



Technische Universität München
Lehrstuhl für organische Chemie II

**Functional analysis of selective aldehyde dehydrogenase 1A1
inhibition by cytotoxic duocarmycin analogs
and LC-MS-based metabolomic profiling of polar metabolites
in bacteria**

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie
der Technischen Universität München

von

M.Sc. Maximilian Felix Koch
geboren am 03.12.1986
in Grünstadt

2015



Technische Universität München
Lehrstuhl für organische Chemie II

**Functional analysis of selective aldehyde dehydrogenase
1A1 inhibition by cytotoxic duocarmycin analogs
and LC-MS-based metabolomic profiling of polar metabolites
in bacteria**

Maximilian Felix Koch

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Tobias A.M. Gulder
Prüfer der Dissertation: 1. Univ.-Prof. Dr. Stephan A. Sieber
2. apl. Prof. Dr. Wolfgang Eisenreich

Die Dissertation wurde am 17.09.2015 bei der bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 29.10.2015 angenommen.

.

Meinen Eltern gewidmet

„Wesentliche Dinge im Leben sind nicht zuletzt der Humor und die Fähigkeit, über sich selbst zu lachen.“ – YEHUDI MENUHIN

Danksagung

In erster Linie möchte ich Herrn Prof. Dr. Stephan A. Sieber sehr herzlich dafür danken, dass ich die Möglichkeit hatte sowohl meine Masterarbeit als auch meine Doktorarbeit an seinem Lehrstuhl machen zu können. In dieser Zeit hatte ich das Glück, diverse komplexe analytische Techniken als auch interessante Fragestellungen bearbeiten zu dürfen. Außerdem möchte ich mich für die Unterstützung und Betreuung bei den vielseitigen Projekten als auch für die stets positive Arbeitsatmosphäre in seinem Labor bedanken.

Den Mitgliedern der Prüfungskommission danke ich für die Bemühungen bei der Bewertung dieser Arbeit.

Besonderer Dank gilt meinen Kooperationspartnern des Lehrstuhls für Biochemie der Technischen Universität München, Dr. Sabine Schneider und Sabrina Harteis für die herausragende Zusammenarbeit und die vielen fruchtbaren Diskussionen. Besonders möchte ich mich dabei bei Dr. Sabine Schneider für die Rund- und Einführung am Paul Scherrer Institute und für die stets positive und produktive Zusammenarbeit bedanken. Mein weiterer Dank gilt den Kooperationspartnern der Ludwig-Maximilians-Universität des Lehrstuhls für Theoretische Chemie, Prof. Dr. Christian Ochsenfeld und Iris Blank sowie den Kooperationspartnern der Technischen Universität München für Medizinische Mikrobiologie und Immunologie, Prof. Dr. Markus Gerhard und Stefanie Wüstner.

Des Weiteren möchte ich mich bei Katja Bäuml und Mona Wolff für euren unermüdlichen Einsatz im Labor bedanken. Dafür, dass Ihr das Labor am Laufen haltet und uns Doktoranden stets unterstützt und zur Seite steht. Ohne euch beide wäre kein Laboralltag möglich.

Ein großer Dank geht an meine lieben Laborkollegen und dabei besonders an Franziska Mandl, alias Franjo, Franz, Frax. Du bist meine Marianne Rosenberg unter den Schlagersängerinnen und mehr als Chemie: nämlich Kunst und für Höheres bestimmt (Mount Everest). Ich werde unsere gemeinsame Zeit im Labor und die während unseres Studiums nicht vergessen und die guten wie die schlechten Zeiten in sehr guter Erinnerung behalten. Außerdem möchte ich mich bei Elena Kunold für ihre russische Art bedanken. Dafür, dass du dich für unseren Lehrstuhl entschieden hast und uns viele unvergessliche Momente beschert hast. Lass dich von dem Elektrospray der Fusion nicht unterdrücken. Mein weiterer Dank gilt meiner zweiten Metabolomix Hälfte Volker Kirsch, mit dem stets gut Kirschen essen war. Wenn du dich mal auf den Weiten der Metabolitidentifizierung verloren fühlen solltest, kannst du dich jederzeit bei mir melden (aber nicht zu oft bitte). Außerdem danke ich meinen Laborkollegen Markus Lakemeyer, für deine vielen Ohrwürmer, Maria Dahmen für dein amüsanter rotwerden und Johannes

Lehmann für deinen unermüdlichen Glauben an den BVB. Ich danke euch für die exzellente Laboratmosphäre und dafür, dass ihr meinen, wenn auch nicht immer einfachen Musikgeschmack und Humor stets stoisch und mit Humor ertragen habt.

Zudem möchte ich mich bei Mathias Hackl, Wolfgang Heydenreuter, Annabell Hoegl, Matthias Stahl, Jan Vomacka, Weining Zhao, Christian Fetzer, Philipp Kleiner, Dr. Megan Wright, Dr. Joanna Krysiak und Dr. Nina Bach für die ausgezeichnete und stets positive Arbeitsatmosphäre im Labor bedanken.

Nicht zu vergessen ist der Dank an die Laborkollegen, die bereits das Labor verlassen haben. Dabei möchte ich besonders Dr. Georg Rudolf zum einen für die grandiose Arbeitsatmosphäre in Labor D und zum anderen für die außerordentliche Zusammenarbeit bei unserem gemeinsamen NASM Projekt bedanken. Des Weiteren möchte ich mich bei Dr. Tanja Wirth, Dr. Matthew Nodwell, Dr. Maximilian Pitscheider, Dr. Roman Kolb und Dr. Bianca Schwanhäuser für eure rege Unterstützung zu Beginn meiner Doktorarbeit und für die stets aufbauenden Worte bedanken.

Außerdem möchte ich Franziska Mandl, Annabelle Hoegl und Dr. Megan Wright fürs Korrekturlesen meiner Arbeit bedanken.

Mein letzter Dank geht an meine Eltern Ruth und Horst Koch, die es mir ermöglicht haben diesen langen Weg des Chemiestudiums mit anschließender Promotion gehen zu dürfen. Vielen Dank für Eure seelische und moralische Unterstützung und dass ihr an mich geglaubt habt. Außerdem danke ich meinen Schwestern Pia und Eva Koch sowie meiner langjährigen Mitbewohnerin Katharina Passlack, die meine ersten englischen Vorträge stets mit aufbauendem Feedback und bemerkenswerter Geduld unterstützt und ertragen hat.

1. Table of contents

1. Table of contents	III
2. Introductory remarks	VII
3. Zusammenfassung	VIII
4. List of abbreviations, acronyms and symbols.....	X

I - Structural, biochemical and computational studies reveal the mechanism of selective aldehyde dehydrogenase 1A1 inhibition by cytotoxic duocarmycin analogs	1
---	----------

1. Introduction.....	3
1.1 The aldehyde dehydrogenase superfamily	3
1.2 Aldehyde dehydrogenases and their role in diseases	5
1.3 Structural insight into the aldehyde dehydrogenase superfamily	7
1.4 Reversible and non-reversible small molecule inhibitors of aldehyde dehydrogenase isoforms	10
2. Scope of this work.....	18
3. Results and Discussion.....	19
4. Summary	26
5. Experimental section.....	28
5.1 Cloning, expression and purification of ALDH1A1 enzymes	28
5.2 Crystallization, data collection and structure determination of ALDH1A1	28
5.3 Aldehyde dehydrogenase activity assay	29
5.4 Determination of K_i and k_{inact} of ALDH1A1	30
5.5 Intact protein mass spectrometry.....	30
5.6 Time dependent intact protein alkylation experiment of human ALDH1A1	31
5.7 Gel-based fluorescent labeling experiment of seco drug 2 probe and human ALDH1A1	31

5.8	Bottom up proteomics for binding site identification.....	31
5.9	MTT Assay with seco drug 2.....	33
5.10	DNA alkylation experiment with seco drug 1, seco drug 2 and duocarmycin SA .	34
5.11	Putative design criteria for selective ALDH1A1 inhibition or DNA-alkylation.....	34
5.12	Details for Molecular Dynamics Simulations.....	35
6.	Appendices.....	36
6.1	Figures.....	36
6.2	Tables.....	47
II -	Metabolomics	49
1.	Introduction	51
1.1	General aspects of metabolomics.....	51
1.2	Targeted and untargeted strategies in metabolomics.....	52
1.3	Metabolic footprinting in <i>Streptomyces</i> extracts as a tool for natural product discovery.....	54
1.4	LC-MS-based analytical platforms in metabolomics	56
1.5	Sample preparation for the determination of endogenous global metabolite levels in bacteria	60
1.6	Data processing and data analysis of LC-MS-based metabolomics experiments	63
2.	Scope of this work.....	66
3.	Results and discussion	68
3.1	Establishment of a targeted and untargeted LC-MS-based metabolite profiling platform for the analysis of metabolite levels in bacteria.....	68
3.1.1	Introduction.....	68
3.1.2	Important aspects of sample preparation for the analysis of endogenous polar metabolites in bacteria.....	69
3.1.3	Establishment of an LC-MS-based analytical platform for the analysis of polar metabolites	72

3.1.4	Development of an in-house database for metabolite identification in complex biological matrices	76
3.1.5	Evaluation of the established metabolite profiling platform in wild-type <i>S. aureus</i> NCTC 8325 as a proof-of-principle.....	79
3.1.6	Summary	85
3.2	Targeted metabolite profiling in wild-type <i>S. aureus</i> USA 300 to elucidate the involvement of the ribokinase ThiD within the pyridoxal salvage pathway.....	87
3.2.1	Introduction	87
3.2.2	Sample preparation protocol and analytical platform for the analysis of PL levels in <i>S. aureus</i> USA 300 extracts	88
3.2.3	Results and discussion	90
3.2.4	Summary	92
3.3	Subclass-Specific Labeling of Protein-Reactive Natural Products with customized nucleophilic probes.....	93
3.3.1	Introduction	93
3.3.2	Probe design and synthesis	95
3.3.3	Probe reaction with prominent electrophilic entities.....	96
3.3.4	Fine tuned naphthalene thiols for maleimide and epoxide capturing	97
3.3.5	Labeling of showdomycin and phosphomycin in metabolic extracts of natural producer strains	100
3.3.6	Summary	102
4.	Experimental Section	103
4.1	Establishment of a targeted and untargeted LC-MS-based metabolite profiling platform for the analysis of metabolite levels in bacteria.....	103
4.1.1	Overnight cultures and cryostock preparation of <i>S. aureus</i> NCTC 8325.....	103
4.1.2	Growth curves of wild-type <i>S. aureus</i> NCTC 8325	103
4.1.3	Sample preparation protocol for the targeted and untargeted analysis of endogenous polar metabolite levels in <i>S. aureus</i> NCTC 8325.....	104
4.1.4	LC-MS-based analytical platform for the detection of endogenous polar metabolite levels in bacteria.....	105
4.1.5	Data analysis	110
4.2	Targeted metabolite profiling in wild-type <i>S. aureus</i> USA 300 to elucidate the involvement of the ribokinase ThiD within the pyridoxal salvage pathway.....	111

4.2.1	Sample preparation for the analysis of extracellular PL levels in <i>S. aureus</i> extracts.....	111
4.2.2	Analytical platform for the LC-MS analysis of PL in <i>S. aureus</i> extracts	111
4.3	Subclass-specific capture of protein-reactive natural products with customized nucleophilic probes	112
4.3.1	General methods.....	112
4.3.2	Evaluation of cap probes with model electrophiles.....	113
4.3.3	Reactivity of Cap6	114
5.	Appendices	116
5.1	Establishment of a targeted and untargeted LC-MS-based metabolite profiling platform for the analysis of metabolite levels in bacteria.....	116
5.2	Subclass-Specific Labeling of Protein-Reactive Natural Products with customized nucleophilic probes	124
5.2.1	Figures	124
5.2.2	Tables	129
III -	Bibliography	131
IV -	Curriculum vitae.....	149

2. Introductory remarks

The present doctoral dissertation was accomplished between February 2012 and October 2015 under the supervision of Prof. Dr. Stephan A. Sieber at the Chair of Organic Chemistry II of the Technische Universität München.

Parts of this thesis have been published as listed below:

Maximilian F. Koch[†], Sabrina Harteis[†], Iris D. Blank, Galina Pestel, Lutz F. Tietze, Christian Ochsenfeld, Sabine Schneider* and Stephan A. Sieber*, “*Structural, biochemical and computational studies reveal the mechanism of selective aldehyde dehydrogenase 1A1 inhibition by cytotoxic duocarmycin analogs*”, *Angew. Int. Ed.*, **2015**, accepted. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

Georg C. Rudolf*, Maximilian F. Koch*, Franziska A. Mandl and Stephan A. Sieber, “*Subclass-specific capture of protein-reactive natural products with customized nucleophilic probes*”, *Chemistry*, **2015**, 21(9), 3701-3707. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

Matthew B. Nodwell, Maximilian F. Koch, Ferdinand Alte, Sabine Schneider and Stephan A. Sieber, “*A sub-family of bacterial ribokinases utilizes a hemithioacetal for pyridoxal phosphate salvage*”, *J. Am. Chem. Soc.*, **2014**, 136(13), 4992–4999.

Publications not highlighted in this thesis:

Wüstner, F., Mejías-Luque, R., Koch, M., Rath, E., Vieth, M., Sieber, S.A., Haller, D., Gerhard, M. “*H. pylori γ -glutamyltranspeptidase impairs T lymphocyte function by compromising metabolic adaptation through inhibition of cMyc and IRF4 expression*”, *Cell. Microbiol.*, **2015**, 17(1), 51–61.

3. Zusammenfassung

Duocarmycine sind Sekretionsprodukte von Bakterien der Gattung *Streptomyces* mit vielversprechenden zytotoxischen, antimikrobiellen und antitumor Eigenschaften, deren Wirkmechanismus in früheren Arbeiten auf die spezifische Alkylierung der DNA zurückgeführt wurde.^{1, 2, 3} Aufbauend auf proteomweiten Studien von *Wirth et al.*, die mithilfe des unnatürlichen Duocarmycin-Derivats seco drug **2** zusätzlich die Aldehyddehydrogenase 1 A1 (ALDH1A1) als Zielenzym von seco drug **2** identifizierten^{4, 5}, wurde in dieser Arbeit der Inhibitionsmechanismus der ALDH1A1 durch Duocarmycinderivate im Detail beleuchtet. Ausgehend von einer Kokristallstruktur wurde das Cys302, welches in direkter Nachbarschaft zu dem aktiven Zentrum Cys303 steht, als Bindestelle bestimmt. Seco drug **2** liegt hierbei passgenau in der einzigartigen hydrophoben Tasche der ALDH1A1 und wechselwirkt mit den drei Interaktionspartnern Phe171, Trp178 und Tyr297. Zur Unterstützung wurden Enzymkinetikparameter von diversen ALDH1A1 Mutanten bestimmt und molekulardynamische Simulationen durchgeführt. Des Weiteren konnte die selektive Inhibition der ALDH1A1 durch seco drug **2** im Vergleich zu den Subfamilienmitgliedern ALDH1A2, ALDH1A3 und ALDH2 bestätigt werden.

Im zweiten Teil der Arbeit wurden Methoden zur Flüssigkeitschromatographie-Massenspektrometrie (LC-MS)-basierten Analyse von Metaboliten implementiert und angewendet. Die sogenannte Metabolomik, ist eine zukunftsweisende Technik, die neben der Genomik, Transkriptomik und Proteomik eine weitere Informationsebene zur Aufklärung zellulärer Stoffwechselwege und Charakterisierung von Proteinfunktionen liefert.⁶ Sie wird zusätzlich vermehrt als Methode für die Identifizierung von Naturstoffen mit potentiellen therapeutischen Eigenschaften eingesetzt.⁷ Zunächst wurde für die weiterführende Charakterisierung von unbekanntem Proteinen ein Protokoll für die globale Analyse von intrazellulären, polaren Metaboliten in *Staphylococcus aureus* (*S. aureus*) NCTC 8325 etabliert. Dieses beinhaltet neben der adäquaten Probenvorbereitung die Implementierung der analytischen Plattform und die Auswertung der generierten umfangreichen Datensätze. Zur Validierung des Protokolls wurde die Reproduzierbarkeit der Ergebnisse untersucht.

Desweiteren wurde ein Protokoll für die gezielte Analyse von Pyridoxal (PL) in extrazellulären Extrakten von *S. aureus* USA 300 etabliert. Ziel der relativen PL-Quantifizierung war es, den Einfluss der Ribokinase SaThiD auf den Pyridoxal-Recycling-Metabolismus zu untersuchen. Basierend auf diesen Metabolomstudien, sowie weiterführenden Sequenzanalysen und aktivitätsbasierten Substratassays wurde SaThiD neu klassifiziert und der Familie der Pyridoxalkinasen zugeordnet.

Zur besseren Identifizierung von niedrig konzentrierten und schwer detektierbaren elektrophilen Naturstoffen wurde im dritten Abschnitt eine LC-MS-basierte Strategie entwickelt, welche auf die kovalente Modifikation der Naturstoffe durch den Einsatz von nukleophilen capture (**Cap**) Sonden basiert. Hierbei wurde eine Bibliothek von **Cap**-Sonden synthetisiert, die jeweils aus einem Naphthalin-Grundgerüst aufgebaut sind, an welches zur Feinabstimmung der Reaktivität, Akzeptor- oder Donor-Gruppen angebracht wurden. Als potenteste Sonde wurde **Cap6** identifiziert, die aus einem Thiol und einer Methoxygruppe aufgebaut ist und vielversprechende Selektivität gegenüber Epoxiden als auch Michaelakzeptoren aufwies. Im Folgenden wurde die Reaktivität von **Cap6** in artifiziellen niedermolekularen Gemischen als auch in komplexen *Streptomyces* Extrakten getestet. Durch die Einführung der **Cap6** Sonde wurde für die beiden Naturstoffe Showdomycin (**Sh**) und Phosphomycin (**Pp**) und damit die resultierenden Additionsprodukte **Cap6-Sh** und **Cap6-Pp** eine deutlich verbesserte Retention bei der Umkehrphasenchromatographie und eine signifikante Steigerung der MS Sensitivität erzielt. Des Weiteren wurde durch die spezifische Absorption im aromatischen Bereich, die Komplexität der Probe reduziert und somit die Identifizierung der **Cap6**-Additionsprodukte signifikant vereinfacht.

4. List of abbreviations, acronyms and symbols

°C	Degree Celsius
Å	Angström
ABPP	Activity-based protein profiling
ADP	Adenosine diphosphate
ADP	Adenosine diphosphate
Ala	L-Alanine
ALDH1A1	Aldehyde dehydrogenase 1A1
AMP	Adenosine monophosphate
AMPc	Cyclic adenosine monophosphate
ATP	Adenosine triphosphate
Av	Acivicin
BHB	Brain-Heart-broth
C18	Octadecyl carbon chain bonded silica
Cap	Capture
CDM	Chemical defined media
CID	Collision induced dissociation
CMP	Cytidine monophosphate
CoA	Cofactor A
CTP	Cytidine triphosphate
Cys	L-Cysteine
dAMP	Deoxyadenosine monophosphate
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
DMSO	Dimethylsulfoxid
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
dsDNA	Double stranded DNA
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphat
e.g.	Exempli gratia = for example
EDG	Electron-donating groups
Epb	4-(4-Ethyl-1-piperazinyI)-4-oxo-2-butenoic acid
eq.	Equivalents
ESI	Electrospray ionization
X	

EtOH	Ethanol
FAD	Flavin adenine dinucleotide
g	Gram
GDP	Guanosine diphosphate
Gly	Glycine
GMP	Guanosine monophosphate
GMP _c	Cyclic guanosine monophosphate
GTP	Guanosine triphosphate
h	Hour
HCl	Hydrochloric acid
HESI	Heated electro spray ionization
HILIC	Hydrophilic interaction chromatography
HMDB	Human metabolome databse
HMP	4-amino-5-hydroxymethyl-2-methylpyrimidine
HPLC	High-performance liquid chromatography
HTCFA	Human tumor colony forming ability assay
IC ₅₀	Half maximal inhibitory concentration
IMP	Inosine monophosphate
kDA	Kilo dalton
kV	Kilo voltage
L-gln-d5	<i>L</i> -Glutamine-2,3,3,4,4-d5
L-glu-d5	<i>L</i> -Glutamic acid-2,3,3,4,4-d5
LB	Luria-broth
LC	Liquid chromatography
Leu	<i>L</i> -Leucine
LTQ-FT-ICR	Linear trap quadrupole - Fourier Transform - ionen cyclotron resonance
M	Molar
m/z	Mass to charge ration
MALDI	Matrix-assisted laser-desorption ionization
MeCN	Acetonitrile
MeOH	Methanol
mg	Milligram
min	Minute
mL	Milliliter
mM	Millimolar
mm	Milimeter
MS	Mass spectrometry

Abbreviations

MS/MS	Tandem mass spectrometry
MW	Molecular weight
NaCl	Sodium chloride
NAD ⁺	Nicotinamide adenine dinucleotide (ox.)
NADH	Nicotinamide adenine dinucleotide (red.)
NADP ⁺	Nicotinamide adenine dinucleotide phosphate (ox.)
NADPH	Nicotinamide adenine dinucleotide phosphate (red.)
NASM	Nucleophilic active site mimetic
nL	Nanoliter
nM	Nanomolar
nm	Nanometer
nmol	Nanomole
NMR	Nuclear magnetic resonance
OD	Optical density
p	Statistical significance
Pa	1,4-phenylendiacrylic acid
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate Buffered Saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PG	Penicillin G
Phe	<i>L</i> -Phenylalanine
PL	Pyridoxal
PLK	Pyridoxal kinase
PLP	Pyridoxal-5'-phosphate
pM	Picomolar
PM	Pyridoxamine
pmol	Picomole
PN	Pyridoxine
Pp	Phosphomycine
ppm	Parts per million
RP	Reversed phase
rpm	Rotations per minute
RT	Room temperature
Ru	Rugulactone
SDS	Sodium dodecyl sulfate
sec	Second
Sh	Showdomycin

SIM	Selected ion monitoring
Su	Sultone
TCEP	Tris-(2-carboxyethyl)-phosphine
TEV	Tobacco Etch Virus
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
Tris	Tris-(hydroxymethyl)-aminomethan
Trp	<i>L</i> -Tryptophane
Tyr	<i>L</i> -Tyrosine
UDP	Uridine diphosphate
UMP	Uridine monophosphate
UTP	Uridine triphosphate
UV	Ultraviolet
Val	<i>L</i> -Valine
WT	Wild-type
ZIC	Zwitterionic
ZIC-HILIC	Zwitterionic-hydrophilic interaction liquid chromatography
μL	Microliter
μM	Micromolar
ρ	Pearson correlation coefficient
σ	Standard deviation

I - Structural, biochemical and computational studies reveal the mechanism of selective aldehyde dehydrogenase 1A1 inhibition by cytotoxic duocarmycin analogs

1. Introduction

1.1 The aldehyde dehydrogenase superfamily

The human genome encodes 19 aldehyde dehydrogenase (ALDH) isozymes (Figure I-1).⁸ They are one of the biggest enzyme superfamilies and the following nomenclature is used: for naming each gene, the root symbol ALDH is followed by an arabic number representing the family and, if needed, a letter which designates the subfamily. The arabic number, which follows the letter denotes the individual gene within the subfamily. An ALDH protein from one gene family is defined as having < 40% amino-acid identity to that from another family. Two members of the same subfamily exhibit approximately $\geq 60\%$ amino-acid identity.⁹ ALDHs are constitutively expressed in mammalian tissues and the highest levels is found in the liver, the kidney, the uterus and the brain.¹⁰ The ALDH superfamily is part of a complex cellular network such that total cellular ALDH activity does not derive from the activity of only one isoform but rather from the interplay of diverse isoforms.¹¹ Cells are capable of activating the expression of specific isoforms when needed, highlighting the complexity of the biological interplay of the superfamily.

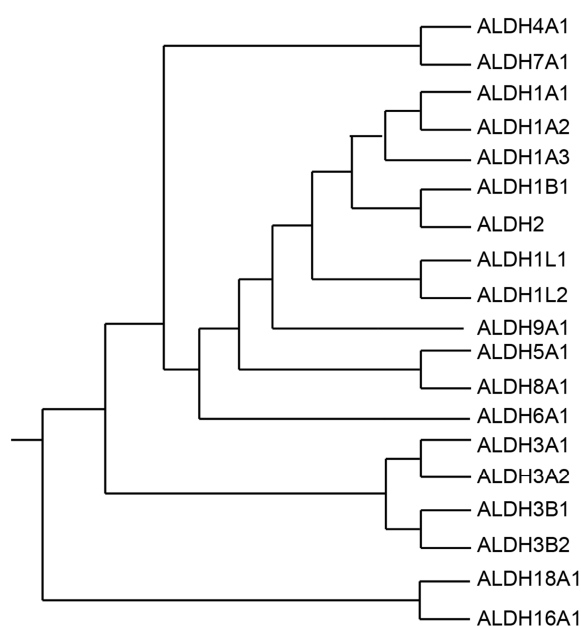


Figure I-1: The 19 ALDH members of the ALDH superfamily of enzymes. Clustered dendrogram, showing the evolution of mammalian ALDHs from a single gene.^{8, 12}

Aldehyde dehydrogenases belong to the group of NAD(P⁺)-dependent oxidoreductases, and catalyze the conversion of endogenous and exogenous aldehydes to their corresponding acids.¹³ In addition to their dehydrogenase activity, ALDHs have some cofactor independent esterase and reductase activity (reducing glyceryl trinitrate (GTN) to glyceryl dinitrate (GDN) and NO₂ for example).^{14, 15, 16} These different functions

demonstrate the importance of ALDHs in physiologically as well as in toxicologically critical roles. In physiology, ALDH1 members play an important role in development and embryogenesis, where they are responsible for the biosynthesis of the important signal molecule retinoic acid (RA).¹⁷ Furthermore, ALDH dependent RA synthesis has been revealed to be crucial for the renewal of stem cells, which is consistent with the increased ALDH activity in these cells in comparison to normal cells.^{18, 19} In addition, ALDHs are relevant in detoxification processes and most prominently in the detoxification of ethanol: ALDH2 catalyzes the chemical transformation from acetaldehyde to acetic acid.²⁰ When ALDH2 activity is decreased less oxidation occurs, leading to an accumulation of toxic acetaldehyde and causing the so-called alcohol-flushing syndrome.²¹ Decreased ALDH2 activity can be a result of diseases or genetic mutations, as for example found in a population of East Asians.²¹ In this population a genetic polymorphism, in which Glu487 is substituted with Lys487 (ALDH2*2), leads to the loss of structural integrity in the coenzyme-binding and active sites, resulting in reduced activity.^{22, 23} Furthermore, ALDHs are involved in cellular protection mechanisms, especially in the detoxification of reactive oxygen species, which are formed during oxidative stress²⁴ as well as during γ -aminobutyric acid²⁵ (GABA) metabolism. ALDHs are also important for the metabolism of exogenous man-made aldehydes, which originate from industry, car exhaust and smoking related gases.²⁶

As a result of their many roles, disorders in the activity of ALDHs can result in dramatic effects. It is not surprising that ALDH disorders have prominent roles in disease, especially in cancer research, as a connection between increased ALDH activity and cancer development has been observed.²⁷ Furthermore, ALDHs are said to have an effect on Parkinson disease, due to the accumulation of neurotoxic substances like 3,4-dihydroxyphenylacetaldehyde.²⁸

Summarizing, ALDHs are pervasive and important key regulators of diverse metabolic reactions. Hence, it is of great interest to increase our knowledge of their roles in cellular metabolism to better understand their involvement in basic cellular biochemistry and to improve the treatment possibilities of ALDH related diseases. In this context it is important to better understand not only their catalytic activity, but also their structural features.

1.2 Aldehyde dehydrogenases and their role in diseases

Recently, ALDHs have been the focus of research due to their involvement in cancer, which is one of the most common causes of death in the developed world.²⁹ Cancer cells occur in a multiplicity of tissue types and are classified by the type of cell that is initially affected.²⁷ Since cancer will be one of the most common diseases in the future, new strategies and a better understanding of the biological process in cancer formation are urgently needed.

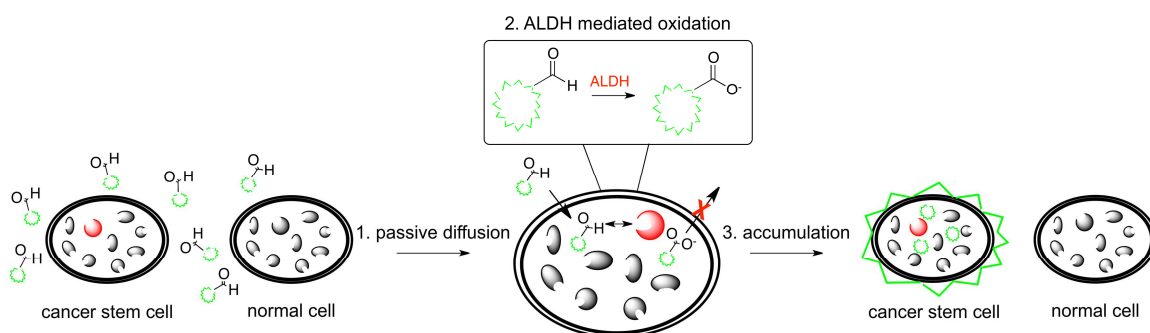


Figure I-2: ALDEFLUOR assay in three steps: **1)** BODIPY®-aminoacetaldehyde is taken up by living cells by passive diffusion. **2)** ALDH mediates conversion of BODIPY®-aminoacetaldehyde to the negatively charged BODIPY®-aminoacetate product, which is retained in the cell cytosol. **3)** Cancer stem cells (CSCs) with high ALDH activity become brightly fluorescent and are therefore called “ALDH bright” in comparison to “ALDH low” cells.³⁰

In recent years a sub-population of cancer cells that plays a potentially critical role in cancer development has been identified. These cells are known as cancer stem cells (CSCs) or tumor initiating cells (TICs) and their presence in most cancer types has been confirmed.^{31, 32} Efforts have been made to investigate and establish strategies to identify and isolate CSCs with the aim of better understanding cancer formation and progression as well as to develop new strategies for cancer treatment. One potential human CSC marker is the membrane antigen CD133, identified in subpopulations of cells in brain, colon and lung tumors.^{33, 34} The expression and activity level of ALDHs has also been proposed as another potential CSC marker.^{35, 36} To investigate the correlation between ALDH activity and CSCs, the ALDEFLUOR assay was developed (Figure I-2).³⁰ Using this method the ALDH activity can be quantified in a variety of cell types e.g. adult tissue cells, primary cancer cells and cultured cells. The assay is based on the ALDH-catalyzed conversion of non-toxic BODIPY aminoacetaldehyde, which can diffuse into intact and viable cells, to the fluorescent reaction product BODIPY aminoacetate. Over time the fluorescent product accumulates in cells with high ALDH activity. These cells are then called ALDH bright cells (ALDH^{br}) and were identified in various cancer tissues including bone marrow, breast, lung, liver, brain and pancreas.³⁷ Proliferation rates, migration and adhesion ability and metastatic potential of ALDH^{br} CSCs were also shown to be higher

than in ALDH low cells.³⁸ These features are related to cancer chemo resistance³⁷ and add to the evidence that ALDH activity can be applied as a universal CSC marker. Additionally, in contrast to what was initially thought, not only ALDH1A1, but also other ALDH isoforms e.g. ALDH2, ALDH3A1 and ALDH9A1, were shown to contribute to ALDH activity in CSCs.^{39, 40}

Subsequently, *Okudela et al.* observed the inverse effect - the observable down regulation of the ALDH1A1 expression - in a considerable number of non-small cell lung carcinomas (NSLC).⁴¹ Although this effect is yet not fully understood, it has been suggested that a loss of ALDH1A1 could promote the progression of lung cancer cells, especially the development of smoking-related adenocarcinomas (ADCs). It is clear that a better understanding of the biological roles of the individual members of ALDH family, in both healthy and cancerous tissues, is urgently needed.

In addition to their involvement in cancer, ALDHs seem to play key roles in other diseases such as intolerance to alcohol and the associated increased risk of ethanol-induced cancer (ALDH1A1 and ALDH2)²⁷, Sjögren-Larsson syndrome (ALDH3A1)^{42, 43}, type II hyperprolinemia (ALDH4A1 and ALDH18A1)^{44, 45}, 4-hydroxybutyric aciduria, mental retardation and seizures (ALDH5A1)²⁷, developmental delay (ALDH6A1)²⁷, hyperammonemia (ALDH18A1)²⁷, pyridoxine-dependent epilepsy (ALDH7A1)⁴⁶, and late-onset Alzheimer's disease (ALDH2)⁴⁷. This frequently and recently discovered involvement of ALDHs in various diseases highlights the importance of increasing our understanding of the biological roles of ALDHs in healthy and disease cell states.

1.3 Structural insight into the aldehyde dehydrogenase superfamily

Within the ALDH superfamily several isoforms share functionalities and catalyze the same reactions, but nevertheless each isoenzyme has its own substrate specificity.⁴⁸ Mammalian ALDHs occur as dimers or tetramers in their active form; each subunit is identical (Figure I-3A and B). These monomeric subunits are built of 500 – 600 amino acids with the exceptions of ALDH16A1, ALDH18A1, ALDH1L1 and ALDH1L2, which are built of 800 – 900 amino acids.^{49, 50}

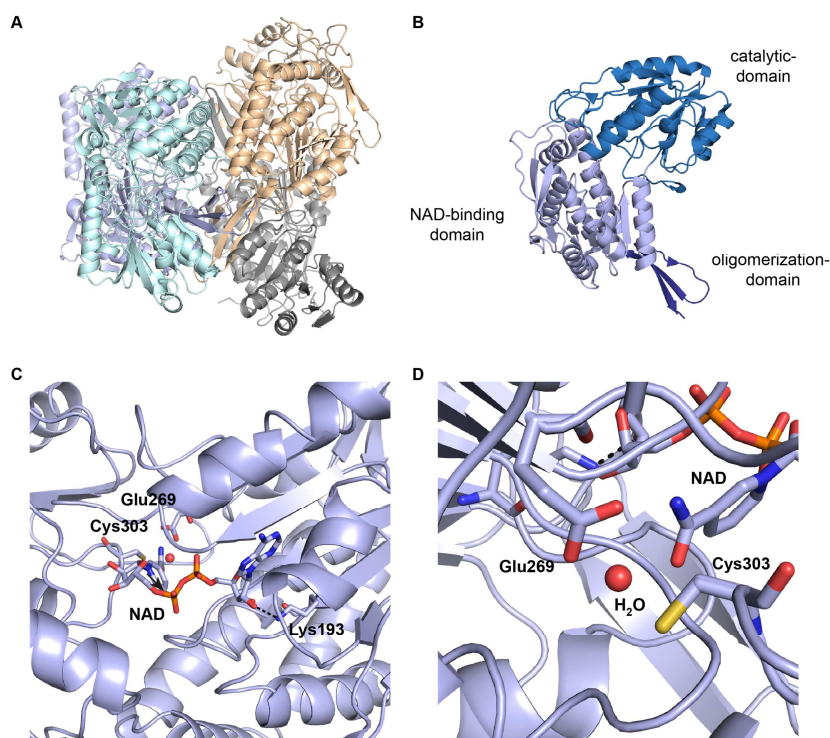


Figure I-3: **A)** Active tetrameric conformation of human ALDH1A1 (PDB code 5ac2). **B)** Monomer subunit of human ALDH1A1 with the three subdomains: catalytic-domain, oligomerization-domain and NAD-binding domain. **C)** Top view of the active site of the human ALDH1A1 with the highly conserved residues: Lys193, Glu269, Cys303 and the cofactor NAD⁺. **D)** Zoomed view into the catalytic domain showing the active side Cys303, Glu269, NAD⁺ and the ordered water molecule in between. Figure generated in Pymol.

Crystallographic studies show that ALDH monomers from all classes share high homology in three distinct functional regions: coenzyme (NAD)-binding domain, catalytic-domain and oligomerization-domain (Figure I-3 B).^{51, 52, 53, 54, 55} In addition to the similarity of the three functional domains, the following catalytic residues are highly conserved: Cys303, Lys193 and Glu269 (Figure I-3C and D).⁸ Further crystallization studies and site directed mutagenesis experiments revealed the importance of these residues in the aldehyde oxidation mechanism.⁵⁶ The residues are highly conserved among the superfamily, with exception of ALDH16A1, which uses CoA as a cofactor. In general the catalytic mechanism of aldehyde oxidation can be divided into five steps and will be discussed

using the example of the human ALDH1A1 (Figure I-4). In the first step the catalytic cysteine (Cys303) is activated via a water molecule mediated proton abstraction with the help of Glu269, which serves as a base. The Cys303 thiolate then attacks the electrophilic aldehyde forming a tetrahedral thiohemiacetale intermediate. The tetrahedral intermediate collapses at the same time as hydride transfer to the pyridine cycle of the cofactor NAD^+ . In the fourth step the thioester intermediate is hydrolysed and finally the reduced cofactor dissociates, so that the enzyme can be regenerated via binding of NAD^+ .

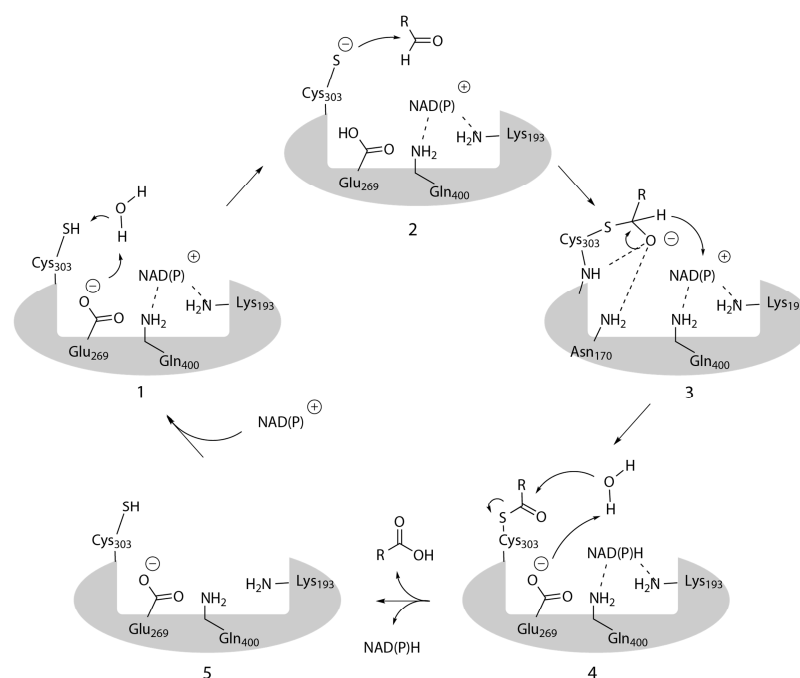


Figure I-4: Five steps of the catalytic aldehyde oxidation mechanism found in human ALDH1A1. Highlighted are the roles of the highly conserved amino acid residues: Lys193, Glu269, Cys303 and Gln400. **1)** Activation of the catalytic active Cys303 via a water molecule mediated proton abstraction with the help of Glu269 which serves as base, **2)** nucleophilic attack by the thiolate of the Cys303 on the electrophilic aldehyde, **3)** collapse of the tetrahedral intermediate at the same time as hydride transfer to the pyridine cycle of the cofactor NAD^+ , **4)** hydrolyses of the thioester intermediate and **5)** dissociation of the reduced cofactor and subsequent regeneration of the enzyme via again binding of NAD^+ .⁵⁶

Due to the high degree of homology in the ALDH superfamily, there is substrate overlap. As an example, ALDH1A1, ALDH1A2 and ALD1A3 are able to oxidize all-*trans*-retinal and 9-*cis*-retinal to their corresponding acids, whereas ALDH1A2 appears to acquire the highest specificity for all-*trans*-retinal.⁸ Despite their similar sequences, ALDH subfamily members differ in the shape of the hydrophobic pocket surrounding the catalytic active site cysteine. This pocket is important for substrate specificity.

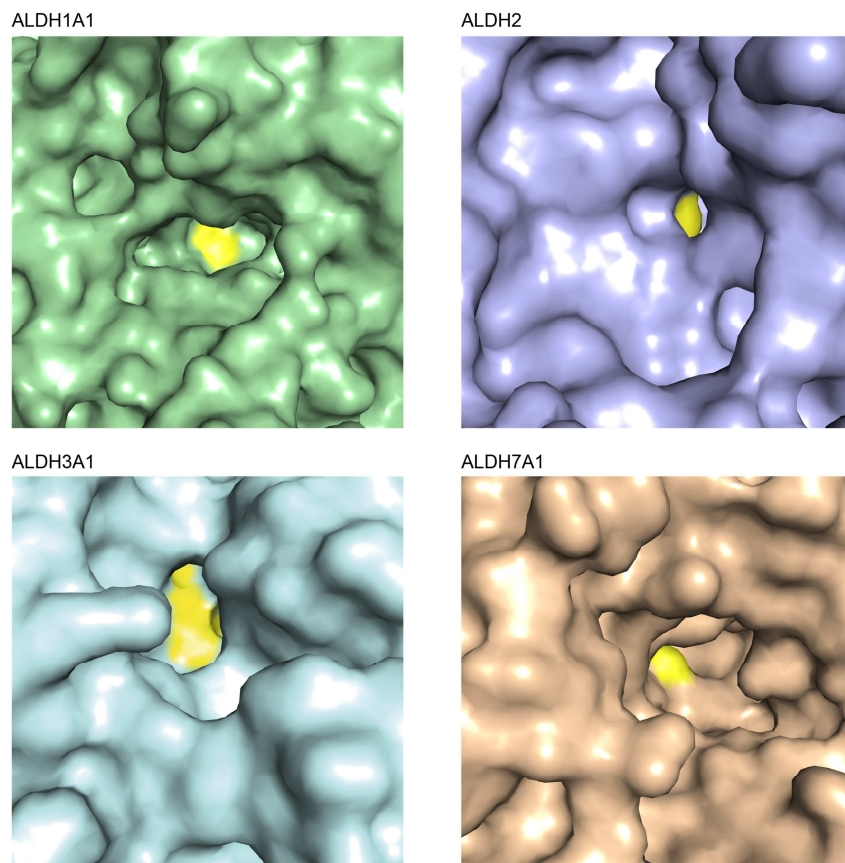
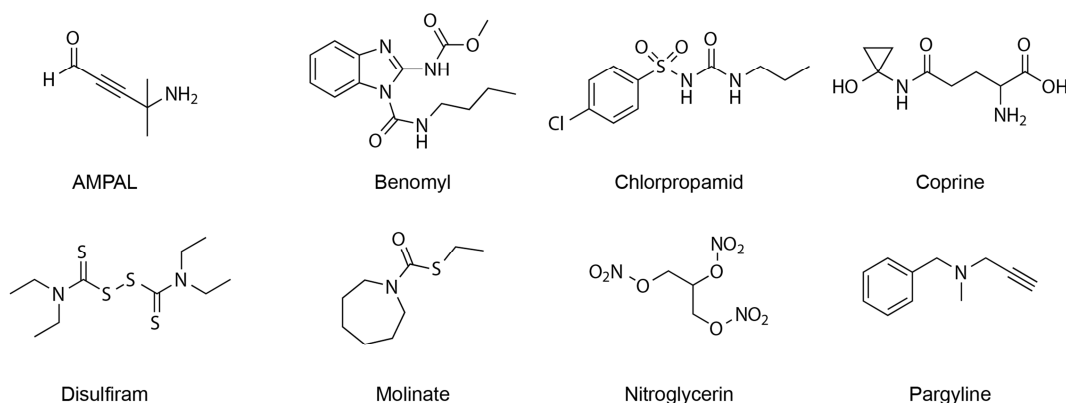


Figure I-5: Insight into the hydrophobic pockets of human ALDH1A1 (PDB code 5ac2), ALDH2 (PDB code 1CW3), ALDH3A1 (PDB code 3SZA) and ALDH7A1 (PDB code 4X0T). Active site cysteines are highlighted in yellow. Figure generated in Pymol.

The different architectures of the hydrophobic pockets are exemplified in ALDH1A1, ALDH2, ALDH3A1 and ALDH7A1 (Figure I-5). The pockets of ALDH2 and ALDH3A1 are smaller in comparison to ALDH1A1 and ALDH7A1. ALDH2 in particular has a highly constricted, cylindrically shaped site. In contrast, ALDH3A1 possesses a wider inner vestibule near the catalytic nucleophile, shown in yellow. The two other ALDH isoforms have larger, hydrophobic pockets and access to the active site cysteine is less restricted. This is in accordance with, their natural substrate specificity: sterically more demanding aldehydes like retinal and (S)-2-amino-6-oxohexanoate are substrates of ALDH1A1 and ALDH7A1; smaller and less sterically demanding aldehydes like acetaldehyde and medium-chain (6 carbons and more) saturated and unsaturated aldehydes, including 4-hydroxy-2-nonenal are substrates for the ALDH2 and ALDH3A1.⁵⁷ This shows that, despite the significant structural conservation, substrate specificity is controlled by the unique shape of the hydrophobic pocket of each isoform. As a consequence, structurally diverse aldehydes can be converted by the superfamily, which is especially beneficial for the detoxification of exogenous aldehydes.^{58, 59}

1.4 Reversible and non-reversible small molecule inhibitors of aldehyde dehydrogenase isoforms

To better understand the roles of the different ALDH isoforms in health and disease, it is useful to be able to selectively inactivate specific enzymes. Therefore genetic knock-outs as well as chemical knock-downs are standard technologies, widely used in protein characterization studies. In genetic deactivation the gene, encoding the specified protein, is deleted or prevented from expressing the protein. In contrast, chemical knock-down does not influence the expression of the protein itself, but a small molecule, which is capable passing the cellular membrane, is used to inhibit the enzymatic activity of the protein. The challenge of a chemical knock-down strategy is the availability of a small molecule to solely address and inhibit the protein of interest, without exhibiting cross reactivity towards other cellular proteins. The discovery of selective inhibitors, especially in the highly homologous ALDH superfamily, is a challenging task. However, applications of such small molecule inhibitors include the investigation of the role of ALDH isoforms in disease cell states. Ultimately, this may enable the development of ALDH inhibitors as drugs.

A

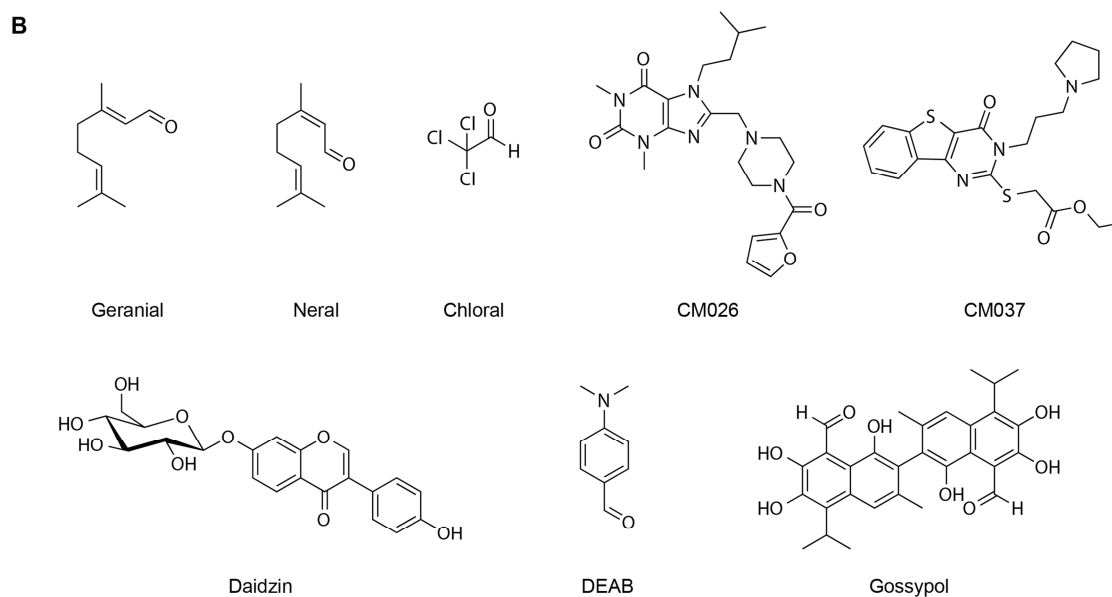


Figure I-6: A) Structures of known irreversible ALDH isoform inhibitors: AMPAL, Benomyl, Chloral, Chlorpropamide, Coprine, Cyanamide, Disulfiram, Molinate, Nitroglycerin, Pargyline. **B)** Structures of known reversible ALDH isoform inhibitors: Citral (Geranial and Neral), Chloral, CM026, CM037, Daidzin, DEAB and Gossypol.

To date many structurally diverse ALDH inhibitors have been discovered (Figure I-6, Table I-1) and can be categorized into reversible and non-reversible binders.⁸ Among them AMPAL, Benomyl, Chloral, Chlorpropamide, Coprine, Disulfiram, Molinate, Nitroglycerin and Pargyline are irreversible, covalently binding inhibitors (Figure I-6 A). They all have in common that they undergo a metabolism-induced activation of the original compound. In the following section the two inhibitors nitroglycerin, also called glyceryl trinitrate (GTN), as well as disulfiram are discussed.

GTN belongs to a class of potent organic nitrates that induce coronary vasodilation by a nitric oxide (NO)-dependent mechanism.⁶⁰ GNT interacts with ALDH2 and is denitrated in different catalytic reaction steps. *Wenzel et al.* and *Lang et al.* clarified the interaction mechanism based on site directed mutagenesis experiments as well as co-crystallization studies (Figure I-7).^{16, 61} The first and central step is the hydrolysis of GTN, induced by nucleophilic attack of the active site Cys303, releasing 1,2-glyceryl dinitrate (1,2-GDN) and resulting in the thionitrate intermediate bound to the ALDH. After this step two different pathways can occur. In pathway A the neighboring cysteine (Cys302) attacks the thiol of Cys303, forming a disulfide bond and releasing a nitrite ion. The nitrite ion is further metabolized and is released as NO, whereas the ALDH2 is oxidized. In pathway B the thionitrite intermediate undergoes a rearrangement reaction resulting in a sulfenyl nitrite species, which leads to the formation of either a sulfenyl radical (C) and NO or to irreversible ALDH inhibition, caused by sulfenyl nitrite rearrangement followed by Glu-269-mediated oxidation (D). As a result, Cys303 is oxidized and inactivated to the sulfinate. of

these complex This example highlights both the variety of reactions, which can be induced by binding of irreversible ALDH inhibitors to ALDHs and the power of combining site directed mutagenesis experiments with crystallography for the elucidation pathways.

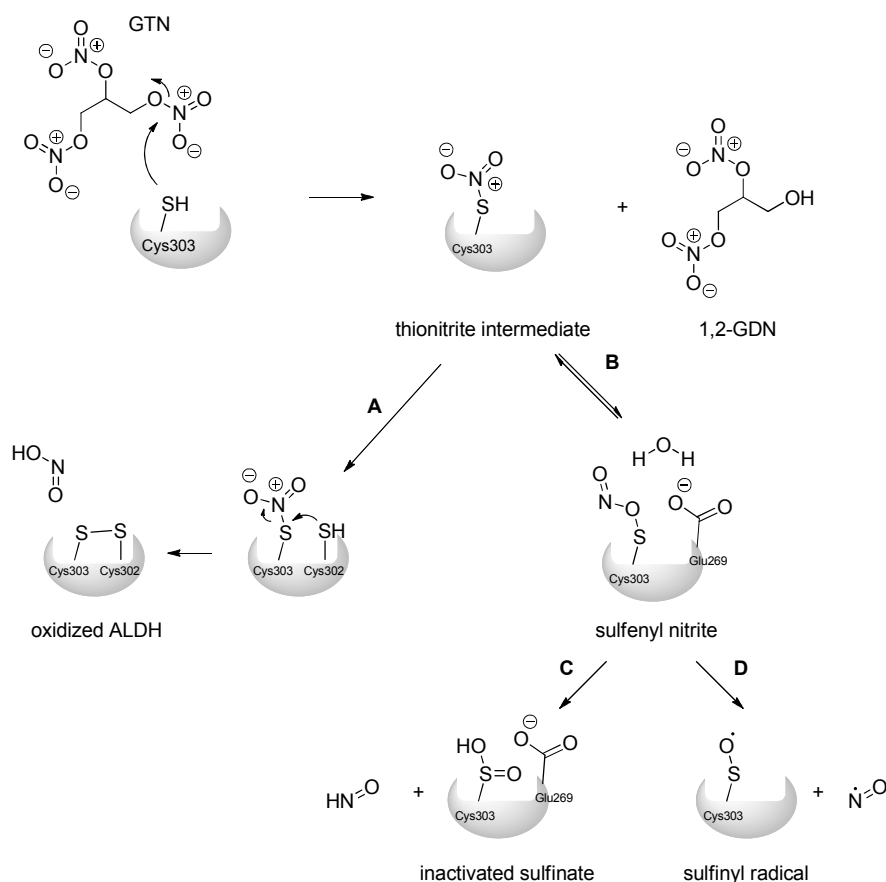


Figure I-7: Nitroglycerin interactions with ALDH. GTN is hydrolyzed by ALDH to form 1,2-glycerlyl dinitrate (1,2-GDN) and a thionitrite intermediate bound to the ALDH active site cysteine (Cys303). Two proposed pathways can then be followed: In pathway **A** the neighboring cysteine (Cys302) attacks the thiol of Cys303, forming a disulfide bond and releasing a nitrite ion. The nitrite ion is further metabolized and is released as NO, whereas the ALDH2 is oxidized. In pathway **B** the thionitrite intermediate undergoes a rearrangement reaction resulting in a sulfenyl nitrite species, which leads to the formation of either a sulfenyl radical (**C**) and NO or to irreversible ALDH inhibition, caused by sulfenyl nitrite rearrangement followed by Glu-269-mediated oxidation (**D**). As a result, Cys303 is oxidized and inactivated to the sulfonate.¹⁶

Since 1949 disulfiram has been applied as an alcohol-aversive agent in the treatment of alcoholism.⁶² In this context the ALDH2 isoform plays a crucial role, as it is important for the detoxification of acetaldehyde, which occurs as a product in ethanol metabolism. Applying disulfiram in combination with alcohol leads to the so-called disulfiram ethanol reaction⁶³ and an accumulation of acetaldehyde with symptoms of low blood pressure, tachycardia, facial flushing, nausea, and vertigo.⁶² The aim of this therapeutic strategy was to discourage the patient, from consuming alcohol. From a mechanistic point of view not disulfiram itself but rather its metabolic products are responsible for the *in vivo* inhibition of ALDH2 (Figure I-8).⁶⁴ Disulfiram is reduced to the diethyldithiocarbamate

(DDTC) and further converted via a hepatic thiol transferase to *S*-methyl-*N,N*-diethyldithiocarbamate (DETC) and *S*-methyl-*N,N*-diethyldithiocarbamate (Me-DDTC). Subsequently P450 enzymes catalyze the oxidation of the metabolites to DETC-sulfoxide (DETC-SO), Me-DDTC-sulfoxide (DDTC-SO) and Me-DDTC-sulfone (Me-DTTC-SO₂). DETC, Me-DDTC, DDTC-SO and DDTC-SO₂ are mitochondrial ALDH inhibitors^{65, 66, 67} and address the catalytic active cysteine by irreversibly modifying it via carbamylation.^{68, 69} In addition to the biologically activated disulfiram derivatives, disulfiram itself is also an inhibitor. Disulfiram is a better inhibitor of ALDH1A1 than of ALDH2 because this bulky compound cannot easily access the smaller entrance tunnel of ALDH2.⁷⁰ So far, the mode of action has not been elucidated but it has been postulated that disulfiram addresses the active site cysteine.

