TECHNISCHE UNIVERSITÄT MÜNCHEN

Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt

Lehrstuhl für Analytische Lebensmittelchemie

Isolation and characterization of antivirals against HIV-1 from plant extracts

Theresa Maria Bader

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Erwin Grill

Prüfer der Dissertation:

      1. apl. Prof. Dr. Philippe Schmitt-Kopplin
      2. Prof. Dr. Karl-Heinz Engel

Die Dissertation wurde am 10.12.2018 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung Umwelt am 20.05.2019 angenommen.

„Für Erik – Ein Hoch auf uns"

Contents

Contents

Contents

I   List of Tables

List of Tables

II   List of Figures

List of Figures

List of Figures

List of Figures

III   List of Abbreviations

1D	One-dimensional
2D	Two-dimensional
%	Percent
(-)	Negative electrospray mode
(+)	Positive electrospray mode
°C	Degree Celsius
µg	Microgram
µ	Microliter
ACN	Acetonitrile
AcOEt	Ethyl acetate
ADME	Activity, stability and absorption, distribution, metabolism and excretion
AIDS	Acquired immunodeficiency syndrome
$AlCl_3$	Aluminium chloride
Anti/ Anti	Antibiotic-antimycotic
API	Atmospheric pressure ionization
Ar	Area
ART	Antiretroviral treatment
Avg	Average
AZT	Azidothymidin
B	Magnetic field strength
BB	Blocking buffer
BBr_3	Boron tribromide
B.C.	Before Christ
C	Circ
CA	Coreceptor antagonists
Casp9	Caspase 9
CCR5	C-C chemokine receptor type 5
CD4	Cluster of differentiation 4-rezeptor
CDC	C-X-C chemokine receptor type 4
CH_2Cl_2	Dichloromethane
Cistus	*Cistus incanus*
CID	Collision induces dissociation
CMC	Chemical medicinal chemistry database
CNE	Cyanopropyl endcapped
CXCR4	C-X-C chemokine receptor type 4
Da	Dalton
DKA	α,γ-diketo acids
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DsRed1	Red fluorescent protein
EASY-HIT	Full-replication exploratory assay system for the discovery of HIV inhibitors
EC_{50}	Half maximal effective concentration
EI	Electron Impact

List of Abbrevations

Env	Envelop
ER	Endoplasmic reticulum
ESI	Electrospray Ionization
eV	Electron volt
F_L	Lorentz Force
FBS	Fetal bovine serum
FC	Fraction
FI	Fusion inhibitor
FID	Free induction decay, flame ionization detector
FT-ICR-MS	Fourier transform ion cyclotron resonance mass spectrometry
Gag	Group-antigen
GC	Gas chromatography
gp120	Envelope glycoprotein 120
gp41	Glycoprotein 41
h	Hour
HAART	Highly active antiretroviral Therapy
HCS	High content screening
HCPP	High content phenotypic profiling
H/C	Hydrogen to carbon ratio
HIV	Human immunodeficiency virus
HIV-1	Human immunodeficiency virus typ-1
HIV-2	Human immunodeficiency virus typ-2
HPLC	High performance liquid chromatography
HSV-1	Herpes simplex virus Typ 1
I	Intensity
IN	Integrase
kb	Kilobase
LC	Liquid chromatography
LOPAC® 1280	Library of Pharmacologically Active Compounds
LT	Long terminal repeat
LWR	Shape measure based on ratio of length to wide of object-aligned bounding box
Lyso	Lysosomes
m	Million
mAu	Milli-absorbance-units
Max.	Maximal
MeOH	Methanol
$MgSO_4$	Magnesium sulfate
m_i	Ion mass
min	Minutes
Mito	Mitochondrium
ml	Milliliter
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromid
MW	Megaword
m/z	Mass to charge ratio
NCp7	Nucleocapsid protein 7
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells

NIAIDS	National Institute of Allergy and Infectious Disease database
nm	Nanometer
(N)NRTI	(Non-)Nucleoside Reverse Transcriptase Inhibitors
NMR	Nuclear magnetic resonance spectroscopy
NP	Normal phase
Nq	Not quantifiable
O/C	Oxygen to carbon ratio
Tulsi	*Ocimum tenuiflorum*
P2A	Shape measure based on ratio of perimeter squared to 4π*area
P53	Protein 53
PAI	Post-attachment inhibitor
PB	Permeabilization Buffer
PBS	Phosphate buffered saline
PDA	Photodiode array detection
PI	Protease Inhibitors
Pol	Polymerase
PP	Polyphenol enriched
PPL	Modified styrene-divinylbenzene polymer
ppm	Part per million
P-value	Probability value
PVPP	Polyvinylpyrrolidone
q	Electrical charge
QC	Quality control
r	Radius
r_S	Spearman correlation coefficient
Rat	Ratio
Rev	Regulator of expression of virion proteins
RNA	Ribonucleic acid
RP	Reversed phase
rpm	Rounds per minute
RSD	Relative standard deviation
RT	Reverse transcriptase, room temperature
Rt	room temperature
SAR	Structure / activity relationship
SD	Standard deviation
SDS	Sodium dodecylsulfate
sec	Second
S/N	Signal-to-noise ratio
Sp	Spot
SpC	Spot Count
SPE	Solid phase extraction
T	Total
Tat	Trans-Activator of Transcription
T-cell	T-lymphocyte
THF	Tetrahydrofuran
TIC	Total ion current
USA	United States of America
UV	Ultraviolet

List of Abbrevations

V	Volt
v	Velocity
ω_c	Angular velocity (cyclotron frequency)
WB	Wash Buffer
WHO	World Health organization

IV   Acknowledgements

V   Summary

The discovery of new bioactive compounds is crucial for the worldwide drug discovery process. Especially HIV research is dependent on new drugs, as the disease is still not curable and around 1 million worldwide HIV related deaths are reported every year. Moreover, novel bioactive compound structures are of common interest due to the appearance of viruses with multiple resistances against the existing drugs. One of the world's richest sources of bioactive compound structures are plant extracts which still exceed a common synthetic laboratory in the number of different substance structures by fare. Whereas the vast majority of bioactivity screenings are primarily based on targeted assays of selected compounds, the identification of novel bioactive structures from plant extracts is necessarily based on a non-targeted high content phenotypic profiling (HCPP) and MS approach.

The aim of this thesis was to investigate the broad bioactivity potential of plant extracts and to identify novel anti-HIV active substances. Therefore, a novel fractionation setup was established using two-dimensional liquid chromatography combined with a PDA detector and ultrahigh resolution mass spectrometry. To monitor the fractionation process and to characterize the bioactivity of the fractions a non-targeted HCPP screening system was combined with a HIV-1 infection assay (EASY-HIT). The biological data were correlated to the compound masses determined by mass spectrometry using different data filtering and enrichment techniques and additional correlation algorithms. These correlations lead to the discovery of a few top correlating compound masses. The fractionation method was performed with two different plants namely *Cistus incanus* (Cistus) and *Ocimum tenuiflorum* (Tulsi). The combination of the fractionation and the non-targeted high content phenotypic profiling resulted in the identification of several distinct bioactivities in both plant extracts.

In Cistus extracts a cluster of an anti-HIV standard and a subset of fractions was identified based on the HCPP data. The anti-HIV-1 activity of the clustered fractions could be confirmed by the targeted EASY-HIT assay verifying the potency of this approach. The correlation of the HCPP and EASY-HIT bioactivity data together with the gained ultrahigh resolution mass spectrometry data resulted in a list of 17 significantly correlating and enriched compound masses. The novel combinational approach leads to the identification and ranking of a few potential bioactive substances from a very complex plant raw extract. Further LC-MS and MS/MS experiments revealed the presence of 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid in *Cistus incanus*. This HIV-1 integrase inhibitor could be isolated and characterized for the first time in *Cistus incanus* and can now explain parts of the shown anti-HIV-1 activity of the raw extract.

In the second plant extract, Tulsi, several bioactivities could be disclosed after the fractionation process and HCPP gives additional insights into the cellular effects. The gained MS data of the bioactive fractions were used to further characterize the enriched and unique metabolites in the bioactive fractions. This approach leads to the discovery of one (out of 180) fraction clustering with three LOPAC® standards containing the well-known bioactive piperazine subunit as lowest common denominator.

All in all, several bioactivities and potential responsible metabolites of the investigated plants were revealed using 2D LC, MS/MS, HCPP and the EASY-HIT. The bioactivities were correlated to a few single

Summary

metabolites out of over 1000 annotated metabolites existing in the plant raw extract, they were ranked and further characterized on a metabolite level using ultrahigh resolution mass spectrometry. Using this approach, we could identify the anti-HIV-1 integrase inhibitor 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid for the first time in *Cistus incanus*

VI   *Zusammenfassung*

Die Entdeckung von Wirkstoffen mit neuen Strukturen ist ausschlaggebend, um das Portfolio bestehender Medikamente zu erweitern. Diese werden vor allem in der HIV Forschung dringend benötigt, da zum einen AIDS immer noch nicht heilbar ist, denn jährlich sterben weltweit 1 Millionen Menschen an den Folgen der Krankheit. Zum anderen viele Viren bereits Resistenzen gegen alle erhältlichen Medikamente zeigen. Um HIV trotz Resistenzen effektiv zu behandeln werden Wirkstoffe mit neuen Strukturen benötigt. Eines der größten Vorkommen dieser Wirkstoffe sind Pflanzenextrakte welche synthetische Labore in ihrer Diversität um ein Vielfaches übertreffen.

Das Ziel dieser Arbeit war es durch die Kombination von phänotypischem Screening und ultrahochauflösender Massenspektrometrie, das bioaktive Potential eines Pflanzenextraktes zu analysieren und neue Anti-HIV aktive Substanzen zu finden. Dafür wurde eine zwei dimensionale (2D) Fraktionierungsmethode entwickelt, welche auf Flüssigchromatographie beruht kombiniert mit einem PDA Detektor und ultrahochauflösender Massenspektrometrie. Um den Fraktionierungsprozess zu bewerten und die Fraktionen auf ihre Bioaktivität hin zu überprüfen wurde ein umfangreiches non-targeted phänotypisches Screening sowie ein Anti-HIV-1 Assay (EASY-HIT) durchgeführt. Die biologischen Daten wurde mit den über das Massenspektrometer ermittelten Metabolit Massen korreliert. Diese Korrelation, basierend auf verschiedene Datenfilterungen, Anreicherungen und Korrelationsalgorithmen, führte zu der Entdeckung einiger signifikant korrelierender potential bioaktiver Massen. Die Fraktionierung wurde mit zwei verschiedenen Pflanzen durchgeführt, *Cistus incanus* (Cistus) and *Ocimum tenuiflorum* (Tulsi). Die Kombination von Fraktionierung und umfangreichem non-targeted phänotypischen Screening, resultierte in der Entdeckung von mehreren Bioaktivitäten und potentiell verantwortlichen Metaboliten in beiden Pflanzenextrakten.

Im Cistus Extrakt wurde ein Anti-HIV-1 aktiver Cluster im phänotypischen Screening identifiziert, welche aus einer fortlaufenden Reihe von Fraktionen und einem Anti-HIV aktiven Standard besteht. Die Anti-HIV-1 Aktivität dieser clusternden Fraktionen konnte durch einen targeted Anti-HIV-1 Test erwiesen werden, was die Stärke der Vorhersage des phänotypischen Screenings noch einmal bestätigt. Die Korrelation der Bioaktivitätsdaten (phänotypisches Screening und EASY-HIT) zusammen mit den Daten des ultrahochauflösenden Massenspektrometers führte zu einer Liste mit 17 signifikant korrelierenden und angereicherter Massen. Diese neue kombinatorische Vorgehensweise führte zu der Identifikation und Klassifizierung von potentiell bioaktiven Massen innerhalb eines sehr komplexen Pflanzenextrakts. Weiterführende LC-MS und MS/MS Experimente zeigten das Vorhandensein von 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid in Cistus incanus. Dieser HIV-1 Integrase Inhibitor konnte mit Hilfe dieser Arbeit das erste Mal in Cistus isoliert und charakterisiert werden.

In dem Pflanzenextrakt von Tulsi konnten nach dem Fraktionierungsprozess mehrere Bioaktivitäten auf zellulärer Basis durch das phänotypische Screening identifiziert werden. Durch die Massenspektrometriedaten der bioaktiven Fraktionen wurde die metabolische Vielfalt in diesen Fraktionen charakterisiert. Dieser Ansatz führte zur Identifikation von Piparazin als gemeinsame Teilstruktur, welche für einen gezeigte bioaktiven Cluster potentiell verantwortlich ist.

Zusammenfassung

Zusammengefasst kann gesagt werden, dass durch die Kombination von 2D LC, MS/MS, phänotypischem Screening und EASY-HIT mehrere Bioaktivitäten und die dazu potenziell verantwortlichen Metaboliten aus den untersuchten Pflanzen isoliert und charakterisiert werden konnten.

Die gezeigten Bioaktivitäten der jeweiligen Pflanzenextrakte wurden mit einzelnen Metaboliten, aus dem sehr komplexen Metabolitengemisch der Pflanze, korreliert. Durch die weitere Charakterisierung der Metaboliten mittels ultrahochauflösenden Massenspektrometer konnten wir den anti-HIV-1 Integrase Inhibitor 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid das erste Mal in *Cistus incanus* nachweisen.

1. General introduction

The number of acquired immunodeficiency syndrome (AIDS) related death was 1 million people in 2016. Even though the count is decreasing from 1.9 million people in 2006, still 890 000 adults and 120 000 children died because of AIDS in 2016. AIDS was first found in 1981 in a group of homosexual man. The research claimed a T-lymphocyte (T-cell) defect for the missing immune response to antigens and the reduction of the lymphocytes (Gottlieb et al., 1981; Maurin, Bailly, Mbemba, Mouscadet, & Cotelle, 2006). 1983 two independent research groups (Barre-Sinoussi et al., 1983; Gallo et al., 1983) proved, that a virus is responsible for the missing immune response and in 1986 the virus was identified and named human immunodeficiency virus (HIV) (Coffin et al., 1986). An untreated infection with HIV is causing AIDS after a latency phase that differ in duration (up to 10 years). HIV destroys the immune cells and impair their function leading to an immunodeficiency. Due to that, an increased susceptibility for a wide range of infections and other diseases e.g. cancer is occurring, HIV infected people can't fight off. By 2016 the World Health organization (WHO) published that estimated 36.7 million (m) people are now living with HIV globally. From that count 25.6 m live in Africa, 3.5 m in South-East Asia, 3.3 m in America, and 2.4 m in Europe to name the highest occurrences. Even though the trend of new infections is decreasing from 3.3 m in 1996 to 1.8 m people in 2016 (2016 globally 0.26% human being from all ages got infected with HIV) still an enormous amount of people is suffering from this disease. Because of the actual medical possibilities, the live expectancy of a HIV infected person is estimated to 63 years if an infected child is born in 2016 (www.who.int/hiv/en/; 20.11.2018). In 2016, 19.5 m people received antiretroviral treatment (ART), signifying a treatment coverage of 53% which is an immense increase compared to the coverage of 23% in 2010.

There is still an urgent need for the discovery of new, safe, efficient and cheap anti-HIV drugs as no healing is currently possible and additionally due to the gained resistances of the virus against the available drugs. It is known, that theoretically drug modeling and drug synthesis in the laboratory towards targets in the HIV replication cycle is limited in its ability to generate new drug scaffolds (David J. Newman & Cragg, 2016). Drugs against almost all disease areas were first found in the plant kingdom, like one of the most famous drugs aspirin (D. J. Newman, Cragg, & Snader, 2000) and morphium (Grabley, Thiericke, & Sattler, 2000). Plant metabolomics studies are necessary to reveal new bioactive scaffolds against HIV (Kashman (Kashiwada et al., 1996; Kashman et al., 1992).

1.1. Metabolomics

The term "metabolomics" was first defined by Oliver Fiehn 2001 and it described the characterization and quantification of all metabolites an organism holds (Fiehn, 2001; Nicholson & Lindon, 2008). Metabolomics studies often deal with the impact of drugs on the whole metabolism (Goodacre, Vaidyanathan, Dunn, Harrigan, & Kell, 2004; Robertson, Watkins, & Reily, 2011) and metabolomics is embedded in the system biology. Other "omics" techniques in the field of system biology like genomics, transcriptomics and proteomics are summarized in Figure 1.

1. Introduction

In general, system biology is investigating a system, defined as molecular collective, biological processes, physiological functions and structures and its changes due to external influences. Metabolomics is nowadays a new main group of this system biology (Faber et al., 2007), see Figure 1, as it monitor the terminal downstream products of the genome. Direct impact of diseases on the metabolite composition, like. e.g. diabetes on the blood sugar concentration can be monitored.
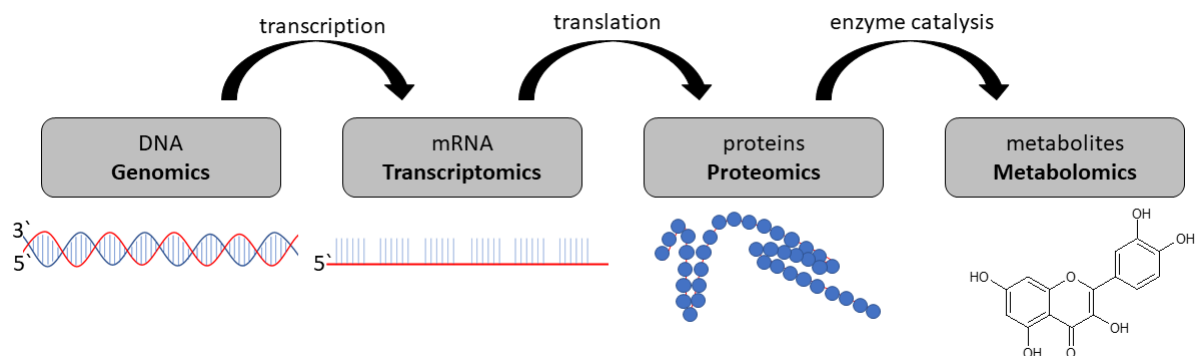
Figure 1: System biology overview and its -omics cascade, deoxyribonucleic acid = DNA, messenger ribonucleic acid = mRNA.

Metabolites are per definition small molecules (< 1000 Dalton (Da)) an organism is synthesizing, absorbing, metabolizing or catabolizing (Oms-Oliu, Odriozola-Serrano, & Martín-Belloso, 2013; Villas-Boas, Mas, Akesson, Smedsgaard, & Nielsen, 2005). The sum of all metabolites in one organism are called metabolome. In general, there are two forms of metabolomics, namely non-targeted metabolomics and targeted metabolomics.

The targeted approach is based on pre-defined metabolite or metabolic class that were selectively extracted, measured and analyzed. The advantages here are the optimized sample preparation and metabolite analysis leading to high time and cost efficiency. One huge disadvantage is the biased view which can result in e.g. undetected impurities or side effects.

In case of non-targeted metabolomics, the entire metabolome is investigated. The challenges of this approach are the time-consuming data processing steps (Bowen & Northen, 2010) and difficulties in identifying and characterizing unknown metabolites (Moco et al., 2007). The advantages lie in the possibility to reveal totally unknown drug scaffolds, drug targets and possible metabolomic side effects. This is possible due to a very unrestricted sample pretreatment and high detector selectivity with broad specificity leading to the identification of new biomarkers for diseases or drug scaffolds. A typical non-targeted metabolomic workflow is illustrated in Figure 2.

6
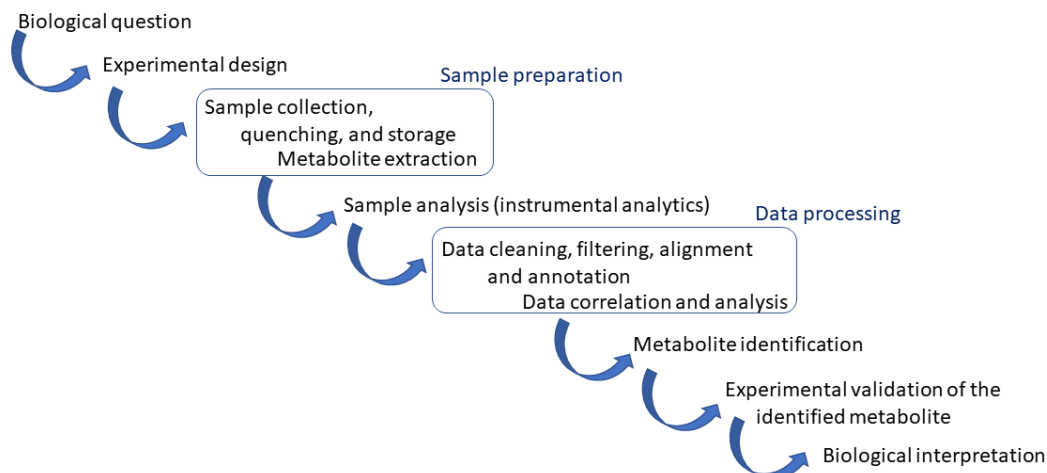
Figure 2: Typical non-targeted metabolomic workflow.

1.2. Techniques in plant metabolomics

Plants produce an extraordinary diversity of 0.2 to 1 million different metabolites (Dixon & Strack, 2003; Saito & Matsuda, 2010). Additionally, different plant family's synthetize distinct sets of specified metabolites like for example *Arabidopsis thaliana* with estimated 3,500 specialized metabolites. In comparison to that, animals are only synthetizing 5,000 to 25,000 different metabolites (Trethewey, 2004). The huge plant metabolomics diversity (amino acids, amines, organic acids, sugars, alkaloids, polyphenols, fatty acids, lipids, isoprenoids, nucleic acids, pigments, volatiles) is also implementing a wide range of different solubilities, polarities, stabilities and metabolite concentrations (Saito & Matsuda, 2010; Sumner, 2010). Taking this into account, an analytical platform with complementary analytical techniques needs to be set up, to cover the different metabolite modalities (Moing et al., 2011). In general, multiple techniques are available like gas chromatography (GC), liquid chromatography (LC) or capillary electrophoresis (CE) coupled to different detectors like, photodiode array detection (PDA), flame ionization detector (FID), mass spectrometry (MS) or nuclear magnetic resonance spectroscopy (NMR) to name the most prominent (Kusano et al., 2011). To cover this diversity in an non-targeted metabolomics experiment a careful experimental design is necessary including plant harvesting, sample extraction and the choice of the analytical techniques (Allwood et al., 2011; Ryan & Robards, 2006).

1.2.1. Extraction

Prior to the extraction the plants need to be carefully harvested. Important variables which might affect the metabolic content are the location, the time point and the developmental stage of the plant (H. K. Kim, Choi, & Verpoorte, 2010; Parcerisa, Codony, Boatella, & Rafecas, 1999; Queiroz & Morel, 1974). Moreover, the samples should be frozen under standardized conditions at -196°C, as different metabolic processes are still active after the harvesting process (Fiehn et al., 2007; Moritz, 2007), and stored

1. Introduction

at - 80°C (Mushtaq, Choi, Verpoorte, & Wilson, 2014). Of note, certain metabolite classes might be more stable in different harvesting conditions for example bioactive phenolic compounds are conserved the best when the sample is air dried and then stored at -20°C or cooler (Mediani, Abas, Tan, & Khatib, 2014). The next step in the extraction workflow is the carefully grinding of the sample, without raising the temperature, to a homogenic powder which is necessary for the following extraction process (Ernst, Silva, Silva, Vencio, & Lopes, 2014; Mushtaq et al., 2014).

The extraction itself is an essential step in the analytical workflow. It needs to be as short as possible to avoid starting biochemicals processes but needs to be reproductive and potent enough to extract all metabolites of interest (Weckwerth, 2007). As it is not possible to extract all plant metabolites with one solvent (H. K. Kim et al., 2010) its crucial to carefully select solvents based on their selectivity, flashpoint, pH value, grade and toxicity. The most abundant solvents for the extractions are methanol, acetonitrile, chloroform, water or mixtures of them (Martin et al., 2014; Villas-Boas et al., 2005; Wolfender, Rudaz, Choi, & Kim, 2013). Other factors like the ratio of solvent volume to sample volume, extraction time, temperature and mechanical impute like ultra-sonification also need to be considered.

Because of the variety of extraction possibilities there is no standardized plant extraction method and therefore the procedure need to be adapted for each plant and specific research question. After the successfully extraction of the metabolites in the extracts they need to be separated using chromatography for further investigation.

1.2.2. Liquid chromatography

The field of chromatography can be divided into gas chromatography and liquid chromatography. Gas chromatography is chosen if the compounds of interest are volatile and thermally stable (fragrances in perfume for example). In contrast, liquid chromatography is employed if the compounds are thermal unstable, large, highly polar and therefore cannot enter the gas phase. For plant metabolomics one of the most common separation technique is liquid chromatography coupled to a mass spectrometer (Allwood & Goodacre, 2010). Liquid chromatography is performed prior to the actual detector because, i) metabolites got separated ii) the matrix effect can be reduced, iii) isomers can be separated with higher probability, iv) additional data (retention time) are recorded v) less ion suppression takes place in e.g. the MS-detector.

1.2.2.1. Two-dimensional liquid chromatography

Due to the high complexity and amount of plant metabolites one-dimensional (1D) separation would not be sufficient as it is tailored to a few tens of compounds (J. Calvin Giddings, 1967; J. C. Giddings, 1984). For the complexity of a plant extract, a two-dimensional (2D) separation is much more efficient to separate the different metabolites. For this 2D separation, different chromatographic methods are available. First, it can be performed on-line, means that after the 1D separation, the fractions are collected and subsequently introduced into the second LC dimension. Second, the fractions are collected after the

1D, can be concentrated or modified if necessary, and were then injected into the 2nd dimension. Additionally, there is the possibility to only forward specific fractions from the 1D into the 2D (heart cut) or the entire 1D is forwarded to the 2D (comprehensive) (Schoenmakers, 2003). Furthermore, the whole variety of column and selectivities can be used (Cubbon, Bradbury, Wilson, & Thomas-Oates, 2007) like reversed phase (C18, C8, C2, PPL (modified styrene-divinylbenzene polymer) etc.) or normal phase (NP). For plants the conventional C18 reversed phase (RP) is widely used (Broeckling et al., 2005; Farag, Huhman, Lei, & Sumner, 2007). A RP C18 column was for example used in 1D to separate flavanone-7-O-glycosides from the citrus juice matrix and a carboxymethylated β-cyclodextrin-based column in the 2D to separate its stereoisomers (Aturki, Brandi, & Sinibaldi, 2004). When choosing a complementary selectivity for the 2D separation a high separation selectivity can be achieved (Farag et al., 2007). Most frequently RP x RP separations were used (Huhman & Sumner, 2002; Kind et al., 2009). Summing up, the maximum peak capacity is achieved when using different methods and columns in the 2D separating system.

1.2.3. Mass spectrometry

After separating the compounds as good as possible from their isomers and matrix, the compounds of interest need to be detected selectively and sensitively. Major challenges of the compound detection could be the low concentrations for compounds of interest and the clustering of compounds with similar physicochemical properties even after 2D separation. To overcome these challenges mass spectrometry is superior over UV and PDA detectors as they need a chromophore group which is not the case in all metabolites. Mass spectrometry is highly specific, sensitive and can identify unknown compounds by using MS/MS. Together with LC, LC-MS cope with the challenges of plant metabolite analysis (de Villiers, Venter, & Pasch, 2016; Pang et al., 2016; Smyth, Smyth, Ramachandran, O'Donnell, & Brooks, 2012).

1.2.3.1. Electrospray ionisation (ESI)

One of the most famous atmospheric pressure ionization (API) for biological samples is Electrospray Ionization (ESI) (Kostiainen & Kauppila, 2009). The technique was established by Dole in 1968 (Dole et al., 1968; Fenn, 2003). As almost no-insource fragmentation, compared to e.g. Electron Impact (EI), takes place it is defined as a soft ionization method (Watson, 2013). Coupled to an LC, the analyte solution is sprayed through a thin capillary (sprayer needle) under atmospheric pressure into an electric field. The applied electric voltage of 3-5 kV causes ionization of the analytes and ion acceleration towards the cathode. Cations, like $[M + nH]^{n+}$, and anions like $[M - nH]^{n-}$ can be produced in this way by protonation and deprotonation of functional groups, respectively. In the positive ionization mode e.g. potassium or calcium adducts can be also formed due to the high probability of salt traces in the solvents (de Hoffmann, 2013). The operating principle of an ESI is shown in Figure 3.

1. Introduction
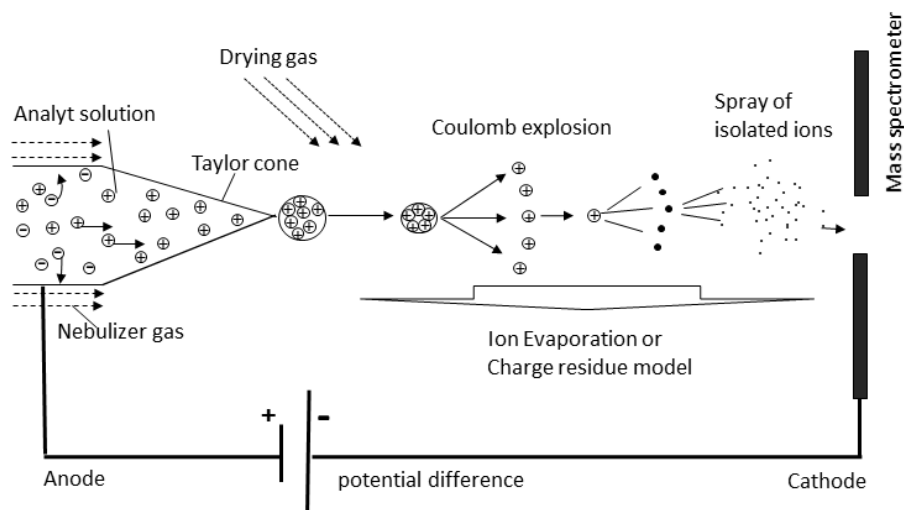
Figure 3: Scheme of an ESI source including Tylor cone, coulomb expoilions and the ion evaporation or charge residue models.

Due to the strong electrical fields on the tip of the spray capillary the negative ions move from the surface of the solution to the side of the capillary and are discharged. The positive ions in contrast can reach the solvent front (Kostiainen & Kauppila, 2009). The so called "taylor cone" is formed by the electrostatical interactions on the tip of the spray capillary. Due to raising repulsion forces in the solution, homogenous charged droplets were released. This aerosol is supported by the added nebulizer gas. Additionally, the also introduced drying gas further reduces the remaining solvent in the droplets and thereby increases the repulsion forces. At one point the repulsion force is bigger than the surface tension and the so called "Rayleigh-Limit" is exceeded. Now a micro explosion, so called Coulomb-Explosion, takes place resulting in a higher number of smaller droplets (Hommerson, Khan, de Jong, & Somsen, 2011). This process is continuing until isolated ions enter the gas phase. Of note, the exact mechanism is not known up to now (S. Nguyen & Fenn, 2007) but two models were developed and discussed: Charge Residue Model and Ion Evaporation Model (Kebarle, 2000).

A major challenge using LC-MS with ESI is still the ion suppression effect. This effect is based on competing co-eluting metabolites in the solution or matrix resulting in a uneven distribution of charge and therefore in a loss of sensitivity (Annesley, 2003). Also the polarity, mass, charge volatility, surface activity, solubility in the mobile phase, buffer concentration and flow rate of the mobile phase have an impact on the ionisation process (Annesley, 2003; King, Bonfiglio, Fernandez-Metzler, Miller-Stein, & Olah, 2000). The most efficient action to lower these effects is the separation of the analytes by LC prior to performing ESI-MS. Depending on the sample and metabolites of interest positive or negative ionization mode can be more sufficient for ion production.

1.2.3.2. Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS)

FT-ICR-MS is the mass spectrometer that offers the highest mass accuracy and ultrahigh resolving power (Nikolaev, Kostyukevich, & Vladimirov, 2014; Schmitt-Kopplin, Gelencsér, et al., 2010). An FT-ICR-MS consists of an ionization source, an ion optic and a strong magnet with an embedded ICR cell. After ion generation, ions are focused in the first quadrupole, accumulated or fragmented in the subsequent collision cell and finally detected in the ICR cell surrounded by an e.g. 12 Tesla magnet.

In the year 1974 Comisarow and Marshall introduced the FT-ICR-MS technique (Comisarow & Marshall, 1974). The used Penning ion trap consists of an electrostatic voltage superposed by a strong homogenic magnetic field alongside the trap axis of symmetry (Comisarow & Marshall, 1974). The chamber consists of six electrodes grouped in three pairs. Two of them limit the chamber on the z-axis and are therefore input and output electrodes (trapping electrodes). The mantle of the cubic form is built by two pair of opposite electrodes. One pair causes the dipolar stimulation of the ions, while the other pair is used for ion detection. Once the ions enter the trap they are forced to rotate in a circular path (cyclotron motion) with small radius in x/y direction because of the Lorenz force, which exists due to the surrounding magnetic field (see Figure 4).
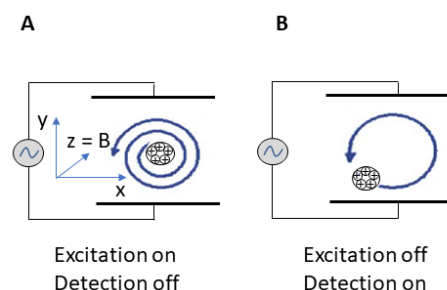
Figure 4: Open ICR cell with ion motion (circular arrow). A) exitation on and detection off, B) exitation off and detection on.

The homogeneity and strength of the magnet is responsible for the stable and reproducible cyclotron trajectory of the rotating ions. The velocity *v* and the electrical charge *q* of the ions undergo the Lorentz Force F_L in the magnetic field *B* which is oriented orthogonal to the ion`s velocity (see Figure 4) (Equation 1):

Equation 1: Lorentz Force F_L, velocity *v*, electrical charge *q* and magnetic field strength *B*

$$F_L = v \cdot q \cdot B$$

This force is responsible for the stable radius *r* of the ions and it is in equilibrium with the centrifugal force (Equation 2):

1. Introduction

Equation 2: Centrifugal force (ions mass m_i, velocity v and radius r) in equilibrium with the Lorentz Force (velocity v, electrical charge q and magnetic field B)

$$m_i \cdot \frac{v_2}{r} = q \cdot v \cdot B$$

$$\frac{m_i}{q} = \frac{r}{v} \cdot B$$

The resulting angular velocity (cyclotron frequency) ω_c can be described with the reciprocal mass to charge relation (q/m) and magnetic field strength (B) (Marshall, Hendrickson, & Jackson, 1998) (Equation 3):

Equation 3: Angular velocity (cyclotron frequency) ω_c, electrical charge q, ions mass m_i and magnetic field B

$$\omega_c = \frac{q}{m_i} B$$

The frequency is therefore dependent on the magnetic field strength and inversely propotional to the m/z (z = q) ratios of the ions. For detection, an additional oscillating radiofrequency field is necessary. This field introduced by mantle electrodes and causes an electrostatic resonant excitation. In case of resonance, the ions were exhilarated and therefore the cyclotron radius of their trajectories are enlarged, whereas the frequency stays constant as it is not dependent on ion`s kinetic energy. After this dipolar excitation, ions with the same m/z ratio are cycling on their enlarged trajectory in ion packets (see Figure 5).
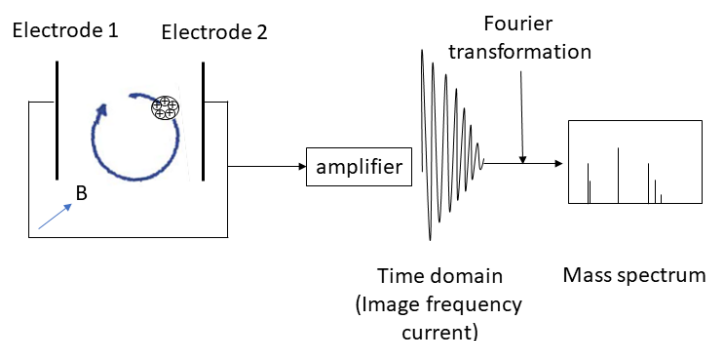
Figure 5: Detection principle of the FT-ICR-MS.

The ion packages are cycling on the radio frequency induced trajectory (Figure 5). The mantle electrodes 1 and 2 are used for the detection. For example, positively charged ions are moving from electrode 1 to electrode 2 (vice versa for negatively charged ions). The induced electrical field pulls electrons in the external power circuit though a resistance in the direction of electrode 2. After that the electrons from electrode 2 move back to electrode 1 and this movement is measured by a preamplifier as image current. This image current frequency is the same as the cyclotron frequency (w_c) of the ions and the amplitude is proportional to the number of ions in the ICR cell. The ions induce a current while going through the electrodes and due to the ion cyclotron motion it is recorded as free induced decay (FID) (Marshall et al., 1998).

When measuring a sample, which has ions of various *m/z* ratios, a short linearly increasing high frequency excitation sweep is enough to exite all the trapped ions for detection in a defined *m/z* range simultaneously. The resulting interferogram is a complex time domain signal consisting of *n* overlapping image frequency currents. Using Fourier transformation, the time signal can now be converted in the frequency domain with can be subsequently assigned to a mass spectrum.

Because of the mathematical depiction of the cyclotron frequency (w_c, Equation 3), the resulting single frequencies can be correlated to *m/z* ratios and a corresponding mass spectrum can be generated. They can be assigned to exact elemental and molecular compositions due to ultrahigh resolution und mass accuracy of the FT-ICR-MS which can be used for structure identification of unknown metabolites (Stephen, Gary, & Jean-Louis, 2005). Although FT-ICR-MS exhibits an extremely high resolution, the separation of isomeric and isobaric metabolites remains one of the main challenges. To overcome that, upstream chromatography is necessary.

1.2.4. Tools for data analysis

When performing research in the field of metabolomics, a huge amount of complex data is generated within hundreds and thousands of metabolites. For their comprehensive evaluation a specific data analysis needs to take place involving cheminformatics and statistic aspects. In addition, the data need to be interpreted with respect to their biochemical causes and physiological consequences (Mehrotra & Mendes, 2006; Robertson et al., 2011). Even though a lot of information is available, it is often very tedious and limited to correlate, implement or compare this data due to i) the complexity of the biological data where additional computational analysis methods are required, ii) the diversity of analytical measuring methods which makes the data comparison very difficult and iii) the lack of complete databases, availability of standards and the tools for integrative analysis (Goble & Stevens, 2008; Gomez-Cabrero et al., 2014).

After the sample preparation and data acquisition, the processing of the gained data is necessary prior to data integration and interpretation. Therefore, different processing steps are necessary such as data pre-processing including noise filtering, retention time correction, peak detection and chromatogram alignment (Kanawati, Bader, Wanczek, Li, & Schmitt-Kopplin, 2017). Afterwards, data preparation including data integrity checking and data normalization and conclusive data interpretation takes place. Here, it is necessary to reveal the molecular composition of the measured metabolites as exact as possible (Tziotis, Hertkorn, & Schmitt-Kopplin, 2011). Additionally for a better visualization of the metabolome data van Krevelen diagrams were used (Krevelen, 1950), which show on the x-axis the oxygen to carbon (O/C) ratio on the y-axis the hydrogen to carbon ratio (H/C).

Afterwards further interpretation like i) functional interpretation including enrichment analysis and pathway analysis, ii) metabolite mapping and iii) data integration in other -omics" fields or knowledge from literature or previous experiments can be employed.

1. Introduction

1.3.   Biological assays

Biological assays are inter alia used to investigate the impact of substances to a living system e.g. cells, tissues, animals or human beings. In general, the assays consist of a certain stimulus (drug) which is applied to a certain subject e.g. cell or animal. Afterwards the response of the subject is measured qualitatively and/or quantitatively. Biological assays are needed to prove the potency of substances in cells and tissues (*in vitro*) or animals and humans (*in vivo*). Historically, substances were often tested using a single target driven high throughput *in vitro* assays (Suzgec-Selcuk et al., 2011) and the selected candidates were then transferred to *in vivo* assays.

Nevertheless i*n vivo* tests are needed to validate in silico modeled or *in vitro* selected compounds to exclude possible off-target effects (Fishman & Porter, 2005). Recently *in vitro* cell tests were used instead of *in vitro* enzymatic tests in an early stage of the drug development process, making the process more effective as toxic effects or unexpected crosstalk from connected pathways can be seen in an early stage of the drug development process (Y. Feng, Mitchison, Bender, Young, & Tallarico, 2009). One effective *in vitro* cell screening test is a non-targeted biological screening called high content phenotypic profiling (HCPP) which predicts the effect on several different biological targets, a compound or mixture of compounds affects. This method established oneself to be very effective in screening potential bioactive single molecules and natural fractions containing a mixture of compounds (Bray et al., 2016; Caicedo, Singh, & Carpenter, 2016; Nichols, 2007; D. W. Young et al., 2008). Any HCPP predicted target needs then to be verified using a specific test, tailored to the revealed target. A strong test in this field of anti-HIV-1 would be the full-replication <u>e</u>xploratory <u>a</u>ssay <u>s</u>ystem for the discover<u>y</u> of <u>H</u>IV <u>i</u>nhibi<u>t</u>ors (EASY-HIT) (Kremb et al., 2011). The non-targeted HCPP and targeted EASY-HIT are used in this thesis and will therefore be introduced in the following chapters.

1.3.1. High content phenotypic profiling (HCPP)

Imagine based high content phenotypic profiling (HCPP) can provide insights into the mode of action, physiology and mechanism of toxicity a metabolite or metabolite mixture is causing on a single cell level by measuring its impact on cellular features (Giuliano et al., 1997; Kremb & Voolstra, 2017; Kurita, Glassey, & Linington, 2015; Swinney & Anthony, 2011; D. W. Young et al., 2008) and is therefore used in commercial drug development (Hoffman & Garippa, 2007). Hence, an automated fluorescence or luminescence microscope together with automated image analysis software is necessary and meanwhile available on the market to characterize the morphological changes like cell and organelle intensity, shape or texture (Rausch, 2006) in a high throughput manner.

When using natural extracts and their fractions consisting of a mixture of compounds, HCPP can reveal the broad biological potential by describing the cellular feature deviation of a sample in correlation with the cellular feature deviation of known standards from the Library of Pharmacologically Active Compounds, LOPAC®1280 (LOPAC®). An additional correlation with MS can reveal the metabolite (class) or metabolic subunit present in the sample, which is probably responsible for the exhibited mode of action

(Bray et al., 2016; Gustafsdottir et al., 2013; Kang et al., 2016; Sumiya et al., 2011). The applied standard sample set in this thesis was the LOPAC® Library, and its usage is described in the following chapter.

1.3.1.1. Sigma's 1280 LOPAC® clustering

To gain knowledge about the possible targets the compound(s) are hitting, the produced sample and fraction HCPP data need to be correlated to HCPP data from derived standards with known bioactivity. The LOPAC® standard library was chosen for that purpose as it consists of a mixture of 1280 pharmacologically active compounds. The standards cover a wide range of targets like gene regulation, neurotransmission, kinases and ion channels. 735 compounds of that library were selected and used in this study (Kremb & Voolstra, 2017). For the target prediction several clustering methods are available to cluster the HCPP data of the sample and LOPAC® standards like Spearman Rank, Pearson and Kendall`s Tau clustering (Reisen, Zhang, Gabriel, & Selzer, 2013; Saeed et al., 2003). An overview of the processed steps from the sample to the correlation and target prediction is displayed in Figure 6.
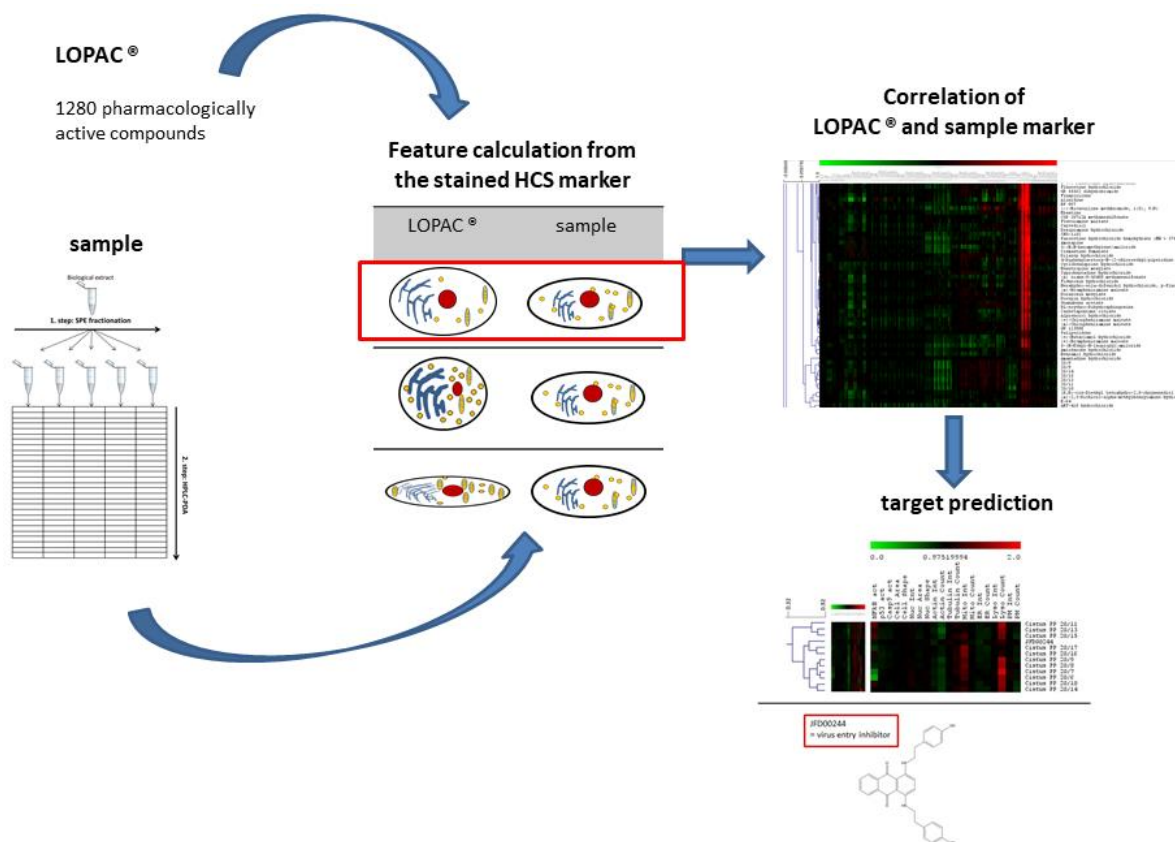
Figure 6: HCPP feature clustering of LOPAC® standards with HCPP features derived from natural sources. After data correlation, a specific biological target gets predicted.

To ensure the quality of the clustering, two different criteria are used. i) the clustering should be repeatable using different clustering algorithms like Spearman Rank, Pearson and Kendall`s Tau, as robust

1. Introduction

clustering shows the independency of the correlation to a specific algorithm and ii) the availability of support trees, which display the probability of the cluster, ensures additionally the statistical quality of the generated clusters. These two criteria help to narrow down the enormous amount of cluster, produced be HCPP, to the significant ones, which can be used for further analysis.

1.3.1.2. Possible clusters

Different types of clusters can appear when using the described clustering method (MacDonald et al., 2006): i) functional similarity cluster, ii) structural similarity cluster and iii) structure activity relationship cluster.

I. Functional similarity cluster

This kind of cluster consists of a sample and one or several standards that show a common mode of action e.g. dopamine receptor agonists (see Figure 7). These clusters share a common mode of action although the involved standards can have a total different molecular structure (Schulze et al., 2013).
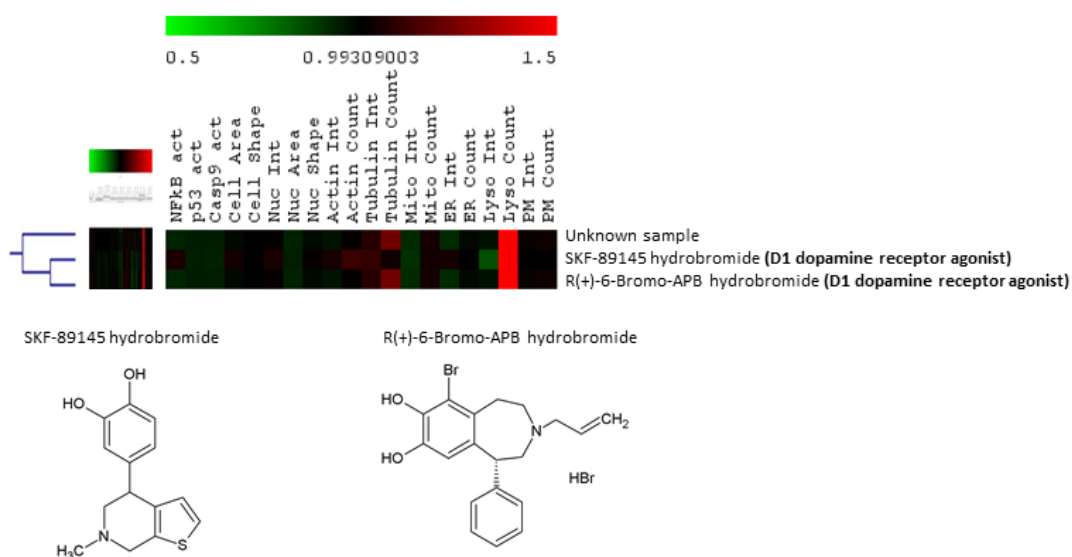
Figure 7: Example of a functional similarity cluster.

II. Structural similarity

In this case a cluster appeared involving standards with a similar substructure (see Figure 8).
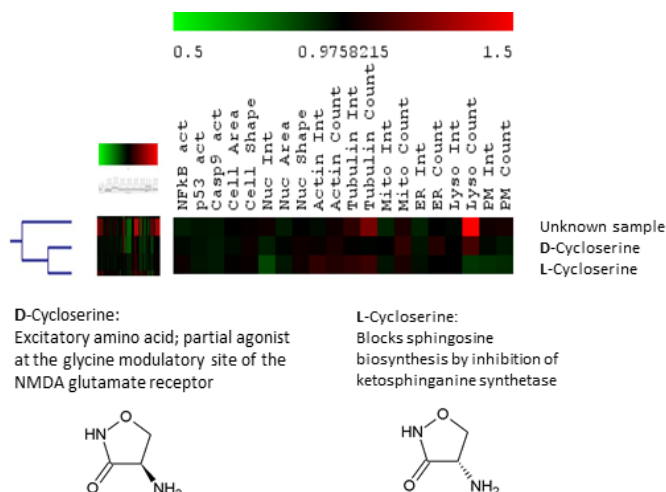
Figure 8: Example of a structural similarity cluster.

When no common mode of action can be identified, these clusters are difficult to biologically analyze. In this case further investigation on common side effects need to be performed or more standards need to be added to get a hint about the common mode of action. This is necessary because a second mode of action test for verification of the predicted target needs to be performed.

