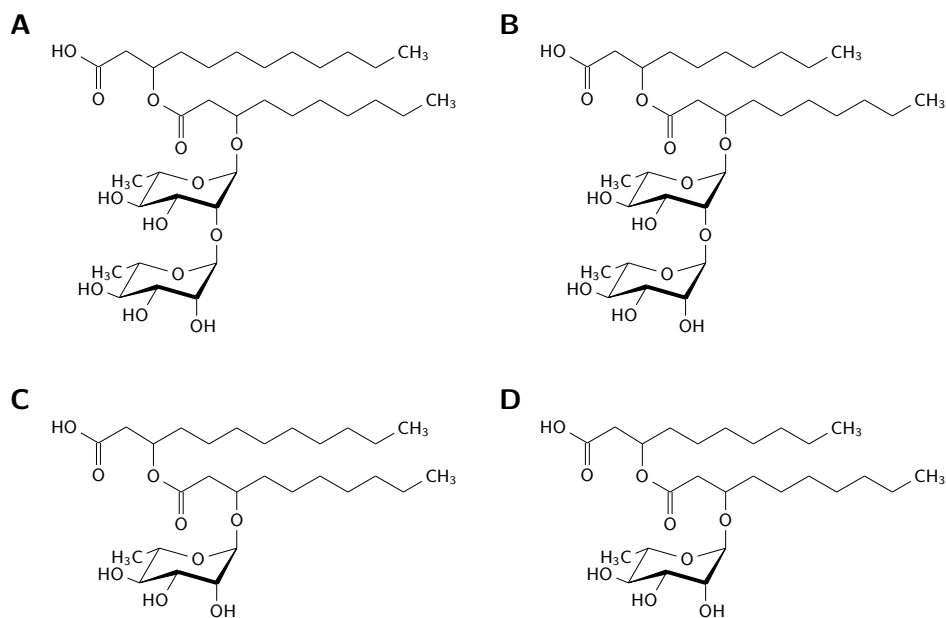
**Figure 7.S5.:** Transcriptional profiles reveal no gene expression pattern associated with the virulence phenotype. The principal component analysis (PCA) plot of transcriptional profiles recorded for clinical isolates grown under planktonic conditions does not cluster according to the observed *in vivo* virulence phenotype in the *G. mellonella* infection model. Each data point represents the transcriptional profile of an individual clinical isolate. Red – virulent cluster A strains, blue – avirulent cluster B strains.



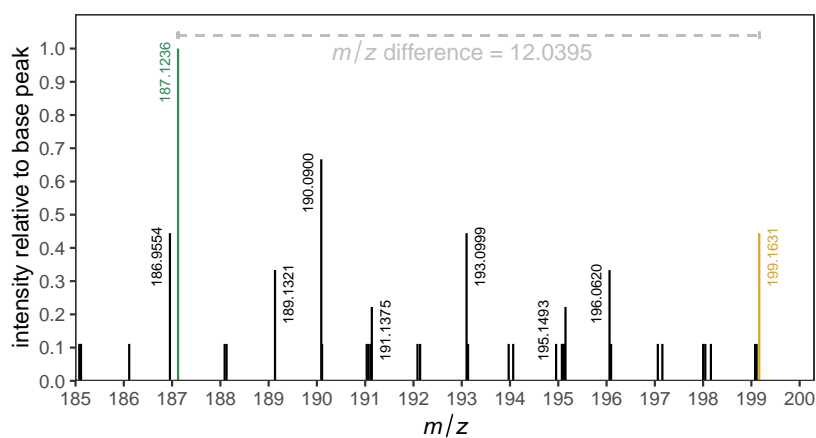
**Figure 7.S6.:** Levels of the two phenazines pyocyanin and phenazine-1-carboxylic acid, in the different strains of the discovery data set. Box plots of the peak areas in arbitrary units in the two phenotypic groups (A). A section of the y-scale with logarithmic scaling is shown to better visualize group differences (B). Both phenazine-1-carboxylic acid and pyocyanin have higher levels in the virulent strains, although there is significant overlap and both groups harbor one high producer strain each. Red – virulent cluster A strains, blue – avirulent cluster B strains.



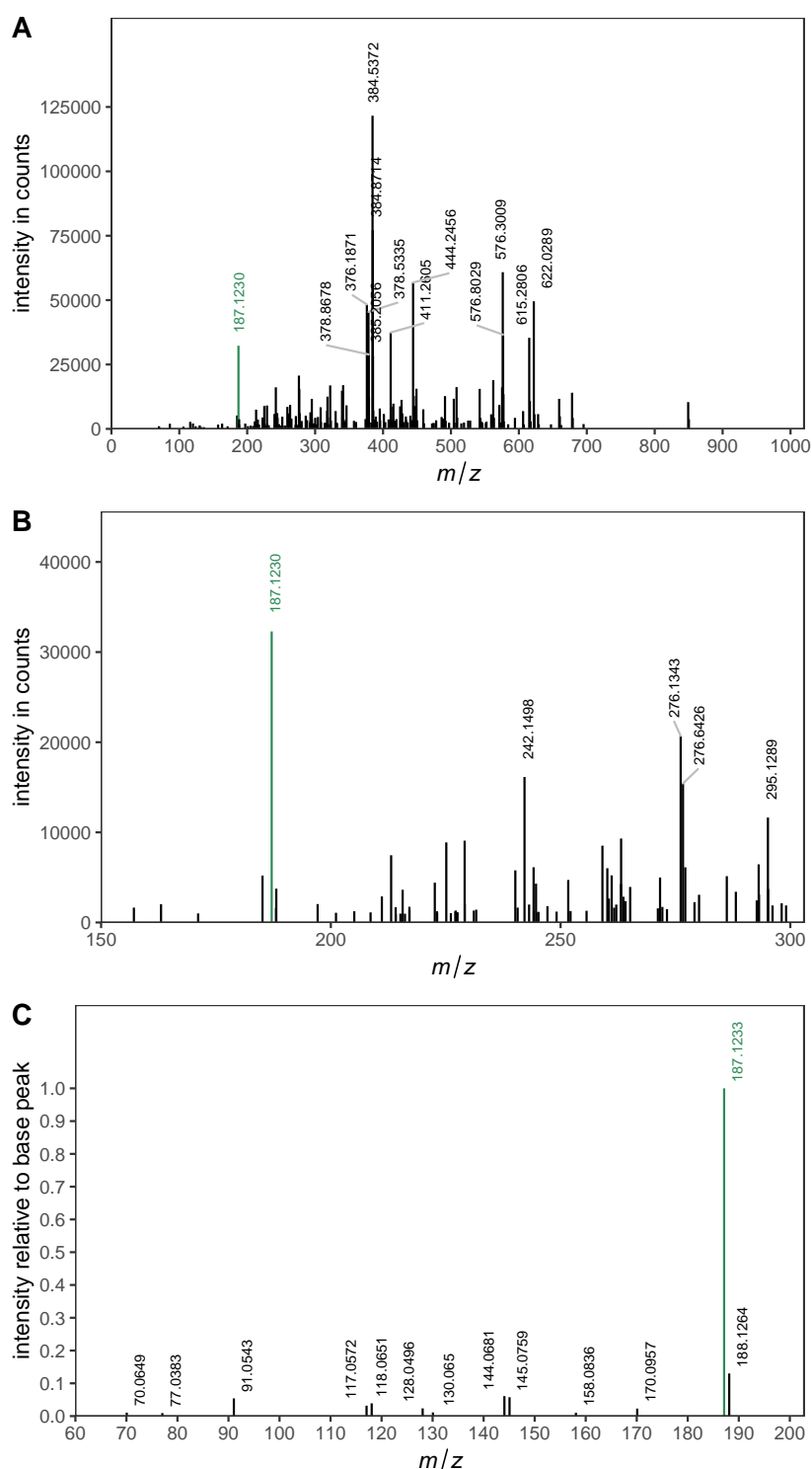
**Figure 7.S7.:** Rhamnolipid levels in the different strains of the discovery data set. Box plots of the peak areas of four annotated rhamnolipids in arbitrary units in the two phenotypic groups. Red – virulent cluster A strains, blue – avirulent cluster B strains.



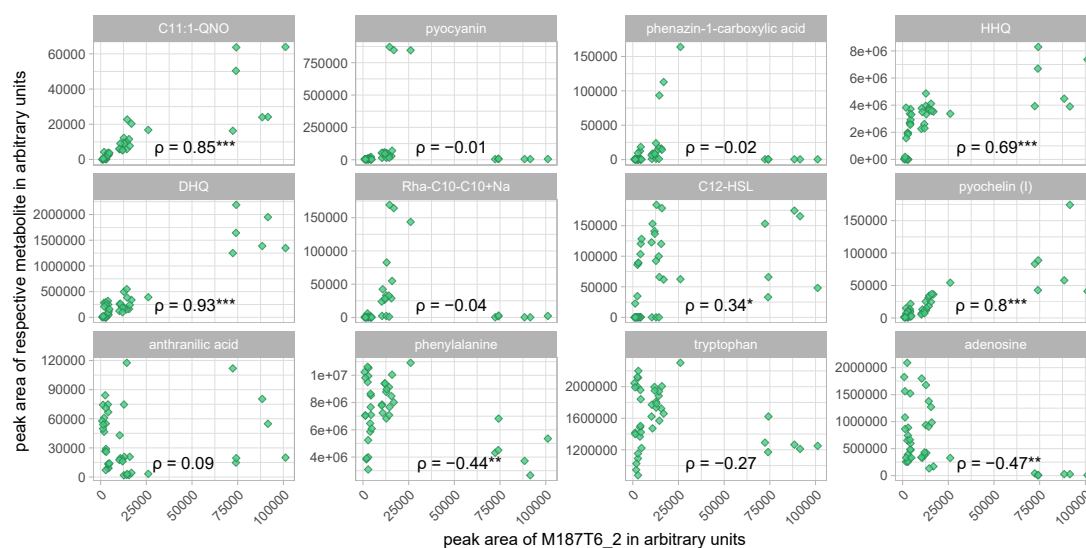
**Figure 7.S8.:** Structures of annotated rhamnolipids (cf. Figure 7.S7). A: Rha-Rha-C10-C12. B: Rha-Rha-C10-C10. C: Rha-C10-C12. D: Rha-C10-C10.



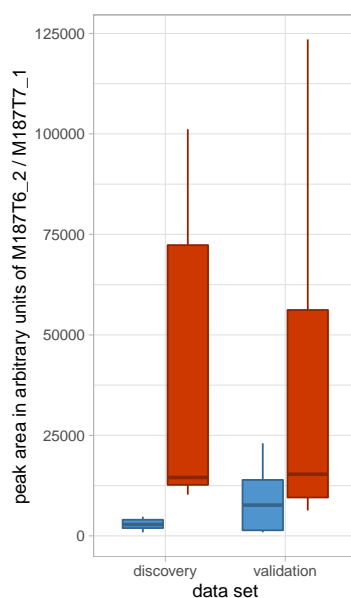
**Figure 7.S9.:** Credentialed peak pair of M187T6\_2 and its  $^{13}\text{C}$ -labeled derivative. Magnified section of a full scan MS spectrum with the peak corresponding to M187T6\_2 marked in green and the one corresponding to its  $^{13}\text{C}$ -labeled derivative marked in yellow. The measured  $m/z$  difference of 12.0395 is in accordance with the expected  $m/z$  difference between  $\text{C}_{12}\text{H}_{15}\text{N}_2^+$  (the assumed formula of M187T6\_2) and  $^{13}\text{C}_{12}\text{H}_{15}\text{N}_2^+$  (theoretical  $m/z$  difference 12.0403). The incorporation of stable isotope labeled carbon from  $^{13}\text{C}_6$ -glucose in the growth medium provides evidence that the feature is a metabolite of biological origin rather than an artifact. The figure was adapted from [1].



**Figure 7.S10.:** Full scan and MS<sup>2</sup> spectrum of the feature M187T6\_2 in the discovery data set. A: Full scan MS spectrum of the full  $m/z$  range from 0 to 1000. B: Same as A magnified to the relevant  $m/z$  range from 150 to 300. C: MS<sup>2</sup> spectrum of the 187.123 ion of M187T6\_2. Peaks for the 187.123 ion are marked in green. The M187T6\_2 feature displays a low abundance and its MS<sup>2</sup> spectrum is rather uninformative as the ion hardly shows fragmentation.

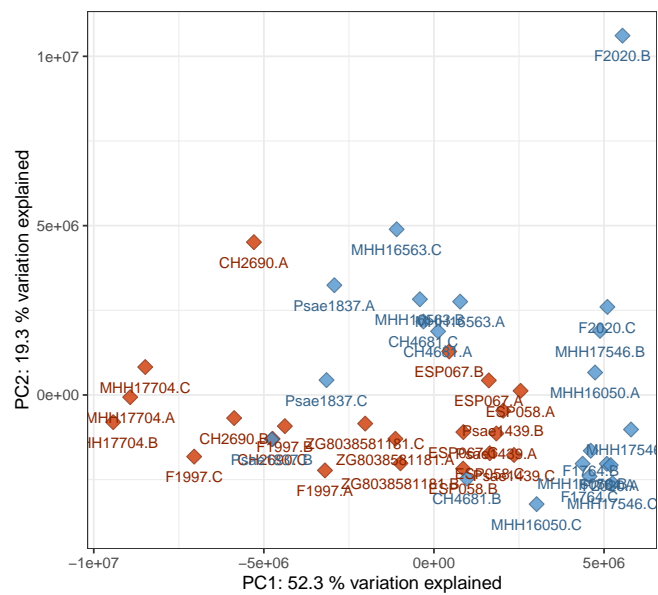


**Figure 7.S11.:** Pearson's correlation of selected feature intensities to those of M187T6\_2. The peak area in arbitrary units of the feature M187T6\_2 is plotted on the x-axis and the peak area of the respective metabolite in the sub-diagram title on the y-axis. Each data point corresponds to a biological replicate of a strain in the discovery data set. Pearson's correlation coefficient between the two respective peak areas is inserted as text in each sub-diagram with asterisks denoting statistical significance of the correlation ( $^{***}$ ,  $p$ -value  $\leq 0.001$ ;  $^{**}$ ,  $p$ -value  $\leq 0.01$ ;  $^*$ ,  $p$ -value  $\leq 0.05$ ; no asterisk,  $p$ -value  $> 0.05$ ). Significant and strongly positive correlations can be found with Aqs and the related DHQ as well as with pyochelin.

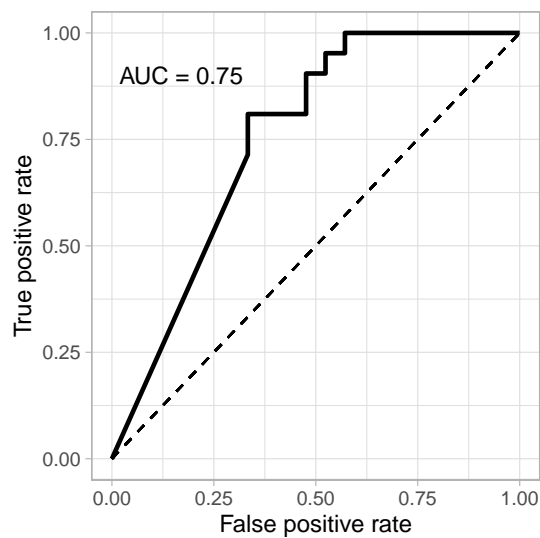


**Figure 7.S12.:** Boxplots of feature intensities for M187T6\_2 in the discovery and validation data set. Due to automatic naming of the features during preprocessing by XCMS online, the respective feature has the identifier M187T7\_1 in the validation data set. Peak areas in arbitrary units are used as a metric for the metabolite levels. While M187T6\_2 is a perfect separator in the discovery data set, there is some overlap in the validation data set, i. e. the highest levels in the avirulent group exceed the lowest levels in the virulent group. The abundances in virulent cluster A and avirulent cluster B strains are significantly different. Red – virulent cluster A strains, blue – avirulent cluster B strains.

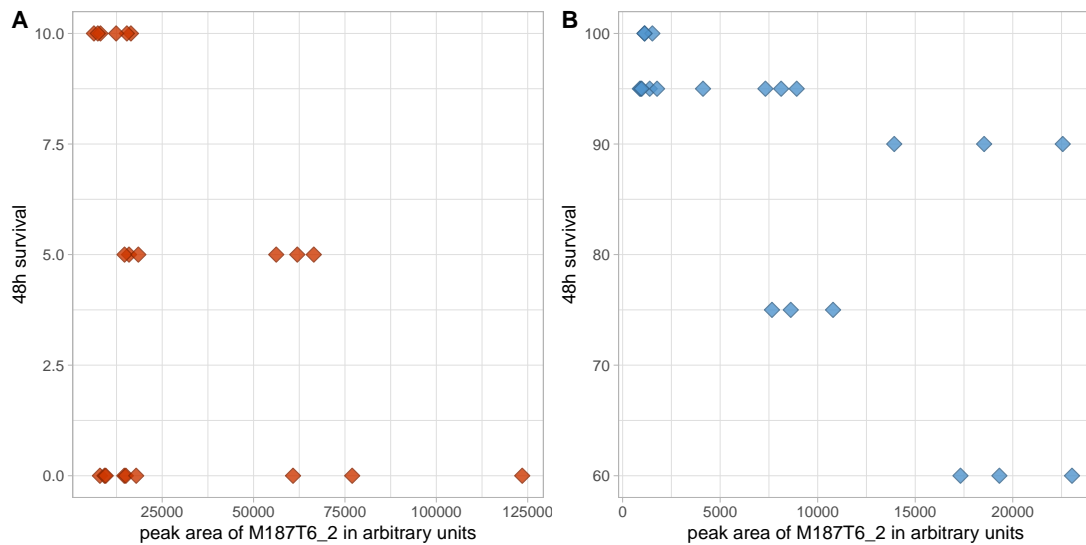




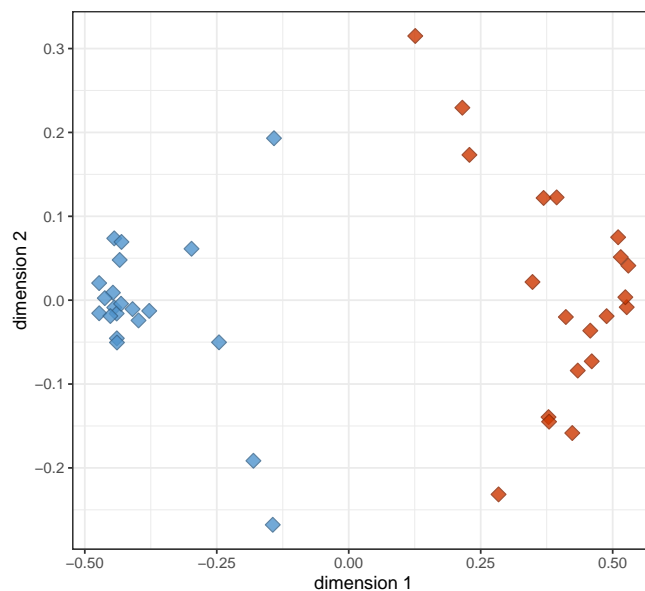
**Figure 7.S13.:** PCA scores plot of the validation data set. The plot was generated analogously to Figure 1 in the main text. Group separation appears to be weaker in the validation data set but is still possible. Red – virulent cluster A strains, blue – avirulent cluster B strains.



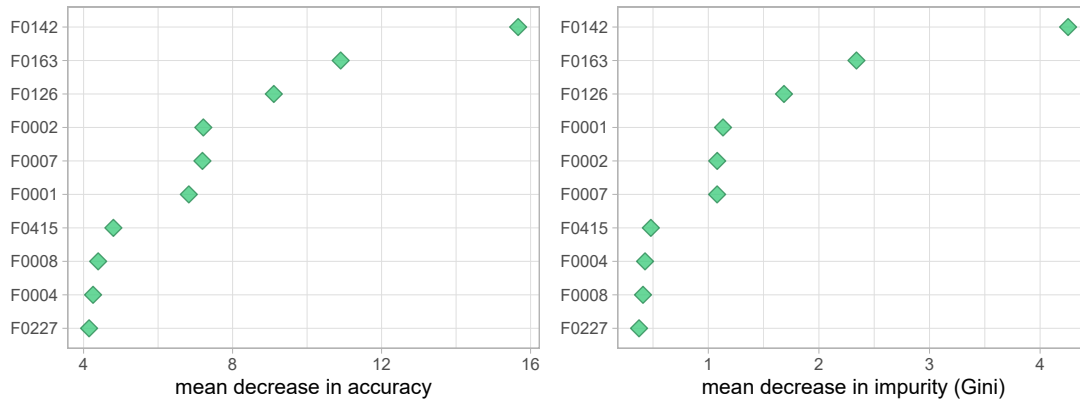
**Figure 7.S14.:** Area under the ROC curve for a logistic regression model using the feature intensity of M187T6\_2 to discriminate virulence phenotypes in the validation data set. The Receiver Operating Characteristics curve was generated analogously to Figure 5 in the main text. An AUC of 0.75 signifies a decent classification performance, but is not sufficient for reliable differentiation of the phenotypes.



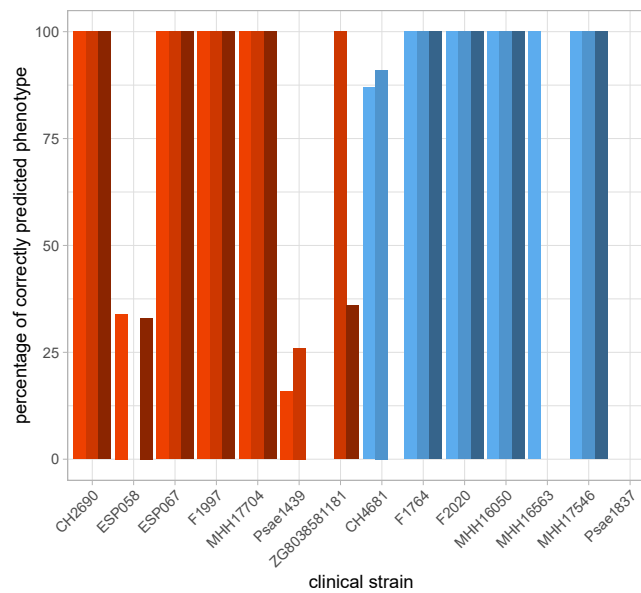
**Figure 7.S15.:** Intra-group correlation of M187T6\_2 with 48h survival in the *Galleria mellonella* assay. The peak area in arbitrary units of M187T6\_2 was plotted against the 48h survival in the *Galleria mellonella* assay. No clear correlation between the abundance of the candidate marker and the extent of virulence in the model could be identified. Red – virulent cluster A strains, blue – avirulent cluster B strains.



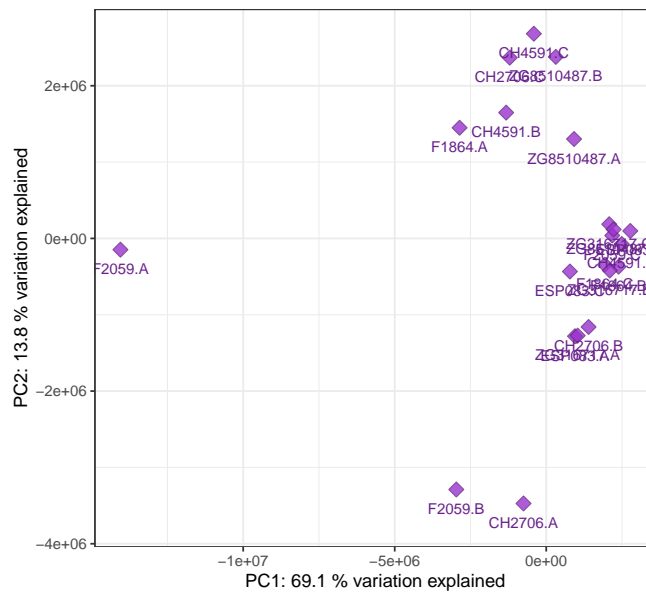
**Figure 7.S16.:** Multidimensional scaling plot visualizing tree distances between the samples of the discovery data set. Red – virulent cluster A strains, blue – avirulent cluster B strains.



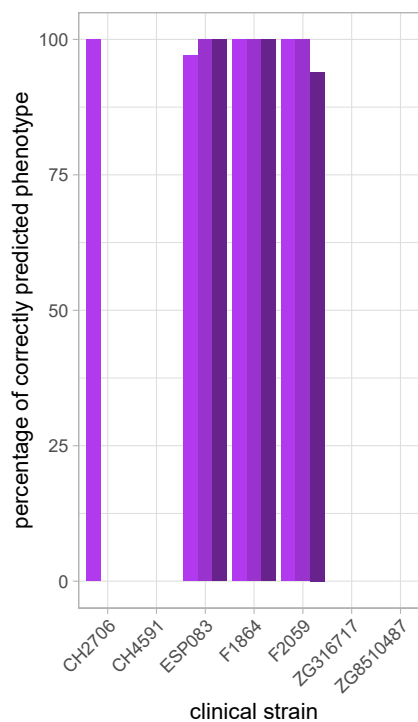
**Figure 7.S17.:** Variable importance plot displaying mean decrease in accuracy and mean decrease in impurity (Gini impurity) of the random forest model constructed from the discovery data set. The ten most important features are shown (identifiers are from the discovery data set): F0142 = M187T6\_2, F0163 = M231T7\_3, F0126 = C9-QNO, F0002 = M85T1\_1, F0007 = C9:1-HQ, F0001 = M126T1\_1, F0415 = M464T9\_3, F0008 = C9:1-HQ, F0004 = M246T3\_1, F0227 = M228T12.



**Figure 7.S18.:** Percentage of correctly predicted virulence phenotype in the validation set if run 100 times independently. While eight strains are reliably assigned to the correct phenotype, three strains appear to be systematically misclassified. Red – virulent cluster A strains, blue – avirulent cluster B strains.



**Figure 7.S19.:** PCA scores plot of the cluster C data set. The plot was generated analogously to Figure 1 in the main text.



**Figure 7.S20.:** Percentage of correctly predicted virulence phenotype in the validation set if run 100 times independently (cf. Figure 7.S18). Only three out of seven strains are reliably assigned to the correct virulence phenotype if the biofilm phenotype differs from those in the discovery data set.

## Supplementary Tables

**Table 7.S1.:** Harvesting data for discovery batch. Each strain was cultivated in three biological replicates and the biomass of all three replicates was harvested at an OD<sub>600</sub> of approximately 2. The exact OD<sub>600</sub> and the timepoint of harvesting in hours after the start of cultivation is for each strain and replicate.

strain	OD <sub>600</sub> at harvesting for individual replicates			timepoint/h
	A	B	C	
CH2860	2.22	2.16	2.12	4.0
CH4528	1.71	1.76	1.52	6.5
ESP046	2.08	2.00	2.29	4.0
ESP088	2.01	1.88	1.84	6.5
F2030	2.05	2.08	1.99	6.5
MHH16798	2.01	1.84	2.10	4.5
ZG302383	1.95	1.99	1.94	4.5
CH2682	2.00	2.07	2.18	4.5
ESP027	2.42	2.23	2.33	6.0
F1959	2.10	1.85	1.98	4.5
F2165	2.85	2.16	2.39	5.0
F2166	2.04	2.50	2.58	6.0
F2224	2.33	2.32	2.25	5.0
MHH17767	1.68	1.77	2.45	5.5

**Table 7.S2.:** Harvesting data for the validation batch. The table is analogous to Table 7.S1.

strain	OD <sub>600</sub> at harvesting for individual replicates			timepoint/h
	A	B	C	
CH2690	1.88	1.73	1.74	5.5
ESP058	1.90	2.08	1.84	6.5
ESP067	1.83	2.01	1.87	5.5
F1997	1.99	1.98	1.78	6.5
MHH17704	1.88	1.86	2.05	6.0
Psae1439	1.91	1.70	1.90	5.0
ZG8038581181	1.93	2.02	1.84	5.5
CH4681	2.13	2.00	1.87	5.0
F1764	2.05	1.85	2.07	6.5
F2020	2.50	1.28	1.46	5.0
MHH16050	2.03	2.00	2.40	6.0
MHH16563	1.64	1.58	1.60	7.5
MHH17546	2.22	1.92	1.96	7.0
Psae1837	1.82	1.95	1.69	5.5

**Table 7.S3.:** Harvesting data for the additional batch. The table is analogous to Table 7.S1.

strain	OD <sub>600</sub> at harvesting for individual replicates			timepoint/h
	A	B	C	
CH2706	1.89	2.07	2.33	4.0
CH4591	1.92	1.98	2.56	4.0
ESP083	1.79	2.06	1.69	4.0
F1864	1.60	2.12	1.94	4.5
F2059	0.85	1.22	2.30	4.5
ZG316717	1.73	2.04	2.11	5.5
ZG8510487	2.40	2.36	2.49	4.5

**Table 7.S4.:** Metabolite identifications. Annotated features in the discovery data set as identified by their median  $m/z$  and median retention time. All annotations are assigned to a metabolite or metabolite class. The identification level is given according to the Metabolomics Standards Initiative [2]: 1, identified by comparison of at least two orthogonal characteristics to an authentic standard; 2, annotated as a distinct compound by comparison with a compound database or the scientific literature; 3, annotated as member of a distinct compound class. The last column states by the use of which properties the compound was identified/annotated: RT, retention time; MS, full scan MS spectrum (exact  $m/z$ , (in-source) fragmentation, isotopic pattern); MSMS, tandem MS fragmentation pattern.