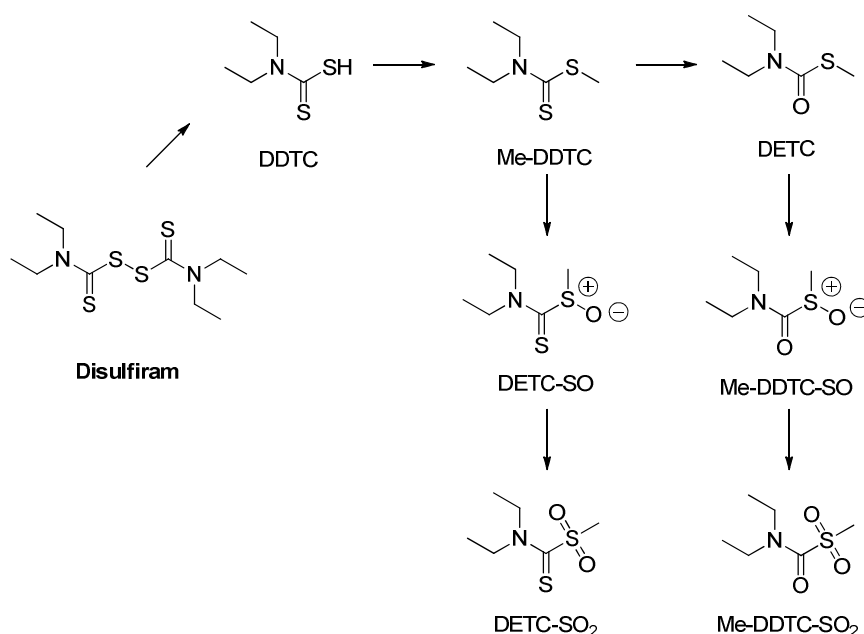


Figure I-8: Disulfiram and its metabolites. DDTC is liberated from disulfiram by disulfide exchange with the catalytic cysteine of ALDH. Hepatic thiomethyl transferases subsequently give rise to all other metabolites.⁶⁴

In addition to GTN and Disulfiram, AMPAL, Benomyl, Chlorpropamid, Coprine, Molinate and Pargyline are also predicted to bind irreversibly.⁸ It is assumed that these compounds address the active site cysteine, with the exception of AMPAL, where no binding site is stated. In depth co-crystallisation studies in combination with mutational studies and MS experiments are needed to investigate and validate the binding mode since the postulated modes of action of these inhibitors often is unclear. With the help of these studies, understanding of the reactivity of the ALDH family can be increased. Based on this understanding, specific, irreversible inhibitors can be designed and tested for their inhibitory efficacy as well as selectivity against ALDH isoforms.

Reversible inhibitors of ALDH isoenzymes include Citral, Chloral, Daidzin, DEAB, Gossypol, CM026 and CM037 (Figure I-6 B). These compounds do not require further activation via cellular metabolism which is why their mode of inhibition is, in comparison to the irreversible inhibitors, more diverse. Gossypol, for example is a rather unselective inhibitor, addressing the NAD-binding domain.^{71, 72} The mode of action is based on inhibiting the annealing of the cofactor NAD⁺, which is essential for the catalytic oxidation mechanism. Testing against ALDH isoforms revealed activity against ALDH1A1, ALDH2 and ALDH3A1.⁷¹ The targeted NAD-binding domain motif, also called the Rossmann fold, is a frequently occurring motif in nucleotide binding proteins, including dehydrogenases.⁷³ It therefore seems unlikely that gossypol is selective for ALDHs.

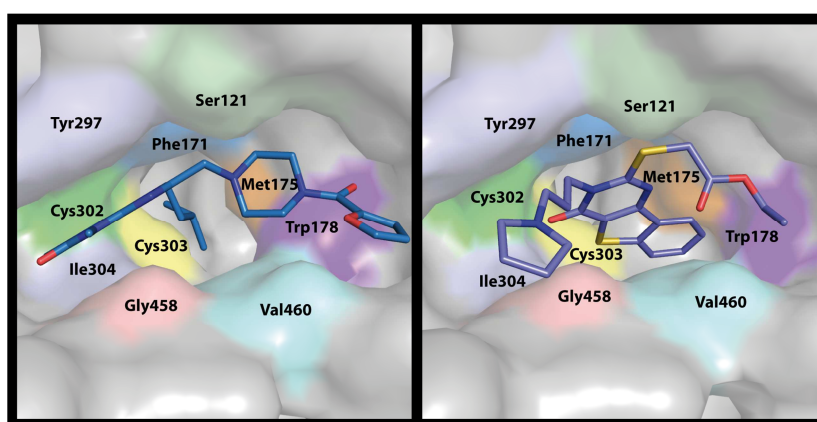


Figure I-9: Human ALDH1A1 inhibitors discovered by Hurley *et al.* CM026 (left, PDB code 4WP7) and CM037 (right, PDB code 4X4L). Figure generated in Pymol.

In contrast to gossypol, Citral, Daidzin, CM026 and CM037 bind to the hydrophobic pockets of ALDHs. Among them the inhibitors reported by Hurley *et al.*, CM026 and CM037, identified in a high-throughput screen, show particularly promising selectivity for ALDH1A1.⁷⁴ Both inhibitors have planar, multiringed structures and co-crystallization studies have revealed their localization in the hydrophobic pocket, exploiting a unique cavity in ALDH1A1, which it is not accessible in other ALDH isoenzymes (Figure I-9). The two inhibitors are interacting with and stabilized by the amino acid residue Gly458. In consequence, the entrance tunnel is blocked, resulting in limited access of the substrate to the active site cysteine explaining the selectivity by amino acid residue Gly458. Among the ALDH members Gly458 is unique for ALDH1A1 and ALDH16A1. The function of ALDH16A1 is still unknown and due to the deletion of the conserved Cys nucleophile, it has been suggested that it may have functions independent of aldehyde dehydrogenase activity.⁷⁵ The selectivity hypothesis for ALDH1A1 could be further validated in activity-based assays among eight ALDH isoform members (ALDH1A2, ALDH1A3, ALDH1B1, ALDH2, ALDH3A1, ALDH4A1 and ALDH5A1), where CM026 only inhibited ALDH1A1.

CM037 also completely inhibits ALDH1A1, with an additional slight inhibition of 20% against ALDH3A1 isoenzyme.⁷⁴

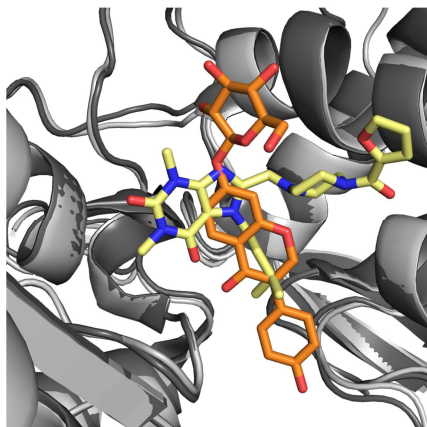


Figure I-10: Comparison of the binding of CM026 to ALDH1A1 (yellow, PDB code 4WP7) with the binding of daidzin to ALDH2 (orange, PDB code 2VLE). Figure generated in Pymol.

Daidzin is a phytochemical derived from the chinese edible vine *Pueraria lobata*. It was identified as a potent inhibitor of ALDH2 with an IC_{50} of 80 nM and some activity against the isoenzyme ALDH1A1.¹⁵ The binding mode of daidzin is similar to that of CM026, blocking the substrate binding site in the hydrophobic pocket. An overlay of the crystal structures (Figure I-10) shows both compounds bound in the same area, but perpendicular to each other. As daidzin fits in the pocket of ALDH2 as well, inhibition for both isoenzymes can be explained.

Another ALDH1 inhibitor, DEAB (4-diethylaminobenzaldehyde), has been applied in the ALDEFLUOR assay mentioned previously, in which it serves as an inhibitor and thereby as a control for background fluorescence. The inhibition mode of DEAB against ALDH1 isoenzymes is predicted to be reversible but has not been completely ascertained.⁸ In contrast to the hypothesis of a reversible mode of action, a recent publication of *Luo et al.* identified for ALDH7A1 an irreversible binding mode of DEAB, as co-crystallization studies revealed the active site Cys302 to be covalently modified by the attachment of DEAB in a thioacyl linkage.⁷⁶ This result contradicts the previously stated reversible mode of action, which needs to be again reviewed and investigated to clarify the mode of action of DEAB against ALDH1 isoenzymes.

The final two reversible inhibitors focused on in this chapter are Citral and Chloral. Citral belongs to a class of volatile α,β -unsaturated aldehydes that occur naturally in herbs as a 2:1 mixture of *cis* and *trans* isomers, termed neral and geranial. It was shown that the mixture shows slow, noncompetitive inhibition of ALDH1, ALDH2 and ALDH3 isozymes.⁷⁷⁷⁸ The predicted binding mode has not yet been validated by crystallization studies. Chloral, originally identified as a hypnotic agent, was identified as a substrate of ALDHs.⁷⁹

It is predicted to reversibly bind to the active site Cys303 forming a thiohemiacetal.⁸ Similar to Citral, the binding mode has not yet been confirmed by in depth co-crystallization studies and therefore remains debatable.

Table I-1: Pharmacologically interesting inhibitors that address ALDH isoforms. Inhibitor names, the type of inhibition, the possibility of metabolic activation of the inhibitor, the postulated binding site and the targeted isoforms are given.⁸

Inhibitors	Type of inhibition	Metabolically activated	Binding site	Targeted ALDH isoform
AMPAL	Non-reversible	Yes	Not known	ALDH1/ALDH3
Benomyl	Non-reversible	Yes	Active site	ALDH2
Chlorpropamid	Non-reversible	Yes	Active site	ALDH2
Coprine	Non-reversible	Yes	Active site	ALDH2
Disulfiram	Non-reversible	Yes	Active site	ALDH1A1/ALDH2
Molinat	Non-reversible	Yes	Active site	ALDH2
Nitroglycerin	Non-reversible	Yes	Active site	ALDH2
Pargyline	Non-reversible	Yes	Active site	ALDH1A1/ALDH2
Citral	Reversible	No	Not known	ALDH1-3
Chloral	Reversible	No	Not known	ALDH
CM026/CM037	Reversible	No	Hydrophobic pocket	ALDH1A1
Daidzin	Reversible	No	Hydrophobic pocket	ALDH2/ALDH1A1
DEAB	Reversible/Non-reversible	No	Active site	ALDH1/ALDH7A1
Gossypol	Reversible	No	NAD domain	ALDH

In summary, there is an urgent need for selective inhibitors of the individual members of the ALDH superfamily. To date reversible as well as irreversible small molecules inhibitors have been identified. Irreversible inhibitors bind in a covalent mode, usually addressing preferentially the active site cysteine and thereby directly blocking the catalytic oxidation reaction of the ALDHs. Reversible inhibitors have more diverse mode of actions. These compounds mainly target the NAD binding domain, or the hydrophobic pocket. In both binding modes the catalytic conversion is interrupted either through inhibition of cofactor binding or by limiting accessibility of the substrates to the active site by blocking the hydrophobic pocket and the entrance tunnel.

In addition to elucidate the binding mode of ALDH inhibitors, the question of their selectivity within the ALDH superfamily needs to be further investigated. Therefore *in vitro* screenings initiated from *Hurley et al.* need to be further expanded among the ALDH superfamily.⁸⁰ Additionally, the *in vivo* selectivity of most of the inhibitors still remains

unclear, especially in the context of the total complex cellular proteome. However, there is scope for many interesting studies such as the elucidation of the role of ALDH isoforms in CSCs as well as in several other diseases like Parkinsons or Alzheimers.⁸¹ To answer the question of *in vivo* selectivity, proteome wide studies need to be carried out to estimate the selectivity of existing reversible as well as irreversible ALDH inhibitors.

In 2012 *Wirth et al.* showed that a duocarmycin derivative had cytotoxic activity in A549 lung cancer cells.⁴ Duocarmycins are natural products isolated from *Streptomyces* strains. Their mode of action was originally thought to be based on irreversible alkylation of the minor groove of DNA.^{82, 83, 84, 85, 86} Surprisingly, *Wirth et al.* identified ALDH1A1 as a selective, covalent target of the duocarmycin derivative and furthermore validated the results by activity-based kinetic studies as well as *in vivo* imaging experiments.⁴ Based on these findings, the aim of the present study was to reveal the binding site of this new ALDH1A1 inhibitor class, thus explaining the compound's high selectivity for ALDH1A1 in the context of the whole proteome. This research is described in the next section.

2. Scope of this work

This chapter discusses work to identify the binding site and illuminate the high selectivity of the duocarmycin derivative seco drug **2** for human ALDH1A1. Previous studies from *Wirth et al.* identified ALDH1A1 as the single covalent target of seco drug **2** in the complex proteome of A549 lung cancer cells.^{4, 5} First step the binding site of the drug was identified by co-crystallization studies of human and sheep ALDH1A1 (sequence homology more than 90%). The identified binding site was then validated by site directed mutagenesis, activity-based kinetic studies, LC-MS bottom-up proteomic experiments as well as molecular dynamic simulations.

Following on from the crystal structure, the next step was to identify the amino acid residues that interact with seco drug **2** inside the hydrophobic pocket. Site directed mutagenesis and determination of the kinetic parameters K_i and k_{inact} were used. Kinetic parameters were determined by activity-based ALDH assays according to Kitz and Wilson.^{87, 88, 89} Finally, the observed selectivity of seco drug **2** for ALDH1A1 was evaluated by comparing the *in vitro* inhibitory efficacy of seco drug **2** for the subfamily members: ALDH1A2, ALDH1A3 and ALDH2.

Author contributions

S.A. Sieber conceived and supervised the project. M.F. Koch performed all experiments unless noted otherwise, including DNA cloning, protein purification, crystallization experiments, kinetic assays and mass spectrometry experiments. S. Harteis and S. Schneider solved the crystal structures. MD simulations were performed by I. Blank.

3. Results and Discussion

Since their discovery in the 1970s duocarmycin natural products such as (+)-CC-1065 and Duocarmycin SA have attracted great attention due to their promising anti-cancer activities (Figure I-11 A).^{1, 2, 3} The duocarmycin mode of action is based on its characteristic curved indole structure and a spirocyclopropylcyclohexadienone electrophile.⁹⁰ This alkylating moiety is in conjugation with a vinylogous amide that tames its intrinsic reactivity. Thus, in solution duocarmycins are remarkably unreactive.⁹¹ Shape selective recognition by the narrow, AT-rich minor groove of DNA induces a conformational change resulting in activation of the cyclopropyl moiety for nucleophilic attack by the 3' adenine N3-nitrogen (Figure I-11C).^{90, 92} Because of this unique target-based activation mechanism duocarmycins were believed to solely address DNA; however, this notion was challenged when activity-based protein profiling (ABPP) with a duocarmycin derived prodrug probe (seco drug **1**, Figure I-11A) in A459 lung cancer cells revealed aldehyde dehydrogenase 1A1 as an additional target.⁴ Moreover, cytotoxic derivatives lacking the DNA binding unit were discovered^{4, 5, 93} that did not exhibit pronounced *in situ* and *in vitro* DNA interactions. The affinity to ALDH1A1 was increased e.g. with seco drug **2** (Figure I-11A) and imaging studies confirmed a cytosolic localization without pronounced nucleus staining.⁵ Several independent RNAi studies showed an important role of ALDH1A1 in cancer cell proliferation.^{5, 94} However, the exact function of ALDH1A1 in human biology is still not fully understood and tools to specifically manipulate its activity in the presence of related ALDH isoforms are needed.⁶³ Recent progress on the development of isoform specific inhibitors *in vitro* has revealed some promising candidates. However, their selectivity in the context of a whole proteome has not yet been investigated.^{74, 95, 96} Up to now the structural prerequisites of ALDH1A1 for seco drug **2** binding were unknown and a specific interaction between a protein and a molecule known to bind DNA questioned.^{97, 98}

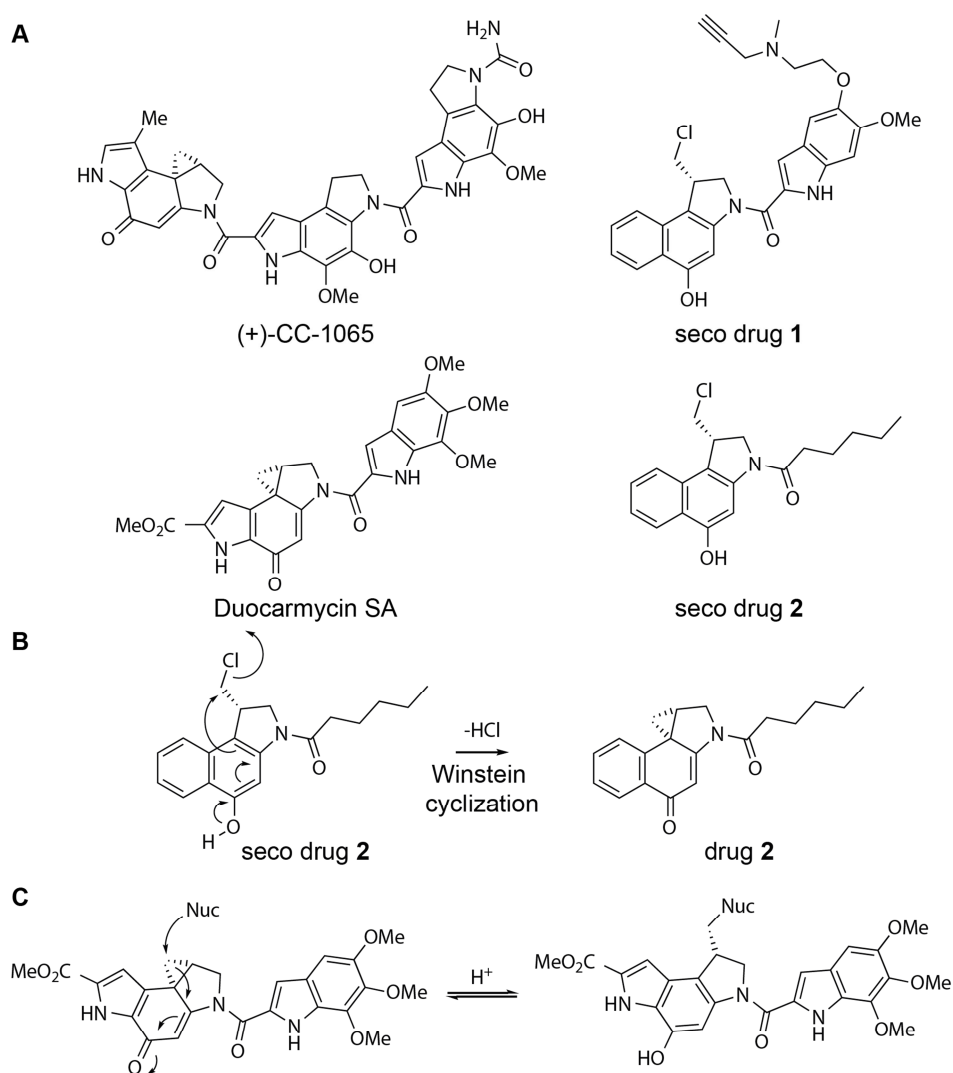


Figure I-11: A) Structures of (+)-CC-1065, Duocarmycin SA, seco drug 1 and 2. **B)** Conversion of seco drug 2 into the active molecule via Winstein cyclization. **C)** Nucleophilic attack at the cyclopropyl ring of Duocarmycin SA by DNA or protein (Nuc = Adenine N3 in DNA or Cys in ALDH1A1).

Since duocarmycin analogs are currently being explored as drug-antibody conjugates (ADC) a precise characterization of all possible target interactions is desirable.⁹⁹ Here we show by high-resolution X-ray co-crystal structures of human and sheep ALDH1A1 a perfectly shaped pocket in the active site that precisely interacts with seco drug 2. Non-catalytic Cys302 covalently traps the compound and several aromatic amino acids engage in crucial interactions via van der Waals forces and π -stacking. The functional role of these residues was analyzed via mutational, kinetic and computational studies.

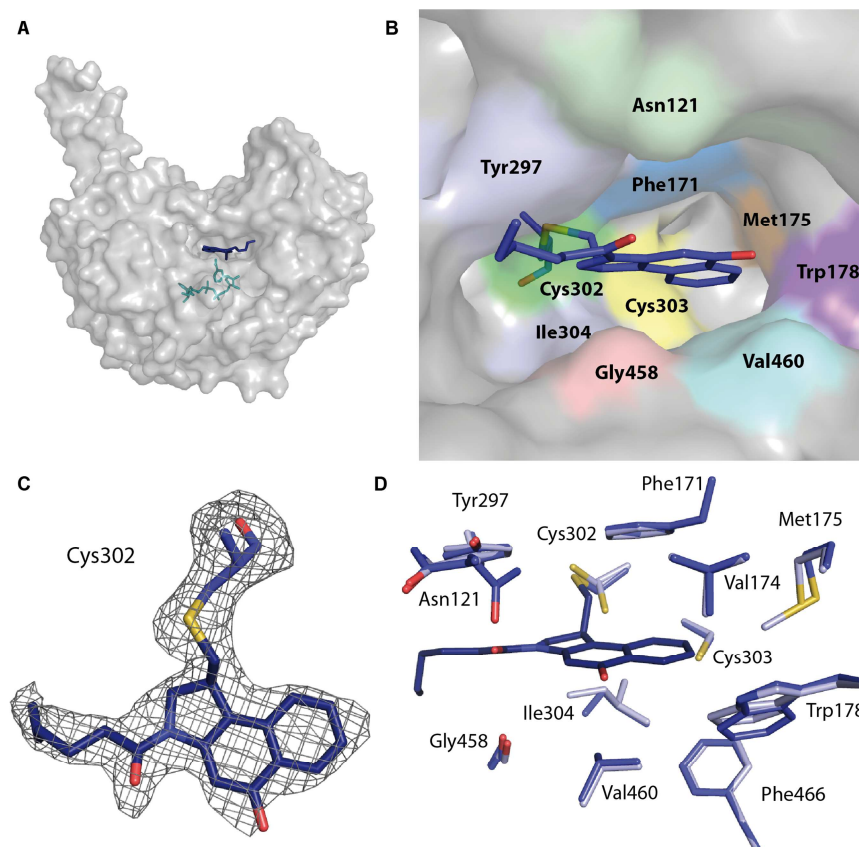


Figure I-12: X-ray crystal structure of seco drug **2** bound to human ALDH1A1 (PDB code 5ac2) **A)** Overall structure of the complex showing ALDH1A1 as surface and the seco drug and the NAD⁺ cofactor as blue and green stick models, respectively. **B)** Close-up view of the hydrophobic catalytic pocket with seco drug **2** bound to Cys302. **C)** Simulated-annealing $F_o - F_c$ difference omit electron density map of seco drug **2** and Cys302 contoured at 2.5σ . **D)** Structural superposition of the active sites of apo ALDH1A1 (light blue, PDB code 4WJ9) and in complex with seco drug **2** (dark blue).

Potency and DNA binding ability of seco drug **2** were evaluated prior to crystallization via MTT cell toxicity and MS-based DNA interaction assays. Seco drug **2** effectively killed A549 cells with an IC_{50} value of 28 nM (Figure I-A1). In contrast to seco drug **1** and duocarmycin SA, which both covalently modified AT-rich double stranded DNA, seco drug **2** showed no such modification suggesting a DNA-independent mode of action (Figure I-A2).¹⁰⁰ As ALDH1A1 is a confirmed protein target with an important role for cell proliferation we investigated this binding in greater detail. Recombinant ALDH1A1s from sheep¹⁰¹ and human⁷⁴ were co-crystallized with seco drug **2** (after Winstein cyclization, Figure I-11B) in different crystal forms and the X-ray structures were determined with 1.8 to 2.1 Å resolution, respectively (Table I-A1). The electron density maps of the compounds reveal binding into the preformed hydrophobic substrate binding pocket next to the NAD⁺ cofactor without the introduction of conformational changes (Figure I-12A-D, Figure I-A3). Apo and holo enzymes superimpose with an RMSD of 0.4 Å in both sheep and human structures (Figure I-A4). The flexible alkyl chain of seco drug **2** points towards

the entrance of the active site pocket and the extent of defined electron density varies in the different structures (Figure I-A5). This is in agreement with inhibition results of indole bearing duocarmycin analogs (seco drug 1) that need extra space in this region to accommodate their bulkier substituent.^{4, 5} In fact, modeling shows that duocarmycin SA fits into the channel without clashes (Figure I-A6). Surprisingly, in both crystal structures the electrophilic cyclopropyl moiety of seco drug 2 did not alkylate the nucleophilic active site Cys303 (numbering according to the human structure⁷⁴) but solely neighboring non-catalytic Cys302.¹⁰² The covalent bond positions the adjacent tricyclic aromatic ring system in a tight-fitting hydrophobic binding pocket that stabilizes the ligand by interactions involving Phe171, Trp178 and Tyr297 (Figure I-12B). In line with the high sequence identity between the human and sheep enzymes (92 % sequence identity, 97% sequence similarity, Figure I-A7) both structures can be superimposed with an RMSD of 0.4 Å.

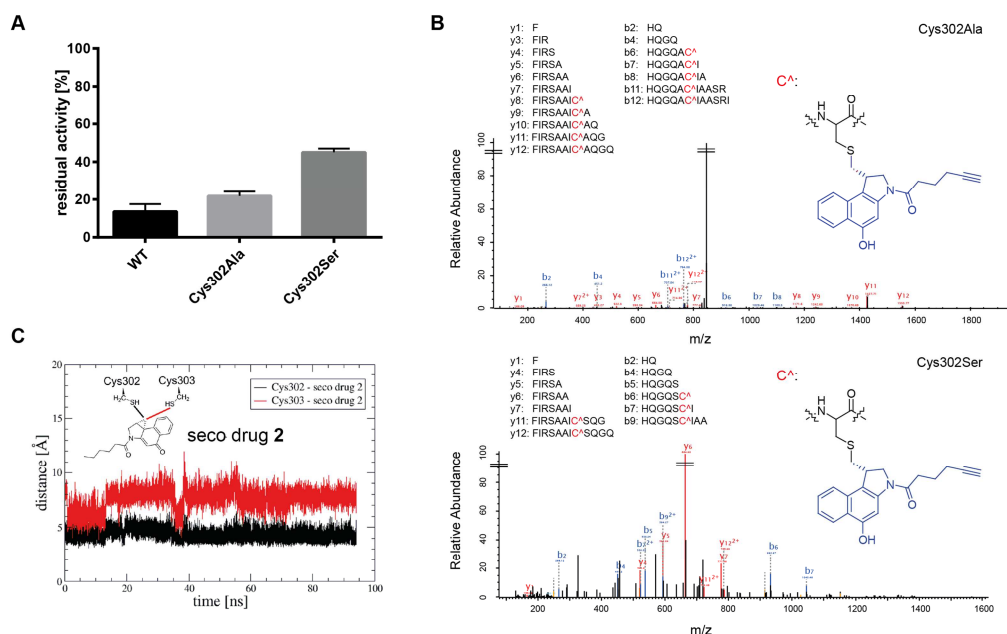


Figure I-13: A) Propionaldehyde-based activity assay with wt and the binding site mutants Cys302Ala and Cys302Ser. Residual activity was measured relative to the uninhibited enzyme activity (shown in %) after 5h of incubation with 50-fold excess of seco drug 2 (Mean +/- std dev, n = 3). **B)** LC-MS/MS binding site identification in human ALDH1A1 for Cys302Ser and Cys302Ala mutants with a probe version of seco-drug 2. Displayed are fragmented peptides (y and b ions) identified by MS/MS sequencing. The modified cysteine residues are indicated by: C^A (highlighted in red). **C)** Distance of the sulfur atoms of Cys302 and Cys303 to the cyclopropyl moiety of seco drug 2 from MD calculations. Cys302 is always closer to the cyclopropyl moiety than Cys303, which supports the selectivity of seco drug 2 for Cys302.

Based on this binding mode we next analyzed the impact of exposed amino acids on interaction with the ligand. Due to the high overall structural similarity we performed all biochemical studies with the human enzyme and independently confirmed the results by MD calculations¹⁰³ with the sheep enzyme. We first evaluated the extent of covalent

alkylation via intact protein mass spectrometry (MS) and gel-based fluorescent labeling with an alkyne probe version of seco drug **2**.⁵ Addition of seco drug **2** or probe (10-fold excess) at pH 7.4 resulted in ALDH1A1 alkylation after >2 h incubation (Figure I-A8) demonstrating a more efficient binding compared to duocarmycin SA (Figure I-A6). MS/MS sequencing with seco drug **2** probe confirmed the crystallographic results with only Cys302 being compound modified (Figure I-A9). Slightly higher pH values (8.0-8.5) were associated with the detection of additional modification sites including Cys456 and Cys464 as we reported earlier.⁴ This assignment can be explained by the elevated nucleophilicity of cysteine thiols at the higher pH which may reflect non-physiological conditions. Previous mutational studies with ALDH1A1 had highlighted only Cys303 as essential for catalysis and clearly showed that neighboring Cys302 is dispensable.¹⁰²

It is thus likely that binding of seco drug **2** to Cys302 obstructs the active site and thereby prevents substrate access and turnover. Unexpectedly, ALDH1A1 Cys302Ala or Cys302Ser mutants were still inhibited by seco drug **2** (Figure I-13A). MS/MS sequencing of both compound treated mutant enzymes revealed exclusive alkylation of the neighboring active site Cys303 suggesting that once the preferred nucleophile is eliminated this proximal residue can substitute in alkylation (Figure I-13B).

To obtain a more quantitative measure of the binding kinetics we determined K_i and K_{inact} values according to the method of Kitz and Wilson for covalent inhibitors.¹⁰⁴ The K_i values for the mutants were up to 10 times higher as compared to the wild-type, suggesting reduced affinity especially in case of the Ser substitution (Figure I-A10). Binding preferences of seco drug **2** to either Cys302 or Cys303 were independently validated by computational studies via molecular dynamics (MD) simulations confirming Cys302 as the favored binding site due to a stable interaction with the cyclopropyl moiety (Figure I-13C). However, Cys303 is within reach of the cyclopropyl moiety and could thus easily bind in the mutants by slight movement of Trp178, as observed in a previous drug-protein co-crystal structure by Hurley *et al* (PDB code 4X4L) (Figure I-A11).⁷⁴ In line with experimental data, the calculations further revealed that the Cys302Ala mutant provides better accessibility of Cys303 compared to the bulkier Cys302Ser mutant (Figure I-A12).

With a mechanistic appreciation of the alkylation reaction we next focused on the hydrophobic pocket and its interactions with the ligand. The structures revealed several aromatic amino acid residues that stabilize the molecule by van der Waals forces and π -stacking (Figure I-12B). To analyze their role in seco drug **2** binding, Phe171, Trp178 and Tyr297 were individually exchanged for Ala and the corresponding mutant proteins tested for compound inhibition. While the k_{inact} values were comparable to wild-type and thus indicative of similar covalent enzyme inactivation rates, the K_i values were significantly

elevated (up to 17-fold) in the mutant proteins suggesting a reduction of stabilizing interactions (Figure I-14A). MD calculations confirmed this hypothesis and revealed π -stacking and van der Waals interactions with Trp178 as well as van der Waals interactions with Phe171 and Tyr297 (Figure I-14B and I-A13).^{105, 106} Interestingly, Tyr297 shows strong and stable van der Waals interactions with the whole aliphatic chain of seco drug **2** (Figure I-14B and I-A14) during all MD simulations. Compared to Trp178 and Phe171 there are more atoms of Tyr297 within reach of seco drug **2**, highlighting this residue as the most important interactor (Figure I-14B). Additional van der Waals interactions with Leu174, Ile304 and Val460 were observed (Figure I-A13).

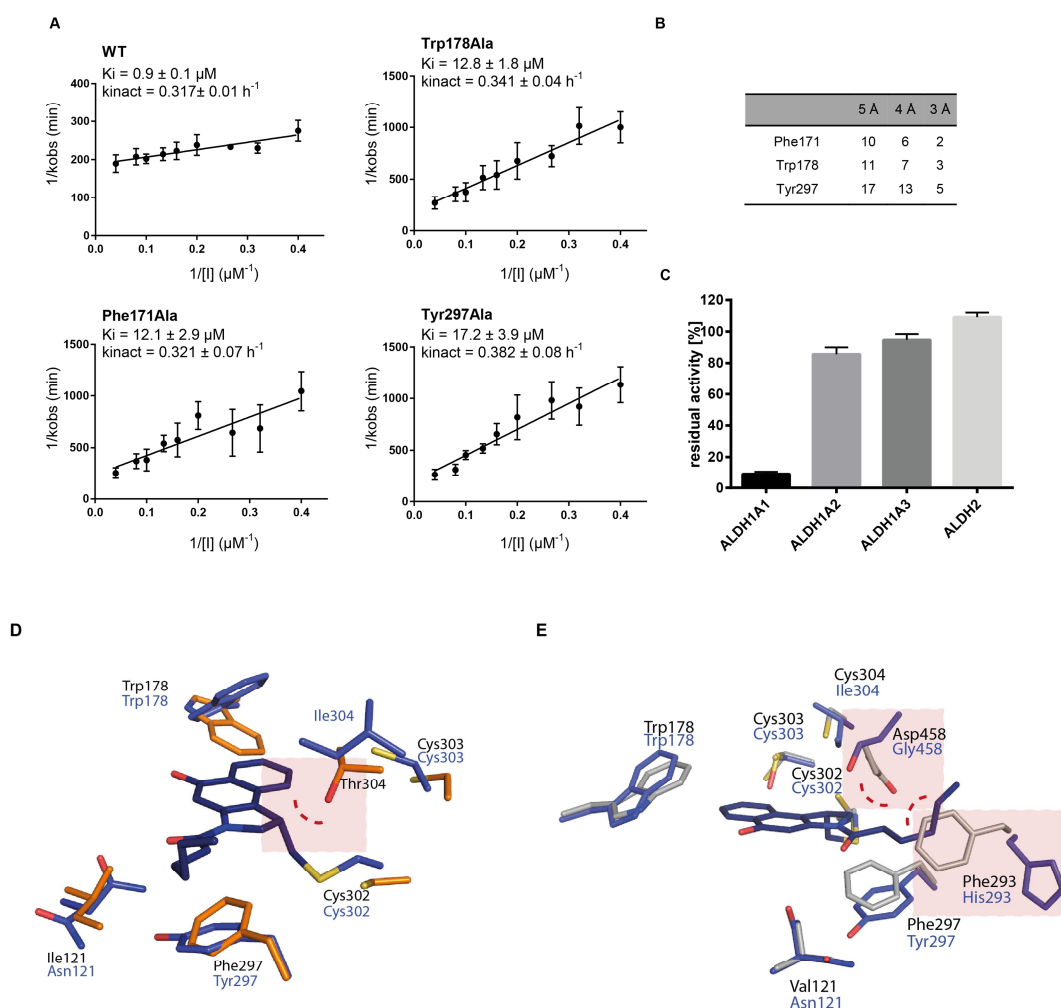


Figure I-14: **A)** Determination of K_i and k_{inact} according to the method of Kitz and Wilson for the wild-type ALDH1A1 and the mutants where the interacting residues of seco drug **2** were mutated to alanine: Phe171Ala, Trp178Ala and Tyr297Ala. **B)** Average number of Phe171, Trp178 and Tyr297 atoms interacting with seco drug **2** at the indicated distances (MD calculations). **C)** Propionaldehyde-based activity assay with related human isoforms ALDH1A2, ALDH1A3 and ALDH2. Residual activity was measured after 5h of incubation with 50 fold excess of seco drug **2**. **D)** Active site of human ALDH1A1 with seco drug **2** (blue, PDB code 5ac2) and the apo structure of ALDH1A2 (orange, PDB code 1B19). **E)** Active site of human ALDH1A1 with seco drug **2** (blue, PDB code 5ac2) and the apo structure of ALDH2 (grey, PDB code 3N80).

Thus the unique fit into this π -stacking network as well as the shape, accessibility and polarity of the corresponding binding pocket provide an explanation for the previously observed proteome selectivity for ALDH1A1^{4, 5}. Interestingly, even closely related ALDH isoforms, ALDH1A2, ALDH1A3 and ALDH2 (all about 70 % sequence identity), were not inhibited or only slightly inhibited (ALDH1A2) by seco drug **2**, likely due to steric clashes - for instance by the substitution of Ile304 (ALDH1A1) with Thr (ALDH1A2/3) or Gly458/His293 (ALDH1A1) by Asp458/Phe293 (ALDH2) (Figure I-14C, D and E). Duocarmycin analogs are thus important tools to analyze the role of ALDH1A1 in cancer cells and stem cell development. In fact such tools are urgently needed since commercially available ALDH1A1 inhibitors like diethylaminobenzaldehyde and disulfiram are known to be unselective and also bind ALDH2.¹⁰⁷ Recent efforts in broad inhibitor screening combined with rational design have provided attractive starting points for more isoform-specific inhibitors. Among those are reversible binders which have been shown by structural studies to address the hydrophobic pocket via a different binding mode as compared to seco drug **2**.^{74, 80} In contrast to seco drug **2** their inhibitory effect and specificity for ALDH1A1 is based on an extension into a cavity next to the entrance to the binding pocket (Figure I-A15). Although these molecules exhibit a preference for ALDH1A1 compared to other ALDH isoforms, their proteome-wide selectivity has not been evaluated yet.

In summary, this study unravels the basis for the selective interaction between a duocarmycin analog and ALDH1A1. A suite of structural, biochemical, kinetic and computational methods reveal binding to a hydrophobic pocket, stabilization via van der Waals interactions and π -stacking, and covalent attachment to a cysteine residue directly adjacent to the catalytic residue. This mechanism is unique to ALDH1A1 and even closely related ALDH isoforms are not capable of seco drug binding due to sterical restrictions at their entrance channel and the catalytic site. The discovery of this unprecedented binding mode is thus not only of utmost importance for drug development, e.g. to enable adjustment of the compound design to be exclusively DNA or ALDH1A1 specific, but also provides the first tool for a focused study of ALDH1A1 function in whole proteomes. In addition, future studies will address if there are non-covalent binders of seco drug that could contribute to the mode of action.

4. Summary

The aim in this work was to identify the binding site of the duocarmycin derivative seco drug **2** in the ALDH1A1. Seco drug **2** is a small molecule, derived from the core structure of duocarmycin analogs, first isolated from *Streptomyces* strains in 1988.¹⁰⁸ This class of natural compounds contains highly cytotoxic compounds with antimicrobial and antitumor activity.^{1, 109, 110, 111} Their mode of action was originally thought to be based on the alkylation of AT-rich sequences in the minor groove of the DNA.^{3, 82, 83, 85} Because of the activity of duocarmycins further structure activity-based studies were carried out, revealing duocarmycin derivatives that were still potent and cytotoxic even though DNA alkylation could no longer be observed.^{4, 5} Whilst investigating this, *Wirth et al.* identified the ALDH1A1 as selective, covalent target in A549 lung cancer cells. These findings were confirmed by activity-based kinetic studies as well as cell imaging and knock-down studies. Both siRNA knock-down and duocarmycin treatment of A549 lung cancer cells revealed significantly decreased cell proliferation, highlighting ALDH1A1 as potential target in cancer research. Therefore it is important to identify the binding site of seco drug **2** to better understand the selective inhibition mode and to structurally fine-tune this class of small molecules for ALDH1A1 inhibition.

To identify the binding mode of seco drug **2** co-crystallization and mass spectrometry (MS) experiments with human and sheep ALDH1A1 were performed. In these studies Cys302 was identified as the binding site. Cys 302 is directly adjacent to the active site Cys303 in both enzymes. Seco drug **2** precisely fits into the hydrophobic pocket, preventing aldehyde substrates from accessing the active site cysteine. To further validate the identified binding site, site directed mutagenesis in combination with activity-based kinetic studies as well as bottom-up proteomics experiments were performed. Cys302 was validated as the favored binding site. Nevertheless the active site cysteine (Cys303) served as an additional binding site in the mutants Cys302Ala and Cys303Ser. Molecular dynamic (MD) simulations supported these findings.

Next the residues that interact with seco drug **2** were further investigated by site directed mutagenesis in combination with activity-based inhibition assays. Phe171, Trp178 and Tyr297 were shown to be interacting residues that influence the binding of seco drug **2** to ALDH1A1. The most prominent interaction partner, Tyr297 was further studied with MD simulations, revealing π -stacking and interactions with the alkyl chain of seco drug **2**.

Additionally, *in vitro* studies with seco drug **2** were performed to assess inhibition of the closest related ALDH isoforms of ALDH1A1 (ALDH1A2, ALDH1A3 and ALDH2). Activity-based assays with a 50 fold excess of seco drug **2** showed a lack of inhibition for ALDH1A2 and ALDH2 and weak inhibition for ALDH1A3. With reference to the solved crystal structure and the architectures of the hydrophobic pockets of the ALDH isoforms this selectivity can be explained by a change of Ile304 (ALDH1A1) to Thr304 (ALDH1A2/-1A3) and Gly458/His293 (ALDH1A1) to Asp458/Phe293 (ALDH2) in the protein sequences, leading to clashes of the residues with seco drug **2** and thereby significantly reducing affinity. This highlights the promising selectivity of seco drug **2** for ALDH1A1 among the ALDH superfamily. The discovery of selective ALDH inhibitors is of high interest and one can think of diverse applications such as the elucidation of the role of ALDH isoforms in cancer as well as in diseases like Parkinson.^{8, 28}

5. Experimental section

5.1 Cloning, expression and purification of ALDH1A1 enzymes

A codon optimized sequence (GeneArt) of the sheep ALDH1A1 (sALDH1A1) was cloned in frame with a sequence coding for an N-terminal Streptacin-affinity and *Tabacco Etch Virus* (TEV) protease cleavage site into pKM-RQ using the Gateway system (Gateway®, life technologies).

Point mutations were introduced in human ALDH1A1 using Quick Change Mutagenesis (QuikChange II, Agilent Technologies) using the oligonucleotides listed in Table I-A2.

For expression, sALDH1A1-pDEST007 was transformed into *Escherichia coli* BL21 (DE3) and cells grown in LB medium supplemented with 100 µg/ml Ampicillin at 37 °C until an OD₆₀₀ of 0.6 was reached. Expression was induced by adding anhydrotetracyclin to a final concentration of 432 nM and the culture was incubated for 12 h at 18 °C. Cells were lysed by sonication and the protein purified by Streptactin-affinity (StrepTrap™ HP 5 mL, Qiagen). After removal of the Strep-tag using TEV protease the enzyme was subject to size exclusion chromatography (HiLoad 16/60 Superdex 200 prep grade, GE healthcare), concentrated in 50 mM Tris-HCl pH 7.5, 10% glycerin and stored at -80 °C. hALDH1A1 was expressed and purified as previously described.⁵

5.2 Crystallization, data collection and structure determination of ALDH1A1

Freshly purified sALDH1A1 was concentrated to 9-10 mg/mL in HEPES (10 mM, pH 7), 1 mM NAD and 5 mM DTT (protein crystallization buffer 1) and crystallized using 100 mM Bis-Tris, pH 6.0, 4.5–7% PEG5000 and 150-225 mM MgCl₂ at 4 C. Crystals typically appeared after 10-14 days.

For co-crystallization of sheep and human ALDH1A1 with seco drug **2**, 570.7 µL seco drug **2** (154.5 µM) in Tris-HCl 50 mM, pH 8.5 were pre-incubated for 2 h at 30 C and 300 rpm. Afterwards the pH is adjusted to 7.4 and than 864 µL ALDH1A1 (28 µM) are added. The reaction mixture of ALDH1A1 and activated seco drug **2** is incubated at 4 C for 3 d. Binding of seco drug **2** was confirmed by intact protein mass spectrometric analysis using an LTQ-FT-ICR. Crystallization of sALDH1A1 in complex with seco drug **2** was accomplished as described above for the apo enzyme. hALDH1A1 in complex with seco drug **2** was concentrated to 4 mg/mL and crystallized in 100 mM sodium Bis-Tris, pH 6.5, 8–12% PEG3350, 200 mM NaCl, and 5-10 mM YbCl₃ at 25 °C over night.

Diffraction data were collected at the synchrotron beam lines PXI (Swiss Light Source, Villigen, Switzerland) and ID29 (European Synchrotron Radiation Facility, Grenoble, France). Crystals of sheep and human ALDH1A1 in complex with or without seco drug **2** belonged to different space groups and the data were processed with XDS¹¹² to 1.7-2.1 Å spacing, with a resolution cut-off of CC 1/2 of 50%.^{113, 114, 115, 116} For data processing statistics see Supplementary Table I-A1. The structures were solved by molecular replacement or different fourier methods using the ALDH1A1 coordinates (PDB code: 1BXS, 4WP7) in PHASER^{117, 118} and Refmac5¹¹⁹, respectively. In order to reduce model bias prior to molecular replacement/rigid body refinement the coordinates for NAD were removed and the temperature factors were reset, followed by simulated annealing in PHENIX.¹²⁰ Clear peaks for seco drug **2** and Cys302 were visible in the simulated-annealing omit Fo-DFc electron density map. Rounds of model building and refinement were carried out in COOT¹²¹ and REFMAC5.¹¹⁹ Restraints for seco drug **2** were created using JLigand¹²² and the number of TLS groups was determined using the TLSMD server.^{123, 124} Ytterbium ions in the structure of the human ALDH1A1 were placed according to the anomalous difference fourier map calculated with FFT.¹¹⁸ Final coordinates were submitted to the PDB_REDO server.¹²⁵ Diffraction data and refinement statistics are summarized in Table I-A1. Structural super positions were done with SSM¹²⁶ and all structural figures were prepared with PyMol (Delano Scientific, San Carlos, CA). Atomic coordinates were submitted to the Protein Data Bank (<http://www.ebi.ac.uk/pdbe/>) with the PDB codes: sALDH1A1 (APO) = 5abm, sALDH1A1 (seco drug **2**, P2₁2₁2) = 5ac0, sALDH1A1 (seco drug **2**, C2₁) = 5ac1 and hALDH1A1 (seco drug **2**) = 5ac2.

5.3 Aldehyde dehydrogenase activity assay

The activity of sALDH1A1, hALDH1A, ALDH1A1-C302A, ALDH1A1-C302S, ALDH1A2 (Origene, ORIGTP323250), ALDH1A3 (life technologies, 11636-H07E-50) and ALDH2 (Abnova, P3479) were carried out as outlined below. To ensure the formation of active duocarmycin drug, initially 500 µM seco drug **2** was pre-incubated in Tris-HCl 50 mM, pH 8.5 for 2 h at 30 °C. This was followed by adding 2.5 µl activated seco drug **2** to 2.5 µl ALDH1A1 (10 µM), resulting in final concentration of 1 µM enzyme and 50 µM seco drug **2** in a total volume of 50 µL at pH 7.4, and was further incubated for 5 h at 30 °C. Subsequently, 50 µL substrate mixture containing 50 mM Tris-HCl, 100 mM KCl, 5 mM β-mercaptoethanol, 1 mM β-NAD⁺ and 10 mM propanal were added to initiate the enzymatic reaction. The product formation of NADH was monitored by measuring the absorption increase at λ = 340 nm at 37 °C in 96-well-plates with an Infinite 200 PRO NanoQuant microplate reader.

5.4 Determination of K_i and k_{inact} of ALDH1A1

To evaluate the influence of residues lining the binding site of hALDH1A1 the kinetic parameters K_i and k_{inact} of wild-type and mutant hALDH1A1 (Cys302Ala, Cys302Ser, Phe171Ala, Trp178Ala, Tyr297Ala) were determined by the method of Kitz and Wilson.^{87, 88} Initially to ensure the formation of active duocarmycin drug, seco drug **2** was pre-incubated in Tris-HCl 50 mM, pH 8.5 for 2 h at 30 °C at different concentrations. Afterwards 2.5 μ l ALDH1A1 (8.2 μ M) was added to give an enzyme concentration of 410 nM in a total volume of 50 μ L (pH 7.4). The incubation proceeded for different time points. Subsequently, 50 μ L substrate mixture containing 50 mM Tris-HCl, 100 mM KCl, 5 mM β -mercaptoethanol, 1 mM β -NAD⁺ and 10 mM propanal were added to initiate the enzymatic reaction. The product formation of NADH was monitored by measuring the absorption increase at $\lambda = 340$ nm at 37 °C in 96-well-plates with an Infinite 200 PRO NanoQuant microplate reader. All measurements were carried out in technical triplicates and biological duplicates. To determine the kinetic parameters the natural log of the remaining enzyme activity after a certain pre-incubation time was plotted against the pre-incubation time for several inhibitor concentrations [I]. The observed rate of inactivation k_{obs} is derived from the negative slopes of the linear fit. A further double reciprocal plot of k_{obs} against [I] leads to K_i and k_{inact} whereas the values are derived from the slope and the intercept of the linear fit.

5.5 Intact protein mass spectrometry

Intact protein measurements were performed on a Dionex Ultimate 3000 HPLC system coupled to Thermo LTQ-FT Ultra mass spectrometer. Samples were on-line desalted using a Massprep desalting cartridge (Waters) and further analyzed in electrospray ionization (ESI) positive mode (spray voltage 4.0 kV, tube lens 110 V, capillary voltage 48 V, sheath gas 60 arb, aux gas 10 arb) with a resolution of 200,000 and a mass range of 600-2000 m/z. Collected data were deconvoluted using ProMass™ for Xcalibur 2.8 software (Thermo Scientific).

5.6 Time dependent intact protein alkylation experiment of human ALDH1A1

Time dependent alkylation of seco drug **2** with human ALDH1A1 was performed via intact protein mass spectrometry measurements as described above. Thereby 160 μM seco drug **2** were pre-incubated in Tris-HCl 50 mM, pH 8.5 for 2 h at 30 °C. Afterwards 13 μl of human ALDH1A1 (16 μM) and 13 μl activated seco drug (160 μM) were added in a total volume of 130 μL PBS (final pH 7.4) and incubated at 30 °C. Every 27 min 2 μL of the reaction mixture were injected into the LC-MS system and analyzed with ProMass™ for Xcalibur 2.8 software (Thermo Scientific).

Alkylation of duocarmycin SA with human ALDH1A1 was performed as described above for seco drug **2**. Thereby 400 μM duocarmycin SA were pre-incubated in Tris-HCl 50 mM, pH 8.5 for 2 h at 30 °C. This was followed by adding 13 μl of human ALDH1A1 (16 μM) and 13 μl activated duocarmycin SA (400 μM) were added in a total volume of 130 μL PBS (final pH 7.4) and consequent incubation at 30 °C for 3 h. 2 μL of the reaction mixture were injected into the LC-MS system and analyzed with ProMass™ for Xcalibur 2.8 software (Thermo Scientific).

5.7 Gel-based fluorescent labeling experiment of seco drug **2** probe and human ALDH1A1

Following pre-incubation of 160 μM seco drug **2** probe in Tris-HCl 50 mM, pH 8.5 for 2 h at 30 °C, 13 μl of activated seco drug **2**, 10 μl of human ALDH1A1 (1.6 μM) and probe (16 μM) were mixed in a total volume of 130 μL at pH 7.4 and incubated for 2.5 h at 30 °C. Afterwards the reaction mixtures were clicked using rhodamine azide, Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA) and applied on a 10% polyacrylamide gel as described by *Wirth et al.*⁵

5.8 Bottom up proteomics for binding site identification

To discover the binding site of seco drug **2** in mutant enzymes ALDH1A1 (Cys302Ala and Cys302Ser), the recombinant enzymes were labeled, tryptic digested and the peptide fragments analyzed by ESI tandem mass spectrometry. 26 μL of seco drug **2** (577 μM) in Tris-HCl 50 mM, pH 8.5 were activated for 2 h at 30 °C and 300 rpm. 86 μL ALDH1A1 (20 μM) and 1.2 mL PBS were added to give a final concentration of ALDH1A1 of 1.3 μM and seco drug **2** of 11.4 μM (final pH 7.4). The reaction mixture was incubated at 30 °C

and 300 rpm. Binding was controlled by intact protein measurement (>90%). To remove unreacted probe the mixture was washed twice with PBS with a centrifugal filter device (10 kDa molecular weight cut-off) and adjusted to a protein concentration of 1 mg/mL. For the reduction of disulfide bonds 1.3 μ L DTT (1 mM) were added and incubated for 15 min at 37 °C and 800 rpm. After reduction, free thiols were alkylated by the addition of 7.1 μ L iodoacetamide (30 mM) and incubated for 30 min at 22 °C and 800 rpm in the dark. Then 2 μ L chymotrypsin (0.01 nmol) and 1.4 μ L CaCl₂ (10 mM) solution were incubated for 15 h at 37 °C and 800 rpm. Tryptic digest was stopped with 7 μ L formic acid (5%). Afterwards the protein solutions were desalted with stage tips (Empore disk-C18, 47 mm, Agilent Technologies) and filtered via modified nylon 0.45 μ M low protein binding centrifugal filter (VWR). The digested peptides were analyzed on a UltiMate 3000 nano HPLC system (Dionex, Sunnyvale, California, USA) coupled to a Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA). Samples were loaded on a Acclaim C18 PepMap100 75 μ m ID x 2 cm trap and transferred to a Acclaim C18 PepMap RSLC, 75 μ m ID x 15 cm separation column (0.1% FA, 5% DMSO, gradient 10 min 3% ACN, 120 min from 3% to 25% ACN, 5 min to 40% ACN, 0.1 min to 90% ACN and 4.9 min hold at 90% ACN, 0.1 min to 3% ACN and 9.9 min 3% ACN). The mass spectrometer was operated in data dependent top speed mode selecting the most intense precursors with a minimal threshold of 5E3. Precursors were measured in the orbitrap at a resolution of 120,000 and an ion target of 4E5 (max inj. time of 50 ms) in a scan range from 300 to 1700 m/z. Monoisotopic precursor selection was enabled. Charge states from 1 to 7 were triggered. Dynamic exclusion duration was set to 60 s with a mass tolerance of 10 ppm. Precursors were isolated in the quadrupole (isol. window 1.6 m/z) and fragmentation was performed using higher-energy collisional dissociation (HCD). Resulting fragments were measured in the ion trap using a rapid scan rate (AGC target: 1E4 and max inj. time 40 ms). Assignment of the measured peptides to proteins was done with MaxQuant 1.4.0.8 software. Default settings were used except for the following: Precursor mass tolerance: 4.5 ppm; Fragment mass tolerance: 0.5 Da. Protein database: human ALDH1A1 Uniprot, enzyme: chymotrypsin (specificity Tyr, Phe and Trp), variable modifications: oxidation (M) +15.995 Da, carbamidomethyl (C) +57.021 Da; SDB6 (C) +291.126 Da.