III.   Structure / activity relationship (SAR)

These clusters consist of a standard set that share a common mode of action and a structural subunit at the same time (see Figure 9). These SAR reveal a subunit of the clustering standards highly probably responsible for the shown mode of action.
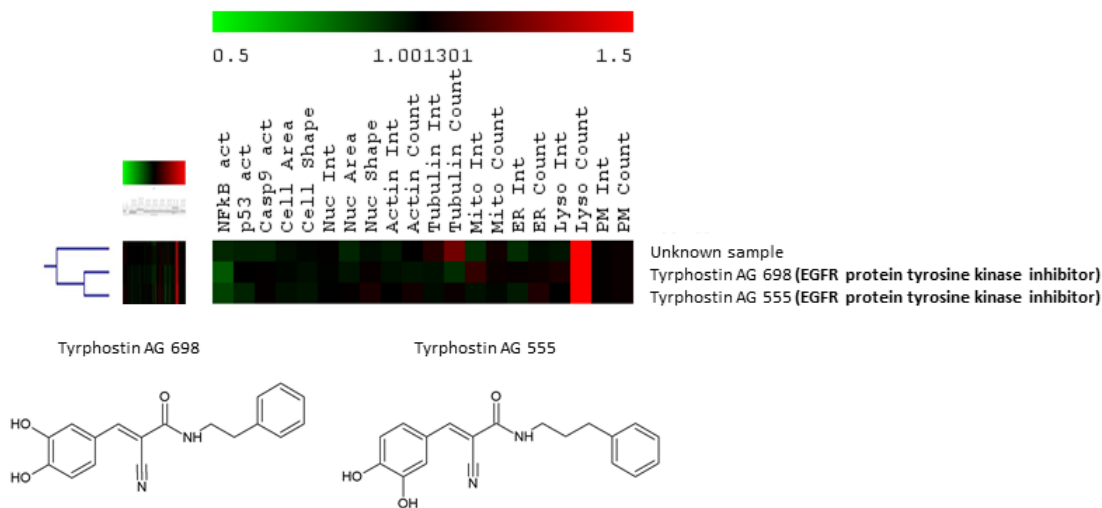
Figure 9: Example of a structure / activity relationship (SAR) cluster.

1. Introduction

These clusters are most interesting in drug discovery as the activity, stability and absorption, distribution, metabolism and excretion (ADME) parameters of the identified compound can be improved by modulating the compound on the inactive part (Bender & Glen, 2004).

HCPP is a predicting tool for biological targets, and any of the shown cluster i) – iii) needs to be verified using a targeted assay. The availability of these tests is another criterion for the effective cluster selection besides the already described necessity of robust clustering and availability of support trees. This additional information helps to select possible lead cluster in a HCPP dataset (Kremb & Voolstra, 2017). The analysis of these described clusters might lead to new bioactive scaffolds, which are important for further drug development (Hajjar et al., 2017; Kremb, Müller, Schmitt-Kopplin, & Voolstra, 2017).

1.3.2. HIV

As HCPP is only a prediction tool every shown cluster need to be verified using a targeted biological assay. HCPP cluster that involves for example one or several anti-HIV agents needs to be verified using an anti-HIV activity test. The used test in this study and the general necessity of anti-HIV assays is shortly explained in the following chapters.

1.3.2.1. HIV, global situation, the virus and the replication cycle

I. The virus

HIV is an enveloped virus from the family of retrovirus and the species of the lentivirus and can be differentiated in human immunodeficiency virus typ-1 and typ-2 (HIV-1 and HIV-2) (Clavel et al., 1986). HIV-2 is causing globally less than 1% of all infections and is almost exclusively occurring in west Africa. Therefore, the comprehensive research and this thesis is focused on HIV-1. With 100 to 120 nm the virus is relatively huge. The virus is covered with a lipoprotein envelope and in this envelope 72 10 nm huge envelop glycoprotein complexes are embedded. Every glycoprotein complex consists of an external envelope glycoprotein 120 (gp120) and one transmembrane glycoprotein 41 (gp41). The bond between gp120, gp41 and the envelop is weak and gp120 can therefore be easily released.

Gp120 can be detected in the serum of HIV patients for diagnostic analysis. In general, every retrovirus and hence HIV-1 necessarily needs three gens namely a group-antigen (gag), polymerase (pol) and envelope (env). The resulting HIV-1 retroviral genome sequence is shown in Figure 10.
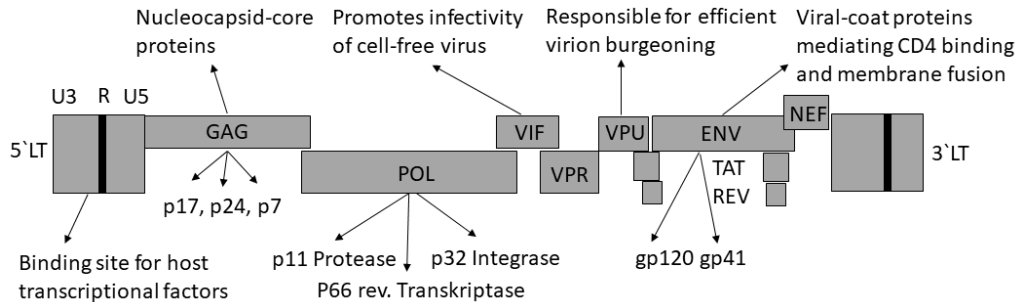
Figure 10: Overview of HIV ribonucleic acid (RNA) and its genes.

The long terminal repeat (LT) regions are parts of the viral genome and are connected in the integration phase to the host DNA. Because of this stable integration the HIV-1 infection is permanent. The classical gene set-up is 5' LT-gag-pol-env-3' LT. A cut out of this specific viral DNA would lead to a curing from HIV and research in that area is ongoing (Hauber et al., 2013). The gene gag (group-antigen) is coding proteins of the virus envelop like p24 - the major capsid protein, p17 - a matrix protein and p7- a nucleocapsid-proteins which stabilizes the viral RNA. Pol (polymerase) is coding the viral enzymes for p11 - protease, p66 - reverse transcriptase (RT) and p32 - integrase. Env (envelop) is coding the glycoproteins of the envelop gp120 and gp41. There are 6 additional accessory genes VIF, VPU, VPR, TAT, REV and NEF in the 9 kilobase (kb) RNA responsible for different regulatory processes. VIF promotes the infectivity of the cell free virus and VPU is responsible for the viral burgeoning. Trans-Activator of Transcription (TAT) and Regulator of Expression of Virion proteins (REV) are coding regulatory proteins that accumulate in the cell nucleus and bind to the RNA on specific parts. Both proteins stimulate the transcription from HIV-DNA to RNA, support the transport of HIV RNA from the nucleus to the cytoplasm and are essential for the translation. NEF for example induces the down regulation of cluster of Differentiation 4-Rezeptor (CD4) and human leukocyte antigen class-1 antigens (Collins, Chen, Kalams, Walker, & Baltimore, 1998) on the surface of the infected cells. Due to that, the infected cell was less often attacked of the cytotoxic T-cells.

II.   The replication cycle

For reproduction, the virus needs host cells with a CD4 receptor and C-C chemokine receptor type 5 (CCR5) or C-X-C chemokine receptor type 4 (CXCR4) proteins as co-receptors on its surface (Dalgleish et al., 1984). These are mainly CD4+ T helper cells, which protect the body from infections. HIV-1 is gradually destroying them, but also memory T cells. This memory T cell, once infected, are the reason why HIV is recrudesced as soon as medication is stopped. Because they serve as a latent infected dormant HIV-1 reservoir (Buzon et al., 2014). The virus uses the CD4 cells for multiplying itself and spreading into the body. This process is called HIV live cycle and can be divided into seven steps. An overview of the different steps is shown in Figure 11.

1. Introduction
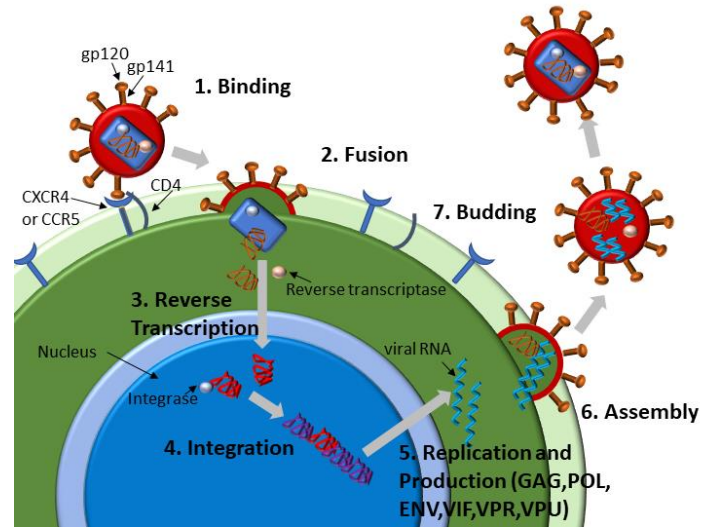
Figure 11: The HIV replication cycle (adapted from https://aidsinfo.nih.gov/understanding-hiv-aids).

1) Binding: HIV binds to the CD4 receptors and CCR5 or CXCR4 coreceptors on the surface of a CD4 + T - cell.

2) Fusion: The envelop of HIV and the cell membrane fuse catalyzed through gp141 rearrangement and the HIV RNA is released into the host cell.

3) Reverse transcription: Protease, Integrase, RNA and RT are released. RT can now convert single-stranded HIV RNA to double-stranded HIV DNA resulting in the transport of the viral DNA into the nucleus.

4) Integration: Integrase is released and insert HIV DNA into the host DNA. The integrated HIV DNA is called provirus and can remain inactive for several years.

5) Replication: The host cell polymerase is used to create RNA and mRNA. When the virus is active the virus mRNA gets produced and long HIV protein chains are produced by the cellular machinery.

6) Assembly: HIV protease cuts the long protein chains into individual proteins which form together with RNA the non-infective HIV at the surface of the host cell.

7) Budding: The immature HIV releases itself of the host cell and take parts of the cell outer envelope equipped with glycoproteins along. These proteins are necessary to bind to the next host cell.

III.   HIV transmission

Besides the CD4 count the viral load is a marker for HIV infection and is measured in virus concentration per ml e.g. blood, means viral RNA per ml blood (Ghani et al., 2001; Phillips & Pezzotti, 2004). The highest virus load is in certain body fluids like blood, semen, rectal fluids, vaginal fluids, and breast milk. For transmission these, fluids must get in contact to fresh wounds, mucous membranes (rectum, vagina,

penis, and mouth), injured sport of the outer skin (Bartholomew & Jones, 2006; Dezzutti & Hladik, 2013; Henderson et al., 1990) or get directly injected into the bloodstream.

Common infection ways are therefore unprotected sexual contact to an HIV infected person, infected blood infusions or syringes and the infection from a mother to her child prior to birth or via breastfeeding. Nevertheless, the Centres for Disease Control and Prevention (CDC) summarized after a large aggregate statistic, that the normal life with an infected person (shared bathroom or drinking from the same glass) does not cause an infection (https://www.cdc.gov/hiv/basics/transmission.html., 20.11.2018).

1.3.2.2.   Known anti-HIV drugs and screening systems

1987 the first anti-HIV-1 drug named Zidovudine, a Nucleoside Reverse Transcriptase Inhibitors (NRTI) was released. Since 1990 (Fischl et al., 1990; Volberding et al., 1990) the effort was intensified to identify potent drugs against HIV. In 1995/1996 Saquinavir, Indinavir and Ritonavir were released which inhibit the HIV protease.

Another drug class was introduced at 1996 with Nevirapin a non-nucleoside Reverse-Transcriptase-Inhibitor (NNRTI). In 1996 a highly active antiretroviral therapy (HAART) also called "AIDS cocktail" was established. A combination of three HIV medicines from 2 or 3 different drug classes were combined in this therapy. In 2003 Enfuvirtide was the first compound inhibiting the virus entry and finally in 2007 Raltegravir was released as the first integrase inhibitor on the market. Five different drug classes are known up to now (May 2018) and 30 single compounds are on the market (https://aidsinfo.nih.gov/understanding-hiv-aids/fact-sheets/21/58/fda-approved-hiv-medicines, 20.11.2018) attacking the following targets:

1. **Nucleoside Reverse Transcriptase Inhibitors (NRTIs)**
   Zidovudine (1987), Didanosine (1991), Zalcitabine (1992), Stavudine (1994), Lamivudine (1995), Abacavir (1998), Didanosine EC (2000), Tenofovir DF (2001), Emtricitabine (2003)
2. **Non-Nucleoside Reverse Transcriptase Inhibitor (NNRTIs)**
   Nevirapine (1996), Delavirdine (1997), Efavirenz (1998), Etravirine (2008), Nevirapine XR (2011), Rilpivirine (2011)
3. **Protease Inhibitors (PIs)**
   Saquinavir (1995), Indinavir (1996), Ritonavir (1996), Nelfinavir (1997), Amprenavir (1999), Atazanavir (2003), Fosamprenavir (2003), Tripanavir (2005), Darunavir (2006)
4. **Entry inhibitors: Coreceptor antagonists (CA), Post-attachment inhibitor (PAI) und Fusion inhibitor (FI)**
   FI: Enfuvirtide (2003)
   CA: Maraviroc (2007)
   PAI: Ibalizumab (2018)
5. **Integrase inhibitor**
   Raltegravir (2007), Dolutegravir (2013), Elvitegravir (2014)

In HAART. The divers combinations of these compounds are possible, and the virus can be inhibited in total at five different stages of its replication cycle. Namely at two stages of the HIV entry in the host cell

1. Introduction

and at the three HIV enzymes Reverse Transcriptase (nucleoside and non-nucleoside), Integrase und Protease. When infected with HIV, the HAAR therapy needs to be pursued for the whole life as otherwise the inhibited virus production is increasing again (Blankson, Persaud, & Siliciano, 2002). Moreover, even when using combined medication like HAART, the appearance of multi resistant HIV is just slowed down but not vanished (Ammaranond & Sanguansittianan, 2012). There are estimations that one year of life due to antiretroviral treatment costs between 13.000 and 23.000 dollar per patient (Freedberg et al., 2001; Gebo et al., 2010). These extra costs can be an enormous financial burden even when health insurance funds are available and help to cover the costs. Comparing the available HIV drugs, their targets and the replication cycle of HIV, it is obvious, that there are still steps in the virus replication cycle with no drug available on the market. Therefore, it is crucial that novel drugs for these targets are identified, produced and used to further improve the life expectancy of affected patients. Moreover, novel drugs against known targets are needed to lower the risk of multi resistant viruses and thereby lower the costs of the treatments.

HIV possesses the ability towards adaptive mutations which was revealed by common resistances against at least one of the 30 available drugs in most infected humans (Taiwo, Hicks, & Eron, 2010). HIV infections cannot be cured and although reaching a normal age with the available medication is possible, strong adverse side effects like liver and renal damages appear in the patients (Dieffenbach & Fauci, 2011). Therefore, new drugs with higher efficacy and lower toxicity are needed. To find those quickly and easy, straight forward screening systems are needed. In general, there are two drug testing systems on the market. The first is using purified HIV enzymes to investigate the effect of compound or compound mixture on the HIV enzymes. Critical parameters like ability to enter the cell, stability or typical side effects were not monitored in these screenings (Prokofjeva et al., 2013). The second type of screening system is using virus isolates in cell systems. Even more laboratory safety regulations are necessary using this infectious HIV, essential information about the further applicability of the tested compounds or compound mixture is gained in the first screening. For example, the EASY-HIT system based on LC5-RIC cells (Kremb et al., 2011). The optical measurement of the HIV infections is ensured through the reporter-gene for a red fluorescent protein (DsRed1) integrated in the cell system. This gene produces a fluorescence protein in the presence of HIV- TAT and REV due to an integrated promotor. Because of that, the infection can easily be quantified. Additionally, the test can distinguish between the early and late phase of the virus inhibition in the replication cycle using a time of addition assay (Daelemans, Pauwels, De Clercq, & Pannecouque, 2011). This test already leads to the discovery of new anti-HIV active scaffolds (Helfer et al., 2014; Kremb et al., 2011; O'Rourke et al., 2016; S. Rebensburg et al., 2016).

1.3.2.3. Natural products against HIV

There is an urgent need for the discovery of new, safe, efficient and cheap anti-HIV drugs as no healing is currently available and due to the gained resistances of the virus against the available drugs. On the one side it is known, that theoretical drug modeling and drug synthesis in the laboratory towards targets in the HIV replication cycle are limited in their ability to generate new drug scaffolds (David J. Newman & Cragg, 2016). On the other side, drugs against almost all disease areas were first found in the plant

kingdom like one of the most famous drugs aspirin (D. J. Newman et al., 2000) and morphium (Grabley et al., 2000). Besides that, plant derived anti-HIV-1 drugs or lead compounds were also identified like calanolide coumarins (Kashman et al., 1992) and betulinic acid (Kashiwada et al., 1996).

Several metabolic classes from plant extracts were found to be anti-HIV-1 active including Alkaloids, Coumarins, Phenolics, Flavonoids, Lignans, Quinones, Saponins, Terpenes/sterols and Xanthones. To name a few alkaloids: Michellamine B from the leafs of *Ancistrocladus korupensis* (family Ancistrocladaceae) inhibits RT and cellular fusion (Manfredi et al., 1991), castanospermine isolated from *Castanospermum australe* (family Fabaceae) has glycosidase inhibitory activity (Karpas et al., 1988) and Nitidine from the roots of *Toddalia asiatica* (family Rutaceae) inhibits the RT (Tan, Pezzuto, Kinghorn, & Hughes, 1991). Additionally coumarins like (+)-Calanolide A, (–)-Calanolide B and its dihydro-derivative, (–)-7,8-dihydrocalanolide B extracted from the fruits and twigs of *C. lanigerum* inhibit the cytopathic effects of HIV-1 in T-cell lines (Xu et al., 2000).

Another huge family of anti-HIV active drugs are based on Phenolics: Corilagin and 1,3,4,6-tetra-O-galloyl-b-D-glucopyranose isolated from *Chamaesyce hyssopifolia* (family Euphorbiaceae) inhibited HIV RT (Yasmina et al., 1997), Repandusinic acid from of *Phyllanthus niruri* (family Euphorbiaceae) inhibited HIV RT (Ogata et al., 1992) and Macrocarpal B isolated from *Eucalyptus globulus* (family Myrtaceae) inhibit also HIV RT (Nishizawa et al., 1992). In this thesis two different plants were used *Ocimum tenuiflorum* and *Cistus incanus* to find new scaffolds, they are described in the following chapters.

1.4. bioactive plants and compounds

1.4.1. Bioactive plants

1.4.1.1. Ocimum tenuiflorum

Ocimum tenuiflorum (Tulsi), also known as *O. sanctum L*, from the family of Lamiaceae is a herbaceous plant with fine haired, 1-3 cm long opposite leafs with light serrated leaf margin (Chowdhury, Mandal, Roy, & De Sarker, 2017). The white to reddish blossoms are arranged in spike-like inflorescences were six blossoms form false whorls. The plant originate from north central India (Saharkhiz et al., 2015) and grows also in the eastern world tropics (Bast, Rani, & Meena, 2014). In Ayurvedic medicine, Tulsi is used due to its various bioactive effects as the "Medicine of Nature", "holy basil" and "The Queen of Herbs" since thousands of years (Das, Raman, Saha, & Singh, 2015; Mondal, Mirdha, & Mahapatra, 2009; Pattanayak, Behera, Das, & Panda, 2010). It has been extensively used in Ayurveda, the world's oldest medical system, among others against colds, malaria, stomach disorders and inflammation (Board, 2005; Leung, 1996; Singh, Majumdar, & Yadav, 1996) but also against arthritis, rheumatism, pain and fever (Godhwani, Godhwani, & Vyas, 1987). Extract of Tulsi leafs showed anti-inflammatory (Chattopadhyay, 1994; Godhwani et al., 1987), anticarcinogenic (Banerjee, Prashar, Kumar, & Rao, 1996), antiulcerogenic (Mandal et al., 1993) anti-diabetic, antioxidant, antimicrobial and immunomodulatory activity (Pattanayak et al., 2010).

1. Introduction

The main metabolic classes of Tulsi extract are 20% phenolics and 7.5% flavonoids. Some bioactive metabolites of the plant are characterized and identified like eugenol (Choudhury et al., 2014), flavonoids like vicenin, galuteolin and cirsilineol (H. Nörr, 1992), phenolic acids like rosmarinic acid and caffeic acid, tannins (Grayer et al., 1996), phenylpropane glycosides, (H. Nörr, 1992) and polyphenols (H. Nguyen, Lemberkovics, É., Tarr, K., Máthé Jr., I. and Petri, G., 1993; Shubha et al., 2017). Latest research show, that Tulsi show activity against pathogens responsible for human infection (Vasudevan P, 1999) and additionally anti-HIV-1 activity (Rege, Ambaye, & Deshmukh, 2010; Usha, Naidu, & Raju, 2003).

1.4.1.2. *Cistus incanus*

Cistus incanus (Cistus) from the family of Cistaceae is a half shrub with leathery, evergreen and often opposite leafs. (Barrajon-Catalan et al., 2011; Witt, 1964). The blossoms, consisting of five crinkled rose petals, have a diameter of 4-6 cm and are individually or collectively arranged (Cheers, 2000). Cistus is a thermophilic plant that requires much light and it therefore grows preferential in calcareous soil regions areas around the Mediterranean area (Gori et al., 2016).

Extracts of Cistus were used since the 4th century B.C. as anti-inflammatory, anti-allergic, antimicrobial and antifungal medicine (Barrajon-Catalan et al., 2011; Kalus, Kiesewetter, & Radtke, 2010; Richard, 2010). A short-cooked water extract of the plant was used against inflammatory skin disease, a long-cooked water extract against diarrhoea and a mash of the plant was used prophylactic against infections in open wounds (Petereit, Kolodziej, & Nahrstedt, 1991). The main active substance in Cistus extracts are phenolic compounds from the group of the flavonoids (Pomponio, Gotti, Santagati, & Cavrini, 2003). Several metabolites like flavanol glycosides, different flavanols, phenolic acids, tannins and ellagitannin are partially described in literature (Barrajon-Catalan et al., 2011; Danne, Petereit, & Nahrstedt, 1993; Droebner, Ehrhardt, Poetter, Ludwig, & Planz, 2007; Petereit et al., 1991; Santagati, Salerno, Attaguile, Savoca, & Ronsisvalle, 2008). In cell culture and animal experiments antiviral effects against influenza viruses like bird flu (H7N7 and H5N1), swine flu (H1N1v), human rhinovirus (HRV14) (Droebner et al., 2007; Ehrhardt et al., 2007; Sakagami, 1995) and HIV (S. Rebensburg et al., 2016) was demonstrated. In this thesis *Cistus incanus* was used in form of the freely available and standardized Cystus052®water extract (Dr. Pandalis Urheimische Medizin).

1.4.2. Bioactive compounds in plats

Besides compounds for the primary plant metabolism, secondary plant metabolites were produced by the plant for e.g. defense or attraction issues. These secondary plant metabolites consist of e.g. carotinoids, saponin, phytosterols, terpens and polyphenols (Watzl, 2005) to only name a few. Some of them, like polyphenols, influence living organism and are therefore called bioactive compounds. Polyphenols are substances which are based on the substructure of phenol. Seven groups of phenols can be distinguished like i) simple phenols, ii) hydroxybenzoic acid, iii) hydroxycinnamic acid, iv) lignan, v) lignin, vi) coumarin and vii) flavonoids (Watzl, 2005) (Figure 12).
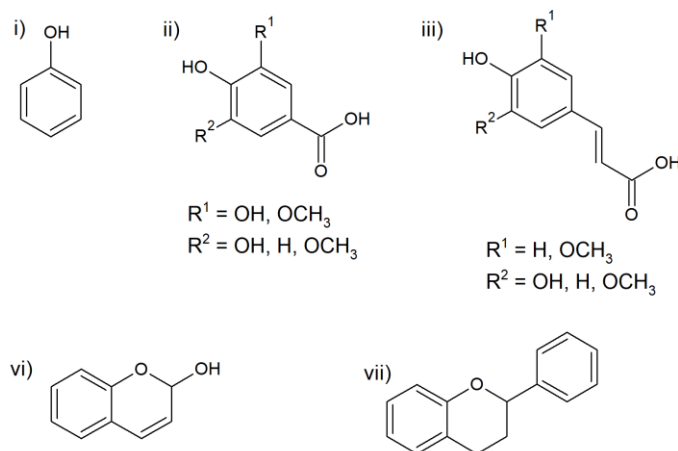
Figure 12: Molecular structure of i) phenol ii) hydroxybenzoic acid, iii) hydroxycinnamic acid, vi) coumarin and vii) flavonoids.

Polyphenols are the main supplier of antioxidants for human beings and therefore these compounds are preventive against cancer, cardiovascular and other age-related disease (Scalbert, Manach, Morand, Remesy, & Jimenez, 2005). In general polyphenols show an anti-cancer, antimicrobial, anti-inflammatory and anti-viral effects (Helfer et al., 2014; Pan, Lai, & Ho, 2010; Zuo, Wang, & Zhan, 2002).

The anti-viral activity of polyphenols could be based on their ability to attach themselves directly to the virus, and thereby block the viral attachment to the cell (De Bruyne et al., 1999). Because of that, the mechanism of polyphenols might be unspecific (Sakagami, 1995). Moreover, subclasses of polyphenols e.g. glycosidic flavoniods from *Crataegus sinaica* exhibit an anti-Herpes simplex virus Typ 1 (HSV-1) activity (Shahat et al., 2002). In addition, (-)-Epigalloctechin-gallat and Theaflavin-di-gallat inhibit influenza virus-A as well as influenza virus-B (Nakayama et al., 1993). The polyphenol rich Cistus052®-extract shows antiviral activity against different influenza viruses (Droebner et al., 2007; Ehrhardt et al., 2007) and HIV-1 (S. Rebensburg et al., 2016). Besides that, a polyphenolic extract of *Pelargonium sidoides* showed anti-HIV-1 activity (Helfer et al., 2014) like other polyphenolic compounds (Andrae-Marobela, Ghislain, Okatch, & Majinda, 2013). The group of polyphenols is therefore an interesting source for compounds with different bioactivies and in particular with anti-HIV activity.

The non-targeted high content phenotypic profiling (HCPP) has the capability to reveal unknown bioactivity potential of plant extracts or plant extract fractions and can underpin known bioactivity by disclosing the underlying mode of action or possible side effects. Correlating these biological results to ultrahigh resolution mass spectrometry data of the analysed plant sample, the metabolic class or single metabolite probable responsible for the shown bioactivity can be characterised and identified. Using this setup, new bioactive drug scaffolds can be found, urgently necessary for further drug development.

1.5.   Structure of the thesis and motivation

The need of new bioactive drugs against e.g. HIV is tremendous, as still millions of people suffer from this disease and no healing was identified until now. Although several drugs are available, they are expensive, cause side effects and the virus is building resistances against these drugs. Therefore, there is an urgent need for new scaffolds of bioactive drugs, with less side effects and the ability to inhibit up to now untouched steps of the virus replication cycle. Plant extracts are often used as a source for new drugs, as plants offer an outstanding variety of different metabolites that overtop the capability in a synthetic lab by far. In this thesis we are investigating a novel way of drug identification in plant extracts by combining the classical LC compound separation with ultrahigh resolution MS metabolite identification. Combined with the clustering of high content phenotypic profiling data (global cellular readouts) to specific standards and an anti-HIV-1 test, a comprehensive metabolite identification and characterization took place. An overview of the thesis structure is shown in Figure 13.
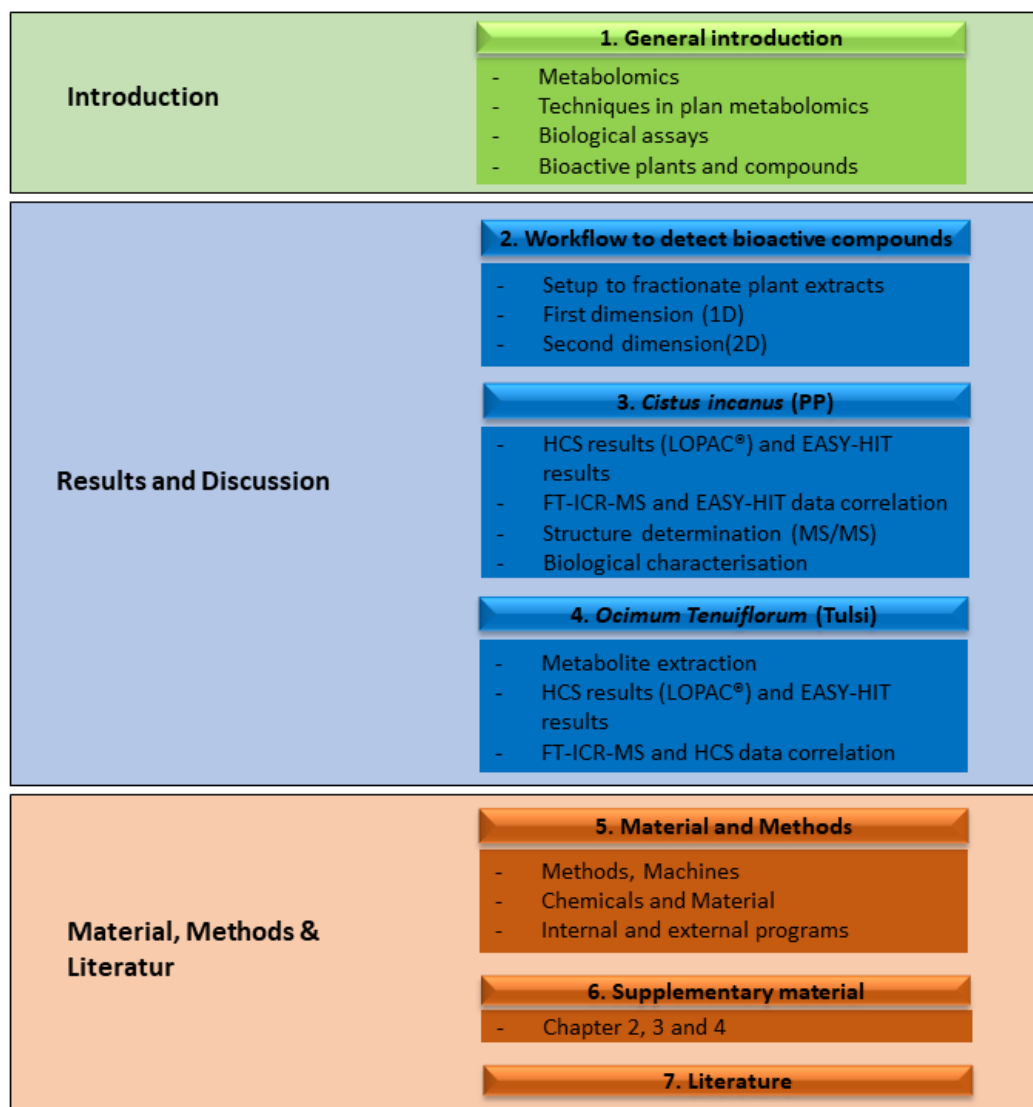
Figure 13: Structure of the thesis.

The general introduction (chapter 1) gave insights about metabolomics, the different techniques used in plant metabolomics, biological assays, bioactive plants and compounds. The background of the used tools is explained and the current research status in the different research fields are summarized. In chapter 2 the used workflow for the detection of bioactive compounds is described highlighting the fractionation process using 2D fractionation. The described workflow was used to characterize bioactive compounds in *Cistus incanus* (chapter 3). Here the produced fractions were investigated for their bioactive capability using HCPP and the EASY-HIT system. The gained FT-ICR-MS data were correlated to the bioactivity tests and the top correlating metabolite was further characterized. In chapter 4 the workflow (chapter 2) was applied to *Ocimum tenuiflorum* (Tulsi). After extracting the metabolites of interest, a high content phenotypic profiling was performed, and the most promising fractions were investigated using FT-ICR-MS. All used material and methods are summed up in chapter 5 and additional information for each chapter is summarized in the supplementary information (chapter 6).

2. *Workflow to detect bioactive compounds*

2.1. Introduction

Many commercial available drugs were initially found in plants e.g. aspirin (D. J. Newman et al., 2000), and morphine (Grabley et al., 2000) but also HIV-1 drugs or lead compounds like calanolide coumarins (Kashman et al., 1992) and betulinic acid (Kashiwada et al., 1996). That's not surprising, since plants are a rich source for the discovery of novel bioactive compounds, as they produce more than 0.2 million different metabolites (Dixon & Strack, 2003).

In this thesis the bioactive plants *Cistus incanus* and *Ocimum tenuiflorum* were used to develop a workflow for the selective detection, isolation and characterization of bioactive compounds. These plants were chosen as they are well described in literature (Barrajon-Catalan et al., 2011; Das et al., 2015) and already used against various diseases e.g. colds, malaria, stomach disorders, inflammation, cancer, diabetes and viral infections (Droebner et al., 2007; Rege et al., 2010; Singh et al., 1996). Their bioactive secondary plant metabolites consist of a variety of different metabolite classes e.g. carotenoids, saponin, phytosterols, terpenes and polyphenol to name only a few (Watzl, 2005). Polyphenols are known to be a very bioactive metabolite subgroup (H. Nguyen, Lemberkovics, É., Tarr, K., Máthé Jr., I. and Petri, G., 1993; Scalbert et al., 2005), in particular they are known to be anti-viral (Ehrhardt et al., 2007; S. Rebensburg et al., 2016; Rege et al., 2010). The huge diversity of plant metabolites and thereby complexity of their extracts (Saito & Matsuda, 2010; Sumner, 2010) causes difficulties in their separation for later analytical or biological analysis. Different analytical techniques are available ranging from different separation procedures GC, LC and capillary electrophorese (CE) coupled to different detection methods like UV absorption, fluorescence, FID, MS and NMR (Kusano et al., 2011). To cover as many plant metabolic classes as possible in a non-targeted metabolic approach the analytical workflow needs to be carefully designed containing suitable fractionation and detection methods. After harvesting the plant and extraction of the metabolites, liquid chromatography is widely used for chromatographic separation, as the sample can be further fractionated, modified or concentrated afterwards if necessary.

Due to the high complexity of the plant metabolites a two-dimensional liquid chromatography is necessary in non-targeted experiments to ensure a wide-ranging metabolite separation (Schoenmakers, 2003). In plant metabolomics a conventional C18 RP column is widely used as separation material (Broeckling et al., 2005) together with a complementary selectivity for the other dimension (1D or 2D) ensuring good separation (Farag et al., 2007). After the fractionation process, the fractions need to be investigated for their metabolic content using ultrahigh resolution MS in non-targeted metabolomics (de Villiers et al., 2016; Pang et al., 2016), as UV and PDA alone are limited in their informative value. MS and PDA can visualize the metabolite distribution over the produced fractions, but a biological assay is necessary to track the bioactivity of each fraction. The sample preparation and separation technique need to be adapted to the metabolite distribution (UV and MS detection) and measured bioactivity of the resulting fractions.

2.2. Objectives

To investigate bioactive plant metabolites a two-dimensional fractionation is recommended due to the high complexity of the plant metabolites (Aturki et al., 2004; Farag et al., 2007; Huhman & Sumner, 2002; Kind et al., 2009). This chapter addresses the research question, weather two-dimensional separation using LC times LC is sufficient to separate the complex mixture of plant metabolites for further biological and analytical analysis. To gain insight into the bioactive metabolite distribution, different SPE materials were investigated. After the 2D HPLC separation, FT-ICR-MS data of the produced fractions were generated to get a deep insight into the metabolic content of the fractions.

Based on the FT-ICR-MS data, the total ion current (TIC) of the produced fractions was simulated to track the metabolite distribution and further was correlated to the gained PDA data. To investigate the bioactivity distribution, an anti-HIV-1 test (EASY-HIT) was exemplary performed to quantify the bioactivity of every fraction. To obtain knowledge about the contribution of polyphenols to the shown bioactivity, an additional polyphenol enrichment (Helfer et al., 2014) of the extracts was performed and characterized.

2.3. Results and Discussion

2.3.1. Setup to fractionate plant extracts

The metabolome of a plant extract is usually very complex, as it includes diverse types of compounds with different polarities and the separation of these metabolites is still a big challenge (Sasidharan, Chen, Saravanan, Sundram, & Yoga Latha, 2011). Frequently occurring plant compounds, for example the group of phenols including flavonoids, flavonoid glycosides, phenolic acids and cumarines, have additionally similar physicochemical properties and are therefore difficult or often impossible to separate using common 1D LC analysis. Moreover, this group of polyphenols and flavonoids are polar and less ionized. Therefore, a comprehensive two-dimensional RP/RP separation with different selectivity was chosen. (Miroslaw, Renata, Monika, Ryszard, & Monika, 2012) (Jandera, 2006). Separation with two different selectivities additionally improve the separation and peak capacity (Jiang et al., 2014; Willemse, Stander, Tredoux, & de Villiers, 2014), reduce matrix effects, differentiate isomers (Lei, Huhman, & Sumner, 2011) and is necessary to overcome ion-suppression effects in the following FT-ICR-MS measurements.

Due to the study design, different framework conditions need to be considered. The produced fractions need to be divided into aliquots for further experiments on the FT-ICR-MS and different bioactivity tests. To ensure enough material for the experiments, an SPE fractionation was chosen for the first-dimension and an semi preparative HPLC-PDA separation for the second-dimension. In Figure 14 the different dimensions are illustrated.

2. Workflow to detect bioactive compounds
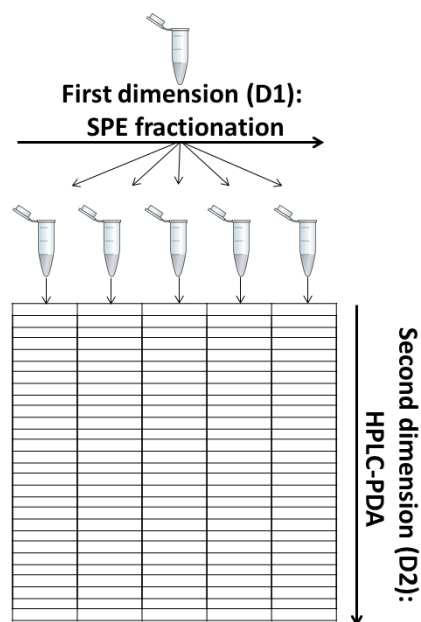
Figure 14: Schematic overview of the fractionation procedure. First-dimension (D1) = SPE-fractionation on the x-axis and second-dimension (D2) = HPLC-PDA fractionation on the y-axis.

For the first-dimension (D1), a SPE-fractionation (x-axis) and for the second-dimension (D2) a HPLC-PDA fractionation (y-axis) were performed. Every produced fraction was aliquoted and additionally tested for its general bioactivity and for its anti-HIV-1 activity. Additionally, FT-ICR-MS measurements were done for every fraction. The results are discussed in the following chapters.

2.3.2. Anti-HIV-1 activity of Cistus, Cistus PP and Tulsi fractions

To investigate the distribution of the active compounds in the fractionation process, it is necessary to have a strong bioactivity test. For this study we choose the EASY-HIT test (Kremb et al., 2011) and started the fractionation process with potent anti-HIV-1 active raw extracts. For this purpose, special extracts of Cistus, Cistus polyphenol enriched (PP), Tulsi and Tulsi PP were investigated. The production of the Tulsi extracts is described in 5.1.2.1.2 and the origin of Cistus in 5.1.2.1.1. The polyphenol enrichment of both extracts is described 5.1.3. To compare their ability for anti-HIV-1 activity, an EASY-HIT test was performed (5.1.1.1 until 5.1.1.3) and the half maximal effective concentration (EC_{50}) of the different extracts was calculated and compared in Table 1.

Table 1: EC_{50} and standard deviation (SD) comparison of Tulsi (PP) and Cistus (PP), n = 3. * values from (Stephanie Rebensburg, 2015), nq = not quantifiable.

	Tulsi			Cistus		
	Raw extract	PP	PP depletion	Raw extract	PP	PP depletion
EC_{50} [µg/ml]	8.66	62.73	nq	7.79*	0.72*	nq
SD [± µg/ml]	0.86	4.58	nq	0.15*	0.02*	nq

30

The comparison revealed, that Cistus PP with an EC_{50} of 0.72 ± 0.02 µg/ml is the strongest anti-HIV-1 active extract followed by Cistus with 7.79 ± 0.15 µg/ml. Investigating the leaf and stem of Tulsi with different extraction methods (5.1.2.1.2), the water extract of the leaf showed a higher anti-HIV-1 potential (EC_{50} of 8.66 ± 0.86 µg/ml) than the polyphenol enriched extract of Tulsi (EC_{50} of 62.73 ± 4.58 µg/ml).

Although the polyphenol enrichment increases the anti-HIV-1 activity in case of Cistus (S. Rebensburg et al., 2016) the same polyphenol enrichment procedure decreases the HIV-1 activity in the Tulsi extract. The difference between Cistus and Tulsi could be grounded on the process itself which inactivate the active metabolites in case of Tulsi. It could also be, that inactive polyphenols were enriched, reducing the overall activity. Tulsi PP which showed the lowest activity among the investigated extracts was therefore excluded from further studies. Tulsi will be further discussed in chapter "4. Tulsi", Cistus and Cistus PP in chapter "3. Cistus and Cistus PP".

2.3.3. First-dimension: SPE fractionation and anti-HIV

2.3.4. HIV-1 activity of Cistus (PP) and Tulsi

As the key criteria for the first fractionation is a high yield and the ability to separati active from non-active metabolites, SPE fractionation was used as a first-dimension. In addition, SPEs with different selectivity were screened to tune the separation of non-active from active metabolites in Tulsi extracts. The extract of Tulsi was used as a model extract to analyze the screening and the SPE fractionation, performed according to the description in 5.1.4. Figure 15 summarizes the tested anti-HIV-1 activity of the model extract and cell toxicity (protocol 5.1.1.1 – 5.1.1.3) after extract fractionation using C18, C8, C2, CNE and PPL-SPE columns and a methanol gradient.
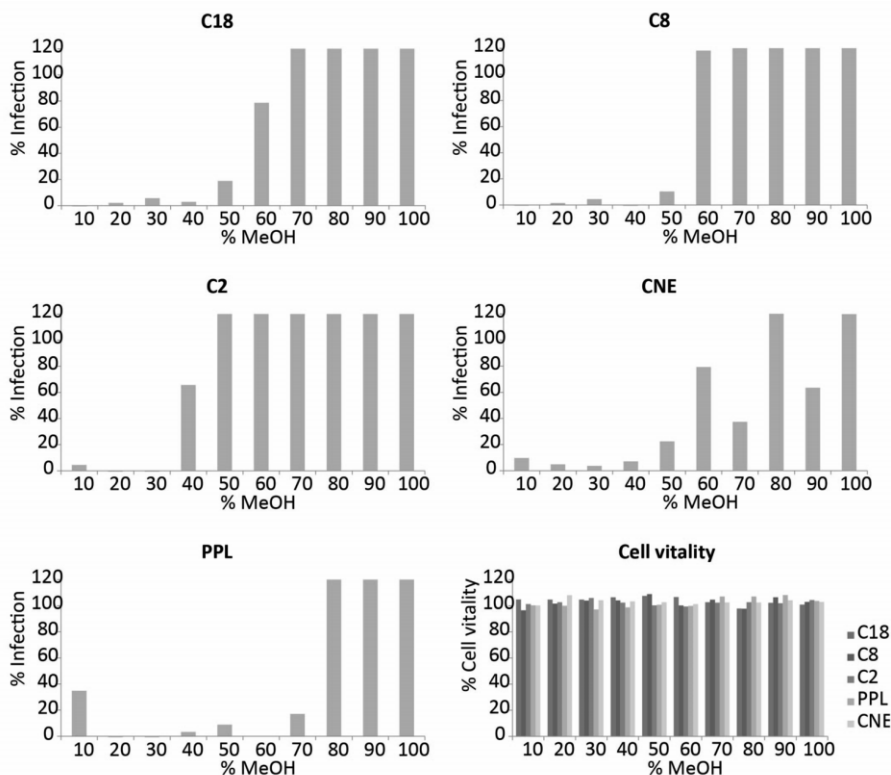
Figure 15: Overview of Tulsi extract fractionation and their effect on cell viability and anti-HIV activity dependent on column material C18, C8, C2, Cyanopropyl endcapped (CNE) and PPL (n=1). Y-axis in [%] infected cells or [%] cell vitality, x-axis in 10% MeOH steps.

The anti-HIV-1 activity distribution was similar in case of C18, C8 and C2 as the highest infection rates were observable in the fractions with lower than 50% MeOH. The CNE-SPE is the only column, with different anti-HIV activity distribution, showing low HIV-1 infection in the fractions from 10% to 40% and two local minima at 70% and 90% MeOH. In contrast, the PPL-SPE spread the activity the most as the first seven fractions show anti-HIV-1 activity under 35%. The used concentrations were not cell toxic as the cell vitality never drops below 90% (Figure 15, right corner).

For the following 2D experiments, it is important to have a clear separation of active and non-active fractions as this is crucial for later data filtering. The active metabolites can only be identified if they can be compared to non-active fractions and therefore non-active metabolites. For this reason, the CNE-SPE was excluded, as it showed besides inhibiting and non-inhibiting to many moderate inhibiting fractions. Another important criterion is, that the D1-fractions must be limited to a minimum of fractions, as they multiply the measuring time, material, costs and measuring capacities for every following experiment. Therefore, the PPL-SPE was excluded due to the increased number of active fractions.

For the following experiments the C2-SPE was chosen, because it provides the highest accumulation of anti-HIV-1 activity in the lowest number of fractions. The increased polarity of the separating material, compared to C8 and C18, leads to a faster elution of the active compounds compared to the C8 or C18.

Moreover, the fractionation was optimized from ten to five fractions using 20% MeOH steps instead of 10% MeOH (Figure 16) for the model extract of Tulsi but also for Cistus and Cistus PP.
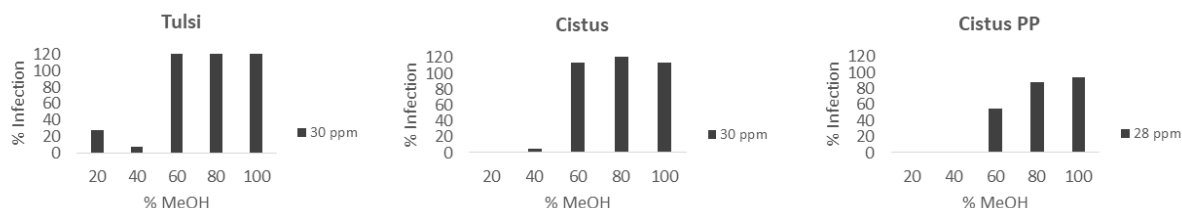
Figure 16: Anti-HIV-1 distribution of the C2-SPE fractions from Tulsi (30 ppm), Cistus (30 ppm) and Cistus PP (28 ppm). The 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromid (MTT) test show no reduction in the cell vitality under 80% (data not shown).

These five fractions (20%, 40%, 60%, 80% and 100% MeOH) clearly separate the active from the non-active fractions for all investigated extracts. Five fractions are a reasonable number for further experiments and the used methods can efficiently elute even strong retaining compounds with the highest percentage of organic solvent. Consequentially, still the whole metabolome from the starting raw extract can be displayed with the resulting fractions.

2.3.5. Second-dimension: HPLC-PDA fractionation and FT-ICR-MS results

The fractionation in the D2 needs to be orthogonal in terms of the column material selectivity compared to the D1, to guarantee effective separation (Jandera, 2006). Therefore, the X-Bridge TM Shield RP C18 was chosen. As polyphenols are of special interest in this study, an additional embedded carbamate group for better polyphenolic separation (as it provides alternate selectivity) was chosen for the RP-C18. A semi preparative column was used, to increase the yield after the fractionation, so that all following experiments (HCPP, EASY-HIT, FT-ICR-MS and FT-ICR-MS/MS) can be performed even in triplicates for each single fraction. To characterize the produced fractions in terms of UV activity, PDA data of the D2 were collected.

The protocol of the high performance liquid chromatography (HPLC)-PDA fractionation is described in 5.2.5 with the chromatographic conditions listed in Table 19 and the used continuously shifted gradient, to ensure good separation and shorter elution time (Table 20) (Jandera, 2012). For quality control a homogenate of the used matrices was prepared and spiked with standards to calculate the intra- and interday precision of the fractionation at different measurements days using the relative standard deviation (RSD%) (See 5.2.5.1. and

Figure 62. The RT RSD% between day precision for the used fractionation setup range from 0.4 - 1.07 RDS% and is therefore under the mandatory limit of 2% (Koop, de Freitas, de Souza, Martinez, & Silveira, 2013; SHABIR, 2004). It is higher compared to the RT RSD% intraday precision because of the different measuring days. The RT RSD% intraday precision for the used fractionation setup range from 0.1 – 0.5 RT RDS% and is still under the mandatory limit of 1% (SHABIR, 2004).

2. Workflow to detect bioactive compounds

To gain deeper knowledge about the metabolite distribution, FT-ICR-MS measurements were performed on the fractions. In Figure 17 the summed-up total ion current (TIC) of each fraction measured on the FT-ICR-MS (conditions see 5.2.1 until 5.2.4) is illustrated. On the z-axis the different SPE fractions (D1) are illustrated from Cistus 20% MeOH until Tulsi 100% MeOH. On the x-axis the corresponding HPLC fractions (2D) are shown and on the y-axis the TIC of every fractions is illustrated.
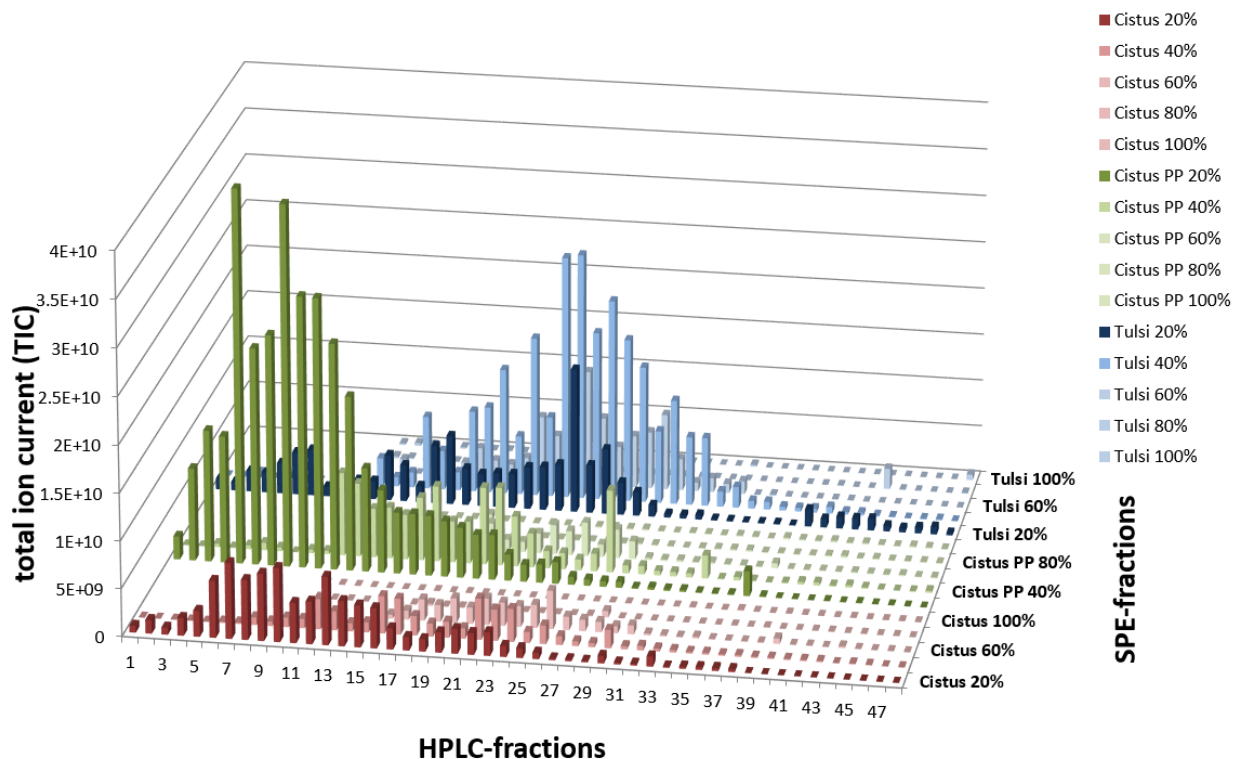
Figure 17: Simulated total ion current (TIC) for all D2 Cistus (red), Cistus PP (green) and Tulsi (blue) fractions. D2 HPLC fractions on the x-axis, the D1 fractionation on the z-axis and the total Ion current (summed up intensity of the measured m/z per fraction) on the y-axis.