Identifier in the discovery data set	Median $m/z$	Median retention time [min]	Annotation	Metabolite	Comment	Identification Level	Identified by
M112T1_4	112.1118	0.96	spermidine (fragment) (I)	spermidine	(in-source) fragment	1	RT, MS, MSMS
M129T1_4	129.1385	0.96	spermidine (fragment) (II)	spermidine	(in-source) fragment	1	RT, MS, MSMS
M146T1_5	146.1651	0.98	spermidine	spermidine		1	RT, MS, MSMS
M89T1	89.1071	0.98	putrescine	putrescine		1	RT, MS
M175T1_6	175.1190	1.09	arginine	arginine		1	RT, MS, MSMS
M156T1_5	156.0766	1.10	histidine	histidine		1	RT, MS
M130T1_4	130.0497	1.12	5-oxoproline (I)	5-oxoproline		1	RT, MS, MSMS
M104T1_3	104.1070	1.12	choline	choline		1	RT, MS, MSMS
M148T1_2	148.0604	1.13	glutamic acid	glutamic acid		1	RT, MS, MSMS
M176T1_2	176.1031	1.14	citrulline	citrulline		1	RT, MS, MSMS
M365T1_4	365.1057	1.14	sugar	undetermined sugar	various possibilities	3	MS, MSMS
M191T1_5	191.1017	1.18	2,6-diaminoheptanedioic acid	2,6-diaminoheptanedioic acid		1	RT, MS
M116T1_3	116.0705	1.18	proline (I)	proline		1	RT, MS
M219T1_4	219.0975	1.19	Glu Ala	Glu Ala		2	MS, MSMS

Identifier in the discovery data set	Median <i>m/z</i>	Median retention time [min]	Annotation	Metabolite	Comment	Identification Level	Identified by
M147T1_3	147.1126	1.26	lysine	lysine		1	RT, MS
M146T1_3	146.0921	1.34	4-guanidinobutyric acid	4-guanidinobutyric acid		1	RT, MS
M535T1_2	535.1880	1.34	Glu Glu Glu Glu (I)	Glu Glu Glu Glu		2	MS, MSMS
M106T1	106.0489	1.34	serine	serine		1	RT, MS
M124T1_2	124.0391	1.34	nicotinic acid (I)	nicotinic acid		1	RT, MS
M136T1_3	136.0615	1.34	adenine	adenine		1	RT, MS
M308T1_2	308.0906	1.35	glutathione	glutathione	double positive charge	1	RT, MS
M130T1_5	130.0499	1.36	5-oxoproline (II)	5-oxoproline		1	RT, MS, MSMS
M123T1_3	123.0550	1.36	nicotinamide (I)	nicotinamide		2	MS, MSMS
M137T1_2	137.0456	1.37	hypoxanthine (I)	hypoxanthine		1	RT, MS
M124T2_1	124.0390	1.54	nicotinic acid (II)	nicotinic acid		1	RT, MS
M190T2_1	190.0707	1.59	N-acetylglutamate	N-acetylglutamate		2	MS, MSMS
M116T2	116.0703	1.61	proline (II)	proline		1	RT, MS
M123T2_2	123.0548	1.65	nicotinamide (II)	nicotinamide		2	MS, MSMS
M118T2	118.0860	1.66	betaine	betaine		1	RT, MS, MSMS
M130T2	130.0498	1.67	5-oxoproline (III)	5-oxoproline		1	RT, MS, MSMS
M169T2	169.0353	1.68	uric acid	uric acid		1	RT, MS
M333T2_1	332.5617	1.76	NAD (2+)	NAD	double positive charge	2	MS, MSMS
M153T2_2	153.0403	1.77	xanthine (I)	xanthine		1	RT, MS
M664T2	664.1162	1.77	NAD	NAD		2	MS, MSMS
M137T2_1	137.0456	1.79	hypoxanthine (II)	hypoxanthine		1	RT, MS
M535T2	535.1877	1.81	Glu Glu Glu Glu (II)	Glu Glu Glu Glu		2	MS, MSMS
M348T2_1	348.0699	1.88	adenosine-5'-monophosphate	adenosine-5'-monophosphate		1	RT, MS, MSMS



Identifier in the discovery data set	Median <i>m/z</i>	Median retention time [min]	Annotation	Metabolite	Comment	Identification Level	Identified by
M132T2_2	132.1019	1.88	Leucine / Isoleucine / Norleucine	Leucine / Isoleucine / Norleucine	the three species could not be distinguished in the experimental setting	3	RT, MS, MSMS
M364T2	364.0649	2.02	guanosine-5'-monophosphate	guanosine-5'-monophosphate		1	RT, MS
M330T2_1	330.0595	2.03	adenosine-2',3'-cyclic monophosphate	adenosine-2',3'-cyclic monophosphate		1	RT, MS
M153T2_1	153.0404	2.04	xanthine (II)	xanthine		1	RT, MS
M164T2	164.0562	2.33	pterine	pterine		2	MS, MSMS
M268T3_1	268.1041	2.68	adenosine	adenosine		1	RT, MS, MSMS
M137T3_2	137.0456	2.74	hypoxanthine (III)	hypoxanthine		1	RT, MS
M140T3	140.0341	2.82	6-hydroxynicotinic acid	6-hydroxynicotinic acid		1	RT, MS
M182T3_2	182.0809	2.98	tyrosine	tyrosine		1	RT, MS, MSMS
M166T3_1	166.0862	3.46	phenylalanine	phenylalanine		1	RT, MS, MSMS
M120T3_2	120.0807	3.46	phenylalanine (fragment) (I)	phenylalanine	(in-source) fragment	1	RT, MS, MSMS
M103T3	103.0542	3.46	phenylalanine (fragment) (III)	phenylalanine	(in-source) fragment	1	RT, MS, MSMS
M219T5	219.1337	5.24	Ser Leu	Ser Leu		2	MS, MSMS
M220T5_2	220.1179	5.28	D-pantothenic acid	D-pantothenic acid		1	RT, MS, MSMS
M598T5_1	597.6777	5.44	UDP-muramyl-pentapeptide	UDP-muramyl-pentapeptide		2	MS, MSMS
M360T6_6	360.2127	5.68	Ile Val Glu / Val Ile Glu (I)	Ile Val Glu / Val Ile Glu	possibly structural isomer	3	MS, MSMS

Identifier in the discovery data set	Median <i>m/z</i>	Median retention time [min]	Annotation	Metabolite	Comment	Identification Level	Identified by
M188T6_2	188.0707	5.93	tryptophan (fragment) (I)	tryptophan		1	RT, MS
M205T6_1	205.0972	5.93	tryptophan	tryptophan		1	RT, MS
M160T6_1	160.0754	5.93	indole-3-acetaldehyde	indole-3-acetaldehyde		1	RT, MS
M144T6_3	144.0807	5.93	tryptamine (I)	tryptamine		1	RT, MS
M298T6_4	298.0968	6.26	5'-methylthioadenosine	5'-methylthioadenosine		1	RT, MS, MSMS
M211T6_3	211.0865	6.30	pyocyanin	pyocyanin		1	RT, MS, MSMS
M188T6_1	188.0703	6.35	tryptophan (fragment) (II)	tryptophan		1	RT, MS
M360T6_7	360.2123	6.36	Ile Val Glu / Val Ile Glu (II)	Ile Val Glu / Val Ile Glu	possibly structural isomer	3	MS, MSMS
M378T7_3	378.2023	6.56	Pro Tyr Val	Pro Tyr Val		2	MS, MSMS
M295T7_4	295.1290	6.56	Glu Phe	Glu Phe		2	MS, MSMS
M144T7_1	144.0806	6.71	tryptamine (II)	tryptamine		1	RT, MS
M328T7_7	328.2233	6.81	Pro Leu Val (I)	Pro Leu Val		2	MS, MSMS
M263T7_2	263.1389	7.01	Pro Phe (I)	Pro Phe		2	MS, MSMS
M138T7	138.0548	7.09	anthranilic acid	anthranilic acid		1	RT, MS
M120T7	120.0444	7.09	anthranilic acid (fragment)	anthranilic acid	(in-source) fragment	1	RT, MS
M328T7_8	328.2233	7.24	Pro Leu Val (II)	Pro Leu Val		2	MS, MSMS
M263T7_3	263.1385	7.29	Pro Phe (II)	Pro Phe		2	MS, MSMS
M342T8_7	342.2389	7.79	Pro Ile Leu or isomer	Pro Ile Leu or isomer	possibly structural isomer	3	MS, MSMS
M344T8_10	344.2540	8.21	Leu Leu Val or isomer	Leu Leu Val or isomer	possibly structural isomer	3	MS, MSMS
M162T8	162.0550	8.35	DHQ	DHQ		1	RT, MS, MSMS

Identifier in the discovery data set	Median <i>m/z</i>	Median retention time [min]	Annotation	Metabolite	Comment	Identification Level	Identified by
M243T9_3	243.0875	9.19	lumichrome	lumichrome	possible riboflavin fragment	2	MS, MSMS
M216T11	216.1382	11.04	C5-HQ	C5-HQ		2	MS, MSMS
M232T11	232.1330	11.30	C5-QNO	C5-QNO		2	MS, MSMS
M225T11	225.0658	11.45	phenazin-1-carboxylic acid	phenazin-1-carboxylic acid		1	RT, MS, MSMS
M325T12_2	325.0674	11.69	pyochelin (I)	pyochelin		1	RT, MS, MSMS
M230T12	230.1537	12.10	C6-HQ	C6-HQ		2	MS, MSMS
M325T12_1	325.0672	12.34	pyochelin (II)	pyochelin		1	RT, MS, MSMS
M288T13	288.1959	12.66	C9-QNO (I)	C9-QNO		2	MS, MSMS
M258T13	258.1487	12.70	C7:1-QNO	C7:1-QNO		2	MS, MSMS
M242T13	242.1541	13.05	C7:1-HQ	C7:1-HQ		2	MS, MSMS
M244T13_1	244.1697	13.10	HHQ	HHQ		1	RT, MS, MSMS
M286T13	286.1798	13.11	C9:1-QNO (I)	C9:1-QNO		2	MS, MSMS
M159T13	159.0676	13.11	HHQ (fragment)	HHQ	(in-source) fragment	1	RT, MS, MSMS
M260T13	260.1647	13.18	C7-QNO	C7-QNO		1	RT, MS, MSMS
M314T14	314.2112	13.64	C11:1-QNO	C11:1-QNO		2	MS, MSMS
M256T14	256.1695	14.01	C8:1-HQ	C8:1-HQ		2	MS, MSMS
M258T14	258.1854	14.04	C8-HQ	C8-HQ		2	MS, MSMS
M274T14	274.1800	14.06	C8-QNO	C8-QNO		2	MS, MSMS
M286T14	286.1802	14.24	C9:1-QNO (II)	C9:1-QNO		2	MS, MSMS
M320T14	320.1833	14.28	C12-HSL	C12-HSL		1	RT, MS, MSMS
M270T14	270.1854	14.33	C9:1-HQ (I)	C9:1-HQ		2	MS, MSMS
M316T14	316.2268	14.40	C11-QNO	C11-QNO		2	MS, MSMS
M268T14	268.1694	14.43	C9:2-HQ	C9:2-HQ		2	MS, MSMS

Identifier in the discovery data set	Median <i>m/z</i>	Median retention time [min]	Annotation	Metabolite	Comment	Identification Level	Identified by
M270T15_2	270.1854	14.57	C9:1-HQ (II)	C9:1-HQ		2	MS, MSMS
M300T15_1	300.1955	14.74	C10:1-QNO (I)	C10:1-QNO		2	MS, MSMS
M284T15	284.2008	14.86	C10:1-HQ (I)	C10:1-HQ		2	MS, MSMS
M288T15	288.1960	14.92	C9-QNO (II)	C9-QNO		2	MS, MSMS
M270T15_1	270.1855	14.94	C9:1-HQ (III)	C9:1-HQ		2	MS, MSMS
M272T15_2	272.2012	14.96	C9-HQ	C9-HQ		2	MS, MSMS
M300T15_2	300.1956	15.27	C10:1-QNO (II)	C10:1-QNO		2	MS, MSMS
M296T16_2	296.2009	15.51	C11:2-HQ (I)	C11:2-HQ		2	MS, MSMS
M314T16_1	314.2116	15.57	C11:1-PQS (I)	C11:1-PQS		2	MS, MSMS
M298T16_1	298.2167	15.61	C11:1-HQ (I)	C11:1-HQ		2	MS, MSMS
M284T16	284.2009	15.67	C10:1-HQ (II)	C10:1-HQ		2	MS, MSMS
M302T16	302.2113	15.76	C10-QNO	C10-QNO		2	MS, MSMS
M296T16_1	296.2010	15.80	C11:2-HQ (II)	C11:2-HQ		2	MS, MSMS
M314T16_3	314.2113	15.84	C11:1-PQS (II)	C11:1-PQS		2	MS, MSMS
M286T16	286.2166	15.84	C10-HQ	C10-HQ		2	MS, MSMS
M298T16_2	298.2166	15.90	C11:1-HQ (II)	C11:1-HQ		2	MS, MSMS
M673T16	673.3766	15.91	Rha-Rha-C10-C10+Na	Rha-Rha-C10-C10	Na adduct	2	MS, MSMS
M454T16_2	454.2929	16.06	PE(16:0/0:0) (I)	PE(16:0/0:0)		1	RT, MS, MSMS
M314T16_2	314.2116	16.12	C11:1-PQS (III)	C11:1-PQS		2	MS, MSMS
M298T16_3	298.2166	16.32	C11:1-HQ (III)	C11:1-HQ		2	MS, MSMS
M328T16	328.2269	16.38	C12:1-QNO	C12:1-QNO		2	MS, MSMS
M454T16_1	454.2930	16.41	PE(16:0/0:0) (II)	PE(16:0/0:0)		1	RT, MS, MSMS
M312T16	312.2321	16.47	C12:1-HQ	C12:1-HQ		2	MS, MSMS
M316T17	316.2273	16.64	C11-PQS	C11-PQS		2	MS, MSMS
M298T17_1	298.2167	16.74	C11:1-HQ (IV)	C11:1-HQ		2	MS, MSMS

Identifier in the discovery data set	Median <i>m/z</i>	Median retention time [min]	Annotation	Metabolite	Comment	Identification Level	Identified by
M300T17	300.2323	16.75	C11-HQ	C11-HQ		2	MS, MSMS
M527T17	527.3190	16.76	Rha-C10-C10+Na	Rha-C10-C10	Na adduct	2	MS, MSMS
M480T17	480.3087	16.76	PE(18:1/0:0)	PE(18:1/0:0)		1	RT, MS, MSMS
M502T17	502.2904	16.76	PE(18:1/0:0) +Na	PE(18:1/0:0)	Na adduct	1	RT, MS, MSMS
M342T17	342.2429	17.06	C13:1-PQS	C13:1-PQS		2	MS, MSMS
M326T17_1	326.2480	17.17	C13:1-HQ (I)	C13:1-HQ		2	MS, MSMS
M701T17	701.4080	17.29	Rha-Rha-C10-C12+Na	Rha-Rha-C10-C12	Na adduct	2	MS, MSMS
M326T17_2	326.2475	17.41	C13:1-HQ (II)	C13:1-HQ		2	MS, MSMS
M555T18	555.3506	18.17	Rha-C10-C12+Na	Rha-C10-C12	Na adduct	2	MS, MSMS
M326T18	326.2477	18.49	C13:1-HQ (III)	C13:1-HQ		2	MS, MSMS
M328T19	328.2635	18.51	C13-HQ	C13-HQ		2	MS, MSMS
M255T19	255.2316	18.60	palmitoleic acid	palmitoleic acid		1	RT, MS
M260T20	260.1646	19.93	PQS	PQS		1	MS, MSMS
M327T20	327.2269	19.93	oleic acid	oleic acid		1	RT, MS

**Table 7.S5.:** XCMS online parameters. The LC-MS data set was (pre-)processed using XCMS online [3]. The parameters were chosen to fit the analytical machinery used to generate the data and partly modified on an empirical basis.

preprocessing step	parameter	value	explanation
feature detection	method	centWave	peak finding algorithm based on continuous wavelet transformation
	ppm	15	allowable $m/z$ deviation in consecutive scans, expressed in parts per million
	minimum peak width	10.9	chromatographic peak widths in s
	maximum peak width	31.12	
	signal/noise threshold	10	minimum signal-to-noise ratio
	mzdiff	0.0155	minimum absolute $m/z$ difference for overlapping chromatographic peaks
	integration method	1	based on Mexican hat filtered data
	prefilter peaks	3	minimum number of peaks with at least "prefilter intensity" to be retained after prefiltering
	prefilter intensity	100	minimum intensity of "prefilter peaks" (see above)
retention time correction	noise filter	0	not necessary for centroided data
	method	obiwarp	chromatographic alignment by "Ordered Bijective Interpolated Warping"
alignment	profStep	1	$m/z$ step size for profile generation
	mzwid	0.026	$m/z$ width of overlapping $m/z$ slices used to group peaks across samples
	bw	5	maximum deviation of retention times in s
	minfrac	1	fraction of samples of one of the sample groups that have to display a group for it to be valid
	minsamp	1	number of samples of one of the sample groups that have to display a group for it to be valid
annotation	max	50	upper threshold for the number of groups in one $m/z$ slice
	search for	isotopes + adducts	CAMERA considers both isotope peaks and possible adducts
	ppm	5	allowable relative $m/z$ deviation between detected and expected peak, expressed in parts per million
	$m/z$ absolute error	0.015	allowable absolute $m/z$ deviation between detected and expected peak

**Table 7.S6.:** Transcriptomic fold changes of proteins associated with phenazine production. Most phenazine biosynthesis enzymes are significantly upregulated in virulent strains.

PA14 ID	Gene name	Product	log <sub>2</sub> (fold change)	adjusted <i>p</i> -value
PA14_09400	<i>phzS</i>	hypothetical protein	2.59	0.0170
PA14_09410	<i>phzG1</i>	pyridoxamine 5'-phosphate oxidase	2.33	0.0274
PA14_09420	<i>phzF1</i>	phenazine biosynthesis protein	2.47	0.0216
PA14_09440	<i>phzE1</i>	phenazine biosynthesis protein PhzE	2.32	0.0338
PA14_09450	<i>phzD1</i>	phenazine biosynthesis protein PhzD	2.68	0.0503
PA14_09460	<i>phzC1</i>	phenazine biosynthesis protein PhzC	2.96	0.0105
PA14_09470	<i>phzB1</i>	phenazine biosynthesis protein	2.46	0.1104
PA14_09480	<i>phzA1</i>	phenazine biosynthesis protein	4.04	0.0001
PA14_09490	<i>phzM</i>	putative phenazine-specific methyltransferase	1.37	0.2548
PA14_39880	<i>phzG2</i>	pyridoxamine 5'-phosphate oxidase	2.33	0.0326
PA14_39890	<i>phzF2</i>	phenazine biosynthesis protein	2.56	0.0156
PA14_39910	<i>phzE2</i>	phenazine biosynthesis protein PhzE	2.36	0.0373
PA14_39925	<i>phzD2</i>	phenazine biosynthesis protein PhzD	2.31	0.0772
PA14_39945	<i>phzC2</i>	phenazine biosynthesis protein PhzC	2.68	0.0242
PA14_39960	<i>phzB2</i>	phenazine biosynthesis protein	3.22	0.0002
PA14_39970	<i>phzA2</i>	phenazine biosynthesis protein	3.47	0.0007

**Table 7.S7.:** Transcriptomic fold changes of proteins associated with pyochelin, rhamnolipid and alkylquinolone production. While corresponding metabolites are significantly more abundant in virulent strains, this difference is not reflected in the transcriptome data.

PA14 ID	Gene name	Product	log <sub>2</sub> (fold change)	adjusted <i>p</i> -value
PA14_09210	<i>pchA</i>	salicylate biosynthesis isochorismate synthase	1.08	0.3191
PA14_09220	<i>pchB</i>	isochorismate-pyruvate lyase	1.43	0.1719
PA14_09230	<i>pchC</i>	pyochelin biosynthetic protein PchC	1.31	0.2339
PA14_09240	<i>pchD</i>	pyochelin biosynthesis protein PchD	1.22	0.2412
PA14_09270	<i>pchE</i>	dihydroaeruginolic acid synthetase	1.50	0.1283
PA14_09280	<i>pchF</i>	pyochelin synthetase	1.24	0.2170
PA14_09290	<i>pchG</i>	pyochelin biosynthetic protein PchG	1.07	0.3105
PA14_09300	<i>pchH</i>	putative ATP-binding component of ABC transporter	1.16	0.2256
PA14_09320	<i>pchI</i>	putative ATP-binding component of ABC transporter	1.11	0.2530
PA14_09700	<i>pqsL</i>	putative monooxygenase	0.81	0.2270
PA14_19100	<i>rhIA</i>	rhamnosyltransferase chain A	1.54	0.2571
PA14_19110	<i>rhIB</i>	rhamnosyltransferase chain B	1.41	0.1992
PA14_19120	<i>rhIR</i>	transcriptional regulator RhIR	0.84	0.4892
PA14_19130	<i>rhII</i>	autoinducer synthesis protein RhII	0.79	0.6093
PA14_30630	<i>pqsH</i>	putative FAD-dependent monooxygenase	0.86	0.4270
PA14_49760	<i>rhIC</i>	rhamnosyltransferase 2	1.23	0.0833
PA14_51340	<i>mvfR</i>	Transcriptional regulator MvfR	0.55	0.2235
PA14_51350	<i>phnB</i>	anthranilate synthase component II	1.16	0.2558
PA14_51360	<i>phnA</i>	anthranilate synthase component I	1.45	0.1544
PA14_51380	<i>pqsE</i>	Quinolone signal response protein	1.10	0.3316
PA14_51390	<i>pqsD</i>	3-oxoacyl-(acyl carrier protein) synthase III	1.18	0.3168
PA14_51410	<i>pqsC</i>	PqsC	1.16	0.3356
PA14_51420	<i>pqsB</i>	PqsB	0.99	0.4767
PA14_51430	<i>pqsA</i>	coenzyme A ligase	1.37	0.2418



**Table 7.S8.:** Primary metabolites annotated in this study (cf. Table 7.S4). Directional fold change and corrected *p*-value refer to the difference in abundance between virulent cluster A and avirulent cluster B isolates in the discovery data set.

Identifier in the discovery set	Median data <i>m/z</i>	Median retention time [min]	Directional fold change	Corrected <i>p</i> -value	Annotation	Metabolite
M191T1_5	191.1017	1.18	1.04	0.862	2,6-diaminoheptanedioic acid	2,6-diaminoheptanedioic acid
M146T1_3	146.0921	1.34	-1.26	0.423	4-guanidinobutyric acid	4-guanidinobutyric acid
M298T6_4	298.0968	6.26	-2.89	0.148	5'-methylthioadenosine	5'-methylthioadenosine
M130T1_4	130.0497	1.12	-1.11	0.643	5-oxoproline (I)	5-oxoproline
M130T1_5	130.0499	1.36	1.08	0.756	5-oxoproline (II)	5-oxoproline
M130T2	130.0498	1.67	1.12	0.646	5-oxoproline (III)	5-oxoproline
M140T3	140.0341	2.82	1.37	0.013	6-hydroxynicotinic acid	6-hydroxynicotinic acid
M136T1_3	136.0615	1.34	1.06	0.847	adenine	adenine
M268T3_1	268.1041	2.68	-1.41	0.297	adenosine	adenosine
M330T2_1	330.0595	2.03	1.23	0.040	adenosine-2',3'-cyclic monophosphate	adenosine-2',3'-cyclic monophosphate
M348T2_1	348.0699	1.88	-1.17	0.674	adenosine-5'-monophosphate	adenosine-5'-monophosphate
M138T7	138.0548	7.09	-1.33	0.390	anthranilic acid	anthranilic acid
M120T7	120.0444	7.09	-1.32	0.398	anthranilic acid (fragment)	anthranilic acid
M175T1_6	175.1190	1.09	1.21	0.578	arginine	arginine
M118T2	118.0860	1.66	1.19	0.206	betaine	betaine
M104T1_3	104.1070	1.12	-1.41	0.010	choline	choline

Identifier in the discovery set	Median <i>m/z</i> data	Median retention time [min]	Directional fold change	Corrected <i>p</i> -value	Annotation	Metabolite
M176T1_2	176.1031	1.14	-1.44	0.179	citrulline	citrulline
M220T5_2	220.1179	5.28	1.28	0.009	D-pantothenic acid	D-pantothenic acid
M148T1_2	148.0604	1.13	-1.14	0.598	glutamic acid	glutamic acid
M308T1_2	308.0906	1.35	1.20	0.265	glutathione	glutathione
M364T2	364.0649	2.02	1.19	0.063	guanosine-5'-monophosphate	guanosine-5'-monophosphate
M156T1_5	156.0766	1.10	-1.06	0.516	histidine	histidine
M137T1_2	137.0456	1.37	1.52	0.125	hypoxanthine (I)	hypoxanthine
M137T2_1	137.0456	1.79	1.45	0.153	hypoxanthine (II)	hypoxanthine
M137T3_2	137.0456	2.74	1.15	0.840	hypoxanthine (III)	hypoxanthine
M160T6_1	160.0754	5.93	1.06	0.527	indole-3-acetaldehyde	indole-3-acetaldehyde
M132T2_2	132.1019	1.88	1.20	0.041	Leucine / Isoleucine / Nor-leucine	Leucine / Isoleucine / Nor-leucine
M243T9_3	243.0875	9.19	1.27	0.003	lumichrome	lumichrome
M147T1_3	147.1126	1.26	1.16	0.098	lysine	lysine
M190T2_1	190.0707	1.59	-1.16	0.444	N-acetylglutamate	N-acetylglutamate
M664T2	664.1162	1.77	1.13	0.178	NAD	NAD
M333T2_1	332.5617	1.76	1.14	0.225	NAD (2+)	NAD (2+)
M123T1_3	123.0550	1.36	1.07	0.455	nicotinamide (I)	nicotinamide
M123T2_2	123.0548	1.65	1.12	0.495	nicotinamide (II)	nicotinamide
M124T1_2	124.0391	1.34	-1.13	0.735	nicotinic acid (I)	nicotinamide
M124T2_1	124.0390	1.54	-1.08	0.834	nicotinic acid (II)	nicotinamide

Identifier in the discovery set	Median <i>m/z</i> data	Median retention time [min]	Directional fold change	Corrected <i>p</i> -value	Annotation	Metabolite
M327T20	327.2269	19.93	-1.32	0.040	oleic acid	oleic acid
M255T19	255.2316	18.60	-1.92	0.006	palmitoleic acid	palmitoleic acid
M454T16_2	454.2929	16.06	1.01	0.910	PE(16:0/0:0) (I)	PE(16:0/0:0)
M454T16_1	454.2930	16.41	-1.31	0.025	PE(16:0/0:0) (II)	PE(16:0/0:0)
M480T17	480.3087	16.76	-1.62	0.006	PE(18:1/0:0)	PE(18:1/0:0)
M502T17	502.2904	16.76	-1.62	0.004	PE(18:1/0:0) +Na	PE(18:1/0:0)
M166T3_1	166.0862	3.46	1.02	0.904	phenylalanine	phenylalanine
M120T3_2	120.0807	3.46	1.02	0.899	phenylalanine (fragment) (I)	phenylalanine
M103T3	103.0542	3.46	1.02	0.918	phenylalanine (fragment) (III)	phenylalanine
M116T1_3	116.0705	1.18	-1.00	0.969	proline (I)	proline
M116T2	116.0703	1.61	1.03	0.867	proline (II)	proline
M89T1	89.1071	0.98	-1.02	0.939	putrescine	putrescine
M106T1	106.0489	1.34	1.16	0.238	serine	serine
M146T1_5	146.1651	0.98	1.20	0.105	spermidine	spermidine
M112T1_4	112.1118	0.96	1.18	0.232	spermidine (fragment) (I)	spermidine
M129T1_4	129.1385	0.96	1.18	0.216	spermidine (fragment) (II)	spermidine
M365T1_4	365.1057	1.14	1.18	0.057	sugar	unidentified sugar
M144T6_3	144.0807	5.93	1.06	0.525	tryptamine (I)	tryptamine
M144T7_1	144.0806	6.71	1.07	0.542	tryptamine (II)	tryptamine
M205T6_1	205.0972	5.93	1.07	0.469	tryptophan	tryptophan

Identifier in the discovery set	Median <i>m/z</i> data	Median retention time [min]	Directional fold change	Corrected <i>p</i> -value	Annotation	Metabolite
M188T6_2	188.0707	5.93	1.07	0.484	tryptophan (fragment) (I)	tryptophan
M188T6_1	188.0703	6.35	1.11	0.665	tryptophan (fragment) (II)	tryptophan
M182T3_2	182.0809	2.98	1.16	0.092	tyrosine	tyrosine
M169T2	169.0353	1.68	1.36	0.001	uric acid	uric acid
M153T2_2	153.0403	1.77	1.22	0.330	xanthine (I)	xanthine
M153T2_1	153.0404	2.04	1.22	0.305	xanthine (II)	xanthine

### Supplementary References

- [1] W. Scherber. “Stable Isotope Labeling to Improve Metabolite Identification in Untargeted Metabolomics of Pathogenic Bacteria”. Master’s Thesis. Hochschule Aalen – Technik und Wirtschaft, 2020.
- [2] L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W.-M. Fan, O. Fiehn, R. Goodacre, J. L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A. N. Lane, J. C. Lindon, P. Marriott, A. W. Nicholls, M. D. Reily, J. J. Thaden, and M. R. Viant. “Proposed minimum reporting standards for chemical analysis”. In: *Metabolomics* 3.3 (Sept. 2007), pp. 211–221. DOI: 10.1007/s11306-007-0082-2.
- [3] R. Tautenhahn, G. J. Patti, D. Rinehart, and G. Siuzdak. “XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data”. In: *Analytical Chemistry* 84.11 (June 2012), pp. 5035–5039. DOI: 10.1021/ac300698c.

## 8 | General Discussion and Outlook

### 8.1. Discussion

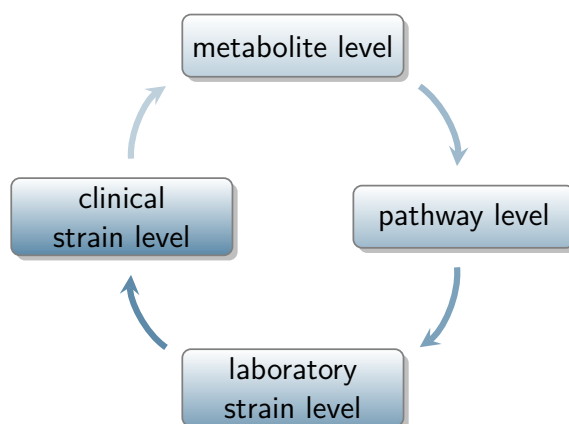
Understanding and characterising phenotypes of microbial pathogens is a prerequisite for mechanistic understanding of infection processes and therefore represents an important element in infection research [1], especially with regard to virulence factors [2]. Along with other –omics methods, microbial metabolomics is a very versatile tool to describe, analyse and scrutinise bacterial phenotypes [3].

The study presented here examined the *Pseudomonas aeruginosa* metabolome and its interplay with pseudomonal virulence on multiple levels, ranging from the annotation of specific metabolites of interest to the study of virulence-associated metabolic pathways and investigations of defined genetically modified organisms as well as clinical *P. aeruginosa* strains on the whole-metabolome level (Figure 8.1). It produced a number of interesting insights into the biosynthesis of virulence-associated secondary metabolites in *P. aeruginosa*, described several hitherto unknown compounds and generated novel knowledge on the interdependence of the pseudomonal metabolome and *in vivo* virulence that enabled the setup of a predictive model for virulence based on untargeted metabolomics data.

The individual subprojects had different foci in terms of content and methodology, but together they led to an overall picture that allows a better metabolome-based understanding of virulence factors in *P. aeruginosa*, as outlined in the following paragraphs.

#### 8.1.1. Chemical inventorying and extended coverage of the *P. aeruginosa* metabolome

The main obstacle on the way towards a meaningful description and interpretation of phenotypes based on untargeted metabolomics data is feature annotation and metabolite identification [4]. Although over the decades, numerous researchers have engaged in the discovery of *P. aeruginosa* secondary metabolites, specifically alkylquinolones [5–7], phenazines [8–10] and rhamnolipids [11, 12], untargeted metabolomics experiments on *P. aeruginosa* extracts still revealed a multitude of unassigned compounds. This



**Figure 8.1.:** Circle of research scope in this study. While the research focus in this study started at the metabolite level with endeavours to annotate more structures in the *P. aeruginosa* metabolome, it gradually moved towards higher organisational levels: The mechanistic and phenotypic examination of AQ biosynthesis and specifically the substrate specificity of PqsBC provide metabolomics data on a pathway level. Subsequently, the global role of PrmC in metabolism and virulence was studied on a laboratory strain level without the restriction to a defined pathway. The untargeted metabolomics project with clinical *P. aeruginosa* isolates then looked at metabolome differences at a clinical strain level using bacteria sampled from human infections. Finally, the identification of a putative biomarker for virulence brought the research back to the metabolite level. Further research could include re-iterating this circle by examining the biosynthetic origin and the regulatory functions of this newly discovered virulence-associated metabolite.

is reflected by the fact that in the dataset presented in Chapter 3, only 150 of 518 MS<sup>2</sup> consensus spectra could be matched to a spectrum in an in-house library and/or online databases. This corresponds to an annotation rate of less than 30% and does not take into account the numerous features that did not produce an MS<sup>2</sup> spectrum in data-dependent fragmentation. Since non-annotated features potentially held valuable information, considerable effort needed to be invested in their annotation.

Software tools to aid metabolite identification that existed at the time, such as ‘Met-Frag’ [13], ‘CSI:FingerID’ [14] and ‘Molecular Networking’ [15], were of limited use for our data set, which prompted the development of a custom tool named CluMSID (see section 8.1.5).

The application of CluMSID led to the annotation of 27 novel members of the alkyl-quinolone family of quorum sensing signalling molecules. In addition to canonical AQs, several short chain congeners and AQs with two double bonds in the alkyl chain have been annotated. Furthermore, alternative structures have been proposed for AQs with an additional oxygen atom that have been assigned as 3-alkyl-2,3-dihydroxy-4-quinolones by Lépine *et al.* [6]: The fragmentation patterns seen in our study suggest

that the respective features are better explained by side chain oxidation products of unsaturated QNOs. These findings represent significant advances in the inventory of the chemodiversity of AQs, which are a very interesting group of secondary metabolites as they are involved in the regulation of various virulence-associated processes, e.g. biofilm formation, toxin production and many other via activation of the PqsR (MvfR) receptor [16] and also act directly in the acquisition of iron [17] and in interspecies relations [18]. These multiple roles underline their significance for pseudomonal virulence and the importance to study these compounds in detail. For instance, an unsaturated alkylquinolone-*N*-oxide has recently been described as bactericidal for *Staphylococcus aureus* [19]. Interestingly, Szamosvári and Böttcher were able to demonstrate a 20-fold higher activity of *trans*- $\Delta^1$ -2-(non-1-enyl)-4-quinolone-*N*-oxide compared to its saturated congeners, emphasising the importance of a thorough knowledge of the different AQ species. Since the publication of our work, even more non-canonical AQs have been discovered, e.g. by Li *et al.* who described AQ congeners with side chains containing sulphur atoms or phenyl groups [20]. Knowing the various AQs in detail is also of interest as the *pqs* system is a promising target for pathoblockers, i.e. drugs that inhibit bacterial virulence instead of exerting bactericidal or bacteriostatic effects [21].

In addition to new discoveries in the AQ family, novel phenazine compounds have been found by CluMSID using hierarchical clustering, namely the putative identifications of tetrahydropyocyanin and pyocyanin carboxylic acid. Phenazines play a vital role in pseudomonal virulence by acting as redox-active toxicants to the human host and other microorganisms [22–27] but also as signalling molecules [28], in biofilm formation and antibiotic tolerance [29–31] as well as in anaerobic survival [32]. Despite the significance of these metabolites, the differential functions of the various phenazine metabolites and the reasons and consequences of phenazine chemodiversity in *P. aeruginosa* remain underexplored [8, 10]. However, broad analytical coverage of pseudomonal phenazines was important to the study presented here, as Publication 4 deals with a *P. aeruginosa* transposon mutant deficient—or rather severely impaired—in pyocyanin production.

Besides generating new knowledge about the composition of the pseudomonal metabolome, CluMSID also enabled the identification of artifact features, e.g. a proton-bound dimer of HQNO and a proton-bound mixed dimer of HHQ and HQNO which could have been mistaken for novel metabolites—especially as their abundance is expected to vary in relation to virulence-associated factors like HQNO and HHQ themselves. While tools like CAMERA are able to detect many common adducts [33], these specific adducts are more difficult to capture. A recent article by Sindelar and Patti

pointed out how crucial it is to uncover such artifacts in untargeted metabolomics data early in the data analysis process to avoid the investment of unnecessary efforts in further examination of a seemingly novel metabolites [34].

Taken together, this subproject contributed significantly to the exploration of the chemodiversity of *P. aeruginosa* virulence-associated secondary metabolites and was a necessary and beneficial groundwork for further metabolome-based studies of pseudomonal virulence factors as it improved the feature annotation process and supported metabolite identification, the key for biological interpretation of metabolomics data [35].

### 8.1.2. Mechanistic insights into *P. aeruginosa* virulence factor biosynthesis

Building on the descriptive data on the chemodiversity of AQs gained in Publication 1, the Thesis aimed at answering the question how *P. aeruginosa* produces such a variety of similar yet distinct AQ species. It became clear early on that this problem was difficult to tackle using MS-based metabolomics alone. Therefore, the AQ biosynthesis steps performed by the enzyme complex PqsBC, which condenses the anthranilate-derived aromatic part of the AQ with an activated fatty acid to form a bicyclic compound with an alkyl side chain, were studied in cooperation with structural biologists and experts in protein bioanalysis. This combination of complementary approaches has provided significant mechanistic insights into AQ biosynthesis.