5.9 MTT Assay with seco drug 2

Varying concentrations of seco drugs **2** were pre-incubated for 2 h in 100 μ L medium (without FBS) with a final DMSO concentration of 1%. A549 cells were grown in 96-well-plates at a concentration of 7000 cells per cavity. After removing the growth medium the pre-incubated medium containing the reactive drug was added and the cells incubated for 24 h at 37 °C and 5% CO₂. After 24 h of exposure, 20 μ L (5 mg/ml) filtered 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) stock solution in PBS were added and the medium thoroughly mixed by gentle pipetting. The cells were incubated at 37 °C and 5% CO₂ for 1 h to allow the MTT to be metabolized and the reaction was controlled under the microscope. Subsequently, the medium and cell debris were removed and the produced formazan resuspended in 200 μ L DMSO by placing the well-plate on a shaking table for 2 min and 650 rpm. The optical density was read out at $\lambda = 570$ nm and $\lambda = 630$ nm with an Infinite 200 PRO NanoQuant microplate reader and the background was subtracted at $\lambda = 630$ nm. Cells incubated with 1% DMSO served as positive control.

For calculation of IC₅₀ values, residual viabilities for the respective compound concentration were fitted to

$$V = \frac{100}{1 + 10^{(\log(IC_{50}) - \log(c)) \cdot N}}$$

with

V : viability [%]

c : Inhibitor concentration [M]

N : Hill slope

using Graphpad Prism 6.0.

Experiments were carried out in triplicates and two independent experiments. IC₅₀ values are given as mean values and 95% confidence interval.

5.10 DNA alkylation experiment with seco drug 1, seco drug 2 and duocarmycin SA

For the DNA alkylation experiment single strand DNA listed in Table I-A3 were dissolved in 150 mM NaCl and 10 mM Tris-HCl pH 7.5 to a final concentration of 1 mM. For hybridization ss-9 and ss-9rc as well as ss-10 and ss-10rc were mixed in a 1:1 ratio and hybridized via a temperature gradient (95 °C to 4 °C) within 2 hours. For the DNA alkylation experiment seco drug 1, seco drug 2 and duocarmycin SA were pre-incubated for activation in 5 µL Tris-HCl 50 mM, pH 8.5 for 2 h at 30 °C with a final concentration of 5 mM. Afterwards 0.5 µl of the activated duocarmycins (5 mM) were mixed with 7.5 µL water and 2 µL hybridized DNA (50 µM) at a final pH of 7.4 and incubated for 24 h at 25 °C. The samples were analyzed via MALDI using a Bruker autoflex II unit with an MTP AnchorChip var/384 target. Prior to the measurements the samples were desalted for 20 min using MF-Millipore membrane filters (0.025 µM). As MALDI matrix a mixture of 3-hydroxypicolinic acid (50 mg), ammonium hydrogencitrate (10 mg) in 500 µL ddH₂O and 500 µL acetonitrile was applied.

5.11 Putative design criteria for selective ALDH1A1 inhibition or DNA-alkylation

By increasing the size of the aromatic ring system bearing the electrophile, binding to ALDH1A1 could be prevented, resulting in compounds solely targeting ds DNA. For instance attachment of an additional 5 or 6 membered ring to the alkylating benzene moiety would most likely result in occlusion from the ALDH1A1 active site due to steric hindrance with Trp178 and Met175. Based on the solution structure of duocarmycin with dsDNA¹²⁷ (PDB code 1DSA) such a modification is unlikely to disturb the interaction with DNA.

In contrast some small alterations of the hydrophobic ring system on seco drug 2, such as addition of methyl groups could be tolerated by the ALDH1A1 active site, and might fine tune the specific interaction.

5.12 Details for Molecular Dynamics Simulations

XLEAP (AmberTool)¹²⁸ has been used to add hydrogen atoms to the X-ray structure and to solvate the solutes (ALDH1A1 from sheep as monomer with ligands) in a box of explicit TIP3P water¹²⁹ with a buffer of 10 Å around the solute. ANTECHAMBER^{130, 131} was used to parameterize seco drug **2** and NAD. For force field minimizations the NAMD engine was used¹³² with Amber10 force field parameters.¹²⁸

Periodic boundary conditions and particle mesh Ewald summation (PME) with a cutoff value of 12 Å were employed. The system was energy minimized (NVT ensemble) using the conjugate gradient algorithm within 20000 steps. A positional constraint of 1 kcal/mol/Å² on non-water atoms was applied.

The system was heated up to 300 K within 30 ps. In the subsequent equilibration step the system was equilibrated for 230 ps. In the second equilibration step we switched to the NPT ensemble employing the Langevin piston Nosé-Hoover method.^{133, 134} At this stage the constraints on non-water atoms are reduced step by step down to zero (0.1 kcal/mol/Å² increments for every 20 ps). Production runs were performed for 10-100 ns using the SHAKE algorithm¹³⁵ with timesteps of 2 fs.

For proper statistics and to obtain a statistically significant analysis, at least 8 simulations have been performed for each system. The average run time over all systems was 600 ns. The simulation time for each system is listed in Table I-A4.

6. Appendices

6.1 Figures

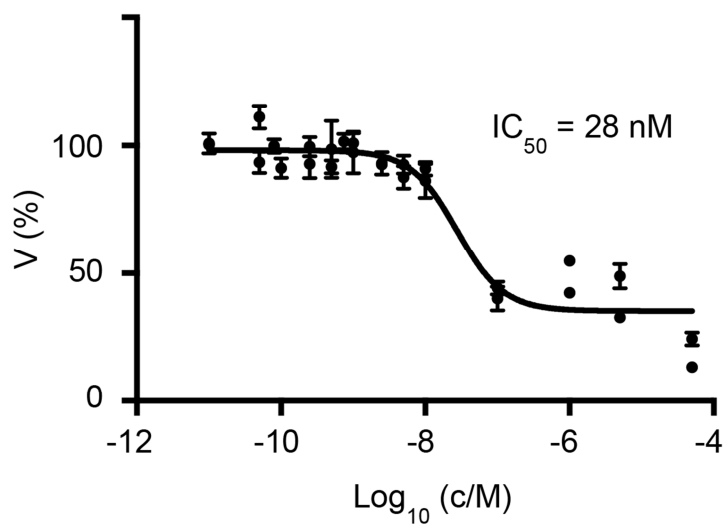
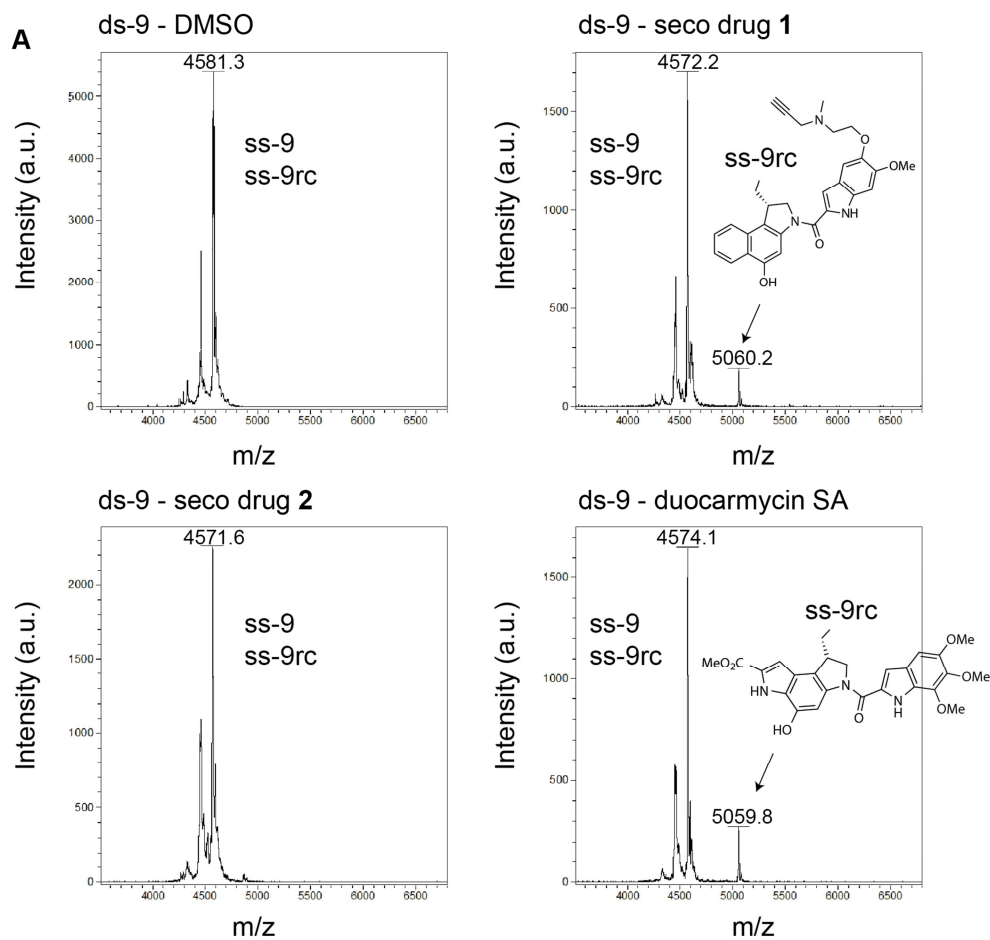


Figure I-A1: Fit of the viability V (MTT assay) for different concentrations c of seco drug 2 ($IC_{50} = 28$ nM with a 95% confidential interval of 12-61 nM).



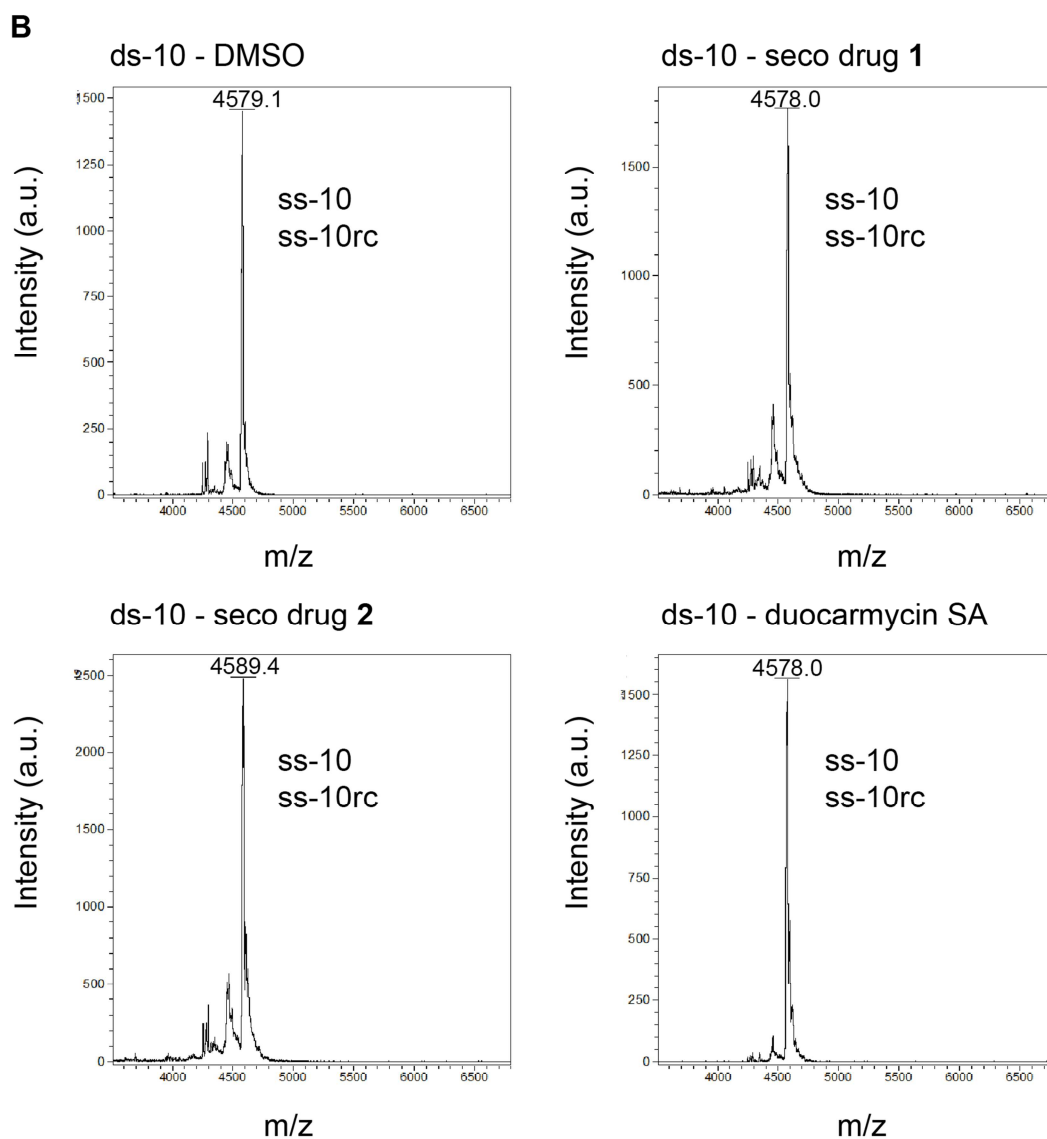


Figure I-A2: MALDI alkylation experiment of *seco drug 1*, *seco drug 2* and *duocarmycin SA* with double stranded DNA as reported by Tietze et al.¹⁰⁰ **A)** Incubation of *duocarmycin* derivatives (25 fold excess) with *ds-9* (AT-rich sequence) for 24 h at 25 °C. Alkylation can be observed for *seco drug 1* and *duocarmycin SA*. For *seco drug 2* no alkylation is detected. **B)** Incubation of *duocarmycin* derivatives with *ds-10* (AT deficient control sequence) for 24 h at 25 °C. No alkylation can be observed for the three derivatives.

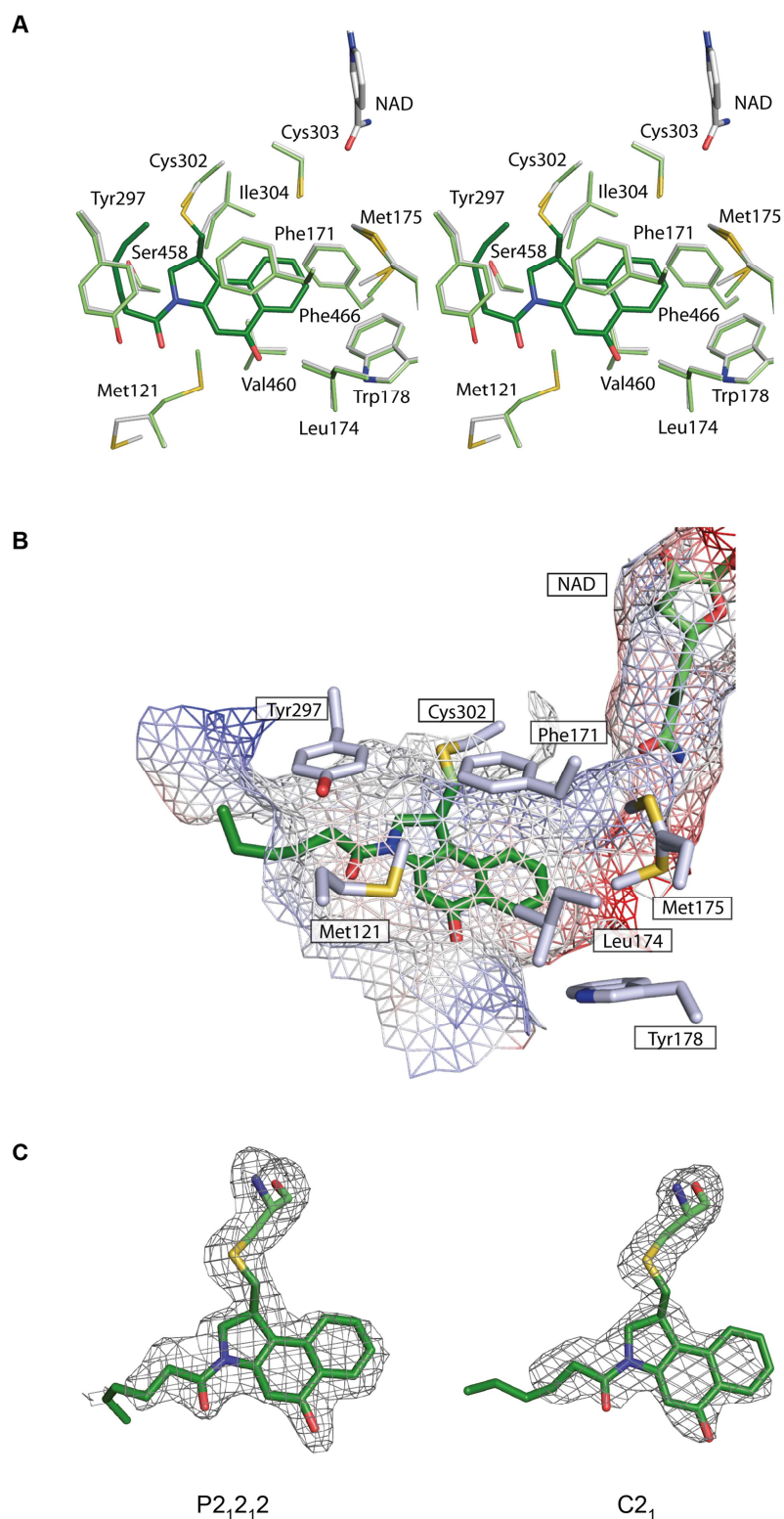


Figure I-A3: **A)** Stereo view of the binding pocket of sheep ALDH1A1, in complex with *seco* drug **2** (green) and the apo structure (white) superimposed. **B)** Surface representation of the binding pocket of sheep ALDH1A1 with bound *seco* drug **2**. The surface of the binding pocket is shown as mesh, with the charge distribution indicated by coloring (red = negative, blue = positive). **C)** Simulated annealing *F_o-DF_c* omit electron density map of *seco* drug **2** bound to Cys301 of sheep ALDH1A1, contoured at 2.5 σ , in the obtained two different crystal forms (PDB codes 5ac1 (P2₁2₁2) and 5ac2 (C₂₁)).

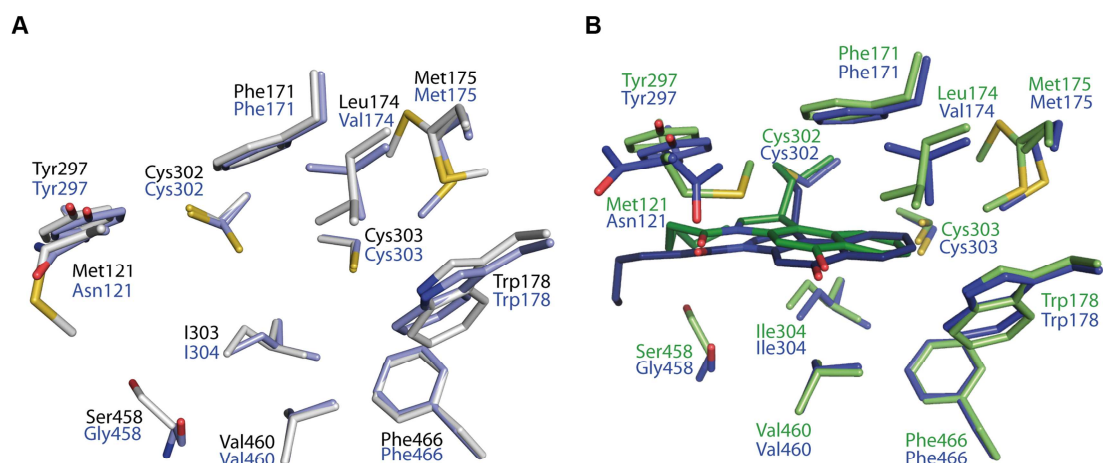


Figure I-A4: **A)** Structural superposition of the apo crystal structures of sheep (grey, this work, PDB code 5abm) and human ALDH1A1 (blue, PDB code 4WJ9). **B)** Comparison of *seco* drug **2** bound to human ALDH1A1 (blue, PDB code 5ac2) and sheep ALDH1A1 (green, PDB code 5ac0/5ac1).

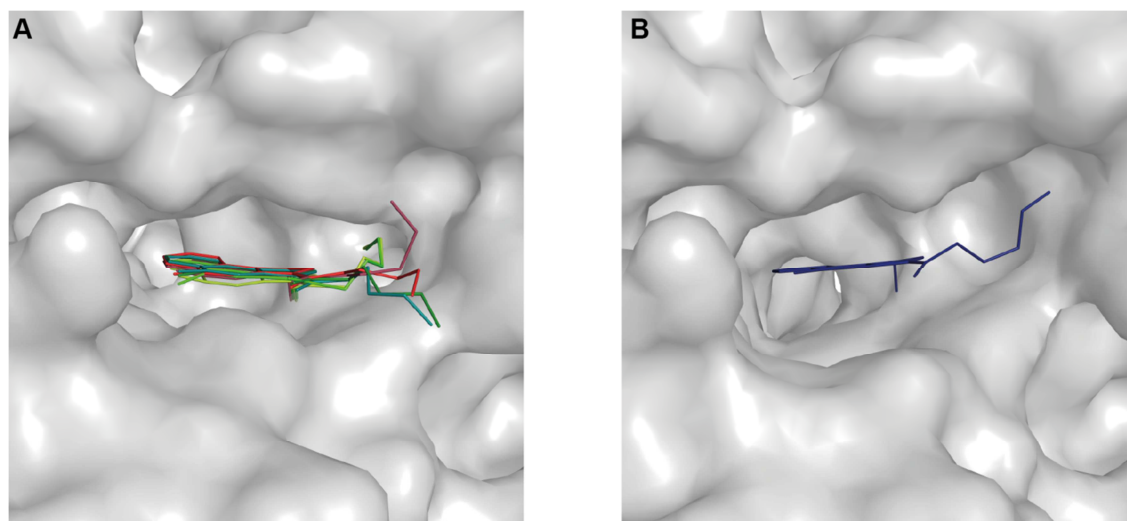


Figure I-A5: **A)** Comparison of *seco* drug **2** bound to the sheep ALDH1A1 monomers in the asymmetric unit of the crystal in the space group $C2_1$ (4 molecules, green/cyan, PDB code 5ac0) and $P2_12_12$ (2 molecules, red, PDB code 5ac1). **B)** *Seco* drug **2** bound to human ALDH1A1 (PDB code 5ac2).

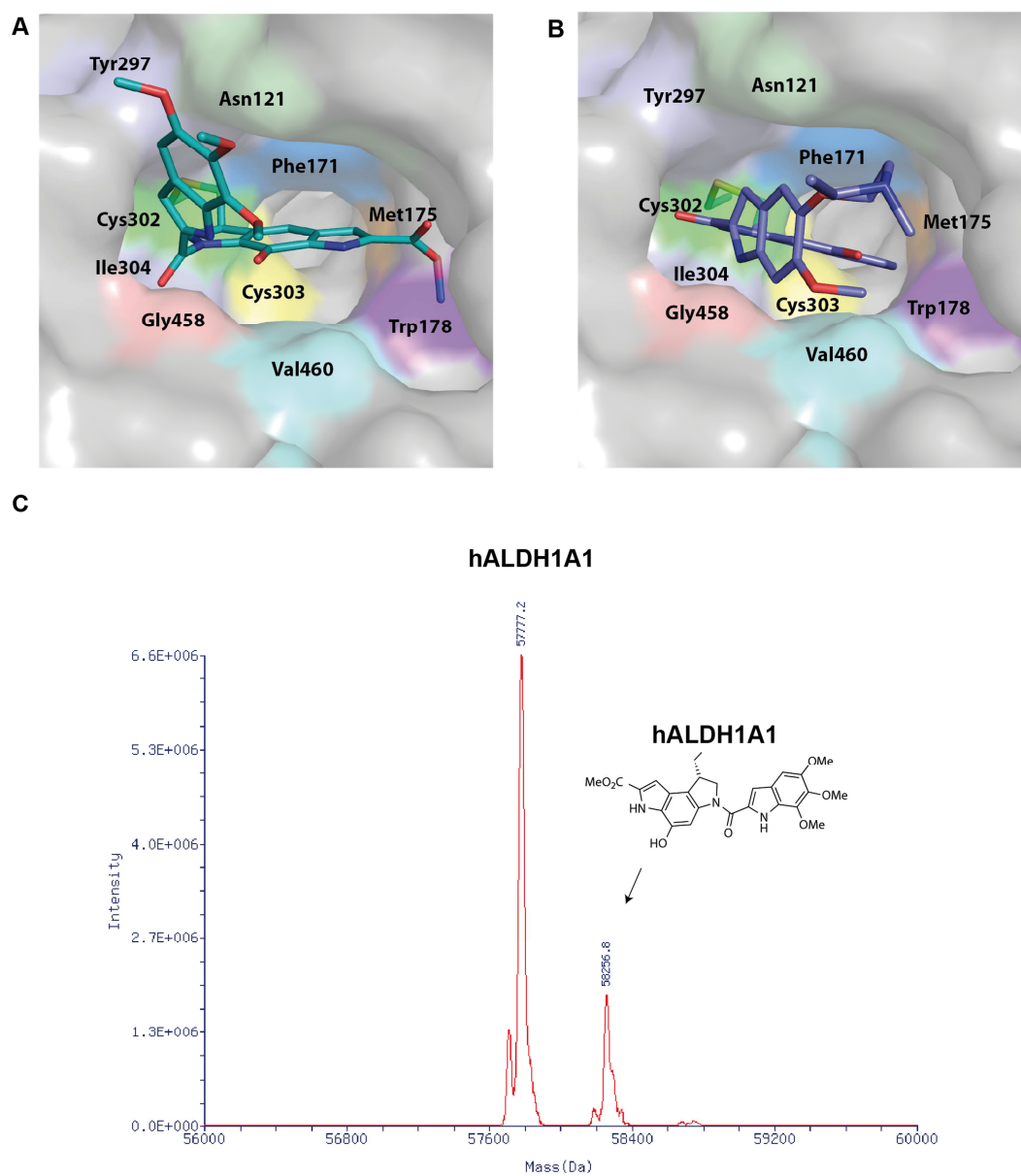


Figure I-A6: Model of Duocarmycin SA bound to Cys302 (**A**) and seco drug **1** (**B**) bound to Cys 302 in the active site pocket of human ALDH1A1. C) Confirmation of alkylation and binding of duocarmycin SA to wild-type hALDH1A1 by ESI-LC-MS. Duocarmycin SA and hALDH1A1 are incubated 3 h at 30 °C with a 25 fold excess of duocarmycin SA.

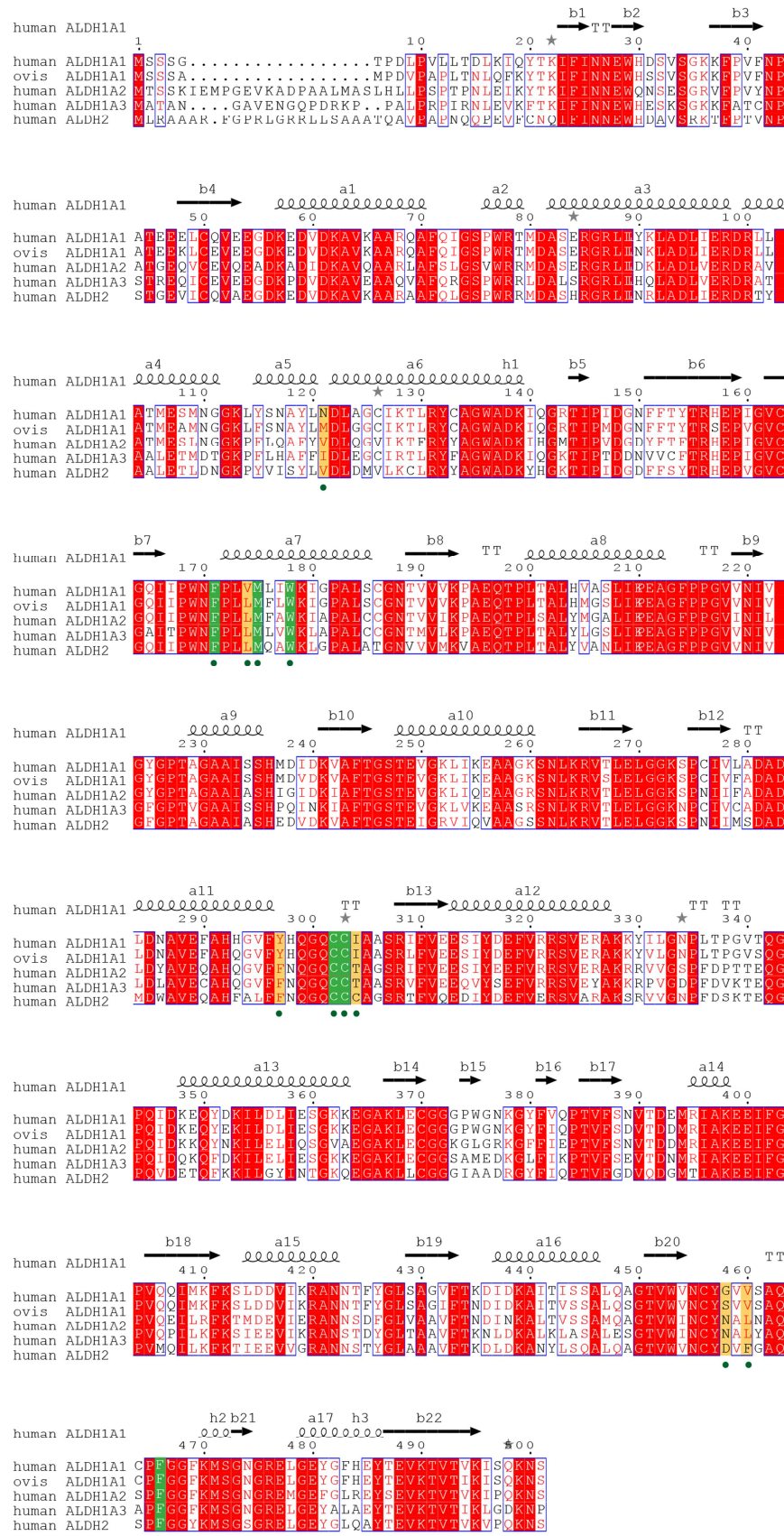


Figure I-A7: Sequence alignment of ALDH1A1 homologues.^{136, 137} Identical amino acids are highlighted in red. Residues lining the binding pocket are marked with dots and are highlighted in yellow (partly conserved) and green (conserved).

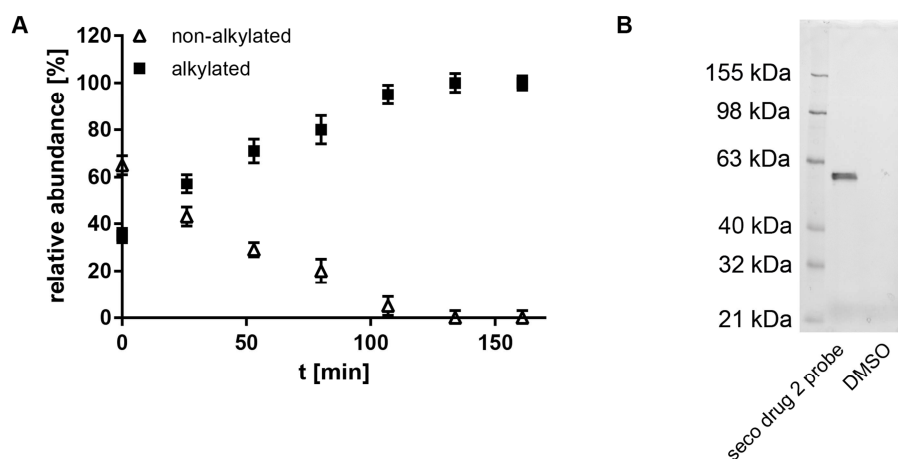


Figure I-A8: *hALDH1A1* alkylation by *seco drug 2*. **A)** Time course of *ALDH1A1* alkylation by relative quantification of by intact protein MS after addition of 10 fold excess of *seco drug 2* to human wild-type *ALDH1A1*. **B)** In-gel fluorescent labeling of human wild-type *ALDH1A1* by click chemistry after addition of a 10 fold excess of *seco drug 2* probe at pH 7.4 and incubation of 2.5 h.

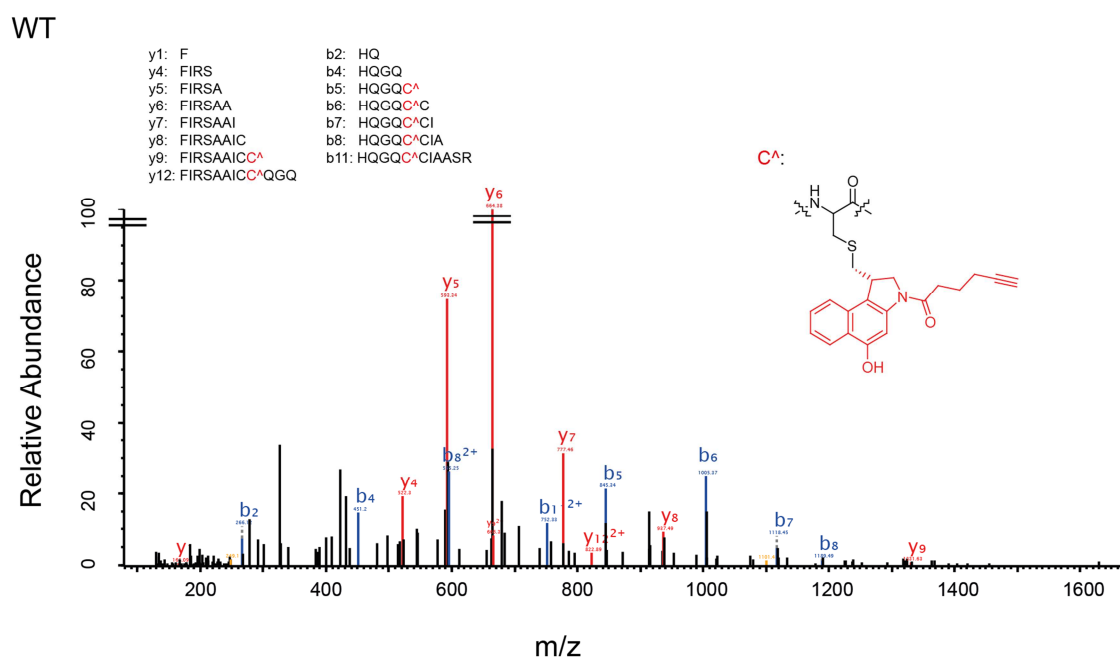


Figure I-A9: LC-MS/MS binding site identification in human *ALDH1A1* with *seco drug 2* probe. Displayed are fragmented peptides (*y* and *b* ions) identified by MS/MS sequencing. The modified cysteine-containing ions are indicated by: *C[^]* (highlighted in red).

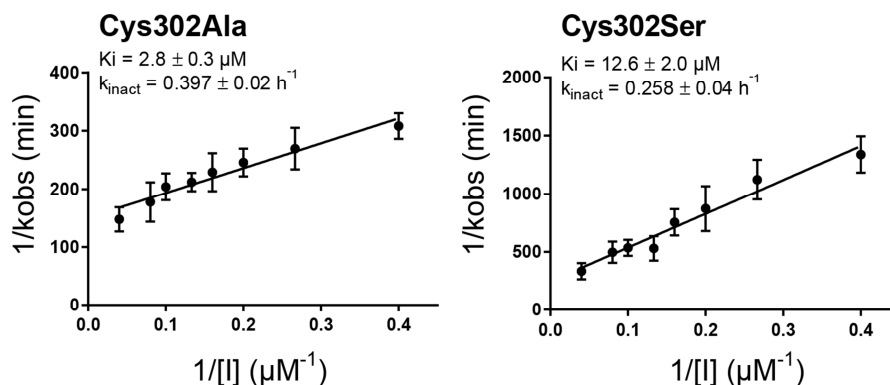


Figure I-A10: Determination of K_i and k_{obs} according the method of Kitz and Wilson⁸⁷ for the ALDH1A1 binding site mutants Cys302Ala and Cys302Ser.

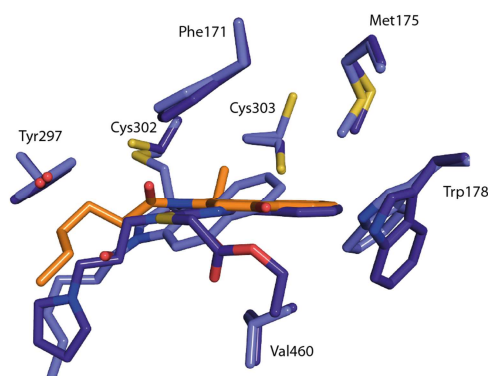


Figure I-A11: Model for seco drug 2 binding to Cys303. In the hALDH1A1 structure bound to the CM037 inhibitor (light blue, PDB code 4X4L)⁷⁴ Trp178 is shifted. Such a shift would allow seco drug 2 (dark blue) to move (orange) into proximity to Cys303.

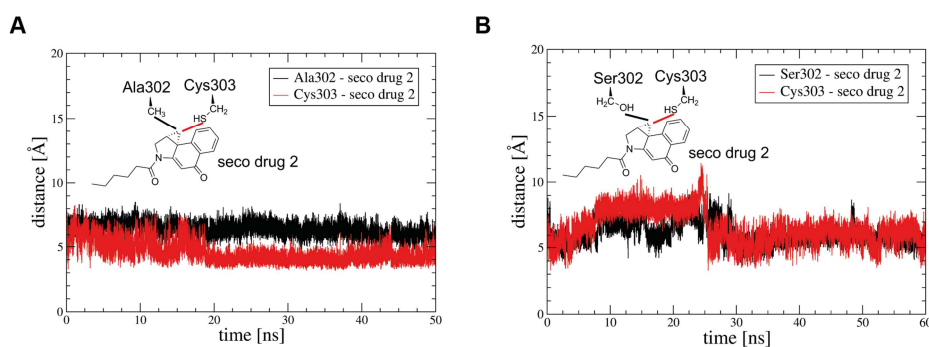
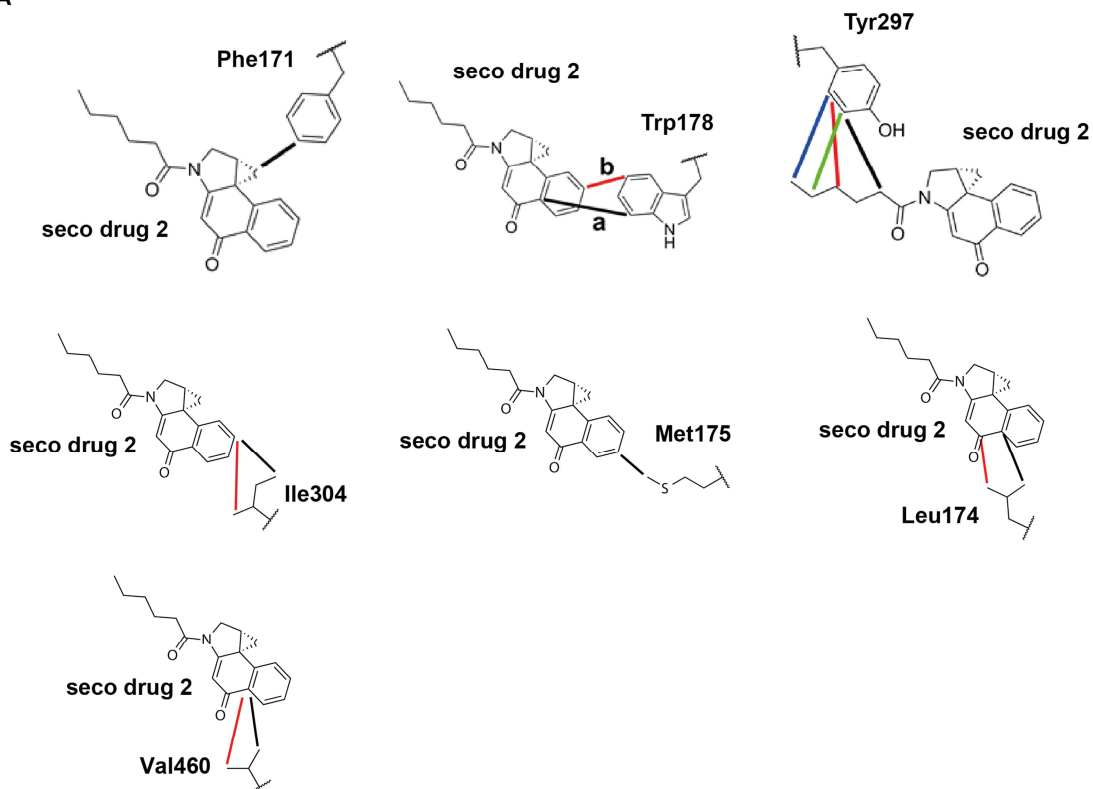
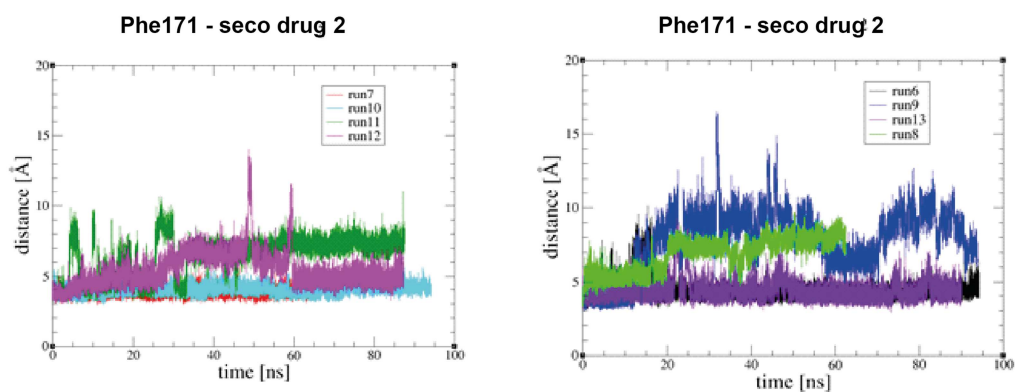


Figure I-A12: Structural analysis of the Cys302Ala and Cys302Ser interaction with seco drug 2 by FF-MD. **A)** Distance of the C_β atom of Cys302 and sulfur atom of Cys303 to the cyclopropyl moiety of seco drug 2 from MD simulations. Cys303 is closer to the cyclopropyl moiety than Ala302 and shows strong and permanent interaction. **B)** Distance of the oxygen side chain atom of Ser302 and sulfur atom of Cys303 to the cyclopropyl moiety of seco drug 2 from MD simulations. Cys303 shows multiple interactions with the cyclopropyl moiety.

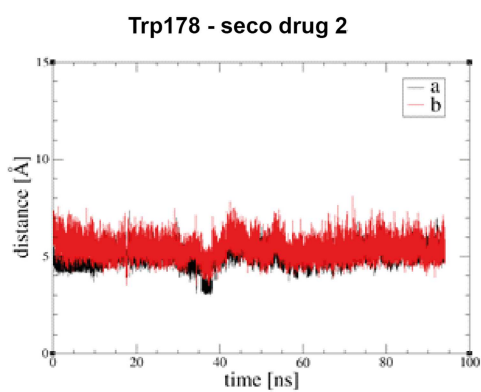
A



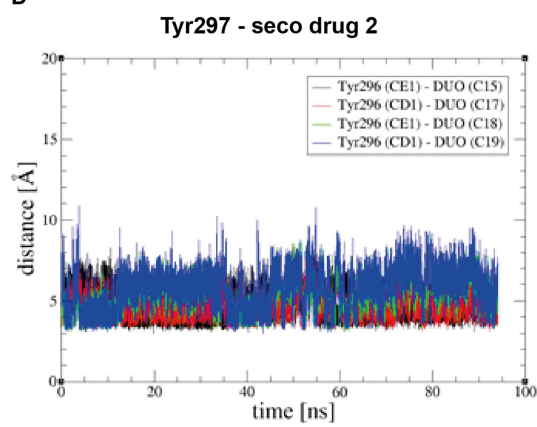
B



C



D



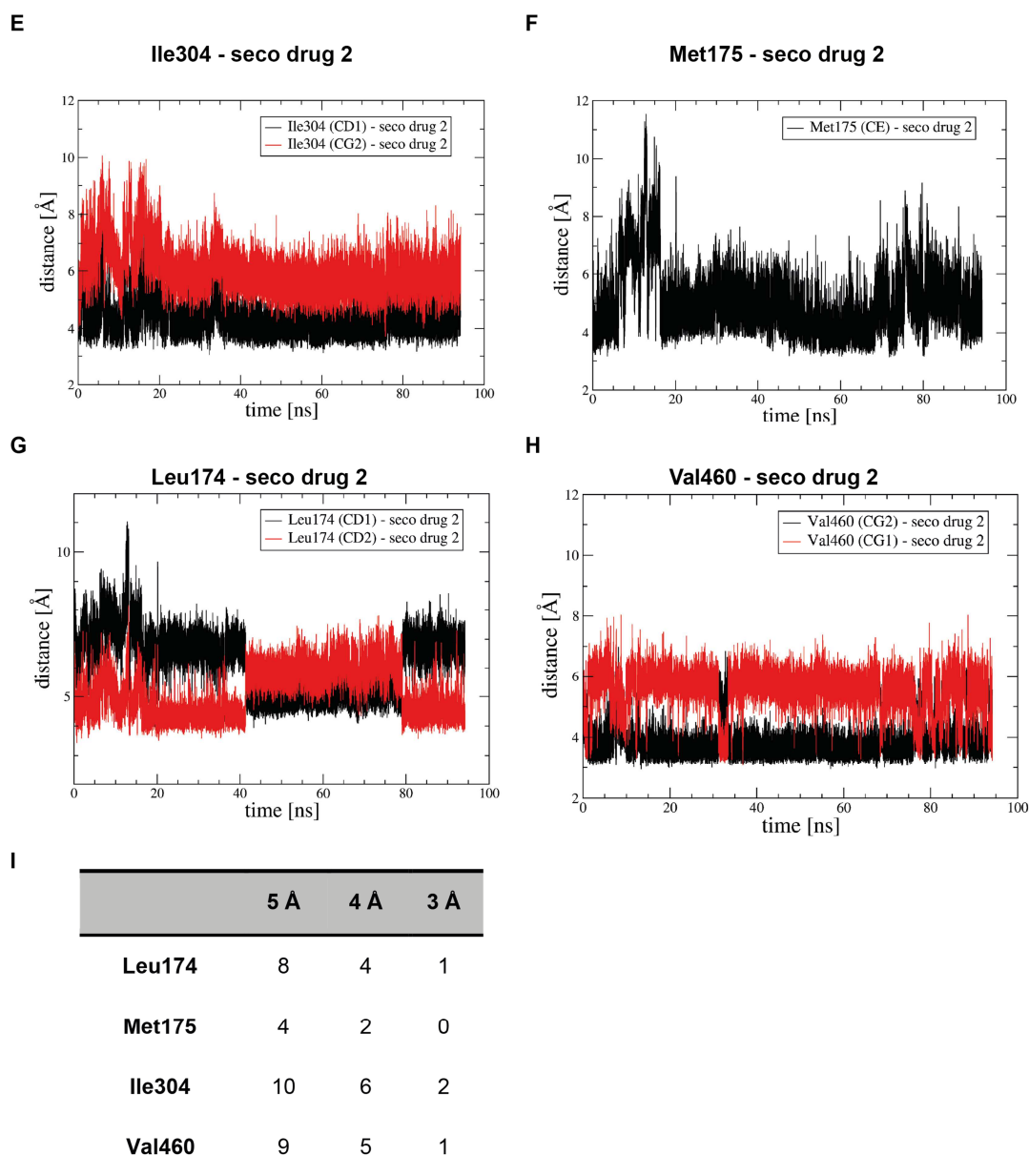


Figure I-A13: **A)** Definition of the plotted distance between the specified atoms for the interaction of Phe171, Trp178, Tyr297, Ile304, Met175, Leu174 and Val460 with seco drug 2. **B)** In 50% of MD simulations Phe171 shows very stable interaction with the cyclopropyl moiety of seco drug 2. **C)** Distances between Trp178 and seco drug 2 for a representative MD simulation. **D)** The interaction of Tyr297 with seco drug 2 out of one representative MD simulation. Tyr297 interacts with all atoms of the aliphatic chain of seco drug 2. This is shown for the distances between tyrosine atoms $C_{\epsilon}1$ and $C_{\delta}1$ and the carbon atoms of the aliphatic chain of seco drug 2. **E)** Distances between Ile304 and seco drug 2 for a representative MD simulation. **F)** Distances between Met175 and seco drug 2 for a representative MD simulation. **G)** Distances between Leu174 and seco drug 2 for a representative MD simulation. **H)** Distances between Val460 and seco drug 2 for a representative MD simulation. **I)** Average number of Leu174, Met175, Ile304 and Val460 atoms interacting with seco drug 2 at the indicated distances (MD calculations).

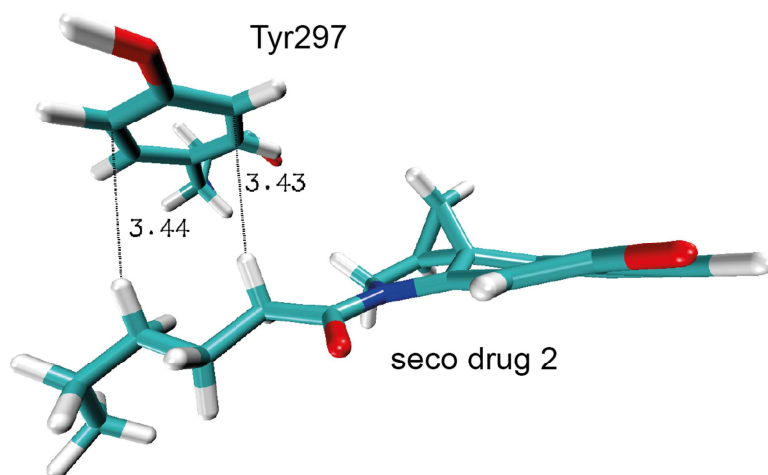


Figure I-A14: MD snapshot of the Interaction between Tyr297 and seco drug 2.

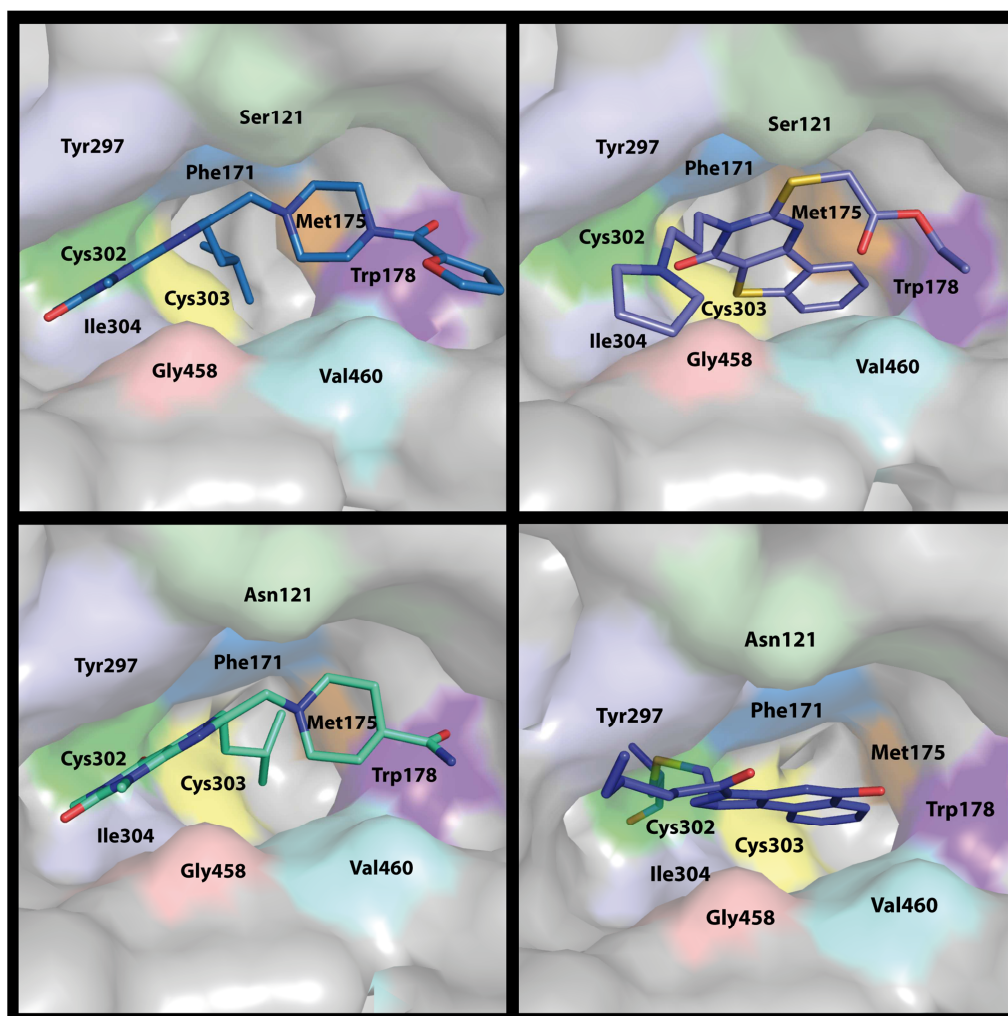


Figure I-A15: Comparison of human ALDH1A1 inhibitors in the active site pocket. The inhibitors are bound at the entrance of the active side. Seco drug 2 is covalently bound to Cys302 located in the inner pocket. **A)** CM026 (PDB code 4WP7), **B)** CM037 (PDB code 4X4L), **C)** CM053 (PDB code 4WPN) and **D)** seco drug 2 (this work, PDB code 5ac2).