Figure 17 shows that most of the ionizable metabolites elute in the D2 20% and 40% MeOH fractions in all plant extracts. A diverse metabolic pattern is shown here. The 60%, 80% and 100% fractions are in comparison to that almost empty. The metabolites are enriched for all fractions of Cistus PP compared to Cistus as expected, after enrichment of the polyphenols (5.1.3). In contrast to Cistus and Cistus PP, the D2 40% MeOH fraction of Tulsi exhibits more metabolites compared to the 20% fractions.

This result does not necessarily mean, that there are no metabolites in the 60%, 80% and 100% fractions of all plants. They might just not be ionizable but could be UV active. The TIC of the D2 fractions were compared to their UV absorbance at the wavelength of 280 nm, as here the most ultraviolet (UV) absorbents was visible (Figure 63). The D2 20% and 40% data are shown in Figure 18, the 60%, 80% and 100% data are summed up in Figure 64 as no UV activity appeared for them.
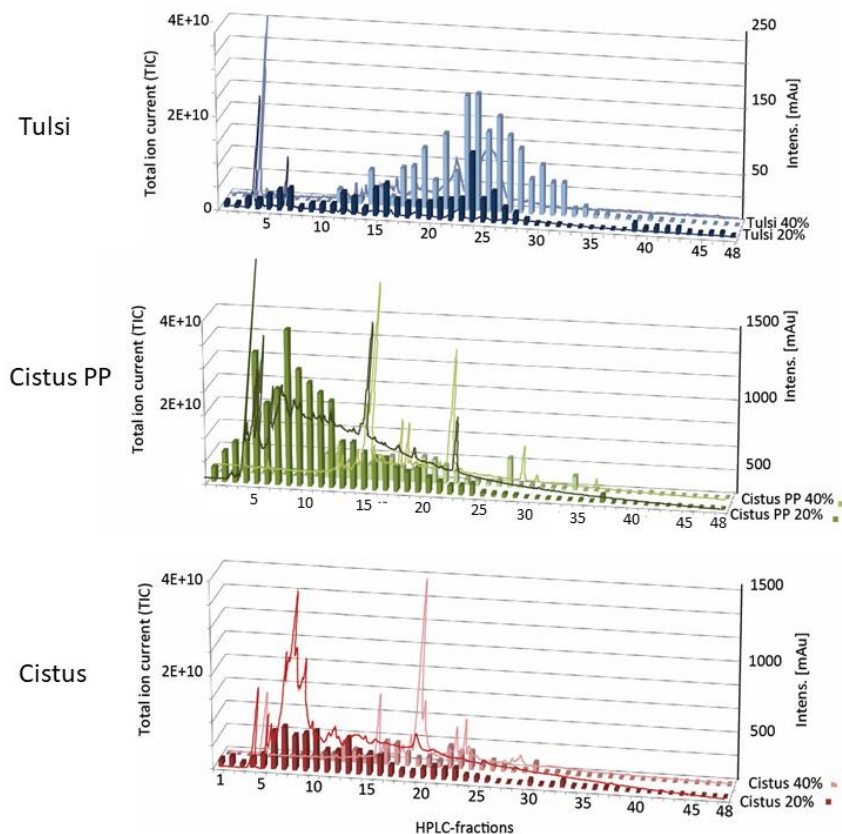
Figure 18: Overlay of the UV absorbents at 280 nm compared to the TIC for Tulsi (blue), Cistus PP (green) and Cistus (red). The D2 20% fractions are always displayed in a stronger color than the D2 40% fractions (e.g. light blue and dark blue). On the left y-axis the TIC is displayed and on the right y-axis the UV intensity in milli-absorbance-units [mAu] is shown and the D2 HPLC fractions are represented on the x-axis.

Here we can see, that for Tulsi D2 20% a very small UV absorbent took place, compared to D2 40% were a higher UV absorbance is visible. This is in agreement with the TIC distribution of that two fractions. In contracts to that in fraction 3, a very high UV peak appear even though the TIC is low. For Cistus PP the UV absorbance is following the TIC except of three UV active peaks, which show no corresponding higher TIC in fraction 4, 14-15 and 22-24. For Cistus, the fractions D2 20% 5-10 and D2 40% 18-19 show a higher UV activity compared to their TIC. As a conclusion we can say, that in fraction 20% and 40% the TIC is following almost all the time the UV absorbance of the fractions. Some fractions show additionally very high UV peaks but for the fractions in 60%, 80% and 100% no TIC appeared, and no UV activity was visible.

This result show, that the UV activity is in the case of Tulsi, Cistus and Cistus PP connected to the metabolite appearance in the fractions measured with the FT-ICR-MS. This is assumed because of the hug content of polyphenolic compounds in the samples but does not peremptory needed to be the case (Roškar & Lušin, 2012).

2. Workflow to detect bioactive compounds

2.4.   Summary and conclusion

Tulsi, Cistus and Cistus PP were chosen as model plant extracts to investigate the metabolic separation in a 2D LC-LC fractionation. This 2D separation is necessary for later isolation, characterization and identification of bioactive compounds in this study. To monitor the distribution of the bioactive compounds during the fractionation process, the anti-HIV-1 test EASY-HIT was chosen. The three used extracts were tested for the anti-HIV-1 activity and showed a clear ranking toward bioactivity. For Cistus we could see that the basic activity of 7.79 ± 0.15 µg/ml could be significantly enhanced by enriching the polyphenols (S. Rebensburg et al., 2016). This leads to a 10-fold enhancement of activity for Cistus PP to an EC_{50} of 0.72 ± 0.02 µg/ml.

Compared to that the most bioactive extract of Tulsi (5.1.2.1.2), the water extract of the leaf had an EC_{50} of 8.66 ± 0.86 µg/ml. The enrichment of this extract for polyphenols using the same method as for Cistus leads to a massive drop in bioactivity (EC_{50} of 62.73 ± 4.58 µg/ml). This effect might either be caused by the procedure, which is deactivating the active metabolites and/or the enrichment of the polyphenols in Tulsi extracts enriches inactive polyphenols leading to a reduced activity. Based on that, Cistus PP is in our case the plant extract with the highest anti-HIV-1 activity followed by the original Cistus extract. Nevertheless, it was possible to produce an extract of Tulsi in the same activity range of Cistus and Cistus PP. These three extracts can be used for the following study. For the D1 fractionation Tulsi was chosen as a model extract, to optimize the system on the lowest anti-HIV-1 active extract. To easily upscale the fractionation, SPE was chosen as method for the D1 fractionation.

To reach on one hand a clear active vs. non active separation and on the other hand a low number of fractions keeping a broad metabolic range, several column materials were tested (C18, C8, C2, PPL and CNE) (Figure 15). For the D1 separation the C2-SPE was chosen with eluting steps of 20% MeOH resulting in 5 fractions. These five fractions (20%, 40%, 60%, 80% and 100% MeOH) clearly separate the active from the non-active fractions, are the lowest possible fraction number and efficiently elute even strong retaining compounds with the highest percentage of organic solvent to still display the whole metabolome from the starting raw extract (Figure 16). The D2 need to be orthogonal to the D1 to ensure sufficient metabolite separation (Dugo, Kumm, Crupi, Cotroneo, & Mondello, 2006). The D2 separation of natural extracts containing phenolic compounds, antioxidants and drugs are optimally separated in an RP/RP system (Jandera, 2006), that's why a RP C-18 column with a carbamate group was chosen for the D2 separation (Appleton, Buss, & Butler, 2007; Tu et al., 2010; Wagenaar, 2008).

During the D2 separation PDA data were collected to observe the distribution of UV active compounds. The analysis revealed, that for Cistus, Cistus PP and Tulsi the UV active compounds are in the 20% and 40% MeOH D2 fractions (Figure 16). Furthermore, FT-ICR-MS measurement were performed, to investigate the metabolic distribution in the produced fractions. The comparison of UV spectra and FT-ICR-MS measurements revealed similar pattern in all fractions. The TIC of Cistus PP is significantly enlarged compared to the TIC of Cistus as we would expect it (S. Rebensburg et al., 2016). The fact that Tulsi originates from a total different plant family compared to Cistus, is visible in its shifted metabolic distribution.

In this context we can sum up, that the target molecules for this study are mainly polar as they elute in the D1 C2 20% and 40% fractions. Additionally, they are UV active, as the TIC of the metabolites correlate strong with the UV activity. Nevertheless, this setup of separation and fractionation can also be used for extracts with more nonpolar compounds (Kremb et al., 2017) and further studies could also focus on that.

3. Cistus incanus (PP)

3.1. Introduction

Cistus incanus from the family of Cistaceae has evergreen leafs, crinkled rose petals and grow in the area around the Mediterranean Sea (Barrajon-Catalan et al., 2011; Gori et al., 2016). Since the 4th century B.C. Cistus extracts, especially water extracts (Petereit et al., 1991) were used as anti-inflammatory, anti-allergic, antimicrobial and antifungal medicine (Barrajon-Catalan et al., 2011; Richard, 2010). Additional anti-viral activity could be shown against influenza viruses like bird flu (H7N7 and H5N1), swine flu (H1N1v), human rhinovirus (HRV14) (Droebner et al., 2007; Ehrhardt et al., 2007) and HIV (S. Rebensburg et al., 2016). Polyphenolic compounds are often responsible for the sown bioactivity (Pomponio (Pomponio et al., 2003) and some of them are already described in literature like flavanol glycosides, different flavanols, phenolic acids, tannins and ellagitannins (Danne et al., 1993; Droebner et al., 2007; Petereit et al., 1991; Santagati et al., 2008). The anti-viral activity of polyphenols could be based on their ability to attach themselves directly to the virus, and thereby block the viral attachment to the cell (De Bruyne et al., 1999). This mechanism could be applicable for different anti-viral substances (Sakagami, 1995).

As the raw extract of *Cistus incanus* is known to strongly inhibit HIV, it is a promising source for the discovery of new anti-HIV drugs. The need for more potent anti-HIV drugs is obvious as still (2016) 36.7 million people are infected with HIV (WHO) and the number of AIDS related death is still one million per year (2016). Now there are 30 single compounds available on the market attacking five different stages in the viral replication cycle. Even though the HAART therapy is using a combination of different drugs, the virus developed drug resistances leading to the appearance of multi resistant HIV viruses whereas no drug for further treatment is on the market right now. There are still steps in the replication cycle of the virus where no inhibitory drugs is available at the moment. The HAART therapy would be significantly enriched, and many multi resistant viruses could be inhibited when drugs against new targets would be available. Drugs with a different target, higher efficacy and lower toxicity are therefore urgently needed.

To improve the identification of those drugs, special screening systems are needed. They need to analyze not only the effect on their target (e.g. the virus) but also the ability of the drugs to enter the cell, their stability, toxicity or potential side effects (Prokofjeva et al., 2013). For a deeper investigation of the shown bioactivity and to investigate the mode of action a HCPP screening characterizing the cellular effects of the compounds is recommended. Such a non-targeted biological cell *in vitro* screening approach can be employed to predict the effects on the target cells. This method is not only effective in screening potential bioactive single molecules but also natural fractions containing a mixture of compounds (Bray et al., 2016; Caicedo et al., 2016; Nichols, 2007; D. W. Young et al., 2008). It provides insights into the mode of action, physiology and mechanism of toxicity a metabolite or metabolite mixture is causing on a single cell level by measuring its impact on cellular features (Giuliano et al., 1997; Kremb & Voolstra, 2017; Kurita et al., 2015; Swinney & Anthony, 2011) and is therefore also used in commercial drug development (Hoffman & Garippa, 2007).

After identifying the interesting bioactive fractions using the described biological tests, the fractions need to be investigated for their metabolic content using ultrahigh resolution FT-ICR-MS. This detection method is highly specific, sensitive and can identify unknown compounds because of its ultrahigh resolution and the possibility to perform MS/MS experiments. Together with LC, LC-MS can cope with the challenges of plant metabolite analysis (de Villiers et al., 2016; Pang et al., 2016; Smyth et al., 2012).

3.2. Objectives

Many drug discovery workflows are limited by not characterizing the drug in a cellular system in an early phase and therefore not reveal important effects e.g. side effects, penetration of the cell membrane and toxicity (Y. Feng et al., 2009).

The aim of this study was to evaluate the whole bioactivity potential of the generated fractions (chapter 2) with a cellular high content phenotypic profiling assay to characterize the distinct modes of action and identify possible side effects. For that purpose, the effective *in vitro*, non-targeted biological screening HCPP was used (Bray et al., 2016; Caicedo et al., 2016; Nichols, 2007; D. W. Young et al., 2008). After the identification of several bioactivity clusters containing the produced fractions and standards of the LOPAC® library with known effects, the predicted targets were verified using EASY-HIT.

A comprehensive metabolic analysis of the clustered fractions was performed using ultrahigh resolution FT-ICR-MS (Kanawati et al., 2017). The MS data were compared to the gained PDA data and afterwards correlated to the data of the biological assays. For that purpose, a R script calculating the Pearson correlation including correlation coefficient was used to evaluate the identified correlations. Further characterization of the top correlating compounds were performed using NetCalc to calculate the molecular composition (Tziotis et al., 2011) and van Krevelen diagrams (Krevelen, 1950) indicating their molecular class. Additional MS/MS experiments were performed to characterize the top correlating compounds using neutral loss, annotation of the produced fragments and retrosynthesis. For the verification of the isolated and characterized molecule additional MS/MS of the pure compound (synthesized) was performed.

All in all, the aim of this study was to evaluate the broad bioactivity potential of the generated fractions and to validate the predicted target with another *in vitro* test. The correlation of the biological data with FT-ICR-MS data lead to the identification of compounds probably responsible for the shown bioactivity. These compounds were further characterized using MS/MS.

3.3. Results and Discussion

3.3.1. Overview of the HCPP results without clustering

To get the phenotypic information of the pharmacology and the general bioactivity of Cistus and Cistus PP a HCPP screening was performed on the raw extract and the D1 and D2 fractions. The screening was

3. *Cistus incanus* (PP)

performed according to the publication of Kremb & Voolstra, 2017 (described in 5.1.6.) In this test HeLa cells were treated with the re-dissolved samples in quadruplicates. For the sample features the mean of each feature was calculated and afterwards normalized to the corresponding, outlier tested, control feature on the same plate. All 133 features were extracted and combined to give the cytological profile for each sample. In the following HCPP figures green indicates always negative (until 0.5) and red positive (until 1.5) deviation from the mean of untreated control cells which has a value of 1.

Besides traditional single target assays (Suzgec-Selcuk et al., 2011), HCPP provides a deeper understanding of the effects the compounds induced in the cellular system. Therefore, HCPP is used in an early stage of the drug discovery process (Y. Feng et al., 2009). The amount of cellular readouts (features) in this assay is outstanding to deeply profile the mode of action, of the compounds in the raw extracts and produced fractions. In Figure 19 the HCPP results of the raw extract and D1 fractions from Cistus and Cistus PP are shown. On the x-axis selected core features are shown (complete list is provided in the supplementary Table 15) for a better visualization of the measured phenotypic deviations. The so-called core features e.g. one for the endoplasmic reticulum (ER), lysosome or membrane are summed up to show major underlying phenotypic attributes. On the y-axis the raw extract and the corresponding D1 fractions are shown.
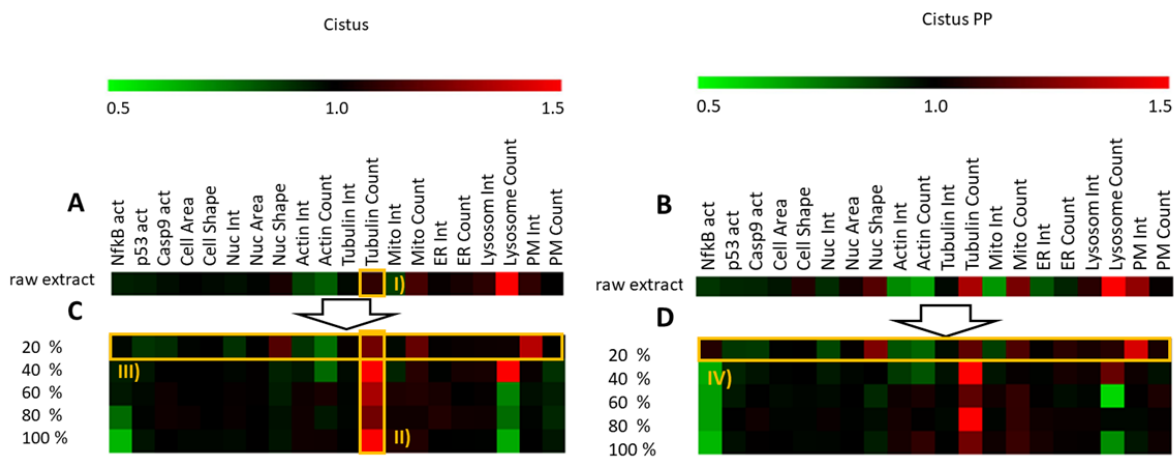
Figure 19: HCPP results of the raw extract and D1 fractions from Cistus A) and C), and Cistus PP B) and D) are shown. The individual extracts and corresponding fractions are represented on the y-axis. On the x-axis 20 cellular core features for better visualization are shown. The mean of n = 4 measurements normalized to the corresponding, outlier tested, control feature on the same plate are shown. Red indicate a positive and green a negative deviation from the mean of untreated control cells with a value = 1. I) marks the up regulation of tubulin in the Cistus raw extract, which shows a lower regulation compared to the D1 fractions II). III) and IV) highlights the D1 20% fraction, which exhibits a different pattern than the other D1 fractions in the HCPP assay.

Cistus and Cistus PP exhibit similar pattern in the majority of HCPP readouts. The raw extracts show a positive effect on the core features tubulin, mitochondria and lysosome count. Additionally, the extracts exhibited a negative effect on the core feature actin count.

In the D1 fractions of Cistus and Cistus PP the 20% fraction exhibit the most distinct pattern of the core features compared to the 40% - 100% fractions (Figure 19 III) and IV)). The positive deviation on the core

features of tubulin count was stronger in the D1 fractions of Cistus compared to the raw extract (Figure 19 I) and II). This effect leads to the assumption, that tubulin core feature down regulating metabolites were relatively enriched in the raw extract, losing their impact as soon as they lose their synergetic effect due to separation (Wagenaar, 2008). Another assumption could be, that the upregulating metabolites build inactive aggregates with other metabolites due to the high metabolite concentration in the raw extract and therefore show no up regulating impact on tubulin (B. Y. Feng & Shoichet, 2006).

To further investigate unique pattern in the core features, the D2 fractions were investigated. After the D2 fractionation, the effects on the core features are more distinct and better traceable due to less possibility for metabolite aggregation and antagonist and agonist mixtures in one sample (Figure 20).
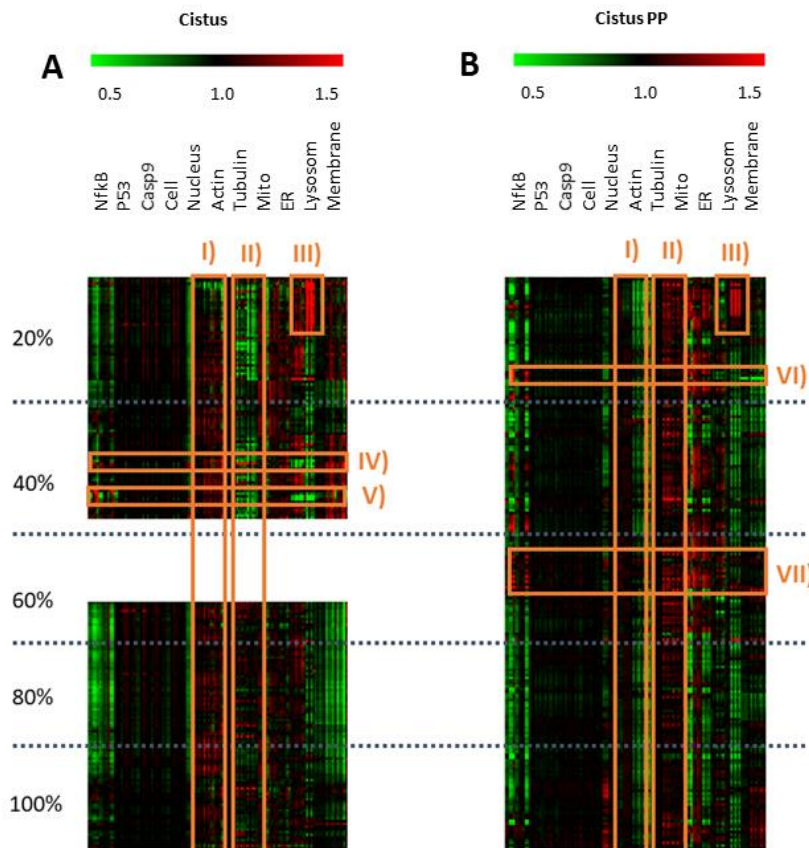
Figure 20: Overview of the HCPP results of the D2 fractions of Cistus A) and Cistus PP B). The mean of n = 4 measurements normalized to the corresponding, outlier tested, control feature on the same plate are shown. Red indicate a positive and green a negative deviation from the mean of untreated control cells with a value = 1. From top down the D2 fractions from 20% to 100 % MeOH are shown in a row and are separated by dotted lines. Due to a problem with the machine, some fractions of Cistus 40% and 60% could not be measured and are therefore left white in the figure. I) indicates the mainly positive deviation for Cistus and negative deviation of actin for Cistus PP. Inversely II) marks the effect on tubulin, which showed a negative deviation in the D2 fractions of Cistus and a positive deviation in the D2 fractions of Cistus PP. In III) the lysosomal marker was strongly upregulated for Cistus and Cistus PP in specific fractions. In IV) and V) two spots with a down regulation for the lysosome and NF-κB features are indicated in Cistus. VI) and VII) mark two spots were NF-κB, actin, tubulin and the ER are up regulated.

3. *Cistus incanus* (PP)

A strong effect on the features of actin was visible for both extracts. For Cistus an upregulation and for Cistus PP a down regulation was visible (Figure 20 I). The tubulin deviations were inversed (Figure 20 II). The polyphenol enrichment might therefore cause a down regulation of the actin and up regulation of the tubulin describing features. A commonality was the strong up regulation of the lysosome features for both extracts (Figure 20 III).

Moreover, the generated HCPP dataset is rich in deviations to untreated HeLa cells like a strong down regulation of the core feature lysosomes together with a down regulation of core feature NF-κB in the 40% D2 fractions of Cistus (Figure 20 IV) and V). In Cistus PP spot VI) and VII) mark two regions were NF-κB, actin, tubulin and the ER are up regulated. The strong up regulation of the lysosomes is unique for the datasets and the magnification of that specific area is displayed in Figure 21.
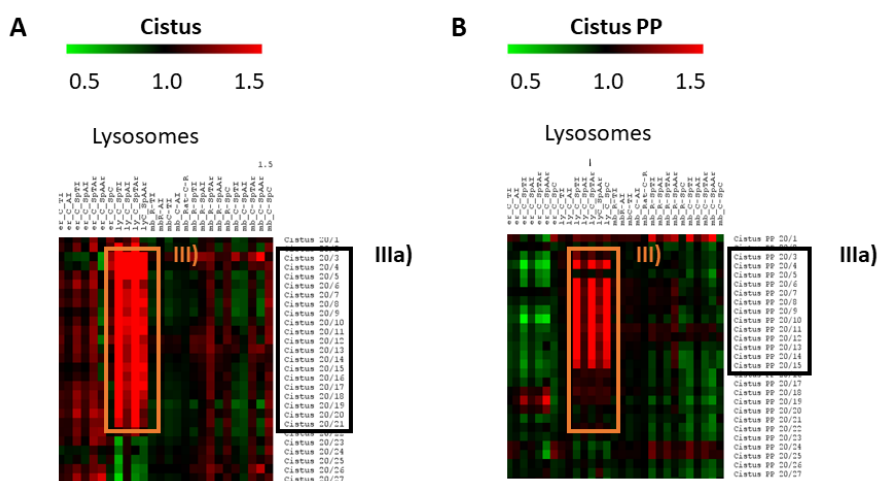
Figure 21: Magnification of the lysosomal effect III) in Cistus A) and Cistus PP B) D2 20%. The mean of n = 4 measurements normalized to the corresponding, outlier tested, control feature on the same plate are shown. Red indicate a positive and green a negative deviation from the mean of untreated control cells with a value = 1. The lysosome marker, see III) was strongly upregulated for Cistus and Cistus PP in specific fractions, see IIIa). This strong effect is unique for these fractions and does not appear in the rest of the data set.

The magnification of this area revealed, that this effect was stronger in Cistus compared to Cistus PP. Moreover, the number of fractions with regulated lysosome core feature was with n=20 higher in Cistus (fractions 2 to 21) than in Cistus PP n=13 (fractions 3 to 15) as displayed in Figure 21 IIIa. To further investigate this HCPP dataset, different cluster methods to standards with known structure and mode of action were applied (3.3.2.).

3.3.2. Overview of the HCPP results with LOPAC® clustering

To narrow down the possible targets for further investigation, the produced high resolution cytological profiles were correlated to those of the reference compounds. The aim of this correlation is the assignment of possible biological targets to certain fractions. The used reference compound library (LOPAC®1280) contained 735 compounds with a variety of cellular targets like neurotransmission,

phosphorylation, ion channels and lipid signaling pathways "LOPAC®1280 Library of Pharmacologically Active Compounds" and were analyzed in the same way as the produced natural fractions (see 5.1.6). In general, several types of clusters with the standard LOPAC® library are possible: (I) cluster with standards that show a certain similarity in their chemical structure, (II) cluster that shows no chemical similarity but the same mode of action or (III) cluster with no structural of functional similarity. In the last case, there could be still a mixed mode of action in the fractions and a further fractionation would be necessary to separate the molecules responsible for the different modes of action (Crisman et al., 2007; Schulze et al., 2013; Daniel W. Young et al., 2007). Or, the fractions might cluster to a set of standards, which are simply not yet included in the database. In Figure 22 the results for the correlation of the raw extracts to the LOPAC® library are displayed.
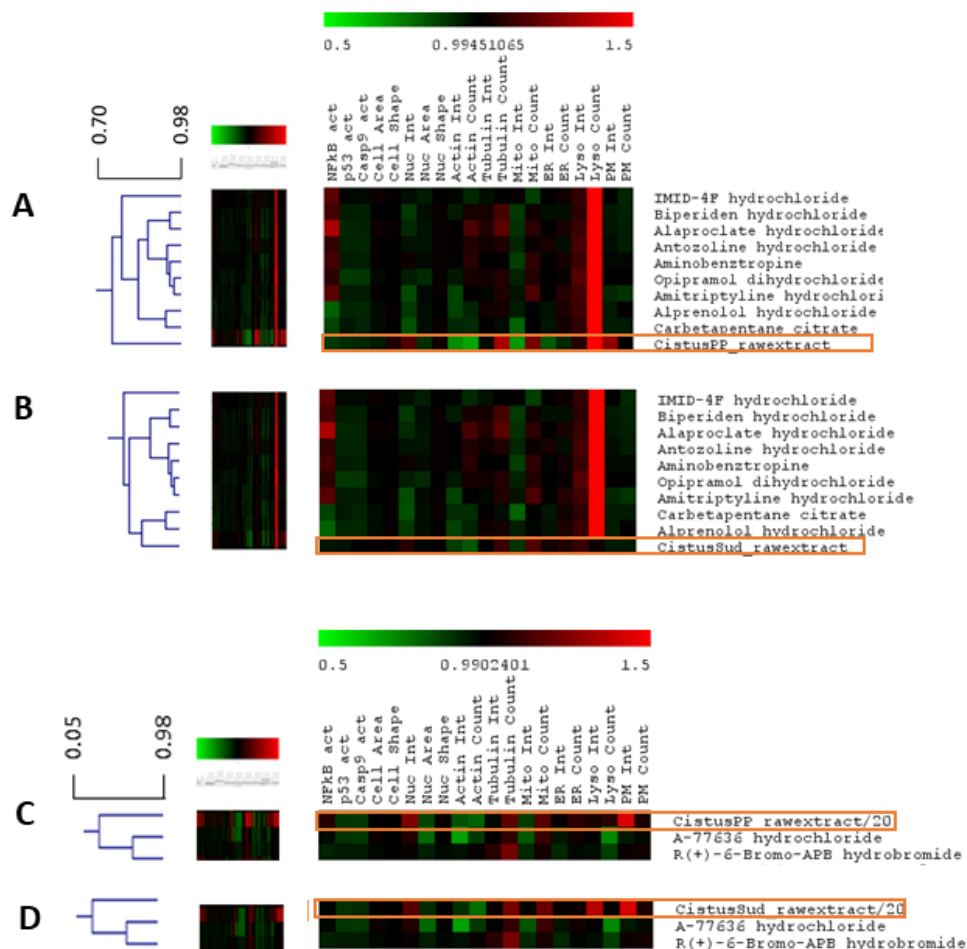
Figure 22 Overview of the raw extract and D1 20% fraction of Cistus and Cistus PP hierarchical clustering of the cytological profiles with LOPAC®. Full cytological profiles were used for the Pearson clustering (left) and 20 core markers are chosen for easier visualization of the cellular effects. The mean of n = 4 measurements normalized to the corresponding, outlier tested, control feature on the same plate are shown. Red indicate a positive and green a negative deviation from the mean of untreated control cells with a value = 1. On the left-hand side in A) the raw extract of Cistus and in C) Cistus PP are shown clustering with the same LOPAC® standards. On the right-hand side in B) the D1 20% MeOH fractions of Cistus and in D) Cistus PP cluster with the same LOPAC® standards (dopamine agonists) R(+)-6-Bromo-APB hydrobromid and A-77636 hydrochlorid.

3. *Cistus incanus* (PP)

Similar to manual investigation of the core features in the Cistus and Cistus PP raw extracts (Figure 19) too many bioactive metabolites were present leading to many hits in the LOPAC® clustering. Due to that, no meaningful biological target could be clustered for the raw extracts even when using different correlations methods like Spearman Rank, Pearson and Kandell`s Tau (Reisen et al., 2013). Both extracts displayed a significantly high up regulation of the lysosomal count. All in all, Cistus and Cistus PP were clustered with the same wide variety of structural and mechanistical different standards revealing that the mode of action between Cistus and Cistus PP was similar but no meaningful tendency towards a biological target could be identified (A) and C)).

Thus after the first fractionation, the D1 20% fractions displayed a correlation to the dopamine agonists (R(+)-6-Bromo-APB hydrobromid and A-77636 hydrochlorid Figure 22 C) and D)). Whether the effect is caused by a single or by multiple compounds could not be distinguished at this point as the fractions were still very complex (Crisman et al., 2007; Schulze et al., 2013). The D1 20% fractionation of Cistus and Cistus PP showed the same core feature pattern in the HCPP. On one side this was predictable as both extracts are derived from the same plant, on the other side the polyphenol enrichment does not yet show any significant difference in the clustering behavior.

For more precise information, the D2 fractions of Cistus and Cistus PP were correlated to the LOPAC® standards to investigate differences and commonalties. Several clusters appeared in the D2 fractions of both extracts. One dataset, namely Cistus PP D2 20% was chosen as an example dataset, as we could find here the most clusters with valid support trees and clear mode of action similarities. A selection of clusters that appeared in the other D2 fractions and could be used for further investigations is summed up in the supplement (see Figure 65 until Figure 68). In Figure 23, an overview of the cluster derived from the D2 20% fractions of Cistus PP are shown. In A) we could see, that fraction 6 until 17 (without 12) are clustering together with the virus inhibitor JFD00244 (from now on named JFD00244 cluster). The positive and negative deviation of the HCPP features will be further characterized in Figure 24.
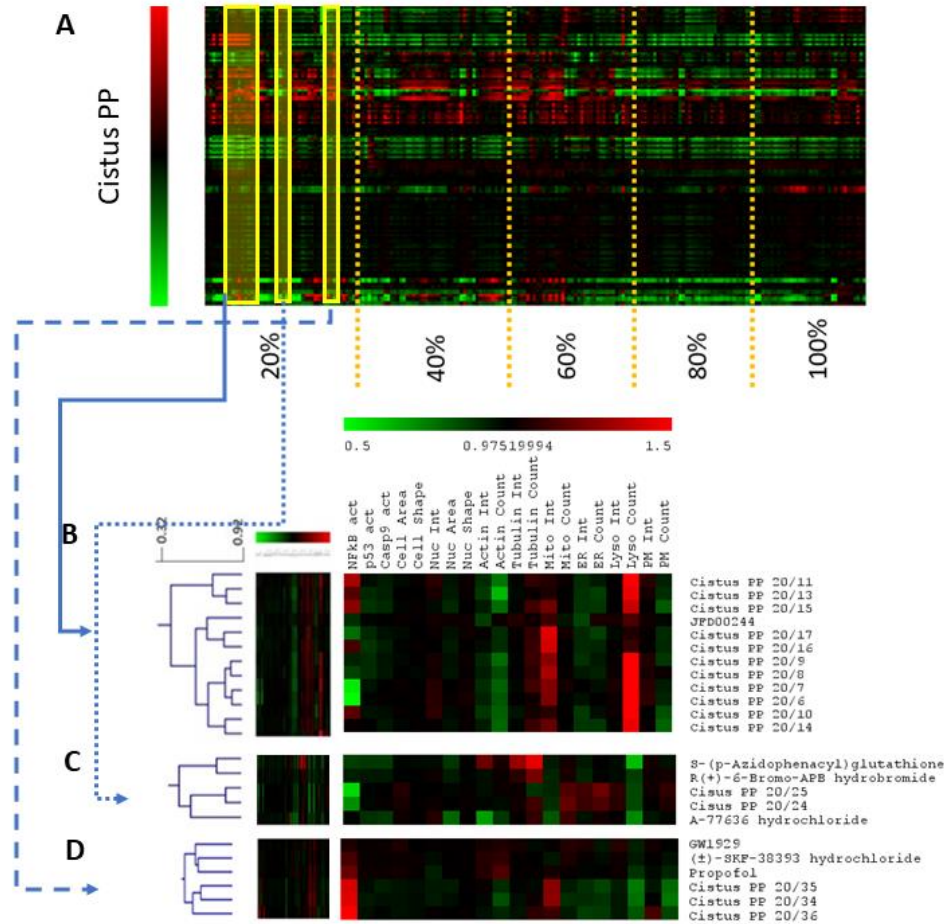
Figure 23: A) Overview of the hierarchical clustering of cytolological profiles from Cistus PP fractions which are separated by dotted lines into the previous 20% - 100% MeOH D2 fractions. Full cytological profiles were used for the Spearman Rank clustering (left) and 20 core markers are chosen for easyier visulatization of the cellular effects. The mean of n = 4 measurements normalized to the corresponding, outlier tested, control feature on the same plate are shown. Red indicate a positive and green a negative deviation from the mean of untreated control cells with a value = 1. A) cluster of fractions 6 until 17 (without 12) (D2 20%) with the virus entry inhibitor standard JFD00244. B) cluster of fractions 24 and 25 (D2 20%) with two dopamine receptor agonists (A-77636 hydrochloride and R(+)-6-Bromo-APB hydrobromide) and a Glyoxalase and glutathione S-transferase inhibitor (S-(p-Azidophenacyl)glutathione). C) cluster of fractions 34, 35 and 36 (D2 20%) with Propofol (neurotransmitter GABA effector), GW1929(PPAR-gamma) agonist) and (+-)-SKF-38393 hydrochloide (dopamine receptor agonist).

Moreover, a second cluster Figure 23 B) includes fractions 24 and 25 and two dopamine receptor agonists (A-77636 hydrochloride and R(+)-6-Bromo-APB hydrobromide) and a Glyoxalase and glutathione S-transferase inhibitor (S-(p-Azidophenacyl)glutathione). These clusters also appeared in the D1 20% fractions of Cistus PP (Figure 22). Although the concentration of the metabolites was different between D1 20% and the fractions D2 20% 24 and 25, the metabolites caused a robust cytological fingerprint, leading to a cluster with standards, even when they appeared in different concentrations (Schulze et al., 2013). In Table 2, an overview of the positive and negative deviations derived from these two clusters are summed up.

3. *Cistus incanus* (PP)

Table 2: Comparison of Cistus PP HCPP results for D1 20% Figure 22 C)) and the D2 20% fraction 24/25 Figure 23 C)) clustering. Both clustered with the LOPAC® dopamine agonists standards R(+)-6-Bromo-APB hydrobromid and A-77636 hydrochlorid. ↑ = up regulation, ↓ = down regulation and ↔ = no regulation. The similar regulations for tubulin and the mitochondrion (mito) are highlighted.

	D1 20% Cistus PP	D2 24/25 Cistus PP
NF-κB	↔	↓
Lysosome count	↑	↔
Protein 53 (p53)	↓	↔
Caspase 9	↓	↔
Actin count & intensity	↓	↔
Membrane intensity	↑	↔
ER count & intensity	↔	↑
Tubulin count	**↑**	**↑(24), ↓(25)**
Tubulin intensity	**↔**	**↔**
Mito count & intensity	**↑(count), ↓(intensity)**	**↑(count), ↓(intensity)**

The clustering of the fractions with the dopamine agonists is most likely based on their effect on the features tubulin, mitochondria and the ER as they are the lowest common denominator after comparing the two clusters (Table 2). The other regulations in the D1 20% cluster might be caused by other bioactive compounds.

Furthermore, fractions 34, 35 and 36 (Figure 23 D)) clustered together with Propofol (positive modulation of the inhibitory function of the neurotransmitter GABA), GW1929 (high affinity peroxisome proliferator-activated gamma (PPAR-gamma) agonist) and (+-)-SKF-38393 hydrochloide (dopamine rezeptor agonist). This cluster is characterized by a clear NF-κB upregulation and a down regulation of the feature lysosomes and the feature membrane. No meaningful cluster with LOPAC® standards in terms of structural or mode of action similarity appeared here, even a common regulation of the features took place.

The identified clusters using different correlations and corresponding positive support tree for B and C in Figure 23 are shown in the supplementary material (Figure 70, Figure 71). As shown in Figure 23, the Cistus PP dataset contains different cluster as it is common for HCPP datasets (Daniel W. Young et al., 2007). More clusters appeared and are summed up in the supplement (see also Figure 65 until Figure 69).

To narrow down the dataset, the identified clusters were filtered according to certain criteria: (I) neighbouring fractions are part of the same cluster to reduce the probability of false positives and negatives as the metabolic content of these fractions must be similar (biological replicate). (II) A strong up and down regulation of the cellular features improves the analysis and thereby the identification of the cellular impact. (III), the cluster should be as stable as possible. In terms of correlation it is favourable that the cluster or sub cluster appear using different correlation algorithms like Spearman Rank, Pearson or Kendall's tau (Reisen et al., 2013). Another criterion for deeper analysis is the availability of a biological test, which can verify the predicted target mode of action. Because 11 subsequent fractions cluster with the JFD00244 standard using the Spearman algorithm, various up and down regulation of the cellular

features are visible, additional clusters with support trees using Kendall's tau and Pearson are present for this cluster (Figure 70) and a biological assay is available, the JFD00244 cluster was chosen for further investigation (Figure 24).
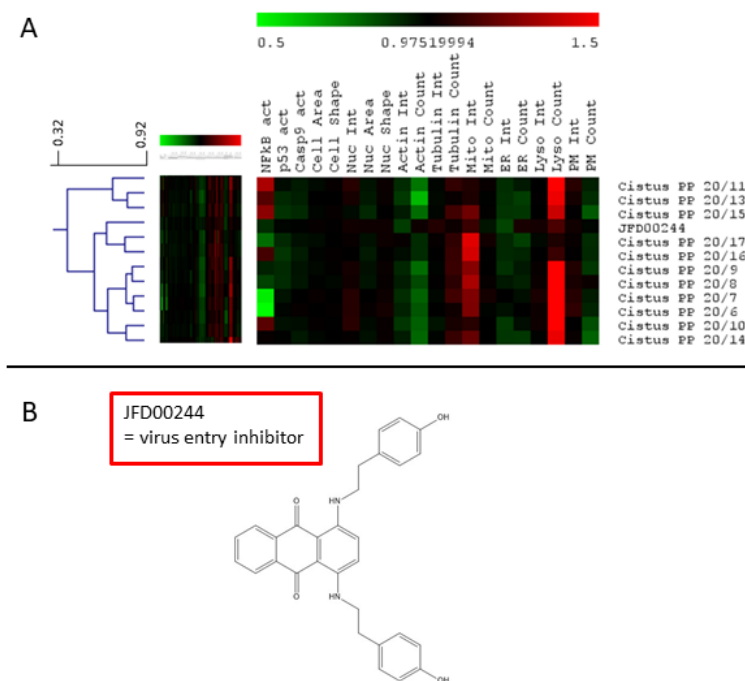
Figure 24: Hierarchical Spearman Rank clustering of cytolological profiles from fractions 6 until 17 (without 12) and JFD00244. Full cytological profiles were used for the Pearson clustering (left) and 20 core markers are chosen for easyier visulatisation of the cellular effects. The mean of n = 4 measurements normalized to the corresponding, outlier tested, control feature on the same plate are shown. Red indicate a positive and green a negative deviation from the mean of untreated control cells with a value = 1. A) HCPP cluster of fractions 6 until 17 (without 12) and the LOPAC® library. B) Structure of the LOPAC® standard JFD00244 (virus entry inhibitor for Lassa, Zika and Marburg virus).

A Spearman Rank clustering of fractions 6 until 17 (without 12) and the LOPAC® standard JFD00244 is shown in Figure 24, but similar clusters and their corresponding support trees are also present using Kendall's tau and Pearson clustering (Figure 70). This cluster is rich in strong up and down regulations of the cellular features. When investigating this cluster in detail, the cellular feature NF-κB leads to a sub classification of the fractions. Similar to the virus inhibitor JFD00244, the fractions 6-9 exhibit a strong down regulation, whereas the remaining fractions, besides 17, showed an upregulation. This finding might be caused by an additional metabolite or metabolic class in the subgroup.

The JFD00244 cluster is characterized by a decreased actin and ER intensity and count. Additionally, a weak up regulation of the tubulin intensity and stronger up regulation of the tubulin count. Moreover, the intensity of the mitochondria and regulation of the lysosomal count were highly up regulated. The virus inhibitor JFD00244 (Figure 24 B) was described in literature to inhibit several viruses e.g. Marburg (20.58 µM, AID=540276), Lassa (5.17 µM, AID=540256) and Zika virus (19.45 µM, 1224854). The common

mode of action is based on the inhibition of the virus cell complex, which leads to a reduction in the cellular uptake of the virus when treated with JFD00244. A protective effect against cell death in human astrocytoma cells could be additionally observed. An additional cellular target is SIRT2 with which JFD00244 forms a hydrogen bond to the backbone carboxyl of Asp95 and Gln167 and interacts with their side chain carboxyl oxygen and with the backbone NH group of Ile169. This leads to an inhibition of SIRT2 with an IC_{50} of 57 μM (Tervo et al., 2006). The HCPP mode of action prediction by comparing standards or single compounds to fractions of plant extracts, works very well as different studies proofed already (Kremb et al., 2017; Kremb & Voolstra, 2017; Schulze et al., 2013; Daniel W. Young et al., 2007). As we could identify the robust JDF00244 cluster we want to validate this prediction with a second test. Therefore, we used the anti-HIV-1 activity test EASY-HIT, as it can measure anti-HIV-1 activity and is additionally able to distinguish between inhibition in the early and late phases of the HIV replication (Kremb et al., 2011).

3.3.3. EASY-HIT

3.3.3.1.   D1 anti-HIV-1 activity distribution and EC_{50} calculation

The fractions of the JDF00244 cluster were analyzed for their anti-HIV-1 activity. Additionally, we included the raw extract, D1 fractions and the D2 20% fractions of the Cistus PP and Cistus datasets to keep track of the anti-HIV-1 activity in the different fractions along the fractionation process. Moreover, the comparison between Cistus and Cistus PP might reveal the impact of the polyphenol enrichment on anti-HIV-1 activity. Using the EASY-HIT we could see, that Cistus and Cistus PP inhibit in the first step of the virus replication cycle (Figure 25).
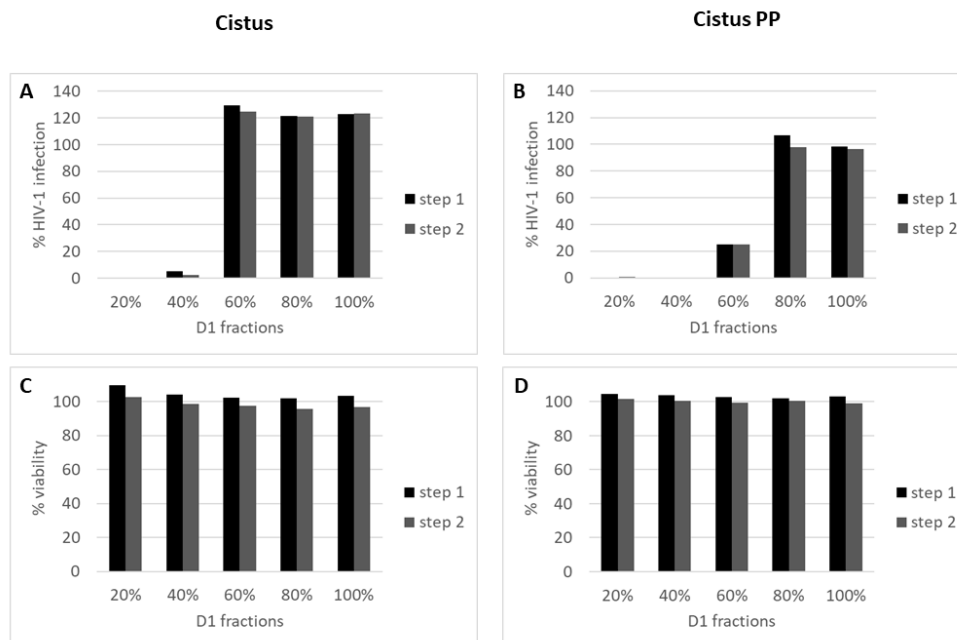
Figure 25: Overview of the anti-HIV-1 activity in the first (black) and second (grey) step of the virus replication cycle of the D1 Cistus A) and Cistus PP B) fractions. In C) and D) the corresponding cell viability test (MTT test) is shown. No reduction of the metabolic process under 90% appeared. n = 1.

As both extracts inhibited in the early phase of the virus replication, the results of the EASY-HIT step 1 were shown from now on (Kremb et al., 2011). These results confirm the HCPP prediction, that fractions of the Cistus PP extract cluster together with JFD00244 (inhibitor for the early phase virus replication). For further investigation the EC_{50} (5.1.1.2) of the raw extracts and D1 fractions was calculated to investigate, how the activity is distributed over the produced fractions. The results are summed up in Table 3.

Table 3: EC_{50} distribution of Cistus and Cistus PP from the raw extract and corresponding SPE fractions, (nq = not quantifiable EC_{50} due to too less inhibition). A cell viability test (MTT test) was carried out for every raw extract and fraction and no reduction of the metabolic process under 80% appeared (supplement, Figure 72). *data from (Stephanie Rebensburg, 2015). The mean and SD of triplicate measurements are shown. The samples were added in µg dry weight per ml cell culture medium.

	Cistus EC_{50} [µg/ml]	Cistus PP EC_{50} [µg/ml]
Raw extract	7.79 ± 0.15*	0.72 ± 0.02*
D1 20%	5.37 ± 0.25	1.06 ± 0.04
D1 40%	8.75 ± 0.42	1.99 ± 0.07
D1 60%	nq	4.97 ± 0.17
D1 80%	nq	5.60 ± 0.46
D1 100%	nq	nq

3. *Cistus incanus* (PP)

Table 3 shows the EC_{50} of Cistus and Cistus PP. In both cases the highest anti-HIV-1 activity of the D1 fractions, which is determined by the lowest EC_{50}, was identified in the 20% MeOH fraction (Cistus 5.37 ± 0.25 µg/ml, Cistus PP 1.06 ± 0.04 µg/ml). Due to that, we assume that we could select the highest amount of active compounds in the D1 20% MeOH fraction. Additionally, the enrichment of polyphenols has a strong impact on the anti-HIV-1 activity by lowering the EC_{50} values from 5.37 µM in Cistus to 1.06 µM in Cistus PP. Additionally we could see, that different classes of polyphenols are enriched, as the strong activity is distributed over 4 fractions, indicating to metabolites with different physicochemical properties (S. Rebensburg et al., 2016).

Interestingly, the main anti-HIV-1 activity was accumulated in the D1 20% MeOH fraction of Cistus PP, which is in line with the identified JFD00244 (virus entry inhibitor) cluster including a set of D2 fractions derived from this 20% D1 fraction. After quality control of the fractions via HPLC-PDA at 200-400 nm (Supplement Figure 63) and TIC (Figure 17) the Cistus PP D1 fractions with 60%, 80% and 100% MeOH were excluded due to the absence of signal.