The studies resulted in the finding that the FabH-like heterodimeric enzyme complex PqsBC exhibits a substrate promiscuity that allows it to produce AQs of different side chain lengths depending on the fatty acyl-CoA used in the reaction. By integrating structural and bioanalytical data it could be demonstrated that the spectrum of AQ side chain lengths found *in vivo* is a function of the availability of fatty acid substrates and the relative substrate specificity of PqsBC that prefers medium-chain fatty acids. Furthermore, we produced evidence that unsaturated AQs are directly biosynthesised by PqsBC using unsaturated fatty acids as substrates. Finally, an intricate structural model of the catalytic cycle of PqsBC was developed based on structural biology data.

The results touch on the highly complex topic of fatty acid metabolism in *P. aeruginosa* which is involved in many biochemical processes, including the production of virulence-associated secondary metabolites: Besides AQs, also rhamnolipids [36] and HSLs [37] contain fatty acyl moieties. It is known in the literature [12, 37, 38] and can also be seen in the untargeted metabolomics data of the other subprojects [39–41] that rhamnolipids and HSLs show a lot less chemodiversity than AQs, with rhamno-



lipids using C10 and C12 fatty acids and HSLs being produced exclusively from C4 and 3-oxo-C12 fatty acids in *P. aeruginosa*. If one also considers fatty acid biosynthesis and degradation by Fab and Fad enzymes, respectively, PqsBC competes for substrates with various other fatty acid metabolising enzymes and presumably profits from its substrate promiscuity by being able to process longer and shorter fatty acyl-CoA molecules when its preferred substrates, octanoyl-CoA and decanoyl-CoA, are scarce. Thereby, sufficient precursor supply for AQ biosynthesis is ensured even if for instance significant amounts of octanoate are channelled into rhamnolipid biosynthesis by condensation with malonyl-ACP to  $\beta$ -ketodecanoyl-ACP that is used to biosynthesise C10-rhamnolipids [36].

The notion that unsaturated AQs are directly produced by PqsBC from unsaturated fatty acid precursors is particularly interesting as the closely related species *Burkholderia thailandensis* uses a specific enzyme named HmqF to introduce double bonds into 4-hydroxy-3-methyl-2-alkylquinolines [42]. As *P. aeruginosa* does not possess a HmqF homologue—the *hmqABCDEFG* operon is homologous to the *pqsABCDE* operon but contains two additional genes coding for HmqF and HmqG [43, 44]—there had to be an alternative biosynthetic route towards unsaturated AQs. As highlighted by the work of Szamosvári and Böttcher mentioned above [19] and further work based on it [45], saturated and unsaturated AQs can have different biological activities. Therefore, an augmented understanding of the biosynthesis of unsaturated AQs holds promise for further insights into the various biological functions of AQs.

This work on PqsBC illustrates how metabolomics can evolve from a descriptive view of the chemical composition of a biological system to inspire and support research on biochemical mechanisms—a key task in the advancement of the field [4]. It added mechanistic understanding of AQ biosynthesis to the chemical inventorying endeavours of the preceding subproject by leveraging cooperations with scientists from other fields and thereby contributed significantly to the overall project aim of gaining a deeper understanding of the interplay of metabolism and virulence in *P. aeruginosa*.

### 8.1.3. Exploration of the interplay of metabolism and virulence in a genetically modified laboratory strain

On the basis of the findings of the first subprojects, the next parts examined the interconnections of metabolism and virulence at a higher organisational level: First, metabolic effects of a virulence-associated enzyme with a partially unexplained mechanism

were studied, and then metabolism–virulence relationships were revealed in different clinical *P. aeruginosa* strains.

In the sophisticated regulatory network that controls pseudomonal virulence, there are several important players whose effects have already been demonstrated, while it is not entirely clear how these effects are achieved mechanistically [16]. Due to their close relationship with the biosynthesis of alkyl quinolones and phenazines, PqsE and PrmC are the two enzymes that were of the highest interest to this project. PqsE, that has been shown to function as a thioesterase directly involved in the alkyl quinolone biosynthesis pathway soon after the onset of this Thesis project [46], is heavily studied by several groups [47–51]. In contrast, the contribution of PrmC towards virulent phenotypes is still underexplored, making it a suitable object of investigation for an untargeted metabolomics study. Moreover, additional complementary –omics data for *prmC* deficient *P. aeruginosa* PA14 strains was available [52, 53] which allowed to conduct multiomics analyses of the untargeted metabolomics data gained in the subproject.

By using both GC-MS and LC-MS and by additionally analysing the exometabolome, a comprehensive data set was generated that described the effect of PrmC on the pseudomonal metabolome for the first time. The most striking outcome was a depletion of various organic metabolites in the *prmC* deficient strain that could be traced back to an impaired shikimate pathway whose key metabolites shikimate and shikimate-3-phosphate were below the level of detection in the *tnprmC* strain. With regard to secondary metabolites, results from previous research [52, 53], including a significantly lower production of AQs and phenazines were confirmed and extended, so that the effects on the individual members of these metabolite classes could be quantified. The lower levels of these metabolites and the decreased activity of the shikimate pathway are assumed to be connected: as both AQs and phenazines are biosynthesised from shikimate-derived educts, a limited precursor supply would explain the lower abundance of those metabolites. This hypothesis is supported by the fact that the *phzC1* and *phzC2* transcripts are significantly down-regulated in *prmC* deficient *P. aeruginosa* [53]. They code for PhzC which is involved in the shikimate pathway and has been discussed as a regulator of precursor flux into the biosynthesis of phenazines [9].

The higher level of detail concerning virulence-associated secondary metabolites enabled the observation that medium chain AQs are less affected by *prmC* deficiency than long chain and short chain congeners. A possible explanation for this phenomenon establishes a link to the PqsBC subproject: It can be assumed that when less aromatic precursors and less enzyme are present—PqsBC was shown to be less abundant in

tn $prmC$  by Krueger *et al.* [53]—the availability of medium-chain fatty acyl-CoAs might be less rate limiting for the enzymatic reaction, whereby the preferred fatty acyl-CoA substrates are used more extensively resulting in an increased percentage of medium chain AQs. Another interconnection to a different subproject is that a putative molecular biomarker for virulent phenotypes identified in *P. aeruginosa* clinical isolates (*vide infra*) was also found to be less abundant in the  $prmC$  deficient strain. These interrelationships within the Thesis project further emphasise the impact of a better metabolome-level understanding for different aspects of *P. aeruginosa* virulence.

#### 8.1.4. Exploration of the interplay of metabolism and virulence in clinical *P. aeruginosa* strains

The analytical approach of the previous two subprojects mainly consisted of investigating individual biological processes important for virulence at the metabolome level, which resulted in important contributions to comprehending the relationship between the metabolome and virulence in *P. aeruginosa*. However, the dissertation project presented here also aims to leverage the possibilities of metabolomics as a systems biology discipline to understand the interconnections between metabolomic characteristics and overall *in vivo* virulence. This global view of virulence is intended to complement and validate the knowledge gained through reductionist approaches and open up the possibility of generating clinically significant results.

To this end, clinical *P. aeruginosa* strains isolated from patients with different types of infections and from different hospitals were subjected to untargeted metabolomics profiling experiments. The strains have been collected by the ‘Molecular Bacteriology’ group of the Helmholtz Centre for Infection Research, and many of them had already undergone a characterisation process that includes genomics, transcriptomics as well as several phenotyping experiments and have been documented in the ‘BACTOME’ database [54].

Generally, the quantification of bacterial virulence is not trivial [55]. Given the fact that *P. aeruginosa* is an opportunistic pathogen that predominantly infects patients with a pre-existing disease such as CF [56] and is often found in co-infections with other bacteria [57], the clinical course of infection does not seem a promising metric for *in vivo* virulence in most cases. Instead, animal experiments are used to quantitatively investigate virulence, and a *Galleria mellonella*-based system is available for *P. aeruginosa*, which correlates well with the effect in the mouse model [58]. Because biofilm formation and specifically the biofilm morphology represent important factors in pseu-

domonal virulence [59], *in vivo* virulence in *G. mellonella* and the biofilm phenotype cluster were used as target variables and surrogates for *P. aeruginosa* virulence in this study.

The choice of analytical methodology for this study was based on the findings of the previous subprojects: The chemical inventorying endeavours described in Chapter 3 proved that untargeted metabolomics using reversed phase chromatography and positive mode electrospray ionisation in combination with QTOF mass spectrometry is an appropriate technology to capture a broad yet detailed picture of many virulence-associated metabolites in *P. aeruginosa*, whose annotation was facilitated by CluMSID. The research on PrmC laid out in Chapter 6 demonstrated that GC-MS metabolomics to uncover changes in primary metabolism and LC-MS exometabolomics to study metabolic footprints are useful additions to the analytical toolbox but also showed that LC-MS endometabolomics is the central approach with the highest information density. Moreover, reports in the literature suggest that the genetic background has a less pronounced effect on the pseudomonal primary metabolome than the cultivation conditions [60]. On these grounds and owing to the number of samples and complexity of data analysis, it was decided to limit this study to the analysis of the cellular metabolome by means of LC-MS.

In total, 35 *P. aeruginosa* clinical strains with three different phenotypes or phenotype combinations have been studied, grouped into a discovery set used for hypothesis generation and model building and two validation sets to test these hypotheses and models. Unsupervised statistical methods enabled a separation of virulent Cluster A and avirulent Cluster B strains based exclusively on their metabolomic profile, which emphasises the power of untargeted metabolomics in unravelling interconnections of virulence and the metabolome. This is further supported by the notion that transcriptional profiles were not indicative of the two phenotypes.

In line with other recent reports [61–63] and with the results of the other subprojects, metabolome differences between the virulence phenotypes were mostly reflected in differential abundance of secondary metabolites that are known to be directly involved in virulence, namely AQS, phenazines and rhamnolipids.

While no clear trends could be detected in the abundance of primary metabolites, it is striking that a number of very important features—judged both by their fold change and *p*-value as well as by their contribution to the random forest classification model—were ‘unknowns’. Most prominently, a featured named M187T6\_2 with the putative sum formula  $C_{12}H_{15}N_2$  for the cation, was identified as a potential biomarker for viru-

lence, as it perfectly separated virulent Cluster A from avirulent Cluster B isolates in the discovery data set. Certainly, caution must be exercised when looking at unknown features as they might well be artifacts [34]. Therefore, considerable effort was invested to verify that the feature is a metabolite produced by *P. aeruginosa* by leveraging analytical technologies such as the verification of the sum formula through isotopic pattern analysis using Bruker's SmartFormula [64] and by demonstrating that the molecule is generated in labelled form when the bacteria are fed with  $^{13}\text{C}$  labelled glucose [65, 66]. Moreover, the feature was detected in *P. aeruginosa* by other researchers which was confirmed by MASST searching [67]. Besides the details mentioned in the publication, extensive work was put into purifying the molecule in the laboratory. The experiments included up-scaling of the cultivation volume followed by liquid-liquid extraction or solid phase extraction and also tuning the cultivation conditions to stimulate the production of the metabolite of interest but in the end, these attempts remained unsuccessful due to the very low abundance of M187T6\_2 and because of the difficulty of handling large quantities of pathogenic bacteria. This failure is a major shortcoming of the study, as the annotation of putative biomarkers is a key step towards clinical applicability [68].

Still, it was possible to construct a statistical model using random forest classification based on the complete metabolome data which discriminated virulent Cluster A and avirulent Cluster B isolates even better than the putative biomarker M187T6\_2 alone. Random forest classification models have been used in the metabolomics field in applications as diverse as Alzheimer's disease [69], Zika virus screening [70], dairy cow conditioning [71], and *Drosophila melanogaster* metabolism [72]. The work presented here demonstrate that it is also a useful data analysis tool for microbial metabolomics and the investigation of *P. aeruginosa* virulence.

The fact that a model that was trained with two phenotypes—high virulence with Cluster A biofilm morphology and low virulence with Cluster B biofilm morphology—performs rather poorly on a third phenotype that exhibits a high virulence in the *G. mellonella* model and a Cluster C biofilm morphology is not necessarily surprising as generalisation of the classification performance to unseen cases (i.e. cases whose class was not present in the training or discovery data set) is a major problem for classification algorithms [73]. This is particularly true if properties are partly linked to each other as is the case for virulence and the biofilm phenotype in *P. aeruginosa*: Thoeming *et al.* have shown that different correlations of biofilm phenotypes and virulence traits exist and that Cluster C isolates share some properties with Cluster A isolates, e.g. twitching motility, and others with Cluster B strains, such as pyocyanin production

[59]. It stands to reason that these differences that exist despite identical virulence results in the *G. mellonella* model are also reflected in the metabolome. Thus, the existing model is not yet applicable to predict pseudomonal virulence as such from metabolomics data alone, while it is successful and superior to transcriptomic data in the discrimination of the two virulence and biofilm phenotypes represented in the discovery data set.

Hence, the research in this subproject provided evidence for the notion that untargeted LC-MS metabolomics is a useful tool for the study of virulence factors in *P. aeruginosa* as it enabled various interesting discoveries concerning the interplay of virulence and metabolism in this versatile pathogen. The identification of a putative virulence biomarker and the construction of a classification model for two distinct virulence phenotypes is a valuable advancement in the understanding of pseudomonal virulence and integrates well with the results of the other subprojects to refine the picture of virulence-associated metabolites in *P. aeruginosa*.

#### **8.1.5. Contribution of a feature annotation tool for the metabolomics and natural product chemistry community**

This Thesis also includes the presentation of a software tool for feature annotation in untargeted metabolomics and other non-targeted MS data. As detailed above, the programme named CluMSID arose from the need to improve the metabolite identification rate in *P. aeruginosa* extracts and hence is tightly connected to the other subprojects described here. It has, however, evolved into an independent software package which can be used for various types of samples and help researchers tackle the highly demanding challenge of metabolite annotation.

Already the prototype of the tool included as an R script in Publication 1 was widely received in the community as reflected in the inclusion of the tool in several review article on metabolomics processing and data analysis software [74–77]. It therefore seemed worthwhile to transform the programme into a more user-friendly and quality-assured version, namely as an R package in the Bioconductor repository. Bioconductor assures high quality standards, considerably improves visibility for other researchers and integrates the tool in an existing ecosystem of other metabolomics-related packages [78]. It is particularly noteworthy that CluMSID is interfaceable with one of the community standard software suites, the ‘xcms’ family of R packages [79] which makes the package particularly convenient for a large number of scientist already working with ‘xcms’. Even if no concrete applications have been published yet, the personal feedback and the

download statistics (accessible at <http://bioconductor.org/packages/stats/bioc/CluMSID/>) allow the conclusion that with CluMSID, a value-added contribution to the community-based software landscape for the analysis of metabolomics data could be accomplished.

## 8.2. Conclusions

The sum of the subprojects that make up this dissertation gives a more detailed account of the diverse and complex relationships between virulence and metabolism in *P. aeruginosa* and opens a comprehensive new perspective on the possibilities of using metabolomics to study bacterial virulence on the metabolite, pathway, organism and strain level.

Significant improvements in the analytical coverage and annotation of pseudomonal small-molecule virulence factors were achieved by advanced data analysis methods with the annotation of i.a. 27 previously undescribed alkylquinolones. These bio- and chemoinformatic efforts were developed into an easy-to-use R package that enables researchers to apply unsupervised clustering methods based on similarity of MS<sup>2</sup> spectra to aid the annotation of ‘unknown’ features in untargeted metabolomics data sets. Building on the extended description of the chemodiversity of AQs, the biosynthetic mechanism leading to the observed diversity of these metabolites was investigated and it was found that the AQ spectrum produced *in vivo* is determined by both the substrate specificity of the relevant enzyme PqsBC and the availability of fatty acid precursors. In the course of these experiments, it could also be shown that unsaturated AQs are synthesised by PqsBC from unsaturated fatty acid substrates. An important virulence-associated enzyme not directly involved in virulence factor production, the peptide chain release factor methyltransferase PrmC, was characterised with respect to influences on the metabolome of *P. aeruginosa* and it was revealed that its dampening effect on virulence includes the depletion of aromatic precursor molecules needed for the biosynthesis of both AQs and phenazines. Finally, the findings were integrated to study interrelations of metabolomic features and *in vivo* virulence in different *P. aeruginosa* clinical strains which led to the discovery of a putative virulence biomarker and allowed for the construction of a random forest classification model based on untargeted metabolomics data that is able to discriminate clinical strains that have distinct virulence and biofilm morphology phenotypes.



In conclusion, this Thesis demonstrates the considerable potential of targeted and untargeted metabolomics for the study of *P. aeruginosa* virulence factors and thereby made important contributions towards the comprehension of the complex interplay of metabolism and virulence in this fascinating pathogen.

### 8.3. Outlook

In addition to the experiments and findings described in this Thesis, the metabolome-based study of virulence factors in *P. aeruginosa* offers other possibilities to discover novel insights, since the interconnections of metabolism and virulence are numerous and highly complex. With respect to metabolome coverage and chemical inventorying, many features in the *P. aeruginosa* metabolome remain unassigned even after extensive annotation efforts. While the current version of CluMSID has proven to be very effective in aiding annotation of metabolites from classes with many homologous members such as AQs, there appears to be considerable potential for the detection of metabolites that do not show spectral similarities to known metabolites. A limitation of CluMSID is its lack of automatic interfaceability with online MS<sup>2</sup> spectral databases—a feature that is, for instance, offered by the GNPS ‘Molecular Networking’ web tool [15]. Future development of the software package should therefore include similarity comparisons with MS<sup>2</sup> spectra deposited in such databases to enable the exploration of clusters that do not contain known metabolites and thereby increase its usefulness for metabolite annotation. Moreover, the extension of analytical coverage of the *P. aeruginosa* metabolome requires ongoing efforts supported by software other than CluMSID, since not all annotation problems can be solved by means of spectral similarity. Beyond improving data analysis strategies, the bioanalytical arsenal leveraged to explore the chemical inventory of *P. aeruginosa* can certainly be extended to include polar separations, negative mode ionisation and other detection principles as shown for *Staphylococcus aureus* [80]. A comprehensive feature credentialing [65] would additionally support the metabolite annotations and thus benefit the scientific quality of the results despite considerable operating expense. A shift of focus away from metabolites known to be involved more or less directly in virulence regulation or exertion might have beneficial effects for metabolome coverage, as well. For instance, in the PrmC subproject, many metabolites involved in the shikimate pathway and subsequent or preceding metabolic reactions have not been detected. Better coverage of metabolites relevant for this part



of primary and intermediary metabolism would contribute to a more concise picture of the influence of PrmC on pseudomonal metabolism.

In the PqsBC subproject, the biosynthesis of mono-unsaturated AQs was studied by means of feeding and labelling experiments. However, AQ species with two double bonds have also been detected and annotated with the help of CluMSID in the *P. aeruginosa* metabolome. The biosynthesis of these AQ congeners was outside the scope of this project although very interesting—especially given that researchers are only beginning to gauge the differential effects of saturated and unsaturated AQs [45].

The groundwork laid in this Thesis along with technical progress in sample preparation, separation, mass spectrometry and data analysis will enable the investigation of a higher number of clinical *P. aeruginosa* strains by means of untargeted metabolomics in order to validate and extend the insights into the interplay of metabolism and *in vivo* virulence and classification models to predict virulence properties. Of particular importance will be further experiments on the putative virulence marker M187T6\_2. Even if its structure may not be elucidated by purification and NMR, its biological role is completely unclear. The fact there might still be unknown metabolites involved in pseudomonal virulence is a huge motivation for further research on the function of this compound. With a broader data basis comprising more isolates and more defined virulence phenotypes, new classification models could be built and potentially developed towards clinical applicability. Although this is a fairly remote prospect, the ability to predict virulence of clinical *P. aeruginosa* isolates based on their metabolomic profile holds promise to help clinicians assess the clinical course of a *P. aeruginosa* infection and thus might contribute to therapeutic progress in this important area of clinical infectiology.

## References

- [1] S. Kreibich and W.-D. Hardt. “Experimental approaches to phenotypic diversity in infection”. In: *Current Opinion in Microbiology* 27 (Oct. 2015), pp. 25–36. DOI: 10.1016/j.mib.2015.06.007.
- [2] H.-J. Wu, A. H.-J. Wang, and M. P. Jennings. “Discovery of virulence factors of pathogenic bacteria”. In: *Current Opinion in Chemical Biology* 12.1 (Feb. 2008), pp. 93–101. DOI: 10.1016/j.cbpa.2008.01.023.
- [3] B. B. Aldridge and K. Y. Rhee. “Microbial metabolomics: innovation, application, insight”. In: *Current Opinion in Microbiology* 19 (June 2014), pp. 90–96. DOI: 10.1016/j.mib.2014.06.009.
- [4] C. H. Johnson, J. Ivanisevic, and G. Siuzdak. “Metabolomics: beyond biomarkers and towards mechanisms”. In: *Nature Reviews Molecular Cell Biology* 17.7 (Mar. 2016), pp. 451–459. DOI: 10.1038/nrm.2016.25.

- [5] E. Deziel, F. Lépine, S. Milot, J. He, M. N. Mindrinos, R. G. Tompkins, and L. G. Rahme. “Analysis of *Pseudomonas aeruginosa* 4-hydroxy-2-alkylquinolines (HAQs) reveals a role for 4-hydroxy-2-heptylquinoline in cell-to-cell communication”. In: *Proceedings of the National Academy of Sciences* 101.5 (Jan. 2004), pp. 1339–1344. DOI: 10.1073/pnas.0307694100.
- [6] F. Lépine, S. Milot, E. Déziel, J. He, and L. G. Rahme. “Electrospray/mass spectrometric identification and analysis of 4-hydroxy-2-alkylquinolines (HAQs) produced by *Pseudomonas aeruginosa*”. In: *Journal of the American Society for Mass Spectrometry* 15.6 (June 2004), pp. 862–869. DOI: 10.1016/j.jasms.2004.02.012.
- [7] C. A. Ortori, J.-F. Dubern, S. R. Chhabra, M. Cámara, K. Hardie, P. Williams, and D. A. Barrett. “Simultaneous quantitative profiling of *N*-acyl-L-homoserine lactone and 2-alkyl-4(1*H*)-quinolone families of quorum-sensing signaling molecules using LC-MS/MS”. In: *Analytical and Bioanalytical Chemistry* 399.2 (Oct. 2010), pp. 839–850. DOI: 10.1007/s00216-010-4341-0.
- [8] D. V. Mavrodi, W. Blankenfeldt, and L. S. Thomashow. “Phenazine Compounds in Fluorescent *Pseudomonas* Spp.: Biosynthesis and Regulation”. In: *Annual Review of Phytopathology* 44.1 (Sept. 2006), pp. 417–445. DOI: 10.1146/annurev.phyto.44.013106.145710.
- [9] M. Mentel, E. G. Ahuja, D. V. Mavrodi, R. Breinbauer, L. S. Thomashow, and W. Blankenfeldt. “Of Two Make One: The Biosynthesis of Phenazines”. In: *ChemBioChem* 10.14 (Sept. 2009), pp. 2295–2304. DOI: 10.1002/cbic.200900323.
- [10] N. Guttenberger, W. Blankenfeldt, and R. Breinbauer. “Recent developments in the isolation, biological function, biosynthesis, and synthesis of phenazine natural products”. In: *Bioorganic & Medicinal Chemistry* 25.22 (Nov. 2017), pp. 6149–6166. DOI: 10.1016/j.bmc.2017.01.002.
- [11] L. Zulianello, C. Canard, T. Kohler, D. Caille, J.-S. Lacroix, and P. Meda. “Rhamnolipids Are Virulence Factors That Promote Early Infiltration of Primary Human Airway Epithelia by *Pseudomonas aeruginosa*”. In: *Infection and Immunity* 74.6 (May 2006), pp. 3134–3147. DOI: 10.1128/iai.01772-05.
- [12] A. M. Abdel-Mawgoud, F. Lépine, and E. Déziel. “Rhamnolipids: diversity of structures, microbial origins and roles”. In: *Applied Microbiology and Biotechnology* 86.5 (Mar. 2010), pp. 1323–1336. DOI: 10.1007/s00253-010-2498-2.
- [13] C. Ruttkies, E. L. Schymanski, S. Wolf, J. Hollender, and S. Neumann. “MetFrag relaunched: incorporating strategies beyond in silico fragmentation”. In: *Journal of Cheminformatics* 8.1 (Jan. 2016). DOI: 10.1186/s13321-016-0115-9.
- [14] K. Dührkop, H. Shen, M. Meusel, J. Rousu, and S. Böcker. “Searching molecular structure databases with tandem mass spectra using CSI:FingerID”. In: *Proceedings of the National Academy of Sciences* 112.41 (Sept. 2015), pp. 12580–12585. DOI: 10.1073/pnas.1509788112.
- [15] M. Wang et al. “Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking”. In: *Nature Biotechnology* 34.8 (Aug. 2016), pp. 828–837. DOI: 10.1038/nbt.3597.
- [16] P. N. Jimenez, G. Koch, J. A. Thompson, K. B. Xavier, R. H. Cool, and W. J. Quax. “The Multiple Signaling Systems Regulating Virulence in *Pseudomonas aeruginosa*”. In: *Microbiology and Molecular Biology Reviews* 76.1 (Mar. 2012), pp. 46–65. DOI: 10.1128/mnbr.05007-11.

- [17] S. P. Diggle, S. Matthijs, V. J. Wright, M. P. Fletcher, S. R. Chhabra, I. L. Lamont, X. Kong, R. C. Hider, P. Cornelis, M. Cámara, and P. Williams. “The *Pseudomonas aeruginosa* 4-Quinolone Signal Molecules HHQ and PQS Play Multifunctional Roles in Quorum Sensing and Iron Entrapment”. In: *Chemistry & Biology* 14.1 (Jan. 2007), pp. 87–96. DOI: 10.1016/j.chembiol.2006.11.014.
- [18] F. J. Reen, M. J. Mooij, L. J. Holcombe, C. M. McSweeney, G. P. McGlacken, J. P. Morrissey, and F. O’Gara. “The *Pseudomonas* quinolone signal (PQS), and its precursor HHQ, modulate interspecies and interkingdom behaviour”. In: *FEMS Microbiology Ecology* 77.2 (May 2011), pp. 413–428. DOI: 10.1111/j.1574-6941.2011.01121.x.
- [19] D. Szamosvári and T. Böttcher. “An Unsaturated Quinolone *N*-Oxide of *Pseudomonas aeruginosa* Modulates Growth and Virulence of *Staphylococcus aureus*”. In: *Angewandte Chemie International Edition* 56.25 (May 2017), pp. 7271–7275. DOI: 10.1002/anie.201702944.
- [20] J. Li, W. Sun, M. Saalim, G. Wei, D. A. Zaleta-Pinet, and B. R. Clark. “Isolation of 2-Alkyl-4-quinolones with Unusual Side Chains from a Chinese *Pseudomonas aeruginosa* Isolate”. In: *Journal of Natural Products* 83.7 (June 2020), pp. 2294–2298. DOI: 10.1021/acs.jnatprod.0c00026.
- [21] C. Schütz and M. Empting. “Targeting the *Pseudomonas* quinolone signal quorum sensing system for the discovery of novel anti-infective pathoblockers”. In: *Beilstein Journal of Organic Chemistry* 14 (Oct. 2018), pp. 2627–2645. DOI: 10.3762/bjoc.14.241.
- [22] G. W. Lau, D. J. Hassett, H. Ran, and F. Kong. “The role of pyocyanin in *Pseudomonas aeruginosa* infection”. In: *Trends in Molecular Medicine* 10.12 (Dec. 2004), pp. 599–606. DOI: 10.1016/j.molmed.2004.10.002.
- [23] G. M. Denning, S. S. Iyer, K. J. Reszka, Y. O’Malley, G. T. Rasmussen, and B. E. Britigan. “Phenazine-1-carboxylic acid, a secondary metabolite of *Pseudomonas aeruginosa*, alters expression of immunomodulatory proteins by human airway epithelial cells”. In: *American Journal of Physiology-Lung Cellular and Molecular Physiology* 285.3 (Sept. 2003), pp. L584–L592. DOI: 10.1152/ajplung.00086.2003.
- [24] C. C. Caldwell, Y. Chen, H. S. Goetzmann, Y. Hao, M. T. Borchers, D. J. Hassett, L. R. Young, D. Mavrodi, L. Thomashow, and G. W. Lau. “*Pseudomonas aeruginosa* Exotoxin Pyocyanin Causes Cystic Fibrosis Airway Pathogenesis”. In: *The American Journal of Pathology* 175.6 (Dec. 2009), pp. 2473–2488. DOI: 10.2353/ajpath.2009.090166.
- [25] B. Cezairliyan, N. Vinayavekhin, D. Grenfell-Lee, G. J. Yuen, A. Saghatelian, and F. M. Ausubel. “Identification of *Pseudomonas aeruginosa* Phenazines that Kill *Caenorhabditis elegans*”. In: *PLoS Pathogens* 9.1 (Jan. 2013). Ed. by D. S. Schneider, e1003101. DOI: 10.1371/journal.ppat.1003101.
- [26] J. Gibson, A. Sood, and D. A. Hogan. “*Pseudomonas aeruginosa*-*Candida albicans* Interactions: Localization and Fungal Toxicity of a Phenazine Derivative”. In: *Applied and Environmental Microbiology* 75.2 (Nov. 2008), pp. 504–513. DOI: 10.1128/aem.01037-08.
- [27] B. Briard, P. Bomme, B. E. Lechner, G. L. A. Mislin, V. Lair, M.-C. Prévost, J.-P. Latgé, H. Haas, and A. Beauvais. “*Pseudomonas aeruginosa* manipulates redox and iron homeostasis of its microbiota partner *Aspergillus fumigatus* via phenazines”. In: *Scientific Reports* 5.1 (Feb. 2015). DOI: 10.1038/srep08220.
- [28] L. E. P. Dietrich, A. Price-Whelan, A. Petersen, M. Whiteley, and D. K. Newman. “The phenazine pyocyanin is a terminal signalling factor in the quorum sensing network of *Pseudomonas aeruginosa*”. In: *Molecular Microbiology* 61.5 (Sept. 2006), pp. 1308–1321. DOI: 10.1111/j.1365-2958.2006.05306.x.

- [29] I. Ramos, L. E. Dietrich, A. Price-Whelan, and D. K. Newman. “Phenazines affect biofilm formation by *Pseudomonas aeruginosa* in similar ways at various scales”. In: *Research in Microbiology* 161.3 (Apr. 2010), pp. 187–191. DOI: 10.1016/j.resmic.2010.01.003.
- [30] T. Das, S. K. Kutty, R. Tavallaie, A. I. Ibugo, J. Panchompoo, S. Sehar, L. Aldous, A. W. S. Yeung, S. R. Thomas, N. Kumar, J. J. Gooding, and M. Manefield. “Phenazine virulence factor binding to extracellular DNA is important for *Pseudomonas aeruginosa* biofilm formation”. In: *Scientific Reports* 5.1 (Feb. 2015). DOI: 10.1038/srep08398.
- [31] K. T. Schiessl, F. Hu, J. Jo, S. Z. Nazia, B. Wang, A. Price-Whelan, W. Min, and L. E. P. Dietrich. “Phenazine production promotes antibiotic tolerance and metabolic heterogeneity in *Pseudomonas aeruginosa* biofilms”. In: *Nature Communications* 10.1 (Feb. 2019). DOI: 10.1038/s41467-019-08733-w.
- [32] Y. Wang, S. E. Kern, and D. K. Newman. “Endogenous Phenazine Antibiotics Promote Anaerobic Survival of *Pseudomonas aeruginosa* via Extracellular Electron Transfer”. In: *Journal of Bacteriology* 192.1 (Oct. 2009), pp. 365–369. DOI: 10.1128/jb.01188-09.
- [33] C. Kuhl, R. Tautenhahn, C. Böttcher, T. R. Larson, and S. Neumann. “CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets”. In: *Analytical Chemistry* 84.1 (Dec. 2011), pp. 283–289. DOI: 10.1021/ac202450g.
- [34] M. Sindelar and G. J. Patti. “Chemical Discovery in the Era of Metabolomics”. In: *Journal of the American Chemical Society* 142.20 (Apr. 2020), pp. 9097–9105. DOI: 10.1021/jacs.9b13198.
- [35] D. C. Sévin, A. Kuehne, N. Zamboni, and U. Sauer. “Biological insights through non-targeted metabolomics”. In: *Current Opinion in Biotechnology* 34 (Aug. 2015), pp. 1–8. DOI: 10.1016/j.copbio.2014.10.001.
- [36] L. Zhang, T. A. Veres-Schalnat, A. Somogyi, J. E. Pemberton, and R. M. Maier. “Fatty Acid Cosubstrates Provide  $\beta$ -Oxidation Precursors for Rhamnolipid Biosynthesis in *Pseudomonas aeruginosa*, as Evidenced by Isotope Tracing and Gene Expression Assays”. In: *Applied and Environmental Microbiology* 78.24 (Oct. 2012), pp. 8611–8622. DOI: 10.1128/aem.02111-12.
- [37] T. A. Gould, H. P. Schweizer, and M. E. A. Churchill. “Structure of the *Pseudomonas aeruginosa* acyl-homoserinelactone synthase LasI”. In: *Molecular Microbiology* 53.4 (July 2004), pp. 1135–1146. DOI: 10.1111/j.1365-2958.2004.04211.x.
- [38] M. R. Parsek, D. L. Val, B. L. Hanzelka, J. E. Cronan, and E. P. Greenberg. “Acyl homoserine-lactone quorum-sensing signal generation.” In: *Proceedings of the National Academy of Sciences of the United States of America* 96 (8 Apr. 1999), pp. 4360–4365. ISSN: 0027-8424.
- [39] T. Depke, R. Franke, and M. Brönstrup. “Clustering of MS<sup>2</sup> spectra using unsupervised methods to aid the identification of secondary metabolites from *Pseudomonas aeruginosa*”. In: *Journal of Chromatography B* 1071 (Dec. 2017), pp. 19–28. DOI: 10.1016/j.jchromb.2017.06.002.
- [40] T. Depke, J. G. Thöming, A. Kordes, S. Häussler, and M. Brönstrup. “Untargeted LC-MS Metabolomics Differentiates Between Virulent and Avirulent Clinical Strains of *Pseudomonas aeruginosa*”. In: *Biomolecules* 10.7 (July 2020), p. 1041. DOI: 10.3390/biom10071041.
- [41] T. Depke, S. Häussler, and M. Brönstrup. “The Peptide Chain Release Factor Methyltransferase PrmC Influences the *Pseudomonas aeruginosa* PA14 Endo- and Exometabolome”. In: *Metabolites* 10.10 (Oct. 2020), p. 417. DOI: 10.3390/metabo10100417.