6.2 Tables

Table I-A1:

	oALDH1A1	oALDH1A1-seco drug 2	oALDH1A1-seco drug 2	hALDH1A1-seco drug 2
Wavelength (Å)	1	1	1	1
Resolution range (Å)	51.8 - 1.7 (1.76 - 1.7)	50.2 - 1.9 (1.96 - 1.9)	49.4 - 2.1 (2.16 - 2.1)	109.1 - 1.85 (1.91 - 1.8)
Space group	C 1 2 1	P 21 21 2	C 1 2 1	P 4 2 2
Unit cell	188.0 80.8 171.4 90 118 90	93.6 150.6 80.7 90 90 90	188.3 80.8 172.3 90 118 90	109.1 109.1 83.2 90 90 90
Total reflections	1,684,395 (158,838)	722,636 (61,494)	903,139 (67,598)	1,135,506 (102,214)
Unique reflections	249,604 (24,273)	90,805 (8,837)	134,836 (11,553)	43,648 (4,180)
Multiplicity	6.7 (6.5)	6.8 (6.0)	6.7 (5.8)	26.0 (24.5)
Completeness (%)	99.7 (97.3)	99.8 (98.1)	98.4 (84.9)	99.8 (98.1)
Mean I/sigma(I)	12.6 (2.1)	10.56 (1.43)	8.9 (1.7)	17.8 (1.2)
Wilson B-factor	19.5	21.8	28.9	27.4
R-merge	0.106 (0.861)	0.125 (1.02)	0.177 (0.921)	0.188 (2.676)
R-meas	0.115	0.136	0.192	0.192
CC1/2	0.998 (0.71)	0.998 (0.71)	0.993 (0.70)	0.999 (0.57)
CC*	0.999 (0.91)	0.999 (0.91)	0.998 (0.91)	1 (0.85)
R-work	0.180 (0.299)	0.153 (0.281)	0.229 (0.338)	0.193 (0.371)
R-free	0.203 (0.308)	0.197 (0.317)	0.265 (0.368)	0.228 (0.387)
Number of non-hydrogen atoms	16,987	8,593	16,425	4,094
macromolecules	15,464	7,760	15,427	3,849
ligands	180	134	268	69
Protein residues	1,984	994	1,988	494
RMS(bonds)	0.017	0.019	0.015	0.010
RMS(angles)	1.73	1.85	1.63	1.42
Ramachandran favored (%)	98	97	97	97
Ramachandran outliers (%)	0	0	0	0.2
Clashscore	2.83	2.5	2.8	1.4
Average B-factor	22.2	25.9	27.9	32.4
macromolecules	21.8	25.2	27.8	32.4
ligands	22.2	31.9	37.9	35.10
solvent	26.2	31.9	25.9	31.6

Table I-A2: Primers for site directed mutagenesis of human ALDH1A1.

Name		Primer Sequence (5' to 3')
C301A	forw	gga tgc ggc tat aca agc ctg gcc ctg gtg gta g
C301A	rev	cta cca cca ggg cca ggc ttg tat agc cgc atc c
C301S	forw	gcg gct ata caa ctc tgg ccc tgg tgg
C301S	rev	cca cca ggg cca gag ttg tat agc cgc
F170A	forw	gca taa cca acg ggg cat tcc aag gaa tga ttt ggc cac ata
F170A	Rev	tat gtg gcc aaa tca ttc ctt gga atg ccc cgt tgg tta tgc
W177A	forw	cag gcc cta tct tcg caa tga gca taa cca acg gga aat t
W177A	Rev	aat ttc ccg ttg gtt atg ctc att gcg aag ata ggg cct g
Y296A	forw	cac tgg ccc tgg tgg gcg aat acc cca tgg tg
Y296A	Rev	cac cat ggg gta ttc gcc cac cag ggc cag tg

Table I-A3: Base sequences of the four single stranded (ss) DNA oligonucleotides used in the present study. A: Adenine, G: Guanine, C: Cytosine, T: Thymine.

Name	Base Sequence (5' to 3')
ss-9	cgg ctt ata tga ccg
ss-9rc	cgg tca tat aag ccg
ss-10	tgg ctg cga gca cct
ss-10rc	agg tgc tcg cag cca

Table I-A4: Overview of performed simulations on different systems.

system	total time [ns]	number of simulations
WT	720	13
C302A	550	8
C302S	550	8

Table I-A5: Peptides identified by mass spectrometry containing modified cysteine residues C⁺ (highlighted in red). The table shows the fragment ion mass (MH⁺), the fragment ion charge (z), the Mass error, the minimal reached p value (PEP) and the Score value.

ALDH1A1	Peptide	MH ⁺	z	Mass error [ppm]	PEP	Score
WT	HQGQC ⁺ CIAASRIF	1432.7	3	-0.31	1E-04	58
C301S	HQGQSC ⁺ IAASRIF	1416.7	3	0.52	9E-08	64
C301A	HQGQAC ⁺ IAASRIF	1400.7	3	1.3	4E-37	103



II - Metabolomics

1. Introduction

1.1 General aspects of metabolomics

Metabolomics is the latest of the so-called *-omics* cascade and was introduced in 1998 in analogy to genomics, transcriptomics and proteomics (Figure II-1).¹³⁸ It represents the endpoint of the cascade coming as close as possible to the resulting phenotype.⁶ In general, metabolomics is about the comprehensive and quantitative analysis of metabolites. More specifically, it is about the study of small molecule metabolite profiles caused by specific cellular processes. The analysis of these unique metabolite fingerprints has diverse applications e.g. global systems biology, clinical toxicology, nutrigenomic and functional genomics.¹³⁹

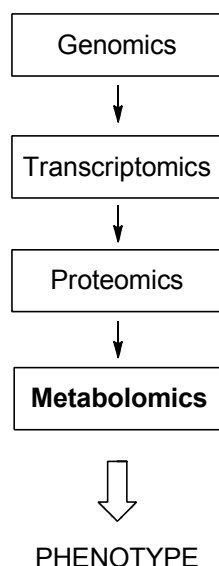


Figure II-1: The *-omics* cascade comprises complex datasets that as an entity comprehensively describe the response of biological systems to disease, genetic, and environmental perturbations. The most powerful database will integrate data from all omic levels. However, of these databases the metabolome is the most predictive of phenotype.⁷

Metabolites are small molecules (< 1500 Da) occurring in diverse biological matrices, e.g. cells, tissues and biofluids. A huge number of metabolites can be found within the different biological systems. For example 200,000 metabolites have been identified in plants⁶, more than 4,000 in human serum¹⁴⁰ and approximately 2,700 in the model organism *Escherichia coli* (*E. coli*).¹⁴¹ Metabolites are the end products of cellular processes and represent the ultimate response to environmental changes. They, are involved in a multitude of biochemical reactions forming a complex, tightly controlled metabolic network.^{142, 143} The complete set of small molecules in a biological system is called the metabolome.⁷ It consists of highly hydrophilic metabolites located in the cytosol,

such as nucleotide triphosphates and sugars, as well as highly hydrophobic metabolites as part of the cell membrane. As a result of the vast structural diversity and high structural complexity, differing metabolite concentration levels can be found. As an example, the concentration levels of metabolites in human serum cover 11 orders of magnitude.¹⁴⁰

With respect to this challenging *-omics* research field, sophisticated analytical platforms are required for a metabolomic workflow. To date, the most commonly applied analytical techniques include nuclear magnetic resonance spectroscopy (NMR), Fourier transform infrared spectroscopy (FTIR) and mass spectrometry (MS).¹³⁹ Due to its constant technical evolution, its outstanding sensitivity, as well as its robustness and accuracy, MS is the most common applied analytical platform today.¹⁴⁴ In combination with liquid chromatography techniques, it allows for sensitive determination of metabolite levels even in complex biological matrices. Beside the analytical platform, sample preparation and data handling are important aspects in the metabolomics workflow, which will be discussed in detail in the following chapters.⁷

1.2 Targeted and untargeted strategies in metabolomics

In recent years, diverse strategies for determining metabolite levels have evolved. In general, they can be divided into targeted and untargeted approaches (Figure II-2). Targeted approaches, including targeted metabolomics and metabolite profiling, aim to analyze and quantify specific, known small molecule metabolites. More precisely, targeted metabolomics is the extraction and analysis of only one specific metabolite.⁷ Contrastingly, in metabolite profiling experiments, a greater number of metabolites belonging to the same metabolite class or to the same metabolic pathway are addressed.⁶ In both approaches, the use of internal standards enables the relative and/or absolute quantification of metabolites.^{145, 146} A significant advantage of the targeted approach is the reduction in complexity of the biological samples within the metabolomic workflow, including sample preparation as well as the analytical platform. This leads to increased sensitivity and metabolite identification even in complex biological samples, due to specialized extraction protocols as well as optimized LC-MS-based methods.

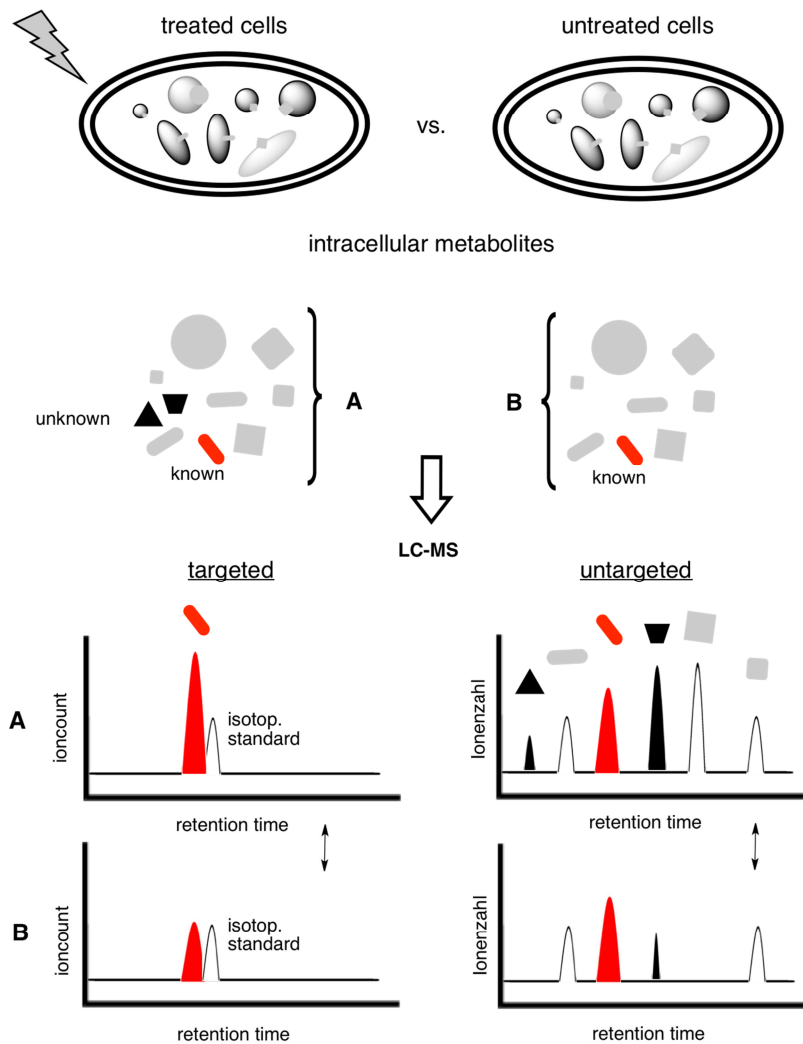


Figure II-2: Targeted and untargeted experiment for the analysis of endogenous metabolites in bacteria. Shown is a comparative workflow, in which treated (A) vs. untreated (B) cells are analyzed via a LC-MS-based platform. In the targeted approach, a specific, known metabolite can be absolutely quantified with the use of an isotopic standard metabolite. In the untargeted approach, both known and unknown metabolites are analyzed and relatively quantified.

In untargeted metabolomic approaches, the aim is to analyze the complete set of metabolites within the biological sample. In addition to the detection and relative quantification of the m/z -values of the metabolite ions, the distinct metabolite identification plays an important role in this metabolomics platform. Therefore sophisticated tandem MS methods are applied, generating unique fragmentation patterns of the analyzed metabolites. In contrast, the so-called metabolite fingerprinting approach aims to detect and analyze the complete set of m/z -values referring to the metabolome. As a result, the quantification is not as crucial as in the metabolomic approach described above, but nevertheless, distinct changes in m/z -values in comparative experiments can be useful for metabolite identification. Both untargeted approaches are confronted with the huge structural as well as polarity-related diversity of metabolites found in biological samples.

This affects the workflow of the untargeted approaches, including the sample preparation, the analytical platform and the data analysis. Herein especially, the full-scale metabolite identification is a challenging aspect even though there are free online databases such as METLIN and the human metabolome database (HMDB) available.

Applications for targeted and untargeted metabolomic strategies are very diverse. Metabolite levels are directly linked to the phenotype and therefore new interactions in biological systems can be revealed. With this, it is possible to functionally characterize specific enzymes or to further investigate complete pathways. Especially in the comparative experimental mode (e.g. treated vs. untreated biological samples), metabolite studies can reveal new insights into the complex and highly dynamic network since metabolites are directly linked to proteins as well as genes.

1.3 Metabolic footprinting in *Streptomyces* extracts as a tool for natural product discovery

In addition to the targeted and untargeted approaches, metabolomic footprinting is a prominent and widely used strategy in metabolomics, especially in the discovery of natural products.^{7, 147} In this approach, the complete set of extracellular, secreted metabolites are the focus of the analysis, including small reactive natural products. Natural products are defined as small molecules synthesized by nature, which often exhibit naturally fine-tuned reactivity.¹⁴⁸ The identification of natural product inspired bioactive scaffolds is of essential need in the drug discovery field, especially in the antibiotic area due to resistance development.^{149, 150, 151} More than two thirds of the clinically useful antibiotics of natural origin including neomycin, chloramphenicol, showdomycin, fosfomycin and penicillin G are produced by bacterial *Streptomyces* strains.¹⁵² *Streptomyces* is the largest genus of actinobacteria and the type genus of the family Streptomycetaceae. More than 500 different species of *Streptomyces* bacteria have been described, and, as with other actinobacteria, are Gram-positive. Under specific stress-related conditions, some of them are able to produce and secrete small reactive molecules as protection mechanisms against environmental enemies.^{152, 153} In this context, the aim of the metabolomic footprinting in *Streptomyces* strains is the discovery of new reactive small molecules. The vast number of *Streptomyces* strains identified to date, in combination with ability to produce interesting natural products, suggests a high potential for discovering unknown reactive metabolites. However, the discovery of these highly interesting and potent chemicals is challenging because the conditions under which

the bacteria are secreting these molecules are mostly unknown and the concentration levels of these metabolites are at the limit of detection. As a result, the identification of further metabolites is particularly difficult despite significant advances in technology and sophisticated analytical platforms in the recent years.

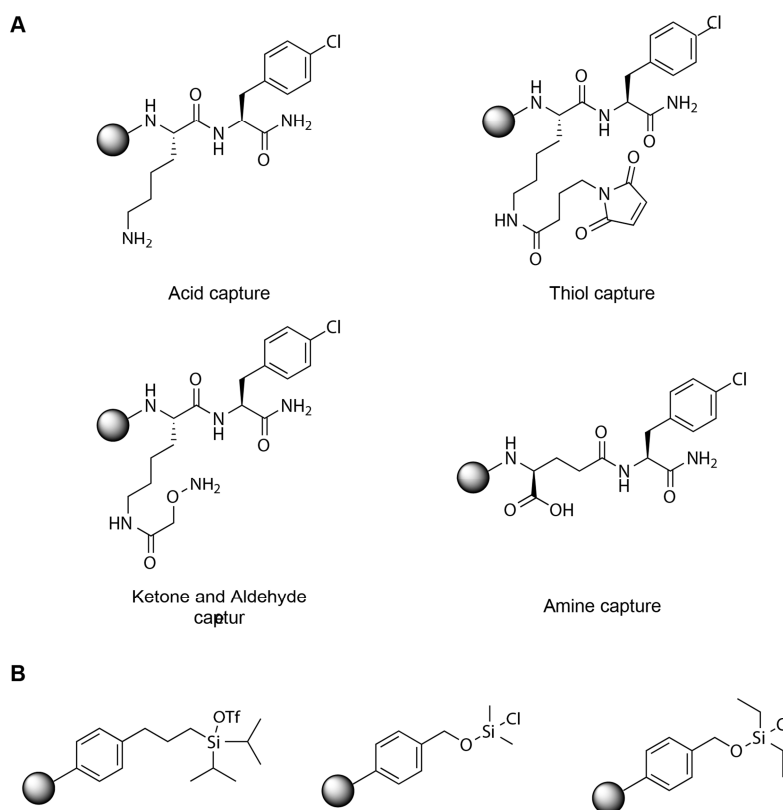


Figure II-3: A) Cl-Phe tags designed by Carlson *et al.* to increase the hydrophobicity of labeled small molecules and improve chromatographic retention of polar metabolites.¹⁵⁴ B) Controllably reversible covalent enrichment tags for chemoselective isolation of alcohol-containing natural products from complex mixtures.¹⁵⁵

To overcome the difficulties associated with the identification of low abundant metabolites, diverse LC-MS-based techniques have been established in the recent years. One prominent approach, introduced by Carlson *et al.*, is the chemoselective identification of highly polar metabolites.¹⁵⁴ In this strategy, a hydrophobic, chlorine-bearing tag is attached to the metabolites (Figure II-3A). This not only enhances the separation of the highly polar metabolites by well-established reversed phase strategies, but also simplifies the identification of tagged metabolites because of the unique chlorine MS pattern. Another promising approach for metabolite identification is the use of controllable, reversible covalent enrichment tags established by Odendaal *et al.* (Figure II-3B).¹⁵⁵ These tags show high chemoselectivity for natural products bearing alcohol functionalities and enable the enrichment and identification of metabolites.

Both strategies highlight the potential of chemoselective, metabolite-capturing enrichment tags in the field of natural product discovery. Thereby, both the separation of highly polar natural products, which are often difficult to separate with common LC methods, and their identification can drastically be facilitated. These enrichment methods are especially useful for low-abundant natural products, and will hopefully result in the discovery of unknown small molecule metabolites.

1.4 LC-MS-based analytical platforms in metabolomics

Metabolomics studies use diverse analytical platforms, largely based on nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS).⁷ NMR is widely used in high-throughput fingerprinting experiments with the advantage of minimal sample preparation requirements as well as the non-discriminating and non-destructive nature of the technique. In comparison to MS-based techniques, drawbacks of NMR are the sensitivity, since only medium to high abundant metabolites can be detected, and the difficulty in identifying metabolites in complex biological matrices. To date, the most widely used technique in the field of metabolomics is a combination of high-performance liquid chromatography (HPLC) coupled to mass spectrometry, which will be discussed in detail in this chapter.¹⁵⁶

The first step in LC-MS-based metabolomics is the time-resolved separation of the biological matrix, an important part of the analytical platform as it highly influences MS sensitivity. The aim of this is to reduce the complexity of the matrix and to minimize co-elution and ion suppression effects. Today, a variety of LC methods, depending on the polarity of the analyzed compounds, are available. The most common and best-established technique is reversed-phase liquid chromatography (RP-LC), which separates compounds according to their hydrophobicity.^{157, 158, 159} The most frequently used stationary phase in RP-LC is octadecylsilane (C18), which is suitable for the separation of nonpolar to slightly polar compounds. However, when examining a complete set of metabolites, suitable methods for the detection of hydrophilic metabolites need to be integrated into the LC-MS-based platform.¹⁶⁰ To analyze polar metabolites, diverse chromatographic methods are available including hydrophilic liquid interaction chromatography (HILIC), ion-exchange chromatography (IEC), aqueous normal-phase chromatography (ANP) and mixed-mode chromatography (MMC).^{161, 162} HILIC especially, is very useful in the separation of polar metabolites, showing good separation capacities of highly polar compounds.^{161, 163, 164, 165} Furthermore, the mobile phase of HILIC contains

a high percentage of organic solvents which improves ionization efficiency and also sensitivity, making it a good partner for MS. Nevertheless, compared to the well-established RP-LC, long gradient times as well as extensive washing and re-equilibration steps need to be taken into account.

In addition to the LC methods, dimensions of the column used are an important factor influencing separation resolution as well as separation time (Table II-1).¹⁴⁴ In general, shorter separation times give narrower peaks since there is less time for diffusive broadening, and therefore decreases the MS detection limit. The key obtaining shorter separation times is the use of shorter columns or higher flow rates without a sacrifice in resolution, which is possible by using columns with smaller particle sizes. Currently the golden standard in metabolomics is the so-called ultra-high performance liquid chromatography (UHPLC), which has a significantly decreased particle size that is smaller than 2 μm .^{144, 160} It has an internal diameter range of 1.5 - 4.6 mm and a length of 30 to 200 mm. Depending on the column dimensions, flow rates of 0.2 mL/min – 1.0 mL/min are usually applied. Furthermore, lower sample amounts are needed for satisfying results. As a consequence of the small particle size, UHPLC systems need to maintain high pressures of 400-1000 bar and are therefore not suitable for standard HPLC systems. In contrast, capillary and normal HPLC with columns equipped with particle sizes between 3 and 5 μm require significantly lower pressure, making them suitable for standard capillary and normal HPLC systems. Both separation techniques use similar column length of 50 – 250 mm for normal HPLC and 30 – 200 mm for capillary HPLC. They mainly differ in their internal diameter (ID); normal HPLC-columns have an ID of 3.0 - 4.6 mm and capillary an ID of 0.3 – 1 mm. Columns with smaller ID's, produce taller peaks, thus providing better detection limits for MS. However due to the reduced internal diameter, capillary HPLC requires significantly lower flow rates between 5 – 50 $\mu\text{L}/\text{min}$ in comparison to normal HPLC (100 – 1000 $\mu\text{L}/\text{min}$). To achieve such low flow rates on standard HPLC systems splitter techniques are necessary. In the case of normal standard HPLC systems can be used without further splitting of the flow rate.

Table II-1: Normal HPLC, capillary HPLC and ultra HPLC systems applied in metabolomic research. Displayed are the range of the internal diameter (ID), the length and the particle size of the column as well as the pressure and the applied flow rates.¹⁴⁴

	ID (mm)	Length (mm)	Particle size (μm)	Pressure (bar)	Flow rate ($\mu\text{L}/\text{min}$)
Normal HPLC	3.0 - 4.6	50 - 250	3-5	< 400	200 – 1000
Capillary HPLC	0.3 - 1	30 - 150	3-5	< 400	5 - 50
Ultra HPLC	1.5 - 4.6	30 - 200	<2	400 - 1500	200 – 1000

In LC-MS-based metabolite analysis, mass spectrometry plays a the key role because of its outstanding sensitivity and its developing potential in the coming years. For MS analysis, metabolites first need to be ionized, in order to be injected into the MS system. The most common ion source is the electrospray ionization (ESI).^{166, 167} This is a soft method of ionization in which a high electric field produces charged droplets from a liquid solution, ultimately leading to the formation of gaseous ions. These ions are then transferred into the MS and are further analyzed. ESI can be performed in positive or negative ionization modes. Advantages of ESI include unnecessary for previously mentioned derivatization steps and the ability to ionize molecules with higher masses. Furthermore non-volatile as well as polar substances can be measured. ESI has proven to be an excellent quantitative ionization method exhibiting good sensitivity properties; however a drawback of using ESI is the ion suppression effect, which can occur by simultaneous ionization of a large number of metabolites. This effect can significantly reduce MS sensitivity, but, with the help of an adequate separation technique, ion suppression can be reduced to a minimum. Further ionization methods include atmospheric pressure ionization (APCI) and atmospheric pressure photoionization (APPI).¹⁴⁴ By comparison to ESI, APCI ionization is a harsher ionization method, often resulting in unwanted fragmentation of the precursor ions. This technique favors the ionization of highly hydrophobic small molecules that are lacking polar groups and is thus less widely used in the field of metabolomics.

Upon ionization metabolites are transported to the analyzer of the mass spectrometer, comprising one or more of the following basic instruments: quadrupoles (Q), ion traps (IT), orbitraps, time-of-flight (TOF) and Fourier transform ion cyclotron resonance devices (FT-ICR, Table II-2). Quadrupole analyzers, are frequently used in MS analysis because of their robustness, high linear dynamic range of 1:10,000 and mass range of up to 4,000 m/z. Disadvantages include their reduced scan speed of 1 sec and comparably low sensitivity. Another analyzer is the TOF, which has fast scanning properties (μ sec), a wide mass range up to 500,000 m/z, a resolution greater than 20,000 and a mass accuracy in between 2-10 ppm. The ion trap is very common and widely used due to its scan speed of 20-200 msec and its ability to perform tandem MS. Drawbacks include low resolution of 200-20,000 and thus a mass accuracy of only 100-1,000 ppm, which makes it unsuitable for global metabolomics, but nevertheless it can be applied for structural, targeted metabolite investigations. In regard to this, Orbitraps as well as FT-ICRs are analyzers with outstanding resolution capabilities greater than 100,000, and leading to mass accuracies smaller than 2 ppm and making them ideal for untargeted metabolite profiling

studies. Drawback of the orbitrap is the low m/z -range of only 2,000 and of the FT-ICR the scan speed of 1 sec needed for a resolution of 1,000,000.

Table II-2: MS analyzers: quadrupoles (Q), time-of-flight devices (TOF), ion traps (IT), orbitraps and Fourier transform ion cyclotron resonance devices (FT-ICR). Listed below are the mass accuracy, resolution, m/z -range, scan speed and the dynamic range of the analyzers.

Analyzer	Q	TOF	IT	orbitrap	FT-ICR
Mass accuracy	100-1,000 ppm	2-10 ppm	100-1,000 ppm	1-2 ppm	1-2 ppm
Resolution	200-2,000	>20,000	200-20,000	>100,000	1,000,000
m/z -range	4,000	>500,000	4,000	2,000	4,000
Scan speed	1 sec	μ sec	20-200 msec	20-200 msec	1 sec
Dynamic range	1:10,000	1:1,000	1:1,000	1:5,000	1:4,000

Significant improvements in the MS technology have been achieved by combining two or more analyzers within one MS system. These hybrid MS systems are able to overcome drawbacks of previous mentioned standard analyzers and are thus interesting for the global analysis of metabolites. Hybrid MS analyzers include Q-TOF, linear ion trap-orbitrap (LTQ-Orbitrap) and LTQ-Fourier transform-ion cyclotron resonance (FT-ICR). Among these, LTQ-FT-ICR is outstanding due to its resolution of up to 1,000,000 and mass accuracy of less than 2 ppm. As it is also able to perform tandem MS, it is highly recommended for metabolomics studies. Herein, the small molecules collide with electrons, resulting in unique fragmentation patterns. Shortcomings over the other two hybrid systems include, the scan speed and the cost intensive maintenance of the system. Also recommended for global metabolite analysis is the Q-TOF, which has a resolution of 20,000 and mass accuracy less than 5 ppm. When combining the TOF with the quadrupole, tandem MS properties as well as the excellent stability and robustness of the quadrupole can be transferred. The LTQ-Orbitrap is also highly recommended for metabolome analysis, combining a high resolution of greater than 100,000 with a mass accuracy within 2 ppm and a m/z range up to 6,000. In addition, maintenance is less costly in comparison to the LTQ-FT-ICR. Nevertheless, the reduced mass accuracy over longer time periods makes it less favorable compared to other systems.

In summary, the global analysis of untargeted metabolomics is currently the most challenging part in the field of metabolomics. Therefore, sophisticated analytical platforms were established in recent years, including ultra HPLC and hybrid MS systems, which dramatically improved the performance. Despite these technical improvements, broad metabolite identification still remains difficult due to the natural diversity of the

metabolites. Therefore, two dimensional LC-MS methods, which are able to analyze hydrophilic as well as hydrophobic metabolites within the same run, need to be further improved. Additionally database-supported metabolite identification needs to be elaborated to facilitate LC-MS standardization and guarantee inter-laboratory reproducibility.

1.5 Sample preparation for the determination of endogenous global metabolite levels in bacteria

The sample preparation in global metabolite profiling experiments is a central aspect which directly impacts the experimental outcome.^{6, 7, 168} Nevertheless it remains highly underestimated and only very few in-depth studies are available. First of all, sample preparation depends on the origin of the biological sample e.g. biofluids, tissues, mammalian cells or bacteria. Second, when working with cells, the analysis of endogenous or exogenous metabolites needs to be discriminated. Additionally, global metabolome studies covering highly hydrophilic as well as highly hydrophobic metabolites are not available to date. Therefore, covering a broad polarity range is not only limited by the analytical platforms, but also by sample preparation including sample extraction. As a result, most metabolite studies have focused on either polar or nonpolar metabolites, depending on the application. In this chapter, the sample preparation of polar endogenous metabolites in bacteria will be discussed. Herein, the workflow can be divided into the following important steps: (1) harvesting and washing, (2) quenching of cellular metabolism, (3) cell lysis, (4) extraction of metabolites, (5) concentration and (6) removal of unwanted macromolecules (Figure II-4).

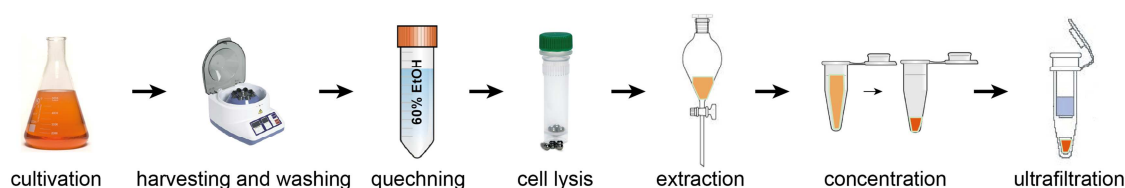


Figure II-4: Sample preparation workflow for the analysis of polar endogenous metabolites including: sampling and washing of the bacteria, quenching cellular metabolism, cell lysis, extraction of metabolites, concentration and removal of unwanted macromolecules.

The first step of the sample preparation workflow is the sampling of intracellular metabolites from bacteria. Commonly applied techniques involve the centrifugation and filtration of bacterial suspensions and have been analyzed in depth by Meyer *et al.* for *Bacillus subtilis* (*B. subtilis*) and by Liebeke *et al.* for *Staphylococcus aureus* (*S.*

aureus).^{169, 170} In this filtration strategy, the bacterial suspension is applied to a 0.45 µm pore size sterile filter which retains the bacteria while allowing the medium to pass through. The significant advantage of this method, is that it is short in time, effective, and additional washing steps can easily be added in order to remove media-related compounds. Contrastingly, the centrifugation strategy requires more time for additional washing. Nevertheless, both harvesting methods are described in literature for global metabolome studies in bacteria. After sampling, washing of the bacteria is important to remove media-related metabolites. In this process, events of intracellular metabolites need to be taken into account. These can be decreased throughout washing by applying a cold (4 °C, isotonic to the cultivation medium) 0.6% NaCl solution as well as phosphate buffered saline (PBS) solutions.

The aim of metabolomics experiments is to obtain a snapshot of the metabolite levels at the moment of harvest. Therefore, the sampling protocol needs to be as fast as possible and the metabolism of the bacteria needs to be arrested to avoid enzymatic degradation processes.^{7, 168} This so-called quenching step can be achieved using cold methanol/ethanol or other organic solvents, which instantly stop the metabolism.^{7, 168, 171} The critical disadvantage of this quenching step is a significant loss of intracellular metabolites, which can occur due to cell lysis in the presence of the organic solvent.^{7, 172, 173} Therefore another established method uses liquid nitrogen (-196 °C) for quenching before or after the separation of cells from the growth medium.^{6, 7, 171} As the given situation assumes a separation of intra- and extracellular metabolites, it is crucial to avoid the loss of significant amounts of intracellular metabolites. Thus, it is advisable to first separate the cells from the medium and then immediately arrest the metabolism.

The cells are subsequently lysed, which can be achieved by various standard methods such as ultrasonication, glass beads, bead mill and the use of organic solvents. Effective and quantitative cell disruption is highly dependent on the type of bacteria used. In general, for Gram-positive bacteria such as *S. aureus*, mechanical cell disruption is necessary since the cell wall will not break up sufficiently using only organic solution.¹⁷¹ Therefore, the glass bead and ultrasonication methods are recommended in literature. During lysis, the heat released as a result of the mechanical disruption must be considered. Therefore, heat-induced degradation processes should be minimized by performing cooling steps in between the disruption cycles.

The fifth step is the extraction of the metabolites, which is one of the critical points in the sample preparation protocol. Requirements for the extraction solvents are that they do not chemically or physically modify the extracted metabolites and that they dissolve a broad

range of metabolites for global studies. For the extraction of polar metabolites, which encompass the majority of endogenous metabolites, several extraction solvents have been tested including 60% methanol^{174, 175}, 60% ethanol¹⁷⁶, 60% methanol with formic acid¹⁷¹, mixtures of chloroform/water/ethanol^{171, 175} as well as water¹⁷⁷. In addition to the composition of the solvents, the influence of temperature on the extraction were tested, covering 4 °C to 100 °C. Meyer *et al.* determine that cold 60% ethanol (4°C) is the best extraction solvent, due to the fact that it not only effectively arrests cellular metabolism, thereby avoiding enzymatic metabolite degradation, but also results in the highest recovered amounts of intracellular compounds.¹⁷¹

Before the final LC-MS-based measurements two further essential steps need to be performed. First, the extracted metabolites need to be concentrated to cover the MS sensitivity range. Concentration methods for polar metabolite solutions can be performed via speed vacuum centrifugation or by lyophilization. The freeze drying method is recommended because it helps to prevent degradation processes by excluding air and thereby avoiding oxidation processes. The last step prior to LC-MS injection, is the removal of macromolecules such as proteins, being careful to avoid reducing sensitivity and damaging the metabolites in the separation process.¹⁶⁸ Usually protein precipitation is performed using organic solvents such as acetone. Due to the reduced solubility of highly polar metabolites in this organic solvent, this method is not recommended for the analysis of polar metabolites. A second method to remove proteins that is easy to handle is using size exclusion filters, which are available in different sizes e.g. 3 kDa, 10 kDa and 30 kDa.^{178, 179, 180} Macromolecules are retained via ultrafiltration and small molecule metabolites are passed through the membrane.

In summary, the sample preparation for the global analysis of polar endogenous metabolites is challenging, but of course essential in metabolomics studies. Due to the natural instability of a great number of metabolites, the sample preparation should be as fast as possible including a minimal number of steps. Nevertheless, aspects such as residual enzymatic activities, which can impact the experimental outcome, as well as the right extraction solvent are of critical importance in the sample preparation protocol and influence the complete metabolomics workflow. In this regard, the metabolomics research field needs to invest more efforts by addressing each step in the sample preparation in order to fully be able to distinguish between artificial and biological changes in the metabolite levels.

1.6 Data processing and data analysis of LC-MS-based metabolomics experiments

Basic metabolomics experiments are in general performed in a comparative mode, in which the effect of one change is investigated. For statistical purposes, at least three independent biological replicate measurements are necessary.¹⁸¹ Especially in metabolomics experiments more replicates are recommended. According to the number of replicates and to the length of LC-MS-based gradients, large, complex data sets are generated, which are difficult to process and interpret. In general, the data handling can be divided into data processing and data analysis.^{182, 183} Both steps are crucial for reliable metabolite identification and quantification, especially to distinguish between interesting biological variances from obscuring sources of variabilities.

In LC-MS-based global profiling strategies, the information generated is accumulated in raw files which represent a collection of histograms.¹⁸⁴ Each histogram represents detected ions over a small time frame. It consists of a number of m/z -values and intensity data points. In order to obtain retention times and intensities of the detected m/z -values in the form of a matrix, these histograms are processed. A typical data processing pipeline is: filtering, feature detection, alignment and normalization.¹⁸² Filtering methods process the raw measurement signal with the aim of removing effects like measurement noise or baseline. Feature detection is used to detect representations of measured ions from the raw signal. Alignment methods cluster measurements across different samples and further normalization removes unwanted systematic variation between samples. To overcome these nontrivial data processing steps, free and commercially-available software such as e.g. MZmine, Xalign, XCMS and Sieve (Thermo Fisher Scientific), MarkerLynx (Waters), BlueFuse (BlueGnome) have been developed. Differences between commercial and free available software include the transparency of the implanted data processing algorithm. Free available software is released as open source software giving details on applied algorithms, which can be manipulated manually. In contrast, most commercially available software do not give insights into data processing steps, with exception of the Software Sieve (Thermo Fisher Scientific, Waltham, MA, USA), which uses ChromAlign for chromatographic alignment. Advantages of commercially available software are the user-friendly interface, allowing high-throughput data processing and data statistics without profound knowledge in biostatistics.

Beside data processing, data analysis is a crucial aspect for the interpretation of metabolomics data sets. Starting from an extracted matrix, which includes information about retention times, intensities as well as the peak areas of the corresponding m/z -

values, univariate or multivariate statistics can be performed.¹⁸⁵ Univariate statistics are the simplest form of statistical analysis, in which only one variable is involved. In contrast, multivariate statistical methods involve the analysis of more than one statistical variable at a time and are therefore more complex. For the analysis of global metabolite profiling experiments, univariate analysis methods are sufficient, including the two-sample *t*-test, which highlights differences between two populations.¹⁸⁶ From this, a *p*-value is determined, which is a function of the observed results that is used for testing a statistical hypothesis. Before the test is performed, a threshold value is chosen, called significance level (α) of the test, which is traditionally 5% or 1%. If the *p*-value is equal to or smaller than the significance level, it suggests that the observed data are inconsistent with the assumption that the null hypothesis is true and that hypothesis must thus be rejected.

A very common univariate visualization method to easily identify changes between two populations is the volcano plot (Figure II-5).¹⁸⁷ Therefore, the two-sample *t*-test is performed using the peak areas of the corresponding *m/z*-values. In the volcano plot, the \log_2 value of the ratio between the two populations is determined and plotted against the $-\log_{10}$ of the *p*-value. *M/z*-values that correlate between the two populations appear in the center of the graph, whereas changing *m/z*-values are shifted to the left (downregulated) or right side (upregulated) of the plot. Furthermore, the definition for significantly or highly significantly changing *m/z*-values can be described by equations I-IV, including the standard deviation among the complete data set (σ_{total}) and the significance level (α), determined within the two-sample *t*-test.

$$y = \frac{1}{(x - \sigma_{total})} + \alpha \quad (I)$$

$$y = \frac{2}{(x - 2 \cdot \sigma_{total})} + \alpha \quad (II)$$

$$y = -\frac{1}{(x + \sigma_{total})} + \alpha \quad (III)$$

$$y = -\frac{2}{(x + 2 \cdot \sigma_{total})} + \alpha \quad (IV)$$

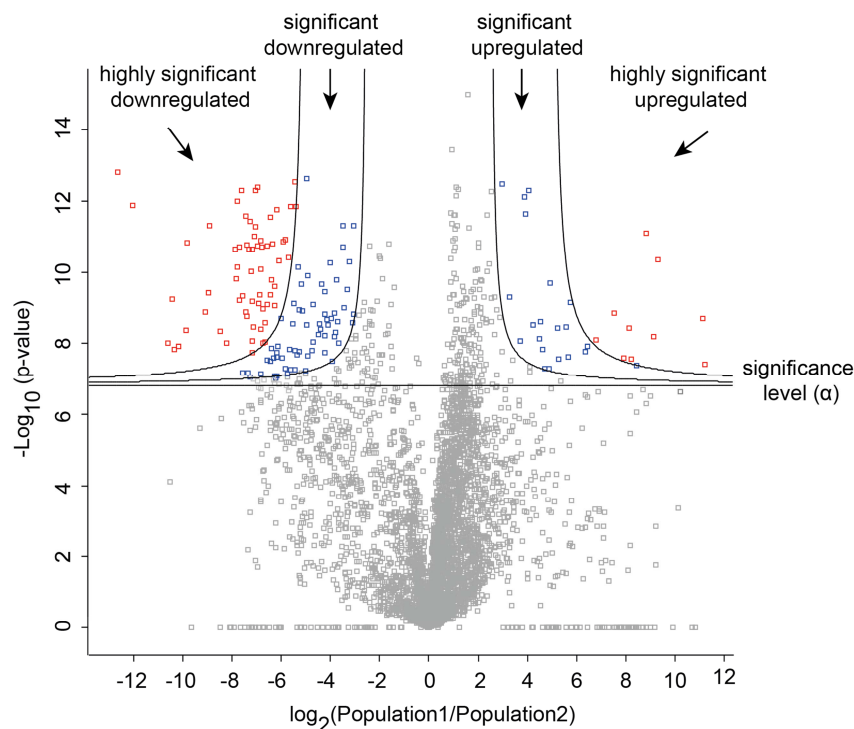


Figure II-5: The volcano plot is a representation of the univariate analysis, the two sample t-test of an untargeted metabolomics experiment to identify up- and downregulated m/z -values within two populations. In the volcano plot the \log_2 value of the ratio of the two populations is plotted against the $-\log_{10}$ p-value. With the equations I-IV, significant (blue) and highly significant (red) changes in m/z -values can be identified.

In summary, in recent years, a dramatic need for analytical methods and software tools has been observed. Therefore, open-source as well as commercially available software has been developed. These software must provide a user-friendly interface in order to allow data processing and data analysis without in-depth knowledge of biostatistics. Additionally, they need to provide high-throughput analysis and visualization tools to help identify specific metabolite changes. Furthermore, they should connect to online databases for metabolite identification. With regards to this, the currently-available software are still in its infancy and requires further optimization. Reliable metabolite analysis can be a starting point for the identification of significantly changing metabolites, thereby improving an understanding of biochemical networks and metabolite fluxes under specific conditions.

2. Scope of this work

(i) The aim of the first project was to establish a targeted and untargeted metabolite profiling platform for the analysis of endogenous polar metabolites within bacteria. Therefore a sample preparation protocol and LC-MS-based analytical platform were investigated. Furthermore an in-house database for defined metabolite identification was developed using tandem MS. As proof-of-principle and to evaluate the sample preparation protocol as well as the analytical platform, wild-type *Staphylococcus aureus* (*S. aureus*) NCTC 8325 extracts were applied.

Author contributions

S.A. Sieber conceived and supervised the project. M.F. Koch performed all experiments including sample preparation, LC-MS experiments and data processing.

(ii) In the second project, the aim was to investigate the role of the ribokinase SaThiD within the pyridoxal salvage pathway. Therefore a targeted metabolite profiling platform for the analysis of extracellular pyridoxal levels of wild-type *S. aureus* USA 300 and two transposon mutants (TnPdxS and SaTnThiD) was generated including a sample preparation protocol and analytical platform itself.

Author contributions

S.A. Sieber and M.B. Nodwell conceived and supervised the project. M.F. Koch performed metabolomics measurements including LC-MS experiments and data processing for relative quantification of pyridoxal in *S. aureus* extracts. MIC measurements, DNA cloning, protein overexpression and purification as well as enzymatic activity assays were performed by M.B. Nodwell. F. Alte and S. Schneider performed crystallization experiments. S. Schneider solved the crystal structure.

(iii) The third project involved the investigation and evaluation of a new natural product discovery strategy using LC and LC-MS-based analytical methods. In the first step, the most promising capture probe was identified. Therefore, previously synthesized capture probes were screened according to their reactivity against a library of small electrophilic compounds. In the second step, the most promising identified capture probe was used in a proof-of-principle study, in order to understand its reactivity and selectivity in complex extracts of *Streptomyces* strains.

Author contributions

S.A. Sieber conceived and supervised the project. M.F. Koch performed reactivity studies with **Cap6** including LC and LC-MS experiments with the artificial compound library as well as the complex extracts of *Streptomyces strains*. **Cap** library was developed, synthesized and initially screened for its reactivity by G. Rudolf.

3. Results and discussion

3.1 Establishment of a targeted and untargeted LC-MS-based metabolite profiling platform for the analysis of metabolite levels in bacteria

3.1.1 Introduction

The aim of this work was to establish an LC-MS-based metabolomics platform for the analysis of endogenous metabolite levels in bacteria. Therefore, it is important to address a broad range of metabolites and not only highly abundant metabolites but also metabolites at the limit of detection. In light of the naturally occurring diversity of metabolites and the differences in their physico-chemical properties, this profiling platform focuses on polar small molecule metabolites e.g. nucleotides, amino acids, sugars and cofactors within a range of 80 -1000 Da.

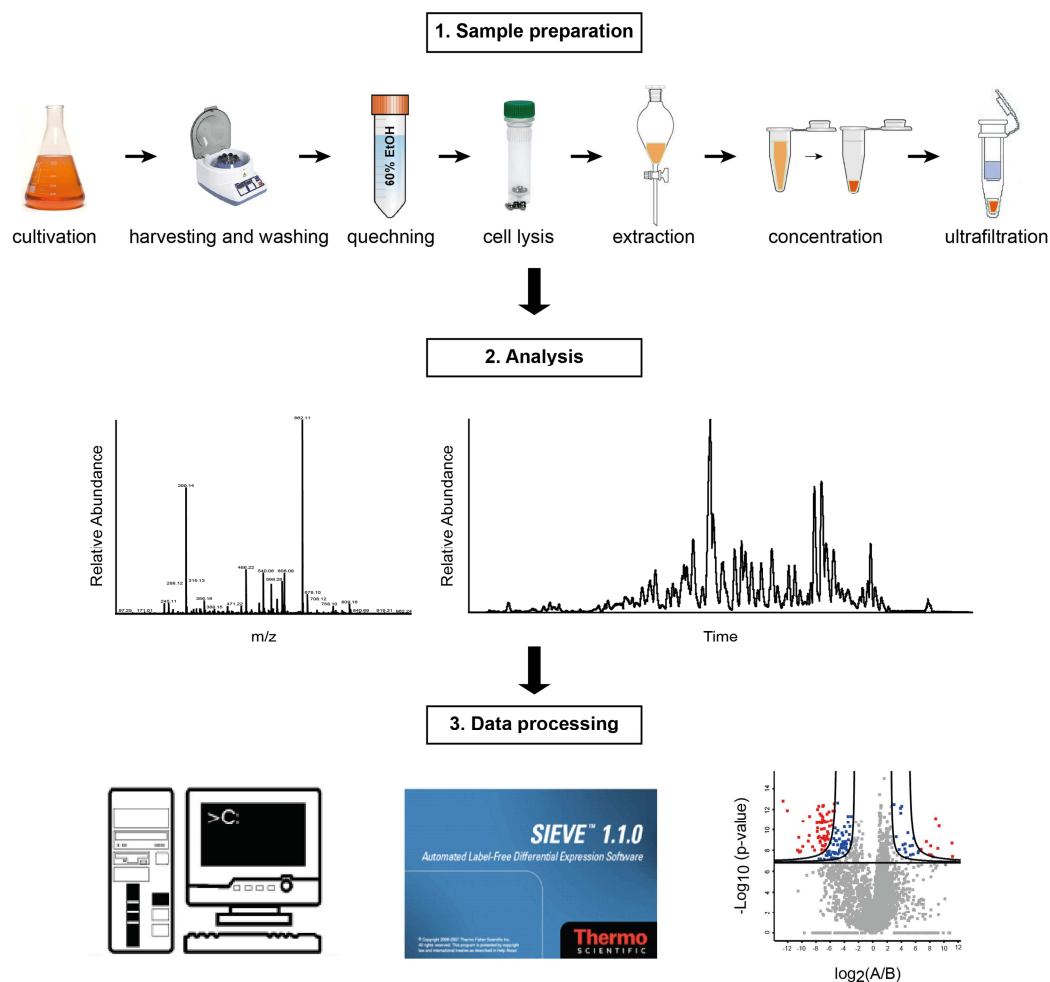


Figure II-6: Workflow of a metabolite profiling experiment including sample preparation, analysis and data processing.

In general, the workflow of a metabolite profiling experiment can be divided into three crucial steps: 1) sample preparation; 2) analysis; 3) data processing (Figure II-6). In this workflow, the reproducibility, accuracy and robustness of the method are important factors, which will be reviewed carefully within the following chapters. Furthermore, metabolite identification via databases is currently still in its infancies. This is why an in-house database was introduced for distinct metabolite identification using tandem MS. As a proof-of-principle and to evaluate the established protocol as well as the database itself, *S. aureus* NCTC 8325 extracts were analyzed to assess the accuracy and reproducibility of the method.

3.1.2 Important aspects of sample preparation for the analysis of endogenous polar metabolites in bacteria

In the metabolomics experiments, sample preparation is the first step within the workflow and therefore a crucial aspect due to its direct influence on the experimental outcome. The general goal of the workflow is to obtain a broad overview of endogenous, polar metabolites and their corresponding concentration levels. This needs to be taken into consideration when designing a sample preparation protocol: (i) harvesting and washing of bacteria; (ii) quenching cellular metabolisms; (iii) cell lysis; (iv) metabolite extraction, (v) concentration and (vi) removal of unwanted macromolecules before LC-MS analysis (Figure II-7). As the aim of the experiment was to reflect the status quo of the cell at a defined state, the preparation protocol must be fast and comprise a minimal number of steps. Additionally, natural degradation of the metabolites as well as degradation due to residual enzymatic activity needs to be considered and minimized. Therefore, a gentle but also effective sample preparation protocol covering a broad range of metabolites needed to be established. In the workflow itself the enzymatic activity should be instantly stopped and natural degradation processes should be reduced to a minimum.

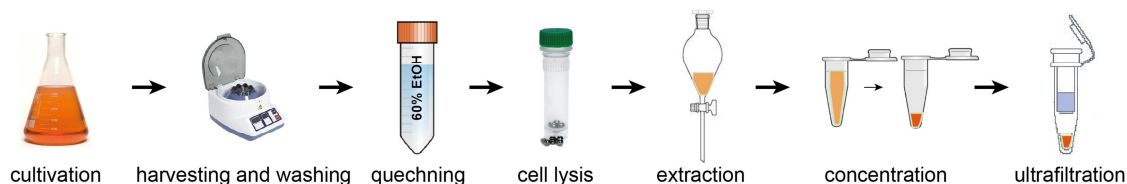


Figure II-7: Sample preparation workflow of a metabolite profiling experiment for the analysis of polar endogenous metabolites in bacteria. The workflow includes the following step: cultivation of the bacteria, harvesting and washing, quenching the metabolism, cell lysis, metabolite extraction, concentration and ultrafiltration.

To begin the protocol, bacteria must first be harvested to separate endogenous from exogenous metabolites and subsequently washed to remove unwanted components from the growth media. Common strategies for this include centrifugation^{171, 188} or filtration.^{171, 189} The filtration technique uses 0.45 µm pore size sterile filters, which retain bacteria while allowing the media to pass through, thereby rapidly separating the from the growth media. Advantages of this method are fast sampling and high-throughput potential. A significant disadvantage of the filtration technique is the difficult handling, especially the complete removal of the bacteria from the filter, which dramatically affects the reproducibility and accuracy of the experimental outcome. In contrast to the filtration method reproducibility and accuracy are less affected within the centrifugation method. However, the drawback of harvesting the bacteria via centrifugation is a significant increase in processing time, which can cause stress-induced changes in the metabolite levels. These, stress-induced changes can be minimized by cooling during centrifugation. After harvesting the bacteria via centrifugation, unwanted media-derived components need to be removed in a washing step. Therefore, bacteria are redissolved in washing buffers and again separated from the supernatant via centrifugation. Common applied washing buffers are isotonic solutions such as phosphate-buffered saline (PBS) or 0.6% NaCl, which help to minimize osmotic-induced cell lysis and therefore leakage of endogenous metabolites into the extracellular extract.¹⁹⁰ However, due to adduct formation and ion suppression effects, which dramatically influence the sensitivity of the following LC-MS-based procedures, high salt concentrations should be avoided.^{7, 191} Therefore bacteria were always washed with LC-MS grade water as a last step.

After harvesting and washing the bacteria, cellular metabolism processes must be arrested instantly in order to avoid alterations in metabolite levels due to residual enzymatic activities. In the literature, different strategies have been introduced to overcome this problem which use organic solvents such as methanol (MeOH) or ethanol (EtOH), or liquid nitrogen (N₂ aq.). In this protocol, the enzymatic activity in the extract is arrested using 60% EtOH.^{192, 193, 194, 195} The next step in the sample preparation is the lysis of the bacterial cells. Common mechanical cell disruption methods include sonication, homogenizing, bead beating, high temperatures and freezing. In this work, cell lysis is performed using glass beads with the help of a homogenizer rotor. Upon cell lysis metabolite extraction is the next crucial step in global metabolite profiling experiments. In the literature, different solvents and conditions for the extraction of polar metabolites have been tested, including methanol, ethanol, water and chloroform. 60% ethanol (4 °C) is highly recommended due to the fact that it is the best compromise between arresting cellular metabolism and extracting polar metabolites.¹⁷¹

Prior to LC-MS analysis two further essential steps need to be performed. First, the extracted metabolites need to be concentrated to cover the MS sensitivity range. Concentration methods for polar metabolite solutions can be performed via speed vacuum centrifugation or by lyophilization. The freeze drying method is recommended because it helps to prevent degradation processes by excluding air and thereby avoiding oxidation processes. Second, the removal of macromolecules such as, for example, proteins and particles is of essential need because they can dramatically affect the performance of the analytical platform. This is because of the high percentages of organic solvent, which can be found at the beginning of the LC-MS gradient, and can lead to protein agglomeration causing pressure instability and blockage of the applied column. Through the use of 3 kDa ultrafiltration devices, particles and macromolecules can be separated from small molecule metabolites. This is particularly useful for the extraction of polar metabolites and frequently used due to its easy-to-handle and high-throughput applicability.

In summary, requirements for a reproducible and reliable sample preparation of a global metabolite profiling experiment are highly demanding and need to be carefully reviewed. Not only the vast diversity of metabolites is a challenging aspect to the protocol, but the rapid degradation of the metabolites. In this work, a protocol to study endogenous polar metabolites was successfully established including the following important steps: (i) harvesting and washing the bacteria via centrifugation; (ii) arresting cellular metabolism using 60% EtOH; (iii) cell lysis using glass beads; (iv) metabolite extraction using 60% EtOH; (v) concentration by lyophilization and (vi) removal of particles and macromolecules by centrifugal ultrafiltration devices.