3.3.3.2. D2 anti-HIV-1 activity distribution and UV spectra

To investigate the potent fractions (20% and 40% MeOH) more in detail, the EASY-HIT was additionally performed on their D2 fractions of both extracts. In Figure 26 the summed-up results are shown.
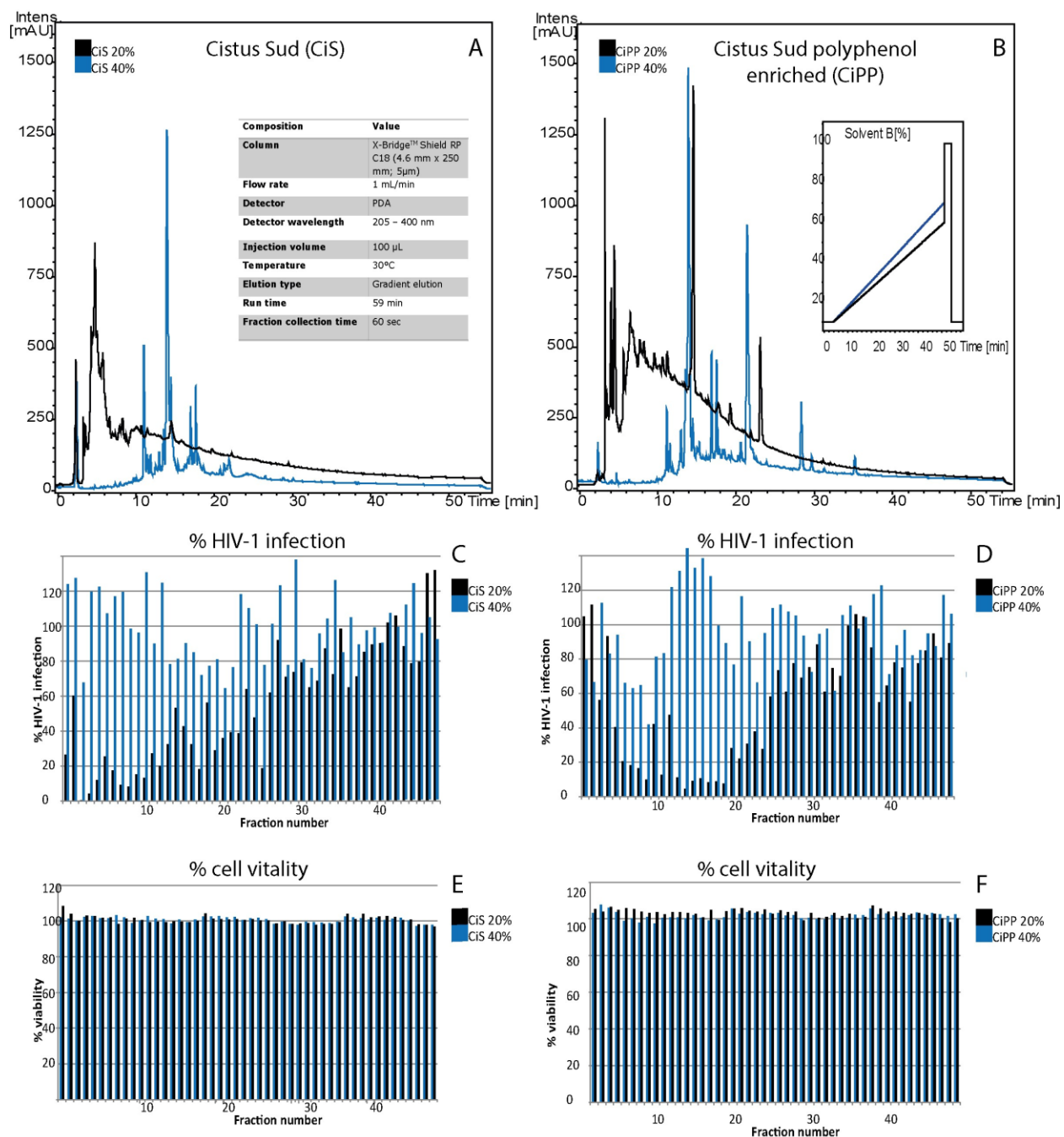
Figure 26: Overview of the results for D2 20% MeOH (black) and 40% MeOH (blue) fractions from Cistus A), C), E) and Cistus PP B), D), F). A)-B) Extracted wavelength of 280 nm of the D2 fractionation with additionally embedded chromatographic conditions and the retention time on the x-axis. C)-D) anti-HIV-1 activity (EASY-HIT) distribution, %HIV-1 infection on the y-axis and the D2 fraction number on the x-axis. E)-F) Overview of the % cell vitality (y-axis) of the tested D2 fractions (x-axis), n = 1.

Figure 26 A), B) contains the HPLC-PDA results for the D2 fractions 20% (black) and 40% MeOH (blue) at 280 nm, an optimum wavelength for phenol detection (González-Rodríguez, Pérez-Juan, & Luque de Castro, 2002) together with the embedded chromatographic conditions. Overall, Cistus and Cistus PP

exhibit a very similar pattern of UV active compounds. Comparing the extracted wavelength at 280 nm the strongest UV active peak was higher for Cistus PP 1400 [mAU] compared to Cistus 1250 [mAU] and the peak area of Cistus PP 20% from min 0 – 20 was also bigger than for Cistus. This result might suggest, that mainly polar polyphenols were enriched using the polyphenol enrichment method (5.1.3) as they eluted in the 20% MeOH fraction. By comparing the overall peak appearance, the peaks are shifted approximately 10 min at 40% MeOH compared to 20%. This was expected, as continuously shifting gradient for separation was chosen to ensure good separation and shorten the separation time (Jandera, 2012). The other wavelength from 200 nm to 400 nm showed no significantly different effects (Figure 63).

The D2 fractions of D1 20% and 40% were tested for their anti-HIV-1 activity and the results are shown in Figure 26 C)-D). Here the anti-HIV-1 activity of the D2 20% fractions was much higher compared to the D2 40% fractions. As similar UV active peaks were available (see A and B) and the % cell vitality was always above 95% in the D2 fractions of 20% and 40% MeOH, we focused for further experiments on the higher anti-HIV-1 active fractions of 20% MeOH.

Starting with the 4th fraction Cistus D2 20% showed a strong inhibition until fraction 25 with a constant HIV-1 activity under 40% for all fractions expect of three. In contrast, Cistus PP D2 20% showed a more bell shaped inhibition (Butler, Fontaine, & Cooper, 2014) where the first 5 fractions were above 40% inhibition. Moreover, a strong inhibition (under 20%) took place for the following 12 fractions and the inhibition changed on the left and right side of these 12 fractions from moderate to no inhibition. Nevertheless, the main inhibition also takes place in the first 25 fractions, like in Cistus, with two exceptions. Of note, the high % of infection of fraction 1-2 for Cistus and Cistus PP verified, that no active compound got lost in the beginning of the fractionation process. Figure 26 E) and F) also displayed, that for all tested fractions the cell viability was above 95%. All inhibitory effects were therefore due to target inhibition and not due toxicity (Helfer et al., 2014). In Figure 27 the anti-HIV-1 activity, UV activity and EC_{50} results from the most active fraction Cistus PP D2 20% 6-9 and 13-19 are shown in more detail.
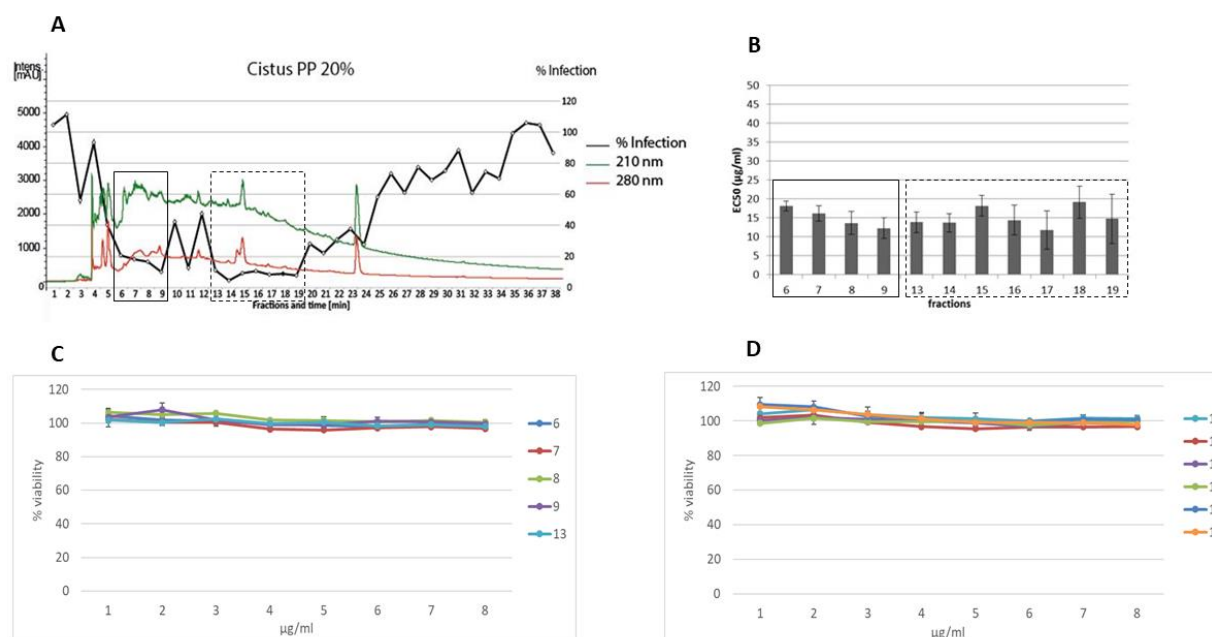
Figure 27: Overview of the combined data from Cistus PP D2 20% 0-38 [min]. A) The wavelength 210 nm is shown in green, 280 nm in red, the anti-HIV-1 activity in black and a section of the fractions (time in [min]) on the x-axis. B) the EC$_{50}$ of the most active fractions 6-9 and 13 – 19 are shown (n = 3). A cell viability test (MTT test) was carried out for every fraction and no reduction of the metabolic process under 80% was visible C) and D). The fractions were added in µg dry weight per ml cell culture medium.

The anti-HIV-1 activity distribution (black line in Figure 27 A) shows a bell-shape with two local maxima at fractions 10 and 12. The highest anti-HIV-1 activity, from the consecutive fractions 6-9 and 13-19 with a calculated EC$_{50}$ between 12 and 19 µg/ml (Figure 27 A), correlate with compounds that have an absorbance at 280 nm and an even higher at 210 nm. In addition, no further activity discrimination of the fractions can be made, as they are in a similar EC$_{50}$ range (Figure 27 B).

Combination and summary of the Cistus PP D2 EASY-HIT and HCPP results are shown in Figure 28. To evaluate the communalities of the Cistus PP fractions clustering with JFD00244 and showing anti-HIV-1 activity, all data are summed up related to those results starting from the raw extract of Cistus PP, to the most active D1 20% fractions and the JFD00244 clustering D2 20% fractions and are shown in Figure 28.

3. *Cistus incanus* (PP)
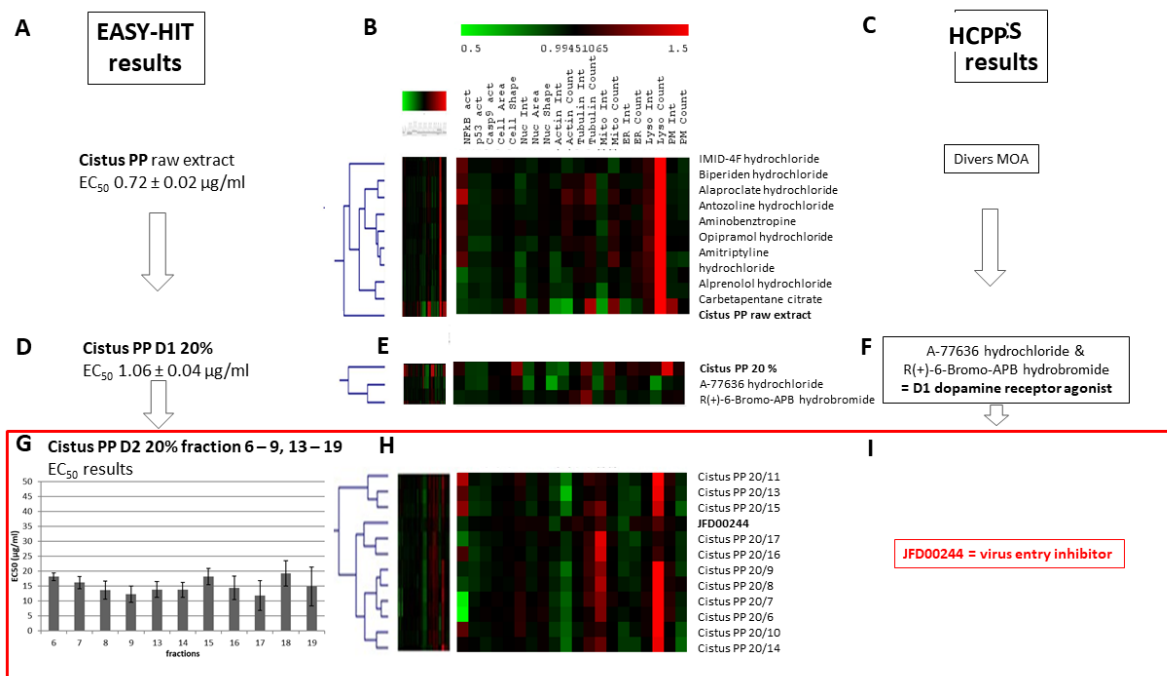
Figure 28: Comparison of the EASY-HIT and HCPP results from Cistus PP raw extract A), B), C). The D1 20% fraction D), E), F) and the D2 20% JFD00244 cluster G), H), I). The EC$_{50}$ values of Cistus PP raw extract, D1 20% fraction and D2 20% 6-9, 13 – 19 fractions are shown A), D) and G). A cell viability test (MTT test) was carried out for every fraction and no reduction of the metabolic process under 80% was visible (see Figure 27), Figure 72). B), E), H) Full cytological profiles were used for the hierarchical clustering (left) and 20 core markers are chosen for easier visualiation of the cellular effects. The mean of n = 4 measurements normalized to the corresponding, outlier tested, control feature on the same plate are shown. Red indicate a positive and green a negative deviation from the mean of untreated control cells with a value = 1. C) divers MOA for the Cistus PP raw extract. F) D1 dopamine receptor agonist cluster for fraction D1 20%. I) virus inhibitor JFD00244 cluster with the D2 fractions 5-11, 13-17. These fractions showed anti-HIV-1 activity at the same time G).

To identify anti-HIV-1 active compounds in the Cistus PP extract, high content phenotypic profiling of cellular readouts and an anti-HIV-1 test (EASY-HIT) were applied. The raw extract of Cistus PP showed a decent EC$_{50}$ in the EASY-HIT test (7.79 ± 0.15 µg/ml), hence it was too complex for a clear HCPP cluster with known standards to identify the substance class or mode of action (Figure 28, A), B), C)). Therefore, further fractionation was necessary to reveal the mode of action for the shown anti-HIV-1 activity. The D1 fractions exhibit a clear ranking in terms of anti-HIV-1 activity with the highest anti-HIV-1 potency in the fraction D1 20% (EC$_{50}$ of 1.06 ± 0.04 µg/ml). After investigating the HCPP results, we could cluster this fraction together with two D1 dopamine receptor agonists (Figure 28, D), E), F)). In addition, this D1 dopamine receptor agonists cluster appeared again in the D2 20% fractions 24/25 (Figure 23) which is characterized by an effect on the test features tubulin, mitochondria and the ER. Testing D1 20% and the D2 20% 25/25 fractions in a dopamine receptor assay would be the next consecutive step and is proposed for further studies.

54

Further fractionation was carried out and the EASY-HIT results and the corresponding HCPP results are shown in Figure 28 G)-H) and I). Anti-HIV-1 active fractions with an EC_{50} between 12 and 19 µg/ml appeared in the 20% sub-fractions. Interestingly, the corresponding HCPP results showed a significant cluster of fractions 5 to 17 with the known viral entry inhibitor for Lassa, Zicka and Marburg virus JFD00244. Table 4 sums up the results of some HCPP core features at different step of the fractionation process from the raw extracts to the D1 20% fraction and the D2 20% fractions (JDF00244 cluster), to retrace the feature deviations.

Table 4: Comparison of Cistus PP HCPP results of the raw extract, D1 20% and the JFD00244 cluster. ↑ = up regulation, ↓ = down regulation and ↔ = no regulation. Similar regulations are highlighted.

	Raw extract	D1 20%	D2 JFD00244 cluster
NF-κB	↓	↔	↑ (10-16), ↓(6-9)
Mito count & intensity	↓	↓	↑
ER count & intensity	↓	↔	↓
Actin count & intensity	↓	↓	↓
Caspase 9	↓	↓	↓
P53	↓	↓	↓
Lysosome count	↑	↑	↑
Membrane intensity	↑	↑	↑

Table 4 shows, that there was a common down regulation for the actin count and intensity, Caspase 9 and p53 in the displayed extracts and fractions. In addition, an up regulation of the lysosomal count and membrane intensity appeared. By further investigating the D2 20% fractions 5-16, we identified a further split of these fractions into a NF-κB down regulating (fractions 6-9) and up regulating (fractions 10-16) part. This might suggest that these two fraction sets possess an at least partly diverse set of metabolites.

The other features are differently up and down regulated between the fractions displayed in Table 4. This might be caused by different relative concentration abundancy of the metabolites in the fractions or the overlaid effect of more bioactive compounds in the raw extract and D1 20% fraction (Wagenaar, 2008). Comparing additionally the cellular features from the identified fractions of the JFD00244 cluster with known HIV IN inhibitors, several communalities and differences could be seen (Figure 29).

3. *Cistus incanus* (PP)
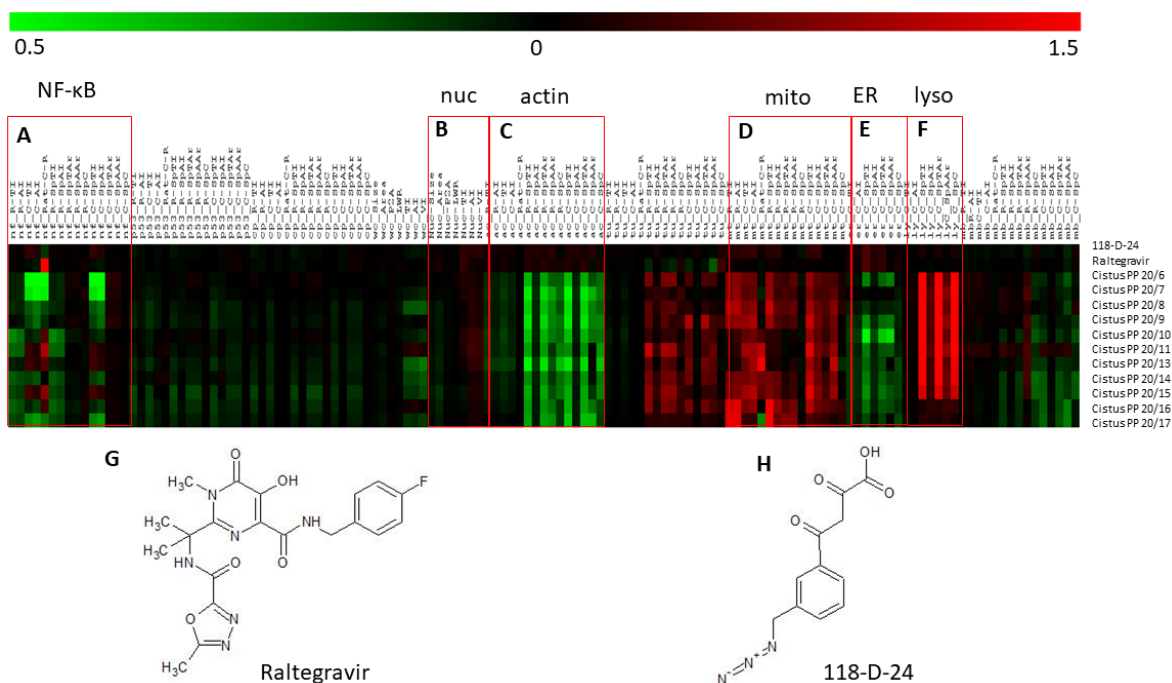
Figure 29: Overview of the HCPP results of 118-D-24, Raltegravir and the D2 fractions of Cistus PP D2 20% 6 – 17 (without 12). The mean of n = 4 measurements normalized to the corresponding outlier tested, control feature on the same plate are shown. Red indicate a positive and green a negative deviation from the mean of untreated control cells with a value = 0.65. Several features showing differences and communalities, A) indicating the NF-κB, B) the nucleus, C) actin, D) the mitochondria, E) the ER, F) the lysosome features. The structure of the IN inhibitors Raltegravir G) and 118-D-24 H) are displayed.

The features for the nucleus A), the mitochondria D) and the lysosomes F) show a common up regulation of the features for the JFD00244 cluster, Raltegravir and 118-D-24. As the up regulation of the lysosomes also appeared for the JFD00244 standard (Figure 24), this effect is the least common denominator effect of IN HIV-1 inhibition in this study. In contrast, the features of NF-κB, actin and ER showed an opposite regulation indicating a new mode of action or the presence of additional bioactive metabolites.

The HCPP predicted anti-viral activity for the JFD00244 cluster was verified with an independent *in vitro* test. The results of the EASY-HIT assay showed a clear inhibition of HIV-1 in the early phase of the virus replication for the predicted set of fractions. To identify the metabolites in these active fractions, ultrahigh resolution mass spectrometry was applied to get more information about the metabolites appearing in this clusters and their impact to the biological effect.

3.3.4. Correlation of mass spectrometry data with anti-HIV-1 and HCPP data

3.3.4.1. FT-ICR-MS measurements of Cistus and Cistus PP

To further characterize the anti-HIV-1 active compounds it was crucial to monitor the chemical space of the active and non-active fractions. Therefore, the D2 Cistus and Cistus PP fractions were additionally

investigated using Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS). This technology allows a more detailed investigation of the chemical diversity due to its extremely high mass accuracy (Marshall et al., 1998) so subsequent prediction of the molecular formulas and their putative annotations is possible (Forcisi et al., 2013; Schmitt-Kopplin, Gabelica, et al., 2010; Tziotis et al., 2011). As the acquired data are very complex, they first need to pass through filtering and alignment steps before they can be correlated to the results of the biological assays. The diligent data processing is shorty summarized in the following chapter [3.3.4.1.1] and [3.3.4.1.2].

3.3.4.1.1. Positive and negative ionization mode and data filtering prior to matrix generation

Prior to the experiment, the ionization mode was monitored to maximize the number of ionizable and detectable substances in the chemical space. Therefore, we tested both ionization modes (positive and negative) and compared the number of measured m/z and the possible annotated metabolites using MassTRIX. In Figure 30 the results for the D1 20% from Cistus PP (chosen as a model sample) are shown.
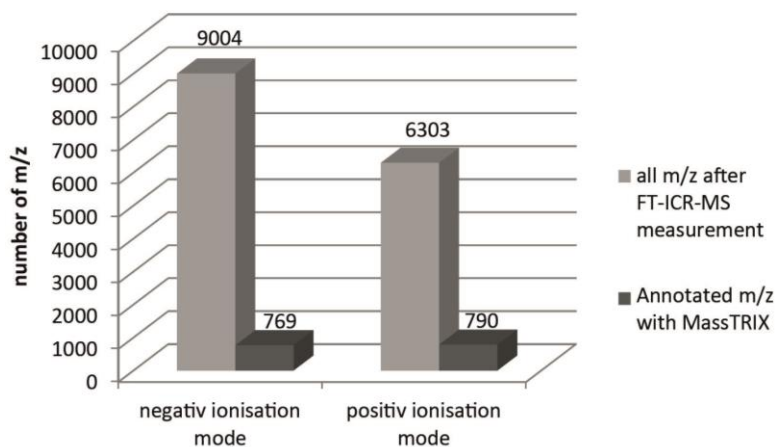
Figure 30: Comparison of the negative and positive ionization modes using D1 20% Cistus PP as a model sample (FT-ICR-MS, 4 MW, 150 Scans). Light grey shows all measured m/z and dark grey shows all annotated m/z using MassTRIX. At a similar amount of annotations much more m/z in total could be measured in the negative mode.

The comparison of both ionization modes revealed that the number of m/z was 9004 (769 annotations) in the negative and 6303 (790 annotations) in the positive ionization mode. Although the annotated metabolites were a bit higher in the positive mode, much more m/z (~2700 additional counts) could be measured in the negative mode.

Moreover, the elevated anti-HIV-1 activity in the Cistus PP compared to the Cistus fractions which was determined by EASY-HIT and HCPP point into the direction, that polyphenols might play a crucial role. Due to the chemical structure of these compounds and the described ionization modes in the literature (Lucci, Saurina, & Núñez, 2017; Schieber, Keller, & Carle, 2001; Thomas, Badr, Desjardins, Gosselin, & Angers, 2018) together with our measurements, the negative ionization mode was chosen. The samples were measured according to the description in 5.2.1 using direct injection (FT-ICR-MS conditions listed in Table 16).

3. *Cistus incanus* (PP)

The measured FT-ICR-MS spectra were further processed and lead to the generation of one data matrix with individual samples on the x-axis and m/z on the y-axis. To create this matrix containing 813 spectra (measured with 4 megaword (MW)), an adapted processing strategy was developed inhouse and published (Kanawati et al., 2017). The description of the filtering is outlined in 5.2.2. The published matrix creation tool possesses several advantages compared to the prior inhouse tool: The matrix generation is faster, has an easy semi-automatic mode, the possibility to work with large studies by aligning more than 800 spectra (4 MW) into one is given matrix and every spectrum can be manually opened for further improvements or quality assurance. The data matrix was generated according to the description in 5.2.3. This matrix was used for further data filtering, correlation and interpretation.

3.3.4.1.2. Visualization of ion suppression and polyphenol enrichment using van Krevelen diagrams

A common problem of complex (plant) extracts is the ion suppression when measured with the FT-ICR-MS (Araujo et al., 2016; Buhrman, Price, & Rudewiczcor, 1996; Taylor, 2005). Available mass signals get lost when they cannot be ionized due to too high competition in the ESI source (Ghaste, Mistrik, & Shulaev, 2016). Compared to the raw extract, our orthogonal fractionation strategy lead to a broader distribution of the chemical structures and thereby a deeper investigation of the chemical space was possible. This improves the chance to identify the metabolites responsible for the anti-HIV-1 activity. The ion suppression is shown in Figure 31, by comparing the chemical space (possible molecule composition calculated with the in-house software NetCalc (Tziotis et al., 2011)) in the raw extract with the chemical space in the corresponding sum of D2 fractions.
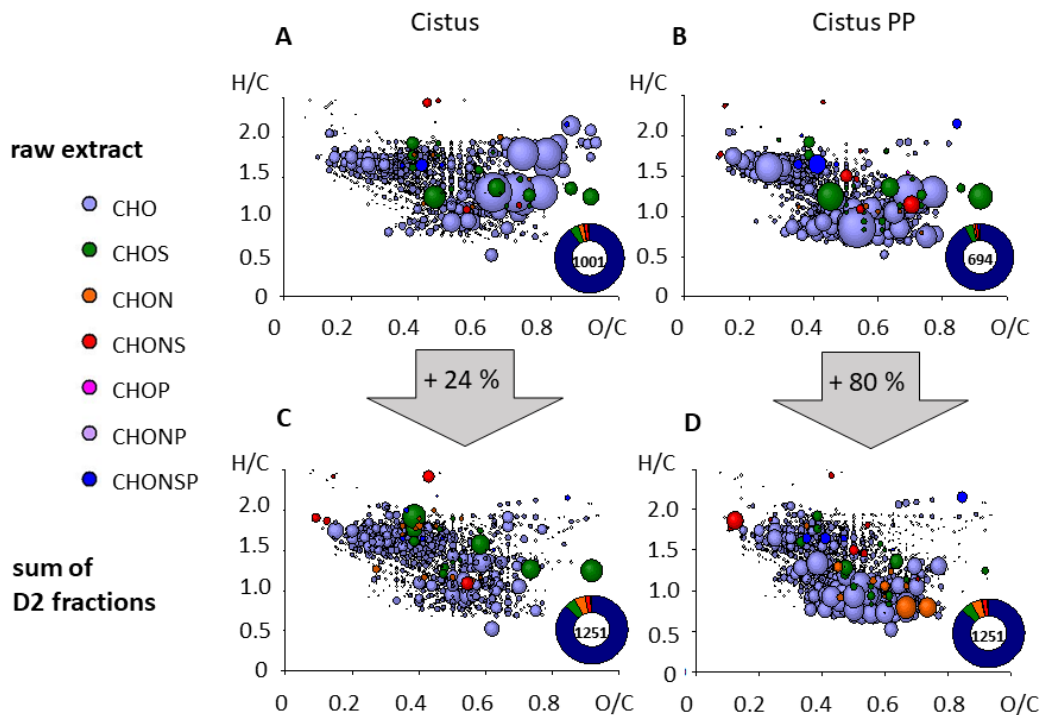
Figure 31: Visualization of the ion suppression effect from comparing the raw extract with the sum of D2 fractions. The increased number of calculated formulas (NetCalc, max 0.1 ppm error) is displayed in % (gray arrows). Krevelen diagrams with H/C ratio on the y-axis and O/C ratio on the x-axis are shown. The size of the bubbles is representing the average intensity. The number of calculated formulas and the van Krevelen diagram for the raw extracts is displayed in A) for Cistus and B) for Cistus PP. The corresponding sum of calculated formulas for the D2 fractions is displayed in C) for Cistus and D) for Cistus PP.

The raw extracts and D2 fractions of Cistus and Cistus PP were covering a wide chemical space. The positive effect of the fractionation on calculated NetCalc formulas are presented in Figure 31. The D2 fractionation leads to additional 250 calculated formulas in Cistus (24% increase compared to the raw extract) and 557 additional calculated formulas in Cistus PP (80% increase). This rise in identified formulas clearly shows the benefit of the extract fractionation when investigating the plant metabolome (Ganzera & Sturm, 2018). Interestingly, the number of metabolites was lower for Cistus PP (694 metabolites) than for Cistus (1001 metabolites). A possible reason could be the higher abundance of certain metabolites, e.g. polyphenolic compounds, supressing other metabolites. Additionally, Cistus PP showed an enrichment of compounds in an specific area (O/C values from 0.3 to 0.8 and H/C values from 0.6 – 1.2)

3. *Cistus incanus* (PP)

compared to Cistus (Figure 31 B)), D)). To visualize the PP enrichment, the ratio of Cistus PP/Cistus in the D2 fractions was calculated (method Helfer et al, 2014) and is displayed in Figure 32.
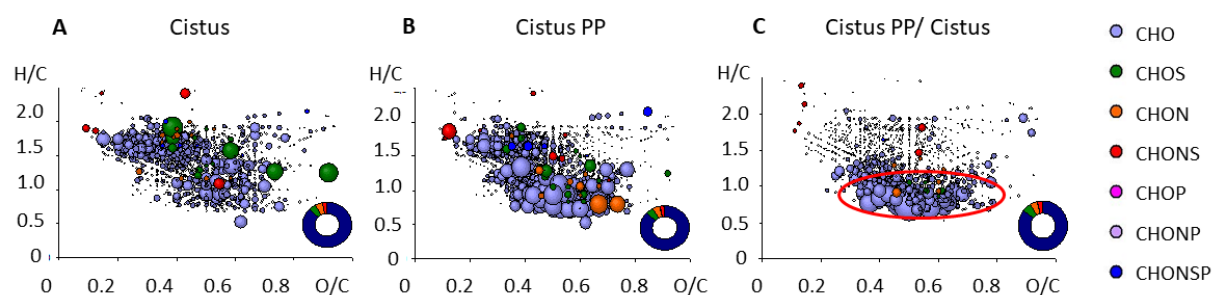
Figure 32: Visualization of the polyphenol enrichment using van Krevelen diagrams with an H/C ratio on the y-axis and O/C ratio on the x-axis. Metabolites with an error of < 0.1 ppm from experimental to NetCalc calculated m/z are shown. Van Krevelen diagrams of D2 fractions of Cistus A) and Cistus PP B) and the intensity ratio for Cistus PP/Cistus C) respectively 1251 molecular masses with calculated chemical composition are presented.

Figure 32 C) nicely illustrates an enrichment of molecular masses in a certain area of the van Krevelen diagram (O/C values of 0.3 to 0.8 and H/C values of 0.6 to 1.2), which is known in literature to contain polyphenols (Kremb et al., 2017; Polyakov et al., 2017). The verification by the van Krevelen plot of the FT-ICR-MS data confirms that the used enrichment procedure for Cistus (5.1.3) traceable enriches masses in the polyphenolic area.

3.3.4.2. Spearman Rank correlation of FT-ICR-MS and anti-HIV-1 data

For a better identification and characterization of the anti-HIV-1 active metabolites the FT-ICR-MS data and anti-HIV-1 activity data determined by the EASY-HIT test of the D2 fractionation were correlated. For the correlation the EASY-HIT results from Cistus and Cistus PP were combined, as any bioactivity pattern visible in Cistus must also be visible in Cistus PP in the same way or even stronger because of the polyphenol enrichment. The usage of both datasets serves therefore as biological replicate.

The Spearman Rank correlation was chosen as correlation algorithm (protocol 5.1.5) to identify similar patterns in the mass signals and anti-HIV-1 results. To rank the correlation results, the corresponding probability values (p-values) were calculated. After investigating all possible matrix combinations for their corresponding Spearman Rank correlation coefficient (r_S) and p-values (Figure 73, Figure 74), the top correlating matrix (D2 20% of Cistus PP and Cistus) with the lowest r_S and p-value was selected for further investigations and is displayed in Figure 33.

60
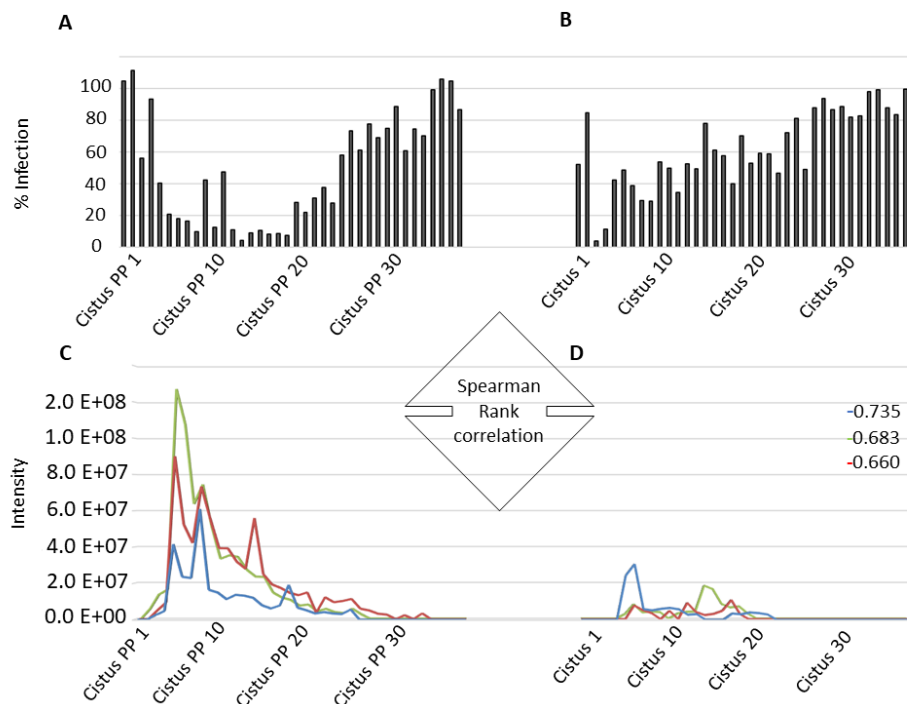
Figure 33: Spearman Rank correlation of EASY-HIT and MS intensity data distribution. (A) % Infection of D2 20% of Cistus PP A) and Cistus B). In C) and D) three representative m/z intensity distributions (r_S) -0.735, r_S -0.683 and r_S -0.660 all from the top 100 correlating compounds) in D2 20% are shown exemplary.

In Figure 33 A) and B) the EASY-HIT activity distributions for the D2 20% Cistus PP and Cistus fractions are shown. Additionally, in Figure 33 C) and D) three out of the 100 top correlating m/z and their distribution over the fractions are shown as an example for the metabolite distribution. The higher intensity in Cistus PP results in a higher HIV-1 inhibition just like for Cistus. The r_S and *p*-values to every correlation of m/z to the specific activity distribution were calculated using a R script for Spearman Rank correlation. In Figure 34 A) the top 100 significantly (p**) correlating compounds are marked and illustrated in a van Krevelen diagram in B) and C). All exact masses and calculated formulas of the top 100 correlating metabolites with associated r_S and *p*-values are listed in Table 25.
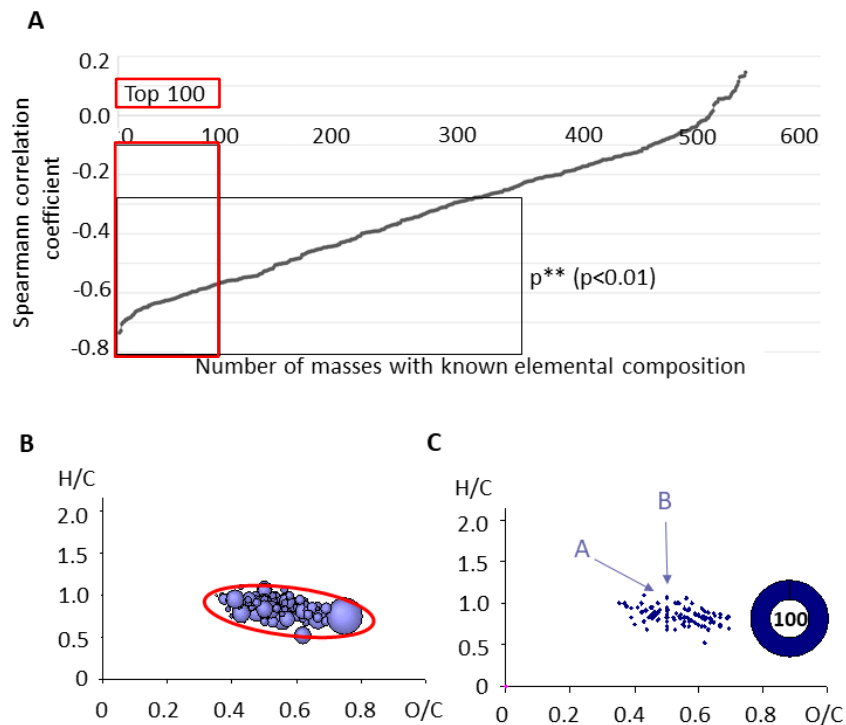
Figure 34: Top 100 masses with known elemental composition and their distribution in the van Krevelen diagram. A) Spearman Rank correlation coefficient on the y-axis and m/z with assigned molecular formulas on the x-axis. The top 100 correlating assigned molecular formulas and the area of significant correlation (p) are marked. B) van Krevelen diagram of the top correlating 100 compounds. Only metabolites with an error of < 0.1 ppm from experimental to NetCalc calculated m/z are shown. C) van Krevelen diagram, which indicates the connection of the compounds via methylation, demethylation or alkyl chain elongation (see arrow A) and hydrogenation or dehydrogenation (see arrow B).**

340 metabolites correlate significantly (p**, p < 0.01) with the measured anti-HIV-1 distribution (Figure 34 A)) The top 100 of these compounds were illustrated in Figure 34 B) and C). Interestingly, the metabolites were clearly located in the area of polyphenols from 0.5 H/C – 1.1 H/C and 0.35 O/C – 0.8 O/C in the van Krevelen diagram (Figure 34 B)) Moreover, a connection of the compounds by methylation, demethylation or alkyl chain elongation (see arrow A) and additionally a reaction path indicating hydrogenation or dehydrogenation (arrow B) (S. Kim, Kramer, & Hatcher, 2003) could be observed Figure 34 C). All in all, the correlation of EASY-HIT and MS revealed to 340 significantly correlating compounds. The top 100 of these compounds were picked for deeper investigation.

3.3.4.3. Characterization of the top 100 correlating compounds using network analysis and the National Institute of Allergy and Infectious Disease database (NIAIDS database)

To identify the compound class of the bioactive metabolites, the chemical properties of all compounds were compared. First hints gave already the van Krevelen diagram showing a connection between the compounds through different reaction pathways Figure 34 E)) Additionally, the whole Cistus metabolome derived from the D2 fractions was plotted in a mass difference network (Figure 35). The nodes represent calculated formulas and the edges were distinct chemical reactions like dehydrogenation or oxidation, showing the connectivity of the compounds to each other. The nodes of the top 100 correlating compounds were marked in red and indicate a clear clustering in Figure 35 B).
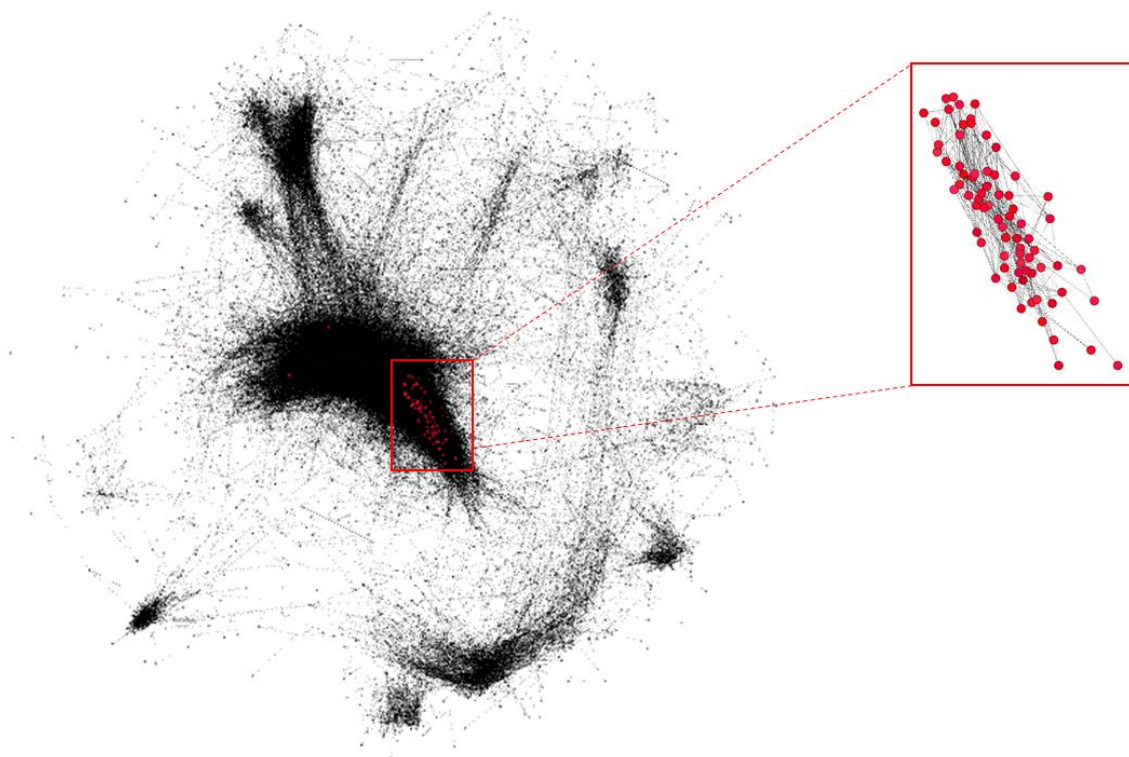
Figure 35: Mass difference network of the Cistus PP metabolome derived from the D2 fractions. The nodes of the top correlating compounds are marked in red.

Figure 35 clearly shows a close clustering of the top correlating compounds. Together with Figure 34 E) it can be concluded that these compounds behave chemically similar. To further characterize these compounds, the exact metabolic compositions found with NetCalc were compared with the National Institute of Allergy and Infectious Disease database (NIAIDS database). 20 of the top 100 compounds were listed in NIAIDS database with an EC_{50} (cell) smaller than 200 ug/ml. Every EC_{50} above that concentration was consider as inactive and therefore is not shown in Table 5 (Tervo et al., 2006).

3. *Cistus incanus* (PP)

Table 5: Detailed overview of the 20 out of the top 100 correlating compounds, which are registered in NIAIDS.

Number	r_S	*p*-value	m/z	Assigned Molecular Formula	Error pmm exp. to theo.	Max. intensity
1	-0.735	7.87×10^{-18}	236.032	$C_{11}H_8O_6$	0.081	6.09E+07
2	-0.683	8.90×10^{-15}	386.064	$C_{19}H_{14}O_9$	-0.021	9.01E+07
3	-0.660	1.29×10^{-13}	474.080	$C_{22}H_{18}O_{12}$	-0.008	1.27E+08
4	-0.657	1.75×10^{-13}	294.074	$C_{14}H_{14}O_7$	-0.028	2.64E+08
5	-0.650	4.04×10^{-13}	376.079	$C_{18}H_{16}O_9$	0.074	1.26E+08
6	-0.**649**	4.33×10^{-13}	306.038	$C_{14}H_{10}O_8$	-0.065	6.74E+07
7	-0.643	8.38×10^{-13}	390.095	$C_{19}H_{18}O_9$	0.021	8.46E+07
8	-0.640	1.12×10^{-13}	430.090	$C_{21}H_{18}O_{10}$	-0.002	1.40E+08
9	-0.633	2.32×10^{-12}	418.090	$C_{20}H_{18}O_{10}$	0.018	1.29E+08
10	-0.631	2.71×10^{-12}	478.075	$C_{21}H_{18}O_{13}$	-0.066	1.67E+08
11	-0.629	3.25×10^{-12}	346.069	$C_{17}H_{14}O_8$	-0.026	3.94E+08
12	-0.626	4.40×10^{-12}	342.074	$C_{18}H_{14}O_7$	-0.021	4.29E+07
13	-0.626	4.48×10^{-12}	388.043	$C_{18}H_{12}O_{10}$	-0.035	3.69E+08
14	-0.626	4.50×10^{-12}	458.085	$C_{22}H_{18}O_{11}$	-0.035	3.69E+08
15	-0.618	1.02×10^{-11}	578.142	$C_{30}H_{26}O_{12}$	0.046	3.86E+07
16	-0.615	1.35×10^{-11}	314.043	$C_{16}H_{10}O_7$	-0.085	1.50E+07
17	-0.605	3.41×10^{-11}	358.069	$C_{18}H_{14}O_8$	0.020	6.38E+07
18	-0.596	7.33×10^{-11}	316.095	$C_{17}H_{16}O_6$	-0.058	9.63E+06
19	-0.595	7.87×10^{-11}	274.048	$C_{14}H_{10}O_6$	0.040	3.48E+08
20	-0.595	8.44×10^{-11}	250.048	$C_{12}H_{10}O_6$	0.061	2.42E+07

The top 100 correlating compounds show an additional clear clustering using the mass difference network, which strengthen the hypothesis of their similarity. Furthermore, from these top 100 correlating compounds 20 exact calculated formulas were registered in the NIADS database to be highly bioactive. The number of metabolites which are highly probable responsible for the bioactivity, is now further narrowed down to 20.

3.3.4.4. Comparison of the top 100 correlating anti-HIV-1 compounds with the metabolites in the HCPP derived JFD00244 cluster

The list of the top 100 correlating and characterized compounds with known chemical composition is now compared to the compounds, which are enriched and/or unique in the HCPP cluster of fractions 5-11, 13-17 (compared to the other D2 20% fractions) with the virus inhibitor JFD00244.
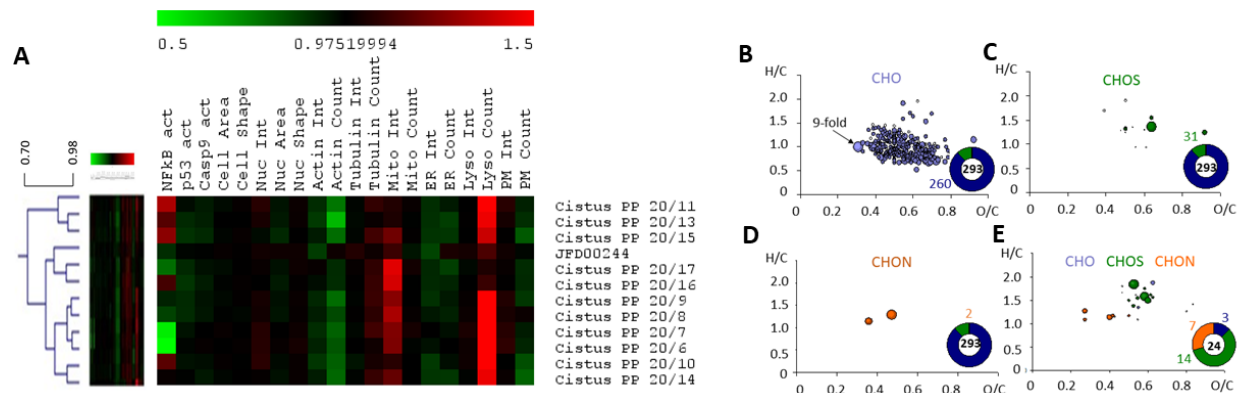
Figure 36: Overview of the JFD00244 cluster and the enriched metabolites in a van Krevelen diagram. Only metabolites with an error of < 0.1 ppm from experimental to NetCalc calculated m/z are shown. In A) the HCPP cluster of Cistus PP D2 20% fractions 5-11, 13-17 with JFD00244 is displayed. The van Krevelen diagram of the metabolite enrichment compared to the other D2 fractions are displayed for CHO B), CHOS C) and CHON D) with a max. enrichment of 9-fold. The unique metabolites are displayed in E).

The average intensity in the clustering fractions divided by the average intensities in the non-clustering fractions (fraction 1-4, 12, 18-38) was calculated and illustrated in a van Krevelen diagram (Figure 36 B) to E)). A polyphenolic area with 260 compounds could be identified in B. These compounds are up to 9-fold enriched. There are two CHON compounds and 31 CHOS compounds. Additionally, a specific area of 14 unique CHOS compounds could be identified. These in total 317 compounds were now compared to the top 100 correlating compounds from the anti-HIV-1 tests and 81 compounds were identified to be in both experiments (Figure 37).
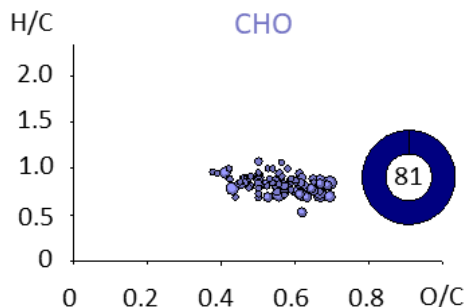
Figure 37: Overlap of 81 metabolites from the top 100 EASY-HIT correlation and the enriched metabolites from the HCPP JFD00244 cluster. Up to five-fold enrichment can be seen. Only metabolites with an error of < 0.1 ppm from experimental to NetCalc calculated m/z are shown.

Furthermore 17 out of the 20 compounds, which are registered in NIAIDS, appeared in both datasets. Just three namely number 12, 15 and 18 from Table 5 were not in the JFD00244 cluster. Because these 17 compounds are the lowest common denominator between the HCPP and EASY-HIT datasets, it is highly probable that they are indeed responsible for the bioactivity and, they were therefore further characterized using MS/MS experiments.

3.3.5. Structure determination

MS/MS experiments often contain several challenges like sufficient peak intensity, isolation of the parent ion and generation of meaningful fragments. Nevertheless, the parental ion [M-H]$^-$, m/z 273 (number 19 in Table 5) was isolated and fragmented from fraction 8 (D2 20%). Among all available parental ions and fractions, the best MS/MS conditions were found in fraction 8 with the parental ion [M-H]$^-$, m/z 273 with respect to available intensity, the possibility of ion isolation and the existence of meaningful fragments (Figure 75 of the supplement contain the MS/MS spectra using 5 eV, 10 eV, 15 eV, 20 eV and 25 eV of this isolated ion). Using the fragment analysis tool MetFrag and retro synthesis fragmentation analysis a preliminary parental ion structure was proposed, namely 4-(6,7-dihydroxy naphthyl)- 2,4-dioxobutanoic acid (Figure 38).
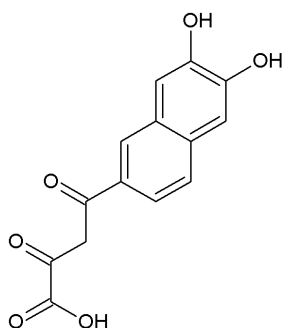
Figure 38: Molecular structure of 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid.