- [42] A. Agarwal, C. Kahyaoglu, and D. B. Hansen. “Characterization of HmqF, a Protein Involved in the Biosynthesis of Unsaturated Quinolones Produced by *Burkholderia thailandensis*”. In: *Biochemistry* 51.8 (Feb. 2012), pp. 1648–1657. DOI: 10.1021/bi201625w.
- [43] L. Vial, F. Lépine, S. Milot, M.-C. Groleau, V. Dekimpe, D. E. Woods, and E. Déziel. “*Burkholderia pseudomallei*, *B. thailandensis*, and *B. ambifaria* Produce 4-Hydroxy-2-Alkylquinoline Analogues with a Methyl Group at the 3 Position That Is Required for Quorum-Sensing Regulation”. In: *Journal of Bacteriology* 190.15 (June 2008), pp. 5339–5352. DOI: 10.1128/jb.00400-08.
- [44] P. M. L. Coulon, M.-C. Groleau, and E. Déziel. “Potential of the *Burkholderia cepacia* Complex to Produce 4-Hydroxy-3-Methyl-2-Alkylquinolines”. In: *Frontiers in Cellular and Infection Microbiology* 9 (Feb. 2019). DOI: 10.3389/fcimb.2019.00033.
- [45] D. Szamosvári, M. Prothiwa, C. L. Dieterich, and T. Böttcher. “Profiling structural diversity and activity of 2-alkyl-4(1*H*)-quinolone *N*-oxides of *Pseudomonas* and *Burkholderia*”. In: *Chemical Communications* 56.47 (2020), pp. 6328–6331. DOI: 10.1039/d0cc02498h.
- [46] S. L. Drees and S. Fetzner. “PqsE of *Pseudomonas aeruginosa* Acts as Pathway-Specific Thioesterase in the Biosynthesis of Alkylquinolone Signaling Molecules”. In: *Chemistry & Biology* 22.5 (May 2015), pp. 611–618. DOI: 10.1016/j.chembio.2015.04.012.
- [47] G. Rampioni, M. Falcone, S. Heeb, E. Frangipani, M. P. Fletcher, J.-F. Dubern, P. Visca, L. Leoni, M. Cámara, and P. Williams. “Unravelling the Genome-Wide Contributions of Specific 2-Alkyl-4-Quinolones and PqsE to Quorum Sensing in *Pseudomonas aeruginosa*”. In: *PLOS Pathogens* 12.11 (Nov. 2016). Ed. by E. Pesci, e1006029. DOI: 10.1371/journal.ppat.1006029.
- [48] M. Zender, F. Witzgall, S. L. Drees, E. Weidel, C. K. Maurer, S. Fetzner, W. Blankenfeldt, M. Empting, and R. W. Hartmann. “Dissecting the Multiple Roles of PqsE in *Pseudomonas aeruginosa* Virulence by Discovery of Small Tool Compounds”. In: *ACS Chemical Biology* 11.6 (Apr. 2016), pp. 1755–1763. DOI: 10.1021/acscchembio.6b00156.
- [49] S. Mukherjee, D. A. Moustafa, V. Stergioula, C. D. Smith, J. B. Goldberg, and B. L. Bassler. “The PqsE and RhlR proteins are an autoinducer synthase–receptor pair that control virulence and biofilm development in *Pseudomonas aeruginosa*”. In: *Proceedings of the National Academy of Sciences* (Sept. 2018), p. 201814023. DOI: 10.1073/pnas.1814023115.
- [50] A. R. McCready, J. E. Paczkowski, J.-P. Cong, and B. L. Bassler. “An autoinducer-independent RhlR quorum-sensing receptor enables analysis of RhlR regulation”. In: *PLOS Pathogens* 15.6 (June 2019). Ed. by V. T. Lee, e1007820. DOI: 10.1371/journal.ppat.1007820.
- [51] M.-C. Groleau, T. de Oliveira Pereira, V. Dekimpe, and E. Déziel. “PqsE Is Essential for RhlR-Dependent Quorum Sensing Regulation in *Pseudomonas aeruginosa*”. In: *mSystems* 5.3 (May 2020). Ed. by E. A. Shank. DOI: 10.1128/msystems.00194-20.
- [52] C. Pustelny, S. Brouwer, M. Müsken, A. Bielecka, A. Dötsch, M. Nimtz, and S. Häussler. “The peptide chain release factor methyltransferase PrmC is essential for pathogenicity and environmental adaptation of *Pseudomonas aeruginosa* PA14”. In: *Environmental Microbiology* 15.2 (Dec. 2012), pp. 597–609. DOI: 10.1111/1462-2920.12040.
- [53] J. Krueger, S. Pohl, M. Preusse, A. Kordes, N. Rugen, M. Schniederjans, A. Pich, and S. Häussler. “Unravelling post-transcriptional PrmC-dependent regulatory mechanisms in *Pseudomonas aeruginosa*”. In: *Environmental Microbiology* 18.10 (July 2016), pp. 3583–3592. DOI: 10.1111/1462-2920.13435.



- [54] K. Hornischer, A. Khaledi, S. Pohl, M. Schniederjans, L. Pezoldt, F. Casilag, U. Muthukumarasamy, S. Bruchmann, J. Thöming, A. Kordes, and S. Häussler. “BACTOME—a reference database to explore the sequence- and gene expression-variation landscape of *Pseudomonas aeruginosa* clinical isolates”. In: *Nucleic Acids Research* 47.D1 (Oct. 2018), pp. D716–D720. DOI: 10.1093/nar/gky895.
- [55] A. Casadevall. “The Pathogenic Potential of a Microbe”. In: *mSphere* 2.1 (Feb. 2017). Ed. by J. A. Alspaugh. DOI: 10.1128/msphere.00015-17.
- [56] J. B. Lyczak, C. L. Cannon, and G. B. Pier. “Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist”. In: *Microbes and infection* 2.9 (2000), pp. 1051–1060.
- [57] G. B. Rogers, L. R. Hoffman, M. Whiteley, T. W. Daniels, M. P. Carroll, and K. D. Bruce. “Revealing the dynamics of polymicrobial infections: implications for antibiotic therapy”. In: *Trends in Microbiology* 18.8 (Aug. 2010), pp. 357–364. DOI: 10.1016/j.tim.2010.04.005.
- [58] G. Jander, L. G. Rahme, and F. M. Ausubel. “Positive Correlation between Virulence of *Pseudomonas aeruginosa* Mutants in Mice and Insects”. In: *Journal of Bacteriology* 182.13 (July 2000), pp. 3843–3845. DOI: 10.1128/jb.182.13.3843-3845.2000.
- [59] J. G. Thöming, J. Tomasch, M. Preusse, M. Koska, N. Grahl, S. Pohl, S. D. Willger, V. Kaefer, M. Müsken, and S. Häussler. “Parallel evolutionary paths to produce more than one *Pseudomonas aeruginosa* biofilm phenotype”. In: *npj Biofilms and Microbiomes* 6.1 (Jan. 2020). DOI: 10.1038/s41522-019-0113-6.
- [60] E. Frimmersdorf, S. Horatzek, A. Pelnikevich, L. Wiehlmann, and D. Schomburg. “How *Pseudomonas aeruginosa* adapts to various environments: a metabolomic approach”. In: *Environmental Microbiology* 12.6 (Feb. 2010), pp. 1734–1747. DOI: 10.1111/j.1462-2920.2010.02253.x.
- [61] E. E. Bardin, S. J. S. Cameron, A. Perdones-Montero, K. Hardiman, F. Bolt, E. W. F. W. Alton, A. Bush, J. C. Davies, and Z. Takáts. “Metabolic Phenotyping and Strain Characterisation of *Pseudomonas aeruginosa* Isolates from Cystic Fibrosis Patients Using Rapid Evaporative Ionisation Mass Spectrometry”. In: *Scientific Reports* 8.1 (July 2018). DOI: 10.1038/s41598-018-28665-7.
- [62] R. A. Quinn, V. V. Phelan, K. L. Whiteson, N. Garg, B. A. Bailey, Y. W. Lim, D. J. Conrad, P. C. Dorrestein, and F. L. Rohwer. “Microbial, host and xenobiotic diversity in the cystic fibrosis sputum metabolome”. In: *The ISME Journal* 10.6 (Dec. 2015), pp. 1483–1498. DOI: 10.1038/ismej.2015.207.
- [63] R. A. Quinn, S. Adem, R. H. Mills, W. Comstock, L. D. Goldasich, G. Humphrey, A. A. Aksenov, A. V. Melnik, R. da Silva, G. Ackermann, N. Bandeira, D. J. Gonzalez, D. Conrad, A. J. O’Donoghue, R. Knight, and P. C. Dorrestein. “Neutrophilic proteolysis in the cystic fibrosis lung correlates with a pathogenic microbiome”. In: *Microbiome* 7.1 (Feb. 2019). DOI: 10.1186/s40168-019-0636-3.
- [64] H. Thiele, G. McLeod, M. Niemitz, and T. Kühn. “Structure verification of small molecules using mass spectrometry and NMR spectroscopy”. In: *Monatshefte für Chemie - Chemical Monthly* 142.7 (Apr. 2011), pp. 717–730. DOI: 10.1007/s00706-011-0486-6.
- [65] N. G. Mahieu, X. Huang, Y.-J. Chen, and G. J. Patti. “Credentialing Features: A Platform to Benchmark and Optimize Untargeted Metabolomic Methods”. In: *Analytical Chemistry* 86.19 (Sept. 2014), pp. 9583–9589. DOI: 10.1021/ac503092d.
- [66] W. Scherber. “Stable Isotope Labeling to Improve Metabolite Identification in Untargeted Metabolomics of Pathogenic Bacteria”. Master’s Thesis. Hochschule Aalen – Technik und Wirtschaft, 2020.

- [67] M. Wang, A. K. Jarmusch, F. Vargas, A. A. Aksenov, J. M. Gauglitz, K. Weldon, D. Petras, R. da Silva, R. Quinn, A. V. Melnik, J. J. J. van der Hooft, A. M. Caraballo-Rodríguez, L. F. Nothias, C. M. Aceves, M. Panitchpakdi, E. Brown, F. D. Ottavio, N. Sikora, E. O. Elijah, L. Labarta-Bajo, E. C. Gentry, S. Shalapour, K. E. Kyle, S. P. Puckett, J. D. Watrous, C. S. Carpenter, A. Bouslimani, M. Ernst, A. D. Swafford, E. I. Zú niga, M. J. Balunas, J. L. Klassen, R. Loomba, R. Knight, N. Bandeira, and P. C. Dorrestein. “Mass spectrometry searches using MASST”. In: *Nature Biotechnology* 38.1 (Jan. 2020), pp. 23–26. DOI: 10.1038/s41587-019-0375-9.
- [68] J. Xia, D. I. Broadhurst, M. Wilson, and D. S. Wishart. “Translational biomarker discovery in clinical metabolomics: an introductory tutorial”. In: *Metabolomics* 9.2 (Dec. 2012), pp. 280–299. DOI: 10.1007/s11306-012-0482-9.
- [69] M. N. Abdullah, B. W. Yap, Y. Zakaria, and A. B. A. Majeed. “Metabolites Selection and Classification of Metabolomics Data on Alzheimer’s Disease Using Random Forest”. In: *Communications in Computer and Information Science*. Springer Singapore, 2016, pp. 100–112. DOI: 10.1007/978-981-10-2777-2\_9.
- [70] C. F. O. R. Melo, L. C. Navarro, D. N. de Oliveira, T. M. Guerreiro, E. de Oliveira Lima, J. Delafiori, M. Z. Dabaja, M. da Silva Ribeiro, M. de Menezes, R. G. M. Rodrigues, K. N. Morishita, C. Z. Esteves, A. L. L. de Amorim, C. T. Aoyagui, P. L. Parise, G. P. Milanez, G. M. do Nascimento, A. R. R. Freitas, R. Angerami, F. T. M. a. Costa, C. W. Arns, M. R. Resende, E. Amaral, R. P. Junior, C. C. Ribeiro-do-Valle, H. Milanez, M. L. Moretti, J. L. Proenca-Modena, S. Avila, A. Rocha, and R. R. Catharino. “A Machine Learning Application Based in Random Forest for Integrating Mass Spectrometry-Based Metabolomic Data: A Simple Screening Method for Patients With Zika Virus”. In: *Frontiers in Bioengineering and Biotechnology* 6 (Apr. 2018). DOI: 10.3389/fbioe.2018.00031.
- [71] M. H. Ghaffari, A. Jahanbekam, H. Sadri, K. Schuh, G. Dusel, C. Prehn, J. Adamski, C. Koch, and H. Sauerwein. “Metabolomics meets machine learning: Longitudinal metabolite profiling in serum of normal versus overconditioned cows and pathway analysis”. In: *Journal of Dairy Science* 102.12 (Dec. 2019), pp. 11561–11585. DOI: 10.3168/jds.2019-17114.
- [72] V. Oza, J. Aicher, and L. Reed. “Random Forest Analysis of Untargeted Metabolomics Data Suggests Increased Use of Omega Fatty Acid Oxidation Pathway in Drosophila Melanogaster Larvae Fed a Medium Chain Fatty Acid Rich High-Fat Diet”. In: *Metabolites* 9.1 (Dec. 2018), p. 5. DOI: 10.3390/metabo9010005.
- [73] M. R. Berthold, C. Borgelt, F. Höppner, and F. Klawonn. *Guide to intelligent data analysis: how to intelligently make sense of real data*. Springer Science & Business Media, 2010.
- [74] I. Blaženović, T. Kind, J. Ji, and O. Fiehn. “Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics”. In: *Metabolites* 8.2 (May 2018), p. 31. DOI: 10.3390/metabo8020031.
- [75] J.-L. Wolfender, J.-M. Nuzillard, J. J. J. van der Hooft, J.-H. Renault, and S. Bertrand. “Accelerating Metabolite Identification in Natural Product Research: Toward an Ideal Combination of Liquid Chromatography–High-Resolution Tandem Mass Spectrometry and NMR Profiling, in Silico Databases, and Chemometrics”. In: *Analytical Chemistry* 91.1 (Nov. 2018), pp. 704–742. DOI: 10.1021/acs.analchem.8b05112.
- [76] B. B. Misra and S. Mohapatra. “Tools and resources for metabolomics research community: A 2017-2018 update”. In: *ELECTROPHORESIS* 40.2 (Nov. 2018), pp. 227–246. DOI: 10.1002/e1ps.201800428.

- [77] S. Rogers, C. W. Ong, J. Wandy, M. Ernst, L. Ridder, and J. J. J. van der Hooft. “Deciphering complex metabolite mixtures by unsupervised and supervised substructure discovery and semi-automated annotation from MS/MS spectra”. In: *Faraday Discussions* 218 (2019), pp. 284–302. DOI: 10.1039/c8fd00235e.
- [78] J. Stanstrup, C. Broeckling, R. Helmus, N. Hoffmann, E. Mathé, T. Naake, L. Nicolotti, K. Peters, J. Rainer, R. Salek, T. Schulze, E. Schymanski, M. Stravs, E. Thévenot, H. Treutler, R. Weber, E. Willighagen, M. Witting, and S. Neumann. “The metaRbolomics Toolbox in Bioconductor and beyond”. In: *Metabolites* 9.10 (Sept. 2019), p. 200. DOI: 10.3390/metabo9100200.
- [79] N. G. Mahieu, J. L. Genenbacher, and G. J. Patti. “A roadmap for the XCMS family of software solutions in metabolomics”. In: *Current Opinion in Chemical Biology* 30 (Feb. 2016), pp. 87–93. DOI: 10.1016/j.cbpa.2015.11.009.
- [80] S. Aros-Calt, B. H. Muller, S. Boudah, C. Ducruix, G. Gervasi, C. Junot, and F. Fenaille. “Annotation of the *Staphylococcus aureus* Metabolome Using Liquid Chromatography Coupled to High-Resolution Mass Spectrometry and Application to the Study of Methicillin Resistance”. In: *Journal of Proteome Research* 14.11 (Oct. 2015), pp. 4863–4875. DOI: 10.1021/acs.jproteome.5b00697.



## A | Supplementary Tables for Publication 1

3.S1	List of theoretical $m/z$ of alkyl quinolone from C1 to C21 with saturated, mono-unsaturated and doubly unsaturated side chain used as preferred mass list in the semi-targeted LC-MS/MS run. . . . .	322
3.S2	List of features that have been identified by comparison of exact mass, retention time and MS <sup>2</sup> fragmentation to our in-house library or putatively annotated by comparison of exact mass and MS <sup>2</sup> fragmentation to metabolite databases. . . . .	323
3.S3	List of cluster ID assignments resulting from density based clustering using the OPTICS algorithm along with the colour coding used in Figure 4. . . . .	331
3.S4	List of cluster ID assignments resulting from hierarchical clustering with average linkage of product ion spectra similarities. . . . .	340
3.S5	List of cluster ID assignments resulting from hierarchical clustering with average linkage of neutral loss pattern similarities. . . . .	348
3.S6	List of cluster ID assignments resulting from hierarchical clustering with average linkage of product ion spectra similarities of the spectra acquired in the alkyl quinolone-biased semi-targeted analysis. . . . .	357
3.S7	List of putatively identified features, their sum formula and isotopic pattern fit in mSigma as determined by the SmartFormula functionality of the Bruker DataAnalysis software. . . . .	369

**Table 3.S1.:** List of theoretical  $m/z$  of alkyl quinolone from C1 to C21 with saturated, mono-unsaturated and doubly unsaturated side chain used as preferred mass list in the semi-targeted LC-MS/MS run.

mass	species	mass	species	mass	species
146.0601	C0-HQ	284.1645	C9:2-QNO	370.2741	C15:1-QNO
160.0757	C1-HQ	284.2009	C10:1-HQ	370.3105	C16-HQ
162.0550	C0-QNO	286.1802	C9:1-QNO	372.2897	C15-QNO
174.0914	C2-HQ	286.2166	C10-HQ	380.2948	C17:2-HQ
176.0706	C1-QNO	288.1958	C9-QNO	382.2741	C16:2-QNO
186.0914	C3:1-HQ	296.2009	C11:1-HQ	382.3105	C17:1-HQ
188.1070	C3-HQ	298.1802	C10:2-QNO	384.2897	C16:1-QNO
190.0863	C2-QNO	298.2166	C11:1-HQ	384.3261	C17-HQ
200.1070	C4:1-HQ	300.1958	C10:1-QNO	386.3054	C16-QNO
202.0863	C3:1-QNO	300.2322	C11-HQ	394.3105	C18:2-HQ
202.1227	C4-HQ	302.2115	C10-QNO	396.2897	C17:2-QNO
204.1019	C3-QNO	310.2166	C12:2-HQ	396.3261	C18:1-HQ
214.1227	C5:1-HQ	312.1958	C11:2-QNO	398.3054	C17:1-QNO
216.1019	C4:1-QNO	312.2322	C12:1-HQ	398.3418	C18-HQ
216.1383	C5-HQ	314.2115	C11:1-QNO	400.3210	C17-QNO
218.1176	C4-QNO	314.2479	C12-HQ	408.3261	C19:2-HQ
228.1383	C6:1-HQ	316.2271	C11-QNO	410.3054	C18:2-QNO
230.1176	C5:1-QNO	324.2322	C13:2-HQ	410.3418	C19:1-HQ
230.1540	C6-HQ	326.2115	C12:2-QNO	412.3210	C18:1-QNO
232.1332	C5-QNO	326.2479	C13:1-HQ	412.3574	C19-HQ
240.1383	C7:2-HQ	328.2271	C12:1-QNO	414.3367	C18-QNO
242.1540	C7:1-HQ	328.2635	C13-HQ	422.3418	C20:2-HQ
244.1332	C6:1-QNO	330.2428	C12-QNO	424.3210	C19:2-QNO
244.1696	C7-HQ	338.2479	C14:2-HQ	424.3574	C20:1-HQ
246.1489	C6-QNO	340.2271	C13:2-QNO	426.3367	C19:1-QNO
254.1540	C8:2-HQ	340.2635	C14:1-HQ	426.3731	C20-HQ
256.1332	C7:2-QNO	342.2428	C13:1-QNO	428.3523	C19-QNO
256.1696	C8:1-HQ	342.2792	C14-HQ	436.3574	C21:2-HQ
258.1489	C7:1-QNO	344.2584	C13-QNO	438.3367	C20:2-QNO
258.1853	C8-HQ	352.2635	C15:2-HQ	438.3731	C21:1-HQ
260.1645	C7-QNO	354.2428	C14:2-QNO	440.3523	C20:1-QNO
268.1696	C9:2-HQ	354.2792	C15:1-HQ	440.3887	C21-HQ
270.1489	C8:2-QNO	356.2584	C14:1-QNO	442.6963	C20-QNO
270.1853	C9:1-HQ	356.2948	C15-HQ	452.3523	C21:2-QNO
272.1645	C8:1-QNO	358.2741	C14-QNO	454.3680	C21:1-QNO
272.2009	C9-HQ	366.2792	C16-HQ	456.3836	C21-QNO
274.1802	C8-QNO	368.2584	C15:2-QNO		
282.1853	C10:2-HQ	368.2948	C16:1-HQ		

**Table 3.S2.:** List of features that have been identified by comparison of exact mass, retention time and MS<sup>2</sup> fragmentation to our in-house library or putatively annotated by comparison of exact mass and MS<sup>2</sup> fragmentation to metabolite databases.

feature ID	med.mz	med.rt	library annotation	putative annotation
M146.17T59.35	146.1653	59.4	spermidine	NA
M129.14T58.57	129.1387	58.6	spermidine (fragment)	NA
M112.11T57.8	112.1119	57.8	spermidine (fragment)	NA
M251.16T60.64	251.1603	60.6	NA	NA
M212.85T65.02	212.8519	65.0	NA	NA
M290.85T64.76	290.8472	64.8	NA	NA
M148.06T69.65	148.0605	69.6	glutamate	NA
M130.05T69.64	130.0500	69.6	glutamate (fragment)	NA
M179.06T71.32	179.0550	71.3	NA	gluconolactone
M197.07T71.57	197.0656	71.6	NA	NA
M116.07T73.51	116.0706	73.5	NA	proline
M242.08T74.02	242.0787	74.0	NA	NA
M324.06T75.32	324.0591	75.3	CMP	NA
M301.11T76.88	301.1141	76.9	NA	NA
M219.11T77.65	219.0974	77.7	NA	Glu Ala
M325.04T78.94	325.0429	78.9	CMP	NA
M428.04T80.88	428.0363	80.9	ADP	NA
M162.08T82.05	162.0761	82.1	NA	NA
M191.05T84.37	191.0483	84.4	NA	NA
M308.06T84.63	308.0636	84.6	dCMP	NA
M189.12T85.8	189.1231	85.8	NA	NA
M85.06T86.7	85.0586	86.7	NA	NA
M372.55T88.26	372.5456	88.3	NADP (2+)	NA
M348.07T90.34	348.0707	90.3	AMP	NA
M136.06T90.6	136.0617	90.6	AMP (fragment)	NA
M695.13T91.88	695.1331	91.9	AMP (2+)	NA
M316.16T94.47	316.1622	94.5	NA	NA
M338.14T94.98	338.1439	95.0	NA	NA
M364.07T97.19	364.0659	97.2	GMP	NA
M299.14T96.54	299.1355	96.5	NA	NA
M99.09T97.84	99.0916	97.8	NA	NA
M254.09T100.71	254.0882	100.7	NA	NA
M110.06T100.45	110.0599	100.4	2-aminophenol	NA
M123.06T103.31	123.0553	103.3	nicotinamide	NA
M159.11T102.26	159.1127	102.3	NA	NA
M317.14T103.57	317.1448	103.6	NA	NA
M290.13T105.39	290.1344	105.4	NA	NA
M332.56T107.48	332.5621	107.5	NAD (2+)	NA
M664.12T108.26	664.1161	108.3	NAD	NA
M542.07T108.52	542.0682	108.5	NAD (fragment)	NA
M278.57T112.67	278.5725	112.7	NA	NA
M137.05T112.16	137.0457	112.2	hypoxanthine	NA
M332.08T113.21	332.0753	113.2	NA	dAMP
M179.07T114.5	179.0672	114.5	NA	NA
M202.18T115.29	202.1798	115.3	NA	NA
M190.07T117.36	190.0708	117.4	N-acetylglutamate	NA
M535.19T118.41	535.1876	118.4	NA	Glu Glu Glu Glu
M180.05T120.87	180.0513	120.9	NA	(iso)xanthopterin
M132.1T122.82	132.1021	122.8	leucine / isoleucine / norleucine	NA
M307.08T126.2	307.0838	126.2	NA	glutathion disulphide (2+)
M182.08T125.93	182.0812	125.9	tyrosine	NA
M613.16T127.23	613.1591	127.2	glutathion disulfide	NA
M322.11T131.92	322.1068	131.9	NA	NA
M304.06T132.18	304.0595	132.2	NA	NA
M323.06T135.83	323.0641	135.8	NA	dTMP
M143.08T138.17	143.0812	138.2	NA	NA
M98.98T141.84	98.9839	141.8	NA	NA
M182.07T141.05	182.0672	141.1	NA	NA



Table 3.S2.: Continued.

feature ID	med.mz	med.rt	library annotation	putative annotation
M388.11T146.83	388.1100	146.8	NA	NA
M164.06T147.09	164.0567	147.1	NA	pt erine
M193.07T150.49	193.0683	150.5	NA	S-(5'-adenosyl)-homocysteine (2+)
M385.13T151.52	385.1290	151.5	S-(5'-adenosyl)-homocysteine	NA
M250.14T156.21	250.1412	156.2	NA	NA
M125.57T156.99	125.5742	157.0	NA	NA
M664.23T165.1	664.2311	165.1	NA	NA
M332.62T165.35	332.6200	165.4	NA	NA
M351.59T166.13	351.5930	166.1	NA	NA
M268.1T171.87	268.1043	171.9	adenosine	NA
M314.09T175.26	314.0913	175.3	NA	NA
M245.18T210.61	245.1762	210.6	NA	NA
M251.15T228.5	251.1504	228.5	NA	NA
M166.09T233.22	166.0864	233.2	phenylalanine	NA
M120.08T233.48	120.0808	233.5	phenylalanine (fragment)	NA
M340.06T231.65	340.0627	231.6	NA	NA
M320.17T232.69	320.1711	232.7	NA	NA
M103.05T235.3	103.0542	235.3	phenylalanine (fragment)	NA
M397.14T289.26	397.1418	289.3	NA	NA
M793.27T273.95	793.2749	273.9	NA	NA
M416.12T271.35	416.1156	271.3	NA	NA
M174.06T277.59	174.0554	277.6	NA	NA
M262.16T278.37	262.1556	278.4	NA	NA
M276.11T276.54	276.1088	276.5	NA	NA
M254.16T280.46	254.1618	280.5	NA	NA
M244.11T293.5	244.1118	293.5	NA	NA
M323.07T305.37	323.0706	305.4	NA	NA
M256.18T309.56	256.1765	309.6	NA	NA
M911.28T323.37	911.2844	323.4	NA	NA
M727.21T324.67	727.2052	324.7	NA	NA
M1235.4T324.66	1235.3957	324.7	NA	NA
M461.66T328.59	461.6627	328.6	NA	NA
M922.32T329.89	922.3165	329.9	NA	NA
M480.64T329.62	480.6359	329.6	NA	NA
M479.19T333.77	479.1887	333.8	NA	NA
M472.65T332.73	472.6535	332.7	NA	NA
M597.68T335.59	597.6785	335.6	NA	UDP-muramyl-pentapeptide
M219.13T335.84	219.1344	335.8	NA	Ser Leu
M382.65T337.93	382.6472	337.9	NA	NA
M220.12T336.88	220.1182	336.9	NA	panthotenate
M568.14T338.32	568.1385	338.3	NA	NA
M194.08T340.15	194.0788	340.1	NA	NA
M526.18T342.1	526.1834	342.1	NA	NA
M1051.36T342.36	1051.3583	342.4	NA	NA
M363.77T343.13	363.7746	343.1	NA	NA
M194.07T346.78	194.0676	346.8	NA	NA
M217.1T347.04	217.1050	347.0	NA	NA
M472.85T348.87	472.8508	348.9	NA	NA
M708.77T349.12	708.7715	349.1	NA	NA
M590.7T351.59	590.7044	351.6	NA	NA
M525.18T352.51	525.1832	352.5	NA	NA
M1180.4T351.98	1180.4000	352.0	NA	NA
M367.64T355.63	367.6418	355.6	NA	NA
M360.21T355.1	360.2133	355.1	NA	Ile Val Glu / Val Ile Glu
M734.27T356.41	734.2747	356.4	NA	NA
M235.07T357.45	235.0665	357.5	NA	NA
M655.23T359.27	655.2268	359.3	NA	NA
M686.76T358.74	686.7596	358.7	NA	NA
M666.22T359.78	666.2163	359.8	NA	NA
M174.06T360.96	174.0554	361.0	NA	NA
M679.29T361.61	679.2916	361.6	NA	NA

Table 3.S2.: Continued.

feature ID	med.mz	med.rt	library annotation	putative annotation
M243.08T362.4	243.0771	362.4	NA	NA
M215.08T362.66	215.0815	362.7	NA	NA
M331.18T363.96	331.1775	364.0	NA	NA
M693.77T367.08	693.7669	367.1	NA	NA
M719.75T366.82	719.7471	366.8	NA	NA
M462.85T367.6	462.8479	367.6	NA	NA
M359.64T368.9	359.6449	368.9	NA	NA
M188.07T371.25	188.0710	371.3	tryptophan (fragment)	NA
M205.1T370.99	205.0975	371.0	tryptophan	NA
M270.19T371.24	270.1926	371.2	NA	NA
M304.18T373.32	304.1769	373.3	NA	NA
M784.27T374.37	784.2691	374.4	NA	NA
M795.26T374.75	795.2586	374.8	NA	NA
M694.26T376.97	694.2594	377.0	NA	NA
M358.2T377.48	358.1985	377.5	NA	Ile Pro Glu / Leu Pro Glu
M195.11T378.66	195.1130	378.7	NA	NA
M848.79T380.87	848.7917	380.9	NA	NA
M211.09T382.17	211.0874	382.2	pyocyanin	NA
M578.85T381.64	578.8462	381.6	NA	NA
M732.26T383.07	732.2604	383.1	NA	NA
M298.1T385.29	298.0978	385.3	5'-methylthioadenosine	NA
M261.13T385.55	261.1317	385.6	NA	NA
M913.31T388.16	913.3134	388.2	NA	NA
M187.12T391.55	187.1235	391.6	NA	NA
M204.12T389.98	204.1234	390.0	NA	NA
M243.18T392.85	243.1839	392.8	NA	NA
M258.58T397.4	258.5822	397.4	NA	NA
M977.84T395.18	977.8352	395.2	NA	NA
M516.16T395.97	516.1560	396.0	NA	NA
M364.62T397.4	364.6189	397.4	NA	NA
M728.23T398.3	728.2285	398.3	NA	NA
M188.12T399.72	188.1199	399.7	NA	NA
M254.09T400.89	254.0930	400.9	NA	NA
M295.13T406.65	295.1293	406.7	NA	Glu Phe
M350.62T409.24	350.6212	409.2	folic acid (fragment)	NA
M268.66T408.73	268.6585	408.7	NA	NA
M145.08T410.28	145.0764	410.3	NA	NA
M190.05T411.45	190.0504	411.4	kynurenate	NA
M214.13T411.31	214.1342	411.3	NA	NA
M378.2T412.48	378.2037	412.5	NA	Pro Tyr Val
M316.22T413.91	316.2235	413.9	NA	Val Val Val
M328.22T437.8	328.2236	437.8	NA	Pro Leu Val
M326.21T414.96	326.2080	415.0	NA	NA
M336.19T415.21	336.1920	415.2	NA	Phe Val Ala / Val Phe Ala
M295.19T417.02	295.1883	417.0	NA	NA
M289.12T417.28	289.1191	417.3	NA	NA
M261.12T419.11	261.1239	419.1	NA	NA
M197.13T423.26	197.1289	423.3	NA	NA
M786.17T426.25	786.1663	426.3	FAD	NA
M316.21T430.54	316.2128	430.5	NA	NA
M231.11T430.8	231.1133	430.8	NA	NA
M160.08T433.66	160.0762	433.7	NA	NA
M457.11T433.91	457.1121	433.9	NA	FMN
M204.1T435.6	204.1025	435.6	NA	NA
M195.09T436.77	195.0882	436.8	caffeine ISTD	NA
M291.15T444.85	291.1461	444.8	trimethoprim ISTD	NA
M275.11T445.88	275.1141	445.9	NA	NA
M263.14T447.96	263.1396	448.0	NA	Pro Phe
M257.15T449.13	257.1456	449.1	NA	NA
M120.04T450.56	120.0447	450.6	anthranilate (fragment)	NA
M138.06T451.33	138.0551	451.3	anthranilate	NA
M243.13T457.64	243.1345	457.6	NA	NA