3.1.3 Establishment of an LC-MS-based analytical platform for the analysis of polar metabolites

In metabolomics, liquid chromatography-mass spectrometry (LC-MS)-based analytical platforms are most frequently used since they can cover a broad range of metabolites and show outstanding sensitivity. Using LC-MS technologies, metabolites are first resolved on the LC and subsequently injected into the MS. Thereby, two dimensions of information are generated wherein the retention times indicate the physico-chemical properties of the metabolites and mass to charge (m/z) values reflect, in case of $m/z = 1$, the molecular weight of the metabolites. During LC, it is important that co-elution and ion suppression effects, which results in reduced sensitivity, are avoided. Therefore, long multi-step gradients are used to obtain satisfying separations even within complex matrices such as metabolomic extracts. In the metabolite profiling experiment the MS instrumentation is one of the key elements and the following aspects are important to consider: (i) high resolution; (ii) fast scanning; (iii) high dynamic range; (iv) accuracy and (v) sensitivity. In addition, the ability to perform tandem mass spectrometry (MS^n) and to thereby generate unique fragmentation patterns is an essential requirement for metabolite identification. In this work, the analytical platform, composed of a Dionex Ultimate 3000 standard liquid chromatography system coupled to a Thermo Finnigan hybrid linear ion trap quadrupole Fourier transform ion cyclotron resonance mass spectrometer (LTQ-FT-ICR-MS), is an ideal equipped MS instrument for metabolomic studies. It combines high resolution of up to 1,000,000 (FT-ICR) with tandem MS properties (LTQ) for distinct metabolite identification.^{144, 196}

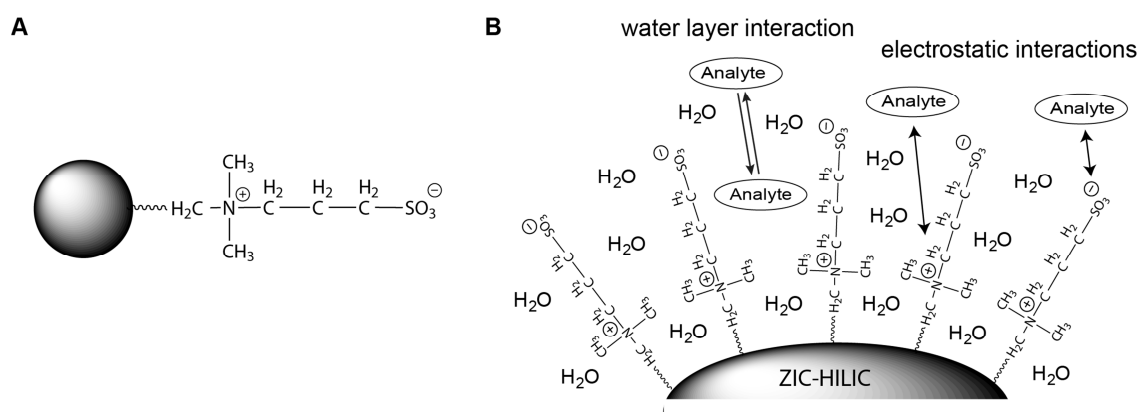
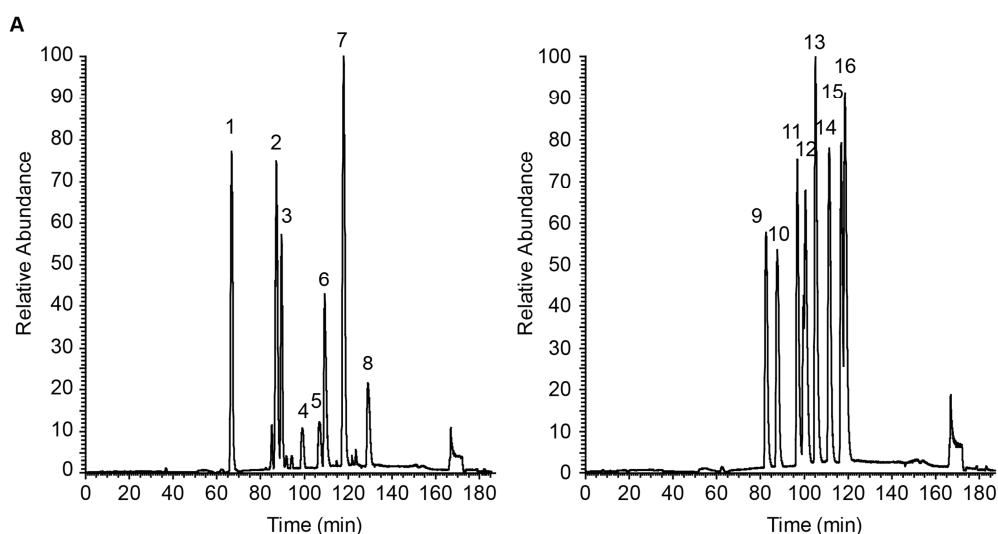


Figure II-8: Column material of the ZIC®-HILIC. **A)** Covalently bound zwitterionic sulfobetain backbone of the stationary phase of the ZIC-HILIC. Sulfobetain is positively charged at the quaternary nitrogen and negatively charged on the sulfonic group. **B)** Interaction possibilities of polar analytes with the stationary phase of the ZIC®-HILIC. Interactions with the water layer as well as electrostatic interactions induced by the charged sulfobetain backbone, are illustrated.¹⁹⁷

For the separation of polar metabolites, a zwitterionic hydrophilic interaction liquid chromatography (ZIC®-HILIC, SeQuant®) was applied.^{162, 165, 198} In this separation technique, zwitterionic sulfobetains, which are positively charged at the quaternary nitrogen and negatively charged on the sulfonic group, are covalently bound to the stationary phase of the column (Figure II-8A). Surrounding the opposite charged stationary phase is a water layer that interacts with polar metabolites. In addition to the water layer, electrostatic effects within the charged sulfobetains impact the separation (Figure II-8B). The recommended eluents for the ZIC®-HILIC are acetonitrile and water mixtures. Additionally, buffer additives like ammonium formate or ammonium acetate are recommended to achieve satisfying separation.

In establishing the analytical platform, the aim was to identify a reproducible LC gradient in which metabolites are retained and separated satisfyingly. Additional aspects that need to be taken into consideration include carry over and reproducibility, as well as adequate washing and equilibration of the ZIC®-HILIC column. In general, LC gradients can be divided into three sections: 1) gradient; 2) washing and 3) equilibration. In the first section, elution of the metabolites must be as distinct as possible and co-elution should be minimized. Important parameters to control this include the flow rate, eluent solvent composition and the slope of the gradient. In the second washing step, highly polar compounds still remaining on the stationary phase are eluted in order to recover the separation ability of the stationary phase. Finally, the column is re-equilibrated to the starting conditions of the gradient for the next run. Especially for HILIC, correct and adequate equilibration is crucial and more time-consuming compared to standard reversed-phase chromatography methods.



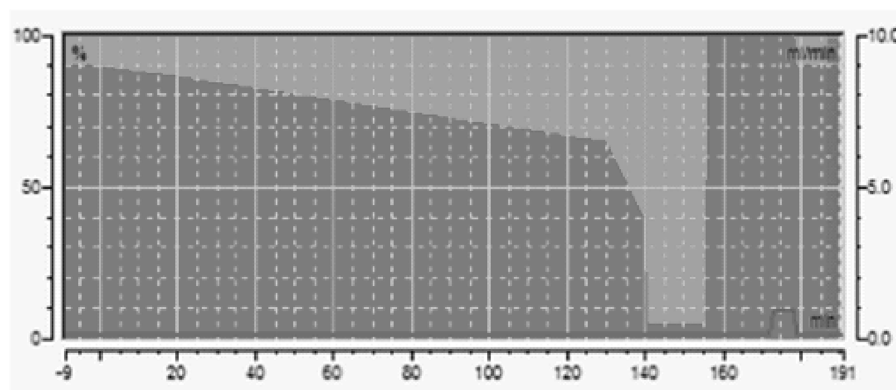
B

Metabolite	Elemental formula	m/z calculated	m/z detected	Deviation (ppm)	Retention time
(4) Fructose-6-phosphate	C ₆ H ₁₂ O ₉ P	259.02134	259.0221	-1.1	100
(5) Glucose-6-phosphate	C ₆ H ₁₂ O ₉ P	259.02134	259.0222	-0.7	107
(10) UMP	C ₉ H ₁₂ N ₂ O ₉ P	323.02749	323.0288	0.6	88
(8) Fructose-1,6-bis-phosphate	C ₆ H ₁₃ O ₁₂ P ₂	338.98768	338.9884	-1.0	129
(9) AMP	C ₁₀ H ₁₃ N ₅ O ₇ P	346.05471	346.0588	0.1	83
(12) UDP	C ₉ H ₁₃ N ₂ O ₁₂ P ₂	402.99382	402.9950	0.1	101
(11) ADP	C ₁₀ H ₁₄ N ₅ O ₁₀ P ₂	426.02104	426.0224	0.6	97
(14) GDP	C ₁₀ H ₁₄ N ₅ O ₁₁ P ₂	442.01596	442.0173	0.6	111
(15) CTP	C ₉ H ₁₅ N ₃ O ₁₄ P ₃	481.97614	481.9779	0.1	117
(6) UTP	C ₉ H ₁₄ N ₂ O ₁₅ P ₃	482.96015	482.9613	0.2	109
(13) ATP	C ₁₀ H ₁₅ N ₅ O ₁₃ P ₃	505.98737	505.9888	0.6	105
(16) GTP	C ₁₀ H ₁₅ N ₅ O ₁₄ P ₃	521.98229	521.9837	0.5	119
(2) NAD ⁺	C ₂₁ H ₂₆ N ₇ O ₁₄ P ₂	662.10075	662.1039	3.1	87
(7) NADP ⁺	C ₂₁ H ₂₇ N ₇ O ₁₇ P ₃	742.06708	742.0701	2.6	118
(3) CoA	C ₂₁ H ₃₅ N ₇ O ₁₆ P ₃ S	766.10683	766.1085	0.7	90
(1) FAD	C ₂₇ H ₃₂ N ₉ O ₁₅ P ₂	784.14876	784.1530	4	67

Figure II-9: A) Elution of standard metabolites using the established multi-step gradient. Shown is the total ion count of the full scan events (60 - 1000 Da) measured in the negative mode. **B)** Table showing elemental formula, m/z calculated, m/z detected, deviation of the detected to the exact mass and retention time of the standard metabolites used.

The analytical platform presented here was established for 16 polar metabolites including nucleotides, amino acids, cofactors and sugars (Figure II-9). A multi-step gradient with an overall time of 3 h and 16 min depicted in Figure II-10 was designed. Acetonitrile/water mixtures (95/5 → A, 5/95 → B) containing 10 mM ammonium formate at neutral pH were used as eluent solvents. MS detection was performed using an FT-ICR analyzer in the full scan mode with a resolution of 25,000. For the precursor ions, mass deviations less than 5 ppm were detected, which highlights the power of the FT-ICR analyzer (Figure II-9B).

A



B

Step	Time [min]	Flow rate [mL/min]	Eluent A [%]	Eluent B [%]	Section
1	-9	0.1	10	90	Equilibration
2	0	0.1	10	90	Gradient
3	130	0.1	35	65	Gradient
4	140	0.1	60	40	Gradient
5	141	0.1	95	5	Washing
6	155	0.1	95	5	Washing
7	156	0.1	0	100	Washing
8	172	0.1	0	100	Washing
9	173	0.9	0	100	Washing
10	178	0.9	0	100	Washing
11	179	0.1	10	90	Equilibration
12	187	0.1	10	90	Equilibration

Figure II-10: A) Established multi-step gradient for the analysis of polar metabolites with the ZIC®-HILIC. **B)** Steps of the established multi-step gradient with the corresponding time, flow rate, eluent percentage and section of the gradient.

3.1.4 Development of an in-house database for metabolite identification in complex biological matrices

The aim of metabolomics experiments is to depict a complete set of small molecule metabolites at a certain cell state. To date, the identification of metabolites remains a limiting factor in the field of metabolomics. A typical metabolite identification workflow is highlighted in Figure II-11. The metabolite is first extracted and isolated, then measured using the analytical platform and the generated data including the m/z -values are further screened against databases of known metabolites. For accurate identification, tandem MS experiments to generate unique fragmentation patterns of the metabolite, need to be performed additionally.¹⁴⁴

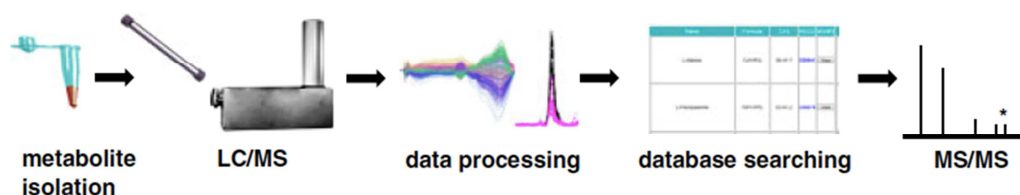


Figure II-11: Workflow of a metabolite identification experiment including: metabolite isolation, LC-MS analysis, data processing, database searching and tandem MS.

Tandem MS can be performed via collision-induced dissociation (CID), in which fragmentation of molecular ions is induced in the gas phase. In this method, molecular ions are usually accelerated to high kinetic energy by an electrical potential and then allowed to collide with neutral molecules (often helium, nitrogen or argon).¹⁹⁷ In this collision, some of the kinetic energy is converted into internal energy, which causes bonds to break and fragmentation of the molecular ion. These fragment ions can then be analyzed via tandem mass spectrometry. Generated fragmentation patterns are unique for each compound and are therefore a tool for definite metabolite identification. Nevertheless, the fragmentation pattern depends on the applied CID energy as well as the applied ionization source and MS instrument. This highlights the difficulties with which metabolite databases are faced and explains the common usage of in-house databases. Standard metabolites are processed by performing additional tandem MS, which adds a third information dimension and thereby enables a distinct metabolite identification.

For the given reason, an in-house database was established in this work. To do so, the data-dependent “ n th order triple play” method was used (Figure II-12). This method consists of three scanning events wherein the first scanning event is a full scan to detect ions with m/z -values in the range of 80 – 1000 Da. In the second scanning event, selected ion monitoring of a previously-specified list of m/z -values and retention times is

completed. The third scanning event is dependent on the data collected in the previous step and consists of a CID fragmentation performed on the m/z -values specified in the list. The first and the second scanning events are performed in the FT-ICR analyzer and the CID fragmentation is induced and detected in the LTQ.

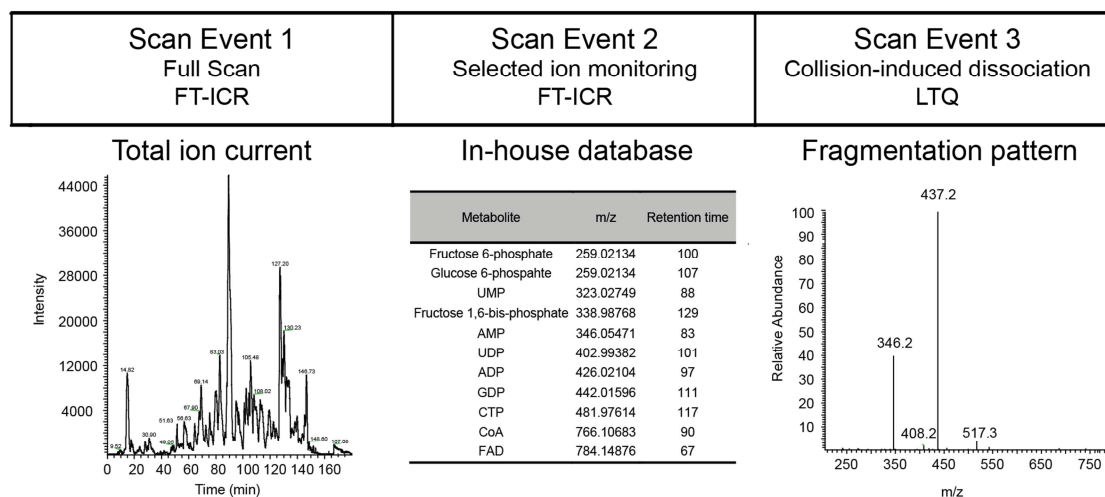


Figure II-12: *nth* order triple play method. The *nth* order triple play method can be divided into three scanning events. In the scan event 1 a full scan is performed with the FT-ICR analyzer using high-resolution. In the second scanning event, selected ion monitoring scans are performed depending on the m/z -values and the retention time of the standard metabolites within the in-house database. In the third scanning event, a data-dependent collision-induced dissociation fragmentation is performed. Thereby, fragmentation is only done for m/z -values specified within the database.

Using this method, an in-house database composed of 129 metabolites was built up, containing information about retention times, m/z -values and fragmentation patterns of each metabolite (Table II-4, Table II-A1 and II-A2). Measurements were performed in positive as well as negative modes. To avoid co-elution and resulting ion suppression effects, a broad metabolite distribution over the length of the gradient was created, as shown in Figure II-13A. Additional CID induced fragmentation was performed, and highlighted in Figure 13-B and C of the two metabolites FAD and AMP. For both metabolites four daughter ions were detected and validated by comparison to tandem MS spectra from the human metabolom database (HMDB).

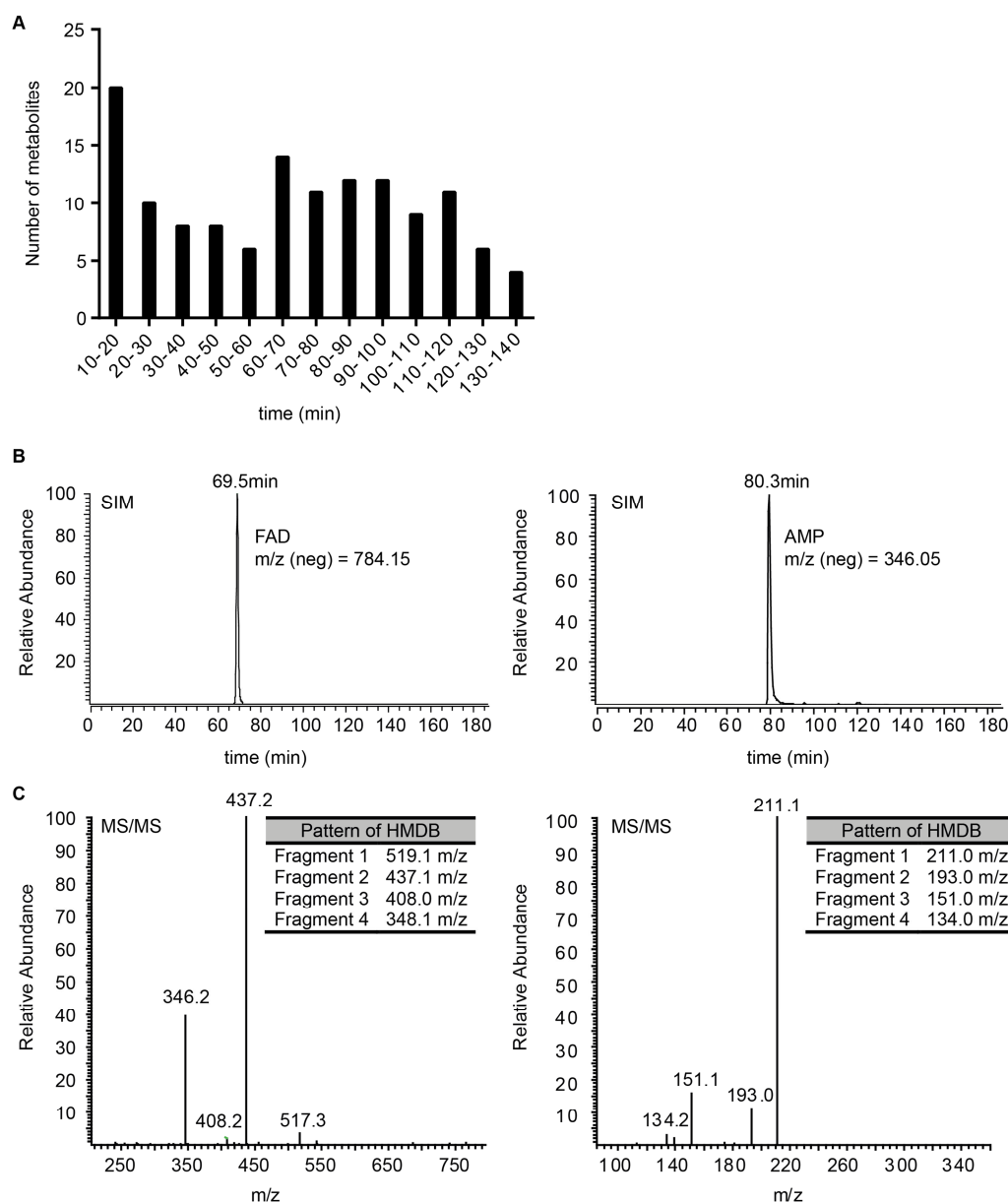


Figure II-13: **A)** Distribution of the metabolites introduced in the in-house database over the length of the established multi-step gradient. **B)** Selected ion monitoring (SIM) scan events of FAD and AMP within the established *n*th order triple play method. **C)** Data-dependent CID induced fragmentation patterns for FAD and AMP. In addition, daughter ions specified according to the HMDB measured on a LC-ESI-ITFT (LTQ Orbitrap XL, Thermo Scientific) in negative mode are shown.

3.1.5 Evaluation of the established metabolite profiling platform in wild-type *S. aureus* NCTC 8325 as a proof-of-principle

After having established the sample preparation protocol and the analytical platform for targeted and untargeted metabolite profiling, the aim was to evaluate our methodology in wild-type *S. aureus* NCTC 8325. The focus of the project was on the reproducibility of the sample preparation, the technical reproducibility of the analytical platform as well as the reproducibility of the experimental design. Furthermore the established in-house database was applied and evaluated for metabolite identification.

First, the reproducibility of the detected metabolite levels was evaluated by harvesting *S. aureus* cells grown to mid-log phase (Figure II-14A and B). By performing three biologically independent replicates on two different days the intra- and the interday reproducibility were evaluated (Figure II-14C). Additionally, to gauge the technical reproducibility, each biological replicate was measured in technical replicates and *L*-glutamine- d_5 (*L*-Gln- d_5) and *L*-glutamate (*L*-Glu- d_5) were spiked in to each sample as internal standards. Bacteria were grown to mid-log phase (OD_{600} (day 1) = 0.35 and OD_{600} (day 2) = 0.5, Figure II-14B), harvested and then prepared for LC-MS measurements (negative mode) as described in chapter II-3.1.2.

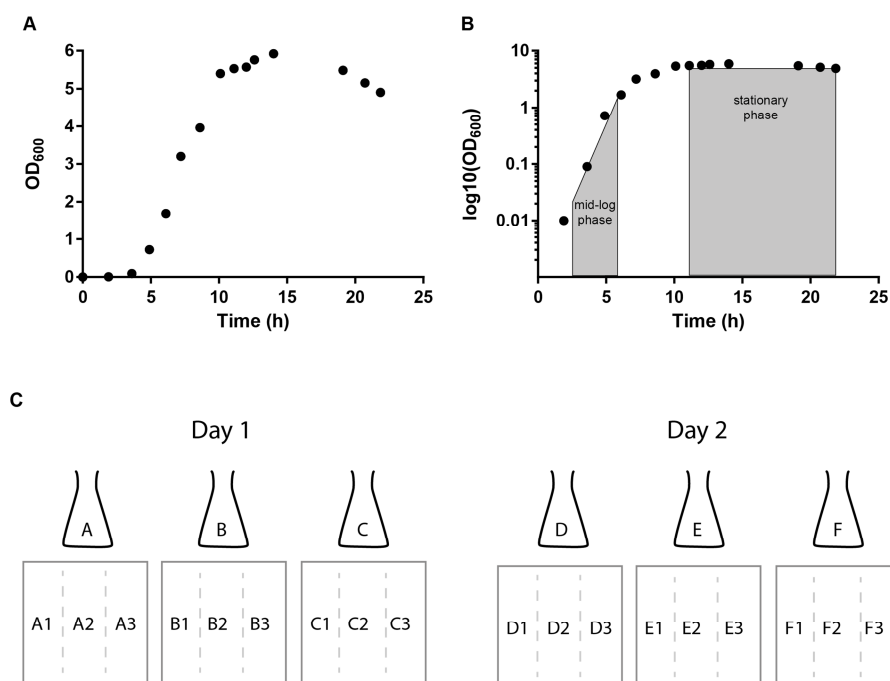


Figure II-14: **A)** Growth curve of wild-type *S. aureus* NCTC 8325 strain over 24 h incubation. **B)** Logarithmic representation of the growth curve to identify the mid-log phase (4 h, OD_{600} 0.3 -0.5) and stationary phase (10 h, OD_{600} > 5.4). **C)** Experimental design of targeted and untargeted metabolite profiling in wild-type *S. aureus* NCTC 8325. Three biological replicates were performed on two different days. For each biological replicate three technical replicates were carried out.

In the first step, the reproducibility and stability of the LC-MS-based platform were evaluated. Therefore, the peak areas of *L*-Gln- d_5 and *L*-Glu- d_5 were compared within the technical and biological replicates. Concerning the technical reproducibility, a deviation of 4% in the case of *L*-Gln- d_5 and 5% in the case of *L*-Glu- d_5 could be observed. Within the biological replicates, (A-F) a deviation of 10% and 15% could be identified. Explanations for the deviations include the time in between each technical replicate measurement (20 h) since long, multi-step gradient were necessary. Furthermore, the different sample matrices that can be found in biological replicates cause variability. Although samples were treated in the same way, slight differences concerning the biological matrices inevitably occur, thus influencing the ionization and also the quantification of the detected metabolites. Nevertheless, the deviations of the internal standards were in an acceptable range, demonstrating good accuracy and reproducibility of the LC-MS-based analytical platform. In the next step, the technical reproducibility as well as the inter- and intraday reproducibility of the biological replicates were evaluated for the complete set of *m/z*-values extracted using the data processing software Sieve2.0 (Thermo Fisher). A scatter plot is commonly used to represent data and address the reproducibility of large data sets.¹⁸⁷ In scatter plot, the extracted intensities of the corresponding *m/z*-values of two different samples are plotted and evaluated using the Pearson correlation coefficient (ρ), which is 0 in the case of no correlation and 1 for perfect correlation of the data sets. Plotting the data according to the described method, the technical reproducibility of the experiment revealed an averaged coefficient of $\rho = 0.92 \pm 0.03$, indicating very good reproducibility (Figure II-15A and B). The intraday reproducibility, meaning biological replicates prepared and measured within the same day, still showed an averaged coefficient of $\rho = 0.85 \pm 0.03$, demonstrating the robustness of the method (Figure II-15C). These results not only confirm the stability of the analytical platform but also highlight the robustness of the experimental design including intraday growth and sample preparation. In contrast, the interday reproducibility, comparing biological replicates between different days, revealed a significant drop in correlation with a coefficient of $\rho = 0.67 \pm 0.06$ (Figure II-15D).

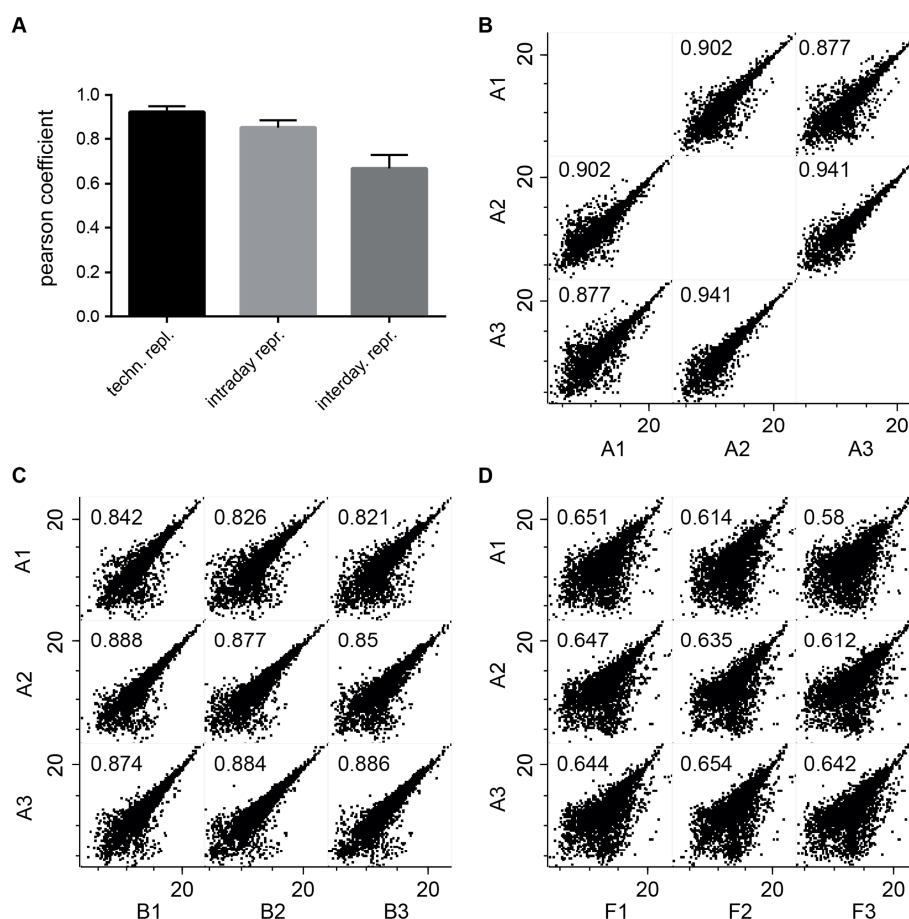


Figure II-15: **A)** Pearson coefficients of technical, intra- and interday biological replicates. Coefficients are determined after the complete data set was extracted using the software Sieve2.0 (Thermo Fisher) with an intensity threshold of $1.0 \cdot 10^4$. **B)** Scatter plots of the technical replicates A1 –A3. **C)** Scatter plots of the biological replicates A and B, which served to evaluate intraday reproducibility. **D)** Scatter plots of the biological replicates A and F to evaluate the interday reproducibility.

In addition to the correlation analysis, volcano plots are a widely used method of representation to illustrate changes of specific m/z-values between two data sets. Therefore, the \log_2 value of the ratio between the samples is determined and plotted against the $-\log_{10}$ of the p-value, which is a parameter of significance. In the volcano plot, correlating m/z-values appear in the center of the graph, whereas non-correlating m/z-values are shifted to the left (downregulated) or right side (upregulated) of the plot.

Figure II-16A shows the volcano plot of biological replicates A and B, which served to compare intraday reproducibility. Referring to equations I-IV (chapter II-1.6), only a small number of m/z-values are significantly (blue) or highly significantly (red) regulated. This result is in line with the previously determined global correlation factor and depicts a more detailed view into m/z-values. In contrast, when a volcano plot is used to compare the interday reproducibility of the biological replicates A and F (Figure II-16B), a vast number

of significantly and highly significantly changing m/z-values were observed. This illustrates a significantly reduced correlation for biological replicates performed on two different days and highlights the difficulty in reproducing the experimental outcome for metabolite level determination within the mid-log phase. Since, bacteria are growing exponentially in mid-log phase, small differences in the measured optical density reflect a large number of bacteria and results in variability in metabolite levels.

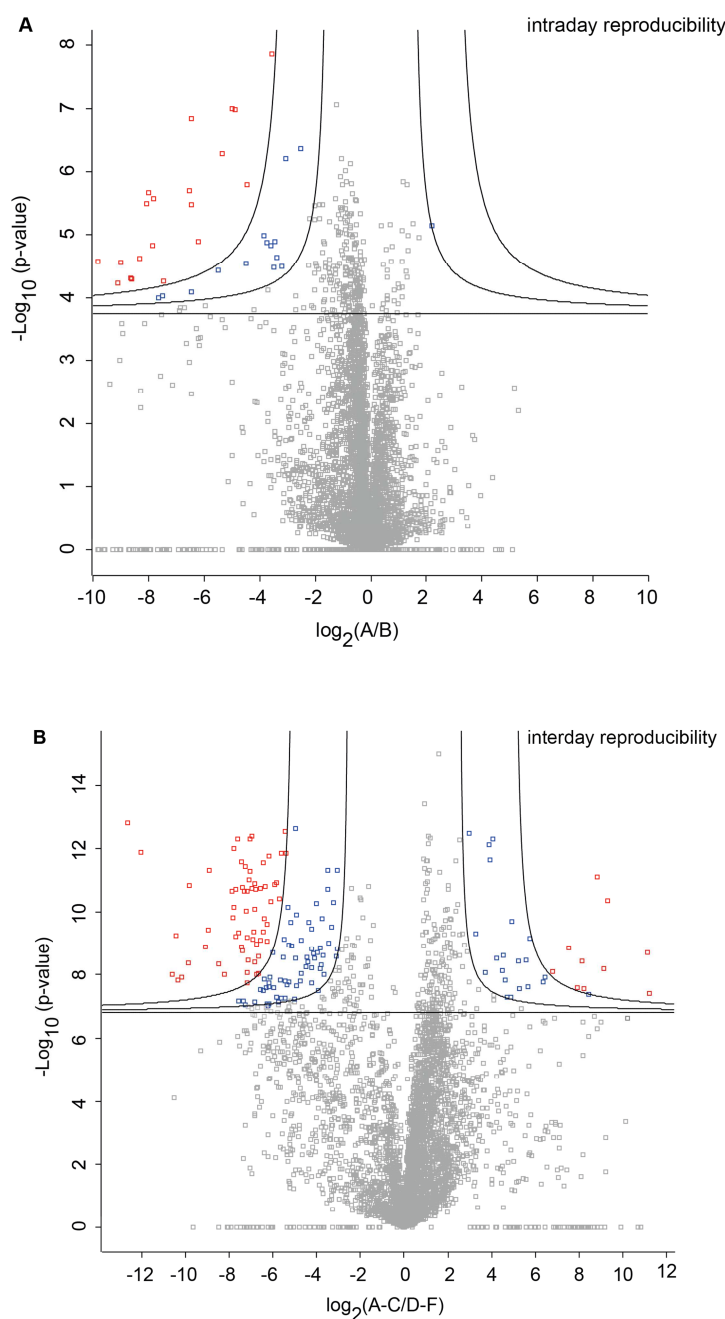


Figure II-16: A) Volcano plot of the intraday reproducibility (A/B). **B)** Volcano plot of the interday reproducibility (A-C/D-F). Significance levels are determined using equations I-IV. Blue and red dots represents significantly and highly significantly regulated m/z-values.

The last aspect of this metabolite study is metabolite identification, which is a critical step in metabolomics. Due to the vast diversity of metabolites, databases are still not very well established. To overcome this shortcoming, an in-house database was set up to distinctly identify the metabolites. The identification was performed using tandem MS and retention times were compared to standard metabolites, that had been deposited in the database. Measurements were performed in negative mode using an intensity threshold for data extraction of $1.0 \cdot 10^4$. In applying the established in-house database to the endogenous polar extracts of *S. aureus* NCTC 8325, 2380 frames with a corresponding m/z -value could be assigned, wherein 49 frames could be distinctly identified. The identified metabolites include sugars, amino acids, nucleotides and cofactors, which covers many important and diverse metabolite pathways. This result is an example for distinct metabolite identification within complex biological matrices and serves as a proof-of-principle study for further identifications. However, it also highlights the urgent need for the development of encompassing databases to enable the assignment of the large number of frames remaining and m/z -values in order to increase the applicability of metabolomic studies.

Table II-3: Metabolites identified within the targeted and untargeted metabolite profiling experiment in *S. aureus* NCTC 8325 using the in-house database. Listed are the IDs, names, deviation of the exact mass to the detected m/z -value and the intensity.

ID	Name	Delta ppm	Intensity
124	NAD ⁺	3.8	4.60E+06
126	NADP ⁺	0.8	2.73E+06
115	ATP	0.7	1.44E+06
99	AMP	0.8	1.26E+06
106	ADP	1.9	1.08E+06
19	<i>L</i> -Aspartic acid	0.2	7.04E+05
31	<i>L</i> -Glutamic acid	0.3	5.02E+05
108	GDP	2.3	4.51E+05
92	UMP	1.8	3.82E+05
120	UDP- <i>D</i> -glucose	3.2	3.78E+05
118	GTP	0.2	3.69E+05
129	FAD	0.3	3.59E+05
30	<i>L</i> -Lysine	0.8	3.18E+05
111	dTTP	0.8	3.07E+05
88	6-Phosphogluconic acid	2.2	2.53E+05
128	CoA	0.8	2.48E+05
42	<i>L</i> -Phenylalanine	0.3	2.40E+05
113	UTP	0.6	2.28E+05

ID	Name	Delta ppm	Intensity
82	Glucose-6-phosphate	2.4	2.23E+05
104	UDP	0.1	2.09E+05
21	Adenine	0.4	2.03E+05
101	GMP	0.9	1.99E+05
70	Ribose-5-phosphate	4.7	1.92E+05
60	Citric acid/Isocitric acid	2.7	1.79E+05
80	Fructose 6-phosphate	3.1	1.28E+05
114	dATP	1.4	1.11E+05
39	<i>L</i> -Histidine	1.0	9.73E+04
64	<i>L</i> -Tryptophan	0.5	7.51E+04
95	dAMP	1.1	4.81E+04
85	Adenosine	0.1	4.16E+04
90	<i>N</i> -Acetylneuraminic acid	1.4	3.93E+04
14	<i>L</i> -Leucine	0.5	3.92E+04
29	<i>L</i> -Glutamine	0.5	3.11E+04
11	Nicotinic acid	0.5	2.81E+04
110	dCTP	0.9	2.76E+04
5	Succinic acid	3.8	2.69E+04
100	IMP	0.3	2.09E+04
101	GMP	3.1	2.02E+04
91	CMP	1.6	2.01E+04
30	<i>L</i> -Lysine	0.9	1.98E+04
112	CTP	1.7	1.83E+04
125	NADH	0.2	1.61E+04
93	cAMP	1.1	1.54E+04
62	Gluconic acid	2.8	1.44E+04
74	Uridine	0.3	1.36E+04
32	<i>L</i> -Methionine	1.1	8.96E+03
67	Pantothenic acid	0.7	7.60E+03
98	cGMP	1.1	6.96E+03

3.1.6 Summary

The goal of this work was to establish an LC-MS-based analytical platform including sample preparation and analysis of metabolite levels in bacteria. As the focus of this protocol was on polar endogenous metabolites, a sample preparation addressing the following important factors was established: (i) harvesting the bacteria via centrifugation; (ii) quenching of residual enzymatic activity by addition of 60% EtOH; (iii) cell lysis with glass beads; (iv) metabolite extraction in 60% EtOH; and finally, (v) removal of macromolecules before LC-MS measurements using 3 kDa filters.

Metabolite extracts were subsequently analyzed via hydrophilic liquid chromatography coupled to a hybrid LTQ-FT-ICR-MS. A zwitterionic HILIC (ZIC®-HILIC) was used for separation of metabolites and a multi-step gradient was established, which avoided co-elution and ion suppression effects. For metabolite identification, an in-house database was built up using the “*n*th order triple play” method. With this method, untargeted metabolite profiling studies can be performed as well as studies of the relative metabolite levels of 129 targeted metabolites from the in-house database.

As a proof-of-principle and to evaluate the performance of an sample preparation protocol and analytical platform, wild-type *S. aureus* NCTC 8325 was used as test organism. The biological reproducibility within the same day (intraday) and on two different days (interday) was analyzed by performing three independent biological experiments on two different days. Additionally, technical reproducibility was reviewed via spiking of two internal standards to the samples prior to LC-MS measurements (*L*-Gln- d_5 and *L*-Glu- d_5). Technical reproducibility proved to be within a good range since internal standards over the complete set of technical replicates had a maximal deviation of less than 20%. Furthermore, Pearson coefficients, which reflect the correlation (1 = perfect correlation, 0 = no correlation) and thereby reproducibility between experiments were analyzed. A Pearson coefficient of 0.92 could be determined for biological replicates, revealing good correlation. Similar results could be found for the intraday, meaning biological replicates performed on the same day, reproducibility with a Pearson correlation coefficient of 0.85. This result highlights the reproducibility of the sample preparation and the experimental design when performed on the same day. In contrast, the interday reproducibility revealed a significantly decreased Pearson coefficient of 0.67. This decreased correlation could be confirmed via significantly and highly significantly changing *m/z*-values between the data sets in the volcano plot. A source of variability in the interday reproducibility may be found in the experimental design, since bacteria were harvested in mid-log phase. In this growth phase, bacteria are dividing exponentially and therefore small changes in optical densities

(day 1 = 0.35 and day 2 = 0.5) can lead to drastic changes in the metabolite levels. In addition evaluating the reproducibility, the metabolite identification using the introduced in-house database was assessed. Thereby, 49 metabolites could be identified through the combination of tandem MS and retention time comparison to standard metabolites.

Applying the established metabolomics workflow, polar metabolites, soluble in 60% EtOH are addressed. To increase the number of metabolites and especially to incorporate the analysis of hydrophobic small molecules, the sample preparation should be adapted accordingly; for example by using chloroform as extraction solvent in combination with reversed-phased chromatography methods. To improve LC-MS performance and to increase the number of detectable low abundant metabolites, capillary HPLC as well as ultra HPLC systems can be introduced in future metabolomics workflows.

3.2 Targeted metabolite profiling in wild-type *S. aureus* USA 300 to elucidate the involvement of the ribokinase ThiD within the pyridoxal salvage pathway

3.2.1 Introduction

In previous studies, *Nodwell et al.* identified a ribokinase (SaThiD) in *S. aureus* as target of the cysteine-reactive natural product antibiotic rugulactone (Ru).¹⁹⁹ ThiD ribokinases are essential for thiamine biosynthesis and play an important role in the pyrophosphorylation of the precursor HMP.²⁰⁰ Surprisingly, further investigations of SaThiD revealed that it did not cluster with the other thiamine biosynthesis genes *tenA*, *thiM*, and *thiE* in *S. aureus* Mu50 and USA300 genomes. Additional substrate studies of SaThiD revealed an increased affinity for pyridoxal (PL, $K_m = 110 \mu\text{M}$) in comparison to the ribokinase substrate HMP ($K_m = 2 \text{ mM}$). In line with these results, *Park et al.* observed similar behavior for *Bacillus subtilis* ThiD²⁰¹, raising the question of ThiD involvement in pyridoxal-5'-phosphate (PLP) production via the PL salvage pathway.

Together, pyridoxal (PL), pyridoxine (PN) and pyridoxamine (PM) comprise vitamin B₆ and are precursors to PLP, an essential cofactor involved in more than 140 enzymatic reactions, or approximately 4% of all classified activities.^{202, 203} While microorganisms and plants are able to synthesize PLP *de novo*, higher organisms are dependent on the uptake of vitamin B₆ via nutrition, which is then further converted to PLP.²⁰⁴ Furthermore, all cells utilize a salvage pathway in which PL is recycled back to PLP by pyridoxal kinases (PLKs).²⁰⁵ In this study, the aim was to clarify the role of SaThiD in PL and PLP metabolism, specifically in the salvage pathway of PL recycling. Therefore, PL levels in the extracellular medium of wild-type *S. aureus* USA 300 and the two transposon mutants TnThiD and TnPdxS were determined in a targeted metabolite profiling experiment.

3.2.2 Sample preparation protocol and analytical platform for the analysis of PL levels in *S. aureus* USA 300 extracts

To determine PL levels in the extracellular medium of *S. aureus* NCTC 8325 and the two transposon mutants TnThiD and TnPdxS, a sample preparation protocol and an analytical platform needed to be established. Therefore, three important aspects within the sample preparation had to be revised: first, cell lysis-induced leakage effects, in which endogenous metabolites get into the extracellular medium. To overcome this problem, bacteria were directly separated from the media via centrifugation, thus avoiding metabolite exchange. Second, metabolite stability needs to be considered for the sample concentration. Before analyzing the extracts via LC-MS, the sample needs to be concentrated to be within the sensitivity range of MS. In parallel, unwanted metabolite degradation should be reduced to a minimum, which can be achieved by a gentle concentration method such as freeze-drying. The third aspect, as previously mentioned in the sample preparation for endogenous metabolites (chapter II-3.1.2), concerns the removal of macromolecules, especially proteins, prior to sample injection onto the LC-MS. Therefore, samples were filtered through a 3 kDa filter such that proteins are retained and metabolites smaller than 3 kDa pass through the filter and can undergo further LC-MS analysis.

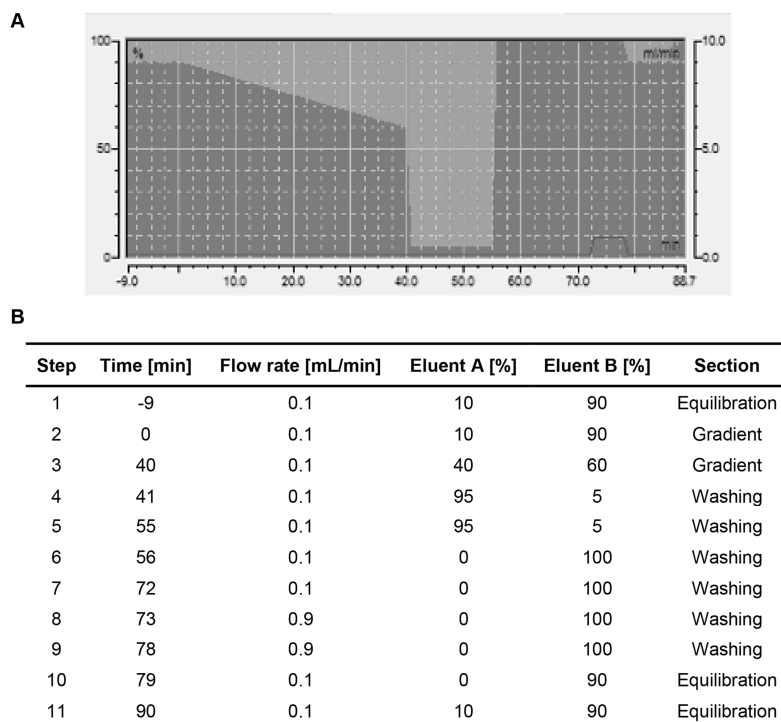


Figure II-17: Established multi-step gradient for the analysis of extracellular PL levels in *S. aureus* USA 300 extracts using the ZIC®-HILIC. Steps are listed with the corresponding time, flow rate, eluent percentage and section of the gradient.

The analytical protocol for the detection of pyridoxal within complex biological matrices utilized a zwitterionic hydrophilic interaction liquid chromatography (ZIC®-HILIC, SeQuant®) coupled to a LTQ-FT-ICR-MS. For relative quantification of pyridoxal levels, a multi-step gradient was established (Figure II-17) which aimed to minimize co-elution and thereby ion suppression effects within the complex biological matrix. MS detection was performed in the high resolution FT-ICR analyzer using the selected ion monitoring (SIM) mode. In the SIM mode, the scan range window can be reduced significantly (167.07 - 169.07 Da) in contrast to the full scan mode (80 – 1000 Da) resulting in increased sensitivity (Figure II-18A, B and C). Additionally, for distinct PL identification in complex biological matrices, tandem MS spectra were generated (Figure II-18D).

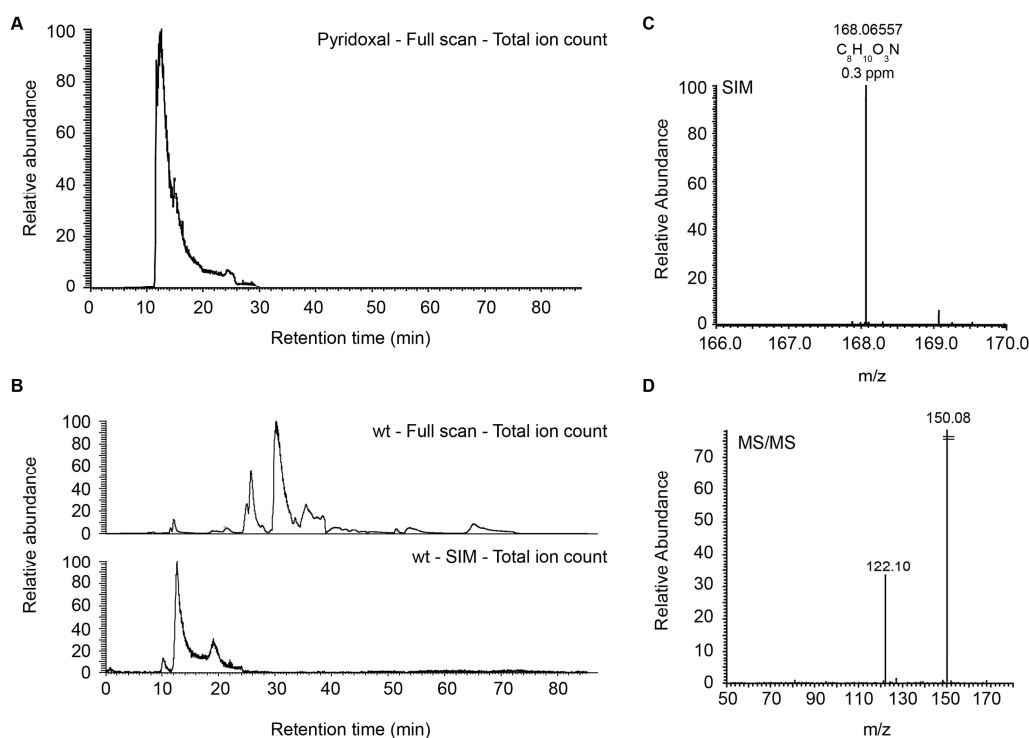


Figure II-18: **A)** Elution of the PL standard using the established LC-MS-based analytical platform. Shown is the total ion count within the SIM scan, using a mass filter of 5 ppm according to the exact mass of PL. The measurement is performed in the positive polarization mode. **B)** PL identification in wild-type *S. aureus* extracts. In the upper lane, the total ion count of the full scan is represented. In the lane below, the SIM scan for PL identification are represented with a mass filter of 5 ppm according to the exact mass of PL. **C)** Mass spectra within the SIM scan at a retention time of 12 min. The mass deviation of the detected peak to the exact mass of PL is shown. **D)** CID induced fragmentation pattern of PL with a normalized collision energy of 40.

3.2.3 Results and discussion

Using the established sample preparation protocol for analysis of extracellular metabolites and the analytical platform optimized for the detection of pyridoxal, targeted metabolite studies in *S. aureus* USA 300 and the two transposon mutants TnThiD (PLK) and TnPdxS were performed. To better characterize and understand the role of SaThiD in the context of the PL salvage pathway, the ThiD mutant (TnThiD) as well as the TnPdxS were included. The TnPdxS mutant is unable to perform *de novo* synthesis of PL and is dependent on PL uptake. The aim was to quantify the level of pyridoxal within the extracellular medium to investigate the role of SaThiD in the uptake of extracellular PL.

First, growth studies of wild-type *S. aureus* USA 300 and the two mutants TnThiD and TnPdxS were performed under supplementation of PL in a chemical defined media. After a 24 h growth period, a final OD₆₀₀ of around 1.2 - 1.3 could be determined for all of the three strains, indicating similar growth behavior (Figure II-19A). To relatively quantify PL levels, bacteria were harvested and separated from chemical defined media by centrifugation. Upon cell extract preparation, PL levels were determined using the established LC-MS-based platform (Figure II-19B). Comparing PL levels of the three *S. aureus* strains, a significantly increased PL level for TnThiD in comparison to the wild-type and TnPdxS could be identified. As wild-type and TnPdxS revealed similar PL levels, PL uptake seems to be favored over *de novo* PL synthesis. In contrast, significant amounts of PL were detected in the media of the TnThiD mutant, indicating that the uptake of PL is affected and PL is built up by *de novo* synthesis.

In summary, the involvement of SaThiD in the PL salvage pathway could be validated using the targeted metabolite profiling platform. Thereby, a significantly increased PL level in studies with the ThiD mutant could be detected, clearly revealing a connection to the PL uptake pathway. This finding is in line with the previous BLAST results and substrate affinity studies and proves the new connection of SaThiD with the PL kinases. SaThiD was thus renamed SaPLK.

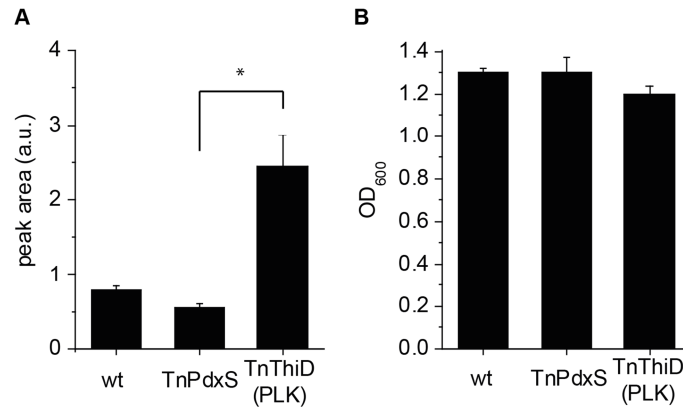


Figure II-19: Targeted metabolite profiling in *S. aureus* USA 300. **A)** Peak areas of PL in extracellular medium for cultures of *S. aureus* WT, TnPdxS and TnThiD after 24 h growth (* p -value = 0.034). **B)** Final OD₆₀₀ for *S. aureus* cultures after 24 h growth. Results are the average of three independent experiments, error bars represent standard deviations.

3.2.4 Summary

The aim of the targeted metabolite profiling study in *S. aureus* USA 300 was to evaluate the involvement of the ribokinase SaThiD within the pyridoxal salvage pathway. Until then, SaThiD was believed to be involved in thiamine biosynthesis by phosphorylating the thiamine precursor 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP). Sequence studies in the genome of *S. aureus* USA 300 and MuA, as well as activity-based substrate studies revealed a connection of SaThiD to the pyridoxal kinase family (PLK). Pyridoxal (PL), as well as pyridoxine (PN) and pyridoxamine (PM), is phosphorylated to pyridoxal-5'-phosphate (PLP) by PLK. PLP is essential for the survival of bacteria and is either recycled via the PL salvage pathway or is synthesized *de novo*. To further investigate the role of SaThiD within the salvage pathway, and thereby in the uptake of PL, PL levels were determined in *S. aureus* extracts. Therefore, wild-type *S. aureus*, the TnPdxS mutant, which is not able to undergo PL *de novo* synthesis, and the TnThiD transposon mutant were grown in chemically-defined media for 24 h. For the analysis of metabolite levels, a sample preparation protocol for extracellular metabolites as well as a LC-MS-based analytical platform was established.

For sample preparation, bacteria were first separated from the chemically-defined media via centrifugation, avoiding leakage effects of endogenous metabolites into the exogenous extracts. Following LC-MS analysis, the extracts were gently concentrated via freeze-drying and macromolecules were removed using a 3 kDa filter. Subsequently, LC-MS measurements were performed on a ZIC®-HILIC coupled to a hybrid LTQ-FT-ICR-MS. For sensitive detection of PL within the complex biological matrices, a multi-step gradient was established, avoiding co-elution and thereby ion suppression effects. For increased sensitivity, the selected ion monitoring mode (SIM) was applied in which only a small range around the *m/z*-value of pyridoxal is covered and detected, leading to a significantly increased sensitivity. Applying the established targeted platform for PL quantification, significant differences within the three *S. aureus* strains could be determined. The wild-type and TnPdxS strains revealed similar PL levels, indicating that under normal conditions the PL salvage pathway in the bacteria is used for PLP production. Through the deletion of SaThiD, a significantly increased PL level could be detected, which indicates, that PL uptake via the salvage pathway is hindered and PLP is produced via the *de novo* synthesis pathway. This result confirms the PL kinase activity of SaThiD and reveals the influence of SaThiD within the salvage pathway, which was until then not known. This highlights the potential of targeted LC-MS-based metabolite profiling studies in the context of protein characterization and increases the knowledge of basic biochemistry and cellular processes in bacteria.

3.3 Subclass-Specific Labeling of Protein-Reactive Natural Products with customized nucleophilic probes

3.3.1 Introduction

Nature provides a rich source of bioactive compounds that comprise a huge diversity of pharmacological activities. In medicine, they constitute up to 50 % of all currently applied drugs and serve as a great source for privileged structures in biomimetic compound screening programs.^{148, 206, 207} These structures have been optimized during evolution to bind and inhibit essential cellular proteins and enzymes. While a large fraction of compounds facilitate this inhibition by reversible interaction a certain subclass of natural products is designed to attach covalently with the active site.^{208, 209} This subclass of so called protein reactive natural products adopts a precise reactivity and specificity profile that ensures target selectivity. Advantage of covalent inhibition is e.g. an increase in potency due to a prolonged duration of the corresponding biological effect.²¹⁰ In fact, many natural product derived drugs, such as β -lactam antibiotics and aspirin rely on this covalent inhibition principle.²¹¹ Scaffolds that exhibit protein reactivity are diverse and include several prominent electrophilic moieties such as Michael-acceptors, β -lactams, β -lactones or epoxides. Some microorganisms produce an additional set of electrophilic scaffolds such as maleimides.²¹² In previous studies, we investigated the targets of these specialized scaffolds such as showdomycin, a potent maleimide antibiotic with activity against *Staphylococcus aureus* as well as multidrugresistant derivatives.²¹³ The molecule inhibits pathogen growth by targeting two essential cell wall biosynthesis enzymes via covalent modification of a cysteine residue. Intrigued by this mode of action, we became interested to explore the subclass of reactive electrophiles in greater detail by establishing a directed isolation procedure.