The molecule was synthetized in cooperation with Prof. Oliver Plettenburg, Institute of medicinal chemistry, Helmholtz Zentrum München, Hannover after a 4.5 month lasting unsuccessful attempt to synthetize this structure in our lab (multi step synthesis protocol of in-house synthesis summarized in 6.2.7.). Following experiments (including this standard), in the laboratory of Prof. Dr. Philippe Schmitt-Kopplin, performed by Dr. Müller and Dr. Kanawati on the successfully synthesized standard, showed, that the experimentally discovered fragments of the parental ion [M-H]$^-$, m/z 273 in fraction 8 (Figure 75) did not match the fragmentation pattern of the proposed standard 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid. The fragmentation pattern of the pure standard at different collision energies is shown in the supplement (Figure 77 and Figure 78)

Nevertheless, literature research showed, that the keto form of 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid was previously synthesized to inhibit the HIV-1 integrase (IN) as a biological target (Maurin et al., 2006). The anti-HIV-1 activity of the standard was experimentally confirmed using an the

EASY-HIT test (chapter 5.1.1.1 - 5.1.1.3). The gained EC_{50} results of 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid is shown in Table 6 and in the supplement Figure 82 (standard EC_{50} curve).

Table 6: EC_{50} of 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid. A cell viability test (MTT test) was carried out and no reduction of the metabolic process under 80% appeared. Experiment was done in the lab of Prof. Dr. Ruth Brack-Werner.

	EC_{50}
4-(6,7-dihydroxy naphthyl)- 2,4-dioxobutanoic acid	21.91 [μg/ml]

Based on the anti-HIV-1 activity we further searched for the appearance of 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid in the Cistus PP extract.

Further experiments showed that the compound 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid is present in the D1 20% fraction of Cistus PP. The parental ion [M-H]$^-$, m/z 273 was successfully isolated and fragmented at different collision energies (Figure 79, Figure 80 and Figure 81). In Figure 39 a compact fragmentation pattern comparison of the standard 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid and the isolated m/z 273, D1 20% sample at the characteristic collision energy of 15 eV is shown.
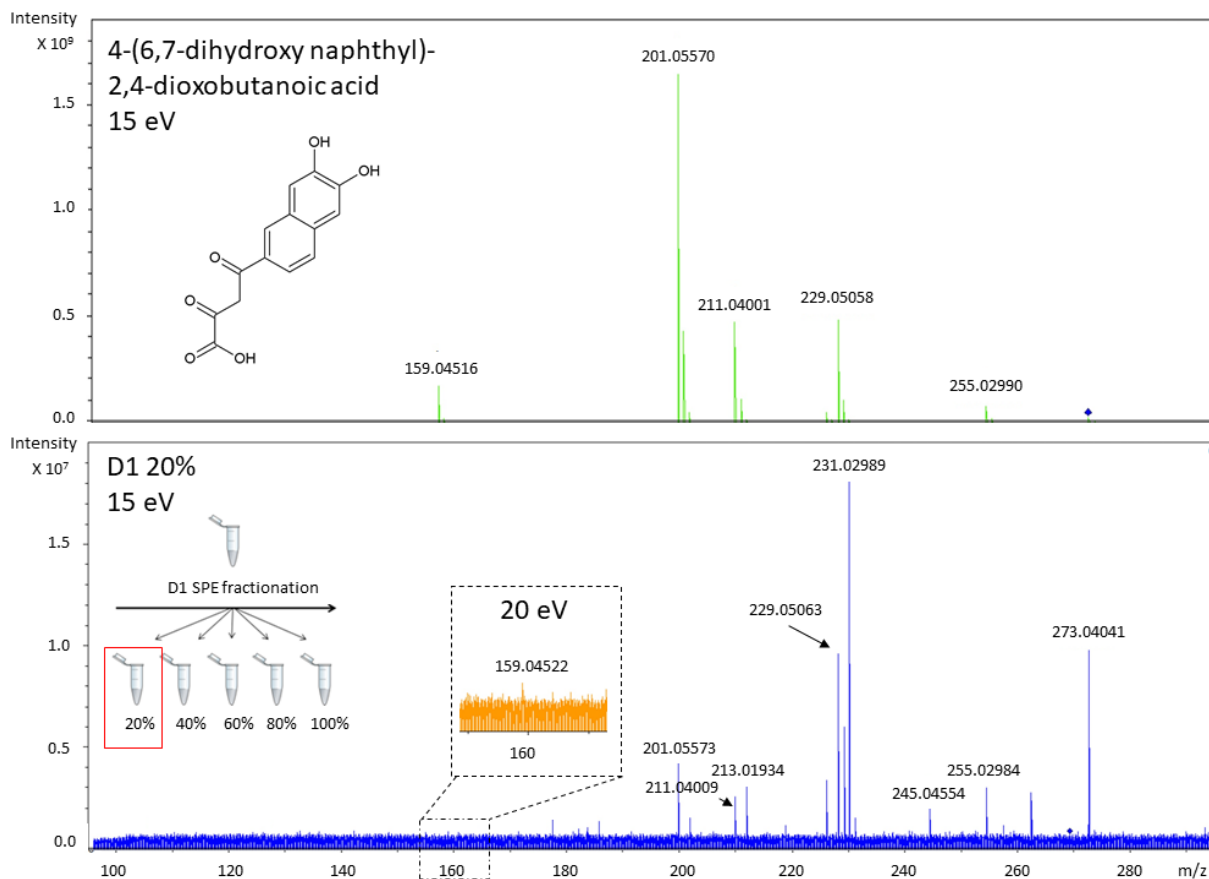
Figure 39: Fragmentation pattern of the standard 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid (top) and the isolated m/z 273.03977, D1 20% (bottom) using (-) ESI-FT-ICR-MS/MS ultrahigh resolution mass spectrometry. A collision energy of 15 electron volt (eV) (and additional 20 eV for D1 20%) was applied to the isolated precursor ions and accurate detected masses are shown.

The comparison of these two fragmentation pattern are leading to two main conclusions. First, the presence of 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid in the sample D1 20% could be verified, due to the detection of the exact parent ion mass and similar fragmentation pattern in the Cistus PP fraction and the standard. Table 7 contains the main fragments of 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid, which were identified in the Cistus PP fraction and the standard.

Table 7: Shared fragment ions (m/z) between Cistus PP D1 20% and the synthetized standard.

Standard m/z	D1 20% m/z	Fragment ion	Elemental composition	Exactly calculated m/z	Standard Error (ppm)	D1 20% Error (ppm)
273.04049	273.04041	[M-H]⁻	$C_{14}H_9O_6+e^-$	273.040462	0.103	0.190
255.02992	255.02984	[M-H-H$_2$O]⁻	$C_{14}H_7O_5+e^-$	255.029897	0.090	0.224
229.05058	229.05063	[M-H-CO$_2$]⁻	$C_{13}H_9O_4+e^-$	229.050662	0.358	0.140
211.04001	211.04009	[M-H-CO$_2$-H$_2$O]⁻	$C_{13}H_7O_3+e^-$	211.040097	0.412	0.033
201.05570	201.05573	[M-H-CO$_2$-CO]⁻	$C_{12}H_9O_3+e^-$	201.055747	0.234	0.085
159.04516	159.04522	[M-H-CO$_2$-CO-CH$_2$-CO]⁻	$C_{10}H_7O_2+e^-$	159.045182	0.138	0.239

The isolated parental ion [M-H]⁻ at m/z 273 isolated from Cistus PP D1 20% generates at 15 and 20 eV very characteristic fragment ions. Due to the exact mass, the corresponding molecular composition of the fragment ions could be calculated. In Figure 40 the proposed fragmentation mechanism A, B and C (color coded) including the corresponding structural loss are shown.
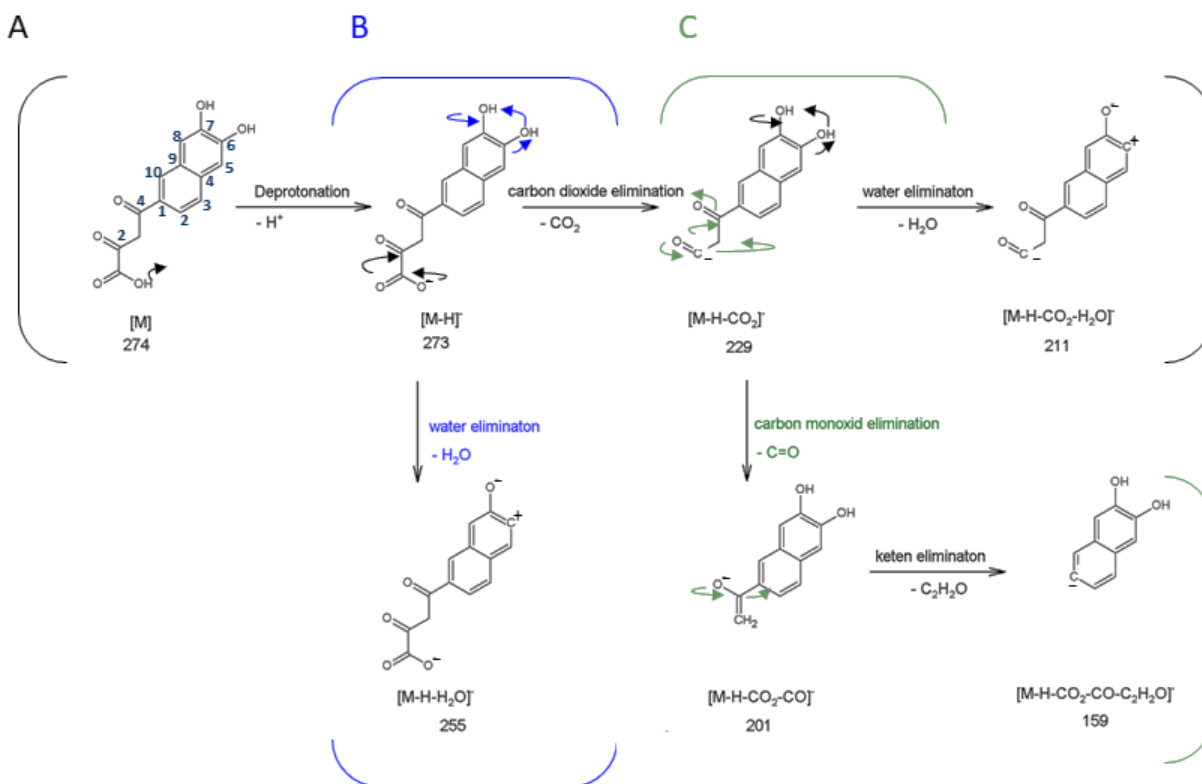
Figure 40: Proposed fragmentation mechanism for 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid.

3. *Cistus incanus* (PP)

The fragmentation starts with the deprotonation of the isolated parent ion [M] m = 274 to yield m/z 273. Subsequently two mechanisms are possible for the generated [M-H]$^-$ at m/z 273. Following the fragmentation mechanism A) in Figure 40, a carbon dioxide elimination from the dioxobutanoic acid side chain take place as the negative charge on the terminal oxygen attacks the C1 carbon of the ketone group (tertiary carbon) in a nucleophilic addition mechanism (black arrows). This leads to a rearrangement whereby CO_2 can be eliminated. Following the fragmentation pattern A), subsequently H_2O can be released from the ion [M-H-CO_2]$^-$ resulting in the stable ion [M-H-CO_2-H_2O]$^-$ m/z 211 (black arrows). Figure 40 path A) show a CO_2 elimination followed by a H_2O elimination, but it can also be inverse.

In the fragmentation path B) a loss of H_2O from the ion [M-H]$^-$ at m/z 273 took place and can form the product ion m/z 255. The remote charge condensation of the two neighbouring hydroxyl groups is here responsible for the water release (blue arrows). Further investigation of the measured MS/MS fragments in Figure 39 proposes another fragmentation path C) (Figure 40). Starting from [M-H-CO_2]$^-$ m/z 229, carbon monoxide was eliminated (green arrows) and thereby the product ion [M-H-CO_2-CO]$^-$ m/z 201 was formed. The remaining atoms of the initial dioxobutanoic acid side were released subsequently in the form of a ketene (green arrows). The elimination of the stable leaving group ketene lead to the product ion [M-H-CO_2-CO-C_2H_2O]$^-$ m/z 159.

The second main conclusion developed out of the fragmentation pattern comprises the existence of additional masses in the fragmentation pattern. The presence of at least one additional isomer in the D1 20% sample, holding the parental ion mas of [M-H]$^-$, m/z 273 is suggested. The additional main fragments in D1 20%, which differ from the standard 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid, are summarized in Table 8.

Table 8: Additional determined fragment ions in the Cistus PP D1 20% fraction.

D1 20% m/z	Fragment ion	Elemental composition	Exactly calculated m/z	D1 20% Error (ppm)
273.04041	[M-H]$^-$	$C_{14}H_9O_6$+e$^-$	273.040462	0.190
245.04554	[M-H-CO]$^-$	$C_{13}H_9O_5$+e$^-$	245.045547	0.029
231.02989	[M-H-CO-CH_2]$^-$	$C_{12}H_7O_5$+e$^-$	231.029897	0.030
213.01934	[M-H-CO-CH_2-H_2O]$^-$	$C_{12}H_5O_4$+e^{--}	213.019332	0.038

To verify the two main outcomes, an additional LC-MS experiment was performed. Beside the exact mass and fragmentation pattern the retention time can be employed as an additional verification marker. The results are shown in Figure 41.
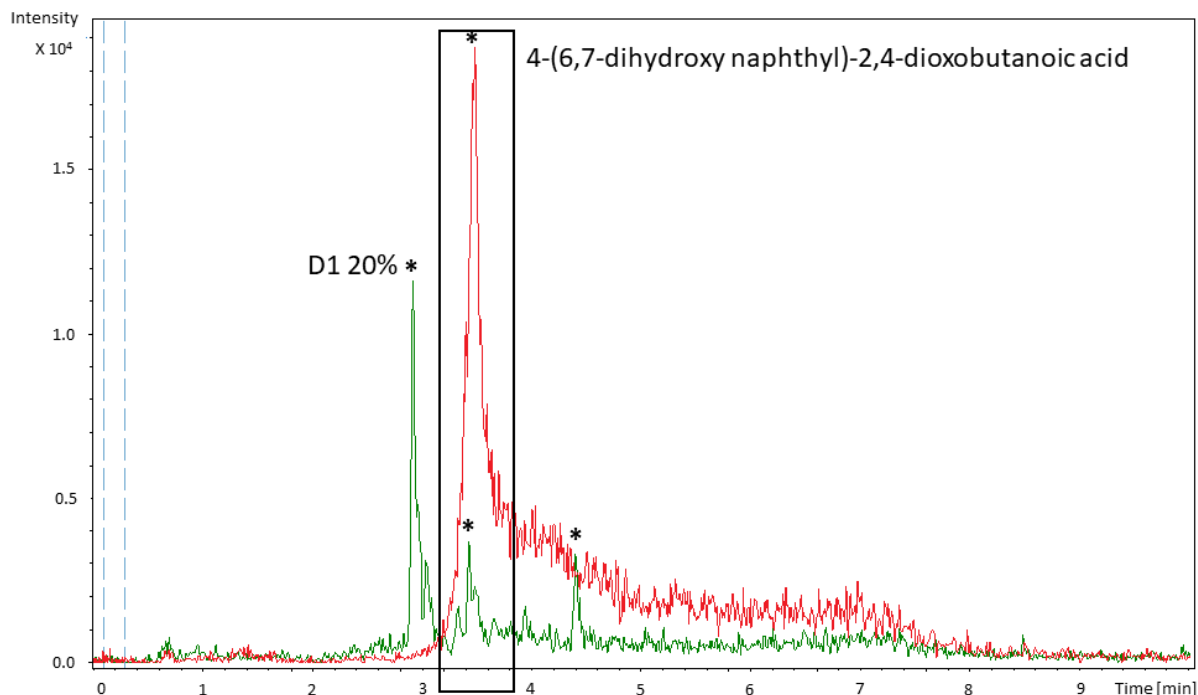
Figure 41: Overview of the results for D1 20% (green) and 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid(red). Extracted ion chromatogram of 273.0390 +/- 0.02 Da, retention time on the x-axis and intensity on the y-axis, * = indicate valid peaks.

The LC-MS experiment clearly shows, that the elution of the standard 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid after 3.5 min, coincident with the elution of a substance with the same properties in the D1 20% fraction using the same chromatographic conditions (Figure 41). Together with the MS/MS experimental data we can conclude that 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid is present in Cistus PP, namely in the fraction D1 20%. This is, to our knowledge, the first time that 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid, an anti-HIV-1 active integrase inhibitor (Maurin et al., 2006), could be identified in a Cistus extract.

Additionally, the elution of two other substances (after 3.0 min and 4.5 min in (Figure 41) was identified. Due to the same exact mass, the very similar retention time in relation to 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid and the identification of additional fragments in the Cistus PP D1 20% fraction (Table 8) these substances might be undescribed isomers of 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid. Although the acquired data already gives hints about, the structure the characterization of these substances would be beyond the scope of this thesis.

3.3.6. Biological characterization of 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid

The verified standard 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid was analyzed for its potential bioactivity. When characterizing molecules in terms of their possible biological impact, certain criteria

need to be proven. One criteria is the comparison of the found molecules to those with known biological activity in terms of their chemical and physiochemical properties (Cramer, Patterson, & Bunce, 1988; A. K. Ghose & Crippen, 1985; Hansch, 1993).

For that issue, several rules were extracted from the the chemical medicinal chemistry database (CMC)(A. Petrauskas & Kolovanov, 2000; A. K. Ghose, Viswanadhan, & Wendoloski, 1999; Lipinski, Lombardo, Dominy, & Feeney, 2001) and applied to the standard 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid (results for the standard = a. in the following list).

(i)   The molecular weight should be between 160 and 480 g/mol molecular mass
          a.   274 g/mol
(ii)   The total number of atoms should be between 20 and 70
          a.   30
(iii)   The molecules should have x < 5 hydrogen bond donors (e.g. OH, NH groups)
          a.   3 donors in the keto and 4 in the enol form
(iv)   The molecules should have x < 10 hydrogen bond acceptors (e.g. O, N atoms)
          a.   6 acceptors
(v)   Additionally, benzene is found to be the most abundant structural unit in CMC
          a.   appears twice in our molecule
(vi)   Additionally, a high count of alcoholic hydroxylic groups frequently appears
          a.   appeared three times
(vii)   The log P of the molecules should be between -0.4 and 5.6
          a.   in our case from 0.13 ± 1.1 to 1.21 ± 1.07 (see Table 9)
(viii) The molar refractivity should be between 40 and 130 cm^3
          a.   between 68.74 ± 0.3cm^3 in the keto and 70.15 cm^3 in the enol from (see Table 9)

Table 9: Calculated log P and molar refractivity for the keto and enol from of 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid (calculated using ChemSketch).

	LogP	Molar refractivity [cm^3]
4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid (keto)	0.13 ± 1.1	68.74 ± 0.3
4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid (enol)	1.21 ± 1.07	70.15 ± 0.3

All seven postulated specifications (i – vii) are applicable for 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid. This suggests, that its physicochemical properties fit to the average bioactive compound. Especially the determined log P and molar refractory indices were important as these values strongly influence e.g. the receptor binding, cellular uptake and bioavailability (Arup K. Ghose & Crippen, 1987). Additionally, the contained benzene rings are a highly abundant subunit in bioactive compounds (A. K. Ghose et al., 1999) due to the lack of stereochemical isomers. Additionally, three alcoholic hydroxylic groups are present. This functional group is often present in bioactive compounds (A. K. Ghose et al., 1999) due to its ability to act as a hydrogen acceptor or donor.

Another criteria for the characterization of biological activity is the understanding of the target protein structure and the active binding site (Kuntz, 1992). In 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid

an α,γ-diketo acids (DKA) subgroup is implemented, which is known to inhibit the HIV-1 integrase. In general, polyphenols served as lead structures for the investigational process of new anti IN drugs over years (Liao, Marchand, Burke, Pommier, & Nicklaus, 2010). Historically, an innovative approach found the patented keto-enols acid-type compounds, also known as α,γ-diketo acids, found by Merck and Shionog. The DKA unit get deprotonated and form a dianion which causes a chelation of the two Mg $^{2+}$ ions present in the IN enzyme (Liao et al., 2010).

As DKA was described as key substructure for IN inhibition, and further DKA compounds were identified as potent and selective inhibitors for HIV-1 integrase over the years (Li & Vince, 2006) e.g. L-731,988, L-708,906 (Anthony, 2004; D. J. Hazuda et al., 2000; Sechi et al., 2004) (see Figure 42)

Three of them, namely S-1360 (Shionogi & Co. Ltd; Billich, 2003), L-870,810 (Daria J. Hazuda et al., 2004) and JTK-303 (GS-9137, Elvitegravir) (Japan Tobacco design)(Allison, Christophe, & Yves, 2004) were part of HIV-1 clinical trials (molecular structure shown in Figure 42) but S-1360 was stopped in phase I/II (Rosemond, St John-Williams, Yamaguchi, Fujishita, & Walsh, 2004) and L-870,810 was discontinued (Little S, 2005) as JTK-303 performed more effective (Liao et al., 2010).
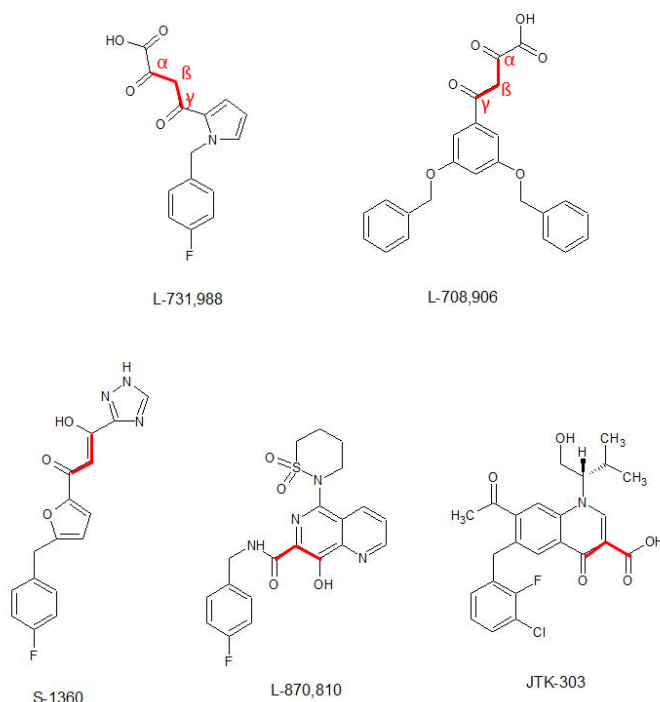
Figure 42: Overview of the molecular structure of L-731,988 (including α, β, γ specification, highlighted in red) L-708,906 and S-1360 (stopped in phase I/II), L-870,810 (stopped) and JTK-303 with α,γ-diketo acids moiety.

Tests with two diketo acid moieties on the same molecule which should interact with the reported two metal ions in the active site of IN failed due to a significant decrease of IN inhibition. This intensive research on compounds with an DKA moiety and its significant findings about the mechanism proofs the relevance of this substructure in the mode of action of the IN inhibition.

3. *Cistus incanus* (PP)

In sum we could identify in this study the anti-HIV-1 active compounds 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid for the first time in *Cistus incanus*. The applicability of all points of the seven rules for bioactive compounds established by Ghose 1999 to this compound could explain the physicochemical impact of the structure to its shown bioactivity. Additionally, two more potential isomers could be detected possessing the same exact mass but a different chromatographic behaviour and therefore probably a different side group allocation. It is likely that the other 2 proposed isomers hit the same biological target, if the DKA subgroup remains stable. Further experiments including isolation of the isomers, MS/MS experiments, standard synthesis, bioactivity tests and verification with LC-MS are necessary to identify these two isomers and use them to study e.g. additionally synergetic effects.

3.3.7. Summary and conclusion

To identify the bioactive compounds with anti-HIV 1 activity in Cistus the raw extract, D1 and D2 fractions of Cistus and Cistus PP were analyzed using HCPP, EASY-HIT assays and ultrahigh resolution mass spectrometry. The correlation of the biological assays and analytical measurements revealed a small set of molecular formulas, which are potentially responsible for the shown activity. Using the combination of ultrahigh resolution mass spectrometry, MS/MS and LC-MS one candidate, 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid was identified and characterized. This proven anti-HIV 1 active compounds could be verified in *Cistus incanus* for the first time.

In the beginning, a high content phenotypic profiling (HCPP) of all extracts was performed and clustered with the LOPAC® standards library to investigate the general bioactivity. Due to the complexity of the raw extract and D1 fractions which might lead to false conclusions (Schulze et al., 2013) (B. Y. Feng & Shoichet, 2006), only the D2 fractions were further investigated. The D2 fractions 6-11 and 13-17 of Cistus PP robustly cluster together with JFD00244, a known SIRT2 and cell entry inhibitor against the Marburg, Lassa, Zika and HIV-1 virus based on (I) Spearman Rank clustering algorithm, (II) Pearson and (III) Kendall's Tau clustering. The clustering of several fractions in a row (similar biological replicates) with JFD00244 qualifies it as a significant cluster for further investigation. Additionally, HCPP reveals a unique pattern for these fractions compared to the control (Table 4).

To confirm the bioactivity against HIV, the fractions of the JFD00244 cluster were tested using the anti-HIV-1 activity test EASY-HIT. This *in vitro* test revealed a strong anti-HIV 1 inhibition in the first step of the virus replication in the identified fractions (EC_{50} between 12 and 19 µg/ml) while not affecting the cell viability.

By correlating the FT-ICR-MS measurements to the anti-HIV-1 inhibition profiles of Cistus and Cistus PP D2 20% using Spearman Rank clustering, several compounds were significantly identified. Interestingly, the top correlating compounds in these bioactive fractions are polyphenolic compounds as indicated in the clustering shown in the van Krevelen diagrams, mass difference network and high UV absorbance (Figure 34, Figure 35, Figure 22).

The comparison of this top 100 correlating compounds to the metabolites enriched in the HCPP derived JDF00244 cluster revealed an overlay of 81 compounds in the EASY-HIT and HCPP cluster, which were enriched up to 5-fold in the JDF00244 cluster. 17 of them are registered in the NIAIDS database. These highly correlating compounds were investigated using MS/MS and LC-MS resulting in the identification of one candidate namely 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid in the Cistus PP D1 20% fractions. Its presence was verified by comparing the fragmentation pattern of the synthesized standard 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid and the isolated ion m/z, [M-H] 273 from the D1 20% fraction. This comparison is further pointing out the presence of at least one additional isomer in the D1 20% sample possessing the exact same mass. Investigation of the additional isomers in the Cistus D1 20% fraction is beyond the scope of this thesis but structural and biological characterization might shed further light on the anti-HIV 1 activity of Cistus extract.

The existence of 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid in Cistus D1 20% fraction was further verified by the comparison of the retention time of the standard to the D1 20% extract using LC-MS. To our knowledge, this was the first time that this substance was isolated and characterized in a Cistus extract. The fact that 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid, was already published to inhibit the HIV 1 integrase (Maurin et al., 2006) strongly underlines the efficiency of the experimental setup to identify single bioactive substances in very complex plant extracts.

All in all, we demonstrated a new and effective way to isolate and characterize bioactive and in particular anti-HIV-1 active metabolites by combining different methods like HCPP LOPAC® clustering, EASY-HIT, LC-MS and FT-ICR-MS. The correlation of the biological assays to the MS data resulted in a list of 17 top correlating and enriched compounds which exhibit a high probability to be responsible for the identified bioactivity. Further LC-MS and MS/MS analysis of these compounds lead to the isolation and characterization of 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid, a compound which is known to inhibit HIV-1 integrase (IN) and which could in this study be identified for the first time in *Cistus incanus*.

Of note, the D2 fractions HCPP LOPAC® clustering revealed many additional biological targets and cluster. In this chapter we focused on the characterization of the most significant one. Further studies might investigate the other identified biological targets and clusters. Therefore, the setup described in this chapter might help to identify the bioactive metabolites from crude plant extract.

4. Ocimum tenuiflorum (Tulsi)

4.1.   Introduction

Ocimum tenuiflorum (Tulsi), also known as *O. sanctum L*, from the family of Lamiaceae is a herbaceous plant with 1-3 cm long leafs and white to reddish blossoms arranged to false whorls (Chowdhury et al., 2017). It originates from north central India and grows in the tropics of the eastern world (Bast et al., 2014; Saharkhiz et al., 2015). Due to its various bioactive effects, Tulsi is known in Ayurvedic medicine since thousands of years (Das et al., 2015; Mondal et al., 2009; Pattanayak et al., 2010), to be active against colds, malaria, stomach disorders, inflammation, arthritis, rheumatism, pain, fever and diabetes (Banerjee et al., 1996; Board, 2005; Godhwani et al., 1987; Leung, 1996; Pattanayak et al., 2010). Latest research show, that Tulsi possesses additional activity against pathogens (Vasudevan P, 1999) and HIV-1 (Rege et al., 2010; Usha et al., 2003). Tulsi consist of 20% phenols and 7.5% flavonoids. Some bioactive compounds extracted from Tulsi are already described in literature like eugenol (Choudhury et al., 2014) luteolin, cirsilineol, phenylpropane glycosides (H. Nörr, 1992) and polyphenols (H. Nguyen, Lemberkovics, É., Tarr, K., Máthé Jr., I. and Petri, G., 1993; Shubha et al., 2017).

For the deeper investigation of the biological activities Tulsi holds, high content phenotypic profiling is recommended. HCPP is an *in vito*, cellular non-targeted biological screening predicting several biological targets the sample is showing activity for. Additionally, the mode of action, possible toxicities and side effects can be monitored in this cellular test (Bray et al., 2016; Caicedo et al., 2016; Fiehn, 2001; Giuliano et al., 1997; Kremb & Voolstra, 2017; Kurita et al., 2015; Nichols, 2007; Swinney & Anthony, 2011; D. W. Young et al., 2008). To characterize the metabolites responsible for the shown bioactivity, ultrahigh resolution FT-ICR-MS data need to be acquired from the samples of interest. The gained amount of complex data including hundreds and thousands of compounds need to be carefully filtered and aligned (Kanawati et al., 2017). The biological assay data can now be correlated to the MS data to characterize the potential bioactive metabolites in detail. A list of top correlating or enriched compounds can be generated and further MS/MS experiments for their identification can be performed.

4.2.   Objectives

Tulsi is known for its various bioactivities and therefore is used since thousands of years especially in the Ayurvedic medicine (Das et al., 2015; Mondal et al., 2009; Pattanayak et al., 2010). Research is going on characterizing the metabolic content of Tulsi and the metabolites responsible for the shown bioactivity. (Banerjee et al., 1996; Board, 2005; Chattopadhyay, 1994; Choudhury et al., 2014; Godhwani et al., 1987; Grayer et al., 1996; H. Nörr, 1992; Leung, 1996; Mandal et al., 1993; H. Nguyen, Lemberkovics, É., Tarr, K., Máthé Jr., I. and Petri, G., 1993; Pattanayak et al., 2010; Rege et al., 2010; Shubha et al., 2017; Singh et al., 1996; Usha et al., 2003; Vasudevan P, 1999). To evaluate the bioactivity of Tulsi in more detail, an extraction method was optimized generating a potent anti-HIV-1 active Tulsi extract.

This anti-HIV-1 active extract was fractionated (chapter 2) and tested in the effective *in vitro*, cellular non-targeted biological screening (HCPP). Clustering the cellular feature deviations from the samples with the

4. *Ocimum tenuiflorum* (Tulsi)

cellular feature deviation of known standards (LOPAC® library) the bioactivity potential of the different fractions can be revealed. Additionally, possible side effects and the distinct mode of action can be described in more detail (Y. Feng et al., 2009). Using Tanimoto Similarity calculations, possible substructures potential responsible for the shown bioactivity can be characterized.

To gain insights on the metabolic content, the clustering fractions of interest were compared to the other fractions and a comprehensive metabolic analysis using ultrahigh resolution FT-ICR-MS with distinct data analysis was performed (Kanawati et al., 2017). After data filtering and enrichment analysis, more knowledge about the top correlating or enriched metabolites was gained using NetCalc (Tziotis et al., 2011) and van Krevelen diagrams (Krevelen, 1950) indicating the molecular class of the identified compounds. The research aim of this study was to generate a bioactive Tulsi extract and fractions and to evaluate the broad bioactivity potential of the generated fractions. The correlation of the biological data with FT-ICR-MS data will lead to compound classes probably responsible for the shown bioactivity.

4.3. Results and Discussion

4.3.1. *Ocimum tenuiflorum* extracts

4.3.1.1. Extraction of anti-HIV-1 active metabolites from *Ocimum tenuiflorum*

Metabolomic studies consist in general of sample collection, metabolite extraction, analysis of the extracted metabolites and subsequently data analysis. The extraction of the metabolites in the beginning of the study is a crucial step, as any following conclusion will be built around the metabolites extracted from the available metabolome of the sample (Weckwerth, 2007).

In this study, we are interested in the bioactive metabolites from the plant *Ocimum tenuiflorum* (Tulsi). The samples were collected in Patiala (India 2013). The extraction solvent was based on water and MeOH as the bioactive plant metabolites are often polar (eg., polyphenols, flavonoids, glycoside flavonoids and tannins). To investigate which plant compartment possesses more bioactive compounds, different plant compartments, in particular the stem and the leaf of Tulsi, were used separately in the extraction. For that purpose, an extraction procedure was optimized (described in 5.1.2.1.2). To test the quality of the produced extracts the anti-HIV-1 activity (EASY-HIT described in 5.1.1.2) and cell vitality was tested (MTT test, described in 5.1.1.3 Viability test). The results of this experiment are summarized in Figure 43.

77

4. *Ocimum tenuiflorum* (Tulsi)
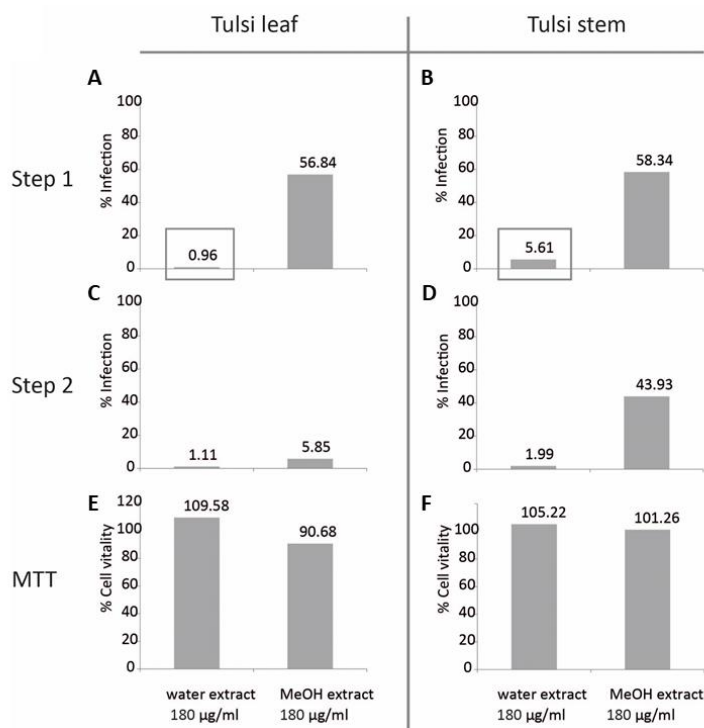
Figure 43: Comparison of anti-HIV-1 activity at the early-phase (step 1) and late-phase (step 2) of the virus replication cycle and cell viability of the leaf and the stem (water and MeOH) extracts of Tulsi at 180 µg/ml. A), B) EASY-HIT results for step 1 and C), D) for step 2 of the viral infection. A), C) extract of the Tulsi leaf, B), D) extract of Tulsi stem. The lowest % infection of 0.96% for Tulsi leaf step 1 and 5.61% for Tulsi stem step 1 are highlighted. In E) and F) the cell viability results are shown. The MTT tests show no reduction under 90% of the metabolic process. n = 1.

By comparing the extracs of Tulsi, the water extracts reveal a higher anti-HIV-1 activity than the MeOH extracts (Figure 43 A)-D)). In addition, a slightly increased anti-HIV-1 activity in the leaf extracts compared to the stem extracts is observable. Due to the small differences between leaf and stem and the high cell viability (Figure 43 E), F)) both water extracts were used in further experiments. The comparison of the leaf extracts (Figure 43 A), C)) to the stem extracts (Figure B, D) reveals an enriched anti-HIV-1 activity in the leaf extracts from Tulsi in both monitored HIV-1 infection steps. Furthermore, at 180 µg/ml the lowest % infection on the cells could be seen in case of the water extract of Tulsi leaf with 0.96% infection followed by the water extract of the stem with 5.61% infection Figure 43 A), B) highlighted). In this experimental setup the infection of 0.96% for the water leaf extract and 5.61% for the water stem extract of Tulsi showed still the lowest HIV-1 infection and was therefore used for further experiments.

Of note, the MeOH extract of the Tulsi leaf shows minor reduction of step 1 (56.84%) but a major reduction of the step 2 (5.85%) of HIV-1 infection suggesting an inhibition of the release mechanism of the virus from the cell. One possible cause of this finding might be compounds like Aglycone flavonoids which possess a known anti-HIV-1 activity in the late phase of the virus replication cycle and exhibit a better solubility in MeOH than in water (Kumar & Pandey, 2013). Nevertheless, further experiments need to be done to prove the involvement of this or other substance groups. All tested extracts of Tulsi show no toxic

effect on the cells at the tested concentration of 180 µg/ml, as the cell vitality was never lower than 90% Figure 43 E) and F). To investigate and compare the leaf and stem extracts in more detailed, a serial dilution from 180 µg/ml to 0.7 µg/ml of both extracts was performed (Figure 44). As the major inhibition takes place in the early phase of the replication cycle step 1 of the EASY-HIT was chosen for further experiments.
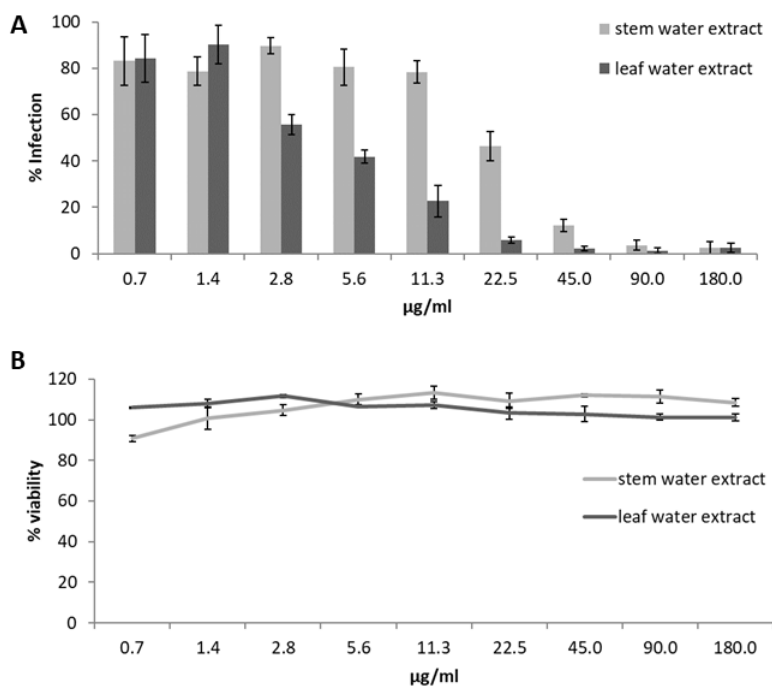
Figure 44: Overview of the EASY-HIT and cell viability of the serial dilution (from 180 µg/ml to 0.7 µg/ml) from Tulsi stem and leaf water extract. A) EASY-HIT test results of the dilution steps of Tulsi stem and leaf water extract (n = 3, mean ± SD). B) Cell viability test (MTT test) of the dilution steps of Tulsi stem and leaf water extract. The fractions were added in µg dry weight per ml cell culture medium.

Similar to the results in Figure 43 A)) and B), Figure 44 A) displays, that the leaf extract always results in more potent infection inhibition compared to the stem extract at the same concentration. At 11.3 µg/ml for example we could see a more than two-fold higher inhibition and at 22.8 µg/ml a more than four-fold higher inhibition for the leaf extract. The MTT test at the same time showed no toxicity for all the measured concentrations (Figure 44 B)). Based on these results, the water based Tulsi leaf extract was chosen as it exhibits the most potent anti-HIV-1 activity which suggests the most suitable extraction of anti-HIV-1 active metabolites and ensure high cell vitality at the same time.

4. *Ocimum tenuiflorum* (Tulsi)

4.3.1.2. EC$_{50}$ of *Ocimum tenuiflorum* and polyphenol enriched *Ocimum tenuiflorum*

After identifying the water extract of Tulsi leafs (from now in named Tulsi raw extract) to give the most anti-HIV-1 active metabolites the EC$_{50}$ of this extract was calculated performing a serial dilution from 0% infection to 100% infection. The EC$_{50}$ was calculated according to 5.1.1.2.
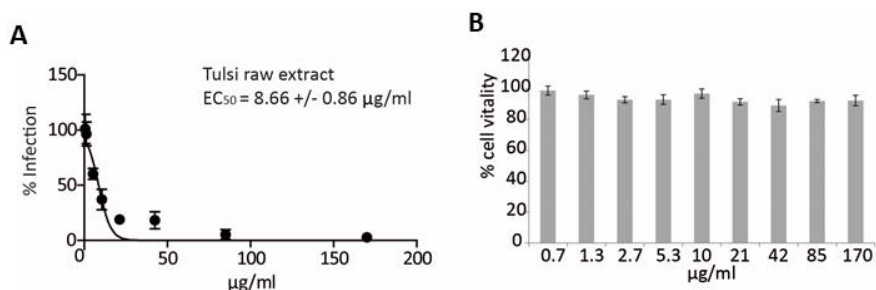
Figure 45: Overview of the EC$_{50}$ and cell viability test results for the Tulsi raw extract (n(experiments)=3, mean±SD). The concentration of the extracts is shown in µg dry weight per ml cell culture medium. A) EC$_{50}$ of Tulsi raw extract = 8.66 ± 0.86 µg/ml. B) Cell viability test (MTT test) for all extract concentrations.

The calculated EC$_{50}$ of 8.66 ± 0.86 µg/ml (Figure 45 A)) was similar to the EC$_{50}$ of the Cistus raw extract 7.79 ± 0.15 µg/ml (Table 1). A polyphenol enrichment of Tulsi raw extract was carried out, to investigate their effect on anti-HIV-1 activity. Therefore, a polyphenol-enriched and polyphenol-depleted fraction of the Tulsi raw extract was produced according to the description in 5.1.3 (Figure 46).
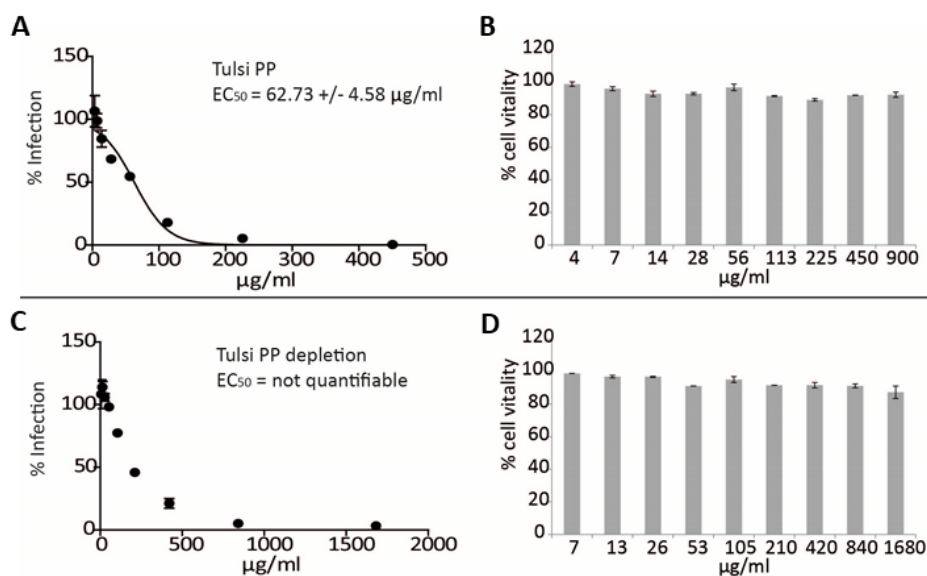
Figure 46: Overview of the EASY-HIT and cell viability test of the polyphenol-enriched fraction of Tulsi (Tulsi PP) and Tulsi PP depleted (n(experiments)=3, mean±SD). The concentration of the extracts is shown in µg dry weight per ml cell culture medium. A), C) EASY-HIT test results of Tulsi PP and Tulsi PP depleted (EC$_{50}$ of Tulsi PP = 62.73

± 4.58 μg/ml and EC$_{50}$ of Tulsi PP depletion = not quantifiable). B) and D) cell viability test (MTT test) of Tulsi PP and Tulsi PP depleted.

The polyphenol-enriched fraction of Tulsi (Tulsi PP) showed an EC$_{50}$ of 62.73 ± 4.58 μg/ml (Figure 46 A)) and the EC$_{50}$ of the polyphenol-depleted fraction of Tulsi could not be quantified due to a non-sigmoidal behaviour of the dilution curve (Figure 46 C)). Nevertheless, the extract concentration, which leads to a 50% infection rate can be roughly estimated at 200 μg/ml. This estimated EC$_{50}$ is much higher than for the Tulsi PP fraction leading to the conclusion, that the active compounds are reduced in the PP depleted fraction. Interestingly, the polyphenol enrichment procedure resulted in no improvement of the anti-HIV-1 activity of Tulsi. At least two causes for that could be possible: (I) that another compound class, beside polyphenols, is responsible for the activity. (II) the compounds responsible for the activity are not stable enough to overcome the applied polyphenol enrichment process. The most anti-HIV-1 active and bioactive extract of Tulsi was now identified, further experiments need to be done to reveal the mode of action of this inhibition. Therefore, the Tulsi raw extract was fractionated in D1 and D2 fractions (see chapter 2) and tested in the HCPP screening to analyze the general bioactivity of the extract and fractions.

4.3.2. Overview of HCPP results without clustering

The HCPP screening was performed like described in 3.3.1 according to the publication of Kremb & Voolstra, 2017. The used method is described in 5.1.6. In Figure 47 the HCPP results of the raw extract and D1 fractions from Tulsi are shown. The complete list of the features is provided in the supplement (Table 15) and for better visualization the measured features are summed up to major underlying phenotypic attributes, e.g. ER, lysosome or membrane.
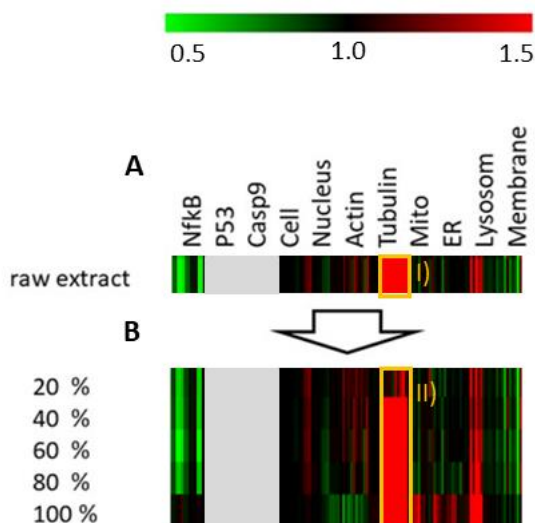
Figure 47: HCPP results of the raw extract A) and D1 fractions B) from Tulsi are shown. The individual extracts and corresponding fractions (20% - 100% MeOH) are represented on the y-axis. On the x-axis 103 features are shown summed up to 9 cellular core features for better visualization. The mean of n = 4 measurements normalized to the corresponding, outlier tested, control feature on the same plate are shown. Red indicate a positive and green a negative deviation from the mean of untreated control cells with a value = 1. Raw extract and D1 fractions of Tulsi (p53 and Cp9 are missing due to experimental problems, the area is marked in gray). I) up regulation of tubulin core feature in the Tulsi raw extract. II) tubulin core feature regulation in the D1.

The raw extract and D1 fractions show a strong positive effect on the core feature tubulin (Figure 47, I) and II)). The finding that all D1 fraction besides the 20% fraction show a strong effect on the tubulin features, suggests an at least partly different metabolic composition present in the 20% fraction compared to 40% until 100%. Additionally, a down regulation of the NF-κB features (fractions 20 to 80%) and an up regulation of the lysosomes features (strongest effect in 100% fraction) is observable.

To further investigate these patterns, the D2 fractions were analyzed. After the D2 (according to 5.2.5) fractionation, the effects on the features are more distinct and better traceable due to less possibility for metabolite aggregation (B. Y. Feng & Shoichet, 2006) and antagonist and agonist mixtures in one sample (Wagenaar, 2008). The HCPP results are displayed in Figure 48.
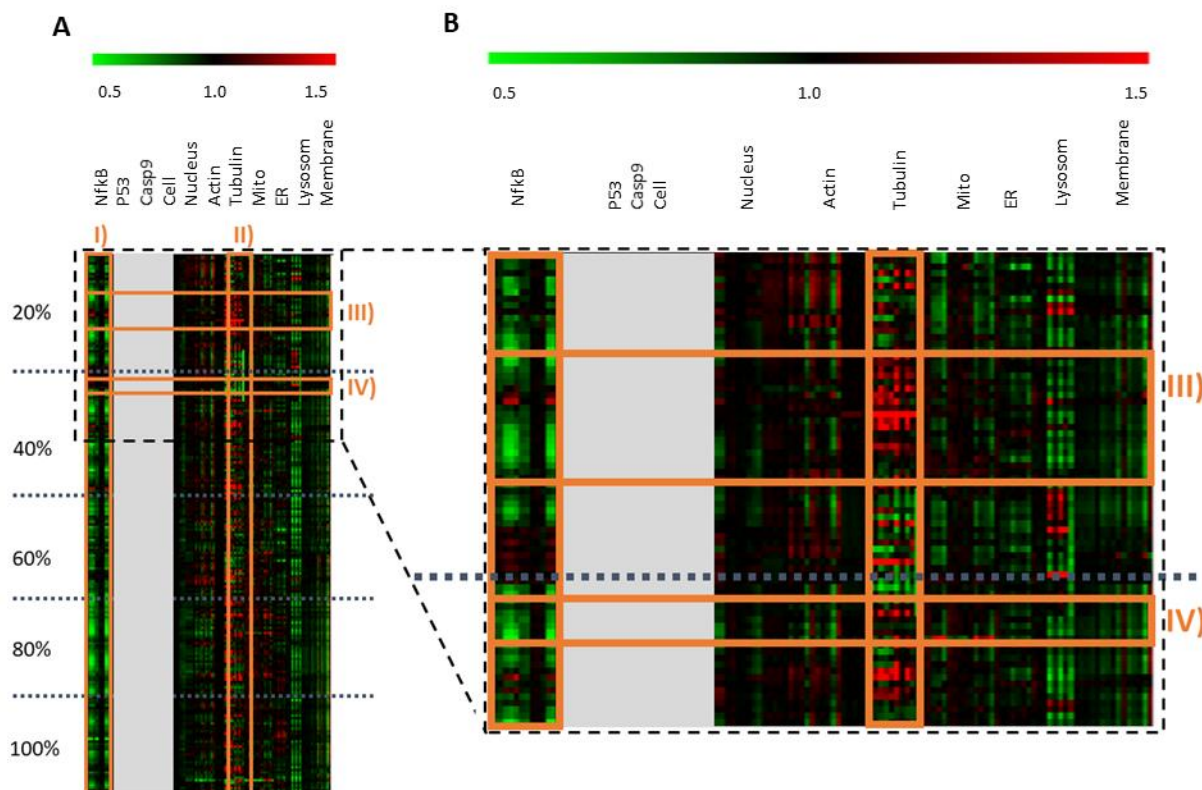
Figure 48: Overview of the HCPP results of the D2 fractions of Tulsi. The mean of n = 4 measurements normalized to the corresponding, outlier tested, control feature on the same plate are shown. Red indicate a positive and green a negative deviation from the mean of untreated control cells with a value = 1. A) from Top down the D2 fractions from 20% to 100% MeOH (D1) are shown in a row and separated by dotted lines. Due to a problem in the experimental part p53, Caspase 9 (Casp9) and the Cell features could not be measured and are therefore left white in the figure. I) negative deviation NF-κB. II) mixed positive and negative deviation on the tubulin features for all D2. III) some fractions from D2 20% showing a distinct upregulation of the tubulin marker. IV) fractions from D2 40% showing a rare up regulation of the lysosome features. B) magnification.