Table 3.S2.: Continued.

feature ID	med.mz	med.rt	library annotation	putative annotation
M360.19T459.98	360.1930	460.0	NA	NA
M300.2T461.28	300.1994	461.3	NA	NA
M399.26T461.02	399.2608	461.0	NA	NA
M291.09T462.05	291.0946	462.1	trimethoprim ISTD	NA
M176.07T465.7	176.0708	465.7	NA	NA
M159.09T464.27	159.0916	464.3	NA	NA
M277.15T466.47	277.1540	466.5	NA	NA
M159.07T466.73	159.0678	466.7	NA	NA
M418.14T474.32	418.1393	474.3	NA	NA
M231.17T475.37	231.1701	475.4	NA	Leu Val
M491.29T496.41	491.2857	496.4	NA	NA
M344.25T476.52	344.2541	476.5	NA	Leu Leu Val or isomer
M255.08T482.73	255.0761	482.7	NA	NA
M211.14T497.45	211.1437	497.5	NA	cyclo(Leu Pro)
M342.24T486.25	342.2386	486.2	NA	Pro Ile Leu or isomer
M427.29T485.86	427.2907	485.9	NA	NA
M245.59T495.11	245.5915	495.1	NA	NA
M232.13T495.36	232.1331	495.4	NA	NA
M145.08T508.11	145.0762	508.1	NA	NA
M439.29T506.82	439.2913	506.8	NA	NA
M831.14T503.95	831.1424	504.0	NA	NA
M187.09T505	187.0865	505.0	NA	NA
M163.09T512.64	163.0866	512.6	NA	NA
M473.3T511.21	473.2970	511.2	NA	NA
M441.31T516.66	441.3069	516.7	NA	NA
M245.13T525.64	245.1278	525.6	NA	cyclo(Phe Pro)
M246.13T525.63	246.1308	525.6	NA	NA
M505.3T526.66	505.3001	526.7	NA	NA
M311.14T527.97	311.1390	528.0	NA	NA
M530.13T538.67	530.1333	538.7	NA	NA
M186.09T531.62	186.0911	531.6	NA	NA
M188.11T535.78	188.1070	535.8	NA	C3-HQ
M275.03T536.04	275.0306	536.0	NA	NA
M321.1T537.6	321.1014	537.6	NA	NA
M202.09T545.21	202.0858	545.2	NA	NA
M265.57T546.37	265.5707	546.4	NA	NA
M243.09T558.67	243.0876	558.7	lumichrome	NA
M204.1T564.12	204.1021	564.1	NA	NA
M260.16T583.7	260.1640	583.7	NA	C7-QNO
M775.9T572.93	775.9025	572.9	NA	NA
M392.25T573.83	392.2541	573.8	NA	Phe Leu Leu or isomer
M263.12T576.07	263.1178	576.1	NA	NA
M200.11T578.31	200.1066	578.3	NA	NA
M202.12T578.44	202.1224	578.4	NA	NA
M136.08T584.09	136.0755	584.1	NA	NA
M118.06T585.64	118.0647	585.6	NA	NA
M323.09T586.42	323.0880	586.4	NA	NA
M485.11T612.66	485.1122	612.7	NA	NA
M215.12T626.24	215.1175	626.2	NA	NA
M263.12T624.79	263.1177	624.8	NA	NA
M224.08T640.69	224.0815	640.7	phenazine-1-carboxamide	NA
M207.05T641.72	207.0547	641.7	phenazine-1-carboxamide (fragment)	NA
M393.12T645.89	393.1224	645.9	NA	NA
M289.14T649.01	289.1381	649.0	NA	NA
M213.07T652.92	213.0657	652.9	NA	NA
M260.16T651.62	260.1642	651.6	HQNO	NA
M297.14T656.95	297.1358	657.0	NA	NA
M186.13T666.88	186.1308	666.9	NA	NA
M214.12T663.32	214.1225	663.3	NA	C5:1-HQ
M216.14T670.01	216.1384	670.0	NA	C5-HQ
M227.08T670.26	227.0815	670.3	NA	NA
M264.18T675.46	264.1750	675.5	nortriptyline ISTD	NA

Table 3.S2.: Continued.

feature ID	med.mz	med.rt	library annotation	putative annotation
M233.13T676.23	233.1326	676.2	nortriptyline ISTD (fragment)	NA
M191.09T676.49	191.0857	676.5	nortriptyline ISTD (fragment)	NA
M276.16T680.26	276.1590	680.3	nortriptyline ISTD +Na	NA
M232.13T683.77	232.1333	683.8	NA	C5-QNO
M1012.42T688.19	1012.4196	688.2	NA	NA
M1012.54T689.35	1012.5450	689.3	NA	NA
M1157.05T687.92	1157.0497	687.9	NA	NA
M900.15T689.22	900.1518	689.2	NA	NA
M1156.91T690.64	1156.9069	690.6	NA	NA
M1012.29T689.74	1012.2926	689.7	NA	NA
M900.04T690	900.0408	690.0	NA	NA
M1156.76T690.25	1156.7629	690.3	NA	NA
M1012.67T693.51	1012.6712	693.5	NA	NA
M225.07T698.07	225.0660	698.1	phenazine-1-carboxylic acid	NA
M207.06T699.1	207.0556	699.1	phenazine-1-carboxylic acid (fragment)	NA
M257.06T704.3	257.0558	704.3	NA	NA
M226.18T703.76	226.1800	703.8	NA	NA
M288.2T705.59	288.1953	705.6	NA	C9-QNO
M269.06T708.74	269.0559	708.7	phenazine-1,6-dicarboxylic acid	NA
M325.07T739.09	325.0676	739.1	NA	pyochelin
M304.19T717.91	304.1904	717.9	NA	NA
M206.03T721.29	206.0268	721.3	NA	NA
M286.18T728.73	286.1799	728.7	NA	C9:1-QNO
M181.08T724.14	181.0761	724.1	NA	NA
M330.19T724.4	330.1930	724.4	NA	NA
M185.1T729.78	185.0960	729.8	NA	naproxen (fragment)
M307.02T733.71	307.0208	733.7	NA	NA
M404.23T734.62	404.2337	734.6	NA	NA
M230.15T735.39	230.1536	735.4	NA	C6-HQ
M255.08T740.91	255.0762	740.9	NA	NA
M317.09T742.2	317.0915	742.2	NA	NA
M446.19T745.32	446.1867	745.3	glipizide ISTD	NA
M468.17T744.02	468.1683	744.0	glipizide ISTD +Na	NA
M321.1T746.09	321.1019	746.1	NA	NA
M913.35T747.39	913.3464	747.4	NA	NA
M891.36T747.39	891.3640	747.4	NA	NA
M328.19T781.65	328.1907	781.7	NA	NA
M929.31T750.51	929.3113	750.5	NA	NA
M344.19T754.17	344.1859	754.2	NA	NA
M231.1T763.28	231.1017	763.3	naproxen ISTD	NA
M185.1T763.53	185.0961	763.5	naproxen ISTD (fragment)	NA
M250.08T762.76	250.0754	762.8	NA	NA
M288.2T765.88	288.1962	765.9	NA	C9-QNO
M258.15T768.47	258.1493	768.5	NA	C7:1-QNO
M270.19T766.9	270.1851	766.9	NA	C9:1-HQ
M274.27T767.94	274.2741	767.9	NA	NA
M328.14T772.12	328.1425	772.1	NA	NA
M655.28T772.63	655.2753	772.6	NA	NA
M242.15T789.6	242.1542	789.6	NA	C7:1-HQ
M309.13T780.23	309.1310	780.2	NA	NA
M353.16T780.48	353.1571	780.5	NA	NA
M113.06T779.69	113.0596	779.7	NA	NA
M304.19T786.75	304.1907	786.8	NA	NA
M184.08T792.48	184.0755	792.5	NA	NA
M260.16T796.66	260.1648	796.7	HQNO	NA
M244.17T796.4	244.1700	796.4	HHQ	NA
M312.2T794.57	312.1951	794.6	NA	NA
M159.07T795.61	159.0681	795.6	HHQ (fragment)	NA
M503.33T797.17	503.3271	797.2	NA	NA
M519.32T799.12	519.3229	799.1	HQNO [2M+H] <sup>+</sup>	NA
M161.1T800.82	161.0960	800.8	NA	NA
M302.17T804.83	302.1749	804.8	NA	NA



Table 3.S2.: Continued.

feature ID	med.mz	med.rt	library annotation	putative annotation
M201.09T806.25	201.0909	806.3	NA	NA
M286.18T808.85	286.1804	808.9	NA	C9:1-QNO
M276.16T811.73	276.1598	811.7	NA	NA
M172.17T813.55	172.1698	813.6	NA	NA
M432.26T815.37	432.2650	815.4	NA	NA
M325.07T819.96	325.0675	820.0	NA	pyochelin
M272.16T822.08	272.1648	822.1	NA	C8:1-QNO
M314.21T825.99	314.2121	826.0	NA	C11:1-QNO
M288.2T824.42	288.1956	824.4	NA	C9-QNO
M284.16T840.09	284.1648	840.1	NA	NA
M326.18T833.85	326.1758	833.8	NA	NA
M358.2T834.1	358.2022	834.1	NA	NA
M274.18T842.43	274.1806	842.4	NA	C8-QNO
M358.2T835.92	358.2016	835.9	NA	NA
M258.19T853.38	258.1855	853.4	NA	C8-HQ
M330.21T844.12	330.2068	844.1	NA	NA
M286.18T864.03	286.1808	864.0	NA	C9:1-QNO
M270.19T873.15	270.1858	873.2	NA	C9:1-HQ
M268.17T875.49	268.1697	875.5	NA	NA
M256.17T851.43	256.1699	851.4	NA	C8:1-HQ
M310.18T853.38	310.1806	853.4	NA	NA
M250.12T856.23	250.1189	856.2	NA	NA
M437.19T860.52	437.1947	860.5	NA	NA
M415.21T859.73	415.2126	859.7	NA	NA
M324.16T867.42	324.1574	867.4	NA	NA
M178.05T867.16	178.0500	867.2	NA	NA
M302.18T879.54	302.1751	879.5	NA	NA
M316.23T869.51	316.2265	869.5	NA	C11-QNO
M316.23T871.07	316.2276	871.1	NA	C11-QNO
M571.35T875.76	571.3534	875.8	NA	NA
M198.09T874.45	198.0914	874.5	NA	NA
M346.2T881.21	346.2014	881.2	NA	NA
M452.28T896.1	452.2779	896.1	NA	NA
M170.1T885.9	170.0967	885.9	NA	NA
M332.22T890.1	332.2219	890.1	NA	NA
M314.21T892.19	314.2112	892.2	NA	C11:1-QNO
M300.2T925.27	300.1960	925.3	NA	C10:1-QNO
M288.2T902.1	288.1965	902.1	NA	C9-PQS
M300.2T894.8	300.1969	894.8	NA	C10:1-QNO
M474.26T899.22	474.2600	899.2	NA	NA
M284.2T904.43	284.2014	904.4	NA	C10:1-HQ
M575.38T904.16	575.3850	904.2	NA	NA
M184.08T908.6	184.0759	908.6	NA	NA
M292.17T907.81	292.1676	907.8	NA	NA
M272.2T910.42	272.2017	910.4	NA	C9-HQ
M541.38T909.89	541.3787	909.9	NA	NA
M543.4T910.68	543.3952	910.7	NA	NA
M563.36T910.94	563.3616	910.9	NA	NA
M342.24T912.12	342.2433	912.1	NA	C13:1-QNO
M646.36T912.5	646.3562	912.5	NA	NA
M312.2T929.44	312.1963	929.4	NA	NA
M327.34T917.18	327.3378	917.2	NA	NA
M304.19T921.87	304.1914	921.9	NA	NA
M286.18T921.6	286.1808	921.6	NA	C9:1-PQS
M360.22T923.7	360.2170	923.7	NA	NA
M358.24T929.17	358.2375	929.2	NA	NA
M672.37T930.21	672.3713	930.2	NA	NA
M499.29T936.54	499.2879	936.5	NA	NA
M314.21T944.87	314.2119	944.9	NA	C11:1-PQS
M296.2T958.11	296.2015	958.1	NA	NA
M627.42T945.38	627.4163	945.4	NA	NA
M298.22T984.33	298.2171	984.3	NA	C11:1-HQ



Table 3.S2.: Continued.

feature ID	med.mz	med.rt	library annotation	putative annotation
M454.29T963.69	454.2938	963.7	NA	PE(16:0/0:0)
M595.43T950.84	595.4256	950.8	NA	NA
M289.15T954.48	289.1541	954.5	NA	NA
M267.17T955.25	267.1722	955.3	NA	NA
M302.21T956.81	302.2120	956.8	NA	C10-QNO
M284.2T959.42	284.2006	959.4	NA	C10:1-HQ
M330.21T962.26	330.2069	962.3	NA	NA
M673.38T966.68	673.3776	966.7	NA	Rha-Rha-C10-C10 +Na
M359.28T966.42	359.2799	966.4	NA	NA
M480.31T982.77	480.3095	982.8	NA	PE(18:1/0:0)
M500.22T974.72	500.2174	974.7	NA	NA
M304.19T977.83	304.1893	977.8	NA	NA
M303.19T979.65	303.1860	979.7	NA	NA
M505.25T982.5	505.2540	982.5	NA	PG(16:1/0:0) +Na
M483.27T981.33	483.2721	981.3	NA	PG(16:1/0:0)
M502.29T985.62	502.2911	985.6	NA	PE(18:1/0:0) +Na
M959.61T986.65	959.6091	986.7	NA	NA
M361.24T989	361.2357	989.0	NA	NA
M328.23T993.95	328.2275	993.9	NA	C12:1-QNO
M99.51T1040.17	99.5123	1040.2	NA	NA
M312.23T1001.35	312.2327	1001.4	NA	C12:1-HQ
M340.23T1007.62	340.2276	1007.6	NA	NA
M316.23T1007.37	316.2276	1007.4	NA	C11-PQS
M310.22T1008.4	310.2166	1008.4	NA	NA
M324.23T1025.79	324.2327	1025.8	NA	NA
M699.39T1013.83	699.3932	1013.8	NA	Rha-Rha-C10-C12:1 / Rha-Rha-C12:1-C10 +Na
M385.3T1014.61	385.2956	1014.6	NA	NA
M677.41T1014.86	677.4117	1014.9	NA	Rha-Rha-C10-C12:1 / Rha-Rha-C12:1-C10
M239.66T1017.2	239.6648	1017.2	NA	NA
M496.34T1018.23	496.3400	1018.2	NA	PC(16:0/0:0)
M527.32T1021.62	527.3199	1021.6	NA	Rha-C10-C10 /Rha-C12-C8 +Na
M359.28T1022.4	359.2802	1022.4	NA	NA
M1031.65T1023.18	1031.6490	1023.2	NA	NA
M292.66T1025.52	292.6600	1025.5	NA	NA
M326.38T1031.82	326.3786	1031.8	NA	NA
M342.24T1034.94	342.2436	1034.9	NA	NA
M314.21T1034.15	314.2120	1034.2	NA	NA
M312.2T1037.01	312.1960	1037.0	NA	NA
M326.25T1043.73	326.2486	1043.7	NA	C13:1-HQ
M522.36T1043.6	522.3553	1043.6	NA	NA
M701.41T1049.33	701.4089	1049.3	NA	Rha-Rha-C10-C12 / Rha-Rha-C12-C10 +Na
M387.31T1049.58	387.3113	1049.6	NA	NA
M679.43T1051.39	679.4271	1051.4	NA	Rha-Rha-C10-C12 / Rha-Rha-C12-C10
M459.23T1054.91	459.2286	1054.9	NA	NA
M341.27T1084.63	341.2667	1084.6	NA	NA
M553.34T1072.25	553.3355	1072.3	NA	Rha-C10-C12:1 / Rha-C12:1-C10 +Na
M385.29T1079.09	385.2944	1079.1	NA	NA
M1083.68T1073.79	1083.6806	1073.8	NA	NA
M381.26T1078.47	381.2624	1078.5	NA	NA
M342.24T1078.72	342.2433	1078.7	NA	C13:1-PQS
M359.28T1078.46	359.2801	1078.5	NA	NA
M727.42T1093.9	727.4246	1093.9	NA	Rha-Rha-C12:1-C12 / Rha-Rha-C12-C12:1 +Na
M473.35T1082.31	473.3455	1082.3	NA	NA
M429.32T1084.63	429.3193	1084.6	cholesteryl acetate	NA
M507.27T1100.59	507.2703	1100.6	NA	NA
M485.29T1092.08	485.2888	1092.1	NA	NA

Table 3.S2.: Continued.

feature ID	med.mz	med.rt	library annotation	putative annotation
M413.33T1095.83	413.3265	1095.8	NA	NA
M705.44T1095.83	705.4419	1095.8	NA	Rha-Rha-C12:1-C12 / Rha-Rha-C12-C12:1
M326.25T1125.41	326.2481	1125.4	NA	C13:1-HQ
M397.26T1099.57	397.2573	1099.6	NA	NA
M311.26T1103.98	311.2565	1104.0	NA	NA
M555.35T1108.39	555.3508	1108.4	NA	Rha-C10-C12 / Rha-C12-C10 +Na
M387.31T1109.17	387.3115	1109.2	NA	NA
M577.33T1107.35	577.3310	1107.3	NA	NA
M1087.71T1109.95	1087.7120	1109.9	NA	NA
M487.36T1112.82	487.3602	1112.8	NA	NA
M443.34T1115.66	443.3351	1115.7	NA	NA
M399.31T1118.65	399.3089	1118.7	NA	NA
M355.28T1117.75	355.2825	1117.8	NA	NA
M99.51T1121.65	99.5125	1121.6	NA	NA
M124.09T1369.75	124.0871	1369.8	NA	NA
M328.26T1127.63	328.2637	1127.6	NA	C13-HQ
M533.29T1139.9	533.2856	1139.9	NA	NA
M729.44T1131.57	729.4401	1131.6	NA	Rha-Rha-C12-C12 +Na
M415.34T1132.34	415.3423	1132.3	NA	NA
M707.46T1132.6	707.4579	1132.6	NA	Rha-Rha-C12-C12
M511.3T1137.29	511.3033	1137.3	NA	NA
M581.37T1153.59	581.3668	1153.6	NA	NA
M282.14T1145.42	282.1367	1145.4	NA	NA
M555.36T1151.77	555.3583	1151.8	NA	Rha-C10-C12 / Rha-C12-C10 +Na
M278.18T1152.3	278.1836	1152.3	NA	NA
M387.31T1159.58	387.3111	1159.6	NA	NA
M409.29T1159.84	409.2935	1159.8	NA	NA
M755.46T1159.83	755.4555	1159.8	NA	NA
M413.32T1163.6	413.3241	1163.6	NA	NA
M369.3T1166.6	369.2980	1166.6	NA	NA
M282.22T1185.24	282.2221	1185.2	NA	NA
M287.19T1183.15	287.1880	1183.2	NA	NA
M583.38T1189.9	583.3816	1189.9	NA	NA
M415.34T1190.29	415.3413	1190.3	NA	NA
M559.13T1193.82	559.1303	1193.8	NA	NA
M354.32T1200.38	354.3185	1200.4	NA	NA
M257.25T1211.04	257.2478	1211.0	NA	NA
M757.47T1212.99	757.4701	1213.0	NA	NA
M705.51T1214.43	705.5122	1214.4	NA	NA
M661.49T1220.54	661.4856	1220.5	NA	NA
M617.46T1226.29	617.4587	1226.3	NA	NA
M573.43T1232.94	573.4323	1232.9	NA	NA
M354.32T1236.97	354.3178	1237.0	NA	NA
M529.41T1240.64	529.4065	1240.6	NA	NA
M283.22T1239.08	283.2167	1239.1	NA	NA
M485.38T1248.74	485.3810	1248.7	NA	NA
M123.09T1250.04	123.0917	1250.0	NA	NA
M284.29T1262.26	284.2949	1262.3	NA	NA
M441.36T1259.15	441.3555	1259.2	NA	NA
M633.15T1261.75	633.1491	1261.7	NA	NA
M532.36T1262.01	532.3577	1262.0	NA	NA
M397.33T1269.79	397.3293	1269.8	NA	NA
M488.33T1270.83	488.3318	1270.8	NA	NA
M136.11T1272.39	136.1118	1272.4	NA	NA
M466.32T1274.46	466.3194	1274.5	NA	NA
M122.1T1276.28	122.0965	1276.3	NA	NA
M865.62T1277.85	865.6232	1277.8	NA	NA
M353.3T1282.53	353.3029	1282.5	NA	NA
M821.6T1283.04	821.5964	1283.0	NA	NA

**Table 3.S2.:** Continued.

feature ID	med.mz	med.rt	library annotation	putative annotation
M353.29T1284.86	353.2862	1284.9	NA	NA
M325.07T1285.12	325.0679	1285.1	NA	NA
M777.57T1286.67	777.5715	1286.7	NA	NA
M609.34T1288.22	609.3412	1288.2	NA	NA
M733.54T1292.1	733.5447	1292.1	NA	NA
M413.27T1297.03	413.2671	1297.0	NA	Bis(2-ethylhexyl)phthalate CONT
M393.3T1294.68	393.2979	1294.7	NA	NA
M149.02T1295.98	149.0235	1296.0	NA	NA
M391.29T1297.28	391.2850	1297.3	NA	NA
M689.52T1298.05	689.5180	1298.1	NA	NA
M645.49T1301.45	645.4914	1301.4	NA	NA
M601.46T1307.16	601.4649	1307.2	NA	NA
M144.98T1409.56	144.9824	1409.6	NA	NA
M265.96T1330.45	265.9630	1330.4	NA	NA
M557.44T1313.67	557.4387	1313.7	NA	NA
M146.98T1316.52	146.9808	1316.5	NA	NA
M513.41T1320.44	513.4136	1320.4	NA	NA
M469.39T1328.01	469.3872	1328.0	NA	NA
M284.3T1328.91	284.2958	1328.9	NA	NA
M324.33T1331.38	324.3270	1331.4	NA	NA
M311.25T1335.27	311.2494	1335.3	NA	NA
M542.42T1359.88	542.4209	1359.9	NA	NA
M338.34T1339.43	338.3423	1339.4	NA	NA
M425.36T1338.38	425.3602	1338.4	NA	NA
M502.35T1343.85	502.3487	1343.8	NA	NA
M121.97T1381.09	121.9663	1381.1	NA	NA
M146.98T1381.23	146.9807	1381.2	NA	NA
M394.35T1366.24	394.3475	1366.2	NA	NA
M122.1T1374.68	122.0965	1374.7	NA	NA
M136.11T1448.2	136.1123	1448.2	NA	NA
M133.96T1397.35	133.9596	1397.4	NA	NA
M122.1T1460.79	122.0966	1460.8	NA	NA
M123.09T1417.81	123.0921	1417.8	NA	NA
M146.98T1450.82	146.9805	1450.8	NA	NA
M469.38T1466.95	469.3779	1467.0	NA	NA
M121.97T1468.13	121.9661	1468.1	NA	NA

**Table 3.S3.:** List of cluster ID assignments resulting from density based clustering using the OPTICS algorithm along with the colour coding used in Figure 4.

feature ID	cluster ID	colour code
M146.17T59.35 - spermidine	1	red
M129.14T58.57 - spermidine (fragment)	1	red
M112.11T57.8 - spermidine (fragment)	1	red
M148.06T69.65 - glutamate	1	red
M130.05T69.64 - glutamate (fragment)	1	red
M116.07T73.51 - (proline)	1	red
M301.11T76.88	1	red
M162.08T82.05	1	red
M316.16T94.47	1	red
M299.14T96.54	1	red
M317.14T103.57	1	red
M290.13T105.39	1	red
M278.57T112.67	1	red
M190.07T117.36 - N-acetylglutamate	1	red
M535.19T118.41 - (Glu Glu Glu Glu)	1	red
M132.1T122.82 - leucine / isoleucine / norleucine	1	red



Table 3.S3.: Continued.

feature ID	cluster ID	colour code
M307.08T126.2 - (glutathion disulphide (2+))	1	red
M182.08T125.93 - tyrosine	1	red
M613.16T127.23 - glutathion disulfide	1	red
M304.06T132.18	1	red
M388.11T146.83	1	red
M664.23T165.1	1	red
M332.62T165.35	1	red
M351.59T166.13	1	red
M314.09T175.26	1	red
M166.09T233.22 - phenylalanine	1	red
M120.08T233.48 - phenylalanine (fragment)	1	red
M103.05T235.3 - phenylalanine (fragment)	1	red
M397.14T289.26	1	red
M793.27T273.95	1	red
M416.12T271.35	1	red
M174.06T277.59	1	red
M262.16T278.37	1	red
M276.11T276.54	1	red
M323.07T305.37	1	red
M461.66T328.59	1	red
M922.32T329.89	1	red
M480.64T329.62	1	red
M472.65T332.73	1	red
M219.13T335.84 - (Ser Leu)	1	red
M382.65T337.93	1	red
M220.12T336.88 - (panthotenate)	1	red
M526.18T342.1	1	red
M1051.36T342.36	1	red
M363.77T343.13	1	red
M590.7T351.59	1	red
M1180.4T351.98	1	red
M367.64T355.63	1	red
M360.21T355.1 - (Ile Val Glu / Val Ile Glu)	1	red
M235.07T357.45	1	red
M655.23T359.27	1	red
M666.22T359.78	1	red
M174.06T360.96	1	red
M243.08T362.4	1	red
M215.08T362.66	1	red
M331.18T363.96	1	red
M719.75T366.82	1	red
M359.64T368.9	1	red
M270.19T371.24	1	red
M784.27T374.37	1	red
M795.26T374.75	1	red
M848.79T380.87	1	red
M578.85T381.64	1	red
M913.31T388.16	1	red
M258.58T397.4	1	red
M977.84T395.18	1	red
M364.62T397.4	1	red
M728.23T398.3	1	red
M295.13T406.65 - (Glu Phe)	1	red
M350.62T409.24 - folic acid (fragment)	1	red
M268.66T408.73	1	red
M316.22T413.91 - (Val Val Val)	1	red
M328.22T437.8 - (Pro Leu Val)	1	red
M326.21T414.96	1	red
M336.19T415.21 - (Phe Val Ala / Val Phe Ala)	1	red
M295.19T417.02	1	red
M289.12T417.28	1	red
M261.12T419.11	1	red

Table 3.S3.: Continued.

feature ID	cluster ID	colour code
M263.14T447.96 - (Pro Phe)	1	red
M257.15T449.13	1	red
M243.13T457.64	1	red
M360.19T459.98	1	red
M399.26T461.02	1	red
M277.15T466.47	1	red
M231.17T475.37 - (Leu Val)	1	red
M491.29T496.41	1	red
M344.25T476.52 - (Leu Leu Val or isomer)	1	red
M211.14T497.45 - (cyclo(Leu Pro))	1	red
M342.24T486.25 - (Pro Ile Leu or isomer)	1	red
M439.29T506.82	1	red
M831.14T503.95	1	red
M473.3T511.21	1	red
M441.31T516.66	1	red
M245.13T525.64 - (cyclo(Phe Pro))	1	red
M246.13T525.63	1	red
M505.3T526.66	1	red
M311.14T527.97	1	red
M265.57T546.37	1	red
M775.9T572.93	1	red
M392.25T573.83 - (Phe Leu Leu or isomer)	1	red
M289.14T649.01	1	red
M297.14T656.95	1	red
M1012.42T688.19	1	red
M1012.54T689.35	1	red
M1157.05T687.92	1	red
M900.15T689.22	1	red
M1156.91T690.64	1	red
M1012.29T689.74	1	red
M900.04T690	1	red
M1156.76T690.25	1	red
M1012.67T693.51	1	red
M288.2T705.59 - (C9-QNO)	1	red
M304.19T717.91	1	red
M286.18T728.73 - (C9:1-QNO)	1	red
M288.2T765.88 - (C9-QNO)	1	red
M270.19T766.9 - (C9:1-HQ)	1	red
M304.19T786.75	1	red
M286.18T808.85 - (C9:1-QNO)	1	red
M288.2T824.42 - (C9-QNO)	1	red
M284.16T840.09	1	red
M286.18T864.03 - (C9:1-QNO)	1	red
M270.19T873.15 - (C9:1-HQ)	1	red
M268.17T875.49	1	red
M571.35T875.76	1	red
M288.2T902.1 - (C9-PQS)	1	red
M284.2T904.43 - (C10:1-HQ)	1	red
M575.38T904.16	1	red
M272.2T910.42 - (C9-HQ)	1	red
M541.38T909.89	1	red
M543.4T910.68	1	red
M286.18T921.6 - (C9:1-PQS)	1	red
M289.15T954.48	1	red
M284.2T959.42 - (C10:1-HQ)	1	red
M459.23T1054.91	1	red
M287.19T1183.15	1	red
M284.29T1262.26	1	red
M284.3T1328.91	1	red
M428.04T80.88 - ADP	2	green
M372.55T88.26 - NADP (2+)	2	green
M348.07T90.34 - AMP	2	green



Table 3.S3.: Continued.

feature ID	cluster ID	colour code
M136.06T90.6 - AMP (fragment)	2	green
M695.13T91.88 - AMP (2+)	2	green
M123.06T103.31 - nicotinamide	2	green
M332.56T107.48 - NAD (2+)	2	green
M664.12T108.26 - NAD	2	green
M542.07T108.52 - NAD (fragment)	2	green
M332.08T113.21 - (dAMP)	2	green
M193.07T150.49 - (S-(5'-adenosyl)-homocysteine (2+))	2	green
M385.13T151.52 - S-(5'-adenosyl)-homocysteine	2	green
M268.1T171.87 - adenosine	2	green
M340.06T231.65	2	green
M298.1T385.29 - 5'-methylthioadenosine	2	green
M786.17T426.25 - FAD	2	green
M321.1T537.6	3	blue
M393.12T645.89	3	blue
M446.19T745.32 - glipizide ISTD	3	blue
M321.1T746.09	3	blue
M891.36T747.39	3	blue
M525.18T352.51	4	cyan
M120.04T450.56 - anthranilate (fragment)	4	cyan
M138.06T451.33 - anthranilate	4	cyan
M188.07T371.25 - tryptophan (fragment)	5	pink
M205.1T370.99 - tryptophan	5	pink
M163.09T512.64	5	pink
M118.06T585.64	5	pink
M245.18T210.61	6	yellow
M145.08T410.28	6	yellow
M214.13T411.31	6	yellow
M145.08T508.11	6	yellow
M202.09T545.21	7	grey
M184.08T792.48	7	grey
M184.08T908.6	7	grey
M260.16T583.7 - (C7-QNO)	8	orange
M260.16T651.62 - HQNO	8	orange
M242.15T789.6 - (C7:1-HQ)	8	orange
M260.16T796.66 - HQNO	8	orange
M244.17T796.4 - HHQ	8	orange
M503.33T797.17	8	orange
M519.32T799.12 - HQNO [2M+H] <sup>+</sup>	8	orange
M232.13T495.36	9	red
M232.13T683.77 - (C5-QNO)	9	red
M176.07T465.7	10	green
M159.09T464.27	10	green
M159.07T466.73	10	green
M159.07T795.61 - HHQ (fragment)	10	green
M161.1T800.82	10	green
M342.24T912.12 - (C13:1-QNO)	11	blue
M324.23T1025.79	11	blue
M342.24T1034.94	11	blue
M326.25T1043.73 - (C13:1-HQ)	11	blue
M342.24T1078.72 - (C13:1-PQS)	11	blue
M326.25T1125.41 - (C13:1-HQ)	11	blue
M328.26T1127.63 - (C13-HQ)	11	blue
M312.2T794.57	12	cyan
M314.21T825.99 - (C11:1-QNO)	12	cyan
M330.21T844.12	12	cyan
M316.23T869.51 - (C11-QNO)	12	cyan
M316.23T871.07 - (C11-QNO)	12	cyan
M332.22T890.1	12	cyan
M314.21T892.19 - (C11:1-QNO)	12	cyan
M312.2T929.44	12	cyan
M314.21T944.87 - (C11:1-PQS)	12	cyan

Table 3.S3.: Continued.