Although technological advancements over the last decades enabled the discovery and structural elucidation of many compounds via NMR and mass spectrometry, the field is confronted with several fundamental challenges: First, the extraction, fractionation and isolation of natural products requires very sensitive analytical methods in order to detect low abundance compounds. Second, the physical properties of some natural products, such as showdomycin (hydrophilic, poorly UV-active) are not compatible with established screening procedures and may evade detection. Third, usually a large number of compounds is detected but only a fraction can be analyzed by detailed follow up studies. Selection criteria have to be established e.g. based on desired activities in order to choose promising compounds.

Here we introduce nucleophilic tags that selectively bind electrophilic moieties of natural product antibiotics (Figure II-20). We synthesized several thiol, hydroxyl and amine nucleophiles and screened these molecules against various protein reactive compounds and natural products. We found a suitable reactivity and selectivity of electron-rich, thiol-substituted naphthalenes with showdomycin and phosphomycin. Application of the probe in the corresponding producer strain extracts facilitated UV detection of tagged natural products at a unique wavelength of 303 nm that significantly reduced the complexity of secretome samples. Moreover, the attached tag increased the molecular weight and hydrophobicity of the molecular adducts improving their separation by HPLC and identification by mass spectrometry (MS). The approach is thus complementary to several excellent previous studies by the *Carlson*, *Cravatt*, *Li* and *Sello* labs which utilized conjugation methods for the ligation of reversible binding natural products with characteristic functionalities (e.g. alcohols, ketones, amines and free carboxylic acid).^{155, 214, 215, 216, 217}

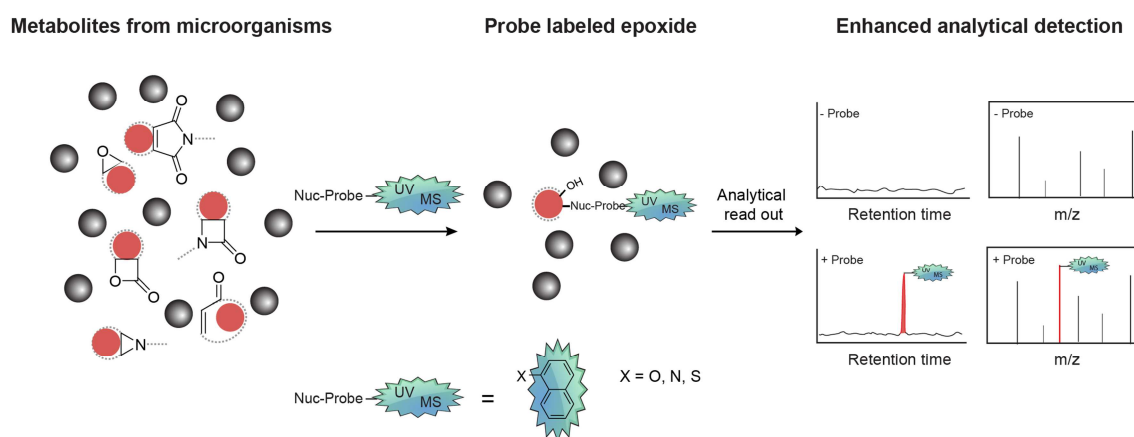


Figure II-20: Selective detection of protein reactive natural products. Metabolic extracts are incubated with a nucleophilic probe based on a naphthalene scaffold. The tagging increases retention on HPLC, UV-detection and analysis by MS.

3.3.2 Probe design and synthesis

Specific probes for electrophilic natural products must fulfil the following criteria: They must contain a representative set of nucleophiles such as thiols, amines or alcohols. Moreover, they need to be incorporated into a rigid scaffold that increases the UV-absorption, the molecular mass, as well as retention on HPLC of trapped molecules. Additionally, this scaffold must be amenable to the incorporation of ligands that enhance the reactivity of the corresponding nucleophile. Naphthalene scaffolds represent such a desired structural backbone that can be equipped with the desired heteroatom nucleophilic centres. The unique UV-absorption characteristics (local maximum around $\lambda = 303$ nm) in combination with its high hydrophobicity enhance the identification of small, hydrophilic, natural products in reversed-phase LC separation and UV detection. Additionally, this medium sized molecular entity can serve as a distinct mass label in mass spectrometry (MS) through a predictable mass shift of the captured natural products, which represents an advantage in case of low molecular weight target compounds. In addition, fine tuning of the nucleophilic reactivity is mediated through the appropriate attachment of electron-donating groups (EDG) with +I (alkyl) and +M (methoxy) effects (Figure II-21A). Finally, reactive sites can be placed in close proximity to an adjacent hydrogen-bond acceptor in *peri*-position which may further facilitate partial deprotonation of the nucleophilic center and thus increase the reactivity.^{218, 219, 220}

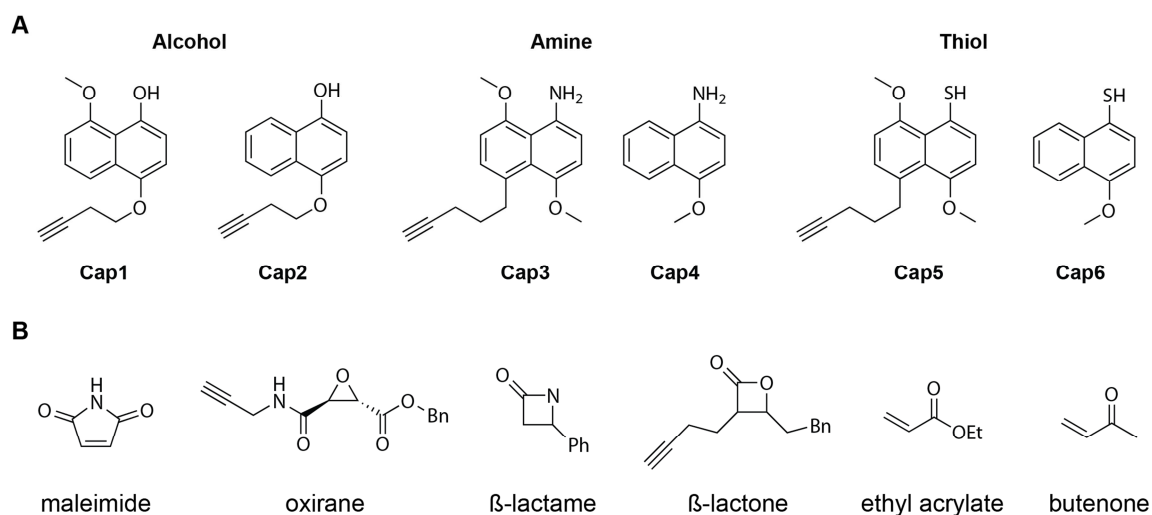


Figure II-21: Selective detection of protein reactive natural products. **A)** Structures of probes with various nucleophilic and electron donating properties. **B)** Electrophilic molecules tested towards reactivity with nucleophilic probes.

A divergent synthetic strategy was chosen to obtain the desired amino- and thiol-substituted naphthalene probes **Cap3** - **Cap6**. The representative synthetic pathways for **Cap3** and **Cap5** are displayed in Figure II-22. Bismethoxy naphthalene **2** was halogen-functionalized followed by selective alkylation in *peri*-position to introduce an alkyne handle to afford general building block **5**. Final installment of the nucleophilic center was accomplished via lithiation and amination for **Cap3** or thiolation for **Cap5** with the respective electrophile. **Cap5** was found to be prone to oxidative dimerization during synthesis, therefore the stable dimer [**Cap5**]₂ served as the synthetic target. Additionally, analogues lacking the acceptor moiety in *peri*-position to the nucleophilic group were prepared to compare the influence of the acceptor moiety on the nucleophilic reactivity (Figure II-21A). Hydroxy-based probes **Cap1** & **Cap2** were accomplished via a natural product based approach starting from Juglone. Synthetic details are provided in the supporting information section 4.

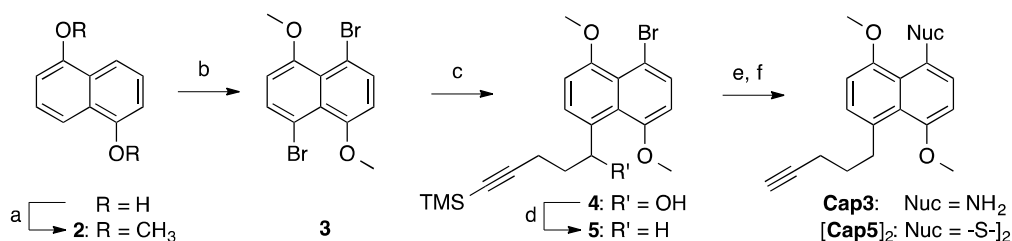


Figure II-22: Representative synthetic pathways for nucleophilic probes **Cap3** & [**Cap5**]₂: a) Me₂SO₄, K₂CO₃, acetone, rfx, 88%; b) NBS, acetonitrile, RT, 52%; c) i) *n*-BuLi (1.00 eq), THF, -78 °C; ii) 5-(trimethylsilyl)pent-4-ynal, THF, -78 °C to RT, 64%; d) TFA, Et₃SiH, CH₂Cl₂, RT, 10 min, 78%; e) for **Cap3**: i) *n*-BuLi, THF, -78 °C; ii) dppa, THF, -78 °C to RT; iii) Red-Al, THF, 0 °C; for [**Cap5**]₂: i) *n*-BuLi, THF, -78 °C, 30 min; ii) sulfur powder, THF, -78 °C to RT, 18 h; iii) Red-Al, THF, -10 °C, 1 h; f) TBAF, THF, **Cap3**: 51% from **5**, [**Cap5**]₂: 13% from **5**.

3.3.3 Probe reaction with prominent electrophilic entities

All probes were tested for reactivity towards a collection of common electrophilic groups (Figure II-21B). Thiol based probes had to be reduced before incubation with electrophiles since the electron enriched naphthalene π-system was prone to oxidation to form thiol dimers. Several reduction methods were screened to identify a high-yielding procedure (Table II-A3). Remarkably, only a TCEP-promoted conversion in slightly acidified aqueous medium developed by *Whitesides et al.* provided the reduced probe **Cap5** in quantitative yield (Figure II-A1).²²¹ As TCEP can act as a nucleophile for Michael acceptor systems as well²²² traces had to be removed after the reduction via filtration over silica.

Upon incubation of all probes with electrophiles for 2 h at equimolar concentration in organic solvent (DMSO/acetonitrile) the reaction products were analyzed by HPLC. Alcohol as well as amine based nucleophilic probes yielded no significant substrate conversion with any electrophile (data not shown). Satisfyingly, thiol based probes quantitatively modified the maleimide scaffold in organic solvent (Figure II-A2) as well as under physiological conditions in aqueous PBS buffer (18 h, 10 to 20 fold excess, pH 7.4) (Table II-A4). In addition, thiol probe **Cap6** without the *peri*-substitution showed a significant higher conversion (50 %) of the epoxide compared with the sterically constrained probe **Cap5**. The results of this study furthermore suggest that the current probe design does not provide a sufficient reactivity of alcohols and amines towards the electrophilic centres. Only thiols as the strongest nucleophile in this series were able to react independent of the presence of the acceptor. As **Cap6** revealed the best results for capturing maleimide and epoxide scaffolds we utilized this probe for all subsequent studies.

3.3.4 Fine tuned naphthalene thiols for maleimide and epoxide capturing

Synthesis of the nucleoside antibiotic showdomycin was adapted from a published procedure with slight modifications.^{213, 223} Incubation of the probes for 3 h with the natural product revealed quantitative conversion with **Cap6**, the compound lacking the methoxy group in *peri*-position (Figure II-23). Importantly, the addition of the probe dramatically altered the physical properties of showdomycin leading to a significantly improved retention on HPLC (there is no retention with unmodified showdomycin on reversed phase LC systems) as well as a specific detection at 303 nm. Of note, as there is no stereo- and regio-control, the nucleophilic attack onto the showdomycin maleimide scaffold results in the formation of isomer peaks, which slightly reduce the overall signal intensity (Figure II-23). As showdomycin is sterically more constrained compared to model substance *N*-ethylmaleimide the trajectory of the attacking thiol group requires more space in order to reach the shielded electrophilic center. Thus compound **Cap5** with an adjacent methoxy-H-acceptor did not show any reactivity likely due to the additional steric bulk (Figure II-A2).

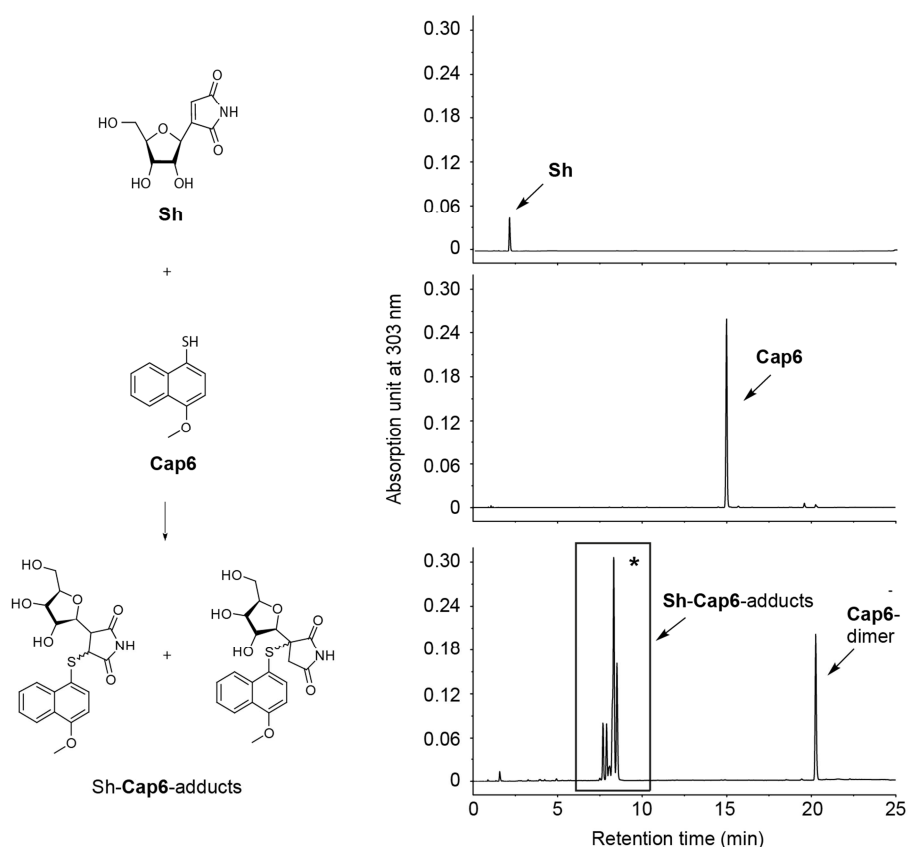


Figure II-23: Reactivity of thiol probe Cap6 with natural product showdomycin (Sh). HPLC analysis of the reaction of showdomycin with **Cap6** in aqueous buffered media (PBS/DMSO 9:1, pH 7.4, 30 °C, 18 h). HRMS of ***Sh-Cap6-adducts** calcd. for $C_{20}H_{21}NO_7S$ [M]: 418.0966; found; 418.09664; diff.: 2.73 ppm.

In order to dissect the reactivity of probe **Cap6** we compared the reaction with naphthalenethiol without electron-donating para-methoxy group. Incubation of **Cap6** and naphthalene-1-thiol with showdomycin revealed an addition product only with the electron rich **Cap6** probe (Figure II-A3) demonstrating that a donor is mandatory for elevated nucleophilicity.

We next investigated the selectivity of capture probe **Cap6** for maleimides and epoxides in a mixture of 8 electrophilic natural products (Figure II-24A and B). Importantly, the probe only converted showdomycin (95%), phosphomycin (36%) and FM209 (83%) emphasizing a desired specificity for maleimide and epoxide containing molecules (Figure II-A4). Again, the tag significantly improved the HPLC retention (Figure II-A5) on reversed phase C18 columns and the ionization of modified showdomycin, phosphomycin and FM209 could be increased under mild ESI conditions 2-3 fold (Figure II-24C, extended gradient shown in Figure II-A6). The MS detection limit for **Sh-Cap6** and **Pp-Cap6** is in the range of 100 to 150 pmol (Figure II-A7).

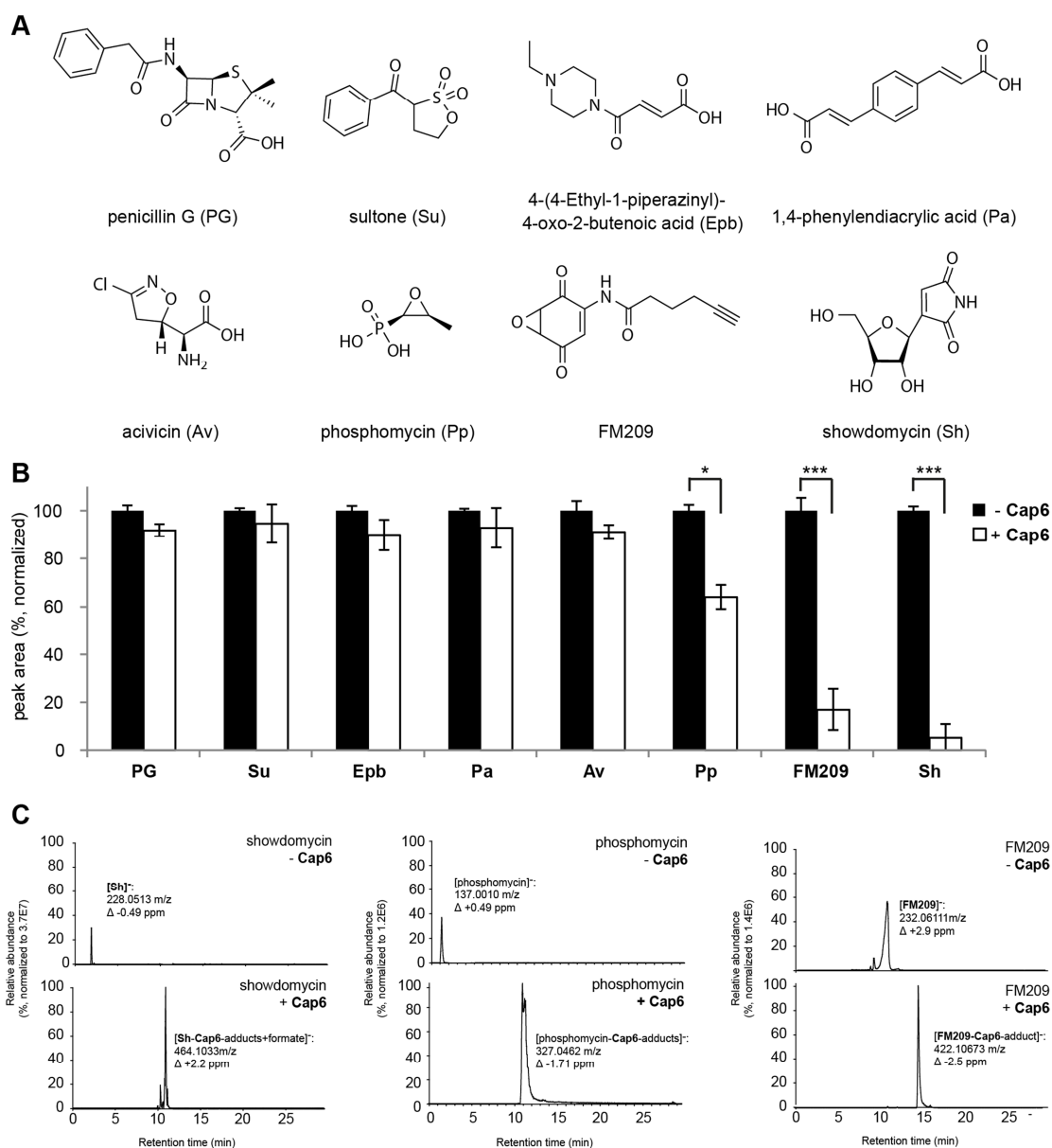


Figure II-24: Profiling of probe specificity towards a panel of electrophiles. **A)** Structures of natural products or natural product inspired compounds. **B)** Conversion of compounds with Cap6. Calculated *p*-values for the different electrophiles: $p = 1.58E-02$ for phosphomycin, $p = 1.49E-04$ for FM209 and $p = 5.02E-05$ for Sh. For FM209 and Sh *p*-values are statistically highly significant ($***p < 0.005$) and for phosphomycin statistically significant ($*p < 0.05$). **C)** Reversed-phased HPLC-ESI-HRMS traces of showdomycin, phosphomycin and FM209 before and after reaction with Cap6.

3.3.5 Labeling of showdomycin and phosphomycin in metabolic extracts of natural producer strains

Due to the insufficient retention of showdomycin and phosphomycin on reversed phase C18 columns (Figure II-24C), direct HPLC-MS analysis of the natural products in the extracts of producing organisms is challenging. With a selective probe in hand we tested the modification of showdomycin and phosphomycin antibiotics in their producing strains *S. showdoensis* and *S. fradiae*, respectively.^{224, 225} HPLC analysis of *S. showdoensis* extract in absence of probe **Cap6** revealed only a few peaks at the characteristic naphthalene absorption wavelength. Importantly, addition of **Cap6** resulted in the formation of intense peaks with the expected retention time of showdomycin-probe isomers (Figure II-25A).

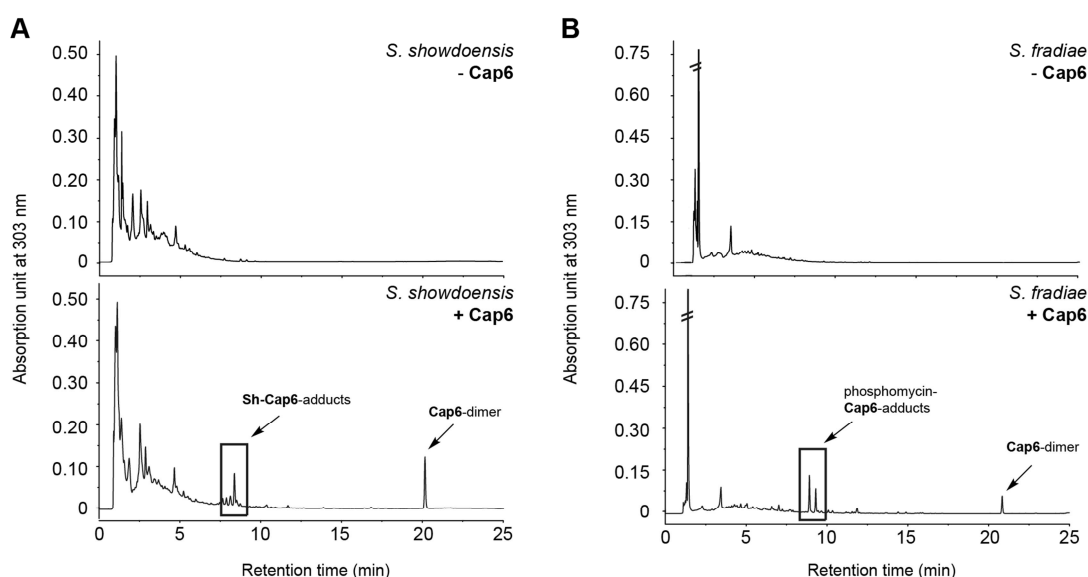


Figure II-25: Showdomycin and phosphomycin identification in *Streptomyces* producer strains. **A)** HPLC traces of *S. showdoensis* extracts before and after treatment with **Cap6**. **B)** HPLC traces of *S. fradiae* extracts before and after treatment with **Cap6**.

In addition, the methoxy-thionaphthalene mass tag led to a characteristic shift in the MS spectrum facilitating an easy showdomycin-probe detection in both the MS base peaks as function of elution time (Figure II-26A) as well as in the sum of all mass spectra (Figure II-26B). Similarly, addition of **Cap6** to the extract of *S. fradiae* resulted in the selective identification of the phosphomycin adducts via HPLC separation and detection via UV and MS (Figure II-25B and Figure II-26C-D). Importantly, neither showdomycin nor phosphomycin could be detected in absence of **Cap6** as the poor retention of both compounds resulted in co-elution with metabolites that suppressed their ionization.

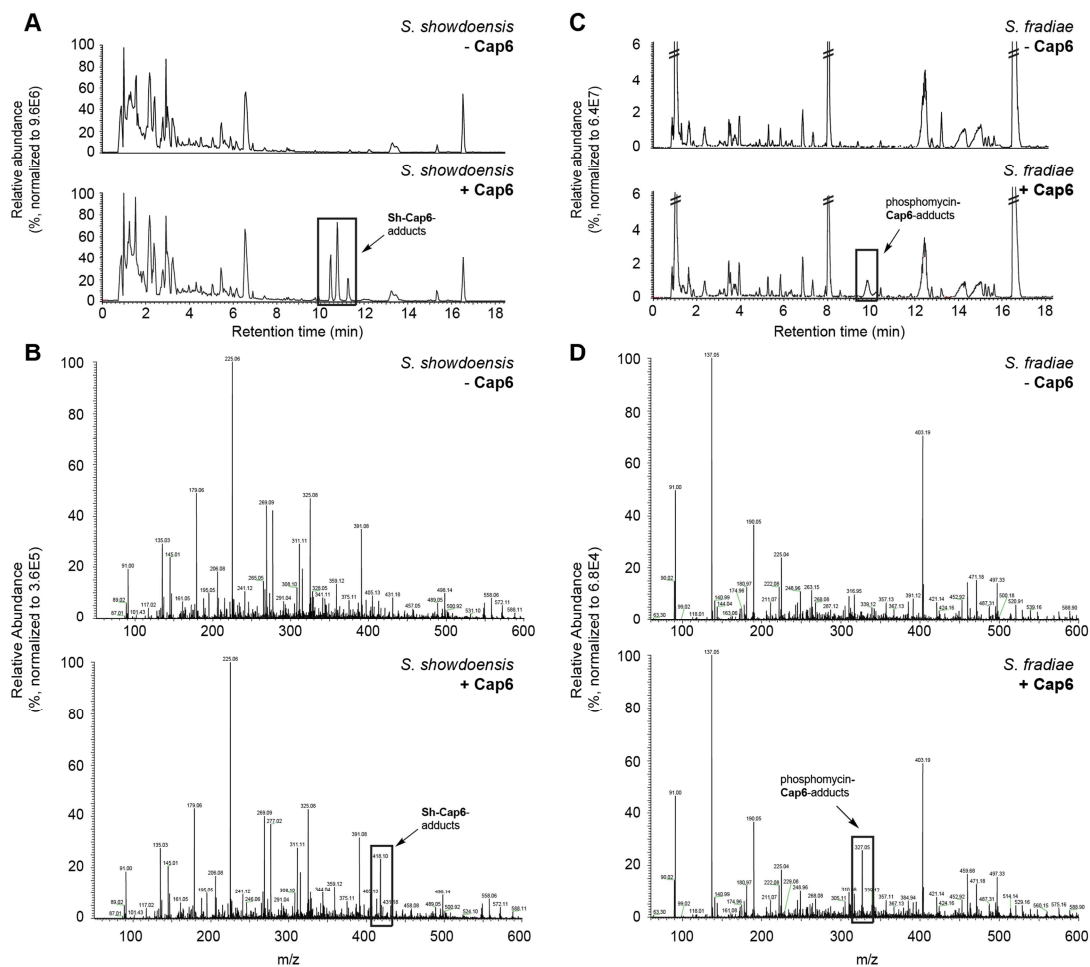


Figure II-26: Showdomycin and phosphomycin identification in *Streptomyces* producer strains. **A)** Base peaks of all mass spectra as a function of elution time (50-600 Da) of *S. showdoensis* extracts before and after treatment with **Cap6**. **B)** Sum of all mass spectra from 1 to 16 mins/ over the whole gradient of *S. showdoensis* extracts before and after treatment with **Cap6**. **C)** Base peaks of all mass spectra as a function of elution time (50-600 Da) of *S. fradiae* extracts before and after treatment with **Cap6**. **D)** Sum of all mass spectra from 1 to 16 mins/ over the whole gradient of *S. fradiae* extracts before and after treatment with **Cap6**.

3.3.6 Summary

The discovery of novel natural products is a major challenge that can only be advanced by improved detection methods.²²⁶ Methods to modify a conserved structural motif such as an amine or carboxylic acid group have been introduced previously via e.g. chemical coupling procedures.^{155, 214, 215} Due to the wealth of bioactive compounds containing an electrophilic center, which is crucial for their mechanism of action, a method to selectively capture these protein reactive natural products by a nucleophilic trap is desired. Recently, *Mitchell et al.* utilized dithiothreitol (DTT) as a small tag for the analysis of natural products containing dehydrated amino acids (DHAAs).²²⁷ In our proof-of-principle study we applied a naphthalene core structure due to its excellent UV absorption, its distinct mass shift, hydrophobic properties, and the wealth of possible modifications that allow to fine tune the nucleophilicity. While the introduction of H-bond acceptors in peri-position lowered the steric accessibility and thus their applicability, the incorporation of electron donating groups led to a significant increase in reactivity. Among the tested nucleophiles one probe with a methoxy acceptor and a nucleophilic thiol group proved to be optimal to selectively react with epoxides and maleimides in presence of other electrophiles such as beta-lactams and Michael-acceptors. The applicability of the probe for capturing natural products was demonstrated in complex extracts. While showdomycin and phosphomycin were difficult to detect in the absence of probe, probe addition resulted in modified compounds with improved retention times, UV absorption and MS ionization. The additional incorporation of a bioorthogonal tag such as an azide or alkyne would further allow the selective enrichment of captured natural products. This could be in particular useful to obtain quantities of low abundance compounds suitable for structural characterization by NMR. In fact, some of our probes are already equipped with alkyne substituents via robust synthetic operations. However, care must be taken in the selection of the alkyne position on the naphthalene scaffold as probe **Cap5** was unreactive probably due to steric restrictions. Moreover, in order to expand the coverage of probes to additional protein reactive natural products such as β -lactams, β -lactones and Michael-acceptors, the design of nucleophilic naphthalene traps will include electronically enriched benzene scaffolds with less steric hindrance, heteroaromatic systems with different EDG to fine tune reactivity and positively charged ligands to further increase MS detection.

4. Experimental Section

4.1 Establishment of a targeted and untargeted LC-MS-based metabolite profiling platform for the analysis of metabolite levels in bacteria

4.1.1 Overnight cultures and cryostock preparation of *S. aureus* NCTC 8325

Overnight cultures are prepared starting from an initial cryostock solution of *Staphylococcus aureus* NCTC 8325. Therefore, a small amount is transferred into 5 mL B medium and incubated overnight at 37 °C and 200 rpm. Overnight cultures are subsequently centrifuged for 10 min at 5000 rpm and 4 °C. The supernatant is discarded and the cells are redissolved in 1 mL fresh B medium. 1 mL sterilized glycerol solution is added and mixed thoroughly. Aliquots of the bacterial solution are immediately frozen in liquid nitrogen and stored at -80 °C.

4.1.2 Growth curves of wild-type *S. aureus* NCTC 8325

Growth curves of wild-type *S. aureus* NCTC 8325 are carried out starting with the preparation of a *S. aureus* preculture. Therefore, 50 µL of an *S. aureus* cryostock aliquot solution are added to 5 mL B medium and incubated for 1.5 h at 37 °C and 200 rpm. After incubation, the *S. aureus* cell number is adjusted to $3.4 \cdot 10^7$ cells and further diluted in three dilution series steps to a cell number of $3.4 \cdot 10^4$. In the first dilution series, 1 mL of the *S. aureus* $3.4 \cdot 10^7$ culture is diluted with 9 mL B medium. In the second dilution series, 5 mL of the *S. aureus* culture with a cell number of $3.4 \cdot 10^6$ are added to 45 mL B medium. In the third dilution series, 25 mL of the $3.4 \cdot 10^5$ culture solution are added to 225 mL B (1 L Erlenmeyer flask). The culture is then incubated at 37 °C and 200 rpm and bacterial growth is recorded via the measurement of the optical density at 600 nm over a period of 24 h.

4.1.3 Sample preparation protocol for the targeted and untargeted analysis of endogenous polar metabolite levels in *S. aureus* NCTC 8325

The sample preparation for the analysis of endogenous metabolite levels within the mid-log phase of *S. aureus* NCTC 8325 can be divided into three days. On the first day, precultures of *S. aureus* are prepared and diluted as described previously to a cell number of $3.4 \cdot 10^4$. Afterwards, 1 L *S. aureus* cultures are grown for 4 h at 37 °C and 200 rpm. The following steps are performed under cooling conditions. Bacteria are harvested by centrifugation for 5 min at 6000 rpm and 4 °C. Pellets are redissolved in 20 mL PBS (4 °C) and centrifuged for 5 min at 6000 rpm and 4 °C. Subsequently, the supernatant is discarded, the pellet is redissolved in ddH₂O (4 °C) and subsequently pelletized for 5 min at 6000 rpm and 4 °C. Afterwards, the supernatant is discarded and the pellets redissolved in 60% ethanol (4 °C). The quenched bacterial suspension is transferred to 7 mL Precellys Ceramic Kit tubes (0.1 mm, Peqlab) and frozen in liquid nitrogen. Samples can be stored at -80 °C until proceeding further.

On the second day, *S. aureus* cultures are lysed using the Precellys Ceramic Kit tubes. Therefore, two cycles (Program 3) on the homogenizer are carried out. Samples are cooled for 2 min in between the two cycles. Afterwards, the supernatants are transferred into 50 mL falcons. The Precellys Ceramic Kit tubes are washed twice with 1 mL H₂O (LC-MS grade) and both washing fractions are added to the 50 ml falcon. Before concentration of the samples via freeze drying, the EtOH percentage is reduced from 60% to 10% by adding 23 mL H₂O (LC-MS grade, 4 °C). Afterwards the solutions are frozen in liquid nitrogen and freeze-dried for 1.5 d.

On the third day, the freeze-dried samples are redissolved in 1 mL H₂O (LC-MS grade) and centrifuged for 10 min, 5000 rpm and 4 °C. For the second concentration step, the solutions are transferred into 1.5 mL Safe-Lock tubes, again frozen in liquid nitrogen and freeze-dried overnight. The crystalline sample can then be stored at -80 °C until LC-MS measurements.

On the day of the LC-MS-based measurements, samples are redissolved in 70 µL H₂O (LC-MS grade) and vortexed for 1 min. Samples are then transferred to Amicon Ultra 3K centrifugal filters to remove macromolecules and proteins. Samples are centrifuged for 2 - 4 h at 5000 g and 4 °C. The metabolite solutions are then transferred to glass vials and are ready for LC-MS analysis.

4.1.4 LC-MS-based analytical platform for the detection of endogenous polar metabolite levels in bacteria

The samples (5 μ L injection volume) are analyzed by LC-HESI-MS/MS on a Thermo Finnigan LTQ FT-ICR and were eluted by a Dionex Ultimate 3000 over a ZIC®-HILIC column (250 x 2.1 mm, 5 μ m particle size). The column temperature is maintained at 30 °C. Eluting buffers are buffer A (10 mM ammonium formate in 95:5 H₂O:MeCN, pH 6.2) and buffer B (10 mM ammonium formate in 5:95 H₂O:MeCN, pH 7.2). The gradient used is -9 \rightarrow 0 min; 90% buffer B (0.1 mL/min); 0 \rightarrow 130 min; 90% \rightarrow 65% buffer B (0.1 mL/min); 130 min \rightarrow 140 min; 40% buffer B (0.1 mL/min); 140 \rightarrow 141 min; 40% \rightarrow 5% buffer B (0.1 mL/min), 141 \rightarrow 155 min; 5% buffer B (0.1 mL/min); 155 \rightarrow 156 min; 5% \rightarrow 100% buffer B (0.1 mL/min); 156 \rightarrow 172 min; 100% buffer B (0.1 mL/min); 172 \rightarrow 173 min; 100% buffer B (0.9 mL/min); 173 \rightarrow 178 min; 100% buffer B (0.9 mL/min); 178 \rightarrow 179 min; 100% \rightarrow 90% buffer B (0.1 mL/min) and 179 \rightarrow 187 min; 90% buffer B (0.1 mL/min). The chromatographic eluent is directly introduced into the ion source without prior splitting. Ions are scanned in negative or positive polarity mode and analyzed with the nth order triple play method. Using this method, lists of the standard metabolites in the database, including m/z-values, retention times and fragmentation patterns (Table II-4 and Table II-A1 and A2) are specified. The “nth order triple play” method is divided into three scan events in which the second and the third scan event are data-dependent. The first scan event is a full scan and performed in the FT-ICR with a resolution of 25,000 and a mass range window of 60 - 1000 Da. The second scan event is performed in the selected ion monitoring (SIM) scan in which the list of the database metabolites is processed (minimal signal threshold 100, nth most intense from list and analyze top 50 peaks). In the third scan event, which is synchronized with the previous SIM scan event, CID activation is performed (default charge state 1, isolation width 2 m/z, normalized collision energy 40 and activation time 30 ms). HESI source parameters in positive polarity mode: spray voltage = 3.5 kV; vaporizer temperature = 60 °C; sheath gas flow rate = 30, aux gas flow rate = 10, capillary voltage = 35 V; capillary temperature = 275 °C; tube lens = 60 V. HESI source parameters in negative polarity mode: spray voltage = 3.5 kV; vaporizer temperature = 60 °C; sheath gas flow rate = 30, aux gas flow rate = 10, capillary voltage = -35 V; capillary temperature = 275 °C; tube lens = -60 V.

Table II-4: List of metabolites specified in the “nth order triple play” method. Specified parameters: metabolite name, m/z-value positive and negative mode, elution time window (start time and end time) of the metabolites within the gradient.

ID	Name	m/z (pos)	m/z (neg)	Start time (min)	End time (min)
1	3-Hydroxybutyric acid	x	103.04	20	34
2	L-Serine	106.05	104.04	63	77
3	L-Proline	116.07	x	43	57
4	Uracil	x	111.02	5	19
5	Succinic acid	x	117.02	59	73
6	2-Hydroxy-3-methylbutyric acid	x	117.06	9	23
7	L-Valine	118.09	x	41	55
8	L-Threonine	120.07	118.05	53	71
9	L-Homoserine	120.07	118.05	53	71
10	L-Cysteine	x	120.01	44	58
11	Nicotinic acid	124.04	122.02	19	33
12	Thymine	x	125.04	3	17
13	Pyroglutamic acid	x	128.04	42	56
14	L-Leucine	132.10	130.09	25	43
15	L-Isoleucine	132.10	130.09	25	43
16	Glutaric acid	x	131.03	72	86
17	L-Asparagine	133.06	131.05	63	77
18	Ornithine	133.10	131.08	129	143
19	L-Aspartic acid	134.04	132.03	80	94
20	L-Malic acid	x	133.01	77	91
21	Adenine	136.06	134.05	7	21
22	Hypoxanthine	137.05	135.03	10	24
23	4-Hydroxybenzoic acid	x	137.02	10	23
24	Acetyl phosphate	x	138.98	87	101
25	4-Amino-5-hydroxymethyl-2-methylpyrimidine (HMP)	140.08	x	11	25
26	4-Guanidinobutanoic acid	146.09	144.08	60	74
27	Spermidine	147.17	x	30	44
28	α -Ketoglutaric acid	x	145.01	69	83
29	L-Glutamine	147.08	145.06	61	75
30	L-Lysine	147.11	145.10	129	143
31	L-Glutamic acid	148.06	146.05	79	93
32	L-Methionine	150.01	148.04	33	47
33	Guanine	152.06	150.04	18	32
34	L-Glutamine-2,3,3,4,4-d ₅	x	150.09	60	75
35	Xanthine	153.04	151.03	10	24
36	L-Glutamic acid-2,3,3,4,4-d ₅	x	151.08	79	93
37	2,3-Dihydroxybenzoic acid	x	153.02	8	22

ID	Name	m/z (pos)	m/z (neg)	Start time (min)	End time (min)
38	<i>L</i> -Lysine-3,3,4,4,5,5,6,6-d ₈	x	153.15	129	143
39	<i>L</i> -Histidine	156.08	154.06	67	81
40	Allantoin	x	157.04	22	36
41	Phenylpyruvic acid	x	163.04	3	17
42	<i>L</i> -Phenylalanine	166.09	164.07	21	35
43	<i>L</i> -β-Phenyllactic acid	x	165.06	5	42143
44	Pyridoxal	168.07	166.05	6	20
45	Phosphoenolpyruvic acid	x	166.98	101	115
46	Pyridoxamine	169.10	167.08	60	74
47	Pyridoxine	170.08	168.07	6	20
48	Dihydroxyacetone phosphate	x	168.99	83	97
49	Glycerol-3-phosphate	173.02	171.01	80	94
50	<i>L</i> -Arginine	175.12	173.10	120	134
51	Citrulline	176.10	174.09	69	83
52	Ascorbic acid	x	175.02	50	64
53	<i>D</i> -Fructose/ <i>D</i> -Glucose	x	179.06	38	57
54	<i>Myo</i> -inositol	x	179.06	61	75
55	<i>L</i> -Tyrosine	182.08	180.07	39	53
56	<i>L</i> -Arginine-d ₁₀	x	183.17	119	134
57	O-P-Serine	186.01	184.00	111	125
58	3-Phosphoglyceric acid	x	184.99	99	113
59	<i>N</i> (6)-acetyl- <i>L</i> -lysine	189.12	187.11	48	62
60	Citric acid /Isocitric acid	x	191.02	86	103
61	Glucuronic acid	x	193.04	72	86
62	Gluconic acid	x	195.05	59	73
63	O-P-Threonine	200.03	198.02	104	118
64	<i>L</i> -Tryptophan	205.10	203.08	23	37
65	2-Deoxyribose-5-phosphate	x	213.02	76	90
66	1-Deoxyxylulose-5-phosphate (DXP)	x	216.03	65	80
67	Pantothenic acid	220.12	218.10	34	48
68	<i>N</i> -Acetyl-glucosamine	220.10	220.08	34	48
69	Deoxycytidine	228.10	226.08	15	29
70	Ribose-5-phosphate	231.03	229.01	85	99
71	Gly-Tyr	239.10	237.09	57	71
72	Thymidine	243.10	241.08	4	18
73	Cytidine	244.09	242.08	24	38
74	Uridine	245.08	243.06	10	24
75	Biotin	245.10	243.08	18	32
76	Isopentenyl pyrophosphate (IPP)	x	246.00	60	74
77	Dimethylallyl pyrophosphate	x	246.00	60	74

ID	Name	m/z (pos)	m/z (neg)	Start time (min)	End time (min)
78	Pyridoxal-5-phosphate	248.03	246.02	63	77
79	Deoxyadenosine	252.11	250.09	5	19
80	Fructose-6-phosphate	261.04	259.02	90	112
81	Glucose-1-phosphate	261.04	259.02	90	112
82	Glucose-6-phosphate	261.04	259.02	90	112
83	O-P-L-Tyrosine	262.05	260.03	112	126
84	Thiamin	265.11	263.10	32	46
85	Adenosine	268.10	266.09	8	31
86	Deoxyguanosine	268.10	266.09	8	31
87	Inosine	269.09	267.07	17	31
88	6-Phosphogluconic acid	277.03	275.02	106	120
89	Glutathione (red)	308.09	306.08	76	90
90	N-Acetylneuraminic acid	310.11	308.10	60	74
91	CMP	324.06	322.04	91	105
92	UMP	325.04	323.03	78	92
93	cAMP	330.06	328.05	35	49
94	Adenosine-2'(3')- monophosphoric acid	330.06	328.05	72	86
95	dAMP	332.08	330.06	67	81
96	Glucose-1,6-bisphosphate	341.00	338.99	120	135
97	Thiamine monophosphate	345.08	343.06	131	145
98	cGMP	346.05	344.04	49	63
99	AMP	348.07	346.06	73	87
100	IMP	349.05	347.04	81	95
101	GMP	364.07	362.05	91	105
102	Riboflavin	377.14	375.13	7	21
103	S-Adenosyl methionine	399.14	x	121	135
104	UDP	405.01	402.99	93	107
105	Thiamine pyrophosphate	425.04	423.03	130	144
106	ADP	428.03	426.02	89	103
107	Folic acid	x	440.13	87	101
108	GDP	444.03	442.02	105	119
109	Flavin mononucleotide	457.11	455.10	62	76
110	dCTP	468.10	465.98	107	121
111	dTTP	x	480.98	94	108
112	CTP	484.10	x	112	126
113	UTP	484.98	482.96	104	118
114	dATP	492.01	489.99	96	110
115	ATP	508.00	505.99	100	122
116	dGTP	508.00	505.99	100	122
117	Tyr-Tyr-Tyr	508.21	506.19	25	39

ID	Name	m/z (pos)	m/z (neg)	Start time (min)	End time (min)
118	GTP	524.10	521.98	114	128
119	CDP-ME	x	523.08	82	95
120	UDP- <i>D</i> -glucose	567.06	565.05	87	101
121	UDP- <i>D</i> -glucuronic acid	581.04	579.03	103	117
122	CDP-MEP	x	603.05	102	116
123	Glutathione (ox)	613.16	611.14	120	134
124	NAD ⁺	664.12	662.10	81	95
125	NADH	666.13	664.12	69	83
126	NADP ⁺	744.08	742.07	111	125
127	NADPH	746.09	744.08	101	115
128	CoA	768.12	766.11	83	97
129	FAD	786.16	784.15	60	74

4.1.5 Data analysis

Raw files of metabolite experiments are processed with the software Sieve2.0 (Thermo Fisher) and are therefore uploaded and processed with settings listed in Table II-5. Generated frame tables of metabolites are exported and further evaluated with Perseus 1.5.1.6. In Perseus, two matrixes are determined: in the first matrix, the \log_2 of the extracted intensities to the corresponding m/z-value is determined, and in the second matrix, the two-sample t-test (Benjamini-Hochberg FDR) with a FDR of 0.05 is performed.

Table II-5: Settings and parameters for the targeted and untargeted metabolite profiling experiment using the software Sieve2.0 (Thermo Fisher). Listed below are the global settings, alignment parameters, frame parameters, global identification parameters and accurate mass identification parameters of the Sieve workflow.

Global settings	
Algorithm	FRAME
Experiment	Metabolomics
Experiment type	A vs. B
Max threads:	4
MZ start	60
MZ stop	1000
RT start	5
RT stop	140
Alignment parameters	
Alignment bypass	False
Alignment intensity	10,000
Correlation bin width	1
Tile size	300
Frame parameters	
Maximum frames	10,000
MZ width ppm	10
PR max charge	5
RT width	10
Threshold	10,000
Global identification parameters	
Assumed charge	-1/+1
Maximum IDs	10,000
Search source	DBLOOKUP
Accurate mass identification parameters	
Adduct	-nH/+nH
DB LookupMethod	COMPMW
MZ tolerance	10

4.2 Targeted metabolite profiling in wild-type *S. aureus* USA 300 to elucidate the involvement of the ribokinase ThiD within the pyridoxal salvage pathway

4.2.1 Sample preparation for the analysis of extracellular PL levels in *S. aureus* extracts

Staphylococcus aureus USA300 and the transposon mutants TnThiD (PLK) (gene: SA300_0562) and TnPdxS (gene: SA300_0504) were obtained from the Nebraska Transposon Mutant Library within the Network on Antimicrobial Resistance in *S. aureus* (NARSA). Chemically defined medium (CDM) is prepared according to *Liebeke et al*²²⁸ and sterilized via filtration. Overnight cultures of wild-type *S. aureus* USA 300, TnThiD (PLK) and TnPdxS grown in brain-heart broth medium (BHB) are pelleted and resuspended in 1 mL CDM. 50 mL CDM are then inoculated with wild-type *S. aureus* and transposon mutants to an initial OD₆₀₀ of 0.08 and grown for 24 h at 37 °C with shaking under aerobic conditions. After reaching a final OD₆₀₀ of 1.2, cells are pelleted and the growth medium is filtered through a 0.2 µm sterile membrane. The medium is then freeze-dried, redissolved in 5 mL deionised H₂O and filtered through an Amicon Ultra 3K centrifugal filter. The samples are again freeze-dried, and redissolved in 1 mL deionised H₂O.

4.2.2 Analytical platform for the LC-MS analysis of PL in *S. aureus* extracts

The samples (5 µL injection volume) are analyzed by LC-HESI-MS on a Thermo Finnigan LTQ FT-ICR and eluted using a Dionex Ultimate 3000 over a ZIC®-HILIC column (250 x 2.1 mm, 5 µm particle size). The column temperature is maintained at 30 °C. Eluting buffers are A (10 mM ammonium formate in 95:5 H₂O:MeCN, pH 6.2) and B (10 mM ammonium formate in 5:95 H₂O:MeCN, pH 7.2). The gradient used is 0 → 40 min; 90% → 60% B (0.1 mL/min); 40 → 41 min; 60% → 5% B (0.1 mL/min); 41 → 55 min; 5% B (0.1 mL/min); 55 → 56 min; 5% → 100% B (0.1 mL/min), 56 → 72 min; 100% B (0.1 mL/min); 72 → 73 min; 100% B (0.9 mL/min); 73 → 78 min; 100% B (0.9 mL/min); 78 → 79 min; 100% → 90% B (0.1 mL/min); 79 → 90 min; 90% B (0.1 mL/min). The eluent is directly introduced into the ion source without prior splitting. Ions are scanned in positive polarity mode. Relative quantification is achieved using single ion monitoring (SIM) with a resolution of 25,000. Identification and validation of pyridoxal is performed by tandem MS (CID). HESI source parameters are: spray voltage, 3.5 kV; capillary temperature, 60 °C; capillary voltage, 48 V; and tube lens, 60 V

4.3 Subclass-specific capture of protein-reactive natural products with customized nucleophilic probes

4.3.1 General methods

Reversed-phase HPLC analysis is performed on a Waters 2695 separation module, equipped with a Waters PDA 2996 and a Waters XBridge C18 column (3.5 μm , 4.6 \times 100 mm, flow = 1.2 mL/min). For preparative scale RP-HPLC separation a Waters 2545 quaternary gradient module in combination with a Waters PDA 2998 and a Waters XBridge C18 (5.0 μm , 30 \times 150 mm, flow = 50 mL/min) column or a YMC Triart C18 (3.5 μm , 10 \times 250 mm, flow = 10 mL/min) column is used. The mobile phase for elution consisted of a gradient mixture of 0.1% (v/v) TFA in water (buffer A, HPLC grade) and 0.1% (v/v) TFA in acetonitrile (buffer B, HPLC grade) unless otherwise noted.

Reversed-phase HPLC-ESI-HR-MS analysis is performed on a Thermo Finnigan LTQ FT-ICR equipped with a Dionex Ultimate 3000 separation module eluting on a Waters XBridge C18 column (3.5 μm , 4.6 \times 100 mm, flow = 1.1 mL/min). The column temperature is maintained at 30 °C. The mobile phase for elution consisted of a gradient mixture of 0.1% (v/v) formic acid in water (buffer A, HPLC-MS grade) and 0.1% (v/v) formic acid in acetonitrile:water 90:10 (buffer B, HPLC-MS grade).

Hydrophilic interaction chromatography (HILIC) HPLC-HESI-HR-MS analysis is performed on a Thermo Finnigan LTQ FT-ICR and are eluted by a Dionex Ultimate 3000 over a ZIC®-HILIC column (2.1 \times 250 mm, 5 μm particle size). The column temperature is maintained at 30 °C. Eluting buffers are buffer A (10 mM ammonium formate in 95:5 water:acetonitrile, pH 6.2) and buffer B (10 mM ammonium formate in 5:95 water:acetonitrile, pH 7.2).

High-resolution mass spectra are obtained on a Thermo Scientific LTQ-FT Ultra via electrospray ionization (ESI/HESI-MS). ESI source parameters for positive polarity mode are: spray voltage, 4.0 kV; capillary temperature, 275 °C; capillary voltage, 48 V; and tube lens, 120 V. ESI source parameters for negative polarity mode are: spray voltage, 3.5 kV; capillary temperature, 275 °C; capillary voltage, -40 V; and tube lens, -120 V. HESI source parameters for negative polarity mode are: spray voltage, 3.5 kV; capillary temperature, 275 °C; capillary voltage, -35 V; vaporizer temperature, 60 °C; and tube lens, -60 V.

4.3.2 Evaluation of cap probes with model electrophiles

Since not otherwise noted, the evaluation experiments for the reactivity screening of nucleophilic probes **Cap1** – **Cap6** is conducted as follows:

a. Disulfide reduction of **Cap5** and **Cap6**:

Disulfane containing nucleophilic probes **Cap5** and **Cap6** are reduced prior to their application in screening experiments: 50 μL of disulfane stock solution (10 mM in DMSO) are diluted in 100 μL acetonitrile and treated with 50 μL TCEP solution (100 mM in aqueous acetate buffer, pH 4.0) for 3 h at 30 °C. Filtration over a small pad of silica with 500 μL acetonitrile and following evaporation *in vacuo* provides an aqueous sulfhydryl containing DMSO stock solution, applicable in the subsequent experiments.

b. Organic solvent medium:

10 μL of nucleophilic probe stock solution (5 mM in DMSO) are incubated with 10 μL of electrophile stock solution (5 mM in DMSO) in 20 μL acetonitrile at 30 °C for a time course of 4 to 24 h as indicated in the experiments' descriptions. The reactions are then analyzed by analytical reversed-phase HPLC (C18 3.5 μm , 4.6 \times 100 mm column, method: gradient 2% B \rightarrow 100% B over 30 min) and mass spectrometry (FLEET-LCQ-MS).

c. PBS-buffered aqueous medium:

10 μL of nucleophilic probe stock solution (50 mM in DMSO) are diluted with 90 μL PBS buffer (pH 7.4) and treated with 1.25 μL of electrophile stock solution (20 mM in DMSO) to afford a final ratio of 20:1 NASM probe to electrophilic starting material. After incubation at 30 °C for 18 h the reactions are analyzed by analytical reversed-phase HPLC (C18 3.5 μm , 4.6 \times 100 mm column, method: gradient 2% B \rightarrow 100% B over 25 min) and mass spectrometry (FLEET-LCQ-MS).

4.3.3 Reactivity of Cap6

The following electrophilic compounds for **Cap6** evaluation are synthesized according to the published literature procedure: acivicin²²⁹, sultone²³⁰. Showdomycin **Sh** and **FM209** are synthesized according to *Rudolf et al.*²³¹ Penicillin G, 4-(4-Ethyl-1-piperazinyl)-4-oxo-2-butenoic acid, 1,4-phenylendiacrylic acid and phosphomycin are obtained from Sigma-Aldrich in reagent grade purity, naphthalene-1-thiol is obtained from Alfa-Aesar.