Similar to the HCPP data of the Cistus D2 fractions the D2 fractionation of Tulsi extract reveals several interesting patterns (Figure 48). A common down regulation on NF-κB was visible for all the D2 fractions shown in Figure 48 I). The tubulin features show a mixture of up and down regulation in the whole Tulsi dataset, Figure 48 II). This feature seems to be sensitive to different extracts and fractions due to the different regulation of this feature in Cistus, Cistus PP and Tulsi (chapter 3.3.1, Figure 20). Moreover, the generated HCPP dataset is rich in deviations to untreated Hela cells leading to consecutive fractions with similar feature deviation. One row is based on the upregulation of the tubulin feature Figure 48 III). The other row contains D2 40% fractions which exhibit a strong upregulation of the lysosome feature, Figure 48 IV).

4.3.3. Overview of the HCPP results with LOPAC® clustering

The clustering of the unknown samples with LOPAC® was performed like previously described for Cistus (chapter 3.3.2). The "LOPAC®1280 Library of Pharmacologically Active Compounds" was used as a standard set. They were analyzed according to the procedure used for the samples (see 5.1.6) and the gained data correlated. Like previously described for Cistus the raw extract and D1 fractions of Tulsi are very rich in different bioactive compounds, too.

Due to that no meaningful biological target could be clustered for these extracts even when using different correlations methods like Spearman Rank, Pearson and Kandell`s Tau (Crisman et al., 2007)(Reisen et al., 2013). For more precise information the D2 fractions of Tulsi were correlated to the LOPAC® standards to investigate differences and commonalties. Several clusters appeared in the D2 fractions of the Tulsi extract. The clusters with a clear mode of action or structural similarity in addition to a valid support tree are shown in Figure 49. In Figure 49 an overview of the cluster derived from the D2 20% fractions of Tulsi are shown.
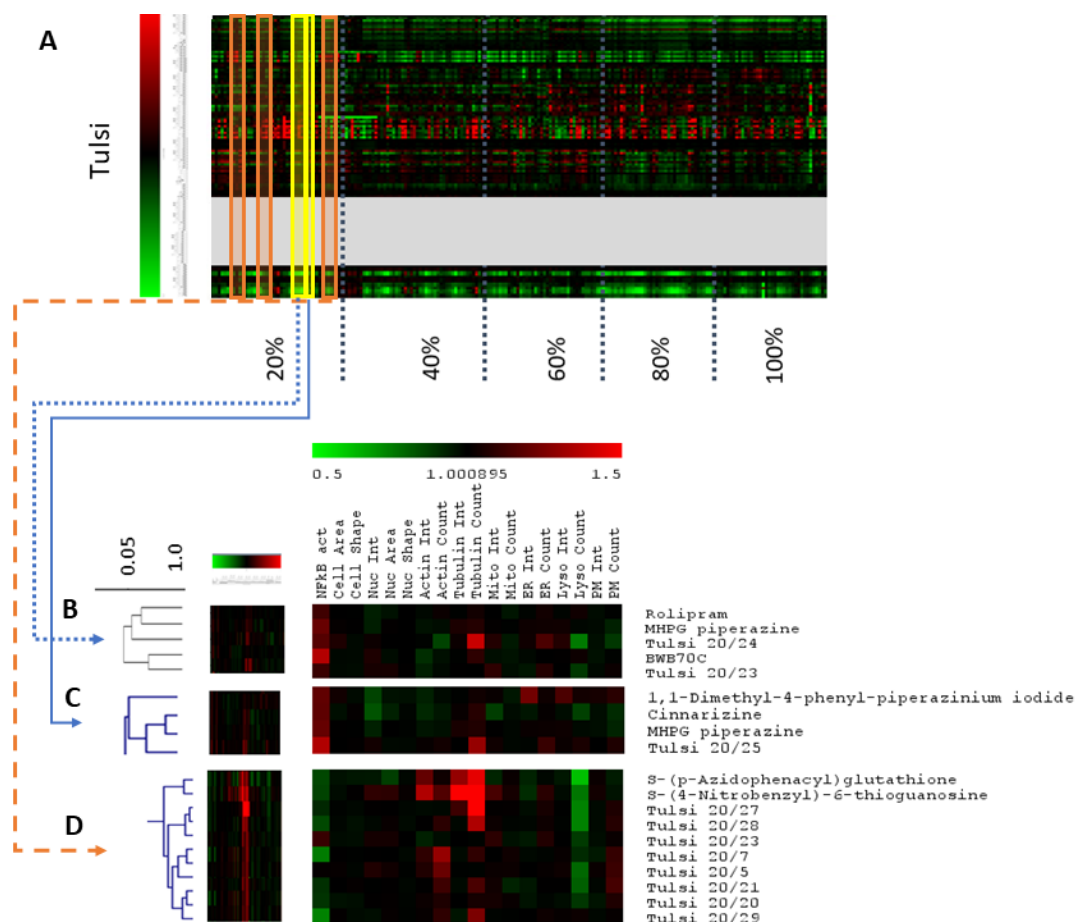
Figure 49: Hierarchical clustering for D2 Tulsi fractions focusing on D2 20%. A) Overview of the hierarchical clustering of cytolological profiles from Tulsi fractions (separated by dotted lines into the previous 20% - 100% MeOH D2 fractions). B), C) and D) Identified clusters of cyctological profiles and overview of the 18 core markers (n(measurements)=4, outlier tested, values normalized to control, red corresponds to up and green to

downregulation compared to control). B) Spearman Rank cluster of 23 and 24 (D2 20%) with Rolipram, MHPG piperazine and BWB70C. C) Pearson cluster of fraction 25 (D2 20%) with 1,1-Dimethyl-4 phenyl-piperazinium iodide, Cinnarizine and MHPG piperazine. D) Pearson cluster of fractions 5,7,20,21,23 and 27-29 (D2 20%) with S-(p-Azidophenacyl) glutathione and S-(4-Nitrobenzyl)-6-thioguanosine. The corresponding positive support tree for B), C) and D) are shown in the supplementary material (Figure 84 until Figure 86).

In Figure 49 B) we could see, that fractions 23 and 24 cluster together with the selective cAMP-dependent Type IV phosphodiesterase cAMP-dependent inhibitor (Rolipram), a norepinephrine metabolite (MHPG piperazine) and a selective inhibitor of 5-lipoxygenase (BWB70C). Although these standards show a common upregulation of the NF-κB and a slight upregulation of the tubulin features no common mode of action could be identified (Crisman et al., 2007).

Besides that, a second cluster (Figure 49 C)) consisting of fraction 25, a nicotinic acetylcholine receptor agonist (1,1-Dimethyl-4-phenyl-piperazinium iodide), a Ca^{2+} channel blocker central and peripheral vasodilator Cinnarizine, and the norepinephrine metabolite (MHPG piperazine) were identified. Like the cluster around fraction 23 and 24, the NF-κB and tubulin features are slightly upregulated. Investigating the LOPAC® structures of this cluster, a common subunit could be identified by calculating the Tanimoto similarity (Figure 50) (D. W. Young et al., 2008).
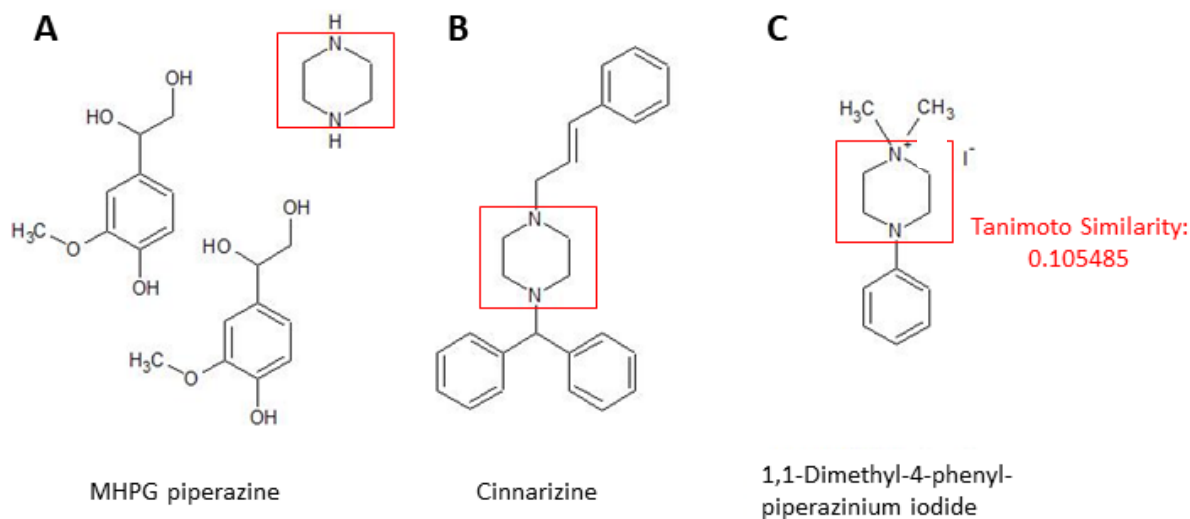
Figure 50: Overview of the molecular structure of A) MHPG piperazine, B) Cinnarizine and C) 1,1-Dimethyl-4-phenyl-piperazinium iodide including the calculated Tanimoto Similarity between A), B) and C) of 0.105485.

The identified subunit piperazine is known to be part of many bioactive compounds showing anticancer, antifungal, antibacterial and antimalarial properties (Guo, Tong, & Li, 2004; Mohammed Al-Ghorbani, 2015). Additionally piperazine derivative drugs are available acting as anti-histamines, anti-depressants and anti-HIV agent indinavir (Guo et al., 2004). Based on the HCPP results it might be worth to investigate the piperazine subunit effect on tubulins and NF-κB signaling. Further experiments like the HCPP results of pure piperazine are necessary to assign the shown feature deviation to the piperazine subunit or other functional groups of the shown standards.

4. *Ocimum tenuiflorum* (Tulsi)

In Figure 49 D) another cluster is shown, including a set of fractions namely 5,7,20,21,23 and 27-29 and the two LOPAC® standards namely a glyoxalase and glutathione S-transferase inhibitor (S-(p-Azidophenacyl) glutathione) and a potent adenosine transport inhibitor (S-(4-Nitrobenzyl)-6-thioguanosine). This cluster is characterized by a common NF-κB and lysosome features downregulation and actin and tubulin feature upregulation. Calculating the Tanimoto similarity a common subunit could be characterized.
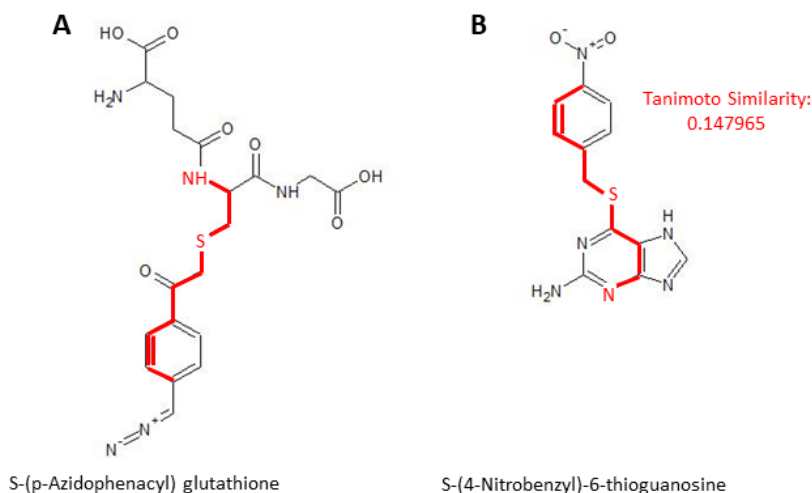
Figure 51: Overview of the molecular structure of A) S-(p-Azidophenacyl) glutathione and B) S-(4-Nitrobenzyl)-6-thioguanosine including the calculated Tanimoto similarity between A) and B) of 0.147965.

Unfortunately, the identified substructure (3-propan-1-amine) is no single compound with available bioactivity data. Nevertheless, it might be that the sulfur connection of two alkyl chains together with a nitrogen functional group could be in total or in parts responsible for the shown core feature regulations. The fact, that a broad range of fractions cluster with these standards leading to the conclusion, that isomers with a common subunit, like the described one, but additional different chemical properties (due to e.g. functional groups) could be responsible for the exhibit features deviation.

4.3.4. EASY-HIT

4.3.4.1. D1 anti-HIV-1 activity distribution and EC$_{50}$ calculation

Although no HCPP clustering of the raw extract, D1 and D2 fractions with anti-viral standards could be identified, the plant was described to possess anti-HIV potential (Rege et al., 2010; Usha et al., 2003). The pre-experiments in this thesis also showed anti-HIV-1 activity (Figure 45, Figure 46). This discrepancy might be based on the different mode of action of the used anti-HIV standards to the anti-HIV biomolecules avaiable in the plant extract. To further investigate the anti-HIV-1 potential of Tulsi extract, the produced D1 and D2 fractions were tested for their anti-HIV-1 activity employing the EASY-HIT test. In Table 10 the EC$_{50}$ results from Tulsi are summarized.

Table 10: EC$_{50}$ distribution of Tulsi of the raw extract and corresponding SPE fractions, (nq = not quantifiable EC$_{50}$ due to too less inhibition). A cell viability test (MTT test) was carried out for every raw extract and fraction and no reduction of the metabolic process under 80% appeared (supplement, Figure 87. The mean and SD of triplicate measurements are shown. The samples were added in µg dry weight per ml cell culture medium.

	EC$_{50}$ [µg/ml]
Raw extract	8.66 ± 0.86
D1 20%	12.74 ± 0.87
D1 40%	7.53 ± 0.64
D1 60%	nq
D1 80%	nq
D1 100%	nq

The EASY-HIT measurements show that anti-HIV-1 activity of the raw extract (EC$_{50}$ 8.66 ± 0.86 µg/ml) was split by the D1 fractionation in a less potent (12.74 ± 0.87 µg/ml (20% MeOH)) and a more potent (7.53 ± 0.64 µg/ml (40% MeOH)) fraction. Compared to the Cistus fractionation, the most potent anti-HIV-1 active compounds are shifted from 20% MeOH to 40% MeOH suggesting another group of compounds being responsible for the main activity. It was not possible to calculate the EC$_{50}$ values for the D1 fraction 60%, 80% and 100% as the inhibitory effect was too low or totally missing.

4.3.4.2. D2 anti-HIV-1 activity distribution and UV spectra

As the D1 fractions 20% and 40% MeOH showed a high HIV-1 inhibition, the corresponding D2 fraction were investigated for their anti-HIV1 activity. The results are summed up in Figure 52.

4. *Ocimum tenuiflorum* (Tulsi)
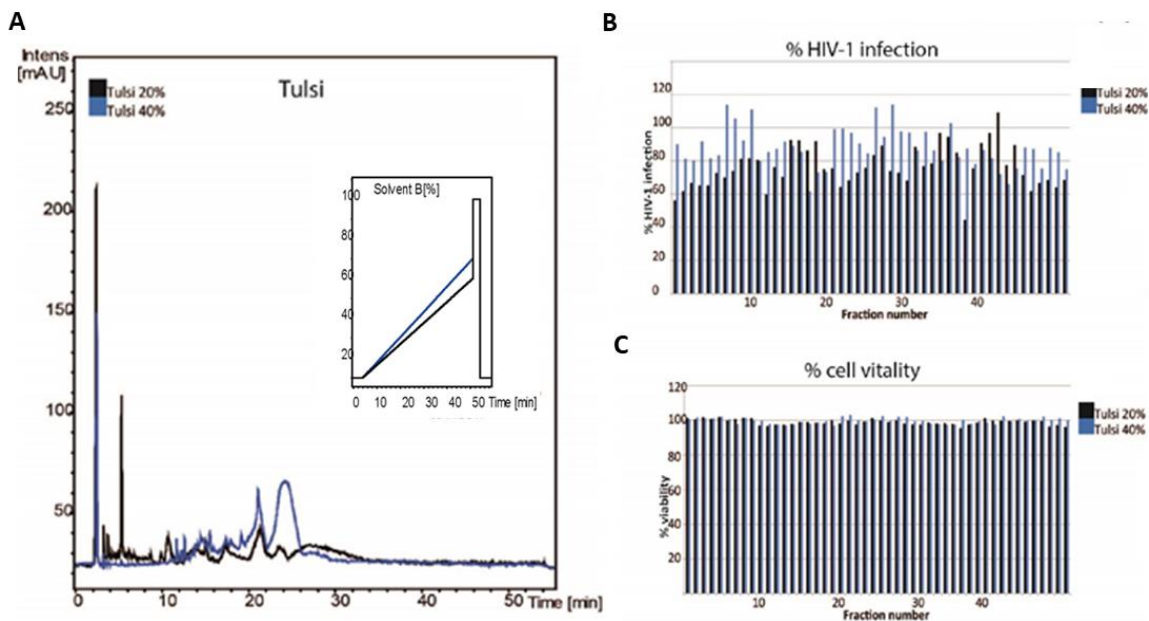
Figure 52: Overview of the results for D2 20% MeOH (black) and 40% MeOH (blue) fractions from Tulsi. A) Extracted wavelength of 280 nm of the D2 fractionation with additionally embedded chromatographic conditions and the retention time on the x-axis. B) anti-HIV-1 activity distribution, %HIV-1 infection on the y-axis and the D2 fraction number on the x-axis in the first step of the EASY-HIT, n = 1. C) overview of the % cell vitality (y-axis) of the tested D2 fractions (x-axis), n = 1.

Figure 52 A) contains the HPLC-PDA results for the D2 fractions 20% (black) and 40% MeOH (blue) at 280 nm together with the embedded chromatographic conditions. In the 40% D2 fractions a higher UV activity compared to the 20% appeared. The wavelength from 260 – 390 nm showed a similar distribution (Figure 63). The wavelength and 280 nm was chosen as model wave length and no UV activity was found in the D2 60%-100% fractions (Figure 64). The D2 fractions of D1 20% and 40% were tested for their anti-HIV-1 activity and the results are shown in Figure 52 B). Interestingly, the D2 fractions of 20% and 40% MeOH exhibit a very similar distribution of inhibition. No bell-shaped activity distribution, like it appeared for Cistus and Cistus PP, could be detected (chapter 3.3.3.2, Figure 26). In contrast a uniform distribution of HIV-1 inhibition over the entire collected fractions was measured. The similar effect of almost all fractions might be based on very different physicochemical properties of the anti-HIV-1 active metabolites leading to the uniform elution from the column. As FC 1 and 48 still show HIV-1 inhibition it cannot be excluded, that bioactive metabolites got lost during the fractionation process eluting prior or after the collected fractions. Nevertheless, no toxicity could be observed with the D2 fractions of 20% and 40% MeOH (Figure 52 C).

All in all, the XBridgeTM Shield RP C 18 did not show a good active, non-active separation for Tulsi extract, although reliable results could be achieved using this column for Cistus and Cistus PP. To display the metabolic distribution, the simulated total ion current (FT-ICR-MS measurements) for the Tulsi D2 20% and 40% fractionation were investigated (Figure 53).
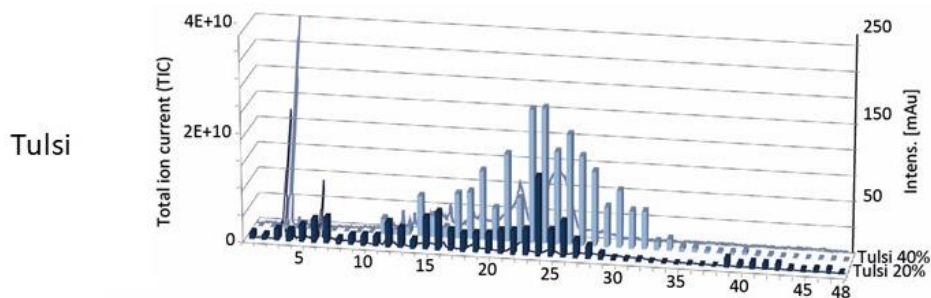
Figure 53: Comparison of the TIC and UV data from Tulsi D2 20% and 40%. The D2 fractions are represented on the x-axis, the D1 fractionation on the z-axis and the total Ion current (summed up intensity of the measured m/z per fraction) is represented on the y-axis (left), the [mAu] from the embedded UV intensity at 280 nm on the right y-axis.

Figure 53 shows, that the metabolites are similarly distributed like it appeared for Cistus and Cistus PP (chapter 2.3.5, Figure 17 and Figure 18). A major difference between the D2 fractions of Tulsi and Cistus are the TIC and UV intensities which are higher in the 40% fractions compared to the 20%. These results fit to the EASY-HIT Tulsi D1 measurements, which exhibit a lower EC_{50} in the 40% than in the 20% D1 fraction. Of note, the UV absorbance is indicating potential interesting fractions which might possess an increased amount of metabolites. Similar to Cistus and Cistus PP, these fractions might serve as starting points to identify the bioactive molecules using mass spectrometry.

4.3.5. Correlation of mass spectrometry data with HCPP data

4.3.5.1. FT-ICR-MS measurements of Tulsi

To further characterize bioactive compounds, it was crucial to monitor the chemical space of the active and non-active fractions. Therefore, the D2 Tulsi fractions were measured using FT-ICR-MS like it was described in chapter 3.3.4.1.

A widespread problem of complex plant extracts is the ion suppression when measured with the FT-ICR-MS (Araujo et al., 2016; Buhrman et al., 1996; Taylor, 2005). Available mass signals get lost when they can not be ionized due to too high competition in the ESI source (Ghaste et al., 2016). Compared to the raw extract our orthogonal fractionation strategy leads to a broader distribution of the chemical structures and thereby in a deeper investigation of the chemical space. This also improves the chance to identify the metabolites responsible for the displayed bioactivity.

4.3.5.1.1.   Visualization of the ion suppression and polyphenol enrichment using van Krevelen diagrams

The ion suppression is shown in Figure 54, by comparing the chemical space (possible molecule composition calculated with the in-house software NetCalc (Tziotis et al., 2011)) in the raw extract to the chemical space in the corresponding D2 fractions.
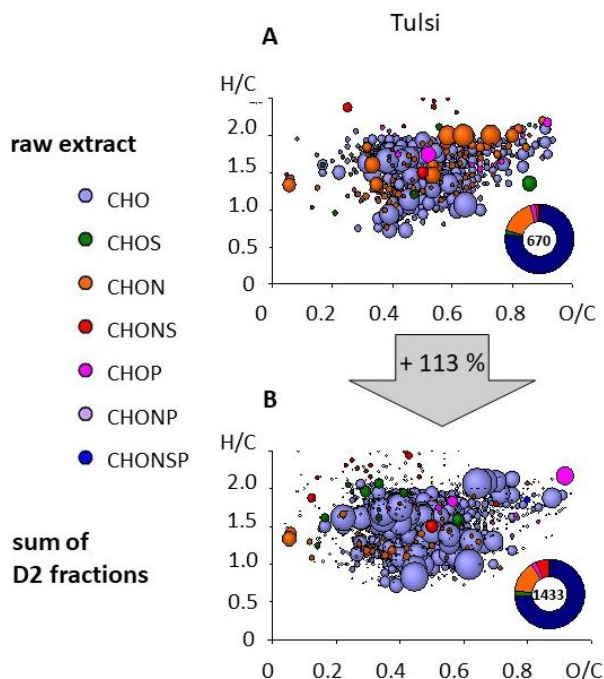
Figure 54: Visualization of the ion suppression effect from raw extract to the sum of D2 fractions. The increased number of calculated formulas (NetCalc, max 0.1 ppm error) is displayed in %. Van Krevelen diagrams with H/C ratio on the y-axis and O/C ratio on the x-axis are shown. The size of the bubbles represents the average intensity. The number of calculated formulas and the van Krevelen diagram for the raw extracts is displayed in A) for Tulsi raw extract. The corresponding sum of calculated formulas for the D2 fractions is displayed in B). The gray arrows represent the additional possible NetCalc calculations in %.

The raw extracts and the D2 fractions of Tulsi cover a wide chemical space. The positive effect of the fractionation process on the number of calculated NetCalc formulas is presented in Figure 54. The D2 fractionation leads to additional 763 calculated formulas in Tulsi (113% increase compared to the raw extract). This rise in the number of identified formulas clearly show the benefit of the extract fractionation when investigating the plant metabolome (Ganzera & Sturm, 2018). A similar tendency of + 24% for Cistus and + 80% for Cistus PP was visible (Figure 31).

4.3.5.2.   Correlation of FT-ICR-MS and HCPP data

For a better identification and characterisation of the bioactive compounds the FT-ICR-MS data and HCPP data of the D2 fractions were correlated. Therefore, the two gained clusters (see Figure 49 B) and C)) were

investigated for their unique and enriched metabolites. The significant metabolic difference between the D2 20% fractions 23, 24 and 25 leading to two different significant clusters with LOPAC® standards, cluster 1 (fractions 23 and 24) and cluster 2 (fraction 25), was remarkable and therefore were deeper investigated. In Figure 55, the HCPP results of cluster 1 and 2 are compared.
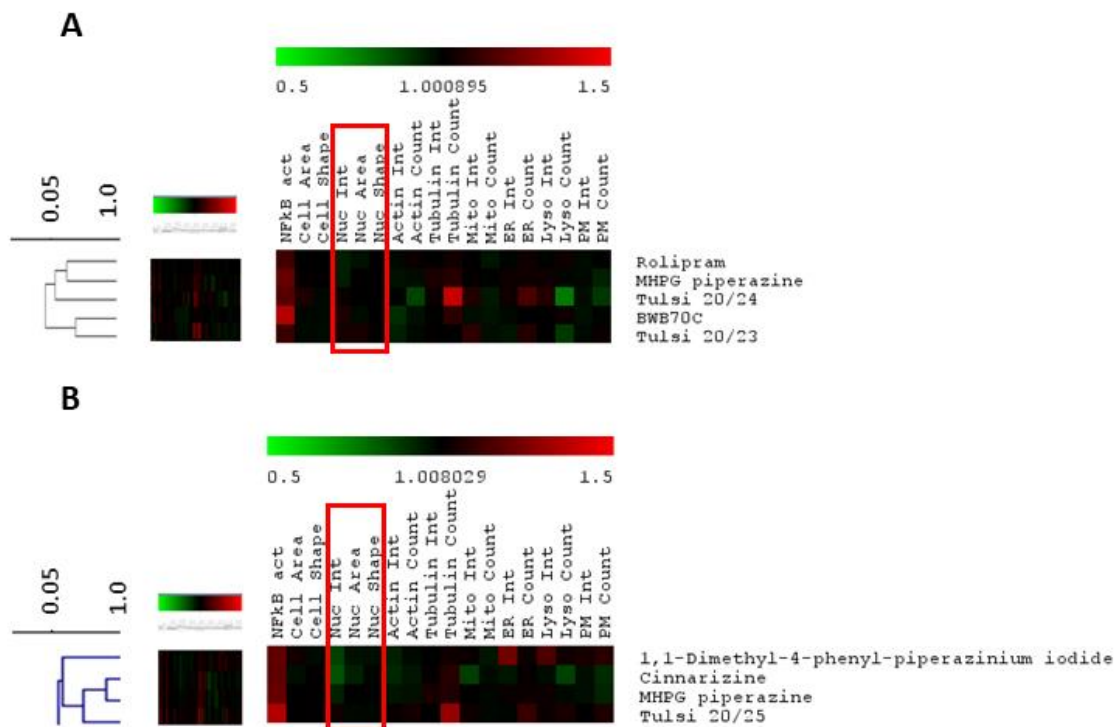
Figure 55: Hierarchical clustering of cytological profiles from cluster 1 and cluster 2. Full cytological profiles were used for the clustering (left) and 20 core markers are chosen for easier visualization of the cellular effects. The mean of n = 4 measurements normalized to the corresponding, outlier tested, control feature on the same plate are shown. Red indicate a positive and green a negative deviation from the mean of untreated control cells with a value = 1. A) Spearman Rank cluster of fractions 23 and 24 D2 20% and the LOPAC® library. B) Pearson cluster of fractions 25 D2 20% and the LOPAC® library.

In Figure 55 A) and B) a common up regulation for the NF-κB and tubulin features as well as a down regulation for the actin and lysosome features were observable. Whereas cluster 1 (Figure 55 A)) was associated with a slightly upregulation in the nucleus features, cluster 2 (Figure 55 B) exhibits a down regulation of these features (highlighted). To investigate the metabolic differences of this effect, the metabolic content, which was revealed by the FT-ICR-MS, like descried in 3.3.4.1, of cluster 1 to cluster 2 were compared. The results were displayed in van Krevelen diagrams in Figure 56.
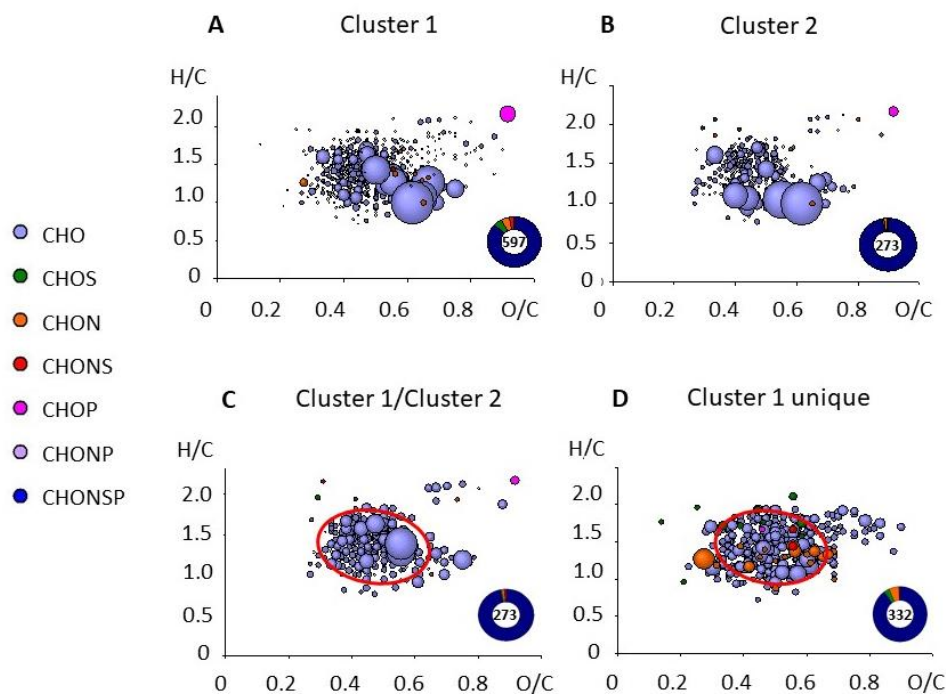
Figure 56: Visualization of the metabolite enrichment and unique metabolites comparing cluster 1 and 2 using van Krevelen diagrams with an H/C ratio on the y-axis and O/C ratio on the x-axis. Metabolites with an error of < 0.1 ppm from experimental to NetCalc calculated m/z are shown. A) total metabolic content of cluster 1. B) total metabolic content of cluster 2. C) enriched metabolites in cluster 1. D) unique metabolites in cluster 1. All molecular masses with calculated chemical composition are presented.

As expected, cluster 1 (Figure 56 A) consists of approximately the double amount of metabolites (597 metabolites) compared to cluster 2 (Figure 56 B)) (273 metabolites) due to the double amount of fractions involved. By determining the enriched and unique metabolites e.g. by the quotient of the metabolic content of cluster 1 and cluster 2, the metabolic difference between the clusters was investigated. Interestingly, a clear enrichment up to 71-fold, of metabolites in a specific van Krevelen area with O/C values from 0.3 to 0.7 and H/C values from 1 – 1.7 was identified (Figure 56 C). Next to the enriched metabolites, additional 332 unique metabolites are present in cluster 1 showing the same O/C and H/C values Figure 56 D). This area is known to contain carboxyl-rich alicyclic molecules, nucleic acids (Hertkorn et al., 2008) and amino acids (Roullier-Gall, Witting, Gougeon, & Schmitt-Kopplin, 2014).

The observation of a 332 metabolites difference between the two clusters consisting of 3 fractions in a row indicate, the remaining complexity of the fractions. Therefore, it is necessary to have a deeper look in the metabolic content of the two cluster revealing differences and communalities. Investigating the metabolites with the highest intensity in cluster 1 several metabolites could be annotated using MassTRIX (Table 11).

Table 11: MassTRIX3 results for the m/z with the highest intensity in cluster 1.

Average intensity	Assigned Molecular Formula	Mass	Database ID	Name	Metabolic class
3.90E+08	$C_{14}H_{14}O_9$	325.0565	C02750 (KEGG)	sinapoyltartronate	sinapoyltartronic acid
3.29E+08	$C_{20}H_{20}O_{10}$	419.0984	LMPK12113361-64, LMPK12113381-83 (Lipid Maps)	5,3',5'-Trihydroxy-3,6,7,8,4'-pentamethoxyflavone	flavones and flavonols
2.82E+08	$C_{16}H_{22}O_8$	341.1242	C00761 (KEGG)	coniferin	glucoside of coniferyl alcohol
1.75E+08	$C_{13}H_{14}O_8$	297.0616	HMDB10324 (HMDB)	benzoyl glucuronid	benzoic acid

Sinapoyltartronate was annotated as putative compound showing an intensity of 3.90E+08 which was the third highest m/z intensity in cluster 1. Two compounds with the highest intensity could not be annotated which is not uncommon in metabolomic studies. Additionally, a Flavones, glycolyzed coniferyl alcohol and benzoic acid could be annotated. All putative compounds possess very high intensities over 1.75 E+08. Investigating the m/z with the highest intensity in cluster 2 we got the following results (Table 12).

Table 12: MassTRIX3 results for the m/z with the highest intensity in cluster 2.

Average intensity	Assigned Molecular Formula	Mass	Database ID	Name	Metabolic class
1.19E+09	$C_{14}H_{14}O_9$	325.0565	325.0565	sinapoyltartronate	sinapoyltartronic acid
2.73E+08	$C_{20}H_{20}O_{11}$	435.0933	C10067 (KEGG)	irisxanthone	xanthen derivative
1.20E+08	$C_{16}H_{16}O_7$	319.0823	LMPK12020133 – 35 (Lipid Maps)	gallocatechine	flavans, flavanols

Again, Sinapoyltartronate with an intensity of 1.19E+09 appeared as putative annotation, it is the compound with the highest intensity in cluster 2. Moreover, a xanthen derivative and flavon appeared among the metabolites with the highest intensity.

In general, both clusters possess a mass, which was annotated to sinapoyltartronate. This mass might be at least partially the reason for the common core feature regulations (up regulation for the NF-κB and tubulin features as well as a down regulation for the actin and lysosome features). Besides that, the annotated compounds with the highest intensities were more diverse between the clusters, which was already suggested by the different trends in the nucleus feature. Further fractionation would be necessary to characterize and identify the compound responsible for the shown bioactivity. Nevertheless, the fractionation method and the identified clusters possess a high potential to identify the compound with bioactivity potential in further studies.

4. *Ocimum tenuiflorum* (Tulsi)

4.3.6. Summary and conclusion

It was possible to generate an anti-HIV-1 active extract out of Tulsi leaf with an $EC_{50} = 8.66 \pm 0.86$ [µg/ml]. This extract is non-toxic even in high concentrations and five times more active compared to the extracts of Tulsi stem and in the same range as the published EC_{50} of Cistus = 7.79 ± 0.15 [µg/ml]. The raw extract, D1 and D2 fractions of Tulsi leaf were analyzed using HCPP, EASY-HIT assays and ultrahigh resolution mass spectrometry to identify the metabolites, which were responsible for the previously shown bioactivity. The correlation of the biological assays and analytical measurement results revealed a couple of significant clusters. One significant cluster including fraction 25 and three LOPAC® standards revealed the famous substructure piperazine as bioactive common subunit using Tanimoto Similarity and showing further cellular regulations important for piperazine derivative drugs in general.

A HCPP screening of all extracts was performed and clustered with the LOPAC® standards library to investigate the general bioactivity. Thus, the raw extract and even the D1 fractions were too complex to see distinct bioactivity (Schulze et al., 2013) (B. Y. Feng & Shoichet, 2006) like it was the case for Cistus and Cistus PP. Subsequently, further fractionation was necessary to reveal the biological target and the modulating metabolite or metabolic class. The D2 20% fraction 25 clustered together with MHPG piperazine, Cinnarizine and 1,1-Dimethyl-4-phenyl-piperazinium iodide using Pearson clustering. The lowest common denominator was piperazine. A substructure that is known to be common in bioactive compounds acting as anticancer, antifungal, antibacterial, antimalarial, anti-histamines, anti-depressants and anti-HIV agent. The anti-HIV-1 activity of fraction 25 was tested using the EASY-HIT but no significant inhibition could be seen.

Interestingly the two previous fractions, namely fraction 23 and 24, clustered together with Rolipram, MHPG piperazine and BWB70C using Spearman Rank clustering. This significant cluster appeared using a different clustering algorithm, showing again MHPG piperazine but no common structural subunit or mode of action could be extracted in this case. The only difference in the feature regulation was a slightly nucleus upregulation for the 23, 24 cluster and a down regulation for the 25 cluster.

The proximity of the both clusters is unique in the Tulsi dataset and the metabolic similarities and differences were investigated using FT-ICR-MS. The cluster containing fraction 23 and 24 shows twice as much metabolites and an enrichment in the van Krevelen diagram up to 71-fold in the specific area with O/C values from 0.3 to 0.7 and H/C values from 1 – 1.7. This area is characteristic for carboxyl-rich alicyclic molecules, nucleic acids and amino acids. The highest intensities in both clusters were annotated using MassTRIX3 and showed Sinapoyltartronate as putative common compound. The other annotated compounds differ in structure and mode of action and might be the reason for the different nucleus feature deviation. Even an annotation for the highest intensities was possible, they might not be responsible for the exhibit effect as metabolites with a lower intensity could be a potent bioactive compounds. As cluster 1 contains 597 metabolites and cluster 2 contain 273, further fractions would be necessary to narrow down the metabolites responsible for the shown effect.

All in all, we found a new and potent anti-HIV-1 active plant extract from Tulsi leaf. It was possible to extract several bioactive HCPP clusters including one clustering with three LOPAC® standards containing the well-known bioactive piperazine subunit as lowest common denominator. Because of the remaining complexity, a further fractionation of the clustering fractions would be necessary to narrow down the responsible metabolites or metabolic class. Of note, the D2 fractions to HCPP LOPAC® clustering revealed many additional biological targets and clusters. In this chapter we focused on the characterization of the most significant one. Further studies might investigate the other identified biological targets and clusters and go one with further fractionation to identify the responsible single compound(s).

5. Material and Methods

5.1. Methods

5.1.1. Cell biological methods

5.1.1.1. Cell cultivation

HCPP screening

HeLa cells were used (parental HeLa cell line, ATCC® CCL-2™) for the HCPP screening. The cells were cultivated in cell culture flask under standard conditions (37°C, 5% CO_2) using Dulbecco's modified eagle's medium (DMEM Life Technologies, Carlsbad, USA) culture medium containing:

- GlutaMAX^{TM}-I (Life Technologies, Carlsbad, USA)
- 10% fetal bovine serum (Life Technologies, Carlsbad, USA)
- 1% antibiotic-antimycotic (anti-anti, life Technologies, Carlsbad, USA)

EASY-HIT

LC5-RIC cells were used for the EASY-HIT system. The red infected cells (RIC) from the LC5 parental line express the DsRed1 reporter in case of HIV-1 infection. The cells were cultivated in cell culture flask (T-185 Nunc Solo Flask, Nunc International, Wiesbaden, Germany) and cultured under standard conditions (37°C, 5% CO_2) using DMEM (Gibco, Karlsruhe, Germany) culture medium containing:

- GlutaMAX^{TM}-I (L-alanyl-L-glutamine), pyruvate and 4.5 g/l glucose (all from Gibco, Karlsruhe, Germany)
- 10% fetal bovine serum (FBS, Biochrom AG, Berlin, Germany)
- 1% anti-anti (Gibco, Karlsruhe, Germany)
- 1% sodium pyruvate (Gibco, Karlsruhe, Germany)

In addition, 0.74 mg/ml geneticin (G418 sulfate; PAA Laboratories, Pasching, Austria) and 0.125 mg/ml Hygromycin B (PAA Laboratories, Paching, Austria) were added every second passage, to ensure stable expression of the CD4 receptor and general stability of the reporter construct. The cells were kept for a minimum of 48 h in medium without geneticin and hygromycin B prior to the experiments. The cells were used for a maximum of ten passages.

When splitting the cells, the medium was removed, and the cells washed two times with 15 ml phosphate buffered saline (PBS) to remove fetal bovine serum residues. Afterwards, 2.5 ml trypsin-EDTA (Gibco, 1X, Karlsruhe, Germany) were added to the cells until they detach from the cell culture flask. 7.5 ml cell culture medium was added afterwards to neutralize the trypsin-EDTA. The cells in solution can now be counted using a counting chamber. The cells in three large squares were counted and their mean was

calculated. The mean was multiplied with 10^4 to get the number of cells per ml for this cell solution. This number was used for further dilutions.

5.1.1.2. Measuring anti-HIV-1 activity with the EASY-HIT

The EASY-HIT system was used according to the publication from (Kremb et al., 2011).

     i.   Cell plating

For the EASY-HIT the LC5-RIC cells were plated in 96-well plates (µCLEAR-Plate Black, Greiner Bio-One, Kremsmuenster, Germany) with a density of 10^4 cells in 100 µL cell culture medium per well. The cells were incubated for 24 h. The used cell culture conditions generate a thermal gradient from the inner wells to the frame wells on the well plates. The so-called edge effect arises in uneven distribution of cells in the frame wells accompanied with different cell adhesion and morphology (Burt, Carter, & Kricka, 1979; Lundholt, Scudder, & Pagliaro, 2003). Because of this edge effect the frame wells were filled with cell culture medium but were not used for samples. Column 11 B-G was always used for the positive controls (containing cells infected with virus, 100% infection) and column 12 B-G for the negative controls (containing cells without virus, 0% infection) like described in Figure 57.
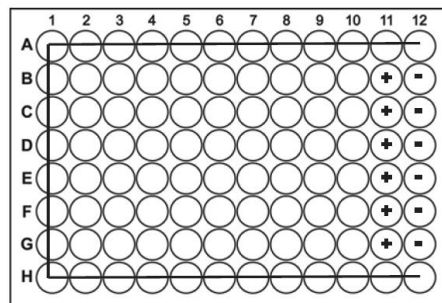
Figure 57: 96- well template for the EASY-HIT

     ii.   First Step

24h after seeding the cells into the well plates, the cell culture medium was removed and the sample (solved in 100 µL cell culture medium) and 20 µL virus were added to the cells. The total volume in the wells is now 120 µL and the cells were incubated for further 48 hours (h). To measure the virus release (iii. Second Step - Virus release assay) 20 µL of the supernatant were now transferred to new LC5-RIC cells in 96-well plates (preparation described in i. Cell plating). To measure the total fluorescence signal intensity, the medium was removed. The fluorescence was measured with a Tecan infinite M200 (Tecan, Crailsheim, Germany) at the monochromator wavelengths of 552 nm for excitation and 596 for emission. Plates were discarded afterwards.

     iii.   Second Step - Virus release assay

The new infected LC5-RIC cells were incubated for further 72 h and the total fluorescence signal intensity was measured as described above.

5. Material and Methods

        iv.   Calculation of HIV-1 infection

To calculate the HIV-1 infection the mean of the positive controls (mean$_{pos}$, cells with virus, 100% infection) and the negative controls (mean$_{neg}$, cells without virus, 0% infection) were calculated. The %HIV-1 infection per well was calculated according to Equation 4. The measured sample values are represented as X$_i$.

Equation 4: Calculation of HIV-1 infection.

$$\text{Infection} = \frac{X_i - \text{mean}_{neg}}{\text{mean}_{pos} - \text{mean}_{neg}}$$

The calculated % infection was now used to calculate the mean and the standard deviations of replicates and the IC$_{50}$ using GraphPad prism 4.0. The IC$_{50}$ was calculated using the dose-response relationship with variable slope and the restriction top = 100 and bottom = 0.

5.1.1.3.   Viability test

To measure the viability of the eukaryotic cells in the EASY-HIT test an MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromid) test was performed (Mosmann, 1983). It measures the viability due to the metabolism rate of the cells. Healthy cells can reduce the added yellow MTT to violet formazan, which would not possible for intoxicated or death cells, as the reducing NAD(P)H dependent ER and mitochondrial enzymes would be missing (Berridge, Tan, McCoy, & Wang, 1996).

After measuring the first step of the EASY-HIT 100 µL of MTT solution (100 µL cell culture medium with 0.5 mg MTT (Sigma, Taufkirchen, Germany) were added and incubated for 2 h at standard conditions. The solution was removed afterwards and 100 µL lysis solution (dimethyl sulfoxide with 10% [wt/vol] sodium dodecylsulfate and 0.6% [vol/vol] acetic acid) was added and mild shacked to dissolve the MTT formazan crystals. The colorimetric assay was measured with a Tecan infinite M200 (Tecan, Crailsheim, Germany) with a test wavelength of 570 nm and a reference wavelength of 630 nm. The untreated positive controls (cells with virus, 100% infection) were related to the treated cells on the plate.

5.1.2. Plant extracts

5.1.2.1.1.   Cystus052®

Cystus052®-Sud was bought from the company Dr. Pandalis Urheimische Medizin. The extract was vaporized under vacuum and had a dry weight of 6.2 mg/ml (Stephanie Rebensburg, 2015). For further experiments the solid material was resolved in water or cell culture medium.

5.1.2.1.2.   Tulsi

The aim of this extraction was to develop a standardized procedure with the maximum output of robust anti-HIV-1 active metabolites from Tulsi (*Ocimum tenuiflorum*). Robust means, that the metabolites must be able to resist vacuum drying (Speedvac), water free storage at -80°C, multiple freeze-thawing cycles and handling at room temperature (rt) for biological experiments. For the solvents it is important to choose one that is suitable for further measurement on the FT-ICR-MS. Therefore, all chlorinated solvents were dismissed, as they can form robust adducts with the metabolites. Tulsi was collected in Patiala (India) from Prof. Dr. Ashok Malik from the Department of Chemistry, Punjabi University on the 19.11.2013. The sample was air dried.

For the maximum anti-HIV-1 active metabolite extraction from Tulsi, several different extraction conditions were tested based on published extraction procedures e.g. (Bondia-Pons et al., 2014), different extraction length (15 min, 1 week), different solvents (water, MeOH) and different extraction temperatures (on ice, rt and 90°C). One week of extraction at rt lead to the release of cell toxic metabolites. The different temperatures showed an increase in anti-HIV-1 activity from 0°C to 25°C (data not shown). The temperature was not further raised in order to not destroy too many metabolites. The overview of the optimized metabolite extraction from Tulsi is shown in Figure 58.
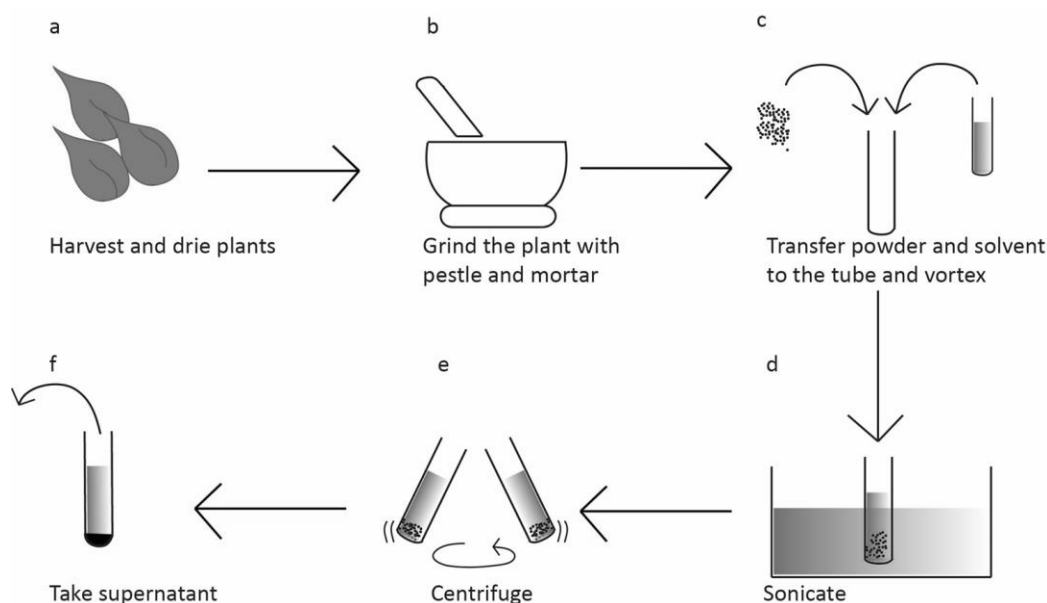
Figure 58: Experimental procedure for metabolite extraction. a) The sample was collected and dried. b) Dried material was grinded with pestle and mortar to powder. c) 6 mg/ml of the powder were transferred to a centrifugation tube and the calculated amount of extraction solvent was added afterwards. The solution was vortexed for 1 min. d) Sonication of the sample for 15 min at 25°C. e) Centrifugation of the sample (25°C, 14 000 rpm, 10 min) to obtain clear supernatant. f) Clear supernatant for further analysis can be taken.

The dried leafs of the plant material were grinded with pestle and mortar to powder. 6 mg/ml of the powder were transferred to centrifugation tubes and the calculated amount of water (MilliQ-Integral,

5. Material and Methods

Merck KGaA, Darmstadt, Germany) were added and the tube closed. The solution was vortex shacked for 1 min at rt. The metabolites were further extracted using an ultrasonic bath for 15 min at 25°C. The sonicated samples were centrifuged (25°C, 14 000 rpm, 10 min) to obtain a clear supernatant. The supernatant was decanted and dried under vacuum (SpeedVac Concentrator, Savant SPD 121P, ThermoFisher Scientific, Waltham, USA). Following this procedure, the extraction rate was 29% +/- 0.5% extracted dry weight of a 6 mg/ml solution.

5.1.3. Polyphenol enrichment for Cystus052® and Tulsi

Polyvinylpyrrolidone (PVPP) was used to enrich polyphenolic metabolites (e.g. catechin, epicatechin, gallic acid and protocatechuic acid) from the original plant extract (Helfer et al., 2014; Magalhaes et al., 2010). Therefore 25 g PVPP and 500 ml plant extract were mixed under permanent stirring (15 min, rt) to allow the attachment of polyphenols to PVPP. Afterwards the suspension was centrifuged (4000 g, 15 min, rt). The pallet represents the polyphenol enriched fraction and the supernatant the polyphenol-depleted fraction. The supernatant was decanted and preserved. The pallet was washed two times with water (Ampuwa®, Fresenius Kabi, Bad Homburg, Germany) and were centrifuged afterwards (4000 g, 15 min, rt). The pallet was vortexed with 15 ml 0.5 M NaOH and again centrifuged (4000 g, 15 min, rt) to elute the polyphenols from the PVPP. The supernatant was preserved, and the pH adjusted to 6 with 2 M HCl.