feature ID	cluster ID	colour code
M296.2T958.11	12	cyan
M627.42T945.38	12	cyan
M298.22T984.33 - (C11:1-HQ)	12	cyan
M595.43T950.84	12	cyan
M316.23T1007.37 - (C11-PQS)	12	cyan
M276.16T811.73	13	pink
M178.05T867.16	13	pink
M302.18T879.54	13	pink
M304.19T921.87	13	pink
M330.21T962.26	13	pink
M170.1T885.9	14	yellow
M314.21T1034.15	14	yellow
M312.2T1037.01	14	yellow
M325.07T739.09 - (pyochelin)	15	grey
M325.07T819.96 - (pyochelin)	15	grey
M325.07T1285.12	15	grey
M211.09T382.17 - pyocyanin	16	orange
M255.08T482.73	16	orange
M330.19T724.4	16	orange
M324.06T75.32 - CMP	17	red
M325.04T78.94 - CMP	17	red
M308.06T84.63 - dCMP	17	red
M317.09T742.2	18	green
M358.2T834.1	18	green
M358.2T835.92	18	green
M185.1T729.78 - (naproxen (fragment))	19	blue
M231.1T763.28 - naproxen ISTD	19	blue
M185.1T763.53 - naproxen ISTD (fragment)	19	blue
M250.08T762.76	19	blue
M264.18T675.46 - nortriptyline ISTD	20	cyan
M233.13T676.23 - nortriptyline ISTD (fragment)	20	cyan
M304.19T977.83	20	cyan
M303.19T979.65	20	cyan
M359.28T966.42	21	pink
M385.3T1014.61	21	pink
M677.41T1014.86 - (Rha-Rha-C10-C12:1 / Rha-Rha-C12:1-C10)	21	pink
M359.28T1022.4	21	pink
M387.31T1049.58	21	pink
M679.43T1051.39 - (Rha-Rha-C10-C12 / Rha-Rha-C12-C10)	21	pink
M385.29T1079.09	21	pink
M359.28T1078.46	21	pink
M413.33T1095.83	21	pink
M387.31T1109.17	21	pink
M415.34T1132.34	21	pink
M707.46T1132.6 - (Rha-Rha-C12-C12)	21	pink
M387.31T1159.58	21	pink
M415.34T1190.29	21	pink
M338.34T1339.43	21	pink
M136.11T1272.39	22	yellow
M136.11T1448.2	22	yellow
M133.96T1397.35	22	yellow
M224.08T640.69 - phenazine-1-carboxamide	23	grey
M207.05T641.72 - phenazine-1-carboxamide (fragment)	23	grey
M225.07T698.07 - phenazine-1-carboxylic acid	23	grey
M207.06T699.1 - phenazine-1-carboxylic acid (fragment)	23	grey
M226.18T703.76	23	grey
M206.03T721.29	23	grey
M672.37T930.21	24	orange
M480.31T982.77 - (PE(18:1/0:0))	24	orange
M959.61T986.65	24	orange
M511.3T1137.29	24	orange
M646.36T912.5	25	red



Table 3.S3.: Continued.

feature ID	cluster ID	colour code
M454.29T963.69 - (PE(16:0/0:0))	25	red
M485.29T1092.08	25	red
M487.36T1112.82	25	red
M452.28T896.1	26	green
M483.27T981.33 - (PG(16:1/0:0))	26	green
M311.26T1103.98	26	green
M99.09T97.84	27	blue
M98.98T141.84	27	blue
M267.17T955.25	27	blue
M393.3T1294.68	28	cyan
M149.02T1295.98	28	cyan
M391.29T1297.28	28	cyan
M186.13T666.88	29	pink
M354.32T1200.38	29	pink
M354.32T1236.97	29	pink
M124.09T1369.75	30	yellow
M123.09T1250.04	30	yellow
M122.1T1276.28	30	yellow
M121.97T1381.09	30	yellow
M122.1T1374.68	30	yellow
M122.1T1460.79	30	yellow
M123.09T1417.81	30	yellow
M121.97T1468.13	30	yellow
M144.98T1409.56	31	grey
M146.98T1316.52	31	grey
M146.98T1381.23	31	grey
M146.98T1450.82	31	grey
M251.16T60.64	0	black
M212.85T65.02	0	black
M290.85T64.76	0	black
M179.06T71.32 - (gluconolactone)	0	black
M197.07T71.57	0	black
M242.08T74.02	0	black
M219.1T77.65 - (Glu Ala)	0	black
M191.05T84.37	0	black
M189.12T85.8	0	black
M85.06T86.7	0	black
M338.14T94.98	0	black
M364.07T97.19 - GMP	0	black
M254.09T100.71	0	black
M110.06T100.45 - 2-aminophenol	0	black
M159.11T102.26	0	black
M137.05T112.16 - hypoxanthine	0	black
M179.07T114.5	0	black
M202.18T115.29	0	black
M180.05T120.87 - ((iso)xanthopterine)	0	black
M322.11T131.92	0	black
M323.06T135.83 - (dTMP)	0	black
M143.08T138.17	0	black
M182.07T141.05	0	black
M164.06T147.09 - (pterine)	0	black
M250.14T156.21	0	black
M125.57T156.99	0	black
M251.15T228.5	0	black
M320.17T232.69	0	black
M254.16T280.46	0	black
M244.11T293.5	0	black
M256.18T309.56	0	black
M911.28T323.37	0	black
M727.21T324.67	0	black
M1235.4T324.66	0	black
M479.19T333.77	0	black



Table 3.S3.: Continued.

feature ID	cluster ID	colour code
M597.68T335.59 - (UDP-muramyl-pentapeptide)	0	black
M568.14T338.32	0	black
M194.08T340.15	0	black
M194.07T346.78	0	black
M217.1T347.04	0	black
M472.85T348.87	0	black
M708.77T349.12	0	black
M734.27T356.41	0	black
M686.76T358.74	0	black
M679.29T361.61	0	black
M693.77T367.08	0	black
M462.85T367.6	0	black
M304.18T373.32	0	black
M694.26T376.97	0	black
M358.2T377.48 - (Ile Pro Glu / Leu Pro Glu)	0	black
M195.11T378.66	0	black
M732.26T383.07	0	black
M261.13T385.55	0	black
M187.12T391.55	0	black
M204.12T389.98	0	black
M243.18T392.85	0	black
M516.16T395.97	0	black
M188.12T399.72	0	black
M254.09T400.89	0	black
M190.05T411.45 - kynurenate	0	black
M378.2T412.48 - (Pro Tyr Val)	0	black
M197.13T423.26	0	black
M316.21T430.54	0	black
M231.11T430.8	0	black
M160.08T433.66	0	black
M457.11T433.91 - (FMN)	0	black
M204.1T435.6	0	black
M195.09T436.77 - caffeine ISTD	0	black
M291.15T444.85 - trimethoprim ISTD	0	black
M275.11T445.88	0	black
M300.2T461.28	0	black
M291.09T462.05 - trimethoprim ISTD	0	black
M418.14T474.32	0	black
M427.29T485.86	0	black
M245.59T495.11	0	black
M187.09T505	0	black
M530.13T538.67	0	black
M186.09T531.62	0	black
M188.11T535.78 - (C3-HQ)	0	black
M275.03T536.04	0	black
M243.09T558.67 - lumichrome	0	black
M204.1T564.12	0	black
M263.12T576.07	0	black
M200.11T578.31	0	black
M202.12T578.44	0	black
M136.08T584.09	0	black
M323.09T586.42	0	black
M485.11T612.66	0	black
M215.12T626.24	0	black
M263.12T624.79	0	black
M213.07T652.92	0	black
M214.12T663.32 - (C5:1-HQ)	0	black
M216.14T670.01 - (C5-HQ)	0	black
M227.08T670.26	0	black
M191.09T676.49 - nortriptyline ISTD (fragment)	0	black
M276.16T680.26 - nortriptyline ISTD +Na	0	black
M257.06T704.3	0	black



Table 3.S3.: Continued.

feature ID	cluster ID	colour code
M269.06T708.74 - phenazine-1,6-dicarboxylic acid	0	black
M181.08T724.14	0	black
M307.02T733.71	0	black
M404.23T734.62	0	black
M230.15T735.39 - (C6-HQ)	0	black
M255.08T740.91	0	black
M468.17T744.02 - glipizide ISTD +Na	0	black
M913.35T747.39	0	black
M328.19T781.65	0	black
M929.31T750.51	0	black
M344.19T754.17	0	black
M258.15T768.47 - (C7:1-QNO)	0	black
M274.27T767.94	0	black
M328.14T772.12	0	black
M655.28T772.63	0	black
M309.13T780.23	0	black
M353.16T780.48	0	black
M113.06T779.69	0	black
M302.17T804.83	0	black
M201.09T806.25	0	black
M172.17T813.55	0	black
M432.26T815.37	0	black
M272.16T822.08 - (C8:1-QNO)	0	black
M326.18T833.85	0	black
M274.18T842.43 - (C8-QNO)	0	black
M258.19T853.38 - (C8-HQ)	0	black
M256.17T851.43 - (C8:1-HQ)	0	black
M310.18T853.38	0	black
M250.12T856.23	0	black
M437.19T860.52	0	black
M415.21T859.73	0	black
M324.16T867.42	0	black
M198.09T874.45	0	black
M346.2T881.21	0	black
M300.2T925.27 - (C10:1-QNO)	0	black
M300.2T894.8 - (C10:1-QNO)	0	black
M474.26T899.22	0	black
M292.17T907.81	0	black
M563.36T910.94	0	black
M327.34T917.18	0	black
M360.22T923.7	0	black
M358.24T929.17	0	black
M499.29T936.54	0	black
M302.21T956.81 - (C10-QNO)	0	black
M673.38T966.68 - (Rha-Rha-C10-C10 +Na )	0	black
M500.22T974.72	0	black
M505.25T982.5 - (PG(16:1/0:0) +Na)	0	black
M502.29T985.62 - (PE(18:1/0:0) +Na)	0	black
M361.24T989	0	black
M328.23T993.95 - (C12:1-QNO)	0	black
M99.51T1040.17	0	black
M312.23T1001.35 - (C12:1-HQ)	0	black
M340.23T1007.62	0	black
M310.22T1008.4	0	black
M699.39T1013.83 - (Rha-Rha-C10-C12:1 / Rha-Rha-C12:1-C10 +Na)	0	black
M239.66T1017.2	0	black
M496.34T1018.23 - (PC(16:0/0:0))	0	black
M527.32T1021.62 - (Rha-C10-C10 /Rha-C12-C8 +Na)	0	black
M1031.65T1023.18	0	black
M292.66T1025.52	0	black
M326.38T1031.82	0	black
M522.36T1043.6	0	black

Table 3.S3.: Continued.

feature ID	cluster ID	colour code
M701.41T1049.33 - (Rha-Rha-C10-C12 / Rha-Rha-C12-C10 +Na)	0	black
M341.27T1084.63	0	black
M553.34T1072.25 - (Rha-C10-C12:1 / Rha-C12:1-C10 +Na)	0	black
M1083.68T1073.79	0	black
M381.26T1078.47	0	black
M727.42T1093.9 - (Rha-Rha-C12:1-C12 / Rha-Rha-C12-C12:1 +Na)	0	black
M473.35T1082.31	0	black
M429.32T1084.63 - cholesteryl acetate	0	black
M507.27T1100.59	0	black
M705.44T1095.83 - (Rha-Rha-C12:1-C12 / Rha-Rha-C12-C12:1)	0	black
M397.26T1099.57	0	black
M555.35T1108.39 - (Rha-C10-C12 / Rha-C12-C10 +Na)	0	black
M577.33T1107.35	0	black
M1087.71T1109.95	0	black
M443.34T1115.66	0	black
M399.31T1118.65	0	black
M355.28T1117.75	0	black
M99.51T1121.65	0	black
M533.29T1139.9	0	black
M729.44T1131.57 - (Rha-Rha-C12-C12 +Na)	0	black
M581.37T1153.59	0	black
M282.14T1145.42	0	black
M555.36T1151.77 - (Rha-C10-C12 / Rha-C12-C10 +Na)	0	black
M278.18T1152.3	0	black
M409.29T1159.84	0	black
M755.46T1159.83	0	black
M413.32T1163.6	0	black
M369.3T1166.6	0	black
M282.22T1185.24	0	black
M583.38T1189.9	0	black
M559.13T1193.82	0	black
M257.25T1211.04	0	black
M757.47T1212.99	0	black
M705.51T1214.43	0	black
M661.49T1220.54	0	black
M617.46T1226.29	0	black
M573.43T1232.94	0	black
M529.41T1240.64	0	black
M283.22T1239.08	0	black
M485.38T1248.74	0	black
M441.36T1259.15	0	black
M633.15T1261.75	0	black
M532.36T1262.01	0	black
M397.33T1269.79	0	black
M488.33T1270.83	0	black
M466.32T1274.46	0	black
M865.62T1277.85	0	black
M353.3T1282.53	0	black
M821.6T1283.04	0	black
M353.29T1284.86	0	black
M777.57T1286.67	0	black
M609.34T1288.22	0	black
M733.54T1292.1	0	black
M413.27T1297.03 - (Bis(2-ethylhexyl)phthalate CONT)	0	black
M689.52T1298.05	0	black
M645.49T1301.45	0	black
M601.46T1307.16	0	black
M265.96T1330.45	0	black
M557.44T1313.67	0	black
M513.41T1320.44	0	black
M469.39T1328.01	0	black
M324.33T1331.38	0	black



Table 3.S3.: Continued.

feature ID	cluster ID	colour code
M311.25T1335.27	0	black
M542.42T1359.88	0	black
M425.36T1338.38	0	black
M502.35T1343.85	0	black
M394.35T1366.24	0	black
M469.38T1466.95	0	black

Table 3.S4.: List of cluster ID assignments resulting from hierarchical clustering with average linkage of product ion spectra similarities.

feature ID	cluster ID
M146.17T59.35 - spermidine	1
M129.14T58.57 - spermidine (fragment)	1
M112.11T57.8 - spermidine (fragment)	1
M251.16T60.64	1
M212.85T65.02	2
M290.85T64.76	3
M148.06T69.65 - glutamate	4
M130.05T69.64 - glutamate (fragment)	4
M179.06T71.32 - (gluconolactone)	5
M197.07T71.57	5
M116.07T73.51 - (proline)	6
M242.08T74.02	4
M324.06T75.32 - CMP	5
M301.11T76.88	4
M219.11T77.65 - (Glu Ala)	4
M325.04T78.94 - CMP	5
M428.04T80.88 - ADP	7
M162.08T82.05	4
M191.05T84.37	8
M308.06T84.63 - dCMP	5
M189.12T85.8	8
M85.06T86.7	9
M372.55T88.26 - NADP (2+)	7
M348.07T90.34 - AMP	7
M136.06T90.6 - AMP (fragment)	7
M695.13T91.88 - AMP (2+)	7
M316.16T94.47	4
M338.14T94.98	10
M364.07T97.19 - GMP	11
M299.14T96.54	4
M99.09T97.84	12
M254.09T100.71	13
M110.06T100.45 - 2-aminophenol	14
M123.06T103.31 - nicotinamide	7
M159.11T102.26	15
M317.14T103.57	4
M290.13T105.39	4
M332.56T107.48 - NAD (2+)	7
M664.12T108.26 - NAD	7
M542.07T108.52 - NAD (fragment)	7
M278.57T112.67	4
M137.05T112.16 - hypoxanthine	7
M332.08T113.21 - (dAMP)	7
M179.07T114.5	16
M202.18T115.29	17
M190.07T117.36 - N-acetylglutamate	4

**Table 3.S4.:** Continued.

feature ID	cluster ID
M535.19T118.41 - (Glu Glu Glu Glu)	4
M180.05T120.87 - ((iso)xanthopterine)	18
M132.1T122.82 - leucine / isoleucine / norleucine	6
M307.08T126.2 - (glutathion disulphide (2+))	4
M182.08T125.93 - tyrosine	6
M613.16T127.23 - glutathion disulfide	4
M322.11T131.92	4
M304.06T132.18	4
M323.06T135.83 - (dTMP)	4
M143.08T138.17	1
M98.98T141.84	12
M182.07T141.05	11
M388.11T146.83	4
M164.06T147.09 - (pterine)	6
M193.07T150.49 - (S-(5'-adenosyl)-homocysteine (2+))	7
M385.13T151.52 - S-(5'-adenosyl)-homocysteine	7
M250.14T156.21	1
M125.57T156.99	19
M664.23T165.1	4
M332.62T165.35	4
M351.59T166.13	4
M268.1T171.87 - adenosine	7
M314.09T175.26	4
M245.18T210.61	20
M251.15T228.5	21
M166.09T233.22 - phenylalanine	6
M120.08T233.48 - phenylalanine (fragment)	6
M340.06T231.65	7
M320.17T232.69	22
M103.05T235.3 - phenylalanine (fragment)	23
M397.14T289.26	4
M793.27T273.95	4
M416.12T271.35	4
M174.06T277.59	14
M262.16T278.37	1
M276.11T276.54	4
M254.16T280.46	6
M244.11T293.5	12
M323.07T305.37	4
M256.18T309.56	6
M911.28T323.37	5
M727.21T324.67	5
M1235.4T324.66	5
M461.66T328.59	4
M922.32T329.89	4
M480.64T329.62	4
M479.19T333.77	4
M472.65T332.73	4
M597.68T335.59 - (UDP-muramyl-pentapeptide)	4
M219.13T335.84 - (Ser Leu)	6
M382.65T337.93	4
M220.12T336.88 - (panthotenate)	6
M568.14T338.32	4
M194.08T340.15	24
M526.18T342.1	4
M1051.36T342.36	4
M363.77T343.13	4
M194.07T346.78	25
M217.1T347.04	26
M472.85T348.87	6
M708.77T349.12	27
M590.7T351.59	4



Table 3.S4.: Continued.

feature ID	cluster ID
M525.18T352.51	14
M1180.4T351.98	4
M367.64T355.63	4
M360.21T355.1 - (Ile Val Glu / Val Ile Glu)	6
M734.27T356.41	28
M235.07T357.45	4
M655.23T359.27	4
M686.76T358.74	29
M666.22T359.78	4
M174.06T360.96	14
M679.29T361.61	30
M243.08T362.4	31
M215.08T362.66	31
M331.18T363.96	1
M693.77T367.08	32
M719.75T366.82	4
M462.85T367.6	6
M359.64T368.9	4
M188.07T371.25 - tryptophan (fragment)	33
M205.1T370.99 - tryptophan	33
M270.19T371.24	34
M304.18T373.32	33
M784.27T374.37	4
M795.26T374.75	4
M694.26T376.97	32
M358.2T377.48 - (Ile Pro Glu / Leu Pro Glu)	6
M195.11T378.66	6
M848.79T380.87	4
M211.09T382.17 - pyocyanin	35
M578.85T381.64	4
M732.26T383.07	36
M298.1T385.29 - 5'-methylthioadenosine	7
M261.13T385.55	37
M913.31T388.16	4
M187.12T391.55	20
M204.12T389.98	38
M243.18T392.85	17
M258.58T397.4	4
M977.84T395.18	4
M516.16T395.97	4
M364.62T397.4	4
M728.23T398.3	14
M188.12T399.72	20
M254.09T400.89	35
M295.13T406.65 - (Glu Phe)	6
M350.62T409.24 - folic acid (fragment)	4
M268.66T408.73	6
M145.08T410.28	20
M190.05T411.45 - kynurenate	39
M214.13T411.31	20
M378.2T412.48 - (Pro Tyr Val)	6
M316.22T413.91 - (Val Val Val)	1
M328.22T437.8 - (Pro Leu Val)	6
M326.21T414.96	6
M336.19T415.21 - (Phe Val Ala / Val Phe Ala)	6
M295.19T417.02	6
M289.12T417.28	6
M261.12T419.11	6
M197.13T423.26	6
M786.17T426.25 - FAD	7
M316.21T430.54	17
M231.11T430.8	40

**Table 3.S4.:** Continued.

feature ID	cluster ID
M160.08T433.66	33
M457.11T433.91 - (FMN)	41
M204.1T435.6	39
M195.09T436.77 - caffeine ISTD	21
M291.15T444.85 - trimethoprim ISTD	42
M275.11T445.88	42
M263.14T447.96 - (Pro Phe)	6
M257.15T449.13	6
M120.04T450.56 - anthranilate (fragment)	14
M138.06T451.33 - anthranilate	14
M243.13T457.64	6
M360.19T459.98	6
M300.2T461.28	17
M399.26T461.02	6
M291.09T462.05 - trimethoprim ISTD	42
M176.07T465.7	43
M159.09T464.27	43
M277.15T466.47	6
M159.07T466.73	43
M418.14T474.32	44
M231.17T475.37 - (Leu Val)	6
M491.29T496.41	6
M344.25T476.52 - (Leu Leu Val or isomer)	6
M255.08T482.73	35
M211.14T497.45 - (cyclo(Leu Pro))	6
M342.24T486.25 - (Pro Ile Leu or isomer)	6
M427.29T485.86	1
M245.59T495.11	35
M232.13T495.36	43
M145.08T508.11	20
M439.29T506.82	6
M831.14T503.95	6
M187.09T505	31
M163.09T512.64	33
M473.3T511.21	6
M441.31T516.66	6
M245.13T525.64 - (cyclo(Phe Pro))	6
M246.13T525.63	6
M505.3T526.66	6
M311.14T527.97	6
M530.13T538.67	45
M186.09T531.62	39
M188.11T535.78 - (C3-HQ)	43
M275.03T536.04	46
M321.1T537.6	23
M202.09T545.21	43
M265.57T546.37	4
M243.09T558.67 - lumichrome	41
M204.1T564.12	43
M260.16T583.7 - (C7-QNO)	43
M775.9T572.93	6
M392.25T573.83 - (Phe Leu Leu or isomer)	6
M263.12T576.07	33
M200.11T578.31	43
M202.12T578.44	43
M136.08T584.09	33
M118.06T585.64	33
M323.09T586.42	47
M485.11T612.66	45
M215.12T626.24	35
M263.12T624.79	33
M224.08T640.69 - phenanzine-1-carboxamide	48



Table 3.S4.: Continued.

feature ID	cluster ID
M207.05T641.72 - phenazine-1-carboxamide (fragment)	48
M393.12T645.89	23
M289.14T649.01	4
M213.07T652.92	49
M260.16T651.62 - HQNO	43
M297.14T656.95	43
M186.13T666.88	50
M214.12T663.32 - (C5:1-HQ)	43
M216.14T670.01 - (C5-HQ)	43
M227.08T670.26	51
M264.18T675.46 - nortriptyline ISTD	33
M233.13T676.23 - nortriptyline ISTD (fragment)	33
M191.09T676.49 - nortriptyline ISTD (fragment)	33
M276.16T680.26 - nortriptyline ISTD +Na	43
M232.13T683.77 - (C5-QNO)	43
M1012.42T688.19	6
M1012.54T689.35	6
M1157.05T687.92	6
M900.15T689.22	6
M1156.91T690.64	6
M1012.29T689.74	6
M900.04T690	6
M1156.76T690.25	6
M1012.67T693.51	6
M225.07T698.07 - phenazine-1-carboxylic acid	48
M207.06T699.1 - phenazine-1-carboxylic acid (fragment)	48
M257.06T704.3	52
M226.18T703.76	48
M288.2T705.59 - (C9-QNO)	43
M269.06T708.74 - phenazine-1,6-dicarboxylic acid	48
M325.07T739.09 - (pyochelin)	53
M304.19T717.91	43
M206.03T721.29	48
M286.18T728.73 - (C9:1-QNO)	43
M181.08T724.14	35
M330.19T724.4	35
M185.1T729.78 - (naproxen (fragment))	54
M307.02T733.71	55
M404.23T734.62	43
M230.15T735.39 - (C6-HQ)	43
M255.08T740.91	35
M317.09T742.2	43
M446.19T745.32 - glipizide ISTD	23
M468.17T744.02 - glipizide ISTD +Na	23
M321.1T746.09	23
M913.35T747.39	23
M891.36T747.39	23
M328.19T781.65	43
M929.31T750.51	56
M344.19T754.17	43
M231.1T763.28 - naproxen ISTD	54
M185.1T763.53 - naproxen ISTD (fragment)	54
M250.08T762.76	54
M288.2T765.88 - (C9-QNO)	43
M258.15T768.47 - (C7:1-QNO)	43
M270.19T766.9 - (C9:1-HQ)	43
M274.27T767.94	57
M328.14T772.12	58
M655.28T772.63	59
M242.15T789.6 - (C7:1-HQ)	43
M309.13T780.23	60
M353.16T780.48	61



Table 3.S4.: Continued.

feature ID	cluster ID
M113.06T779.69	62
M304.19T786.75	43
M184.08T792.48	43
M260.16T796.66 - HQNO	43
M244.17T796.4 - HHQ	43
M312.2T794.57	43
M159.07T795.61 - HHQ (fragment)	43
M503.33T797.17	43
M519.32T799.12 - HQNO [2M+H] <sup>+</sup>	43
M161.1T800.82	43
M302.17T804.83	43
M201.09T806.25	58
M286.18T808.85 - (C9:1-QNO)	43
M276.16T811.73	36
M172.17T813.55	17
M432.26T815.37	43
M325.07T819.96 - (pyochelin)	53
M272.16T822.08 - (C8:1-QNO)	43
M314.21T825.99 - (C11:1-QNO)	43
M288.2T824.42 - (C9-QNO)	43
M284.16T840.09	43
M326.18T833.85	43
M358.2T834.1	43
M274.18T842.43 - (C8-QNO)	43
M358.2T835.92	43
M258.19T853.38 - (C8-HQ)	43
M330.21T844.12	43
M286.18T864.03 - (C9:1-QNO)	43
M270.19T873.15 - (C9:1-HQ)	43
M268.17T875.49	43
M256.17T851.43 - (C8:1-HQ)	43
M310.18T853.38	43
M250.12T856.23	63
M437.19T860.52	64
M415.21T859.73	33
M324.16T867.42	53
M178.05T867.16	36
M302.18T879.54	36
M316.23T869.51 - (C11-QNO)	43
M316.23T871.07 - (C11-QNO)	43
M571.35T875.76	43
M198.09T874.45	43
M346.2T881.21	14
M452.28T896.1	12
M170.1T885.9	14
M332.22T890.1	43
M314.21T892.19 - (C11:1-QNO)	43
M300.2T925.27 - (C10:1-QNO)	43
M288.2T902.1 - (C9-PQS)	43
M300.2T894.8 - (C10:1-QNO)	43
M474.26T899.22	65
M284.2T904.43 - (C10:1-HQ)	43
M575.38T904.16	43
M184.08T908.6	43
M292.17T907.81	66
M272.2T910.42 - (C9-HQ)	43
M541.38T909.89	43
M543.4T910.68	43
M563.36T910.94	67
M342.24T912.12 - (C13:1-QNO)	43
M646.36T912.5	12
M312.2T929.44	43



Table 3.S4.: Continued.

feature ID	cluster ID
M327.34T917.18	1
M304.19T921.87	36
M286.18T921.6 - (C9:1-PQS)	43
M360.22T923.7	43
M358.24T929.17	43
M672.37T930.21	12
M499.29T936.54	68
M314.21T944.87 - (C11:1-PQS)	43
M296.2T958.11	43
M627.42T945.38	43
M298.22T984.33 - (C11:1-HQ)	43
M454.29T963.69 - (PE(16:0/0:0))	12
M595.43T950.84	43
M289.15T954.48	43
M267.17T955.25	12
M302.21T956.81 - (C10-QNO)	43
M284.2T959.42 - (C10:1-HQ)	43
M330.21T962.26	36
M673.38T966.68 - (Rha-Rha-C10-C10 +Na )	69
M359.28T966.42	63
M480.31T982.77 - (PE(18:1/0:0))	12
M500.22T974.72	70
M304.19T977.83	33
M303.19T979.65	33
M505.25T982.5 - (PG(16:1/0:0) +Na)	71
M483.27T981.33 - (PG(16:1/0:0))	12
M502.29T985.62 - (PE(18:1/0:0) +Na)	71
M959.61T986.65	12
M361.24T989	72
M328.23T993.95 - (C12:1-QNO)	43
M99.51T1040.17	73
M312.23T1001.35 - (C12:1-HQ)	43
M340.23T1007.62	43
M316.23T1007.37 - (C11-PQS)	43
M310.22T1008.4	43
M324.23T1025.79	43
M699.39T1013.83 - (Rha-Rha-C10-C12:1 / Rha-Rha-C12:1-C10 +Na)	74
M385.3T1014.61	63
M677.41T1014.86 - (Rha-Rha-C10-C12:1 / Rha-Rha-C12:1-C10)	63
M239.66T1017.2	75
M496.34T1018.23 - (PC(16:0/0:0))	6
M527.32T1021.62 - (Rha-C10-C10 /Rha-C12-C8 +Na)	69
M359.28T1022.4	63
M1031.65T1023.18	69
M292.66T1025.52	5
M326.38T1031.82	76
M342.24T1034.94	43
M314.21T1034.15	14
M312.2T1037.01	14
M326.25T1043.73 - (C13:1-HQ)	43
M522.36T1043.6	6
M701.41T1049.33 - (Rha-Rha-C10-C12 / Rha-Rha-C12-C10 +Na)	69
M387.31T1049.58	63
M679.43T1051.39 - (Rha-Rha-C10-C12 / Rha-Rha-C12-C10)	63
M459.23T1054.91	43
M341.27T1084.63	77
M553.34T1072.25 - (Rha-C10-C12:1 / Rha-C12:1-C10 +Na)	74
M385.29T1079.09	63
M1083.68T1073.79	74
M381.26T1078.47	69
M342.24T1078.72 - (C13:1-PQS)	43
M359.28T1078.46	63

**Table 3.S4.:** Continued.

feature ID	cluster ID
M727.42T1093.9 - (Rha-Rha-C12:1-C12 / Rha-Rha-C12-C12:1 +Na)	78
M473.35T1082.31	79
M429.32T1084.63 - cholesteryl acetate	80
M507.27T1100.59	71
M485.29T1092.08	12
M413.33T1095.83	63
M705.44T1095.83 - (Rha-Rha-C12:1-C12 / Rha-Rha-C12-C12:1)	63
M326.25T1125.41 - (C13:1-HQ)	43
M397.26T1099.57	81
M311.26T1103.98	12
M555.35T1108.39 - (Rha-C10-C12 / Rha-C12-C10 +Na)	69
M387.31T1109.17	63
M577.33T1107.35	82
M1087.71T1109.95	69
M487.36T1112.82	12
M443.34T1115.66	83
M399.31T1118.65	84
M355.28T1117.75	85
M99.51T1121.65	73
M124.09T1369.75	86
M328.26T1127.63 - (C13-HQ)	43
M533.29T1139.9	71
M729.44T1131.57 - (Rha-Rha-C12-C12 +Na)	87
M415.34T1132.34	63
M707.46T1132.6 - (Rha-Rha-C12-C12)	63
M511.3T1137.29	12
M581.37T1153.59	88
M282.14T1145.42	89
M555.36T1151.77 - (Rha-C10-C12 / Rha-C12-C10 +Na)	90
M278.18T1152.3	91
M387.31T1159.58	63
M409.29T1159.84	69
M755.46T1159.83	92
M413.32T1163.6	93
M369.3T1166.6	94
M282.22T1185.24	95
M287.19T1183.15	43
M583.38T1189.9	87
M415.34T1190.29	63
M559.13T1193.82	96
M354.32T1200.38	50
M257.25T1211.04	25
M757.47T1212.99	97
M705.51T1214.43	98
M661.49T1220.54	99
M617.46T1226.29	100
M573.43T1232.94	101
M354.32T1236.97	50
M529.41T1240.64	102
M283.22T1239.08	103
M485.38T1248.74	104
M123.09T1250.04	86
M284.29T1262.26	43
M441.36T1259.15	83
M633.15T1261.75	105
M532.36T1262.01	106
M397.33T1269.79	84
M488.33T1270.83	107
M136.11T1272.39	108
M466.32T1274.46	109
M122.1T1276.28	86
M865.62T1277.85	110



Table 3.S4.: Continued.

feature ID	cluster ID
M353.3T1282.53	85
M821.6T1283.04	111
M353.29T1284.86	85
M325.07T1285.12	53
M777.57T1286.67	112
M609.34T1288.22	113
M733.54T1292.1	114
M413.27T1297.03 - (Bis(2-ethylhexyl)phthalate CONT)	93
M393.3T1294.68	115
M149.02T1295.98	115
M391.29T1297.28	115
M689.52T1298.05	116
M645.49T1301.45	117
M601.46T1307.16	118
M144.98T1409.56	86
M265.96T1330.45	119
M557.44T1313.67	120
M146.98T1316.52	86
M513.41T1320.44	121
M469.39T1328.01	119
M284.3T1328.91	43
M324.33T1331.38	122
M311.25T1335.27	103
M542.42T1359.88	14
M338.34T1339.43	63
M425.36T1338.38	123
M502.35T1343.85	124
M121.97T1381.09	86
M146.98T1381.23	86
M394.35T1366.24	95
M122.1T1374.68	86
M136.11T1448.2	108
M133.96T1397.35	108
M122.1T1460.79	86
M123.09T1417.81	86
M146.98T1450.82	86
M469.38T1466.95	125
M121.97T1468.13	86

Table 3.S5.: List of cluster ID assignments resulting from hierarchical clustering with average linkage of neutral loss pattern similarities.

feature ID	cluster ID
M146.17T59.35 - spermidine	1
M129.14T58.57 - spermidine (fragment)	2
M112.11T57.8 - spermidine (fragment)	3
M251.16T60.64	4
M212.85T65.02	5
M290.85T64.76	6
M148.06T69.65 - glutamate	7
M130.05T69.64 - glutamate (fragment)	7
M179.06T71.32 - (gluconolactone)	7
M197.07T71.57	7
M116.07T73.51 - (proline)	7
M242.08T74.02	8
M324.06T75.32 - CMP	9
M301.11T76.88	10