The evaluation experiments of the reactivity of **Cap6** towards the electrophilic compounds are conducted as follows:

a. Disulfide reduction of Cap6:

Disulfane containing nucleophilic probe **Cap6** is reduced prior to its application under the following conditions: 50 μL of **Cap6** stock solution (50 mM in DMSO) are diluted in 100 μL acetonitrile and treated with 50 μL TCEP solution (100 mM in aqueous acetate buffer, pH 4.0) for 3 h at 30 °C (reduced **Cap6** stock solution).

b. Evaluation of Cap6 with synthetically prepared showdomycin Sh:

40 μL of showdomycin **Sh** stock solution (625 μM in PBS buffer, pH 7.4) are treated with 10 μL of reduced **Cap6** stock solution (25 mM) which is filtered prior over a small pad of silica with 500 μL acetonitrile to afford a final ratio of 10:1 NASM probe to electrophilic starting material. Samples are evaporated to dryness, redissolved in 50 μL of 20% DMSO in PBS buffer (pH 7.4) and filtered through a 0.45 μm centrifugal filter. Reactions are then analyzed by analytical reversed-phase HPLC (C18 3.5 μm , 4.6 \times 100 mm column, method: gradient 2% B \rightarrow 100% B over 25 min) and reversed-phase HPLC-ESI-HRMS (C18 3.5 μm , 4.6 \times 100 mm column, method: gradient 2% B \rightarrow 100% B over 30 min).

c. Evaluation of Cap6 with phosphomycin:

40 μL of phosphomycin stock solution (625 μM in PBS buffer, pH 7.4) are treated with 10 μL of reduced **Cap6** stock solution (25 mM) which is filtered prior over a small pad of silica with 500 μL acetonitrile to afford a final ratio of 10:1 NASM probe to electrophilic starting material. Samples are evaporated to dryness, redissolved in 50 μL of 20% DMSO in PBS buffer (pH 7.4) and filtered through a 0.45 μm centrifugal filter. Reactions are then analyzed by analytical reversed-phase HPLC (C18 3.5 μm , 4.6 \times 100 mm column, method: gradient 2% B \rightarrow 100% B over 25 min) and reversed-phase HPLC-ESI-HRMS (C18 3.5 μm , 4.6 \times 100 mm column, method: gradient 2% B \rightarrow 100% B over 30 min).

d. Comparative elucidation of conversion of showdomycin **Sh with **Cap6** and naphthalene-1-thiol:**

40 μL of showdomycin **Sh** stock solution (625 μM in PBS buffer, pH 7.4) are treated with 10 μL of reduced **Cap6** stock solution (25 mM) which is filtered prior over a small pad of silica with 500 μL acetonitrile to afford a final ratio of 10:1 NASM probe to electrophilic starting material. Samples are evaporated to dryness and redissolved in 50 μL 20% DMSO in PBS buffer (pH 7.4) and filtered through a 0.45 μm centrifugal filter. Showdomycin conversion and product formation are then analyzed by HILIC-HPLC-HESI-HRMS (ZIC®-HILIC, 5 μm , 2.1 x 250 mm column, method: gradient 100% B \rightarrow 5% B over 30 min).

e. Reaction control of **Cap6 in a mixture of 8 different electrophiles:**

To a solution of 32 μL PBS buffer (pH 7.4) treated with 1 μL of a solution of showdomycin **Sh**, phosphomycin, penicillin G, acivicin, 1,4-phenylenediacrylic acid, sultone, **FM209** and 4-(4-Ethyl-1-piperazinyl)-4-oxo-2-butenoic acid (25 mM each) are added 10 μL of **Cap6** stock solution (25 mM) which is filtered prior over a small pad of silica with 500 μL acetonitrile to afford a final ratio of 10:1 NASM probe to electrophilic starting materials. Samples are evaporated to dryness and redissolved in 50 μL of 20% DMSO/PBS buffer solution (pH 7.4). Reactions are filtered through a 0.45 μm centrifugal filter and then analyzed by reversed-phase HPLC-ESI-HRMS (C18 3.5 μm , 4.6 x 100 mm column, method: gradient 2% B \rightarrow 100% B over 30 min) and HILIC-HPLC-HESI-MS (ZIC®-HILIC, 5 μm , 2.1 x 250 mm column, method: gradient 100% B \rightarrow 5% B over 30 min).

5. Appendices

5.1 Establishment of a targeted and untargeted LC-MS-based metabolite profiling platform for the analysis of metabolite levels in bacteria

Table II-A1: Metabolites measured in the established in-house database in negative ionization mode. Listed is the ID of each metabolite, the name, molecular formula, exact mass in negative mode, retention time in the established multi-step gradient and the first three daughter ions resulting from the CID of the precursor ion.

ID	Metabolite	Molecular formula	M (exact)	RT (min)	D1	D2	D3
Metabolome A1							
50	L-Arginine	C ₆ H ₁₄ N ₄ O ₂	173.10	127	131.14	x	x
19	L-Aspartic acid	C ₄ H ₇ NO ₄	132.03	87	88.03	114.98	114.06
10	L-Cysteine	C ₃ H ₇ NO ₂ S	120.01	51	x	x	x
31	L-Glutamic acid	C ₅ H ₉ NO ₄	146.04	86	128.06	102.14	x
29	L-Glutamine	C ₅ H ₁₀ N ₂ O ₃	145.06	68	127.05	109.16	x
39	L-Histidine	C ₆ H ₉ N ₃ O ₂	154.06	74	137.07	136.13	93.12
15	L-Isoleucine	C ₆ H ₁₃ NO ₂	130.09	36	x	x	x
14	L-Leucine	C ₆ H ₁₃ NO ₂	130.09	32	x	x	x
Metabolome A2							
30	L-Lysine	C ₆ H ₁₄ N ₂ O ₂	145.10	136	97.02	99.05	x
32	L-Methionine	C ₅ H ₁₁ NO ₂ S	148.04	40	100.09	x	x
42	L-Phenylalanine	C ₉ H ₁₁ NO ₂	164.07	28	147.13	x	x
3	L-Proline	C ₅ H ₉ NO ₂	114.06	x	x	x	x
2	L-Serine	C ₃ H ₇ NO ₃	104.04	70	74.00	72.00	x
8	L-Threonine	C ₄ H ₉ NO ₃	118.05	60	74.07	72.03	x
64	L-Tryptophan	C ₁₁ H ₁₂ N ₂ O ₂	203.08	30	159.09	116.10	142.18
55	L-Tyrosine	C ₉ H ₁₁ NO ₃	180.07	46	163.18	93,11	x
7	L-Valine	C ₅ H ₁₁ NO ₂	116.07	x	x	x	x
Metabolome B1							
99	AMP	C ₁₀ H ₁₄ N ₅ O ₇ P	346.06	80	211.08	151.06	x
91	CMP	C ₉ H ₁₄ N ₃ O ₈ P	322.04	98	211.16	279.18	181.20
101	GMP	C ₁₀ H ₁₄ N ₅ O ₈ P	362.05	98	211.11	x	x
92	UMP	C ₉ H ₁₃ N ₂ O ₉ P	323.03	85	211,08	280,13	x
106	ADP	C ₁₀ H ₁₅ N ₅ O ₁₀ P ₂	426.02	96	328.15	134.07	408.28
115	ATP	C ₁₀ H ₁₆ N ₅ O ₁₃ P ₃	505.99	107	408.20	273.07	x
112	CTP	C ₉ H ₁₆ N ₃ O ₁₄ P ₃	481.98	119	384.15	159.11	x
108	GDP	C ₁₀ H ₁₅ N ₅ O ₁₁ P ₂	442.02	112	344.18	150.11	x
118	GTP	C ₁₀ H ₁₆ N ₅ O ₁₄ P ₃	521.98	121	424.19	273,20	x

104	UDP	C ₉ H ₁₄ N ₂ O ₁₂ P ₂	402.99	100	305.13	385.16	291.15
Metabolome B2							
113	UTP	C ₉ H ₁₅ N ₂ O ₁₅ P ₃	482.96	111	385.30	273.17	x
128	CoA	C ₂₁ H ₃₆ N ₇ O ₁₆ P ₃ S	766.11	90	419.22	408.27	426.27
80	Fructose-6-phosphate	C ₆ H ₁₃ O ₉ P	259.02	97	96.94	168.99	78.95
82	Glucose-6-phosphate	C ₆ H ₁₃ O ₉ P	259.02	105	96.94	199.04	168.98
129	FAD	C ₂₇ H ₃₃ N ₉ O ₁₅ P ₂	784.15	67	437.23	346.19	x
124	NAD ⁺	C ₂₁ H ₂₈ N ₇ O ₁₄ P ₂	662.10	88	540.38	x	x
125	NADH	C ₂₁ H ₂₉ N ₇ O ₁₄ P ₂	664.12	76	408.16	397.18	346.25
126	NADP ⁺	C ₂₁ H ₂₉ N ₇ O ₁₇ P ₃	742.07	118	620.34	x	x
127	NADPH	C ₂₁ H ₃₀ N ₇ O ₁₇ P ₃	744.08	108	426.18	646.36	x
Metabolome C1							
9	<i>L</i> -Homoserine	C ₄ H ₉ NO ₃	118.05	64	98.03	100.08	x
21	Adenine	C ₅ H ₅ N ₅	134.05	14	107.07	92.12	x
85	Adenosine	C ₁₀ H ₁₃ N ₅ O ₄	266.09	15	134.12	x	x
73	Cytidine	C ₉ H ₁₃ N ₃ O ₅	242.08	31	109.07	152.11	x
79	2-Deoxyadenosine	C ₁₀ H ₁₃ N ₅ O ₃	250.09	12	134.17	x	x
74	Uridine	C ₉ H ₁₂ N ₂ O ₆	243.06	17	200.11	x	x
100	IMP	C ₁₀ H ₁₃ N ₄ O ₈ P	347.04	88	211.11	x	x
53	<i>D</i> -Glucose	C ₆ H ₁₂ O ₆	179.06	50	89.09	x	x
68	<i>N</i> -Acetyl-glucosamine	C ₈ H ₁₅ NO ₆	220.08	41	119.20	x	x
14	<i>L</i> Leucine	C ₆ H ₁₃ NO ₂	130.08	32	84.11	x	x
Metabolome C2							
45	Phosphoenolpyruvic acid	C ₃ H ₅ O ₆ P	166.98	108	78.92	x	x
25	HMP	C ₆ H ₉ N ₃ O	138.07	x	x	x	x
11	Nicotinic acid	C ₆ H ₅ NO ₂	122.02	26	78.05	x	x
67	Pantothenic acid	C ₉ H ₁₇ NO ₅	218.10	41	87.97	x	x
44	Pyridoxal	C ₈ H ₉ NO ₃	166.05	13	138.10	x	x
78	Pyridoxal-5-phosphate	C ₈ H ₁₀ NO ₆ P	246.02	70	x	x	X
47	Pyridoxine	C ₈ H ₁₁ NO ₃	168.07	13	150.14	138.17	x
46	Pyridoxamine	C ₈ H ₁₂ N ₂ O ₂	167.08	67	x	x	x
84	Thiamine	C ₁₂ H ₁₇ N ₄ OS+	263.10	40	233.18	147.20	x
Metabolome C3							
60	Citric acid	C ₆ H ₈ O ₇	191.02	93	111.00	173.07	x
28	α-Ketoglutaric acid	C ₅ H ₆ O ₅	145.01	76	100.97	x	x
123	Glutathione (ox)	C ₂₀ H ₃₂ N ₆ O ₁₂ S ₂	611.14	127	306.19	272.21	338.23
89	Glutathione (red)	C ₁₀ H ₁₇ N ₃ O ₆ S	306.08	83	254.17	272.23	288.20
23	4-Hydroxybenzoic acid	C ₇ H ₆ O ₃	137.02	16	93.00	x	x
52	Ascorbic acid	C ₆ H ₈ O ₆	175.02	57	115.06	x	x
Metabolome D1							

1	Hydroxybutyric acid	C ₄ H ₈ O ₃	103.04	27	57.00	85.15	x
61	Glucuronic acid	C ₆ H ₁₀ O ₇	193.03	79	131.07	113.12	175.16
69	2-Hydroxy-3-methylbutyric acid	C ₅ H ₁₀ O ₃	117.06	16	70.99	x	x
69	2-Deoxycytidine	C ₉ H ₁₃ N ₃ O ₄	226.08	22	93.05	183.13	135.11
117	Tyr-Tyr-Tyr	C ₂₇ H ₂₉ N ₃ O ₇	506.19	32	462.44	400.40	180.22
65	2-Deoxyribose-5-phosphat	C ₅ H ₁₁ O ₇ P	213.02	83	96.99	195.08	x
13	Pyroglutamic acid	C ₅ H ₇ NO ₃	128.04	50	82.08	84.05	x
Metabolome D2							
5	Succinic acid	C ₄ H ₆ O ₄	117.02	66	72.93	99.10	x
87	Inosine	C ₁₀ H ₁₂ N ₄ O ₅	267.07	24	135.05	x	x
43	<i>L</i> -β-Phenyllactic acid	C ₉ H ₁₀ O ₃	165.06	12	147.29	x	x
121	UDP- <i>D</i> -Glucuronic acid	C ₁₅ H ₂₂ N ₂ O ₁₈ P ₂	579.03	110	403.19	323.17	255.11
97	Thiamine monophosphate	C ₁₂ H ₁₇ N ₄ O ₄ PS	343.06	138	245.16	218.25	x
54	<i>Myo</i> -inositol	C ₆ H ₁₂ O ₆	179.06	68	161.08	117.08	x
Metabolome D3							
86	Deoxyguanosine	C ₁₀ H ₁₃ N ₅ O ₄	266.09	24	150.12	x	x
72	Thymidine	C ₁₀ H ₁₄ N ₂ O ₅	241.08	11	125.06	151.07	x
48	Dihydroxyacetonphosphate	C ₃ H ₇ O ₆ P	168.99	90	96.91	x	x
120	UDP- <i>D</i> -glucose	C ₁₅ H ₂₄ N ₂ O ₁₇ P ₂	565.05	94	323.20	x	x
49	Glycerol-3-phosphate	C ₃ H ₉ O ₆ P	171.01	87	78.96	x	x
88	6-Phosphogluconic acid	C ₆ H ₁₃ O ₁₀ P	275.02	113	257.14	96.96	x
70	Ribose-5-phosphate	C ₅ H ₁₁ O ₈ P	229.01	92	96.97	78.95	x
17	<i>L</i> -Asparagine	C ₄ H ₈ N ₂ O ₃	131.05	70	113.06	114.06	x
18	Ornithine	C ₅ H ₁₂ N ₂ O ₂	131.08	136	85.02	x	x
Metabolome D4							
27	Spermidine	C ₇ H ₁₉ N ₃	144.15	x	x	x	x
16	Glutaric acid	C ₅ H ₈ O ₄	131.03	75	87.04	113.13	x
9	<i>L</i> -Homoserine	C ₄ H ₉ NO ₃	118.05	64	97.98	100.04	x
100	IMP	C ₁₀ H ₁₃ N ₄ O ₈ P	347.04	87	211.07	x	x
81	Glucose-1-phosphate	C ₆ H ₁₃ O ₉ P	259.02	98	241.08	97.00	x
96	Glucose-1,6-bisphosphate	C ₆ H ₁₄ O ₁₂ P ₂	338.99	127	241.28	x	x
105	Thiamine pyrophosphate	C ₁₂ H ₁₉ N ₄ O ₇ P ₂ S	423.03	137	176.96	302.08	x
Metabolome D5							
41	Phenylpyruvic acid	C ₉ H ₈ O ₃	163.04	10	91.03	x	x
4	Uracil	C ₄ H ₄ N ₂ O ₂	111.02	12	x	x	x
12	Thymine	C ₅ H ₆ N ₂ O ₂	125.04	10	x	x	x
60	Isocitric acid	C ₆ H ₈ O ₇	191.02	96	173.10	111.03	x
90	<i>N</i> -Acetylneuraminic acid	C ₁₁ H ₁₉ NO ₉	308.10	67	290.21	170.06	119.18
71	Gly-Tyr	C ₁₁ H ₁₄ N ₂ O ₄	237.09	64	180.13	193.19	x

24	Acetylphosphate	C ₂ H ₅ O ₅ P	138.98	94	78.99	x	x
79	Deoxyadenosine	C ₁₀ H ₁₃ N ₅ O ₃	250.09	12	134.10	x	x
95	dAMP	C ₁₀ H ₁₄ N ₅ O ₆ P	330.06	74	195.04	177.12	x
Metabolome D6							
33	Guanine	C ₅ H ₅ N ₅ O	150.04	25	133.10	x	x
35	Xanthine	C ₅ H ₄ N ₄ O ₂	151.03	17	108.02	x	x
94	Adenosin-2,'(3')- monophosphoric acid	C ₁₀ H ₁₂ N ₅ O ₆ P	328.05	79	134.16	x	x
82	Glucose-6-phopsphate	C ₆ H ₁₃ O ₉ P	259.02	104	96.93	199.04	x
Metabolome E1							
62	Gluconic acid	C ₆ H ₁₂ O ₇	195.05	66	129.09	177.09	x
57	O-P-Serine	C ₃ H ₈ NO ₆ P	184.00	118	96.97	166.05	79.06
51	Citrulline	C ₆ H ₁₃ N ₃ O ₃	174.09	76	131.12	x	x
5	Succinic acid	C ₄ H ₆ O ₄	117.02	x	x	x	x
109	Flavin mononucleotide	C ₁₇ H ₂₁ N ₄ O ₉ P	455.10	69	213.16	199.10	x
98	cGMP	C ₁₀ H ₁₂ N ₅ O ₇ P	344.04	56	150.07	133.18	x
26	4-Guanidinobutanoic acid	C ₅ H ₁₁ N ₃ O ₂	144.08	67	102.06	127.13	x
58	3-Phosphoglyceric acid	C ₃ H ₇ O ₇ P	184.99	106	96.93	167.13	x
63	O-P-Threonine	C ₄ H ₁₀ NO ₆ P	198.02	111	97.00	79.01	x
83	O-P-L-Tyrosine	C ₉ H ₁₂ N ₁ O ₆ P	260.03	119	79.03	x	x
Metabolome E2							
59	<i>N</i> (6)-Acetyl- <i>L</i> -Lysine	C ₈ H ₁₆ N ₂ O ₃	187.11	55	145.18	x	x
532	<i>D</i> -Fructose	C ₆ H ₁₂ O ₆	179.06	45	89.19	119.13	x
13	Pyroglutamic acid	C ₅ H ₇ NO ₃	128.04	49	82.11	84.02	x
40	Allantoin	C ₄ H ₆ N ₄ O ₃	157.04	29	114.11	139.99	x
37	2,3-Dihydroxybenzoic acid	C ₇ H ₆ O ₄	153.02	15	109.03	x	x
16	Glutaric acid	C ₅ H ₈ O ₄	131.03	79	87.03	113.00	x
20	<i>L</i> -Malic acid	C ₄ H ₆ O ₅	133.01	84	115.03	x	x
Metabolome F1							
107	Folic acid	C ₁₉ H ₁₉ N ₇ O ₆	440.13	94	311.24	422.36	x
22	Hypoxanthine	C ₅ H ₄ N ₄ O	135.03	17	92.02	x	x
93	cAMP	C ₁₀ H ₁₂ N ₅ O ₆ P	328.05	42	134.06	x	x
114	dATP	C ₁₀ H ₁₆ N ₅ O ₁₂ P ₃	489.99	103	392.15	158.98	x
110	dCTP	C ₉ H ₁₆ N ₃ O ₁₃ P ₃	465.98	114	368.12	158.99	448.13
116	dGTP	C ₁₀ H ₁₆ N ₅ O ₁₃ P ₃	505.99	115	408.15	488.13	x
111	dTTP	C ₁₀ H ₁₇ N ₂ O ₁₄ P ₃	480.98	101	383.17	159.10	463.10
Miscellaneous							
75	Biotin	C ₁₀ H ₁₆ N ₂ O ₃ S	243.08	25	200.20	199.23	166.21
102	Riboflavin	C ₁₇ H ₂₀ N ₄ O ₆	375.13	14	255.22	x	x

Table II-A2: Metabolites measured in the established in-house database in positive ionization mode. Listed is the ID of each metabolite, the name, molecular formula, exact mass in negative mode, retention time in the established multi-step gradient and the first three daughter ions resulting from the CID of the precursor ion.

ID	Name	Molecular formula	M (exact)	RT (min)	D1	D2	D3
Metabolome A1							
50	L-Arginine	C ₆ H ₁₄ N ₄ O ₂	175.12	127	158.09	157.08	130.10
19	L-Aspartic acid	C ₄ H ₇ NO ₄	134.04	87	116.01	88.02	x
10	L-Cysteine	C ₃ H ₇ NO ₂ S	122.03	x	x	x	x
31	L-Glutamic acid	C ₅ H ₉ NO ₄	148.06	87	129.91	101.93	83.94
29	L-Glutamine	C ₅ H ₁₀ N ₂ O ₃	147.08	68	130.01	x	
39	L-Histidine	C ₆ H ₉ N ₃ O ₂	156.08	73	109.99	x	
15	L-Isoleucine	C ₆ H ₁₃ NO ₂	132.10	36	85.93	x	
14	L-Leucine	C ₆ H ₁₃ NO ₂	132.10	32	86.00	x	x
Metabolome A2							
30	L-Lysine	C ₆ H ₁₄ N ₂ O ₂	147.11	137	130.08	129.26	x
32	L-Methionine	C ₅ H ₁₁ NO ₂ S	150.06	40	132.96	103.92	x
42	L-Phenylalanine	C ₉ H ₁₁ NO ₂	166.09	28	120.12	x	x
3	L-Proline	C ₅ H ₉ NO ₂	116.07	50	69.94	x	x
2	L-Serine	C ₃ H ₇ NO ₃	106.05	71	60.03	x	x
8	L-Threonine	C ₄ H ₉ NO ₃	120.07	62	x	x	x
64	L-Tryptophan	C ₁₁ H ₁₂ N ₂ O ₂	205.10	30	188.04	x	x
55	L-Tyrosine	C ₉ H ₁₁ NO ₃	182.08	46	165.04	136.10	x
7	L-Valine	C ₅ H ₁₁ NO ₂	118.09	48	71.92	x	x
Metabolome B1							
99	AMP	C ₁₀ H ₁₄ N ₅ O ₇ P	348.07	80	135.93	x	x
91	CMP	C ₉ H ₁₄ N ₃ O ₈ P	324.06	98	111.91	x	x
101	GMP	C ₁₀ H ₁₄ N ₅ O ₈ P	364.07	98	151.98	x	x
92	UMP	C ₉ H ₁₃ N ₂ O ₉ P	325.04	85	96.96	212.99	x
106	ADP	C ₁₀ H ₁₅ N ₅ O ₁₀ P ₂	428.04	97	136.02	348.10	x
115	ATP	C ₁₀ H ₁₆ N ₅ O ₁₃ P ₃	508.00	107	410.12	348.16	x
112	CTP	C ₉ H ₁₆ N ₃ O ₁₄ P ₃	483.99	119	324.12	208.03	306.15
108	GDP	C ₁₀ H ₁₅ N ₅ O ₁₁ P ₂	444.03	112	152.04	248.21	x
118	GTP	C ₁₀ H ₁₆ N ₅ O ₁₄ P ₃	524.00	121	152.02	248.20	426.19
104	UDP	C ₉ H ₁₄ N ₂ O ₁₂ P ₂	405.01	101	275.06	307.03	177.04
Metabolome B2							
113	UTP	C ₉ H ₁₅ N ₂ O ₁₅ P ₃	484.98	111	x	x	x
128	CoA	C ₂₁ H ₃₆ N ₇ O ₁₆ P ₃ S	768.12	90	428.20	261.13	x
80	Fructose-6-phosphate	C ₆ H ₁₃ O ₉ P	261.04	97	242.96	244.08	224.76

82	Glucose-6-phosphate	C ₆ H ₁₃ O ₉ P	261.04	105	243.03	224.19	x
129	FAD	C ₂₇ H ₃₃ N ₉ O ₁₅ P ₂	786.16	67	348.15	439.24	x
124	NAD ⁺	C ₂₁ H ₂₈ N ₇ O ₁₄ P ₂	664.12	88	524.18	542.01	x
125	NADH	C ₂₁ H ₂₉ N ₇ O ₁₄ P ₂	666.13	76	649.25	347.91	x
126	NADP ⁺	C ₂₁ H ₂₉ N ₇ O ₁₇ P ₃	744.08	118	604.02	621.99	x
127	NADPH	C ₂₁ H ₃₀ N ₇ O ₁₇ P ₃	746.10	107	729.18	x	x
Metabolome C1							
9	L-Homoserine	C ₄ H ₉ NO ₃	120.07	64	73.89	101.99	x
21	Adenine	C ₅ H ₅ N ₅	136.06	13	x	x	x
85	Adenosine	C ₁₀ H ₁₃ N ₅ O ₄	268.10	15	136.01	x	x
73	Cytidine	C ₉ H ₁₃ N ₃ O ₅	244.09	31	111.88	x	x
79	2-Deoxyadenosine	C ₁₀ H ₁₃ N ₅ O ₃	252.11	13	135.96	x	x
74	Uridine	C ₉ H ₁₂ N ₂ O ₆	245.08	18	112.94	x	x
100	IMP	C ₁₀ H ₁₃ N ₄ O ₈ P	349.05	88	136.99	233.19	x
53	D-Glucose	C ₆ H ₁₂ O ₆	181.07	x	x	x	x
68	N-Acetyl-glucosamine	C ₈ H ₁₅ NO ₆	222.10	41	203.99	x	x
14	L-Leucine	C ₆ H ₁₃ NO ₂	132.10	32	86.00	x	x
Metabolome C2							
45	Phosphoenolpyruvic acid	C ₃ H ₅ O ₆ P	168.99	x	x	x	x
25	HMP	C ₆ H ₉ N ₃ O	140.08	18	121.98	81.01	x
11	Nicotinic acid	C ₆ H ₅ NO ₂	124.04	26	80.08	x	x
67	Pantothenic acid	C ₉ H ₁₇ NO ₅	220.12	41	202.17	89.98	184.10
44	Pyridoxal	C ₈ H ₉ NO ₃	168.07	13	150.06	x	x
78	Pyridoxal-5-phosphate	C ₈ H ₁₀ NO ₆ P	248.03	70	x	x	x
47	Pyridoxine	C ₈ H ₁₁ NO ₃	170.08	13	152.03	x	x
46	Pyridoxamine	C ₈ H ₁₂ N ₂ O ₂	169.10	68	x	x	x
84	Thiamine	C ₁₂ H ₁₇ N ₄ OS+	265.11	39	122.10	144.03	
Metabolome C3							
60	Citric acid	C ₆ H ₈ O ₇	193.03	x	x	x	x
28	α-Ketoglutaric acid	C ₅ H ₆ O ₅	147.03	x	x	x	x
123	Glutathione (ox)	C ₂₀ H ₃₂ N ₆ O ₁₂ S ₂	613.16	127	355.22	484.14	595.35
89	Glutathione (red)	C ₁₀ H ₁₇ N ₃ O ₆ S	308.09	83	179.05	162.12	233.24
23	4-Hydroxybenzoic acid	C ₇ H ₆ O ₃	139.04	x	x	x	x
52	Ascorbic acid	C ₆ H ₈ O ₆	177.04	x	x	x	x
Metabolome D1							
1	Hydroxybutyric acid	C ₄ H ₈ O ₃	105.05	x	x	x	x
61	Glucuronic acid	C ₆ H ₁₀ O ₇	195.05	x	x	x	x
69	2-Hydroxy-3-methylbutyric acid	C ₅ H ₁₀ O ₃	119.07	x	x	x	x
69	2-Deoxycytidine	C ₉ H ₁₃ N ₃ O ₄	228.10	22	111.96	x	x

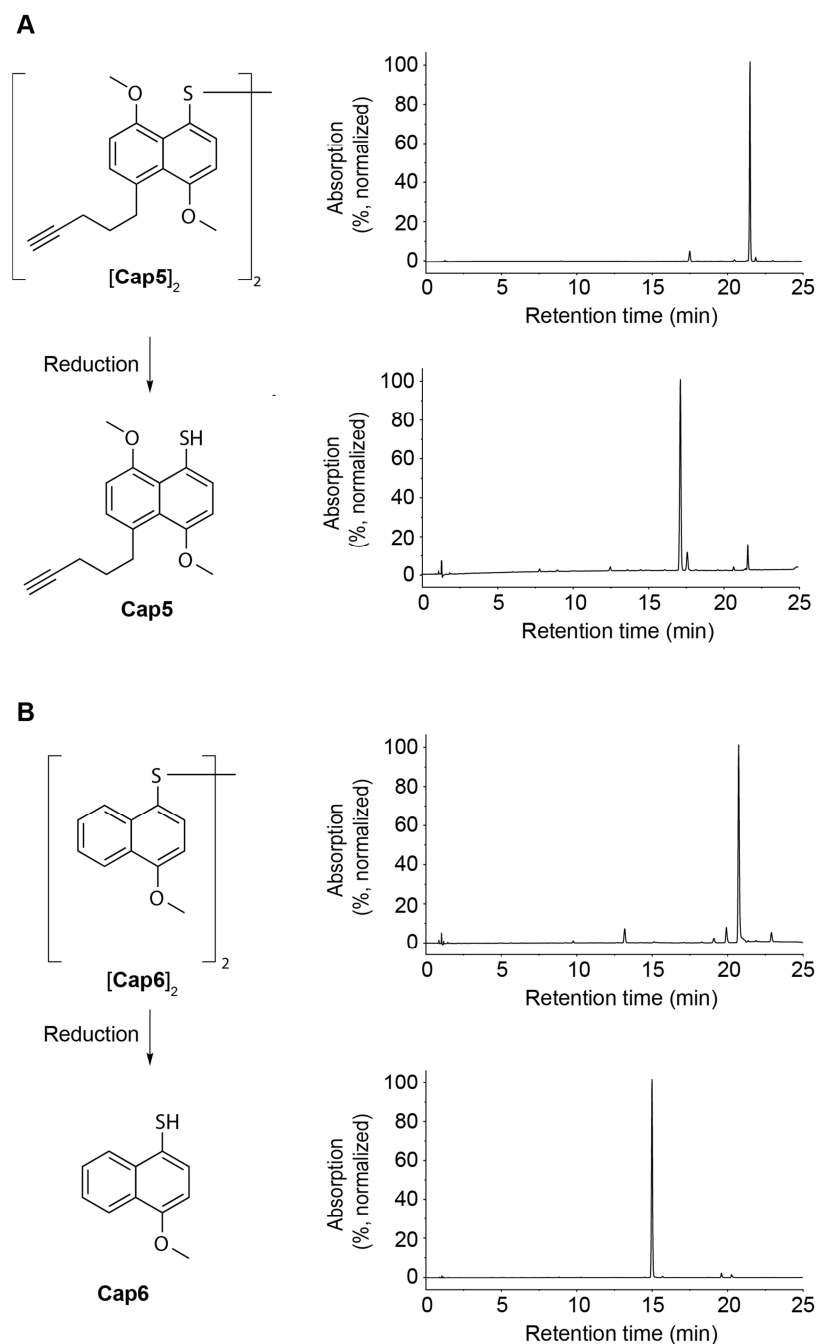
Chapter II – Appendices

117	Tyr-Tyr-Tyr	C ₂₇ H ₂₉ N ₃ O ₇	508.21	32	327.08	299.10	x
65	2-Deoxyribose-5-phosphate	C ₅ H ₁₁ O ₇ P	215.03	x	x	x	x
13	Pyroglutamic acid	C ₅ H ₇ NO ₃	130.05	x	x	x	x
Metabolome D2							
5	Succinic acid	C ₄ H ₆ O ₄	119.03	x	x	x	x
87	Inosine	C ₁₀ H ₁₂ N ₄ O ₅	269.09	24	137.18	x	x
43	L-β-Phenyllactic acid	C ₉ H ₁₀ O ₃	167.07	x	x	x	x
121	UDP-D-Glucuronic acid	C ₁₅ H ₂₂ N ₂ O ₁₈ P ₂	581.04	110	325.18	405.19	563.09
97	Thiamine monophosphate	C ₁₂ H ₁₇ N ₄ O ₄ PS	345.08	138	122.01	224.09	247.16
54	Myo-inositol	C ₆ H ₁₂ O ₆	181.07	x	x	x	x
Metabolome D3							
86	Deoxyguanosine	C ₁₀ H ₁₃ N ₅ O ₄	268.10	24	151.98	x	x
72	Thymidine	C ₁₀ H ₁₄ N ₂ O ₅	243.10	11	126.99	x	x
48	Dihydroxyacetonephosphate	C ₃ H ₇ O ₆ P	171.01	x	x	x	x
120	UDP-D-glucose	C ₁₅ H ₂₄ N ₂ O ₁₇ P ₂	567.06	94	x	x	x
49	Glycerol-3-phosphate	C ₃ H ₉ O ₆ P	173.02	88	98.92	x	x
88	6-Phosphogluconic acid	C ₆ H ₁₃ O ₁₀ P	277.03	113	259.05	x	x
70	Ribose-5-phosphate	C ₅ H ₁₁ O ₈ P	231.03	92	212.91	x	x
17	L-Asparagine	C ₄ H ₈ N ₂ O ₃	133.06	70	86.87	x	x
18	Ornithine	C ₅ H ₁₂ N ₂ O ₂	133.10	136	115.04	115.97	x
Metabolome D4							
27	Spermidine	C ₇ H ₁₉ N ₃	146.17	37	129.07	72.01	x
16	Glutaric acid	C ₅ H ₈ O ₄	133.05	x	x	x	x
9	L-Homoserine	C ₄ H ₉ NO ₃	120.07	x	x	x	x
100	IMP	C ₁₀ H ₁₃ N ₄ O ₈ P	349.05	87	136.98	233.18	x
81	Glucose-1-phosphate	C ₆ H ₁₃ O ₉ P	261.04	98	224.87	98.92	x
96	Glucose-1,6-bisphosphate	C ₆ H ₁₄ O ₁₂ P ₂	341.00	126	322.31	243.01	x
105	Thiamine pyrophosphate	C ₁₂ H ₁₉ N ₄ O ₇ P ₂ S	425.04	137	346.22	305.07	x
Metabolome D5							
41	Phenylpyruvic acid	C ₉ H ₈ O ₃	165.05	x	x	x	x
4	Uracil	C ₄ H ₄ N ₂ O ₂	113.03	x	x	x	x
12	Thymine	C ₅ H ₆ N ₂ O ₂	127.05	x	x	x	x
60	Isocitric acid	C ₆ H ₈ O ₇	193.03	x	x	x	x
90	N-Acetylneuraminic acid	C ₁₁ H ₁₉ NO ₉	310.11	67	292.15	274.08	x
71	Gly-Tyr	C ₁₁ H ₁₄ N ₂ O ₄	239.10	64	193.07	182.06	221.11
24	Acetylphosphate	C ₂ H ₅ O ₅ P	140.99	x	x	x	x
79	Deoxyadenosine	C ₁₀ H ₁₃ N ₅ O ₃	252.11	12	135.95	x	x
95	dAMP	C ₁₀ H ₁₄ N ₅ O ₆ P	332.08	75	135.96	x	x
Metabolome D6							
33	Guanine	C ₅ H ₅ N ₅ O	152.06	25	135.07	109.91	x

35	Xanthine	C ₅ H ₄ N ₄ O ₂	153.04	17	135.97	110.06	x
94	Adenosin-2&3-monophosphoric acid	C ₁₀ H ₁₂ N ₅ O ₆ P	330.06	x	x	x	x
82	Glucose-6-phosphphate	C ₆ H ₁₃ O ₉ P	261.04	104	243.01	225.24	x
Metabolome E1							
62	Gluconic acid	C ₆ H ₁₂ O ₇	197.07	x	x	x	x
57	O-P-Serine	C ₃ H ₈ NO ₆ P	186.02	117	87.92	x	x
51	Citrulline	C ₆ H ₁₃ N ₃ O ₃	176.10	76	159.06	x	x
5	Succinic acid	C ₄ H ₆ O ₄	119.03	x	x	x	x
109	Flavin mononucleotide	C ₁₇ H ₂₁ N ₄ O ₉ P	457.11	69	439.20	359.20	x
98	cGMP	C ₁₀ H ₁₂ N ₅ O ₇ P	346.05	55	152.00	135.13	x
26	4-Guanidinobutanoic acid	C ₅ H ₁₁ N ₃ O ₂	146.09	67	86.98	128.01	104.03
58	3-Phosphoglyceric acid	C ₃ H ₇ O ₇ P	187.00	x	x	x	x
63	O-P-Threonine	C ₄ H ₁₀ NO ₆ P	200.03	111	101.93	x	x
83	O-P-L-Tyrosine	C ₉ H ₁₂ N ₁ O ₆ P	262.05	119	216.07	245.09	x
Metabolome E2							
59	N(6)-Acetyl-L-Lysine	C ₈ H ₁₆ N ₂ O ₃	189.12	55	126.10	143.01	x
32	D-Fructose	C ₆ H ₁₂ O ₆	181.07	x	x	x	x
13	Pyroglutamic acid	C ₅ H ₇ NO ₃	130.05	x	x	x	x
40	Allantoin	C ₄ H ₆ N ₄ O ₃	159.05	x	x	x	x
37	2,3-Dihydroxybenzoic acid	C ₇ H ₆ O ₄	155.03	x	x	x	x
16	Glutaric acid	C ₅ H ₈ O ₄	133.05	x	x	x	x
20	L-Malic acid	C ₄ H ₆ O ₅	135.03	x	x	x	x
Metabolome F1							
107	Folic acid	C ₁₉ H ₁₉ N ₇ O ₆	442.15	x	x	x	x
22	Hypoxanthine	C ₅ H ₄ N ₄ O	137.05	16	x	x	x
93	cAMP	C ₁₀ H ₁₂ N ₅ O ₆ P	330.06	42	312.13	136.14	x
114	dATP	C ₁₀ H ₁₆ N ₅ O ₁₂ P ₃	492.01	103	394.15	322.18	x
110	dCTP	C ₉ H ₁₆ N ₃ O ₁₃ P ₃	468.00	114	192.11	450.03	x
116	dGTP	C ₁₀ H ₁₆ N ₅ O ₁₃ P ₃	508.00	115	232.14	152.00	x
111	dTTP	C ₁₀ H ₁₇ N ₂ O ₁₄ P ₃	483.00	x	x	x	x
Miscellaneous							
75	Biotin	C ₁₀ H ₁₆ N ₂ O ₃ S	245.10	25	227.13	x	x
102	Riboflavin	C ₁₇ H ₂₀ N ₄ O ₆	377.15	14	243.20	x	x

5.2 Subclass-Specific Labeling of Protein-Reactive Natural Products with customized nucleophilic probes

5.2.1 Figures



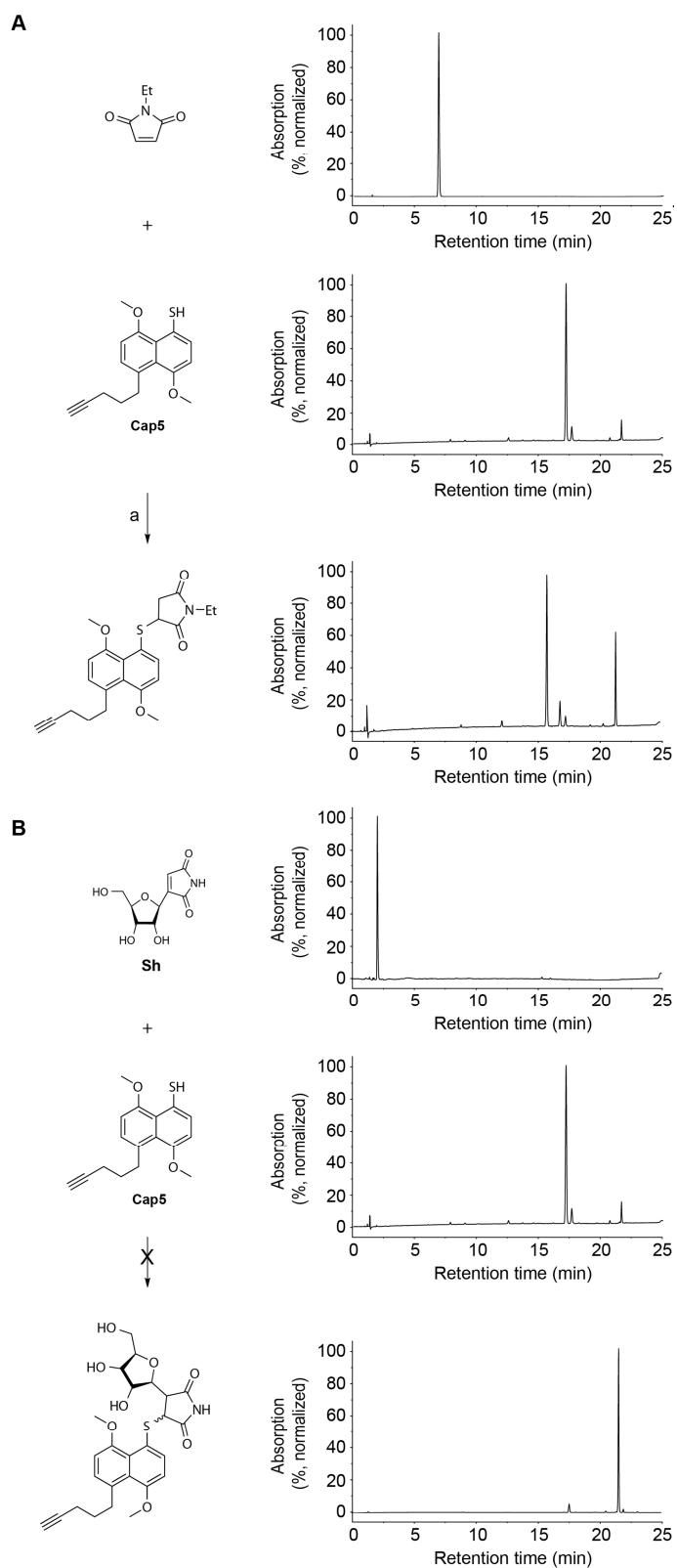


Figure II-A2: A) *N*-ethylmaleimide captured by thiol-based nucleophilic probe **Cap5**: a) acetonitrile/DMSO, 18 h, 30 °C. HPLC chromatograms at 303 nm wavelength (normalized to highest peak) display the quantitative conversion of *N*-ethylmaleimide (thiol spare signal due to excess starting material in the reaction). B) HPLC analysis of the reaction of showdomycin with **Cap5** in aqueous buffered media (PBS/DMSO 9:1, pH 7.4, 30 °C, 18 h). No conversion of **Cap5** with **Sh** observed.

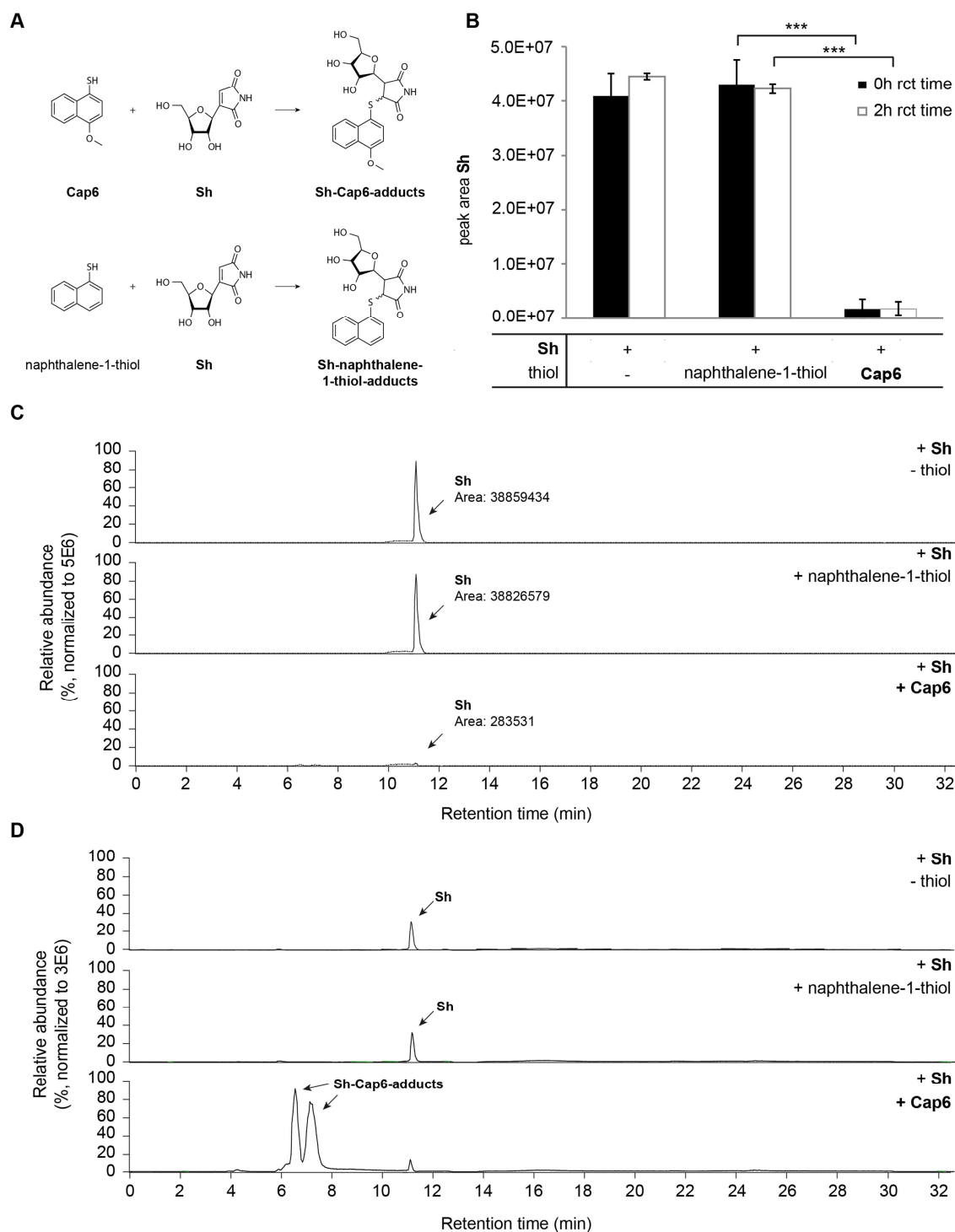
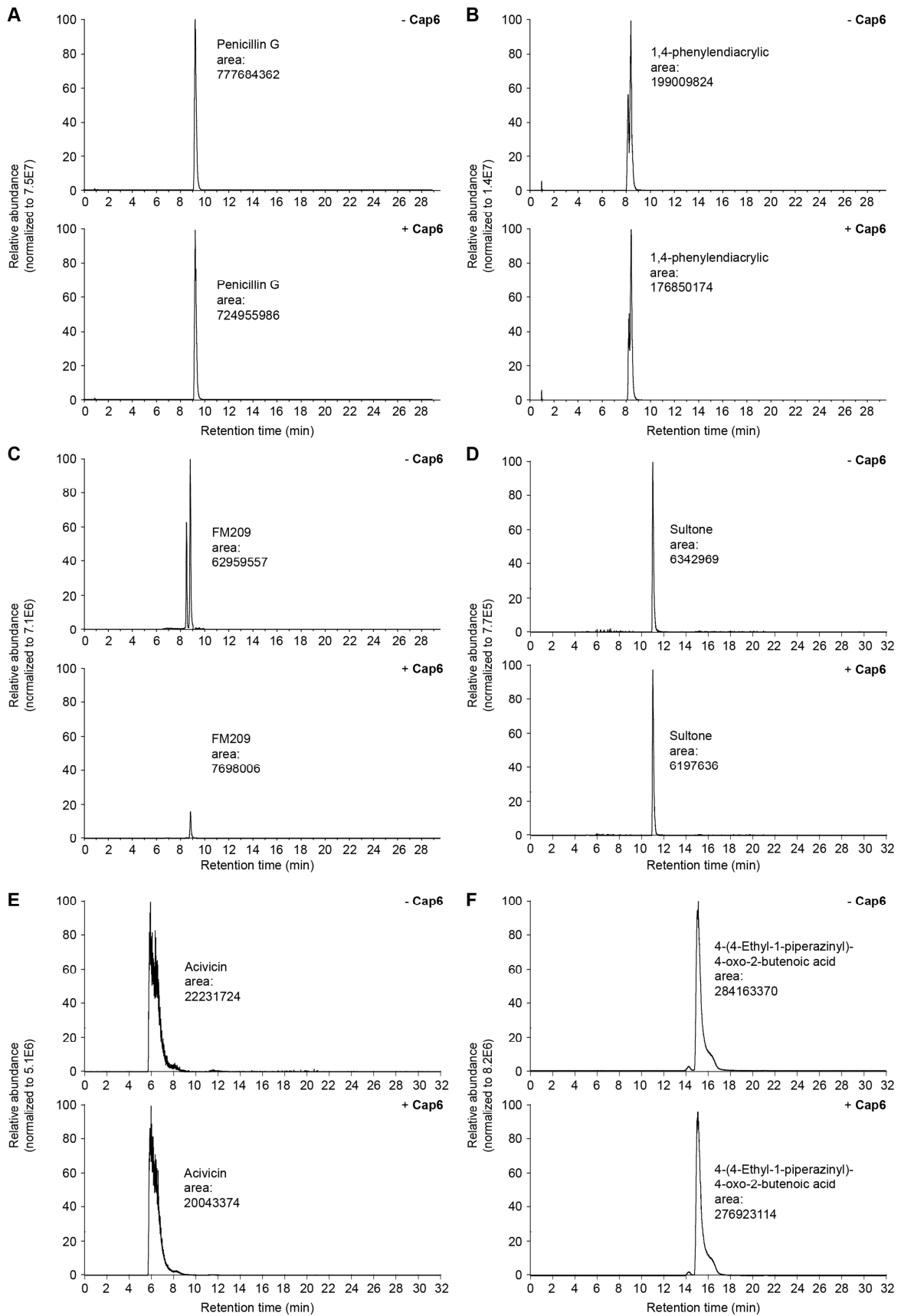


Figure II-A3: A) Reaction of naphthalene-1-thiol and Cap6 with showdomycin Sh. **B)** Bar graphs represent the ratio (in %) of detected showdomycin Sh compared to control. Calculated *p*-values for the different reaction times are: $p = 4.85E-04$ for 0 h rct time and $p = 5.58E-06$ for 2 h rct time. In both experiments *p*-values are statistically highly significant ($***p < 0.005$). Results are the average of three independent measurements, and error bars represent standard deviation. **C)** SIM-scan HILIC-HESI-HRMS traces of the conversion of showdomycin with Cap6 and naphthalene-1-thiol compared to a showdomycin control. **D)** Corresponding MS-Full Scans (80-1000 Da) of the conversion shown in C).



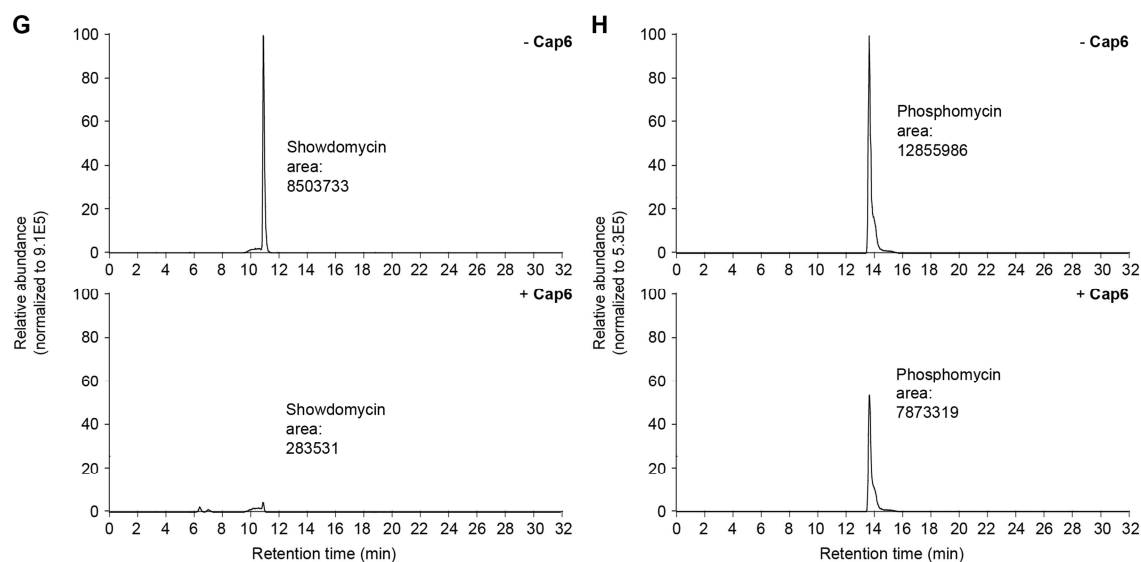


Figure II-A4: Reaction of Cap6 in a mixture of 8 electrophiles **A) – C)** Reversed-phase HPLC-ESI-HRMS (negative polarity) traces of the reaction penicillin G, phenylendiacrylic acid and FM209. **D) – F)** HILIC-HESI-HRMS (positive polarity) traces of sultone, acivicin and 4-(4-Ethyl-1-piperazinyl)-4-oxo-2-butenoic acid. **G)** and **H)** HILIC-HESI-HRMS (negative polarity) traces of showdomycin and phosphomycin.

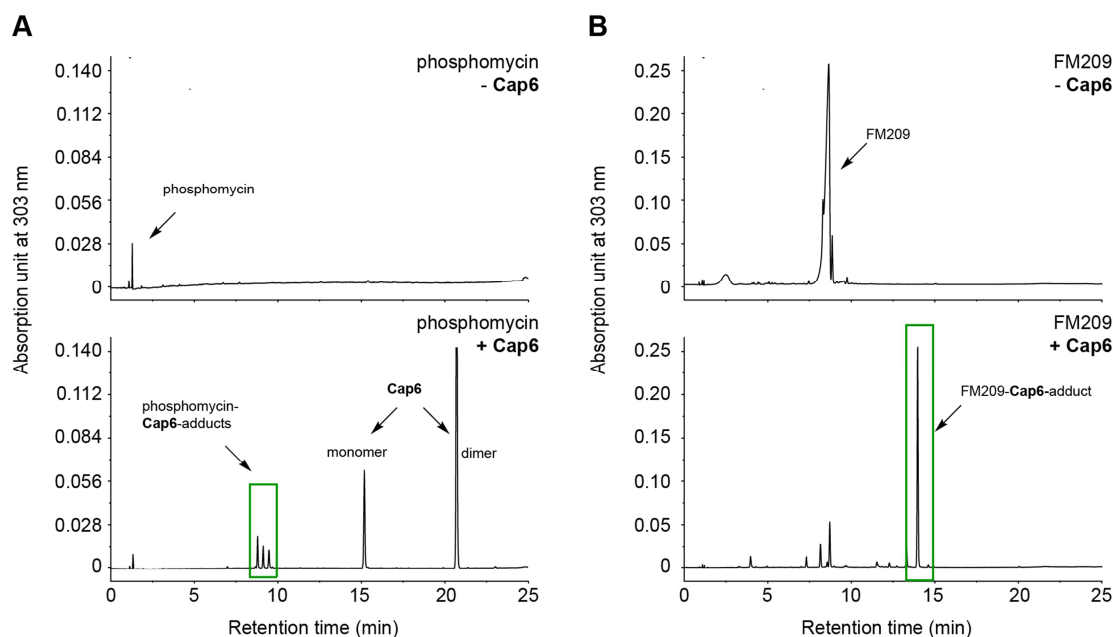


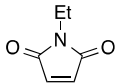
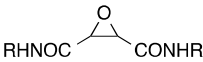

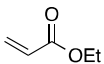


Figure II-A5: A) Reversed-phase HPLC traces of phosphomycin and of phosphomycin incubated with Cap6. B) Reversed-phase HPLC traces of FM209 and of FM209 incubated with Cap6.