For the removal of salts from the polyphenol-enriched and the polyphenol-depleted fraction, a C-18 SPE (2g Bond Elut C18, Agilent Technologies, USA) was used. The C-18 SPE was equilibrated with 4 ml MeOH containing 1% formic acid and washed afterwards with 4 ml H_2O containing 1% formic acid. 1 ml of 10 mg/ml polyphenol-depleted or polyphenol-enriched fraction was loaded. The C-18 SPE was washed with 3 ml H_2O containing 1% formic acid and was eluted afterwards with 3 ml of MeOH. The MeOH eluate was dried under vacuum and resolved in water (Ampuwa®, Fresenius Kabi, Bad Homburg, Germany). A 10 mg/ml stock solution was prepared and stored at -80°C for further experiments.

5.1.4. SPE fractionation

Prior to the fractionation with different 100 mg SPE (see Table 13), the sample was resolved in 1 ml of acidified (pH = 2, 32%, p.a., Merck KGaA, Darmstadt, Germany) water.

Table 13: Used 100 mg SPE.

Name	Supplier
Bond Elut™ C-18	Varian, Inc., Walnut creek cA, USA
Bond Elut™ C-8	Varian, Inc., Walnut creek cA, USA
Bond Elut™ C-2	Varian, Inc., Walnut creek cA, USA
Bond Elut™ PPL	Varian, Inc., Walnut creek cA, USA
Bond Elut™ CN-E	Varian, Inc., Walnut creek cA, USA

The used 100 mg SPE was conditioned using 3 x 1 ml MeOH (Chromasolv® LC-MS grade, Sigma Aldrich, St. Louis, USA) and 3 x 1 ml HCl (pH = 2). There upon 1 ml of the sample was loaded and subsequently 1 ml HCl (pH = 2) was added and eluted to wash all salts from the SPE. The sample was gravity fed through the cartridge. Different elution series were eluted using always 1 ml solution of different concentrations, and were either in steps of 10% MeOH (e.g. 10% MeOH, 20% MeOH, 30% MeOH etc.) or in steps of 20% MeOH eluted (e.g. 20% MeOH, 40% MeOH, 60% MeOH, 80% MeOH and 100% MeOH). The gained fractions were vacuum dried and stored at -80°C. For the EASY-HIT and cell vitality tests, the fractions were reconstituted in water. The Bond Elut™ C-2 fractionation was used for further experiments. For the Bond Elut™ C-2 (5 g) the procedure was up scaled to a loaded dried sample amount of 100 mg (2% of column material) and a volume of 50 ml for the conditioning, washing and eluting steps.

5.1.5. Spearman Rank correlation

The association between m/z features and infection rate (distributed among the fractions) was calculated by Spearman Rank correlation. *P*-values were calculated via the asymptotic t approximation (Wolfe, 1973). It was tested, whether the Spearman Rank correlation is robust enough for this data set. Therefore, the EASY-HIT and FT-ICR-MS data of the 20% and 40% MeOH fractions of Cistus PP were used as a training set. For time and cost reasons the anti-HIV-1 activity of the D2 fractions was only measured ones and not in triplicates. The SD of the positive and negative controls in the EASY-HIT tests was ± 5%. This ± 5% biological variance was simulated on the EASY-HIT results of the 20% and 40% MeOH D2 fractions. The results of the top 100 correlating mass signals of the original and simulated EASY-HIT distribution were very similar. The top 10 correlating m/z stayed the same and 90 of the top 100 correlating mass signals were again present in the simulated correlation just in a slightly different order (data not shown).

5.1.6. High content phenotypic profiling (HCPP)

The used high content phenotypic profiling (HCPP) was designed and validated by Dr. Stephan Kremb and Prof. Dr. Christian Voolstra and published in (Kremb & Voolstra, 2017). For the HCPP a HeLa cell line (parental HeLa cell line, ATCC® CCL-2) was used. They were kept under standard conditions (37°C in 5 % CO_2) in cell culture medium (DMEM containing Glutamax-1, 10% fetal bovine serum and 1% antibiotic-antimycotic solution (all from Life Technologies)).

The HeLa cells were plated in 384-well plates at a density of 2000 cells in 25 μL cell culture medium per well. The plates were kept under standard conditions for 24 h. Afterwards the cells were treated with 2 μL of the re-dissolved samples in quadruplicates. After 24 h of incubation four different cell staining protocols were used to visualize specific cell targets. The staining protocols are called panels and an overview of each panel with its cellular target, description and used staining solutions are summarized in Table 14.

5. Material and Methods

Table 14: Overview of the different panels and their cellular targets together with a description and the used staining solutions.

Panel	Cellular target	Description	Staining solutions
Panel 1	Nucleus	Nuclear parameter	Hoechst33342 (Thermo Fisher Scientific)
	Actin	Cytoskeleton	Phalloidin-FITC (Sigma Aldrich)
	Tubulin	Cytoskeleton	Beta-tubulin antibody and GAM-DyLight 550 (Thermo Fisher Scientific)
	Mitochondria	Mitochondrial membrane potential	MitoTracker® Orange CMTMRos (Life Technologies)
Panel 2	Whole cell	Whole cell morphology	Wheat Germ Agglutinin, Alexa Fluor® 488 Conjugate (Life Technologies)
	ER	Endoplasmic reticulum	ER-Tracker Blue-White DPX (Life Technologies)
	Lysosomes	Lysosomes	LysoTracker Red DND-99 (Life Technologies)
	Membrane	Membrane	Wheat Germ Agglutinin, Alexa Fluor® 488 Conjugate (Life Technologies)
Panel 3	Nucleus	Nuclear parameters	Hoechst33342 (Thermo Fisher Scientific)
	NF-κB	NF-κB activation/distribution	Antibody for NF-κB and GAR-DyLight 550 (Thermo Fisher Scientific)
Panel 4	Nucleus	Nuclear parameters	Hoechst33342 (Thermo Fisher Scientific)
	Caspase 9	Caspase 9 activation/distribution	Caspase 9 antibody and GAM-DyLight 550 (Thermo Fisher Scientific)
	p53	p53 activation/distribution	p53 antibody and GAR- DyLight 488 (Thermo Fisher Scientific)

For the blocking, permeabilization and washing steps in the panels, optimized Cellomics HCPP reagents were used e.g. Wash Buffer (WB), Wash Buffer II (WBII), Blocking Buffer (BB) and Permeabilization Buffer (PB) (all from Thermo Fisher Scientific, Waltham, MA, USA). The following µL descriptions always refer to the amount of solution added to one well and the cells were always incubated under standard conditions (37°C in 5% CO_2).

Panel 1: Nucleus – Actin – Tubulin - Mitochondria

The cell culture medium was removed, and the cells were incubated with 12.5 µl of a mitochondrial dye solution (0.2 µl/ml, MitoTracker® Orange CMTMRos, Life Technologies), in pre-warmed cell culture medium for 30 min under standard cell culture conditions. Afterwards the cell culture medium was removed, and the cells were fixed using 4% formaldehyde in PBS for 20 min. After removing the formaldehyde solution, the cells were washed two times with PBS. Afterwards the cells were permeabilized for 15 min using PB (1:10 in PBS) and blocked for 15 min using BB (1:10 in PBS). The cells were incubated for 1 h under standard conditions using 12.5 µL primary staining solution (3.6 µl/ml

Phalloidin-FITC (Sigma Aldrich)) and 1.3 µl/ml beta-tubulin antibody (Thermo Fischer Scientific), in BB 1:10 in PBS. Afterwards the cells were washed two times with BB and subsequently 12.5 µL of the secondary staining solution (1:500 GAM-DyLight 550, Thermo Fischer Scientific, in BB 1:10 in PBS) was incubated for one hour. The cells were afterwards washed for three times with PBS and stained for 10 min using 25 µL Hoechst33342 (0.8 µl/ml Hoechst33342, Thermo Fisher Scientific, in PBS) and subsequently washed three times with PBS.

Panel 2: Whole cell – ER – Lysosomes - Membrane

The cell culture medium was removed, and the ER and lysosomes were stained for 30 min under standard conditions using 12.5 µL (including 1 µl/ml ER-Tracker Blue-White DPX and 0.2 µl/ml LysoTracker Red DND-99, both from Life Technologies, in pre-warmed cell culture medium). The staining solution was removed, and the cells were washed two times with PBS. After fixation of the cells using 4% formaldehyde in PBS for 20 min the cells were washed again two times with PBS. The cells were further incubated for 10 min at rt using 12.5 µL labelled wheat germ agglutinin (5 µl/ml Wheat Germ Agglutinin, Alexa Fluor® 488 Conjugate, Life Technologies, in PBS) and washed for two times with PBS afterwards.

Panel 3: Nucleus – NF-κB

The cell culture medium was removed, and the cells were fixed using 4% formaldehyde in PBS for 20 min. Afterwards the cells were permeabilized for 17 min using PB (1:10 in water) and washed two times with PBS. The cells were incubated for 1 h under standard conditions using 12.5 µL of NF-κB primary staining solution (6.81 µl/ml antibody for NF-κB, Thermo Fisher Scientific, in PBS). After removal of the antibody solution the cells were incubated for 15 min with WBII and washed two times with PBS. For the secondary staining solution 12.5 µL (including 2 µL/mL 1:500 GAR-DyLight 550, Thermo Fischer Scientific, in WB) was incubated for an additional hour. Afterwards the cells were incubated for 10 min with WBII, stained for 10 min using 25 µL Hoechst33342 (0.8 µl/ml Hoechst33342, Thermo Fisher Scientific, in PBS) and washed two times with PBS afterwards.

Panel 4: Nucleus – Caspase 9– p53

The cell culture medium was removed, and the cells were fixed using 4% formaldehyde in PBS for 20 min. After removing the formaldehyde solution, the cells were permeabilized for 17 min using PB (1:10 in PBS) and washed afterwards two times with PBS. For blocking, BB (BB 1:10 in PBS) was added for 17 min. The BB solution was removed and 12.5 µL of the primary antibody solution (5.5 µl/ml p53 antibody and 1.5 µl/ml caspase 9 antibody, both Thermo Fisher Scientific, in BB) was incubated for 1 h. The cells were afterwards washed two times with WBII and one time with PBS. 12.5 µL of the secondary staining solution (1:500 GAR-DyLight 488 and 1:500 GAM-DyLight 550, both Thermo Fischer Scientific, in BB) was incubated for additional 1 h. After removal of the staining solution and washing one time with WB the cells were stained for 10 min using 25 µL Hoechst33342 (0.8 µl/ml Hoechst33342, Thermo Fisher Scientific, in PBS) and washed two times with PBS afterwards.

In this study, the cellular target mitochondria were stained and measured in Panel 1 instead of Panel 3 (Kremb & Voolstra, 2017).

5. Material and Methods

The CellomicsArrayScan VTI (Thermo Fisher Scientific) platform was used for the high content analysis. Images were made with a 10x objective (Zeiss Plan Neofluar, NA 0.3) and analyzed using the Compartment Analysis Bio Application (Cellomics, Thermo Fisher Scientific). For each well, at least 500 valid objects were analyzed.

The used cell culture conditions (37°C in 5% CO_2) generate a thermal gradient from the inner wells to the frame wells on the well plates. The so-called edge effect arises in uneven distribution of cells in the frame wells accompanied with different cell adhesion and morphology (Burt et al., 1979; Lundholt et al., 2003). Because of this edge effect in this study the frame wells were filled with cell culture medium but not used for calculation.

After verifying that the control features are normally distributed a Grubbs outlier test was applied for every set of control features (Equation 5).

Equation 5: Grubbs' Outlier Test.

$$g = \frac{\max |X_i - \overline{X}|_{i=1\ldots n}}{SD}$$

g	= test statistic
X_i	= tested values
$\overline{X}$	= mean of the values
SD	= standard deviation

Every set of control features contains n = 14 features and was tested with a 2,5% significance Level (α = 0.05) resulting in a g-crit = 2.5073 (Grubbs, 1969; Stefansky, 1972). If the resulting test statistic g is greater than the calculated g-crit the value is defined as an outlier and removed from the dataset. After removing these outliers from the dataset, the test must be repeated using the same g-crit.

For the sample features the mean of each feature was calculated and afterwards normalized to the corresponding, outlier tested, control feature. The control features were set to 1. All features were combined to give the cytological profile for each sample. For Data analysis 133 features were extracted from the different panels as described in Figure 59.
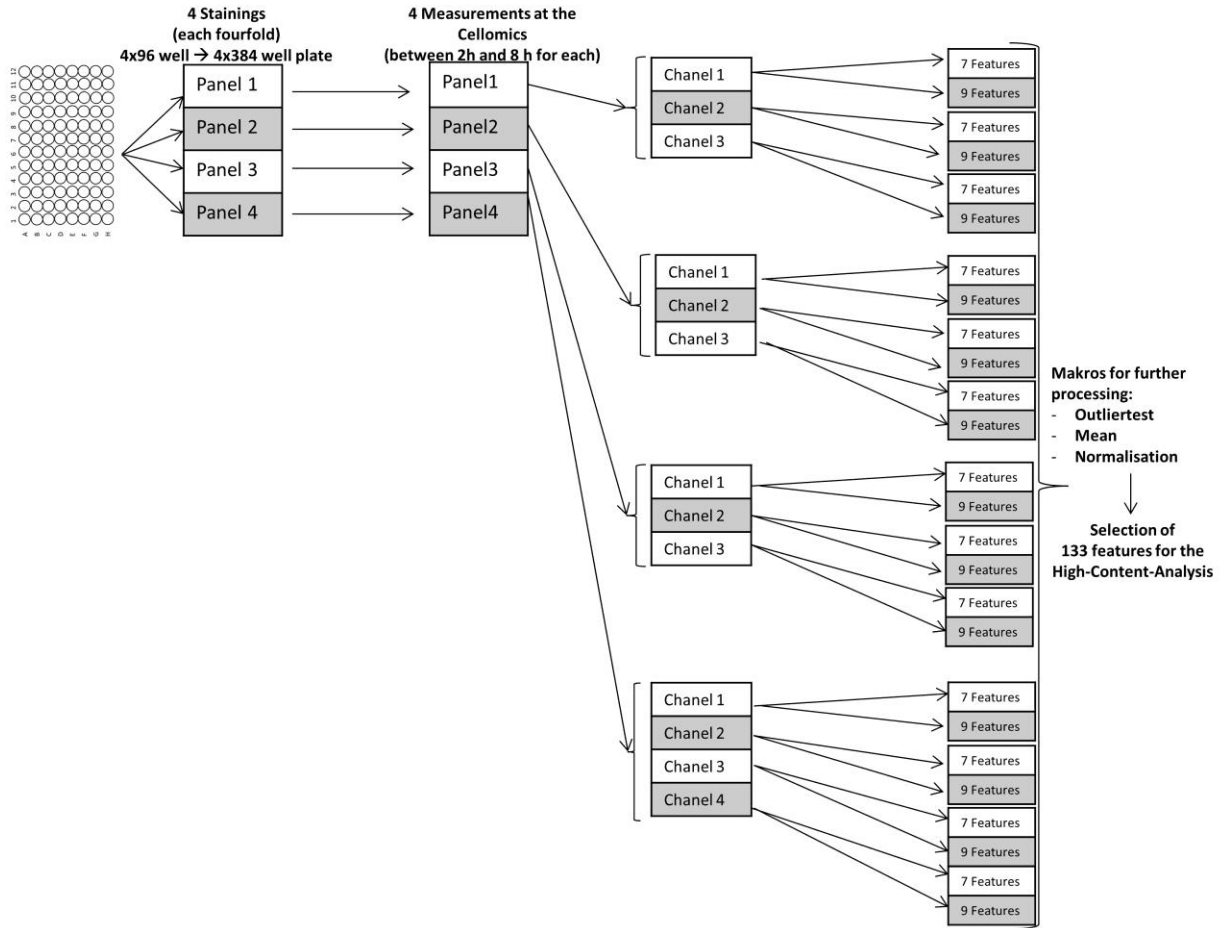
Figure 59: Overview of feature extraction from the different panels.

The 133 features were selected according to the description in Table 15.

5. Material and Methods

Table 15: Overview of the 133 extracted and selected features. Abbreviations: T = Total, A = Average, I = Intensity, R = Ring, C = Circ, Rat = ratio, Sp = Spot, Ar = Area, SpC = Spot Count, P2a = shape measure based on ratio of perimeter squared to 4π*area, LWR = shape measure based on ratio of length to wide of object-aligned bounding box.

Whole Cell	whC_Size	p53	p53_R-TI	Actin	ac_R-TI	Mitochondria	mt_R-TI	Membrane	mb_R-TI
	whC_Area		p53_R-AI		ac_R-AI		mt_R-AI		mb_R-TI
	whC_P2A		p53_C-TI		ac_C-TI		mt_C-TI		mbR-AI
	whC_LWR		p53_C-AI		ac_C-AI		mt_C-AI		mbC-TI
	whC_TI		p53_Rat-C-R		ac_Rat-C-R		mt_Rat-C-R		mb_C-AI
	whC_AI		p53_R-SpTI		ac_R-SpTI		mt_R-SpTI		mb_Rat-C-R
	whC_VI		p53_R-SpAI		ac_R-SpAI		mt_R-SpAI		mb_R-SpTI
Nucleus	Nuc-Size		p53_R-SpTAr		ac_R-SpTAr		mt_R-SpTAr		mb_R-SpAI
	Nuc_Area		p53_R-SpAAr		ac_R-SpAAr		mt_R-SpAAr		mb_R-SpTAr
	Nuc-P2A		p53_R-SpC		ac_R-SpC		mt_R-SpC		mb_R-SpAAr
	Nuc-LWR		p53_C-SpTI		ac_C-SpTI		mt_C-SpTI		mb_R-SpC
	Nuc-TI		p53_C-SpAI		ac_C-SpAI		mt_C-SpAI		mb_C-SpTI
	Nuc-AI		p53_C-SpTAr		ac_C-SpTAr		mt_C-SpTAr		mb_C-SpAI
	Nuc-VI		p53_C-SpAAr		ac_C-SpAAr		mt_C-SpAAr		mb_C-SpTAr
NFkB	nf_R-TI		p53_C-SpC		ac_C-SpC		mt_C-SpC		mb_C-SpAAr
	nf_R-AI	Caspase 9	cp_R-TI	Tubulin	tu_R-TI	ER	er_C_TI		mb_C-SpC
	nf_C-TI		cp_R-AI		tu_R-AI		er_C_AI		
	nf_C-AI		cp_C-TI		tu_C-TI		er_C_SpTI		
	nf_Rat-C-R		cp_C-AI		tu_C-AI		er_C_SpAI		
	nf_R-SpTI		cp_Rat-C-R		tu_Rat-C-R		er_C_SpTAr		
	nf_R-SpAI		cp_R-SpTI		tu_R-SpTI		er_C_SpAAr		
	nf_R-SpTAr		cp_R-SpAI		tu_R-SpAI		er_C_SpC		
	nf_R-SpAAr		cp_R-SpTAr		tu_R-SpTAr	Lysosomes	ly_C_TI		
	nf_R-SpC		cp_R-SpAAr		tu_R-SpAAr		ly_C_AI		
	nf_C-SpTI		cp_R-SpC		tu_R-SpC		ly_C_SpTI		
	nf_C-SpAI		cp_C-SpTI		tu_C-SpTI		ly_C_SpAI		
	nf_C-SpTAr		cp_C-SpAI		tu_C-SpAI		ly_C_SpTAr		
	nf_C-SpAAr		cp_C-SpTAr		tu_C-SpTAr		lyC_SpAAr		
	nf_C-SpC		cp_C-SpAAr		tu_C-SpAAr		ly_C_SpC		
			cp_C-SpC		tu_C-SpC				

For further data analysis, the cytological profiles of the sample were clustered using average linkage clustering with optimized gene leaf order and a Pearson, Spearman Rank or Kendall's Tao correlation (Reisen et al., 2013). For this purpose the Multi Experiment Viewer28 (MeV v4.9, Dana-Farber Cancer Institute, Boston, MA, USA) was used (Saeed et al., 2003).

5.1.6.1. Toxicity and concentration effects

The used cell culture conditions generate a thermal gradient from the inner wells to the frame wells on the well plates. The so-called edge effect arises in uneven distribution of cells in the frame wells accompanied with different cell adhesion and morphology (Burt et al., 1979; Lundholt et al., 2003). Because of this edge effect the frame wells were filled with cell culture medium were not used for calculations. An example of the edge effect is displayed in Figure 60.
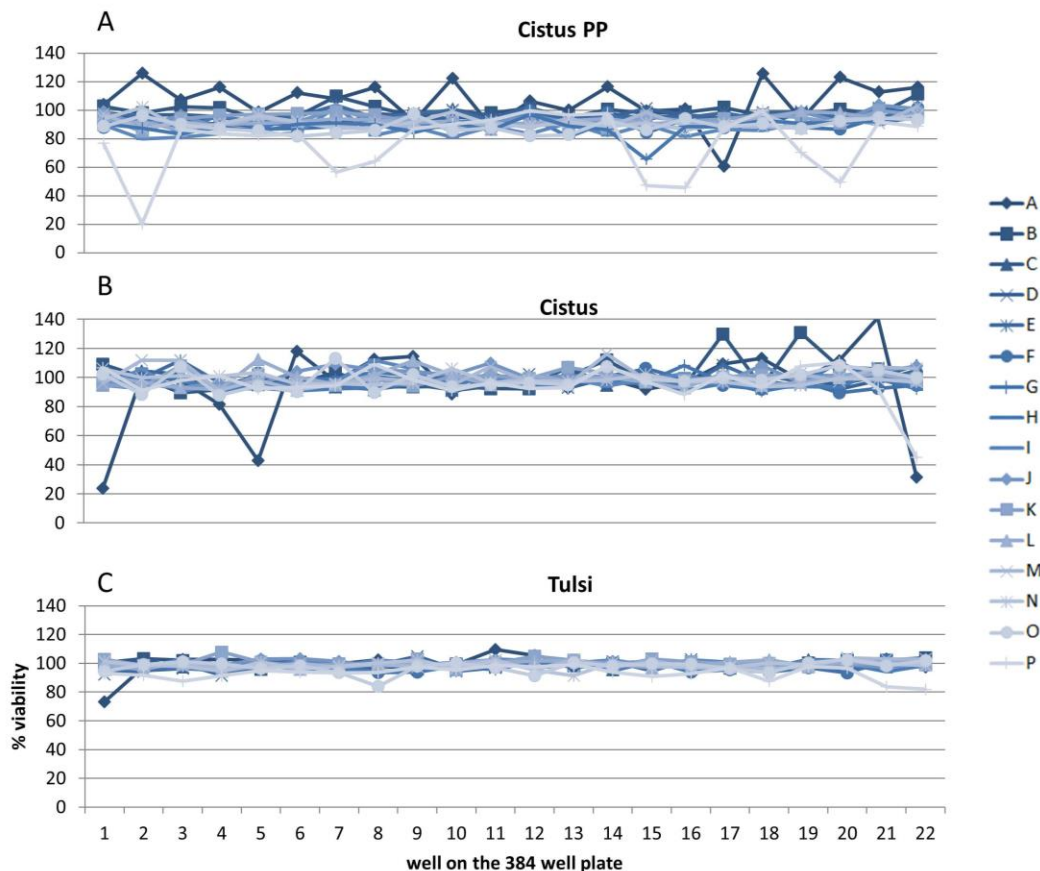
Figure 60: The viability of the cells with the used amount of sample were tested prior to the HCPP and the data are shown in A) for Cistus PP, B) for Cistus and C) for Tulsi. On the x-axis the column on the 384 well (A-P x 1-22 = 384) plate from 1 to 22 are represented and the corresponding viability for every well in the rows from A to P on the y-axis. The used concentrations of the sample is not toxic for the cells. Outliers were detected in the frame wells of the plate see row A and P.

5.2. Machines

5.2.1. FT-ICR-MS analysis

Ultrahigh resolution mass spectra were acquired using a Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR-MS) (SolariX®, Bruker Daltonik GmbH, Bremen, Germany) equipped with a 12 Tesla superconducting magnet (Magnex Scientific, Oxford, Great Britain). For sample injection into the electrospray ionization source (Apollo2, Bruker Daltonik GmbH, Bremen, Germany) a microliter pump with flow rate of 2 μL/min was used. The samples were measured in negative ionization mode. They were diluted with MeOH and water to reach the target concentration. The used FT-ICR-MS parameters for the measurement of the samples are listed in Table 16.

5. Material and Methods

Table 16: Parameters for the FT-ICR-MS measurements.

Parameters	Values
Acquired Scans	250
Analyzer Entrance [V]	10
Capillary[V]	3600
Collision Voltage (Entrance) [V]	4
Drying Gas Flow Rate [L/min]	4
Drying Gas Temperature [°C]	180
Ion Accumulation Time [sec]	0.3
Mass range [m/z]	122.9 - 1200
Nebulizer gas flow rate [bar]	2
Sidekick [V]	-5
Sidekick Offset [V]	3
Spray Shield [V]	-500
Source Accumulation [sec]	0.03
Time of flight [ms]	0.7
Time domain transient [MW]	4

Prior to every measurement, the system was calibrated using L-Arginine (2 ppm, mass error < 100 ppb) clusters (Geller & Lifshitz, 2003; Nemes, Schlosser, & Vekey, 2005). The samples were randomized and after every 10 samples one blank (MeOH) was added to minimize the risk of cross contamination. A Gilson auto sampler (sample changer 223, Gilson Inc., Middleton, USA) was used to inject the samples. The samples were additionally cooled to 8°C. After acquiring the spectra, they were further processed using DataAnalysis 4.2.386.0 (Bruker Daltonics GmbH, Bremen, Germany) and an in house developed data processing strategy (Kanawati et al., 2017).

In Kanawati et al., 2017 the power function uses the specific resolution at every m/z to delete satellite peaks (Amster, 1996; van der Burgt et al., 2007) and applies further a mass defect filter to the dataset. Because of the fraction character of the samples in this study the metabolite composition changes a lot from sample to sample. That's why only external calibration is possible. To guarantee perfect calibration, plant specific calibration lists including fatty acids and MeOH specific compounds (Table 35) were created. This resulted in a maximum calibration error of 0.05 ppm. For the Cistus calibration list (Table 33 supplement) published compounds (Barrajon-Catalan et al., 2011; Ehrhardt et al., 2007; Riehle, Rusche, Saake, & Rohn, 2014; Riehle, Vollmer, & Rohn, 2013; Santagati et al., 2008) were verified via exact mass and isotopic pattern in the measured mass spectra and additionally used together with fatty acids for calibration. The same was done for the Tulsi calibration list (Farhin Inam, 2014; Kothari, Bhattacharya, & Ramesh, 2004; Zheljazkov, Cantrell, Tekwani, & Khan, 2008)(Table 34, Supplement). Additionally, to the carefully calibration of the machine and of the MS spectra, the quality of the MS spectra was further assured by checking the FID of the acquired spectra. This is necessary due to the fact, that the spectra differ in their molecular composition because of their fraction character. To assure that in no spectra an oversaturation of the ICR cell took place, the FID was accurately monitored and, in this study, never

exceeded 2×10^4 counts in each time-domain transient. Because the filtering of the satellite peaks and the mass defect took place in the first place, it was possible to align 750 mass spectra in one matrix using the in-house written software Matrix Generator (1 ppm window). The possible formulae compositions of the measured masses were calculated using the in house software NetCalc (Tziotis et al., 2011). For better visualization of the data van Krevelen diagrams were used (Krevelen, 1950) which show on the x-axis the oxygen to carbon (O/C) ratio and on the y-axis the hydrogen to carbon ratio (H/C). The D2 fractions were resolved in 80% MeOH prior to the experiment and were further diluted 1:10 using 80% MeOH to avoid over saturation of the ICR cell.

5.2.2. Resolution and mass defect filter prior to matrix generation

The acquired data were filtered according to (Kanawati et al., 2017) and are briefly described here. After the samples were measured on the FT-ICR-MS, further data processing steps were necessary. These processing steps lead to the generation of one data matrix with individual samples on the x-axis and m/z on the y-axis. With this matrix, further data filtering and interpretation can take place. As so far, no tool for matrix generation was able to generate one matrix containing 750 spectra (measured with 4 MW) representing this study. For this purpose, an adapted processing strategy was developed in collaboration with Dr. Basem Kanawati.

The main aspect of the new MS data processing is the filtering of the data prior to the matrix generation, to ensure that only valid peaks are used for the m/z alignment and generation of the matrix. The error happening by the alignment of no-valid peaks to valid peaks in the alignment algorithm used by Matrix Generator program is thus minimized, by adopting this pre-filtering of individual mass spectra, before the matrix is being generated. Therefore, the well-known data filtering steps i) deleting satellite peaks and ii) mass defect filtering (Hughey, Hendrickson, Rodgers, Marshall, & Qian, 2001) need to be processed on the raw data for each individual mass spectrum. Figure 61 A) visualizes the developed consecutive steps which lead to the processing of 750 spectra and the successful generation of one matrix out of 750 spectra (4 MW) with the possibility to align even more spectra. To save time and to avoid careless mistakes, the single steps were programmed to work automatically.

The first step ("Upload") is the upload of up to 60 4 MW FT-ICR-MS spectra from one measurement day into DataAnalysis. Afterwards ("Find"), the standard parameters (S/N = 4, relative intensity threshold = 0.01%, absolute intensity threshold = 100) need to be applied for every spectrum. The next step ("Calibrate") is the external calibration using one carefully calibrated spectra of the uploaded measuring day. Due to the measurement of fractions in this study the external calibration is a crucial step for spectra calibration, as in some fractions only few compounds for internal calibration are available, which was inadequate to perform internal calibration. Subsequently ("Export") the spectra are exported in .xml format for further processing. The .xml format does allow the last step ("Filter") which let the above-mentioned power-function resolution mass filter to automatically recognise and eliminate the satellite peaks and also perform mass defect data filtering afterwards.

The output of the filtered spectra is in .txt format as this can be further used for the generation of the matrix using Matrix Generator. The entire process is semi-automatic as the 5 single steps (Upload, Find,

5. Material and Methods

Calibrate, Export and Filter) need to be started individually but work then automatically. The full process takes around 60 min for processing 60 spectra in a rate of 1 min/1 spectrum.
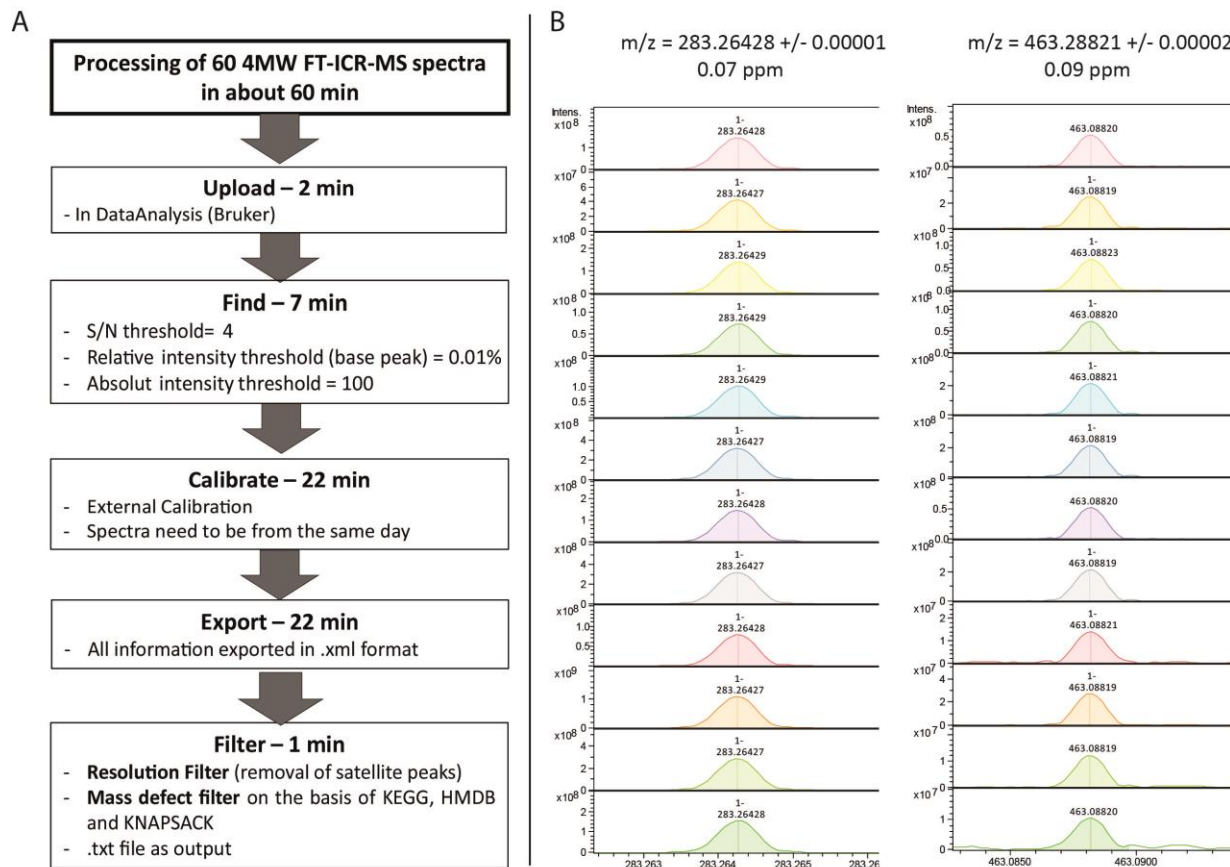
Figure 61: A) overview of the data FT-ICR-MS data processing, PC specification, B) calibration proof in low and high mases with careful external calibration.

The quality of the externally calibrated spectra is represented in Figure 61 B). Twelve randomized calibrated spectra of one measurement day were picked and the max error of two m/z, after using plant specific calibration lists was calculated. The error for one lower mass at 283.26428 ± 0.00001 Da (0.07 ppm) and for one higher mass at 463.28821 ± 0.00002 Da (0.09 ppm) was calculated. The error of 0.07 and 0.09 ppm is comparable to the conventional error of internal calibrated spectra. In addition to the advantages of the time and careless error saving in the semi-automatic mode, the possibility to work with huge studies by aligning 750 spectra and more (4 MW) into one matrix exists. Moreover, every MS spectrum after this above metioned automatic external calibration can be manually opened for further improvements or quality assurance.

5.2.3. Matrix generation, filtering and NetCalc

After filtering and exporting the spectra they were aligned using the in-house software Matrix Generator, which aligns the spectra within a max mass error of 1 ppm. The generated matrix was containing 32.436 mass signals. In a next step, all signals occurring in the blank MeOH spectra, so called zero controls, were excluded from the matrix. This filtration lead to a matrix containing 30.613 mass signals. The matrix was filtered afterwards for metabolites, which only occur once in the data set, so called unique metabolites, resulting in a matrix of 13.968 mass signals. Now the matrix was divided and in case of Tulsi one metabolite must be present in the Tulsi matrix (raw extract, SPE fractions and HPLC fractions of Tulsi) for at least three times to be valid. All the others occurring only two times were excluded. In case of Cistus and Cistus PP only mass signals were valid which occur two times in the Cistus matrix and additionally two times in the Cistus PP matrix. As any further filtering would exclude meaningful information generated by the D1 and D2 fractionation, this is the final matrix for any further interpretation.

For molecular formula calculation, the in-house written software NetCalc (Tziotis et al., 2011) was used. The calculation resulted in 7864 valid molecular formulas used for further investigations.

5.2.4. FT-ICR-MS/MS analysis

The machine was calibrated as described in 5.2.1. The used parameters for the measurement are listed in Table 17.

Table 17: Parameters for the FT-ICR-MS/MS measurements.

Parameters	Values
Acquired Scans	50
Analyzer Entrance [V]	10
Capillary[V]	3600
Collision Voltage (Entrance) [V]	4
Collision Voltage collision cell [V]	0 - 30
Drying Gas Flow Rate [L/min]	4
Drying Gas Temperature [°C]	180
Ion Accumulation Time [sec]	2 - 4
Nebulizer gas flow rate [bar]	2
Sidekick [V]	-5
Sidekick Offset [V]	2.5
Spray Shield [V]	-500
Source Accumulation [sec]	0.01
Time of flight [ms]	0.7
Time domain transient [MW]	4

In general, the fragmentation was achieved by the collision of the ions with argon atoms. For this purpose, the ions were accumulated in the first hexapol and pulsed through the quadrupole into the collision cell. In the collision cell the collision-induced dissociation (CID) took place using gradually increasing voltage

5. Material and Methods

(0 – 30 eV). The acquired mass spectra were externally calibrated, and the molecular formulas were calculated using NetCalc.

5.2.5. HPLC-PDA conditions

For the injection into the HPLC a similar proportion of each elution step was injected. In general, from 100 mg dried sample material, that was introduced into the D1 C2-SPE, 25 tubes (each 2 ml) were eluted. Three tubes were vacuum dried, and each tube was resolved in 60 µL water for concentration. The three tubes were combined and from the resulting 180 µL, 100 µL were injected into the HPLC. The resulting dry weight per injection are displayed in Table 16.

Table 18: Concentration of D1 fractions of Cistus PP injected (100µl) into the HPLC for further D2 fractionation. N/a = not available.

Sample	20% MeOH	20% MeOH	20% MeOH	20% MeOH	100% MeOH
Cistus PP	3 mg	1.6 mg	0.3 mg	0.3 mg	n/a
Cistus	2 mg	0.3 mg	0.16 mg	n/a	0.16 mg
Tulsi	0.16 mg	0.16 mg	n/a	n/a	0.16 mg

The vacuum dried samples were resolved in water and filtered using 0.20 µm cellulose acetate membrane (Whatman® GD/X, Sigma Aldrich, St. Louis, USA) (Ivanova-Petropulos et al., 2015) before injection into the HPLC. Afterwards Acetonitril (ACN) (Chromasolv® LC-MS grade, Sigma Aldrich, St. Louis, USA) was added until the starting concentration of the used gradient (10% acetonitrile (ACN), 90% water) was achieved. For the LC separation, the column XBridge™ Shield RP (C18, 5 µm, 4.6 mm x 250 mm) (Waters,Milford, Massachusetts,USA) was used under the conditions displayed in (Table 19 and Table 20).

Table 19: Chromatographic conditions.

Parameters	Values/Description
Colum	X-Bridge ™ Shield RP C18 (4.6 mm x 250 mm; 5µm)
Flow rate	1 ml/min
Detector	PDA
Detector wavelength	205 – 400 nm
Injection volume	100 µL
Temperature	30°C
Pressure limit	400 bar
Elution type	Gradient elution
Run time	59 min
Fraction collection time	60 sec

Table 20: For the separation A = 90/10 H$_2$O/ACN and B = 10/90 ACN/H$_2$O was used. The gradient for the separation of C2-20% SPE fraction from min 0 until min 59 is shown in the table. * = the gradient was adapted for better seperation C2-40% (B 70%, A 30%), C2-60% (B 80%, A 20%), C2-80% (B 90%, A 10%) and C2-100% (B 100%, A 0%).

Time [min]	B [%]	A [%]
0	10	90
4	10	90
51	60*	40*
51.01	100	0
54	100	0
54.01	10	90
59	10	90

5.2.5.1.   Quality control, intra- and interday precision using RSD%

For the quality control of the HPLC-PDA separation a homogenate of the used sample matrixes Cistus, Cistus PP and Tulsi was prepared in 10% ACN/ 90% water and standards similar to the target compound class were spiked into it. Therefore (+)-catechin hydrat and 3,5 Dihydrobenzoic acid (both from Sigma Aldrich, St. Louis, USA) were dissolved in water and diluted to a concentration of 500 ppm 10% ACN solution. The column was conditioned using this homogenate. Prior to every fractionation measurement, the homogenate was measured for three times and the spiked homogenate for additional two times to ensure and monitor the stable fractionation conditions.

5. Material and Methods
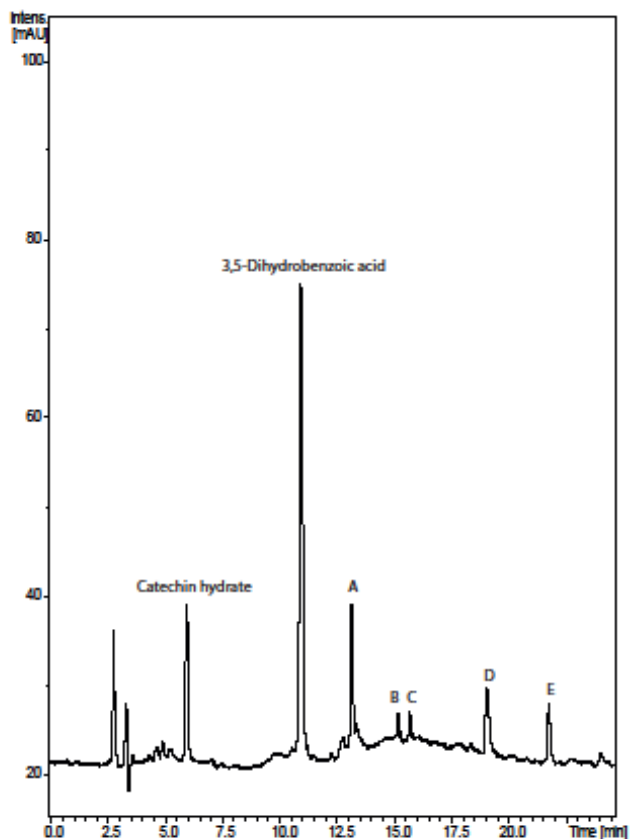
Peaks	Catechin hydrate	3,5 Dihydrobenzoic acid	A	B	C	D	E
Inter day precision							
RSD [%]	1.07	0.81	0.56	0.4	0.45	0.54	0.79
Intra day precision							
RSD [%]	0.5	0.27	0.25	0.15	0.17	0.11	0.1

Figure 62: The chromatogram shows an example spectrum of Quality control (QC) with Standards UV Chromatogram 205-400 nm. The table gives the inter day precision (between day precision): RSD% in [min] (criteria <2%) of n = 3 following measurement days (Koop et al., 2013) and intraday precision (repeatability): average and RSD% (criteria <1%); A-E are always measured for five times, Catechin hydrat and 3,5-Dihydrobenzoic acid for two times per day.

The RSD% in a between day precision is under 2% as requested (SHABIR, 2004). Its higher compared to the intraday precision because of the different measuring days and sample matrices. The RSD% of an intraday precision is under 1% as requested (SHABIR, 2004).

5.2.6. Other used Machines

Table 21: Machines used in the cell culture.

EASY-HIT	
Name	**Supplier**
Tecan Infinite M200 (monochromator)	Tecan Group
Incubator	Memmert
Gas supply incubator	Heraeus
HeraSafe Laminar Flow Bank	Heraeus
Liquidator96 Benchtop Liquid Handling Device	Steinbrenner Laborsysteme GmbH
Multispeed Vortex	Kisker Biotech
HandyStep Electronic	Brand
Eppendorf set of Pipetts "research"	Eppendorf
HCPP screening	
Name	**Supplier**
Cellomics ArrayScan VTI HCR Reader	Thermo Fisher Scientific
10x objective	Zeiss Plan Neofluar, NA 0.3

5.3. Chemicals and Material

5.3.1. Chemicals

Table 22: Substances used.

EASY-HIT	
Name	**Supplier**
DMEM (Dulbecco's modified eagle's medium; GlutamMAX-I, 4,5 g/l glucose, pyruvate)	Gibco
Fetal bovine serum (FBS)	Gibco
Antibiotic- antimycotic (anti-anti, 10000 units/ml penicillin G; 10000 units/ml streptomycin; 25 µg/ml amphotericin)	Gibco
Sodium pyruvate	Gibco
Geneticin G418	Gibco
Hygromycin B	Invitrogen
MTT (3-(4,5-dimethyldiazol-2-yl)-2,5 diphenyl tetrazolium romid)	Sigma-Aldrich
Trypsin-EDTA	Gibco
Phosphat buffered saline (PBS)	140 mM NaCl, 5,4 mM KCl, 9,7 mM Na_2HPO_4 $2xH_2O$, 2mM KH_2PO_4
Cystus052®brew unfiltered	Dr. Pandalis Urheimische Medizin
Dimethyl sulfoxide (DMSO)	Chromasolv®, Sigma-Aldrich
Sodium dodecylsulfate (SDS)	Sigma-Aldrich
Acetic acid	Chromasolv®, Sigma-Aldrich
HCPP screening	

5. Material and Methods

Name	Supplier
LOPAC® [1280]	Sigma-Aldrich
Cellomics HCS reagents Wash Buffer (WBI)	Thermo Fisher Scientific
Cellomics HCS reagents Wash Buffer II (WBII)	Thermo Fisher Scientific, Waltham, MA, USA).
Cellomics HCS reagents Blocking Buffer (BB)	Thermo Fisher Scientific
Cellomics HCS reagents Permeabilization Buffer (PB)	Thermo Fisher Scientific
Phalloidin-FITC	Sigma Aldrich
Beta-tubulin antibody	Thermo Fisher Scientific
GAM-DyLight 550	Thermo Fisher Scientific
Hoechst33342	Thermo Fisher Scientific
ER-Tracker Blue-White DPX	Life Technologies
LysoTracker Red DND-99	Life Technologies
Wheat Germ Agglutinin, Alexa Fluor® 488 Conjugate	Life Technologies
MitoTracker® Orange CMTMRos	Life Technologies)
Antibody for NF-κB	Thermo Fisher Scientific
GAR-DyLight 488	Thermo Fisher Scientific
p53 antibody	Thermo Fisher Scientific
Caspase 9 antibody	Thermo Fisher Scientific
Analytical lab	
Name	**Supplier**
3,5 Dihydrobenzoic acid	Sigma Aldrich, St. Louis, USA
(+)-catechin hydrat	Sigma Aldrich, St. Louis, USA
Hydrochloric acid (32%)	Merck KGaA, Darmstadt, Germany
MeOH Chromasolv® LC-MS grade	Fluka® Analytical (Sigma-Aldrich-Aldrich, St. Louis,USA)
Acetonitrile Chromasolv® LC-MS grade	Fluka® Analytical (Sigma-Aldrich-Aldrich, St. Louis,USA)

5.3.2. Material

Table 23: Material used.

EASY-HIT	
Name	**Supplier**
96-well culture plates, μ-clear ground, black wall	Greiner Bio-One
Combitips 1-10 ml for HandyStep®	Eppendorf
Reaction tubes 0,5 ml, 1,5 ml, 2 ml	Eppendorf
Falcon® 15 ml, 50 ml tubes	BD Biosciences
Fast-Read 102 counting camber	Biosigma
96-well microtiter plates	Greiner Bio-One
Pipette tips (10-1000 μl)	Eppendorf

Nunc Solo Flask T-175 cell culture flask	Nunc International
C-18 SPE; 1 mg; Varian Bond Elute	Agilent Technologies
Analytical lab	
Name	**Supplier**
X-Bridge TM Shield RP C18 (4.6 mm x 250 mm; 5µm)	Waters, Milford, Massachusetts,USA
Whatman® GD/X syringe filters	Sigma Aldrich, St. Louis, USA
100 mg SPE (C-18, C-8, C-2, PPL and CN-E) Bond Elute^{TM}	Varian, Inc., Walnut creek cA, USA

5.4. Used external programs

Table 24: External Software.