**Table 3.S5.:** Continued.

feature ID	cluster ID
M219.1T77.65 - (Glu Ala)	11
M325.04T78.94 - CMP	9
M428.04T80.88 - ADP	12
M162.08T82.05	7
M191.05T84.37	13
M308.06T84.63 - dCMP	14
M189.12T85.8	1
M85.06T86.7	15
M372.55T88.26 - NADP (2+)	16
M348.07T90.34 - AMP	9
M136.06T90.6 - AMP (fragment)	2
M695.13T91.88 - AMP (2+)	17
M316.16T94.47	18
M338.14T94.98	1
M364.07T97.19 - GMP	9
M299.14T96.54	19
M99.09T97.84	20
M254.09T100.71	4
M110.06T100.45 - 2-aminophenol	11
M123.06T103.31 - nicotinamide	11
M159.11T102.26	4
M317.14T103.57	18
M290.13T105.39	10
M332.56T107.48 - NAD (2+)	21
M664.12T108.26 - NAD	22
M542.07T108.52 - NAD (fragment)	23
M278.57T112.67	24
M137.05T112.16 - hypoxanthine	11
M332.08T113.21 - (dAMP)	14
M179.07T114.5	25
M202.18T115.29	26
M190.07T117.36 - N-acetylglutamate	27
M535.19T118.41 - (Glu Glu Glu Glu)	28
M180.05T120.87 - ((iso)xanthopterine)	25
M132.1T122.82 - leucine / isoleucine / norleucine	7
M307.08T126.2 - (glutathion disulphide (2+))	29
M182.08T125.93 - tyrosine	7
M613.16T127.23 - glutathion disulfide	10
M322.11T131.92	10
M304.06T132.18	11
M323.06T135.83 - (dTMP)	30
M143.08T138.17	31
M98.98T141.84	11
M182.07T141.05	32
M388.11T146.83	33
M164.06T147.09 - (pterine)	2
M193.07T150.49 - (S-(5'-adenosyl)-homocysteine (2+))	34
M385.13T151.52 - S-(5'-adenosyl)-homocysteine	35
M250.14T156.21	2
M125.57T156.99	36
M664.23T165.1	28
M332.62T165.35	37
M351.59T166.13	38
M268.1T171.87 - adenosine	39
M314.09T175.26	40
M245.18T210.61	41
M251.15T228.5	42
M166.09T233.22 - phenylalanine	7
M120.08T233.48 - phenylalanine (fragment)	2
M340.06T231.65	43
M320.17T232.69	44
M103.05T235.3 - phenylalanine (fragment)	45



Table 3.S5.: Continued.

feature ID	cluster ID
M397.14T289.26	46
M793.27T273.95	28
M416.12T271.35	38
M174.06T277.59	32
M262.16T278.37	47
M276.11T276.54	48
M254.16T280.46	49
M244.11T293.5	50
M323.07T305.37	51
M256.18T309.56	49
M911.28T323.37	52
M727.21T324.67	53
M1235.4T324.66	54
M461.66T328.59	37
M922.32T329.89	28
M480.64T329.62	55
M479.19T333.77	56
M472.65T332.73	57
M597.68T335.59 - (UDP-muramyl-pentapeptide)	58
M219.13T335.84 - (Ser Leu)	7
M382.65T337.93	59
M220.12T336.88 - (panthotenate)	60
M568.14T338.32	61
M194.08T340.15	62
M526.18T342.1	46
M1051.36T342.36	28
M363.77T343.13	63
M194.07T346.78	2
M217.1T347.04	64
M472.85T348.87	65
M708.77T349.12	66
M590.7T351.59	37
M525.18T352.51	28
M1180.4T351.98	28
M367.64T355.63	37
M360.21T355.1 - (Ile Val Glu / Val Ile Glu)	10
M734.27T356.41	67
M235.07T357.45	68
M655.23T359.27	46
M686.76T358.74	69
M666.22T359.78	70
M174.06T360.96	71
M679.29T361.61	72
M243.08T362.4	32
M215.08T362.66	73
M331.18T363.96	74
M693.77T367.08	75
M719.75T366.82	37
M462.85T367.6	65
M359.64T368.9	76
M188.07T371.25 - tryptophan (fragment)	27
M205.1T370.99 - tryptophan	19
M270.19T371.24	49
M304.18T373.32	49
M784.27T374.37	46
M795.26T374.75	77
M694.26T376.97	78
M358.2T377.48 - (Ile Pro Glu / Leu Pro Glu)	79
M195.11T378.66	80
M848.79T380.87	37
M211.09T382.17 - pyocyanin	81
M578.85T381.64	82

**Table 3.S5.:** Continued.

feature ID	cluster ID
M732.26T383.07	28
M298.1T385.29 - 5'-methylthioadenosine	83
M261.13T385.55	84
M913.31T388.16	85
M187.12T391.55	86
M204.12T389.98	60
M243.18T392.85	87
M258.58T397.4	88
M977.84T395.18	89
M516.16T395.97	10
M364.62T397.4	90
M728.23T398.3	91
M188.12T399.72	92
M254.09T400.89	93
M295.13T406.65 - (Glu Phe)	10
M350.62T409.24 - folic acid (fragment)	94
M268.66T408.73	95
M145.08T410.28	96
M190.05T411.45 - kynurenate	7
M214.13T411.31	86
M378.2T412.48 - (Pro Tyr Val)	97
M316.22T413.91 - (Val Val Val)	98
M328.22T437.8 - (Pro Leu Val)	98
M326.21T414.96	79
M336.19T415.21 - (Phe Val Ala / Val Phe Ala)	99
M295.19T417.02	100
M289.12T417.28	47
M261.12T419.11	80
M197.13T423.26	32
M786.17T426.25 - FAD	101
M316.21T430.54	39
M231.11T430.8	25
M160.08T433.66	81
M457.11T433.91 - (FMN)	102
M204.1T435.6	11
M195.09T436.77 - caffeine ISTD	103
M291.15T444.85 - trimethoprim ISTD	60
M275.11T445.88	102
M263.14T447.96 - (Pro Phe)	104
M257.15T449.13	105
M120.04T450.56 - anthranilate (fragment)	32
M138.06T451.33 - anthranilate	11
M243.13T457.64	106
M360.19T459.98	107
M300.2T461.28	108
M399.26T461.02	42
M291.09T462.05 - trimethoprim ISTD	109
M176.07T465.7	7
M159.09T464.27	110
M277.15T466.47	42
M159.07T466.73	111
M418.14T474.32	112
M231.17T475.37 - (Leu Val)	98
M491.29T496.41	113
M344.25T476.52 - (Leu Leu Val or isomer)	98
M255.08T482.73	114
M211.14T497.45 - (cyclo(Leu Pro))	42
M342.24T486.25 - (Pro Ile Leu or isomer)	106
M427.29T485.86	113
M245.59T495.11	88
M232.13T495.36	11
M145.08T508.11	2



Table 3.S5.: Continued.

feature ID	cluster ID
M439.29T506.82	79
M831.14T503.95	115
M187.09T505	116
M163.09T512.64	25
M473.3T511.21	98
M441.31T516.66	98
M245.13T525.64 - (cyclo(Phe Pro))	80
M246.13T525.63	117
M505.3T526.66	118
M311.14T527.97	80
M530.13T538.67	10
M186.09T531.62	107
M188.11T535.78 - (C3-HQ)	119
M275.03T536.04	120
M321.1T537.6	121
M202.09T545.21	4
M265.57T546.37	88
M243.09T558.67 - lumichrome	31
M204.1T564.12	122
M260.16T583.7 - (C7-QNO)	102
M775.9T572.93	123
M392.25T573.83 - (Phe Leu Leu or isomer)	106
M263.12T576.07	124
M200.11T578.31	125
M202.12T578.44	119
M136.08T584.09	11
M118.06T585.64	31
M323.09T586.42	126
M485.11T612.66	4
M215.12T626.24	127
M263.12T624.79	124
M224.08T640.69 - phenazine-1-carboxamide	25
M207.05T641.72 - phenazine-1-carboxamide (fragment)	73
M393.12T645.89	128
M289.14T649.01	129
M213.07T652.92	32
M260.16T651.62 - HQNO	11
M297.14T656.95	130
M186.13T666.88	19
M214.12T663.32 - (C5:1-HQ)	119
M216.14T670.01 - (C5-HQ)	131
M227.08T670.26	11
M264.18T675.46 - nortriptyline ISTD	132
M233.13T676.23 - nortriptyline ISTD (fragment)	133
M191.09T676.49 - nortriptyline ISTD (fragment)	134
M276.16T680.26 - nortriptyline ISTD +Na	11
M232.13T683.77 - (C5-QNO)	26
M1012.42T688.19	135
M1012.54T689.35	136
M1157.05T687.92	137
M900.15T689.22	138
M1156.91T690.64	139
M1012.29T689.74	140
M900.04T690	141
M1156.76T690.25	142
M1012.67T693.51	143
M225.07T698.07 - phenazine-1-carboxylic acid	11
M207.06T699.1 - phenazine-1-carboxylic acid (fragment)	32
M257.06T704.3	11
M226.18T703.76	144
M288.2T705.59 - (C9-QNO)	11
M269.06T708.74 - phenazine-1,6-dicarboxylic acid	11



**Table 3.S5.:** Continued.

feature ID	cluster ID
M325.07T739.09 - (pyochelin)	145
M304.19T717.91	11
M206.03T721.29	73
M286.18T728.73 - (C9:1-QNO)	146
M181.08T724.14	147
M330.19T724.4	148
M185.1T729.78 - (naproxen (fragment))	148
M307.02T733.71	149
M404.23T734.62	150
M230.15T735.39 - (C6-HQ)	151
M255.08T740.91	11
M317.09T742.2	152
M446.19T745.32 - glipizide ISTD	153
M468.17T744.02 - glipizide ISTD +Na	153
M321.1T746.09	121
M913.35T747.39	154
M891.36T747.39	154
M328.19T781.65	155
M929.31T750.51	154
M344.19T754.17	156
M231.1T763.28 - naproxen ISTD	7
M185.1T763.53 - naproxen ISTD (fragment)	148
M250.08T762.76	157
M288.2T765.88 - (C9-QNO)	11
M258.15T768.47 - (C7:1-QNO)	26
M270.19T766.9 - (C9:1-HQ)	158
M274.27T767.94	159
M328.14T772.12	152
M655.28T772.63	39
M242.15T789.6 - (C7:1-HQ)	26
M309.13T780.23	160
M353.16T780.48	161
M113.06T779.69	162
M304.19T786.75	11
M184.08T792.48	163
M260.16T796.66 - HQNO	155
M244.17T796.4 - HHQ	26
M312.2T794.57	164
M159.07T795.61 - HHQ (fragment)	165
M503.33T797.17	150
M519.32T799.12 - HQNO [2M+H] <sup>+</sup>	150
M161.1T800.82	166
M302.17T804.83	155
M201.09T806.25	27
M286.18T808.85 - (C9:1-QNO)	167
M276.16T811.73	155
M172.17T813.55	26
M432.26T815.37	150
M325.07T819.96 - (pyochelin)	145
M272.16T822.08 - (C8:1-QNO)	26
M314.21T825.99 - (C11:1-QNO)	11
M288.2T824.42 - (C9-QNO)	11
M284.16T840.09	26
M326.18T833.85	155
M358.2T834.1	146
M274.18T842.43 - (C8-QNO)	155
M358.2T835.92	146
M258.19T853.38 - (C8-HQ)	164
M330.21T844.12	11
M286.18T864.03 - (C9:1-QNO)	155
M270.19T873.15 - (C9:1-HQ)	167
M268.17T875.49	26

Table 3.S5.: Continued.

feature ID	cluster ID
M256.17T851.43 - (C8:1-HQ)	26
M310.18T853.38	146
M250.12T856.23	10
M437.19T860.52	168
M415.21T859.73	169
M324.16T867.42	167
M178.05T867.16	7
M302.18T879.54	170
M316.23T869.51 - (C11-QNO)	171
M316.23T871.07 - (C11-QNO)	11
M571.35T875.76	150
M198.09T874.45	148
M346.2T881.21	172
M452.28T896.1	8
M170.1T885.9	173
M332.22T890.1	102
M314.21T892.19 - (C11:1-QNO)	174
M300.2T925.27 - (C10:1-QNO)	164
M288.2T902.1 - (C9-PQS)	174
M300.2T894.8 - (C10:1-QNO)	164
M474.26T899.22	175
M284.2T904.43 - (C10:1-HQ)	167
M575.38T904.16	176
M184.08T908.6	152
M292.17T907.81	177
M272.2T910.42 - (C9-HQ)	167
M541.38T909.89	178
M543.4T910.68	178
M563.36T910.94	178
M342.24T912.12 - (C13:1-QNO)	4
M646.36T912.5	179
M312.2T929.44	155
M327.34T917.18	180
M304.19T921.87	174
M286.18T921.6 - (C9:1-PQS)	181
M360.22T923.7	182
M358.24T929.17	11
M672.37T930.21	179
M499.29T936.54	170
M314.21T944.87 - (C11:1-PQS)	174
M296.2T958.11	167
M627.42T945.38	183
M298.22T984.33 - (C11:1-HQ)	174
M454.29T963.69 - (PE(16:0/0:0))	8
M595.43T950.84	184
M289.15T954.48	185
M267.17T955.25	186
M302.21T956.81 - (C10-QNO)	187
M284.2T959.42 - (C10:1-HQ)	167
M330.21T962.26	11
M673.38T966.68 - (Rha-Rha-C10-C10 +Na )	188
M359.28T966.42	189
M480.31T982.77 - (PE(18:1/0:0))	8
M500.22T974.72	190
M304.19T977.83	41
M303.19T979.65	41
M505.25T982.5 - (PG(16:1/0:0) +Na)	191
M483.27T981.33 - (PG(16:1/0:0))	8
M502.29T985.62 - (PE(18:1/0:0) +Na)	175
M959.61T986.65	192
M361.24T989	193
M328.23T993.95 - (C12:1-QNO)	182

**Table 3.S5.:** Continued.

feature ID	cluster ID
M99.51T1040.17	194
M312.23T1001.35 - (C12:1-HQ)	186
M340.23T1007.62	174
M316.23T1007.37 - (C11-PQS)	174
M310.22T1008.4	195
M324.23T1025.79	174
M699.39T1013.83 - (Rha-Rha-C10-C12:1 / Rha-Rha-C12:1-C10 +Na)	188
M385.3T1014.61	189
M677.41T1014.86 - (Rha-Rha-C10-C12:1 / Rha-Rha-C12:1-C10)	196
M239.66T1017.2	197
M496.34T1018.23 - (PC(16:0/0:0))	198
M527.32T1021.62 - (Rha-C10-C10 /Rha-C12-C8 +Na)	170
M359.28T1022.4	189
M1031.65T1023.18	199
M292.66T1025.52	200
M326.38T1031.82	201
M342.24T1034.94	202
M314.21T1034.15	155
M312.2T1037.01	155
M326.25T1043.73 - (C13:1-HQ)	186
M522.36T1043.6	203
M701.41T1049.33 - (Rha-Rha-C10-C12 / Rha-Rha-C12-C10 +Na)	188
M387.31T1049.58	189
M679.43T1051.39 - (Rha-Rha-C10-C12 / Rha-Rha-C12-C10)	204
M459.23T1054.91	47
M341.27T1084.63	205
M553.34T1072.25 - (Rha-C10-C12:1 / Rha-C12:1-C10 +Na)	206
M385.29T1079.09	189
M1083.68T1073.79	196
M381.26T1078.47	170
M342.24T1078.72 - (C13:1-PQS)	202
M359.28T1078.46	189
M727.42T1093.9 - (Rha-Rha-C12:1-C12 / Rha-Rha-C12-C12:1 +Na)	188
M473.35T1082.31	207
M429.32T1084.63 - cholesteryl acetate	208
M507.27T1100.59	198
M485.29T1092.08	8
M413.33T1095.83	189
M705.44T1095.83 - (Rha-Rha-C12:1-C12 / Rha-Rha-C12-C12:1)	204
M326.25T1125.41 - (C13:1-HQ)	186
M397.26T1099.57	209
M311.26T1103.98	210
M555.35T1108.39 - (Rha-C10-C12 / Rha-C12-C10 +Na)	170
M387.31T1109.17	189
M577.33T1107.35	170
M1087.71T1109.95	204
M487.36T1112.82	211
M443.34T1115.66	212
M399.31T1118.65	213
M355.28T1117.75	214
M99.51T1121.65	215
M124.09T1369.75	216
M328.26T1127.63 - (C13-HQ)	217
M533.29T1139.9	203
M729.44T1131.57 - (Rha-Rha-C12-C12 +Na)	188
M415.34T1132.34	189
M707.46T1132.6 - (Rha-Rha-C12-C12)	204
M511.3T1137.29	8
M581.37T1153.59	206
M282.14T1145.42	152
M555.36T1151.77 - (Rha-C10-C12 / Rha-C12-C10 +Na)	171
M278.18T1152.3	218



Table 3.S5.: Continued.

feature ID	cluster ID
M387.31T1159.58	189
M409.29T1159.84	170
M755.46T1159.83	219
M413.32T1163.6	220
M369.3T1166.6	221
M282.22T1185.24	131
M287.19T1183.15	222
M583.38T1189.9	170
M415.34T1190.29	189
M559.13T1193.82	223
M354.32T1200.38	180
M257.25T1211.04	186
M757.47T1212.99	224
M705.51T1214.43	225
M661.49T1220.54	226
M617.46T1226.29	227
M573.43T1232.94	228
M354.32T1236.97	180
M529.41T1240.64	229
M283.22T1239.08	174
M485.38T1248.74	230
M123.09T1250.04	231
M284.29T1262.26	232
M441.36T1259.15	233
M633.15T1261.75	234
M532.36T1262.01	235
M397.33T1269.79	236
M488.33T1270.83	237
M136.11T1272.39	148
M466.32T1274.46	238
M122.1T1276.28	239
M865.62T1277.85	240
M353.3T1282.53	241
M821.6T1283.04	242
M353.29T1284.86	243
M325.07T1285.12	145
M777.57T1286.67	244
M609.34T1288.22	245
M733.54T1292.1	246
M413.27T1297.03 - (Bis(2-ethylhexyl)phthalate CONT)	222
M393.3T1294.68	247
M149.02T1295.98	32
M391.29T1297.28	222
M689.52T1298.05	248
M645.49T1301.45	249
M601.46T1307.16	250
M144.98T1409.56	251
M265.96T1330.45	252
M557.44T1313.67	253
M146.98T1316.52	251
M513.41T1320.44	254
M469.39T1328.01	255
M284.3T1328.91	256
M324.33T1331.38	257
M311.25T1335.27	186
M542.42T1359.88	258
M338.34T1339.43	180
M425.36T1338.38	259
M502.35T1343.85	260
M121.97T1381.09	19
M146.98T1381.23	251
M394.35T1366.24	261

**Table 3.S5.:** Continued.

feature ID	cluster ID
M122.1T1374.68	262
M136.11T1448.2	133
M133.96T1397.35	263
M122.1T1460.79	262
M123.09T1417.81	1
M146.98T1450.82	251
M469.38T1466.95	264
M121.97T1468.13	19

**Table 3.S6.:** List of cluster ID assignments resulting from hierarchical clustering with average linkage of product ion spectra similarities of the spectra acquired in the alkyl quinolone-biased semi-targeted analysis.

feature ID	cluster ID
M623.03T1.85	1
M146.17T62 - spermidine	2
M129.14T59.08 - spermidine (fragment)	2
M112.11T59.58 - spermidine (fragment)	2
M251.16T62.63	2
M212.85T65.69	3
M272.98T67.98	4
M268.16T71.04	2
M258.11T71.54	5
M244.09T73.32	6
M232.12T73.82	6
M258.15T76.24	6
M244.13T77.38	6
M162.08T82.86	7
M146.09T82.86	7
M278.57T80.94	7
M316.16T83.6	7
M348.07T90.6 - AMP	7
M535.19T105.08 - (glutamyl-glutamyl-glutamyl-glutamate)	7
M317.15T90.72	7
M307.08T124.25 - (glutathion disulphide (2+))	7
M695.14T92.88 - AMP (2+)	7
M160.12T94.16	2
M228.1T97.33	8
M286.15T99.5	7
M613.16T127.81 - glutathion disulfide	7
M190.07T111.57	7
M664.12T108.26 - NAD	7
M332.56T110.04 - NAD (2+)	7
M214.09T110.18	9
M332.08T114.24	7
M388.11T112.96	7
M179.07T114.74	10
M202.18T118.3	11
M132.1T120.95 - leucine / isoleucine / norleucine	6
M182.08T124.39 - tyrosine	12
M165.06T124.26	12
M664.23T160.49	7
M332.62T150.95	7
M182.07T141.54	9
M143.08T140.01	2
M322.09T141.4	7

Table 3.S6.: Continued.

feature ID	cluster ID
M164.06T146.64	13
M388.11T146.24	7
M314.09T175.36	7
M193.07T151.1	7
M170.07T151.6	7
M385.13T152.99 - S-(5'-Adenosyl)-L-Homocysteine	7
M250.14T157.33	10
M125.57T157.83	10
M233.12T158.33	10
M226.16T166.32	6
M351.59T169.88	7
M268.11T172.44 - adenosine	7
M270.11T172.44	7
M292.17T179.56	14
M343.1T179.67	6
M100.11T183.91	5
M166.09T229.79 - phenylalanine	13
M344.18T191.66	6
M254.16T206.28	6
M228.17T205.89	6
M397.14T212.12	7
M793.28T213.9	7
M416.12T214.39	7
M388.14T214.4	7
M409.12T217.95	7
M245.18T219.37	15
M340.06T226.34	7
M246.18T223.56	15
M146.08T224.06	15
M114.58T224.56	15
M251.15T246.65	16
M120.08T234.08	13
M103.05T235.22	13
M320.17T239.4	14
M190.11T243.09	17
M100.08T265.58	5
M199.11T252.49	6
M397.14T266.07	7
M125.07T254.76	18
M793.28T267.35	7
M276.11T261.88	7
M408.13T268.99	7
M416.12T270.77	7
M388.14T273.44	7
M254.16T280.72	6
M309.1T283.5	19
M174.06T290.61	20
M262.16T288.83	2
M244.11T292.28	15
M176.06T288.33	20
M115.04T300.3	21
M245.15T303.96	6
M246.15T305.63	6
M323.07T306.13	7
M397.14T307.91	7
M256.18T310.97	22
M138.06T311.47	7
M169.1T311.97	6
M561.21T315.52	7
M793.28T314.52	7
M344.18T316.8	6
M461.66T325.19	7

**Table 3.S6.:** Continued.

feature ID	cluster ID
M922.32T325.7	7
M597.68T324.42 - (UDP-muramyl-pentapeptide)	7
M311.12T328.86	12
M479.19T334.2	7
M382.65T337.51	7
M316.19T338.55	6
M219.14T338.66	6
M174.06T351.23	20
M526.19T343.1	7
M218.11T347.55	23
M194.07T347.54	24
M242.19T348.32	6
M525.19T352.37	20
M590.71T353.51	7
M537.1T354.65	7
M367.64T357.96	7
M667.31T357.96	7
M326.17T359.88	6
M218.1T358.6	15
M655.23T361.27	7
M664.18T361.77	7
M232.12T363.05	25
M216.09T364.18	26
M326.21T370.56	6
M302.2T371.31	11
M270.19T369.77	6
M312.19T376.01	6
M204.07T380.34	13
M694.26T379.56	6
M549.08T381.98	7
M195.11T381.34	6
M211.09T384.65 - pyocyanin	27
M256.18T385.68	24
M240.15T386.18	24
M214.1T384.9	27
M204.12T389.24	17
M186.11T389.74	15
M298.1T390.24 - 5'-methylthioadenosine	7
M188.13T392.8	15
M254.09T397.99	6
M243.18T393.79	11
M244.19T397.35	11
M424.31T396.35	12
M258.58T398.62	7
M364.62T399.12	7
M302.17T401.7	6
M188.12T402.2	15
M333.18T405.75	6
M308.56T406.25	7
M316.19T408.81	6
M260.14T419.09	7
M214.13T411.09	15
M190.05T411.59 - kynurenate	13
M328.22T437.8	6
M300.2T412.87	24
M284.18T413.37	24
M316.22T416.42	2
M302.18T415.15	24
M326.21T421.26	6
M296.19T420.87	6
M286.18T421.37	24
M197.13T424.18	6



Table 3.S6.: Continued.

feature ID	cluster ID
M415.26T424.68	6
M279.14T426.45	28
M298.18T426.6	2
M209.07T427.59	27
M138.06T429.38 - anthranilate	7
M316.21T431.95	11
M232.12T432.45	25
M214.09T432.95	25
M232.12T444.26	25
M344.23T436.41	24
M324.2T437.16	15
M204.1T438.3	29
M195.09T441.85 - caffeine ISTD	16
M542.14T443.63	30
M272.19T444.41	31
M244.14T457.75	6
M230.12T446.3	29
M258.11T447.33	29
M246.11T447.83	6
M310.17T451.14	6
M326.18T450.64	6
M300.16T450.74	13
M258.15T453.8	6
M257.15T456.08	6
M159.09T457.86	12
M240.15T460.03	6
M360.19T459.64	6
M160.1T463.19	12
M340.22T470.45	6
M300.2T465.86	11
M232.1T467.53	25
M176.07T468.67 - [C1-QNO]	29
M160.07T468.52 - [C1-HQ]	12
M159.07T469.81	12
M514.33T470.3	6
M220.1T472.09	29
M498.15T474.75	32
M344.26T479.98	6
M215.08T480.98	15
M174.09T482.9 - [C2-HQ]	29
M347.13T482.75	33
M211.14T485.83	6
M216.07T486.47	6
M183.09T486.33	27
M342.24T491.56	6
M427.29T489	6
M394.24T492.31	12
M327.2T491.67	6
M258.19T493.84	6
M418.11T492.96	7
M328.22T495.62	6
M232.13T497.65 - [C5-QNO]	29
M214.08T498.79	22
M146.08T513.91	15
M491.29T500.57	6
M146.06T503.24	15
M354.24T505.8	6
M342.2T509.6	6
M188.09T506.55	26
M246.13T524.47	6
M230.12T508.96 - [C5:1-QNO]	29
M204.07T510.24	13



**Table 3.S6.:** Continued.

feature ID	cluster ID
M162.06T511.63	12
M190.09T514.94 - [C2-QNO]	29
M163.09T517.22	12
M441.31T519.14	6
M242.11T522.69	34
M497.24T525.47	35
M245.13T526.75	6
M439.29T527.24	6
M246.19T528.91	6
M188.11T533.2 - [C3-HQ]	29
M505.3T529.02	6
M311.14T530.79	13
M246.11T531.57	6
M232.13T533.84 - [C5-QNO]	29
M186.09T533.45	29
M190.11T538.04	29
M275.03T539.68	36
M321.1T539.68	13
M214.16T538.68	2
M186.12T542.24	29
M185.12T544.52	29
M342.24T546.7	6
M288.16T548.09	37
M286.13T549.37	37
M530.14T549.86	38
M265.57T550.36	7
M412.22T552.14	13
M279.17T555.2	13
M453.31T556.48	6
M243.09T556.98 - lumichrome	13
M244.09T559.4	13
M230.08T561.04	6
M244.09T561.82	13
M215.21T565.23	2
M247.14T564.59	13
M204.1T566.76 - [C3-QNO]	29
M657.4T570.71	6
M260.17T585.83 - [C7-QNO]	29
M225.06T573.87	20
M392.26T576.54	13
M846.35T577.04	6
M258.15T607.57 - [C7:1-QNO]	29
M202.12T581.87 - [C4-HQ]	29
M264.12T578.82	12
M240.08T580.6	39
M274.14T612.9	29
M485.28T583.01	40
M323.09T587.23	20
M118.07T585.43	12
M202.09T586.73 - [C3:1-QNO]	29
M638.36T587.73	6
M428.25T591.43	6
M489.31T591.28	6
M276.16T592.57	29
M227.18T593.07	20
M464.19T596.62	41
M552.38T600.18	6
M228.18T601.46	6
M187.09T603.23	26
M230.08T608.07	6
M878.42T610.35	13
M485.12T627.62	38



Table 3.S6.: Continued.

feature ID	cluster ID
M288.16T618.24	37
M213.07T623.07	27
M879.35T623.57	6
M781.64T624.08	6
M215.12T624.84	27
M263.12T625.34	12
M210.06T627.62	20
M216.12T627.76	27
M297.14T633.98	29
M733.07T631.17	6
M260.17T652.4 - [C7-QNO]	29
M467.1T635.51	38
M240.08T636.01	39
M224.08T641.73 - phenazine-1-carboxamide	42
M259.11T639.56	29
M275.18T640.07	43
M207.06T642.73	42
M393.12T646.18	13
M415.11T647.18	44
M214.12T652.01 - [C5:1-HQ]	27
M289.14T651.37	29
M394.13T650.73	13
M242.15T653.79 - [C7:1-HQ]	29
M298.14T659.11	29
M230.12T657.83 - [C5:1-HQ]	29
M224.04T660.39	20
M297.64T661.39	29
M120.04T664.06	20
M288.2T682.62 - [C9-QNO]	29
M216.14T667.11 - [C5-HQ]	29
M222.02T668.5	20
M227.08T671.17	2
M218.14T672.2	29
M264.18T680.95 - nortriptyline ISTD	12
M233.13T673.84 - nortriptyline ISTD (fragment)	12
M286.18T733.29 - [C9:1-QNO]	29
M218.11T678.04	29
M204.09T679.17	12
M284.16T676.9 - [C9:2-QNO]	29
M232.13T684.65 - [C5-QNO]	12
M276.16T682.73	29
M208.04T684.01	15
M190.03T684.51	12
M186.13T687.18	45
M423.18T689.35	46
M900.04T689.85	6
M228.14T693.4 - [C6:1-HQ]	42
M274.14T692.9	29
M258.15T696.46 - [C7:1-PQS]	29
M225.07T700.4 - phenazine-1-carboxylic acid	42
M207.06T701.4 - phenazine-1-carboxylic acid (fragment)	42
M244.19T704.46	47
M266.17T703.57	12
M257.06T704.07	48
M256.12T705.35	48
M269.06T707.12 - phenazine-1,6-dicarboxylic acid	42
M325.07T711.18 - (pyochelin)	15
M270.06T710.68	42
M226.07T732.51	42
M302.18T770.02	29
M260.17T714.88 - [C7-QNO]	29
M146.03T716.52	15

**Table 3.S6.:** Continued.

feature ID	cluster ID
M230.15T725.8 - [C6-HQ]	42
M181.08T724.27	2
M206.03T723.63	42
M330.19T726.2	27
M225.07T729.59	42
M185.1T728.95 - (naproxen ISTD (fragment))	49
M186.1T730.23	49
M256.08T738.23	29
M242.15T764.53 - [C7:1-HQ]	29
M258.15T759.2 - [C7:1-PQS]	29
M246.15T740.92 - [C6-QNO]	29
M328.19T744.87	29
M314.21T750.7 - [C11:1-QNO]	29
M302.18T743.2	29
M288.2T766.31 - [C9-QNO]	29
M256.11T746.25	29
M240.11T746.75	27
M260.17T748.53 - [C7-PQS]	29
M330.2T757.31	29
M325.07T754.75 - (pyochelin)	15
M146.03T755.14	15
M316.21T758.2	29
M312.2T757.81	29
M330.2T769.77	29
M330.21T773.32	29
M284.16T827.44 - [C9:2-QNO]	29
M274.14T769.87	43
M260.16T771.8	29
M328.14T771.17	50
M515.29T771.67	29
M270.22T773.83	51
M309.13T780.55	52
M330.2T784.89	29
M256.17T785.39 - [C8:1-HQ]	29
M244.17T785.89 - HHQ	29
M328.19T786.68	29
M326.17T789.35	29
M244.16T792.51	29
M184.08T793.01	29
M316.23T812.44 - [C11-QNO]	29
M246.18T794.28	29
M312.2T797.34 - [C11:2-QNO]	29
M260.16T800.1 - HQNO	29
M300.16T820.96	29
M202.09T806.33	50
M288.2T839.88 - [C9-PQS]	29
M298.14T811.8	53
M288.19T810.28 - [C9-PQS]	29
M272.16T816.11	54
M326.17T832.92	29
M330.21T842.8	29
M314.21T821.35 - [C11:1-QNO]	29
M316.22T827.57 - [C11-QNO]	29
M296.2T827.18 - [C11:2-HQ]	29
M270.12T828.57	54
M270.19T832.13 - [C9:1-HQ]	29
M274.18T838.74 - [C8-QNO]	29
M328.18T835.58	29
M300.18T835.19	55
M258.19T837.46 - [C8-HQ]	29
M268.17T862.37 - [C9:2-HQ]	29
M456.34T842.69	56



Table 3.S6.: Continued.

feature ID	cluster ID
M344.24T842.3	57
M312.2T889.43 - [C11:2-QNO]	29
M288.19T864.53	29
M312.2T847.63 - [C11:2-QNO]	29
M310.18T851.58	29
M312.19T855.14	29
M260.19T857.02	29
M312.16T858.31	29
M437.19T858.8	58
M302.18T881.92	43
M178.05T865.03	43
M270.19T868.98 - [C9:1-HQ]	29
M324.16T869.48	59
M316.23T870.25 - [C11-QNO]	29
M272.19T885.48	29
M316.23T873.17 - [C11-QNO]	29
M296.2T874.81 - [C11:2-HQ]	29
M342.21T873.81	29
M326.16T880.92	59
M314.21T915.2 - [C11:1-QNO]	29
M310.18T888.04	29
M539.36T889.04	29
M332.22T890.32	29
M300.2T918.12 - [C10:1-QNO]	29
M352.24T896.04	60
M302.2T897.04	29
M312.26T898.71	55
M284.2T904.14	29
M246.12T899.71	29
M314.18T901.62	29
M288.2T905.42 - [C9-QNO]	29
M274.21T910.5	29
M270.19T912.14 - [C9:1-HQ]	29
M186.09T910.36	29
M342.24T912.03 - [C13:1-QNO]	29
M342.24T914.7 - [C13:1-QNO]	29
M328.18T920.94	29
M326.17T922.08	29
M296.2T950.38 - [C11:2-HQ]	29
M302.21T942.76 - [C10-QNO]	29
M358.24T928.92	29
M314.2T931.21 - [C11:1-QNO]	29
M672.37T932.34	61
M518.21T933.48	62
M499.29T935.65	63
M331.25T936.15	64
M350.17T938.82	59
M316.22T943.14	29
M330.21T962.73	43
M298.21T943	29
M284.16T945.04 - [C9:2-QNO]	29
M298.21T945.43	29
M300.22T949.63	29
M286.22T952.15 - [C10-HQ]	29
M595.43T951.65	29
M454.29T960.16 - (PE(16:0/0:0))	55
M617.41T953.04	65
M298.22T954.32 - [C11:1-HQ]	29
M286.18T957 - [C9:1-QNO]	29
M284.2T956.6 - [C10:1-HQ]	29
M272.2T958.39 - [C9-HQ]	29
M310.22T962.84 - [C12:2-HQ]	29