5.2.2 Tables

Table II-A3: Reduction methods tested for the conversion of disulfane [Cap5]2 to free thiol Cap5 (n.c. = no conversion observed).

Reduction Method	Conditions	Yield (HPLC)
NaBH ₄ , MeOH	DMF, 40 °C	“thiol odour”
PPh ₃ , HCl/H ₂ O	Dioxane, 40 °C	n.c.
TCEP, H ₂ O	MeCN/MeOH, 30 °C	slow
TCEP, acetate buffer (pH 4)	MeCN/MeOH, 30 °C	> 95%
TCEP (immobilized)	MeCN/MeOH, 30 °C	n.c.

Table II-A4: Screening results for reactivity of nucleophilic probes in aqueous buffered media (PBS/DMSO 9:1, pH 7.4, 30 °C, 18 h, HPLC analysis, conversion yield attributed to consumption of the electrophilic starting material; n.c. = no conversion, R = aliphatic rest).

Nucleophilic probe:	Cap5	Cap6
	quant.	quant.
	< 10%	~ 50%
	< 10%	< 10%
	< 10%	< 10%
	n.c.	n.c.
	n.c.	n.c.

III - Bibliography

1. Martin DG, Chidester CG, Duchamp DJ, Mizesak SA. Structure of CC-1065 (NSC-298223), a new antitumor antibiotic. *The Journal of antibiotics* 1980, **33**(8): 902-903.
2. Yasuzawa T, Saitoh Y, Ichimura M, Takahashi I, Sano H. Structure of duocarmycin SA, a potent antitumor antibiotic. *The Journal of antibiotics* 1991, **44**(4): 445-447.
3. Boger DL, Johnson DS. CC-1065 and the duocarmycins: unraveling the keys to a new class of naturally derived DNA alkylating agents. *Proceedings of the National Academy of Sciences of the United States of America* 1995, **92**(9): 3642-3649.
4. Wirth T, Schmuck K, Tietze LF, Sieber SA. Duocarmycin analogues target aldehyde dehydrogenase 1 in lung cancer cells. *Angew Chem Int Ed Engl* 2012, **51**(12): 2874-2877.
5. Wirth T, Pestel GF, Ganai V, Kirmeier T, Schuberth I, Rein T, *et al.* The Two Faces of Potent Antitumor Duocarmycin-Based Drugs: A Structural Dissection Reveals Disparate Motifs for DNA versus Aldehyde Dehydrogenase 1 Affinity. *Angewandte Chemie International Edition* 2013, **52**(27): 6921-6925.
6. Fiehn O. Metabolomics – the link between genotypes and phenotypes. *Plant Mol Biol* 2002, **48**(1-2): 155-171.
7. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews* 2007, **26**(1): 51-78.
8. Koppaka V, Thompson DC, Chen Y, Ellermann M, Nicolaou KC, Juvonen RO, *et al.* Aldehyde Dehydrogenase Inhibitors: a Comprehensive Review of the Pharmacology, Mechanism of Action, Substrate Specificity, and Clinical Application. *Pharmacological Reviews* 2012, **64**(3): 520-539.
9. Sophos NA, Vasiliou V. Aldehyde dehydrogenase gene superfamily: the 2002 update. *Chemico-biological interactions* 2003, **143–144**: 5-22.
10. Alnouti Y, Klaassen CD. Tissue Distribution, Ontogeny, and Regulation of Aldehyde Dehydrogenase (Aldh) Enzymes mRNA by Prototypical Microsomal Enzyme Inducers in Mice. *Toxicological Sciences* 2008, **101**(1): 51-64.
11. Riveros-Rosas H, González-Segura L, Julián-Sánchez A, Díaz-Sánchez ÁG, Muñoz-Clares RA. Structural determinants of substrate specificity in aldehyde dehydrogenases. *Chemico-biological interactions* 2013, **202**(1–3): 51-61.
12. Marchitti SA, Deitrich RA, Vasiliou V. Neurotoxicity and Metabolism of the Catecholamine-Derived 3,4-Dihydroxyphenylacetaldehyde and 3,4-Dihydroxyphenylglycolaldehyde: The Role of Aldehyde Dehydrogenase. *Pharmacological Reviews* 2007, **59**(2): 125-150.
13. Black W, Vasiliou V. The Aldehyde Dehydrogenase Gene Superfamily Resource Center. *Human Genomics* 2009, **4**(2): 136-142.

14. Suzuki Y, Ogawa S, Sakakibara Y. Chaperone Therapy for Neuronopathic Lysosomal Diseases: Competitive Inhibitors as Chemical Chaperones for Enhancement of Mutant Enzyme Activities. *Perspectives in Medicinal Chemistry* 2009, **3**: 7-19.
15. Lowe ED, Gao G-Y, Johnson LN, Keung WM. Structure of Daidzin, a Naturally Occurring Anti-Alcohol-Addiction Agent, in Complex with Human Mitochondrial Aldehyde Dehydrogenase. *Journal of medicinal chemistry* 2008, **51**(15): 4482-4487.
16. Wenzl MV, Beretta M, Griesberger M, Russwurm M, Koesling D, Schmidt K, *et al.* Site-Directed Mutagenesis of Aldehyde Dehydrogenase-2 Suggests Three Distinct Pathways of Nitroglycerin Biotransformation. *Molecular Pharmacology* 2011, **80**(2): 258-266.
17. Yoshida A, Hsu LC, Davé V. Retinal oxidation activity and biological role of human cytosolic aldehyde dehydrogenase. *Enzyme* 1992, **46**(4-5): 239-244.
18. Hess DA, Wirthlin L, Craft TP, Herrbrich PE, Hohm SA, Lahey R, *et al.* Selection based on CD133 and high aldehyde dehydrogenase activity isolates long-term reconstituting human hematopoietic stem cells. *Blood* 2006, **107**(5): 2162-2169.
19. Rice KL, Izon DJ, Ford J, Boodhoo A, Kees UR, Greene WK. Overexpression of stem cell associated ALDH1A1, a target of the leukemogenic transcription factor TLX1/HOX11, inhibits lymphopoiesis and promotes myelopoiesis in murine hematopoietic progenitors. *Leukemia Research* 2008, **32**(6): 873-883.
20. Vasiliou V, Pappa A, Petersen DR. Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism. *Chemico-biological interactions* 2000, **129**(1-2): 1-19.
21. Higuchi S, Matsushita S, Murayama M, Takagi S, Hayashida M. Alcohol and aldehyde dehydrogenase polymorphisms and the risk for alcoholism. *Am J Psychiatry* 1995, **152**(8): 1219-1221.
22. Larson HN, Weiner H, Hurley TD. Disruption of the coenzyme binding site and dimer interface revealed in the crystal structure of mitochondrial aldehyde dehydrogenase 'Asian' variant. *The Journal of biological chemistry* 2005, **280**(34): 30550-30556.
23. Larson HN, Zhou J, Chen Z, Stamler JS, Weiner H, Hurley TD. Structural and Functional Consequences of Coenzyme Binding to the Inactive Asian Variant of Mitochondrial Aldehyde Dehydrogenase: Roles of Residues 475 and 487. *The Journal of biological chemistry* 2007, **282**(17): 12940-12950.
24. Esterbauer H, Schaur RJ, Zollner H. Chemistry and biochemistry of 4-hydroxynonenal, malonaldehyde and related aldehydes. *Free Radical Biology and Medicine* 1991, **11**(1): 81-128.
25. Kurys G, Ambroziak W, Pietruszko R. Human aldehyde dehydrogenase. Purification and characterization of a third isozyme with low Km for gamma-aminobutyraldehyde. *Journal of Biological Chemistry* 1989, **264**(8): 4715-4721.

26. O'Brien PJ, Siraki AG, Shangari N. Aldehyde Sources, Metabolism, Molecular Toxicity Mechanisms, and Possible Effects on Human Health. *Critical Reviews in Toxicology* 2005, **35**(7): 609-662.
27. Allahverdiyev AM, Bagirova M, Oztel ON, Yaman S, Abamor ES, Koc RC, *et al.* *Aldehyde Dehydrogenase: Cancer and Stem Cells*, 2012.
28. Michel TM, Käsbauer L, Gsell W, Jecel J, Sheldrick AJ, Cortese M, *et al.* Aldehyde dehydrogenase 2 in sporadic Parkinson's disease. *Parkinsonism & Related Disorders* 2014, **20**, **Supplement 1**: S68-S72.
29. Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer Statistics, 2007. *CA: A Cancer Journal for Clinicians* 2007, **57**(1): 43-66.
30. Storms RW, Trujillo AP, Springer JB, Shah L, Colvin OM, Ludeman SM, *et al.* Isolation of primitive human hematopoietic progenitors on the basis of aldehyde dehydrogenase activity. *Proceedings of the National Academy of Sciences of the United States of America* 1999, **96**(16): 9118-9123.
31. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med* 1997, **3**(7): 730-737.
32. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences* 2003, **100**(7): 3983-3988.
33. Eramo A, Lotti F, Sette G, Pillozzi E, Biffoni M, Di Virgilio A, *et al.* Identification and expansion of the tumorigenic lung cancer stem cell population. *Cell Death Differ* 2007, **15**(3): 504-514.
34. Salnikov AV, Gladkich J, Moldenhauer G, Volm M, Mattern J, Herr I. CD133 is indicative for a resistance phenotype but does not represent a prognostic marker for survival of non-small cell lung cancer patients. *International Journal of Cancer* 2010, **126**(4): 950-958.
35. Alison MR, Guppy NJ, Lim SML, Nicholson LJ. Finding cancer stem cells: are aldehyde dehydrogenases fit for purpose? *The Journal of Pathology* 2010, **222**(4): 335-344.
36. Mitchell JB, McIntosh K, Zvonic S, Garrett S, Floyd ZE, Kloster A, *et al.* Immunophenotype of Human Adipose-Derived Cells: Temporal Changes in Stromal-Associated and Stem Cell-Associated Markers. *STEM CELLS* 2006, **24**(2): 376-385.
37. Siclari VA, Qin L. Targeting the osteosarcoma cancer stem cell. *Journal of Orthopaedic Surgery and Research* 2010, **5**: 78-78.
38. Serrano D, Bleau A-M, Fernandez-Garcia I, Fernandez-Marcelo T, Iniesta P, Ortiz-de-Solorzano C, *et al.* Inhibition of telomerase activity preferentially targets aldehyde dehydrogenase-positive cancer stem-like cells in lung cancer. *Molecular Cancer* 2011, **10**: 96-96.
39. Marchitti SA, Brocker C, Stagos D, Vasiliou V. Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. *Expert opinion on drug metabolism & toxicology* 2008, **4**(6): 697-720.

40. Marcato P, Dean CA, Giacomantonio CA, Lee PWK. Aldehyde dehydrogenase: Its role as a cancer stem cell marker comes down to the specific isoform. *Cell cycle* 2011, **10**(9): 1378-1384.
41. Okudela K, Woo T, Mitsui H, Suzuki T, Tajiri M, Sakuma Y, *et al.* Downregulation of ALDH1A1 expression in non-small cell lung carcinomas – its clinicopathologic and biological significance. *International Journal of Clinical and Experimental Pathology* 2013, **6**(1): 1-12.
42. Rizzo WB, Dammann AL, Craft DA. Sjögren-Larsson syndrome. Impaired fatty alcohol oxidation in cultured fibroblasts due to deficient fatty alcohol:nicotinamide adenine dinucleotide oxidoreductase activity. *Journal of Clinical Investigation* 1988, **81**(3): 738-744.
43. Rizzo WB, Carney G. Sjögren-Larsson syndrome: Diversity of mutations and polymorphisms in the fatty aldehyde dehydrogenase gene (ALDH3A2). *Human Mutation* 2005, **26**(1): 1-10.
44. Geraghty MT, Vaughn D, Nicholson AJ, Lin W-W, Jimenez-Sanchez G, Obie C, *et al.* Mutations in the Δ^1 -Pyrroline 5-Carboxylate Dehydrogenase Gene Cause Type II Hyperprolinemia. *Human molecular genetics* 1998, **7**(9): 1411-1415.
45. Farrant RD, Walker V, Mills GA, Mellor JM, Langley GJ. Pyridoxal Phosphate Deactivation by Pyrroline-5-carboxylic Acid: INCREASED RISK OF VITAMIN B6 DEFICIENCY AND SEIZURES IN HYPERPROLINEMIA TYPE II. *Journal of Biological Chemistry* 2001, **276**(18): 15107-15116.
46. Mills PB, Struys E, Jakobs C, Plecko B, Baxter P, Baumgartner M, *et al.* Mutations in antiquitin in individuals with pyridoxine-dependent seizures. *Nat Med* 2006, **12**(3): 307-309.
47. Grünblatt E, Riederer P. Aldehyde dehydrogenase (ALDH) in Alzheimer's and Parkinson's disease. *J Neural Transm* 2014: 1-8.
48. Wang M-F, Han C-L, Yin S-J. Substrate specificity of human and yeast aldehyde dehydrogenases. *Chemico-biological interactions* 2009, **178**(1–3): 36-39.
49. Min H, Shane B, Stokstad ELR. Identification of 10-formyltetrahydrofolate dehydrogenase-hydrolase as major folate binding protein in liver cytosol. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1988, **967**(3): 348-353.
50. Strickland KC, Krupenko NI, Dubard ME, Hu CJ, Tsybovsky Y, Krupenko SA. Enzymatic properties of ALDH1L2, a mitochondrial 10-formyltetrahydrofolate dehydrogenase. *Chemico-biological interactions* 2011, **191**(1–3): 129-136.
51. Liu Z-J, Hempel J, Sun J, Rose J, Hsiao D, Chang W-R, *et al.* Crystal Structure of a Class 3 Aldehyde Dehydrogenase at 2.6Å Resolution. In: Weiner H, Lindahl R, Crabb D, Flynn TG (eds). *Enzymology and Molecular Biology of Carbonyl Metabolism* 6, vol. 414. Springer US, 1997, pp 1-7.
52. Steinmetz CG, Xie P, Weiner H, Hurley TD. Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion. *Structure* 1997, **5**(5): 701-711.

53. Colby TD, Bahnson BJ, Chin JK, Klinman JP, Goldstein BM. Active Site Modifications in a Double Mutant of Liver Alcohol Dehydrogenase: Structural Studies of Two Enzyme–Ligand Complexes. *Biochemistry* 1998, **37**(26): 9295-9304.
54. Cobessi D, Tête-Favier F, Marchal S, Branlant G, Aubry A. Structural and biochemical investigations of the catalytic mechanism of an NADP-dependent aldehyde dehydrogenase from *Streptococcus mutans*1. *Journal of molecular biology* 2000, **300**(1): 141-152.
55. D'Ambrosio K, Pailot A, Talfournier F, Didierjean C, Benedetti E, Aubry A, *et al.* The First Crystal Structure of a Thioacylenzyme Intermediate in the ALDH Family: New Coenzyme Conformation and Relevance to Catalysis†. *Biochemistry* 2006, **45**(9): 2978-2986.
56. Hempel J, Perozich J, Chapman T, Rose J, Boesch J, Liu Z-J, *et al.* Aldehyde Dehydrogenase Catalytic Mechanism. In: Weiner H, Maser E, Crabb D, Lindahl R (eds). *Enzymology and Molecular Biology of Carbonyl Metabolism 7*, vol. 463. Springer US, 1999, pp 53-59.
57. Pappa A, Brown D, Koutalos Y, DeGregori J, White C, Vasiliou V. Human Aldehyde Dehydrogenase 3A1 Inhibits Proliferation and Promotes Survival of Human Corneal Epithelial Cells. *Journal of Biological Chemistry* 2005, **280**(30): 27998-28006.
58. Moreb JS, Ucar D, Han S, Amory JK, Goldstein AS, Ostmark B, *et al.* The enzymatic activity of human aldehyde dehydrogenases 1A2 and 2 (ALDH1A2 and ALDH2) is detected by Aldefluor, inhibited by diethylaminobenzaldehyde and has significant effects on cell proliferation and drug resistance. *Chemico-biological interactions* 2012, **195**(1): 52-60.
59. Moreb J, Schweder M, Suresh A, Zucali JR. Overexpression of the human aldehyde dehydrogenase class I results in increased resistance to 4-hydroperoxycyclophosphamide. *Cancer Gene Ther* 1996, **3**(1): 24-30.
60. Daiber A, Harrison DG, Münzel T. Doubt about an essential role for constitutive nitric oxide synthase in nitroglycerin-mediated vasodilation. *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**(47): E92-E92.
61. Lang BS, Gorren AC, Oberdorfer G, Wenzl MV, Furdul CM, Poole LB, *et al.* Vascular bioactivation of nitroglycerin by aldehyde dehydrogenase-2: reaction intermediates revealed by crystallography and mass spectrometry. *The Journal of biological chemistry* 2012, **287**(45): 38124-38134.
62. Bell RG, Smith HW. CLINICAL TRIALS OF ANTABUSE. *Canadian Medical Association Journal* 1949, **60**(3): 286-288.
63. Marcato P, Dean CA, Giacomantonio CA, Lee PW. Aldehyde dehydrogenase: its role as a cancer stem cell marker comes down to the specific isoform. *Cell cycle* 2011, **10**(9): 1378-1384.
64. Pike MG, Mays DC, Macomber DW, Lipsky JJ. Metabolism of a Disulfiram Metabolite, S-MethylN,N-Diethylthiocarbamate, by Flavin Monooxygenase in

- Human Renal Microsomes. *Drug Metabolism and Disposition* 2001, **29**(2): 127-132.
65. Hart BW, Faiman MD. In vitro and in vivo inhibition of rat liver aldehyde dehydrogenase by s-methyl N, N-diethylthiocarbamate sulfoxide, a new metabolite of disulfiram. *Biochemical pharmacology* 1992, **43**(3): 403-406.
66. Mays DC, Nelson AN, Lam-Holt J, Fauq AH, Lipsky JJ. S-Methyl-N,N-Diethylthiocarbamate Sulfoxide and S-Methyl-N,N-Diethylthiocarbamate Sulfone, Two Candidates for the Active Metabolite of Disulfiram. *Alcoholism: Clinical and Experimental Research* 1996, **20**(3): 595-600.
67. Mays DC, Ortiz-Bermudez P, Lam JP, Tong IH, Fauq AH, Lipsky JJ. Inhibition of Recombinant Human Mitochondrial Aldehyde Dehydrogenase by Two Intermediate Metabolites of Disulfiram. *Biochemical pharmacology* 1998, **55**(7): 1099-1103.
68. Staub RE, Quistad GB, Casida JE. S-methyl N-butylthiocarbamate sulfoxide: Selective carbamoylating agent for mouse mitochondrial aldehyde dehydrogenase. *Biochemical pharmacology* 1999, **58**(9): 1467-1473.
69. Lipsky JJ, Shen ML, Naylor S. In vivo inhibition of aldehyde dehydrogenase by disulfiram. *Chemico-biological interactions* 2001, **130–132**: 93-102.
70. Moore SA, Baker HM, Blythe TJ, Kitson KE, Kitson TM, Baker EN. Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. *Structure* 1998, **6**(12): 1541-1551.
71. Rekha G, Sladek N. Inhibition of Human Class 3 Aldehyde Dehydrogenase, and Sensitization of Tumor Cells that Express Significant Amounts of this Enzyme to Oxazaphosphorines, by the Naturally Occurring Compound Gossypol. In: Weiner H, Lindahl R, Crabb D, Flynn TG (eds). *Enzymology and Molecular Biology of Carbonyl Metabolism* 6, vol. 414. Springer US, 1997, pp 133-146.
72. Dodou K. Investigations on gossypol: past and present developments. *Expert Opinion on Investigational Drugs* 2005, **14**(11): 1419-1434.
73. Liu Z-J, Sun Y-J, Rose J, Chung Y-J, Hsiao C-D, Chang W-R, *et al.* The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. *Nature structural & molecular biology* 1997, **4**(4): 317-326.
74. Morgan CA, Hurley TD. Characterization of two distinct structural classes of selective aldehyde dehydrogenase 1A1 inhibitors. *Journal of medicinal chemistry* 2015, **58**(4): 1964-1975.
75. Vasiliou V, Sandoval M, Backos DS, Jackson BC, Chen Y, Reigan P, *et al.* ALDH16A1 is a novel non-catalytic enzyme that may be involved in the etiology of gout via protein–protein interactions with HPRT1. *Chemico-biological interactions* 2013, **202**(1–3): 22-31.
76. Luo M, Gates KS, Henzl MT, Tanner JJ. Diethylaminobenzaldehyde Is a Covalent, Irreversible Inactivator of ALDH7A1. *ACS chemical biology* 2015.

77. Boyer CS, Petersen DR. The metabolism of 3,7-dimethyl-2,6-octadienal (citral) in rat hepatic mitochondrial and cytosolic fractions. Interactions with aldehyde and alcohol dehydrogenases. *Drug Metabolism and Disposition* 1991, **19**(1): 81-86.
78. Kikonyogo A, Abriola DP, Dryjanski M, Pietruszko R. Mechanism of inhibition of aldehyde dehydrogenase by citral, a retinoid antagonist. *European Journal of Biochemistry* 1999, **262**(3): 704-712.
79. Poon R, Nakai J, Yagminas A, Benoit F, Moir D, Chu I, *et al.* Subchronic toxicity of chloral hydrate on rats: a drinking water study. *Journal of Applied Toxicology* 2002, **22**(4): 227-236.
80. Morgan CA, Hurley TD. Development of a high-throughput in vitro assay to identify selective inhibitors for human ALDH1A1. *Chemico-biological interactions* 2015, **234**: 29-37.
81. Allen EMG, Anderson DGR, Florang VR, Khanna M, Hurley TD, Doorn JA. Relative Inhibitory Potency of Molinate and Metabolites with Aldehyde Dehydrogenase 2: Implications for the Mechanism of Enzyme Inhibition. *Chemical Research in Toxicology* 2010, **23**(11): 1843-1850.
82. Chidester CG, Krueger WC, Mizensak SA, Duchamp DJ, Martin DG. The structure of CC-1065, a potent antitumor agent and its binding to DNA. *Journal of the American Chemical Society* 1981, **103**(25): 7629-7635.
83. Hurley L, Reynolds V, Swenson D, Petzold G, Scahill T. Reaction of the antitumor antibiotic CC-1065 with DNA: structure of a DNA adduct with DNA sequence specificity. *Science* 1984, **226**(4676): 843-844.
84. Boger DL, Johnson DS. CC-1065 and the duocarmycins: unraveling the keys to a new class of naturally derived DNA alkylating agents. *Proceedings of the National Academy of Sciences* 1995, **92**(9): 3642-3649.
85. Boger DL, Bollinger B, Hertzog DL, Johnson DS, Cai H, Mésini P, *et al.* Reversed and Sandwiched Analogs of Duocarmycin SA: Establishment of the Origin of the Sequence-Selective Alkylation of DNA and New Insights into the Source of Catalysis. *Journal of the American Chemical Society* 1997, **119**(21): 4987-4998.
86. Boger DL, Ishizaki T, Wysocki RJ, Munk SA, Kitos PA, Suntornwat O. Total synthesis and evaluation of (.+-.)-N-(tert-butoxycarbonyl)-CBI, (.+-.)-CBI-CDPI1, and (.+-.)-CBI-CDPI2: CC-1065 functional agents incorporating the equivalent 1,2,9,9a-tetrahydrocyclopropa[1,2-c]benz[1,2-e]indol-4-one (CBI) left-hand subunit. *Journal of the American Chemical Society* 1989, **111**(16): 6461-6463.
87. Kitz R, Wilson IB. Esters of Methanesulfonic Acid as Irreversible Inhibitors of Acetylcholinesterase. *Journal of Biological Chemistry* 1962, **237**(10): 3245-3249.
88. Ghanbari F, Rowland-Yeo K, Bloomer JC, Clarke SE, Lennard MS, Tucker GT, *et al.* A Critical Evaluation of the Experimental Design of Studies of Mechanism Based Enzyme Inhibition, with Implications for In Vitro-In Vivo Extrapolation. *Current Drug Metabolism* 2006, **7**(3): 315-334.
89. Maurer T, Fung H-L. Comparison of methods for analyzing kinetic data from mechanism-based enzyme inactivation: Application to nitric oxide synthase. *AAPS PharmSci* 2000, **2**(1): 68-77.

90. MacMillan KS, Boger DL. Fundamental relationships between structure, reactivity, and biological activity for the duocarmycins and CC-1065. *Journal of medicinal chemistry* 2009, **52**(19): 5771-5780.
91. Tietze LF, Krewer B, Frauendorf H. Probing the mechanism of action of potential anticancer agents at a molecular level using electrospray ionisation Fourier transform ion cyclotron resonance mass spectrometry. *Eur J Mass Spectrom* 2009, **15**: 661-672.
92. Boger DL, Garbaccio RM. Catalysis of the CC-1065 and duocarmycin DNA alkylation reaction: DNA binding induced conformational change in the agent results in activation. *Bioorganic & medicinal chemistry* 1997, **5**(2): 263-276.
93. Tietze LF, von Hof JM, Muller M, Krewer B, Schuberth I. Glycosidic prodrugs of highly potent bifunctional duocarmycin derivatives for selective treatment of cancer. *Angew Chem Int Ed Engl* 2010, **49**(40): 7336-7339.
94. Yue L, Huang ZM, Fong S, Leong S, Jakowatz JG, Charruyer-Reinwald A, *et al.* Targeting ALDH1 to decrease tumorigenicity, growth and metastasis of human melanoma. *Melanoma research* 2015, **25**(2): 138-148.
95. Kimble-Hill AC, Parajuli B, Chen CH, Mochly-Rosen D, Hurley TD. Development of selective inhibitors for aldehyde dehydrogenases based on substituted indole-2,3-diones. *Journal of medicinal chemistry* 2014, **57**(3): 714-722.
96. Khanna M, Chen CH, Kimble-Hill A, Parajuli B, Perez-Miller S, Baskaran S, *et al.* Discovery of a novel class of covalent inhibitor for aldehyde dehydrogenases. *The Journal of biological chemistry* 2011, **286**(50): 43486-43494.
97. Tercel M, McManaway SP, Leung E, Liyanage HD, Lu GL, Pruijn FB. The cytotoxicity of duocarmycin analogues is mediated through alkylation of DNA, not aldehyde dehydrogenase 1: a comment. *Angew Chem Int Ed Engl* 2013, **52**(21): 5442-5446.
98. Tietze LF, Sieber SA. Duocarmycin analogues without a DNA-binding indole unit associate with aldehyde dehydrogenase 1A1 and not DNA: a reply. *Angew Chem Int Ed Engl* 2013, **52**(21): 5447-5449.
99. Elgersma RC, Coumans RG, Huijbregts T, Menge WM, Joosten JA, Spijker HJ, *et al.* Design, Synthesis, and Evaluation of Linker-Duocarmycin Payloads: Toward Selection of HER2-Targeting Antibody-Drug Conjugate SYD985. *Molecular pharmaceutics* 2015.
100. Tietze L, Krewer B, Von Hof JM, Frauendorf H, Schuberth I. Determination of the Biological Activity and Structure Activity Relationships of Drugs Based on the Highly Cytotoxic Duocarmycins and CC-1065. *Toxins* 2009, **1**(2): 134.
101. Moore SA, Baker HM, Blythe TJ, Kitson KE, Kitson TM, Baker EN. Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. *Structure* 1998, **6**(12): 1541-1551.
102. Farres J, Wang TT, Cunningham SJ, Weiner H. Investigation of the active site cysteine residue of rat liver mitochondrial aldehyde dehydrogenase by site-directed mutagenesis. *Biochemistry* 1995, **34**(8): 2592-2598.

103. Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annual Review of Biophysics* 2012, **41**(1): 429-452.
104. Lam JP, Mays DC, Lipsky JJ. Inhibition of recombinant human mitochondrial and cytosolic aldehyde dehydrogenases by two candidates for the active metabolites of disulfiram. *Biochemistry* 1997, **36**(44): 13748-13754.
105. Martinez CR, Iverson BL. Rethinking the term "pi-stacking". *Chemical Science* 2012, **3**(7): 2191-2201.
106. Paton RS, Goodman JM. Hydrogen Bonding and π -Stacking: How Reliable are Force Fields? A Critical Evaluation of Force Field Descriptions of Nonbonded Interactions. *Journal of Chemical Information and Modeling* 2009, **49**(4): 944-955.
107. Moreb JS, Ucar D, Han S, Amory JK, Goldstein AS, Ostmark B, *et al.* The enzymatic activity of human aldehyde dehydrogenases 1A2 and 2 (ALDH1A2 and ALDH2) is detected by Aldefluor, inhibited by diethylaminobenzaldehyde and has significant effects on cell proliferation and drug resistance. *Chemico-biological interactions* 2012, **195**(1): 52-60.
108. Yasuzawa T, Iida T, Muroi Ki, Ichimura M, Takahashi K, Sano H. STRUCTURES OF DUOCARMYCINS, NOVEL ANTITUMOR ANTIBIOTICS PRODUCED BY STREPTOMYCES SP. *CHEMICAL & PHARMACEUTICAL BULLETIN* 1988, **36**(9): 3728-3731.
109. Hanka LJ, Dietz A. U-42, 126, a new antimetabolite antibiotic: production, biological activity, and taxonomy of the producing microorganism. *Antimicrobial agents and chemotherapy* 1973, **3**(3): 425-431.
110. Boger DL, Munk SA. DNA alkylation properties of enhanced functional analogs of CC-1065 incorporating the 1,2,9,9a-tetrahydrocyclopropa[1,2-c]benz[1,2-e]indol-4-one (CBI) alkylation subunit. *Journal of the American Chemical Society* 1992, **114**(14): 5487-5496.
111. Bhuyan BK, Newell KA, Crampton SL, Von Hoff DD. CC-1065 (NSC 298223), a Most Potent Antitumor Agent: Kinetics of Inhibition of Growth, DNA Synthesis, and Cell Survival. *Cancer research* 1982, **42**(9): 3532-3537.
112. Kabsch W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D: Biological Crystallography* 2010, **66**(Pt 2): 133-144.
113. Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography* 2013, **69**(Pt 7): 1204-1214.
114. Evans P. Resolving Some Old Problems in Protein Crystallography. *Science* 2012, **336**(6084): 986-987.
115. Diederichs K, Karplus PA. Better models by discarding data? *Acta Crystallographica Section D: Biological Crystallography* 2013, **69**(Pt 7): 1215-1222.

116. Karplus PA, Diederichs K. Linking crystallographic model and data quality. *Science (New York, NY)* 2012, **336**(6084): 1030-1033.
117. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *Journal of Applied Crystallography* 2007, **40**(Pt 4): 658-674.
118. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* 2011, **67**(Pt 4): 235-242.
119. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography* 2011, **67**(Pt 4): 355-367.
120. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* 2010, **66**(Pt 2): 213-221.
121. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* 2010, **66**(Pt 4): 486-501.
122. Lebedev AA, Young P, Isupov MN, Moroz OV, Vagin AA, Murshudov GN. JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallographica Section D: Biological Crystallography* 2012, **68**(Pt 4): 431-440.
123. Painter J, Merritt EA. TLSMD web server for the generation of multi-group TLS models. *Journal of Applied Crystallography* 2006, **39**(1): 109-111.
124. Winn MD, Murshudov GN, Papiz MZ. Macromolecular TLS Refinement in REFMAC at Moderate Resolutions. In: Charles W. Carter, Jr., Robert MS (eds). *Methods in Enzymology*, vol. Volume 374. Academic Press, 2003, pp 300-321.
125. Joosten RP, Long F, Murshudov GN, Perrakis A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* 2014, **1**(Pt 4): 213-220.
126. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D* 2004, **60**(12 Part 1): 2256-2268.
127. Lin CH, Patel DJ. Solution Structure of the Covalent Duocarmycin A-DNA Duplex Complex. *Journal of molecular biology* 1995, **248**(1): 162-179.
128. Case D, Darden TA, Cheatham TE, Simmerling C, Wang J, Duke R, *et al.* *Amber 11*.
129. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 1983, **79**(2): 926-935.
130. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *Journal of Computational Chemistry* 2004, **25**(9): 1157-1174.

131. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* 2006, **25**(2): 247-260.
132. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 2005, **26**(16): 1781-1802.
133. Nosé S. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* 1984, **81**(1): 511-519.
134. Hoover WG. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* 1985, **31**(3): 1695-1697.
135. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* 1977, **23**(3): 327-341.
136. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* 2007, **23**(21): 2947-2948.
137. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* 2014, **42**(W1): W320-W324.
138. Oliver SG, Winson MK, Kell DB, Baganz F. Systematic functional analysis of the yeast genome. *Trends in biotechnology* 1998, **16**(9): 373-378.
139. Rochfort S. Metabolomics Reviewed: A New “Omics” Platform Technology for Systems Biology and Implications for Natural Products Research. *Journal of natural products* 2005, **68**(12): 1813-1820.
140. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, *et al.* The Human Serum Metabolome. *PLoS ONE* 2011, **6**(2): e16957.
141. Guo AC, Jewison T, Wilson M, Liu Y, Knox C, Djombou Y, *et al.* ECMDB: The E. coli Metabolome Database. *Nucleic Acids Research* 2013, **41**(D1): D625-D630.
142. Nielsen J. It Is All about Metabolic Fluxes. *Journal of bacteriology* 2003, **185**(24): 7031-7035.
143. Kim Y-G, Saghatelian A. Functional Analysis of Protein Targets by Metabolomic Approaches. In: Sieber SA (ed). *Activity-Based Protein Profiling*, vol. 324. Springer Berlin Heidelberg, 2012, pp 137-162.
144. Lu X, Zhao X, Bai C, Zhao C, Lu G, Xu G. LC–MS-based metabonomics analysis. *Journal of Chromatography B* 2008, **866**(1–2): 64-76.
145. Bennett BD, Yuan J, Kimball EH, Rabinowitz JD. Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. *Nature protocols* 2008, **3**(8): 1299-1311.
146. Baillie TA, Rettenmeier AW. Recent Advances in the Use of Stable Isotopes in Drug Metabolism Research. *The Journal of Clinical Pharmacology* 1986, **26**(6): 481-484.

147. Sheridan H, Krenn L, Jiang R, Sutherland I, Ignatova S, Marmann A, *et al.* The potential of metabolic fingerprinting as a tool for the modernisation of TCM preparations. *Journal of Ethnopharmacology* 2012, **140**(3): 482-491.
148. Clardy J, Walsh C. Lessons from natural molecules. *Nature* 2004, **432**(7019): 829-837.
149. Barbosa TM, Levy SB. The impact of antibiotic use on resistance development and persistence. *Drug Resistance Updates* 2000, **3**(5): 303-311.
150. Hand WL. Current challenges in antibiotic resistance. *Adolesc Med* 2000, **11**(2): 427-438.
151. Cars O, Högberg LD, Murray M, Nordberg O, Sivaraman S, Lundborg CS, *et al.* *Meeting the challenge of antibiotic resistance*, vol. 337, 2008.
152. Bibb Mervyn J. *Understanding and manipulating antibiotic production in actinomycetes*, vol. 41, 2013.
153. Scherlach K, Hertweck C. Triggering cryptic natural product biosynthesis in microorganisms. *Organic & biomolecular chemistry* 2009, **7**(9): 1753-1760.
154. Carlson EE, Cravatt BF. Enrichment Tags for Enhanced-Resolution Profiling of the Polar Metabolome. *Journal of the American Chemical Society* 2007, **129**(51): 15780-15782.
155. Odendaal AYT, D. J.; Carlson, E. E. Chemoselective enrichment for natural products discovery. *Chem Sci* 2011(2).
156. Zhang A, Sun H, Wang P, Han Y, Wang X. Modern analytical techniques in metabolomics analysis. *Analyst* 2012, **137**(2): 293-300.
157. Theodoridis GA, Gika HG, Want EJ, Wilson ID. Liquid chromatography–mass spectrometry based global metabolite profiling: A review. *Analytica Chimica Acta* 2012, **711**: 7-16.
158. Theodoridis G, Gika HG, Wilson ID. Mass spectrometry-based holistic analytical approaches for metabolite profiling in systems biology studies. *Mass Spectrometry Reviews* 2011, **30**(5): 884-906.
159. Theodoridis G, Gika HG, Wilson ID. LC-MS-based methodology for global metabolite profiling in metabolomics/metabolomics. *TrAC Trends in Analytical Chemistry* 2008, **27**(3): 251-260.
160. Gika HG, Theodoridis GA, Plumb RS, Wilson ID. Current practice of liquid chromatography–mass spectrometry in metabolomics and metabolomics. *Journal of pharmaceutical and biomedical analysis* 2014, **87**: 12-25.
161. Patti GJ. Separation strategies for untargeted metabolomics. *Journal of Separation Science* 2011, **34**(24): 3460-3469.
162. Rojo D, Barbas C, Rupérez FJ. LC–MS metabolomics of polar compounds. *Bioanalysis* 2012, **4**(10): 1235-1243.

163. Fairchild JN, Horvath K, Gooding JR, Campagna SR, Guiochon G. Two-dimensional liquid chromatography/mass spectrometry/mass spectrometry separation of water-soluble metabolites. *Journal of Chromatography A* 2010, **1217**(52): 8161-8166.
164. Guo Y, Gaiki S. Retention and selectivity of stationary phases for hydrophilic interaction chromatography. *Journal of Chromatography A* 2011, **1218**(35): 5920-5938.
165. Spagou K, Tsoukali H, Raikos N, Gika H, Wilson ID, Theodoridis G. Hydrophilic interaction chromatography coupled to MS for metabonomic/metabolomic studies. *Journal of Separation Science* 2010, **33**(6-7): 716-727.
166. Gaskell SJ. Electrospray: Principles and Practice. *Journal of Mass Spectrometry* 1997, **32**(7): 677-688.
167. Huhman DV, Sumner LW. Metabolic profiling of saponins in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer. *Phytochemistry* 2002, **59**(3): 347-360.
168. Vuckovic D. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography–mass spectrometry. *Anal Bioanal Chem* 2012, **403**(6): 1523-1548.
169. Meyer H, Weidmann H, Lalk M. Methodological approaches to help unravel the intracellular metabolome of *Bacillus subtilis*. *Microbial Cell Factories* 2013, **12**: 69-69.
170. Liebeke M, Meyer H, Donat S, Ohlsen K, Lalk M. A Metabolomic View of *Staphylococcus aureus* and Its Ser/Thr Kinase and Phosphatase Deletion Mutants: Involvement in Cell Wall Biosynthesis. *Chemistry & biology* 2010, **17**(8): 820-830.
171. Meyer H, Liebeke M, Lalk M. A protocol for the investigation of the intracellular *Staphylococcus aureus* metabolome. *Analytical biochemistry* 2010, **401**(2): 250-259.
172. Wittmann C, Krömer JO, Kiefer P, Binz T, Heinzle E. Impact of the cold shock phenomenon on quantification of intracellular metabolites in bacteria. *Analytical biochemistry* 2004, **327**(1): 135-139.
173. Villas-Bôas SG, Højer-Pedersen J, Åkesson M, Smedsgaard J, Nielsen J. Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast* 2005, **22**(14): 1155-1169.
174. Prasad Maharjan R, Ferenci T. Global metabolite analysis: the influence of extraction methodology on metabolome profiles of *Escherichia coli*. *Analytical biochemistry* 2003, **313**(1): 145-154.
175. Winder CL, Dunn WB, Schuler S, Broadhurst D, Jarvis R, Stephens GM, *et al.* Global Metabolic Profiling of *Escherichia coli* Cultures: an Evaluation of Methods for Quenching and Extraction of Intracellular Metabolites. *Analytical Chemistry* 2008, **80**(8): 2939-2948.

176. Letisse F, Lindley ND. An intracellular metabolite quantification technique applicable to polysaccharide-producing bacteria. *Biotechnology Letters* 2000, **22**(21): 1673-1677.
177. Hiller J, Franco-Lara E, Weuster-Botz D. Metabolic profiling of Escherichia coli cultivations: evaluation of extraction and metabolite analysis procedures. *Biotechnology Letters* 2007, **29**(8): 1169-1178.
178. Pesek JJ, Matyska MT, Loo JA, Fischer SM, Sana TR. Analysis of hydrophilic metabolites in physiological fluids by HPLC-MS using a silica hydride-based stationary phase. *Journal of Separation Science* 2009, **32**(13): 2200-2208.
179. Courant F, Pinel G, Bichon E, Monteau F, Antignac J-P, Le Bizec B. Development of a metabolomic approach based on liquid chromatography-high resolution mass spectrometry to screen for clenbuterol abuse in calves. *Analyst* 2009, **134**(8): 1637-1646.
180. Myint KT, Uehara T, Aoshima K, Oda Y. Polar Anionic Metabolome Analysis by Nano-LC/MS with a Metal Chelating Agent. *Analytical Chemistry* 2009, **81**(18): 7766-7772.
181. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, *et al.* Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics : Official journal of the Metabolomic Society* 2007, **3**(3): 211-221.
182. Katajamaa M, Orešič M. Data processing for mass spectrometry-based metabolomics. *Journal of Chromatography A* 2007, **1158**(1-2): 318-328.
183. Zhou B, Xiao JF, Tuli L, Resson HW. LC-MS-based metabolomics. *Molecular bioSystems* 2012, **8**(2): 470-481.
184. Hilario M, Kalousis A, Pellegrini C, Müller M. Processing and classification of protein mass spectra. *Mass Spectrometry Reviews* 2006, **25**(3): 409-449.
185. Saccenti E, Hoefsloot HJ, Smilde A, Westerhuis J, Hendriks MWB. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* 2014, **10**(3): 361-374.
186. Gowda H, Ivanisevic J, Johnson CH, Kurczy ME, Benton HP, Rinehart D, *et al.* Interactive XCMS Online: Simplifying Advanced Metabolomic Data Processing and Subsequent Statistical Analyses. *Analytical Chemistry* 2014, **86**(14): 6931-6939.
187. Patti GJ, Tautenhahn R, Rinehart D, Cho K, Shriver LP, Manchester M, *et al.* A View from Above: Cloud Plots to Visualize Global Metabolomic Data. *Analytical Chemistry* 2013, **85**(2): 798-804.
188. Villas-Bôas SG, Noel S, Lane GA, Attwood G, Cookson A. Extracellular metabolomics: A metabolic footprinting approach to assess fiber degradation in complex media. *Analytical biochemistry* 2006, **349**(2): 297-305.
189. Sáez M, Lagunas R. Determination of intermediary metabolites in yeast. Critical examination of the effect of sampling conditions and recommendations for obtaining true levels. *Mol Cell Biochem* 1976, **13**(2): 73-78.

190. Bolten CJ, Kiefer P, Letisse F, Portais J-C, Wittmann C. Sampling for Metabolome Analysis of Microorganisms. *Analytical Chemistry* 2007, **79**(10): 3843-3849.
191. Mallet CR, Lu Z, Mazzeo JR. A study of ion suppression effects in electrospray ionization from mobile phase additives and solid-phase extracts. *Rapid Communications in Mass Spectrometry* 2004, **18**(1): 49-58.
192. Weibel KE, Mor J-R, Fiechter A. Rapid sampling of yeast cells and automated assays of adenylate, citrate, pyruvate and glucose-6-phosphate pools. *Analytical biochemistry* 1974, **58**(1): 208-216.
193. Gonzalez B, François J, Renaud M. A rapid and reliable method for metabolite extraction in yeast using boiling buffered ethanol. *Yeast* 1997, **13**(14): 1347-1355.
194. Hans M, Heinzle E, Wittmann C. Quantification of intracellular amino acids in batch cultures of *Saccharomyces cerevisiae*. *Applied microbiology and biotechnology* 2001, **56**(5-6): 776-779.
195. Hajjaj H, Blanc PJ, Goma G, François J. *Sampling techniques and comparative extraction procedures for quantitative determination of intra- and extracellular metabolites in filamentous fungi*, vol. 164, 1998.
196. Marshall AG, Hendrickson CL, Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews* 1998, **17**(1): 1-35.
197. Cubbon S, Antonio C, Wilson J, Thomas-Oates J. Metabolomic applications of HILIC–LC–MS. *Mass Spectrometry Reviews* 2010, **29**(5): 671-684.
198. Cook HA, Hu W, Fritz JS, Haddad PR. A Mechanism of Separation in Electrostatic Ion Chromatography. *Analytical Chemistry* 2001, **73**(13): 3022-3027.
199. Nodwell MB, Menz H, Kirsch SF, Sieber SA. Rugulactone and its Analogues Exert Antibacterial Effects through Multiple Mechanisms Including Inhibition of Thiamine Biosynthesis. *Chembiochem : a European journal of chemical biology* 2012, **13**(10): 1439-1446.
200. Begley TP, Downs DM, Ealick SE, McLafferty FW, Van Loon APGM, Taylor S, et al. Thiamin biosynthesis in prokaryotes. *Arch Microbiol* 1999, **171**(5): 293-300.
201. Park J-H, Burns K, Kinsland C, Begley TP. Characterization of Two Kinases Involved in Thiamine Pyrophosphate and Pyridoxal Phosphate Biosynthesis in *Bacillus subtilis*: 4-Amino-5-Hydroxymethyl-2-Methylpyrimidine Kinase and Pyridoxal Kinase. *Journal of bacteriology* 2004, **186**(5): 1571-1573.
202. Eliot AC, Kirsch JF. PYRIDOXAL PHOSPHATE ENZYMES: Mechanistic, Structural, and Evolutionary Considerations. *Annual review of biochemistry* 2004, **73**(1): 383-415.
203. Percudani R, Peracchi A. A genomic overview of pyridoxal-phosphate-dependent enzymes. *EMBO Reports* 2003, **4**(9): 850-854.
204. Mooney S, Leuendorf J-E, Hendrickson C, Hellmann H. Vitamin B6: A Long Known Compound of Surprising Complexity. *Molecules* 2009, **14**(1): 329.

205. di Salvo ML, Contestabile R, Safo MK. Vitamin B6 salvage enzymes: Mechanism, structure and regulation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 2011, **1814**(11): 1597-1608.
206. Clardy J, Fischbach MA, Walsh CT. New antibiotics from bacterial natural products. *Nature biotechnology* 2006, **24**(12): 1541-1550.
207. Newman DJ, Cragg GM. Natural products as sources of new drugs over the last 25 years. *Journal of natural products* 2007, **70**(3): 461-477.
208. Drahl C, Cravatt BF, Sorensen EJ. Protein-reactive natural products. *Angew Chem Int Ed Engl* 2005, **44**(36): 5788-5809.
209. Gersch M, Kreuzer J, Sieber SA. Electrophilic natural products and their biological targets. *Natural product reports* 2012, **29**(6): 659-682.
210. Singh J, Petter RC, Baillie TA, Whitty A. The resurgence of covalent drugs. *Nature reviews Drug discovery* 2011, **10**(4): 307-317.
211. Johnson DS, Weerapana E, Cravatt BF. Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future medicinal chemistry* 2010, **2**(6): 949-964.
212. Uehara Y, Fisher JM, Rabinovitz M. Showdomycin and its reactive moiety, maleimide. A comparison in selective toxicity and mechanism of action in vitro. *Biochemical pharmacology* 1980, **29**(16): 2199-2204.
213. Böttcher T, Sieber SA. Showdomycin as a versatile chemical tool for the detection of pathogenesis-associated enzymes in bacteria. *Journal of the American Chemical Society* 2010, **132**(20): 6964-6972.
214. Carlson EE, Cravatt BF. Enrichment tags for enhanced-resolution profiling of the polar metabolome. *Journal of the American Chemical Society* 2007, **129**(51): 15780-15782.
215. Carlson EE, Cravatt BF. Chemoselective probes for metabolite enrichment and profiling. *Nat Methods* 2007, **4**(5): 429-435.
216. Yuan W, Edwards JL, Li S. Global profiling of carbonyl metabolites with a photo-cleavable isobaric labeling affinity tag. *Chem Commun (Camb)* 2013, **49**(94): 11080-11082.
217. Totaro KA, Okandeji BO, Sello JK. Use of a multicomponent reaction for chemoselective derivatization of multiple classes of metabolites. *Chembiochem : a European journal of chemical biology* 2012, **13**(7): 987-991.
218. Kirby AJ, Lima MF, da Silva D, Nome F. Nucleophilic attack by oxyanions on a phosphate monoester dianion: the positive effect of a cationic general Acid. *Journal of the American Chemical Society* 2004, **126**(5): 1350-1351.
219. Kirby AJ, Lima MF, da Silva D, Roussev CD, Nome F. Efficient intramolecular general acid catalysis of nucleophilic attack on a phosphodiester. *Journal of the American Chemical Society* 2006, **128**(51): 16944-16952.

220. Kirby AJ, Tondo DW, Medeiros M, Souza BS, Priebe JP, Lima MF, *et al.* Efficient intramolecular general-acid catalysis of the reactions of alpha-effect nucleophiles and ammonia oxide with a phosphate triester. *Journal of the American Chemical Society* 2009, **131**(5): 2023-2028.
221. Burns JA, Butler, J.C., Moran, J., Whitesides, G.M. Selective reduction of disulfides by tris(2-carboxyethyl)phosphine. *The Journal of organic chemistry* 1991, **56**: 2648-2650.
222. Shafer DE, Inman JK, Lees A. Reaction of Tris(2-carboxyethyl)phosphine (TCEP) with maleimide and alpha-haloacyl groups: anomalous elution of TCEP by gel filtration. *Analytical biochemistry* 2000, **282**(1): 161-164.
223. Barrett AGM, Broughton HB, Attwood SV, Gunatilaka AAL. Two total syntheses of showdomycin and related studies. *J Org Chem* 1986, **51**: 495-503.
224. Hendlin D, Stapley EO, Jackson M, Wallick H, Miller AK, Wolf FJ, *et al.* Phosphonomycin, a new antibiotic produced by strains of streptomyces. *Science* 1969, **166**(3901): 122-123.
225. Nishimura H, Mayama M, Komatsu Y, Kato H, Shimaoka N, Tanaka Y. Showdomycin, a New Antibiotic from a Streptomyces Sp. *The Journal of antibiotics* 1964, **17**: 148-155.
226. Roemer T, Xu D, Singh SB, Parish CA, Harris G, Wang H, *et al.* Confronting the challenges of natural product-based antifungal discovery. *Chemistry & biology* 2011, **18**(2): 148-164.
227. Cox CL, Tietz JI, Sokolowski K, Melby JO, Doroghazi JR, Mitchell DA. Nucleophilic 1,4-additions for natural product discovery. *ACS chemical biology* 2014, **9**(9): 2014-2022.
228. Liebeke M, Dorries K, Zuhlke D, Bernhardt J, Fuchs S, Pane-Farre J, *et al.* A metabolomics and proteomics study of the adaptation of Staphylococcus aureus to glucose starvation. *Molecular Biosystems* 2011, **7**(4): 1241-1253.
229. Orth R, Bottcher T, Sieber SA. The biological targets of acivicin inspired 3-chloro- and 3-bromodihydroisoxazole scaffolds. *Chem Commun* 2010, **46**(44): 8475-8477.
230. Priem T, Bouteiller C, Camporese D, Romieu A, Renard P-Y. Synthesis and reactivity of a bis-sultone cross-linker for peptide conjugation and [¹⁸F]-radiolabelling via unusual "double click" approach. *Org Biomol Chem* 2012, **10**(5): 1068-1078.
231. Rudolf GC, Koch MF, Mandl FAM, Sieber SA. Subclass-Specific Labeling of Protein-Reactive Natural Products with Customized Nucleophilic Probes. *Chemistry – A European Journal* 2015, **21**(9): 3701-3707.

IV - Curriculum vitae

Maximilian Felix Koch, born 3rd December 1986 in Grünstadt, Germany

Doctoral studies	10/2015	<p>Doctoral studies: Supervisor Prof. Dr. Stephan A. Sieber, Chair of Organic Chemistry II, Technische Universität München.</p> <p>Projects:</p> <p>(I) <i>Establishment of a targeted and untargeted metabolite profiling platform for the analysis of metabolite levels in bacteria</i></p> <p>(II) <i>Structural, biochemical and computational studies reveal the mechanism of selective aldehyde dehydrogenase 1A1 inhibition by cytotoxic duocarmycin analogs</i></p>
Academic education	09/2009-10/2011	<p>Master of Science in chemistry Technische Universität München Focus: Organic Chemistry and Biological Chemistry</p> <p>Master's Thesis Supervisor Prof. Dr. Stephan A. Sieber, Chair of Organic Chemistry II, Technische Universität München.</p> <p>Project: <i>Development of an LC-MS based global metabolome analysis approach for the study of Bacillus subtilis</i></p>
	09/2006-09/2009	<p>Bachelor of Science in Chemistry Technical University Munich</p> <p>Bachelor's Thesis Supervisor Katrin Kloth, PhD student at the Institut of Hydrochemistry and Chair of Analytic chemistry, Technische Universität München</p> <p>Project: <i>Optimierung der Tetracyclin-Detektion in Rohmilch am Munich Chip Reader 3</i></p>

**School
education**

- | | |
|-----------------|---|
| 09/1998-06/2006 | Deutsche Schule Paris , Paris
Higher education entrance qualification
("Abitur" certificate) |
| 09/1997-07/1998 | Leininger Gymnasium , Grünstadt |
| 09/1993-09/1997 | Dekan-Ernst-Schule , Grünstadt |

Conferences

- | | |
|---------|---|
| 07/2014 | LMU-Harvard Young Scientist Forum,
Cambridge |
| 10/2013 | Chrom-Forum, Ismaning |
| 07/2013 | LMU-Harvard Young Scientist Forum, Munich |

Seminars

- | | |
|---------|---|
| 04/2014 | <i>Orbitrap Fusion Tribrid Biotech Operations</i> ,
training course, Technical University Munich |
| 01/2013 | <i>Project Management in Biotech Industries</i> ,
Technical University Munich |
| 11/2012 | <i>Führung von Teams in Projekten</i> , TUM
Graduate School |
| 06/2011 | <i>Chemical Biology – Interactive Seminar</i> with
Prof. Waldmann, <i>TUM Graduate School</i> |