**Table 3.S6.:** Continued.

feature ID	cluster ID
M452.31T965.4	55
M342.27T966.79	64
M314.28T969.7	29
M326.21T970.34	66
M340.29T989	61
M270.19T972.62 - [C9:1-HQ]	29
M314.24T1002.33 - [C12-HQ]	29
M358.26T976.07	64
M316.22T979.73	29
M302.18T983.43	12
M300.22T993.94	29
M328.23T993.45 - [C12:1-QNO]	43
M312.23T998.78 - [C12:1-HQ]	29
M340.23T1000.06 - [C13:2-QNO]	29
M304.19T1002.83	29
M342.21T1004.49	29
M316.23T1005.88 - [C11-QNO]	29
M342.23T1009.83	29
M316.23T1008.8 - [C11-QNO]	29
M310.22T1008.16 - [C12:2-HQ]	29
M338.21T1010.32	29
M324.23T1028.97 - [C13:2-HQ]	29
M386.3T1014.28	64
M326.24T1016.54	29
M260.16T1204.08 - PQS	29
M300.23T1017.04 - [C11-HQ]	29
M302.24T1020.09	29
M298.22T1019.7 - [C11:1-HQ]	29
M342.27T1022.75	64
M452.31T1022.25	64
M300.19T1024.13 - [C10:1-QNO]	29
M328.23T1027.83	43
M314.21T1033.02 - [C11:1-QNO]	29
M388.25T1031.24	20
M342.24T1032.02 - [C13:1-QNO]	29
M330.21T1035.19	20
M344.25T1036.47	29
M316.22T1034.8	29
M314.16T1038.36	29
M326.38T1040.78	29
M301.17T1041.91	29
M326.25T1042.69 - [C13:1-HQ]	29
M312.26T1043.19	55
M328.25T1046.75	29
M312.26T1045.47	55
M370.3T1051.59	64
M459.23T1054.36	29
M286.18T1089.93 - [C9:1-PQS]	29
M270.19T1056.78 - [C9:1-HQ]	29
M300.2T1056.92 - [C10:1-QNO]	29
M314.28T1063.15	55
M507.27T1060.59	61
M300.16T1063.65	29
M342.24T1067.59 - [C13:1-QNO]	29
M386.3T1072.93	64
M312.2T1069.87	55
M312.23T1072.54 - [C12:1-HQ]	55
M382.27T1076.99	64
M414.33T1080.04	64
M342.27T1085.24	29
M344.25T1081.18	29
M326.25T1097.04 - [C13:1-HQ]	29



Table 3.S6.: Continued.

feature ID	cluster ID
M314.21T1083.71	55
M312.26T1091.61	55
M298.24T1088.79	29
M533.29T1092.6	61
M298.15T1094.77	29
M398.26T1097.81	67
M338.25T1098.31	61
M368.26T1100.73	29
M288.2T1101.87	29
M270.19T1102.37 - [C9:1-HQ]	29
M286.16T1110.12	29
M370.3T1109.37	64
M344.26T1112.05 - [C13-QNO]	29
M344.26T1115.61	29
M272.16T1118.53 - [C8:1-QNO]	29
M543.32T1120.16	29
M408.28T1122.72	64
M352.26T1122.58 - [C15:2-HQ]	29
M386.3T1123.22	64
M328.26T1127.17 - [C13-HQ]	29
M416.28T1129.06	20
M398.34T1132.5	67
M370.27T1132.11	68
M414.33T1154.68	64
M288.2T1193.17 - [C9-PQS]	29
M581.37T1145.93	56
M312.26T1140.51	55
M282.14T1141.5	69
M282.14T1143.65	69
M354.28T1145.81 - [C15:1-HQ]	29
M278.18T1150.74	29
M410.3T1158.24	56
M256.26T1157.46	29
M457.35T1173.22	70
M545.34T1162.8	29
M286.18T1167.63 - [C9:1-PQS]	29
M571.35T1166.36	71
M282.15T1189.47	72
M282.22T1177.03	72
M370.3T1186.31	68
M370.27T1181.36	68
M282.15T1182.36	72
M284.23T1187.7	72
M326.27T1189.61	73
M583.38T1192.53	56
M452.3T1193.81	74
M837.59T1198.37	75
M278.18T1200.78	29
M758.57T1200.14	29
M793.57T1202.55	29
M300.2T1205.35 - [C10:1-PQS]	29
M314.21T1209.8	29
M705.51T1214.75	76
M700.56T1215.63	24
M757.47T1216.13	29
M661.49T1216.91	77
M656.53T1217.91	24
M617.46T1225.42	78
M612.51T1224.53	24
M300.25T1226.31	79
M398.31T1227.59	67
M384.32T1228.97	80

**Table 3.S6.:** Continued.

feature ID	cluster ID
M573.44T1233.42	81
M568.48T1234.81	24
M356.3T1236.48 - [C15-HQ]	29
M284.22T1238.76	55
M529.41T1240.78	82
M452.37T1240.04	24
M286.18T1240.54 - [C9:1-PQS]	29
M524.45T1242.81	24
M408.34T1248.04	24
M382.31T1249.42 - [C17:1-HQ]	29
M436.4T1256.04	24
M438.41T1257.82	24
M310.31T1260.6	64
M270.19T1266.82 - [C9:1-HQ]	29
M282.15T1264.15	83
M398.33T1269.37	67
M314.21T1285.47 - [C11:1-PQS]	29
M392.37T1269.48	24
M466.32T1272.54	84
M458.33T1273.04	20
M298.22T1274.31 - [C11:1-HQ]	29
M441.36T1276.58	24
M354.31T1283.2	85
M865.62T1280.14	86
M338.29T1283.7	72
M733.54T1291.2	87
M728.59T1292.59	24
M394.3T1294.76	88
M413.27T1295.4 - (Bis(2-ethylhexyl)phthalate CONT)	89
M689.52T1297.42	90
M684.56T1297.93	24
M645.49T1300.98	91
M640.54T1302.37	24
M601.47T1306.82	92
M596.51T1308.61	24
M365.3T1310.39	93
M557.44T1313.44	94
M552.49T1313.94	24
M521.32T1315.21	29
M298.31T1317.63	29
M513.41T1321.04	95
M508.46T1319.27	24
M288.2T1332.99 - [C9-PQS]	29
M535.33T1322.96	96
M271.26T1323.71	29
M330.28T1323.6	20
M469.39T1327.66	97
M297.28T1326.38	29
M284.3T1328.8	29
M270.18T1331.71 - [C9:1-HQ]	29
M314.21T1333.38	55
M426.36T1337.83	98
M312.25T1336.55	55
M338.34T1340.11	64
M360.32T1340.61	93
M428.31T1339.61	93
M675.68T1342.38	99
M502.35T1344.54	100
M733.54T1344.15	87
M480.34T1347.2	101
M382.34T1349.62	102
M344.3T1350.12	20



Table 3.S6.: Continued.

feature ID	cluster ID
M286.18T1349.48 - [C9:1-PQS]	29
M849.63T1353.92	103
M805.6T1356.46	104
M761.58T1359.12	105
M756.62T1360.12	24
M717.55T1363.18	106
M394.35T1365.35	72
M673.52T1369.02	107
M396.35T1368.02	72
M342.3T1369.15	79
M629.5T1372.71	108
M585.47T1377.79	109
M358.31T1378.18	20
M133.96T1384.9	110
M541.44T1383.51	111
M486.36T1389.48	20
M701.43T1389.98	93
M497.42T1393.29	93
M550.39T1393.68	20
M131.96T1394.67	110
M428.26T1397.24	93
M512.37T1404.63	20
M699.41T1405.27	112
M553.39T1407.68	113
M355.31T1411.24	79
M356.31T1414.3	79
M568.44T1417.36	93
M133.96T1425.52	110
M442.28T1430.26	114
M339.28T1433.6	55
M629.37T1436.79	115
M121.97T1440.85	110
M131.96T1445.03	110
M469.38T1459.31	116
M608.42T1458.17	116
M456.29T1464.28	70
M369.33T1465.67	79
M512.37T1465.28	20
M500.38T1491.86	20
M133.96T1480.41	110
M353.3T1491.47	55
M121.97T1495.28	110
M131.96T1498.6	110



**Table 3.S7.:** List of putatively identified features, their sum formula and isotopic pattern fit in mSigma as determined by the SmartFormula functionality of the Bruker DataAnalysis software. Isotopic pattern fit is an additional way of validating sum formulae of putatively identified compounds. 50 mSigma is a conventional threshold.

feature ID	putative annotation	ion sum formula	isotopic pattern fit (mSigma)	comment
M179.06T71.32	gluconolactone	C <sub>6</sub> H <sub>11</sub> O <sub>6</sub> <sup>+</sup>	4.2	
M116.07T73.51	proline	C <sub>5</sub> H <sub>10</sub> NO <sub>2</sub> <sup>+</sup>	6.3	
M219.1T77.65	Glu Ala	C <sub>8</sub> H <sub>15</sub> N <sub>2</sub> O <sub>5</sub> <sup>+</sup>	10.5	
M332.08T113.21	dAMP	C <sub>10</sub> H <sub>15</sub> N <sub>5</sub> O <sub>6</sub> P <sup>+</sup>	26.2	
M535.19T118.41	Glu Glu Glu Glu	C <sub>20</sub> H <sub>31</sub> N <sub>4</sub> O <sub>13</sub> <sup>+</sup>	6.2	
M180.05T120.87	(iso)xanthopterin	C <sub>6</sub> H <sub>6</sub> N <sub>5</sub> O <sub>2</sub> <sup>+</sup>	13.8	
M307.08T126.2	glutathion disulphide (2+)	C <sub>20</sub> H <sub>34</sub> N <sub>6</sub> O <sub>12</sub> S <sub>2</sub> <sup>++</sup>	4.8	
M323.06T135.83	dTMP	C <sub>10</sub> H <sub>16</sub> N <sub>2</sub> O <sub>8</sub> P <sup>+</sup>	6.5	
M193.07T150.49	S-(5'-adenosyl)-homocysteine (2+)	C <sub>14</sub> H <sub>22</sub> N <sub>6</sub> O <sub>5</sub> S <sup>++</sup>	5.2	
M597.68T335.59	UDP-muramyl-pentapeptide	C <sub>41</sub> H <sub>67</sub> N <sub>9</sub> O <sub>28</sub> P <sub>2</sub> <sup>++</sup>	34.9	
M219.13T335.84	Ser Leu	C <sub>9</sub> H <sub>19</sub> N <sub>2</sub> O <sub>4</sub> <sup>+</sup>	9.8	
M220.12T336.88	panthotenate	C <sub>9</sub> H <sub>18</sub> NO <sub>5</sub> <sup>+</sup>	3	
M360.21T355.1	Ile Val Glu / Val Ile Glu	C <sub>16</sub> H <sub>30</sub> N <sub>3</sub> O <sub>6</sub> <sup>+</sup>	25.5	
M358.2T377.48	Ile Pro Glu / Leu Pro Glu	C <sub>16</sub> H <sub>28</sub> N <sub>3</sub> O <sub>6</sub> <sup>+</sup>	n.d.	signal too weak
M254.09T400.89	aeruginosin A	C <sub>14</sub> H <sub>12</sub> N <sub>3</sub> O <sub>2</sub> <sup>+</sup>	23.3	
M295.13T406.65	Glu Phe	C <sub>14</sub> H <sub>19</sub> N <sub>2</sub> O <sub>5</sub> <sup>+</sup>	10.6	
M378.2T412.48	Pro Tyr Val	C <sub>19</sub> H <sub>28</sub> N <sub>3</sub> O <sub>5</sub> <sup>+</sup>	13.3	
M316.22T413.91	Val Val Val	C <sub>15</sub> H <sub>30</sub> N <sub>3</sub> O <sub>4</sub> <sup>+</sup>	n.d.	signal too weak
M328.22T437.8	Pro Leu Val	C <sub>16</sub> H <sub>30</sub> N <sub>3</sub> O <sub>4</sub> <sup>+</sup>	3.4	
M336.19T415.21	Phe Val Ala / Val Phe Ala	C <sub>17</sub> H <sub>26</sub> N <sub>3</sub> O <sub>4</sub> <sup>+</sup>	n.d.	signal too weak
M457.11T433.91	FMN	C <sub>17</sub> H <sub>22</sub> N <sub>4</sub> O <sub>9</sub> P <sup>+</sup>	11.2	
M263.14T447.96	Pro Phe	C <sub>14</sub> H <sub>19</sub> N <sub>2</sub> O <sub>3</sub> <sup>+</sup>	19.3	
M231.17T475.37	Leu Val	C <sub>11</sub> H <sub>23</sub> N <sub>2</sub> O <sub>3</sub> <sup>+</sup>	26.3	
M344.25T476.52	Leu Leu Val or isomer	C <sub>17</sub> H <sub>34</sub> N <sub>3</sub> O <sub>4</sub> <sup>+</sup>	13.8	
M255.08T482.73	pyocyanin carboxylic acid	C <sub>14</sub> H <sub>10</sub> N <sub>2</sub> O <sub>3</sub> <sup>+</sup>	5.9	
M211.14T497.45	cyclo(Leu Pro)	C <sub>11</sub> H <sub>19</sub> N <sub>2</sub> O <sub>2</sub> <sup>+</sup>	14.4	
M342.24T486.25	Pro Ile Leu or isomer	C <sub>17</sub> H <sub>32</sub> N <sub>3</sub> O <sub>4</sub> <sup>+</sup>	8.0	
M245.13T525.64	cyclo(Phe Pro)	C <sub>14</sub> H <sub>17</sub> N <sub>2</sub> O <sub>2</sub> <sup>+</sup>	6.1	
M188.11T535.78	C3-HQ	C <sub>12</sub> H <sub>14</sub> NO <sup>+</sup>	15.0	
M260.16T583.7	C7-QNO	C <sub>16</sub> H <sub>22</sub> NO <sub>2</sub> <sup>+</sup>	n.d.	signal too weak
M392.25T573.83	Phe Leu Leu or isomer	C <sub>21</sub> H <sub>34</sub> N <sub>3</sub> O <sub>4</sub> <sup>+</sup>	18.4	
M200.11T578.31	C4:1-HQ	C <sub>13</sub> H <sub>14</sub> NO <sup>+</sup>	17.0	
M202.12T578.44	C4-HQ	C <sub>13</sub> H <sub>16</sub> NO <sup>+</sup>	15.7	
M215.12T626.24	tetrahydropyocyanin	C <sub>13</sub> H <sub>15</sub> N <sub>2</sub> O <sup>+</sup>	18.4	
M214.12T663.32	C5:1-HQ	C <sub>14</sub> H <sub>16</sub> NO <sup>+</sup>	10.5	
M216.14T670.01	C5-HQ	C <sub>14</sub> H <sub>18</sub> NO <sup>+</sup>	4.7	
M232.13T683.77	C5-QNO	C <sub>14</sub> H <sub>18</sub> NO <sub>2</sub> <sup>+</sup>	4.8	
M288.2T705.59	C9-QNO	C <sub>18</sub> H <sub>26</sub> NO <sub>2</sub> <sup>+</sup>	25.7	
M325.07T739.09	pyochelin	C <sub>14</sub> H <sub>17</sub> N <sub>2</sub> O <sub>3</sub> S <sub>2</sub> <sup>+</sup>	16.7	
M304.19T717.91	C9-QNO side chain oxidation product	C <sub>18</sub> H <sub>26</sub> NO <sub>3</sub> <sup>+</sup>	16.1	
M286.18T728.73	C9:1-QNO	C <sub>18</sub> H <sub>24</sub> NO <sub>2</sub> <sup>+</sup>	23.2	
M185.1T729.78	naproxen (fragment)	C <sub>13</sub> H <sub>13</sub> O <sup>+</sup>	6.0	
M230.15T735.39	C6-HQ	C <sub>15</sub> H <sub>20</sub> NO <sup>+</sup>	3.0	
M288.2T765.88	C9-QNO	C <sub>18</sub> H <sub>26</sub> NO <sub>2</sub> <sup>+</sup>	25.7	
M258.15T768.47	C7:1-QNO	C <sub>16</sub> H <sub>20</sub> NO <sub>2</sub> <sup>+</sup>	3.8	
M270.19T766.9	C9:1-HQ	C <sub>18</sub> H <sub>24</sub> NO <sup>+</sup>	5.3	
M242.15T789.6	C7:1-HQ	C <sub>16</sub> H <sub>20</sub> NO <sup>+</sup>	2.7	
M304.19T786.75	C9-QNO side chain oxidation product	C <sub>18</sub> H <sub>26</sub> NO <sub>3</sub> <sup>+</sup>	33.5	
M312.2T794.57	C11:2-QNO	C <sub>20</sub> H <sub>26</sub> NO <sub>2</sub> <sup>+</sup>	17.1	
M503.33T797.17	HHQ/HQNO mixed dimer	C <sub>32</sub> H <sub>43</sub> N <sub>2</sub> O <sub>3</sub> <sup>+</sup>	37.4	
M286.18T808.85	C9:1-QNO	C <sub>18</sub> H <sub>24</sub> NO <sub>2</sub> <sup>+</sup>	17.7	
M325.07T819.96	pyochelin	C <sub>14</sub> H <sub>17</sub> N <sub>2</sub> O <sub>3</sub> S <sub>2</sub> <sup>+</sup>	10.5	
M272.16T822.08	C8:1-QNO	C <sub>17</sub> H <sub>22</sub> NO <sub>2</sub> <sup>+</sup>	22.2	
M314.21T825.99	C11:1-QNO	C <sub>20</sub> H <sub>28</sub> NO <sub>2</sub> <sup>+</sup>	22.9	



Table 3.S7.: Continued.

feature ID	putative annotation	ion sum formula	isotopic pattern fit (mSigma)	comment
M288.2T824.42	C9-QNO	$C_{18}H_{26}NO_2^+$	3.4	
M274.18T842.43	C8-QNO	$C_{17}H_{24}NO_2^+$	15.0	
M258.19T853.38	C8-HQ	$C_{17}H_{24}NO^+$	19.8	
M330.21T844.12	C11:1-QNO side chain oxidation product	$C_{20}H_{28}NO_3^+$	25.6	
M286.18T864.03	C9:1-QNO	$C_{18}H_{24}NO_2^+$	7.7	
M270.19T873.15	C9:1-HQ	$C_{18}H_{24}NO^+$	5.5	
M256.17T851.43	C8:1-HQ	$C_{17}H_{22}NO^+$	146.3	low signal, no other sum formula possible
M316.23T869.51	C11-QNO	$C_{20}H_{30}NO_2^+$	13.3	
M316.23T871.07	C11-QNO	$C_{20}H_{30}NO_2^+$	6.2	
M314.21T892.19	C11:1-QNO	$C_{20}H_{28}NO_2^+$	22.9	
M300.2T925.27	C10:1-QNO	$C_{19}H_{26}NO_2^+$	23.7	
M288.2T902.1	C9-PQS	$C_{18}H_{26}NO_2^+$	23.3	
M300.2T894.8	C10:1-QNO	$C_{19}H_{26}NO_2^+$	17.4	
M284.2T904.43	C10:1-HQ	$C_{19}H_{26}NO^+$	34.3	
M272.2T910.42	C9-HQ	$C_{18}H_{26}NO^+$	10.9	
M342.24T912.12	C13:1-QNO	$C_{22}H_{32}NO_2^+$	35.3	
M286.18T921.6	C9:1-PQS	$C_{18}H_{24}NO_2^+$	25.5	
M358.24T929.17	C13:1-QNO side chain oxidation product	$C_{22}H_{32}NO_3^+$	10.0	
M314.21T944.87	C11:1-PQS	$C_{20}H_{28}NO_2^+$	2.7	
M298.22T984.33	C11:1-HQ	$C_{20}H_{28}NO^+$	31.9	
M454.29T963.69	PE(16:0/0:0)	$C_{21}H_{45}NO_7P^+$	6.6	
M302.21T956.81	C10-QNO	$C_{19}H_{28}NO_2^+$	20.4	
M284.2T959.42	C10:1-HQ	$C_{19}H_{26}NO^+$	13.5	
M673.38T966.68	Rha-Rha-C10-C10 +Na	$C_{32}H_{58}NaO_{13}^+$	13.7	
M480.31T982.77	PE(18:1/0:0)	$C_{23}H_{47}NO_7P^+$	7.4	
M505.25T982.5	PG(16:1/0:0) +Na	$C_{22}H_{43}NaO_9P^+$	11.8	
M483.27T981.33	PG(16:1/0:0)	$C_{22}H_{44}O_9P^+$	26.9	
M502.29T985.62	PE(18:1/0:0) +Na	$C_{23}H_{46}NNaO_7P^+$	49.4	
M328.23T993.95	C12:1-QNO	$C_{21}H_{30}NO_2^+$	40.6	
M312.23T1001.35	C12:1-HQ	$C_{21}H_{30}NO^+$	8.4	
M316.23T1007.37	C11-PQS	$C_{20}H_{30}NO_2^+$	3.5	
M699.39T1013.83	Rha-Rha-C10-C12:1 / Rha-Rha-C12:1-C10 +Na	$C_{34}H_{60}NaO_{13}^+$	34.8	
M677.41T1014.86	Rha-Rha-C10-C12:1 / Rha-Rha-C12:1-C10	$C_{34}H_{61}O_{13}^+$	49.8	
M496.34T1018.23	PC(16:0/0:0)	$C_{24}H_{51}NO_7P^+$	16.3	
M527.32T1021.62	Rha-C10-C10 /Rha-C12-C8 +Na	$C_{26}H_{48}NaO_9^+$	8.4	
M326.25T1043.73	C13:1-HQ	$C_{22}H_{32}NO^+$	6.5	
M701.41T1049.33	Rha-Rha-C10-C12 / Rha-Rha-C12-C10 +Na	$C_{34}H_{62}NaO_{13}^+$	13.3	
M679.43T1051.39	Rha-Rha-C10-C12 / Rha-Rha-C12-C10	$C_{34}H_{63}O_{13}^+$	27.3	
M553.34T1072.25	Rha-C10-C12:1 / Rha-C12:1-C10 +Na	$C_{28}H_{50}NaO_9^+$	17.6	
M342.24T1078.72	C13:1-PQS	$C_{22}H_{32}NO_2^+$	25.9	
M727.42T1093.9	Rha-Rha-C12:1-C12 / Rha-Rha-C12-C12:1 +Na	$C_{36}H_{64}NaO_{13}^+$	43.2	
M705.44T1095.83	Rha-Rha-C12:1-C12 / Rha-Rha-C12-C12:1	$C_{36}H_{65}O_{13}^+$	16.9	
M326.25T1125.41	C13:1-HQ	$C_{22}H_{32}NO^+$	96.3	low signal, no other sum formula possible
M555.35T1108.39	Rha-C10-C12 / Rha-C12-C10 +Na	$C_{28}H_{52}NaO_9^+$	23.5	
M328.26T1127.63	C13-HQ	$C_{22}H_{34}NO^+$	17.3	
M729.44T1131.57	Rha-Rha-C12-C12 +Na	$C_{36}H_{66}NaO_{13}^+$	15.3	
M707.46T1132.6	Rha-Rha-C12-C12	$C_{36}H_{67}O_{13}^+$	39.5	
M555.36T1151.77	Rha-C10-C12 / Rha-C12-C10 +Na	$C_{28}H_{52}NaO_9^+$	7.2	

## Acknowledgements

I would like to thank my doctoral supervisor Prof. Dr. Mark Brönstrup and my co-supervisor Prof. Dr. Susanne Häußler, who have made this research project possible and have supported me competently and reliably over the years. Thanks also go to Prof. Dr. Wulf Blankenfeldt, who as a member of my Thesis Committee was a valuable discussion partner and source of ideas.

I especially thank Dr. Raimo Franke, who not only accompanied this work in terms of content and methodology, but was also a very good teacher and, not least, a great support when there had been a little dip in the motivation curve from time to time.

I would like to express my gratitude to the other members of the CBIO Metabolomics subgroup, above all Dr. Volker Berndt, with whom I have been through a lot, but of course also Mariel García, Heike Overwin, Ulrike Beutling, Dr. Aamna Habib, Dr. Federica Fiorini and Dr. Michael Hensler, without whose support this work would not have been possible. I am also very grateful to Tabea Linde, who has supported our small metabolomics unit with her great work.

I owe thanks to Aditya Shekhar, Jana Krull, Yi-Hui Lai, Manuel Orozco and Dr. Hans Prochnow, who, together with the above-mentioned, ensured a good atmosphere during the lunch breaks and on other occasions. I would also like to express my thanks to the other CBIOs and BISAs/COPs, including Prof. Dr. Ursula Bilitewski, Dr. Charlotte Grandclaoudon, Dr. Kevin Ferreira, Dr. Verena Fetz, Dr. Sven-Kevin Hotop, Bettina Mehner and all those who, for reasons of space, are not listed here.

Special thanks go to Dr. Florian Witzgall of the SFPR group, with whom the cooperation was really great fun. Many thanks to my colleagues in the MOBA group, especially Dr. Janne Thöming, Dr. Adrian Kordes, Dr. Ariane Khaledi and Lorena Petzold, for their excellent cooperation and support.

I would also like to acknowledge the HZI Graduate School for their guidance from the very beginning and for organising instructive and/or fun events.

Last but not least I owe thanks to Marilena and my family and friends for their support and patience during this rather long and exhausting time: this work is dedicated to you.

# Curriculum vitae

## Personal Data

Name **Tobias Depke**  
Date of Birth **14 January 1990**  
Place of Birth **Bünde (Westfalen), Germany**

## Education

since 2015 **PhD Student**, *Gottfried-Wilhelm-Leibniz-Universität, Hannover*  
Title: **Metabolome-based studies of virulence factors in *Pseudomonas aeruginosa***  
Supervision: Prof. Dr. rer. nat. Mark Brönstrup und Prof. Dr. med. Susanne Häußler

2015 **License to practise pharmacy (Approbation als Apotheker)**, *Bezirksregierung Münster*

2015 **3<sup>rd</sup> State Examination in Pharmacy**, *Westfälische Wilhelms-Universität, Münster*

2014 **2<sup>nd</sup> State Examination in Pharmacy**, *Westfälische Wilhelms-Universität, Münster*

2011 **1<sup>st</sup> State Examination in Pharmacy**, *Westfälische Wilhelms-Universität, Münster*

2009-2014 **Student of Pharmacy**, *Westfälische Wilhelms-Universität, Münster*

2009 **University entrance qualification (Abitur)**, *Widukind-Gymnasium, Enger*

## Professional Experience

09/2020–present **Scientist**, *Institute of Clinical Pharmacology, RWTH Univeristy Hospital, Aachen, Germany*  
Drug bioanalytics, LC–MS/MS, targeted metabolomics

02/2019–08/2020 **Pharmacist**, *Hospital Pharmacy, Univeristy Hospital, Cologne, Germany*  
E-medication stewardship, data analysis, clinical pharmacy

07/2015–01/2019 **Doctoral researcher**, *Department of Chemical Biology, Helmholtz-Zentrum für Infektionsforschung, Braunschweig, Germany*  
Microbial metabolomics, bioinformatics/chemometrics, research and teaching

11/2014–6/2015 **Trainee pharmacist**, *St. Michael-Apotheke am Ochsenmarkt, Flensburg, Germany*  
General pharmacy practice

5/2014–10/2014 **Trainee pharmacist**, *Department of Drug Delivery & Innovation, Merck KGaA, Darmstadt, Germany*  
Analytical development, pre-clinical development, drug delivery systems

4/2013–9/2013 **Erasmus+ internship**, *Division of Molecular Structural Biology, Karolinska Institutet, Stockholm, Sweden*  
Protein biochemistry, structural biology

## Awards and Scholarships

2016–2019 **PhD Scholarship of the German National Academic Foundation (Studienstiftung des deutschen Volkes)**

2014 **Prize of the Pharmacist Foundation Westfalen-Lippe for special achievements in the 2nd State Examination**

2009–2014 **Scholarship of the German National Academic Foundation (Studienstiftung des deutschen Volkes)**

## List of Publications

### Journal articles

1. T. Depke, R. Franke and M. Brönstrup. "Clustering of MS<sup>2</sup> spectra using unsupervised methods to aid the identification of secondary metabolites from *Pseudomonas aeruginosa*". In: *Journal of Chromatography B* 1071 (Dec. 2017), pp. 19–28. DOI: 10.1016/j.jchromb.2017.06.002
2. T. Depke, R. Franke and M. Brönstrup. "CluMSID: an R package for similarity-based clustering of tandem mass spectra to aid feature annotation in metabolomics". In: *Bioinformatics* (Jan. 2019). ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btz005
3. F. Witzgall, T. Depke, M. Hoffmann, M. Empting, M. Brönstrup, R. Müller and W. Blankenfeldt. "The Alkylquinolone Repertoire of *Pseudomonas aeruginosa* is Linked to Structural Flexibility of the FabH-like 2-Heptyl-3-hydroxy-4(1H)-quinolone (PQS) Biosynthesis Enzyme PqsBC". in: *ChemBioChem* 19.14 (May 2018), pp. 1531–1544. DOI: 10.1002/cbic.201800153
4. T. Depke, S. Häußler and M. Brönstrup. "The Peptide Chain Release Factor Methyltransferase PrmC Influences the *Pseudomonas aeruginosa* PA14 Endo- and Exometabolome". In: *Metabolites* 10.10 (Oct. 2020), p. 417. DOI: 10.3390/metabo10100417
5. T. Depke, J. G. Thöming, A. Kordes, S. Häußler and M. Brönstrup. "Untargeted LC-MS Metabolomics Differentiates Between Virulent and Avirulent Clinical Strains of *Pseudomonas aeruginosa*". In: *Biomolecules* 10.7 (July 2020), p. 1041. DOI: 10.3390/biom10071041

### Conference presentations

1. T. Depke, S. Häußler and M. Brönstrup. *Metabolome-based studies of virulence factors in Pseudomonas aeruginosa*. Poster presented at the 7<sup>th</sup> Annual Retreat, HZI Graduate School, Harzhöhe, Goslar–Hahnenklee. June 2016
2. T. Depke, S. Häußler and M. Brönstrup. *Metabolome-based studies of virulence factors in Pseudomonas aeruginosa*. Talk presented at the 8<sup>th</sup> Annual Retreat, HZI Graduate School, Harzhöhe, Goslar–Hahnenklee. May 2017
3. T. Depke, S. Häußler and M. Brönstrup. *Metabolome-based studies of virulence factors in Pseudomonas aeruginosa*. Talk presented at the 9<sup>th</sup> Annual Retreat, HZI Graduate School, Hotel Schlossmühle, Quedlinburg. Apr. 2018
4. T. Depke, S. Häußler and M. Brönstrup. *Metabolome-based studies of virulence factors in Pseudomonas aeruginosa*. Poster presented at the 8<sup>th</sup> International PhD Symposium, HZI Graduate School, Braunschweig. Dec. 2015
5. T. Depke, R. Franke and M. Brönstrup. *Clustering of MS<sup>2</sup> spectra using unsupervised methods to aid the identification of secondary metabolites from Pseudomonas aeruginosa*. Poster presented at the 9<sup>th</sup> International PhD Symposium, HZI Graduate School, Braunschweig. Dec. 2016
6. T. Depke, R. Franke and M. Brönstrup. *Clustering of MS<sup>2</sup> spectra using unsupervised methods to aid the identification of secondary metabolites from Pseudomonas aeruginosa*. Poster presented at the 10<sup>th</sup> International PhD Symposium, HZI Graduate School, Braunschweig. Dec. 2017
7. T. Depke, J. Thöming, S. Häußler and M. Brönstrup. *Untargeted LC-MS metabolomics differentiates between virulent and avirulent clinical strains of Pseudomonas aeruginosa*. Talk presented at the 11<sup>th</sup> International PhD Symposium, HZI Graduate School, Braunschweig. Dec. 2018
8. T. Depke, S. Häußler and M. Brönstrup. *Influence of the peptide chain release factor methyltransferase PrmC on virulence associated metabolic pathways of Pseudomonas aeruginosa PA14*. Poster presented at the Annual Meeting of the German Pharmaceutical Society (DPHG), Munich. Oct. 2016
9. T. Depke, J. Thöming, S. Häußler and M. Brönstrup. *Untargeted LC-MS Metabolomics Differentiates Between Virulent and Avirulent Clinical Strains of Pseudomonas aeruginosa*. Talk and Poster presented at the Herrenhausen Symposium "Individualized Infection Medicine – The future is now", Hannover–Herrenhausen. June 2018
10. T. Depke, J. Thöming, S. Häußler and M. Brönstrup. *Untargeted LC-MS Metabolomics Differentiates Between Virulent and Avirulent Clinical Strains of Pseudomonas aeruginosa*. Poster presented at the Metabolomics Society Conference 2018, Seattle, USA. June 2018

1	General Introduction .....	1
2	Aims of the thesis .....	41
3	Publication 1: Clustering of MS <sup>2</sup> Spectra to Aid Metabolite Identification .....	45
4	Publication 2: CluMSID: an R package for clustering of tandem mass spectra .....	95
5	Publication 3: PqsBC determines the AQ repertoire of <i>P. aeruginosa</i> .....	153
6	Publication 4: The influence of PrmC on metabolism and virulence .....	215
7	Publication 5: Metabolomics of virulent and avirulent clinical <i>P. aeruginosa</i> strains .....	241
8	General Discussion and Outlook .....	301
A	Supplementary Tables for Publication 1 .....	321