EASY-HIT	
Name	**Supplier**
GraphPad Prism 5	GraphPad Software Inc.
Tecan i control (Tecan Group)	Tecan Group
HCPP screening	
Name	**Supplier**
Cellomics, Compartmental Analysis Bio Application	Thermo Fisher Scientific
Analytical lab	
Name	**Supplier**
Endnote X.7	Thomson Reuters
Illustrator CS5	Adobe Systems
Microsoft Office 2010	Microsoft Corp.
Adobe Acrobat Pro 9	Adobe Systems
Adobe Illustrator CS 3	Adobe Systems
Data Analysis 4.2	Bruker Daltonics
MassTRIX	HelmholtzZentrum munich
Multi Experiment viewer (MeV v. 4.9)	Dana-Farber Cancer Institute, Boston, MA, USA

6. Supplementary Information

6.1. Chapter 2

6.1.1. UV data for Tulsi, Cistus and Cistus PP
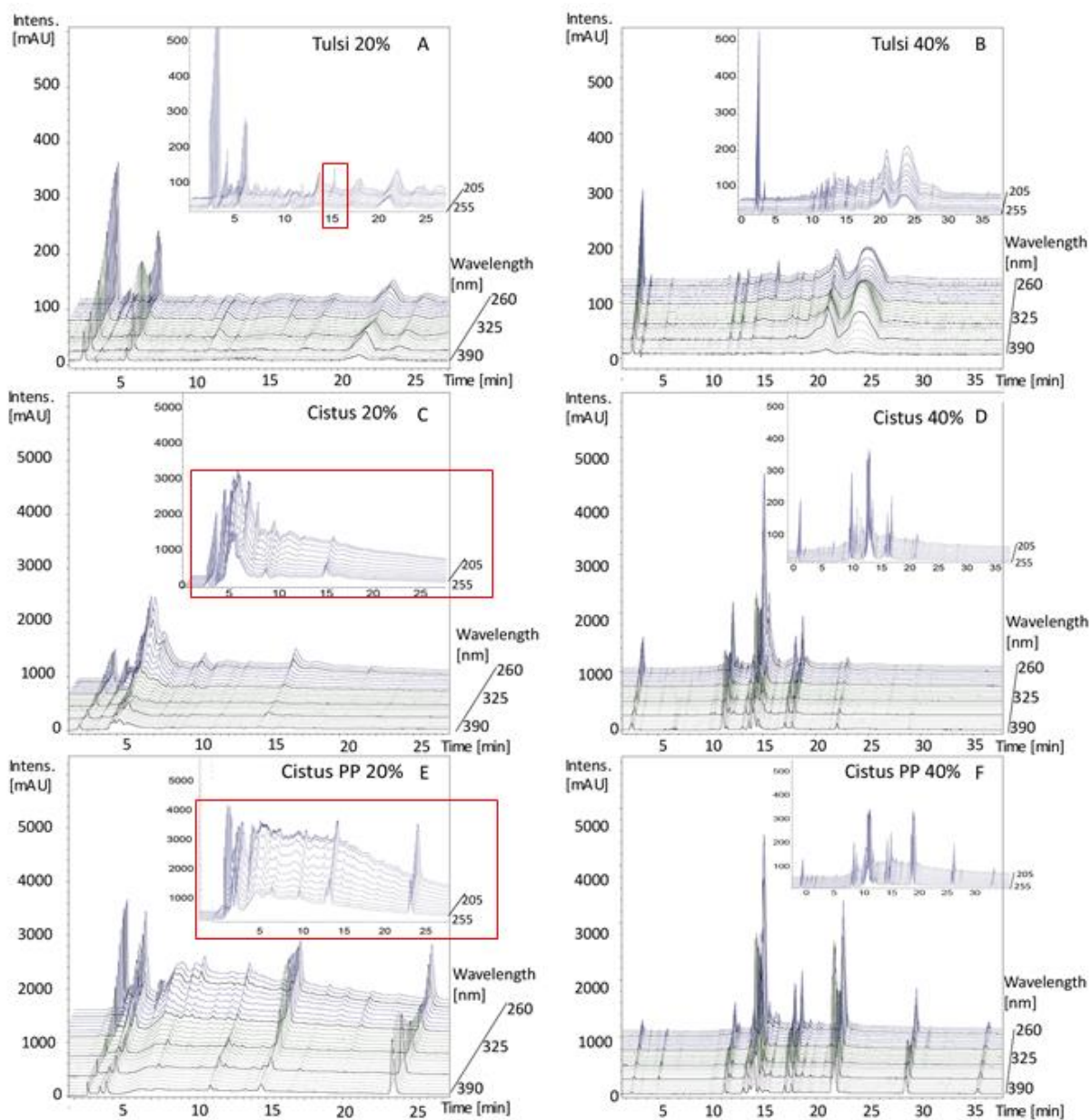
Figure 63: Overview of the wavelength from 260 nm in steps of 5 nm until 390 nm on the z-axis and an enlargement for 260 bis 205 nm. On the x-axis the retention time in [min] and on the y-axis the Intensity in [mAU] is given. A) and B) wavelength distribution for Tulsi 20% and 40% including a 10-fold enlargement for the intensity in [mAU] to a maximum of 500. B) and C) wavelength distribution for Cistus 20% and 40% with a maximum intensity of 5000 [mAU] D) and E) wavelength distribution for Cistus PP 20% and 40% with a maximum intensity of 5000 [mAU].
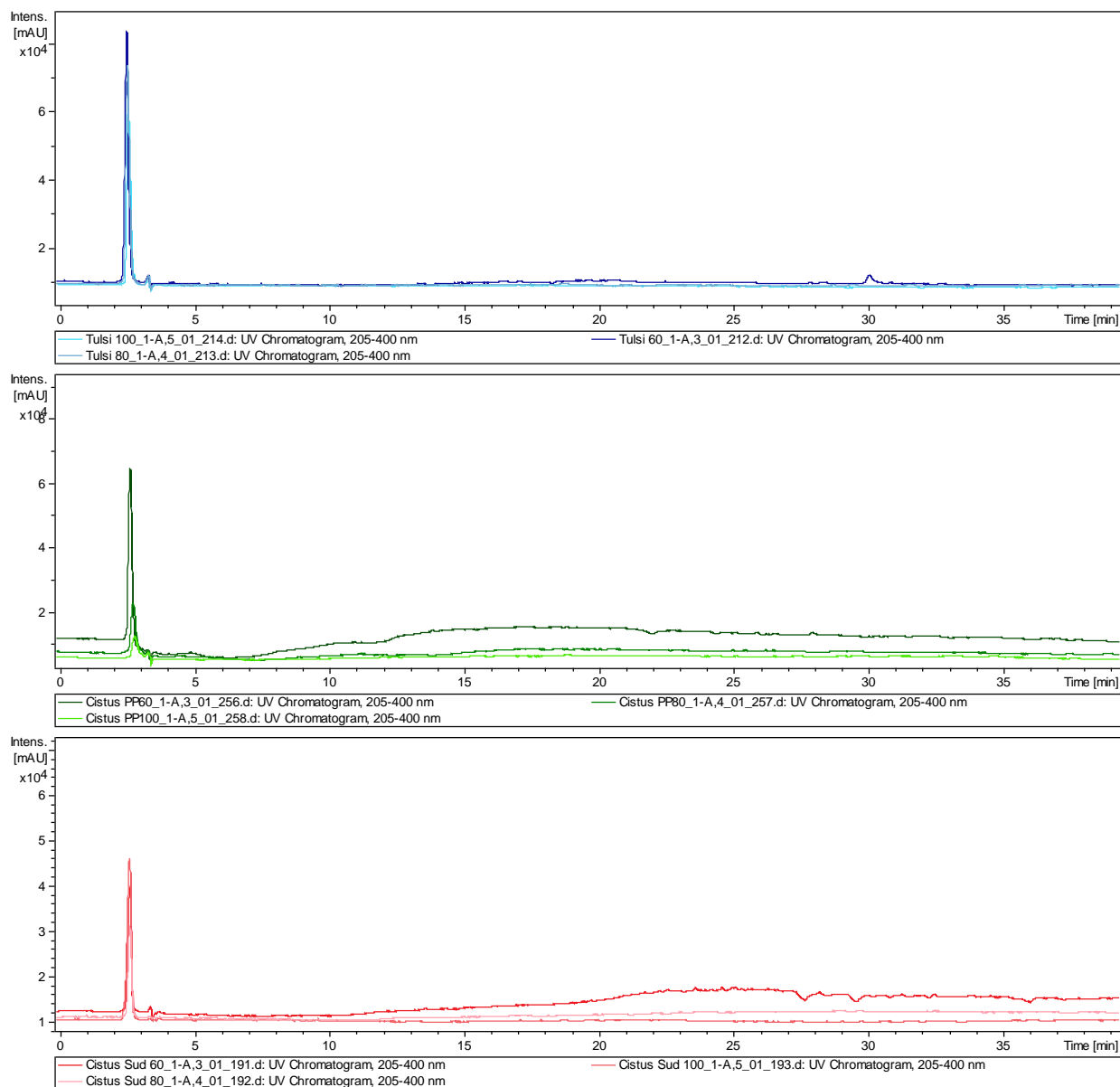
Figure 64: UV activity of the D2 60%, 80% and 100% fractions of Tulsi (blue), Cistus PP (green) and Cistus (red) at 205-400 nm.

6. Supplementary Information

6.2. Chapter 3

6.2.1. Cistus and Cistus PP HCPP results
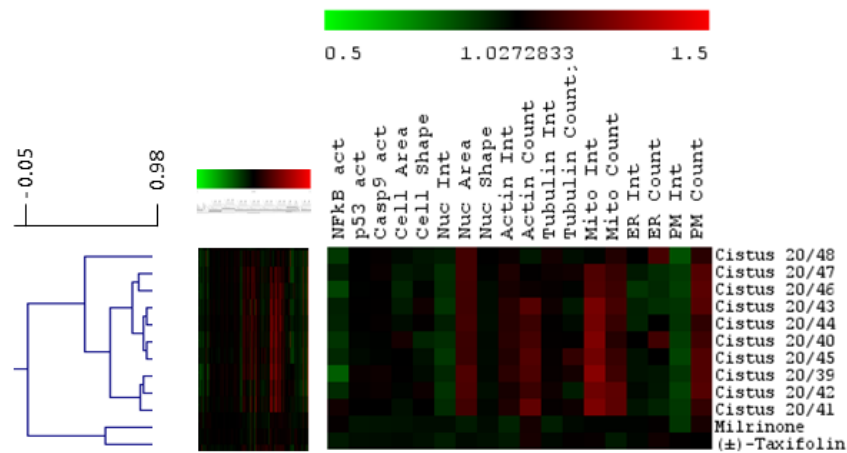
Figure 65: Cistus D2 20% hierarchical Pearson clustering of cytolological profiles from fractions 39 until 48 and Milrinone and (±)-Taxifolin. Full cytological profiles were used for the clustering (left) and 20 core markers are chosen for easier visualization of the cellular effects. The mean of n = 4 measurements normalized to the corresponding, outlier tested, control feature on the same plate are shown. Red indicate a positive and green negative deviation from the mean of untreated control cells with a value = 1. The cluster also appeared using Kendalls Tau and Spearman Rank clustering.
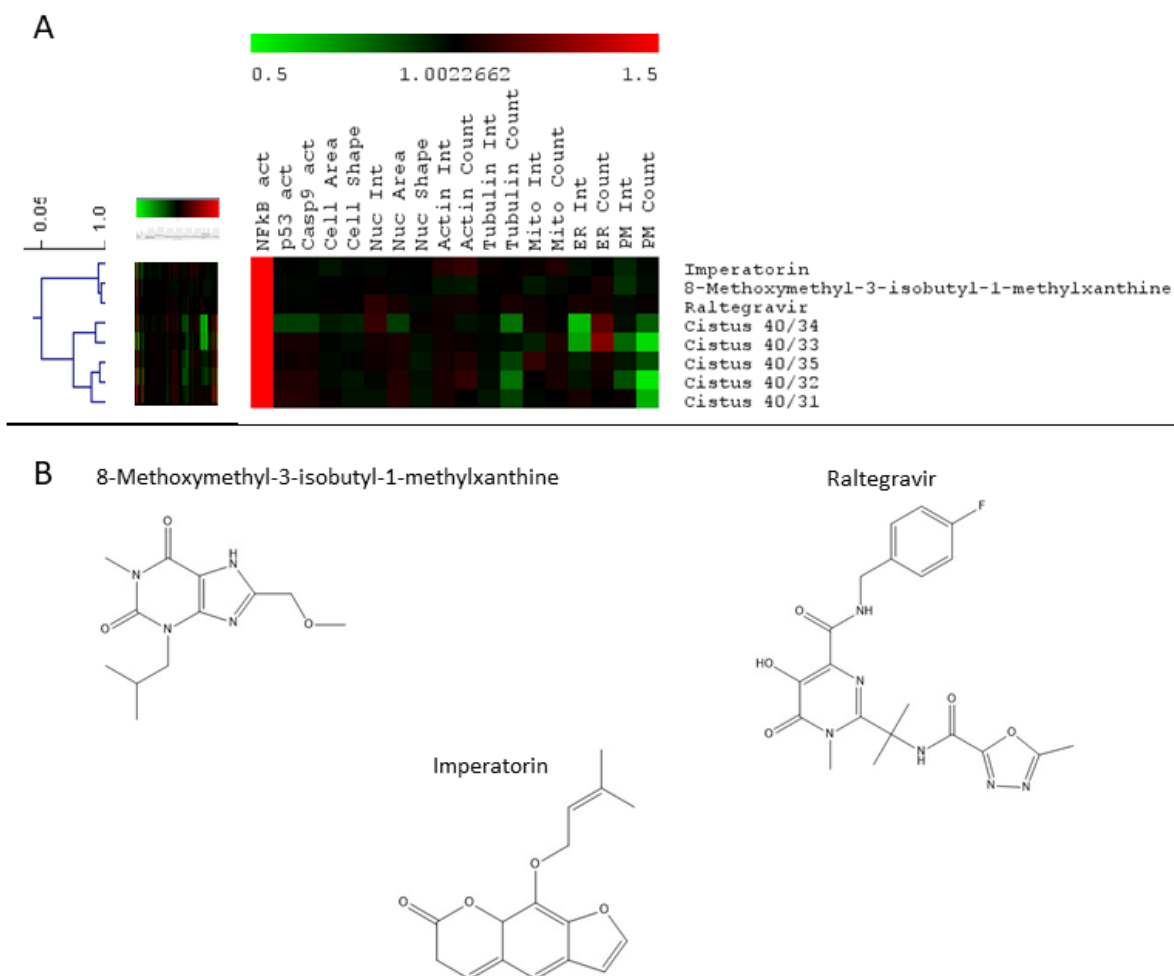
Figure 66: A) Cistus D2 40% hierarchical Spearman Rank clustering of cytolological profiles from fractions 31 until 35 and B) Imperatorin, 8-Methoxymethyl-3-isobutyl-1-methylxanthine and Raltegravir. Full cytological profiles were used for the clustering (left) and 20 core markers are chosen for easier visualization of the cellular effects. The mean of n = 4 measurements normalized to the corresponding, outlier tested, control feature on the same plate are shown. Red indicate a positive and green a negative deviation from the mean of untreated control cells with a value =1. The clustering also appeared using Pearson clustering.

6. Supplementary Information
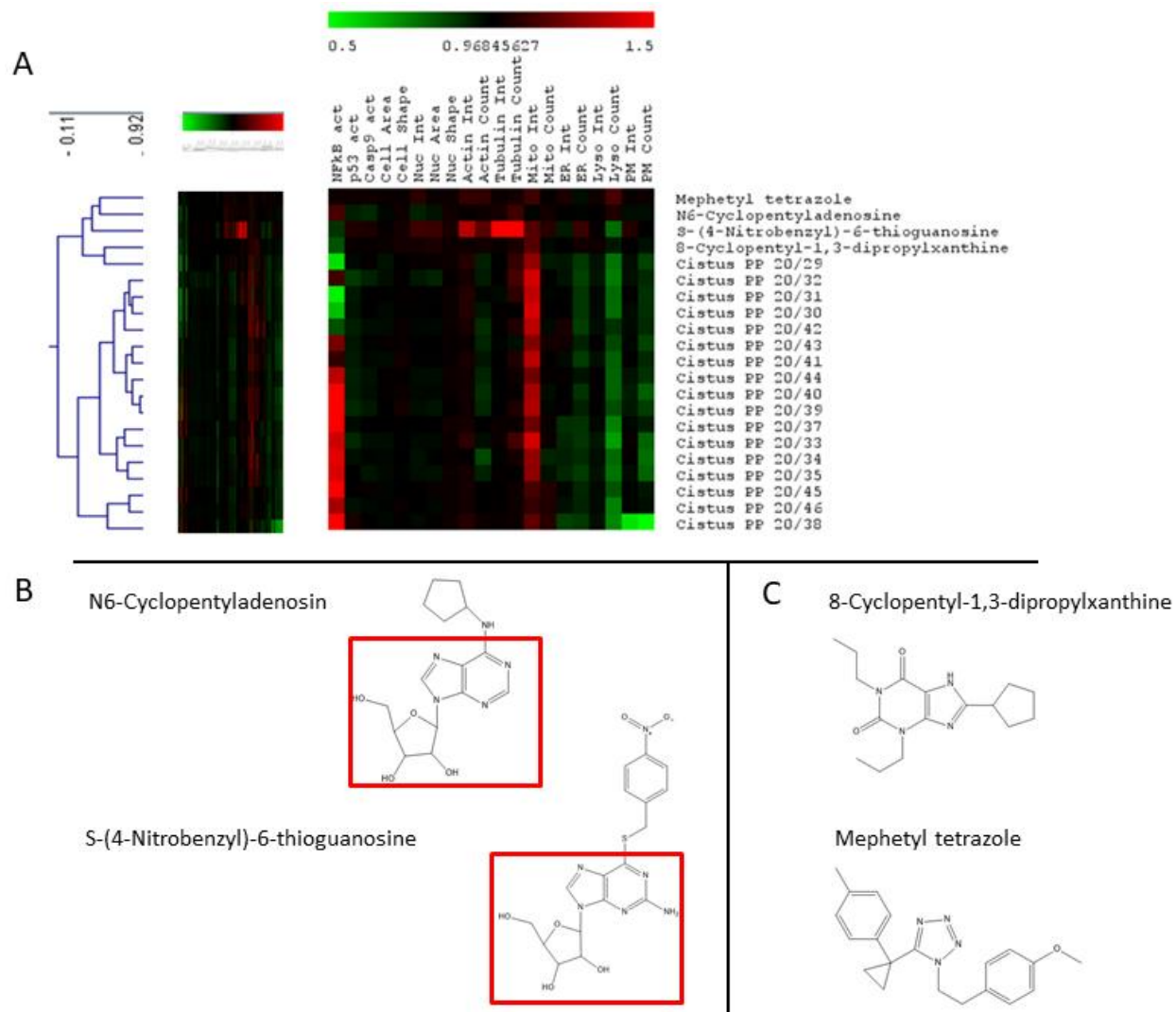
Figure 67: A) Cistus PP D2 20% fraction 29-46 hierarchical Kendalls Tau clustering of cytolological profiles with B) N6-Cyclopentyladenosin, S-(4-Nitrobenzyl)-6-thioguanosine showing a similar structural subunit and C) 8-Cyclopentyl-1,3-dipropylxanthine and Mephetyl tetrazole. Full cytological profiles were used for the clustering (left) and 20 core markers are chosen for easier visualization of the cellular effects. The mean of n = 4 measurements normalized to the corresponding, outlier tested, control feature on the same plate are shown. Red indicate positive and green a negative deviation from the mean of untreated control cells with a value =1.
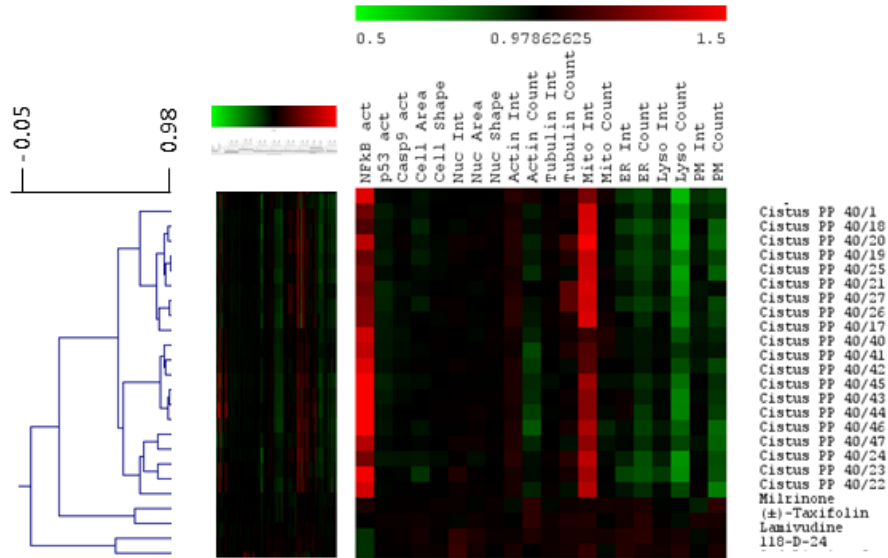
Figure 68: Cistus PP D2 40% fraction 1, 17-27,40-47 hierarchical Pearson clustering of cytolological profiles with Milrinone, (±)-Taxifolin, Lamivudine and 118-D-24. Full cytological profiles were used for the clustering (left) and 20 core markers are chosen for easier visualization of the cellular effects. The mean of n = 4 measurements normalized to the corresponding, outlier tested, control feature on the same plate are shown. Red indicate positive and green a negative deviation from the mean of untreated control cells with a value =1. The cluster also appeared using Kendalls Tau clustering.

6. Supplementary Information
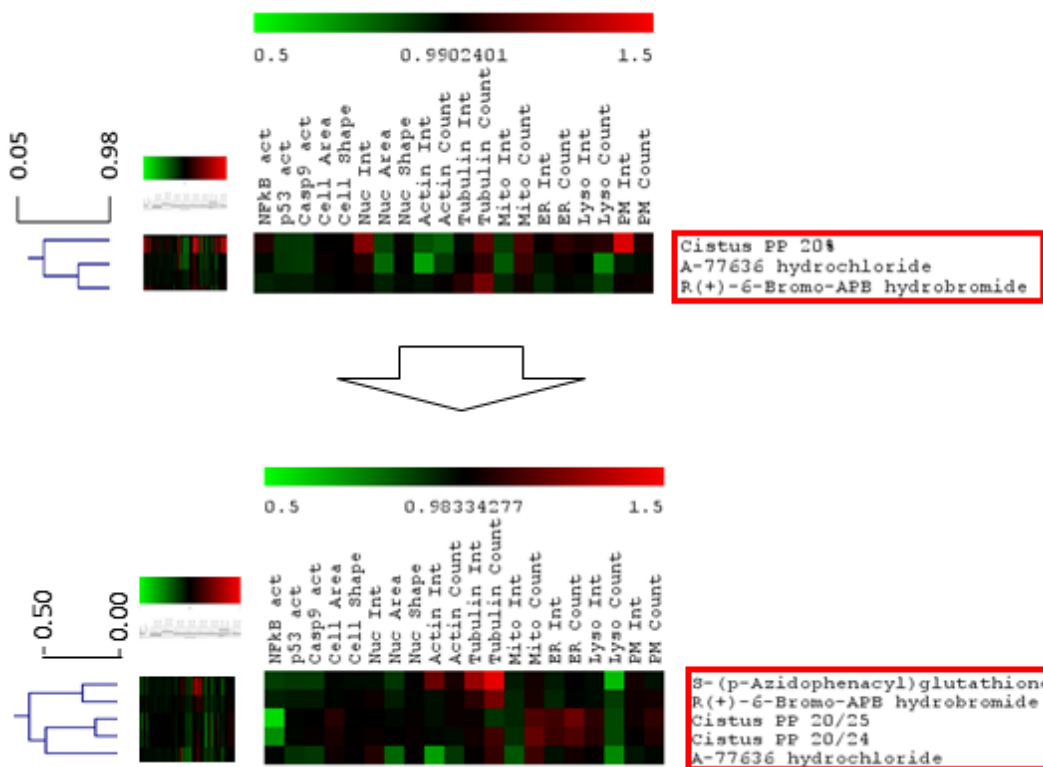
Figure 69: Cistus PP D1 20% and Cistus PP D2 20% fraction 25-26 hierarchical Pearson clustering of cytolological profiles with R(+)-6-Bromo-APB hydrobromid and A-77636 hydrochlorid. Clustering with the same standard even different concentrations appeared (Schulze et al., 2013). Full cytological profiles were used for the Pearson clustering (left) and 20 core markers are chosen for easier visualizstructureation of the cellular effects. The mean of n = 4 measurements normalized to the corresponding, outlier tested, control feature on the same plate are shown. Red indicate positive and green a negative deviation from the mean of untreated control cells with a value =1.

6.2.2. Support trees
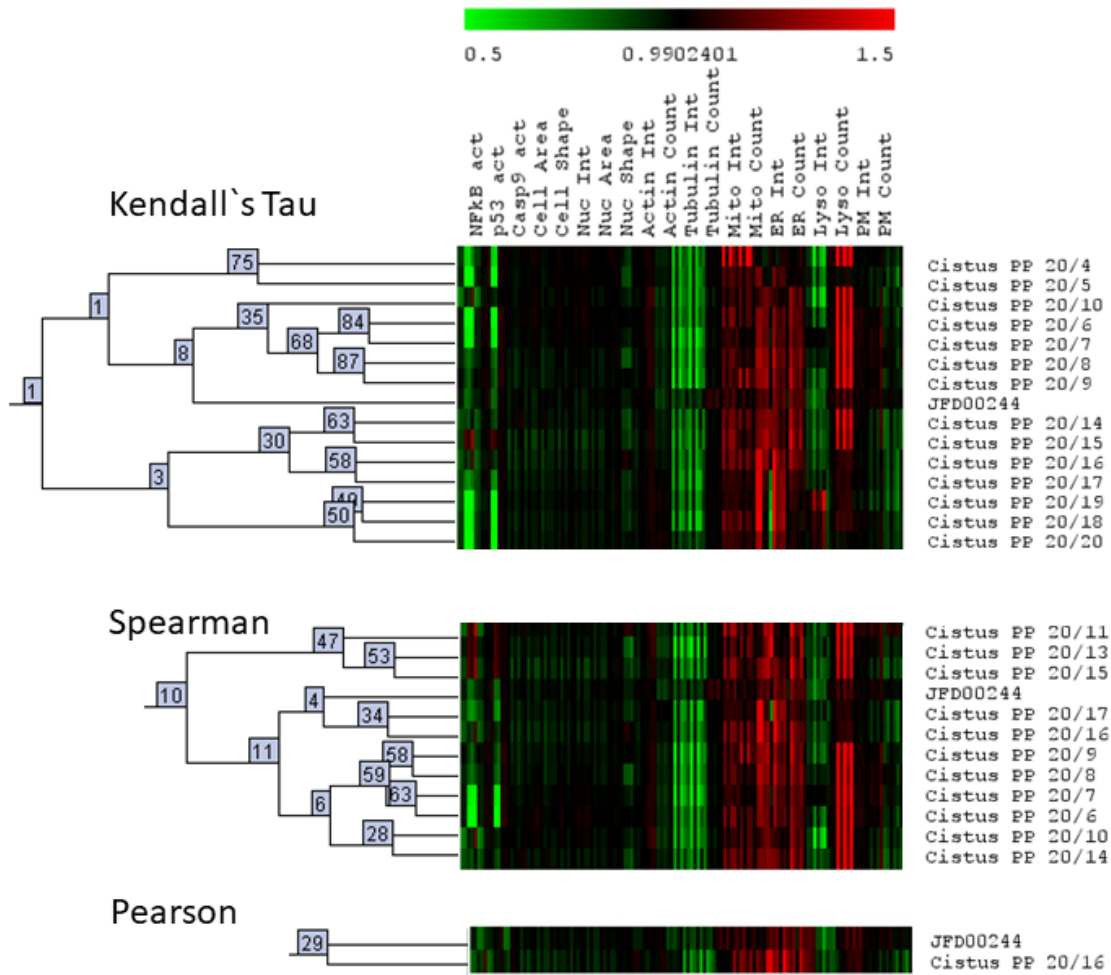
Figure 70: Support trees for the JFD00244 cluster using Kendall`s Tau, Spearman Rank and Pearson clustering.
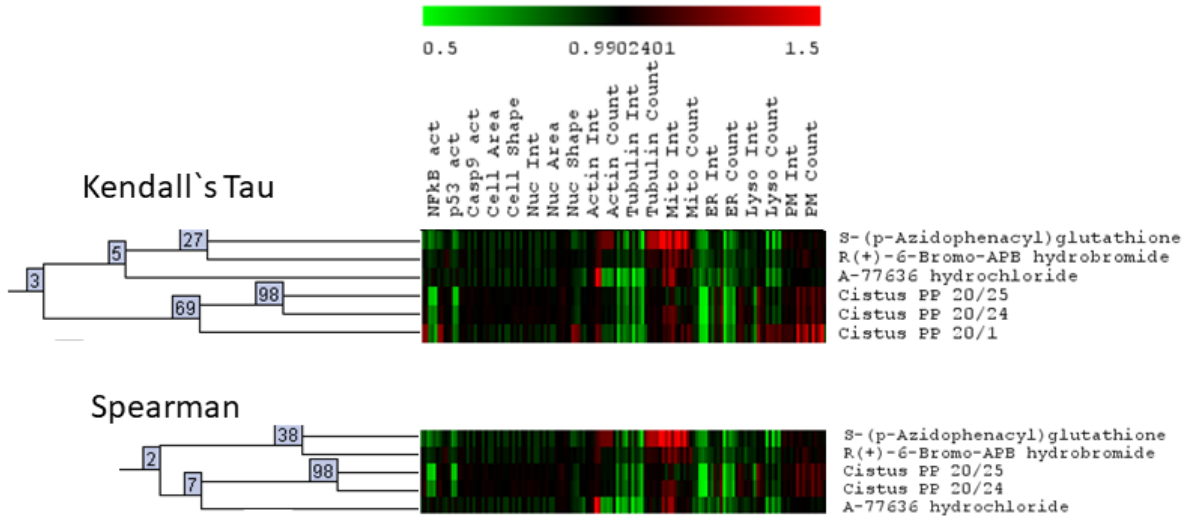
Figure 71: Support trees for the Cistus PP D1 20% 24/25 cluster using Kendall`s Tau and Spearman Rank clustering.

6.2.3. EASY-HIT results
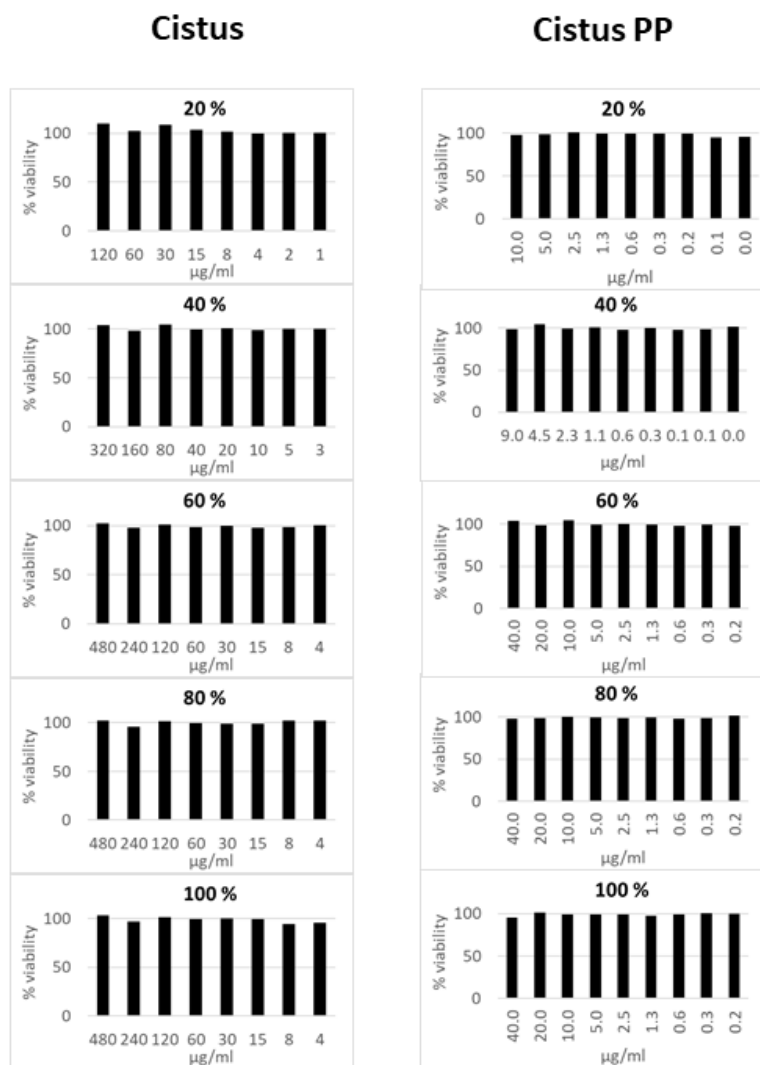
Figure 72: Overview of the cell viability tests (MTT tests) for the D1 fractions of Cistus and Cistus PP. No reduction of the metabolic process under 80% appeared.

6.2.4. Correlations using combined and submatrices

The used study design allows the investigation of different hypothesis. Therefore, the EASY-HIT and FT-ICR-MS data of Tulsi, Cistus and Cistus PP were characterized together and separately. The data of Cistus and Cistus PP were combined in one matrix and resulting submatrices, because any correlation visible in Cistus PP must be **also present in Cistus**. Figure 73 gave an overview of the different possible matrices consisting of the different D2 fractions sets e.g. 20% MeOH and 40% MeOH.

6. Supplementary Information
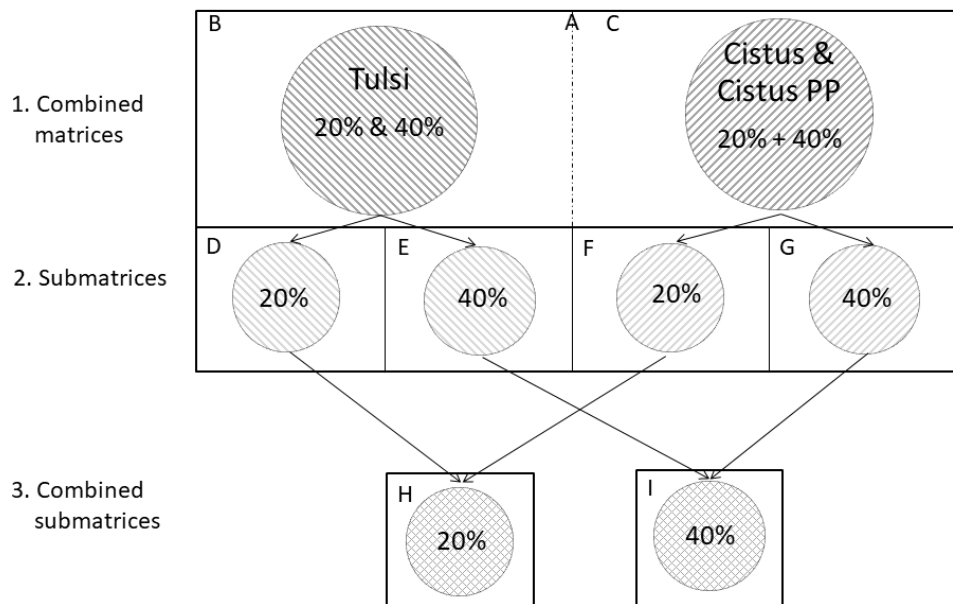
Figure 73: Overview of the different matrices and submatrices from Tulsi, Cistus and Cistus PP.

Four hypotheses can be tested with these matrices:

i) The question if there is a common class of metabolites responsible for the anti-HIV-1 activity in the two plants Tulsi and Cistus can be answered with the results of the correlation of matrix A.

ii) Matrix B and C are answering the question if there is a plant specific active metabolite class following the anti-HIV-1 distribution in the 20% and 40% MeOH fraction.

iii) The matrices D-G are dealing with the question if there are fraction specific metabolites classes responsible for the anti-HIV-1 activity.

iv) Matrix H and I are investigating, if there is a common fraction specific metabolite class in Tulsi and Cistus responsible for the anti-HIV-1 activity there.

The Spearman Rank correlation was calculated for Matrix A – I (Figure 73) according to the description in 5.1.5. The results are summed up in Figure 74. The closer r_S is to -1, the stronger the correlation. The different strength of the correlations are described using the following guide:

0.00 to -0.19	very weak
-0.20 to -0.39	weak
-0.40 to -0.59	moderate
-0.60 to -0.79	strong
-0.80 to -1.00	very strong

The correlation is significant ($\alpha = 0.001$) until $r_S \leq 0.4$ (GLASSER & WINTER, 1961).

Matrix	A	B	C	D	E	F	G	H	I
Min r_S	- 0.46	- 0.37	- 0.64	- 0.43	- 0.40	- 0.73	- 0.29	- 0.60	- 0.20
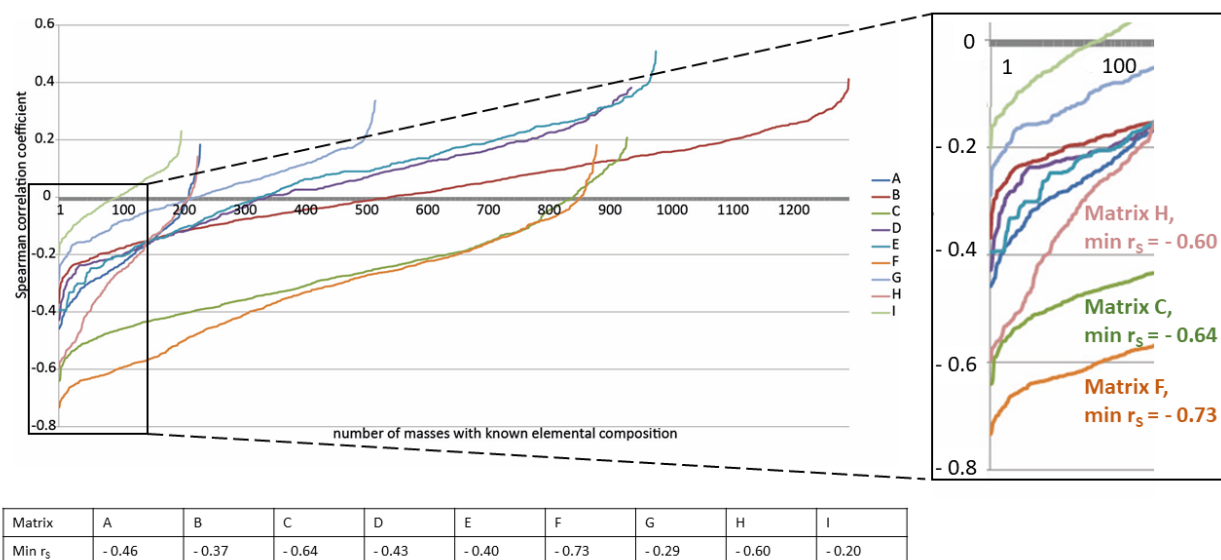
Figure 74: Overview of the Spearman Rank correlation coefficients (r_S) for Matrix A)-I) on the y-axis and the number of masses with known elemental composition (NetCalc) on the x-axis. The enlargement on the right show the matrixes with the lowest minimal r_S . The table give an overview of all minimal r_S .

The previous questions i-iv can now be answered.

i) Combining all Matrixes of Tulsi and Cistus in matrix A, only a moderate correlation (-0.46) could be seen.

ii) Only for Matrix C a strong correlation could be seen (-0.64) indicating strong Cistus specific anti-HIV-1 active metabolites

iii) The results of the submatrices D-G show, that matrix F show a strong correlation (-0.73) indicating Cistus PP 20% specific anti-HIV-1 active metabolites

iv) Common correlating metabolites in the 20% fraction of Tulsi and Cistus can be seen due to a strong to moderat (-0.60) correlation in matrix H.

As a conclusion, the matrices C, F and H provide the best correlations (Figure 74 enlargement). Here we have a specific number of metabolites correlating strong with the corresponding anti-HIV-1 distribution. The best correlation for Tulsi, as a single plant, could be seen in matrix D (-0.43). The effect on the Spearman Rank correlation coming from of the different length of the matrixes (202 to 1299 metabolites) is neglectable, as one of the shortest matrices (matrix H) show one of the best correlations and one of the longest (matrix B) only weak correlations.

6. Supplementary Information

6.2.5. MS/MS

Table 25: Overview of the data for the top 100 Spearman Rank correlations of Cistus PP 20%. r_S is the Spearman Rank coefficient with corresponding *p*-value in the next column. Followed by the max. Intensity of the m/z, the experimental (exp.) m/z, the theoretical (theo.) assigned molecular formulae using NetCalc. The error of the experimental and theoretical assigned molecular formula is<0.1 ppm.

Top 100	r_S	p-value	Max Intensity	Experimental m/z	Theo. assigned Molecular Formula	Error [pmm] exp. to theo.
1	-0.735	7.87E-18	6.09E+07	235.025	$C_{11}H_8O_6$	0.081
2	-0.728	2.12E-17	8.12E+07	431.026	$C_{19}H_{12}O_{12}$	-0.078
3	-0.706	4.89E-16	1.72E+08	417.046	$C_{19}H_{14}O_{11}$	-0.047
4	-0.699	1.14E-15	6.86E+07	363.036	$C_{16}H_{12}O_{10}$	-0.014
5	-0.699	1.24E-15	3.28E+08	469.005	$C_{21}H_{10}O_{13}$	-0.082
6	-0.692	3.04E-15	1.40E+08	375.036	$C_{17}H_{12}O_{10}$	-0.032
7	-0.691	3.25E-15	2.63E+08	361.057	$C_{17}H_{14}O_9$	0.016
8	-0.687	5.62E-15	1.16E+08	279.015	$C_{12}H_8O_8$	0.004
9	-0.687	5.62E-15	1.16E+08	293.030	$C_{13}H_{10}O_8$	-0.025
10	-0.683	8.90E-15	9.01E+07	385.057	$C_{19}H_{14}O_9$	-0.021
11	-0.682	1.01E-14	3.32E+08	377.051	$C_{17}H_{14}O_{10}$	-0.019
12	-0.676	2.08E-14	4.22E+08	489.031	$C_{21}H_{14}O_{14}$	-0.072
13	-0.675	2.32E-14	4.27E+07	403.031	$C_{18}H_{12}O_{11}$	-0.049
14	-0.668	5.35E-14	1.90E+08	443.062	$C_{21}H_{16}O_{11}$	0.037
15	-0.663	9.22E-14	9.28E+07	429.046	$C_{20}H_{14}O_{11}$	-0.063
16	-0.661	1.18E-13	4.85E+07	391.031	$C_{17}H_{12}O_{11}$	-0.034
17	-0.660	1.29E-13	1.27E+08	473.073	$C_{22}H_{18}O_{12}$	-0.008
18	-0.658	1.69E-13	1.52E+08	391.067	$C_{28}H_{16}O_{10}$	-0.032
19	-0.657	1.75E-13	2.64E+08	293.067	$C_{14}H_{14}O_7$	-0.028
20	-0.654	2.38E-13	1.00E+08	307.010	$C_{13}H_8O_9$	-0.057
21	-0.653	2.94E-13	9.75E+07	445.041	$C_{20}H_{14}O_{12}$	-0.074
22	-0.650	4.04E-13	1.26E+08	375.072	$C_{18}H_{16}O_9$	0.074
23	-0.649	4.16E-13	6.72E+07	389.051	$C_{18}H_{14}O_{10}$	0.000
24	-0.649	4.33E-13	6.74E+07	305.030	$C_{14}H_{10}O_8$	-0.065
25	-0.648	4.65E-13	2.71E+08	335.041	$C_{15}H_{12}O_9$	-0.056
26	-0.643	8.03E-13	1.67E+08	459.057	$C_{21}H_{16}O_{12}$	0.025
27	-0.643	8.38E-13	8.46E+07	389.088	$C_{19}H_{18}O_9$	0.021
28	-0.640	1.12E-12	1.40E+08	429.083	$C_{21}H_{18}O_{10}$	-0.002
29	-0.639	1.23E-12	1.20E+08	387.072	$C_{19}H_{16}O_9$	0.028
30	-0.637	1.52E-12	1.12E+08	489.067	$C_{22}H_{18}O_{13}$	0.019
31	-0.635	1.78E-12	6.00E+07	399.072	$C_{20}H_{16}O_9$	0.044
32	-0.635	1.79E-12	8.82E+07	321.025	$C_{14}H_{10}O_9$	-0.042
33	-0.634	1.98E-12	9.71E+07	321.062	$C_{15}H_{14}O_8$	-0.089

34	-0.634	2.00E-12	8.19E+07	525.104	$C_{26}H_{22}O_{12}$	-0.066
35	-0.633	2.32E-12	1.29E+08	417.083	$C_{20}H_{18}O_{10}$	0.018
36	-0.633	2.32E-12	2.48E+08	307.046	$C_{14}H_{12}O_8$	-0.065
37	-0.631	2.71E-12	1.67E+08	477.068	$C_{21}H_{18}O_{13}$	-0.066
38	-0.629	3.25E-12	3.94E+08	345.062	$C_{17}H_{14}O_8$	-0.026
39	-0.629	3.36E-12	1.00E+08	349.020	$C_{15}H_{10}O_{10}$	-0.061
40	-0.626	4.40E-12	4.29E+07	341.067	$C_{18}H_{14}O_7$	-0.021
41	-0.626	4.48E-12	5.40E+08	387.036	$C_{18}H_{12}O_{10}$	-0.028
42	-0.626	4.50E-12	3.69E+08	457.078	$C_{22}H_{18}O_{11}$	-0.035
43	-0.625	4.83E-12	8.43E+07	405.046	$C_{18}H_{14}O_{11}$	-0.018
44	-0.623	5.81E-12	6.00E+07	473.036	$C_{21}H_{14}O_{13}$	-0.094
45	-0.622	6.53E-12	4.06E+08	427.067	$C_{21}H_{16}O_{10}$	-0.015
46	-0.621	7.07E-12	2.23E+08	377.088	$C_{18}H_{18}O_9$	0.003
47	-0.620	7.99E-12	2.07E+07	219.030	$C_{11}H_8O_5$	0.091
48	-0.620	8.36E-12	4.78E+07	279.051	$C_{13}H_{12}O_7$	0.003
49	-0.618	9.74E-12	3.86E+07	261.040	$C_{13}H_{10}O_6$	0.046
50	-0.616	1.23E-11	2.35E+08	291.051	$C_{14}H_{12}O_7$	-0.029
51	-0.615	1.35E-11	1.50E+07	313.035	$C_{16}H_{10}O_7$	-0.085
52	-0.613	1.53E-11	2.68E+08	207.030	$C_{10}H_8O_5$	-0.008
53	-0.610	2.03E-11	2.21E+07	379.031	$C_{16}H_{12}O_{11}$	-0.017
54	-0.610	2.05E-11	5.64E+07	501.068	$C_{23}H_{18}O_{13}$	-0.073
55	-0.608	2.53E-11	1.37E+09	295.010	$C_{12}H_8O_9$	-0.082
56	-0.606	3.17E-11	2.58E+08	457.041	$C_{21}H_{14}O_{12}$	-0.033
57	-0.605	3.19E-11	1.34E+08	483.093	$C_{24}H_{20}O_{11}$	-0.050
58	-0.605	3.31E-11	1.55E+08	399.036	$C_{19}H_{12}O_{10}$	-0.007
59	-0.605	3.41E-11	6.38E+07	357.062	$C_{18}H_{14}O_8$	0.020
60	-0.602	4.51E-11	6.24E+07	447.057	$C_{20}H_{16}O_{12}$	-0.095
61	-0.601	4.85E-11	2.66E+08	475.052	$C_{21}H_{16}O_{13}$	-0.038
62	-0.599	5.51E-11	6.19E+07	405.083	$C_{19}H_{18}O_{10}$	-0.019
63	-0.596	7.33E-11	9.63E+06	315.087	$C_{17}H_{16}O_6$	-0.058
64	-0.596	7.59E-11	4.55E+07	393.046	$C_{17}H_{14}O_{11}$	-0.090
65	-0.595	7.81E-11	1.87E+07	381.062	$C_{20}H_{14}O_8$	-0.011
66	-0.595	7.87E-11	3.48E+08	273.040	$C_{14}H_{10}O_6$	0.040
67	-0.595	8.44E-11	2.42E+07	249.040	$C_{12}H_{10}O_6$	0.061
68	-0.591	1.16E-10	3.17E+08	485.073	$C_{23}H_{18}O_{12}$	-0.047
69	-0.589	1.36E-10	6.79E+07	289.035	$C_{14}H_{10}O_7$	0.025
70	-0.588	1.50E-10	2.25E+08	443.098	$C_{22}H_{20}O_{10}$	0.023
71	-0.587	1.62E-10	1.06E+08	407.062	$C_{18}H_{16}O_{11}$	-0.023
72	-0.587	1.68E-10	3.88E+07	357.098	$C_{19}H_{18}O_7$	-0.005
73	-0.586	1.75E-10	1.39E+08	453.119	$C_{24}H_{22}O_9$	-0.071
74	-0.585	1.87E-10	6.01E+07	401.088	$C_{20}H_{18}O_9$	0.017

6. Supplementary Information

75	-0.585	1.94E-10	3.46E+07	455.098	$C_{23}H_{20}O_{10}$	-0.098
76	-0.583	2.23E-10	6.30E+07	459.093	$C_{22}H_{20}O_{11}$	0.083
77	-0.582	2.56E-10	8.19E+07	421.078	$C_{19}H_{18}O_{11}$	-0.070
78	-0.581	2.66E-10	6.96E+07	371.041	$C_{18}H_{12}O_9$	-0.020
79	-0.578	3.38E-10	1.05E+08	251.020	$C_{11}H_8O_7$	0.062
80	-0.577	3.93E-10	4.93E+07	343.082	$C_{18}H_{16}O_7$	-0.013
81	-0.575	4.56E-10	7.99E+07	277.035	$C_{13}H_{10}O_7$	0.020
82	-0.573	5.14E-10	5.45E+07	351.072	$C_{16}H_{16}O_9$	-0.024
83	-0.572	5.71E-10	4.60E+07	435.057	$C_{19}H_{16}O_{12}$	-0.026
84	-0.572	5.92E-10	7.15E+07	351.036	$C_{15}H_{12}O_{10}$	-0.072
85	-0.571	6.02E-10	4.58E+07	193.014	$C_9H_6O_5$	-0.100
86	-0.568	8.21E-10	2.93E+07	567.114	$C_{28}H_{24}O_{13}$	-0.068
87	-0.567	8.83E-10	3.57E+07	329.030	$C_{16}H_{10}O_8$	-0.038
88	-0.566	9.37E-10	9.54E+07	509.109	$C_{26}H_{22}O_{11}$	-0.003
89	-0.565	1.00E-09	1.48E+08	463.052	$C_{20}H_{16}O_{13}$	-0.080
90	-0.564	1.13E-09	2.87E+08	623.104	$C_{30}H_{24}O_{15}$	-0.093
91	-0.563	1.16E-09	5.92E+07	503.083	$C_{23}H_{20}O_{13}$	0.093
92	-0.561	1.38E-09	8.18E+07	453.046	$C_{22}H_{14}O_{11}$	-0.037
93	-0.560	1.46E-09	5.18E+07	371.077	$C_{19}H_{16}O_8$	0.030
94	-0.559	1.64E-09	2.47E+07	421.041	$C_{18}H_{14}O_{12}$	-0.062
95	-0.558	1.69E-09	8.70E+07	373.093	$C_{19}H_{18}O_8$	0.043
96	-0.558	1.70E-09	3.36E+07	417.119	$C_{21}H_{22}O_9$	0.018
97	-0.558	1.76E-09	1.05E+08	503.047	$C_{22}H_{16}O_{14}$	-0.047
98	-0.558	1.79E-09	7.65E+07	471.093	$C_{23}H_{20}O_{11}$	-0.038
99	-0.558	1.80E-09	3.84E+08	427.103	$C_{22}H_{20}O_9$	-0.020
100	-0.557	1.83E-09	1.55E+08	309.025	$C_{13}H_{10}O_9$	-0.069

6.2.6. Retro synthesis
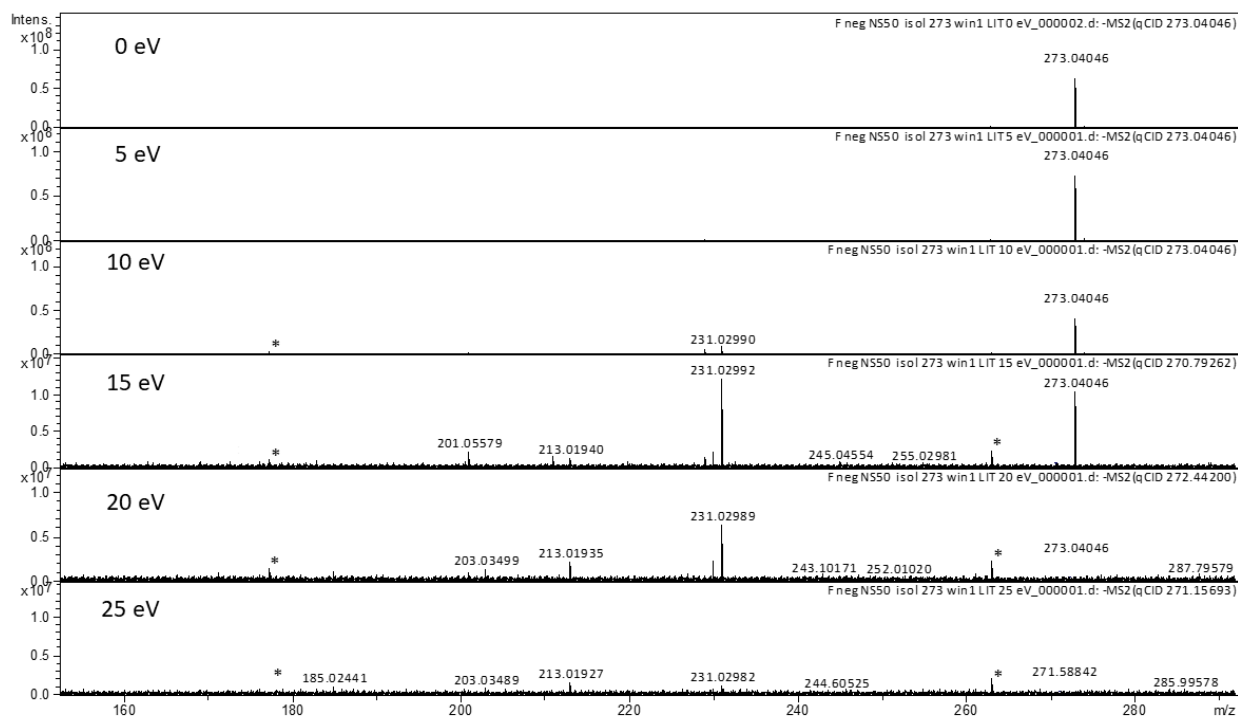
Figure 75: Fragmentation of the m/z 273 fraction 8 (D2 20%) using (-) ESI-FT-ICR-MS/MS ultrahigh resolution mass spectrometry. From top to the bottom 5 eV, 10 eV, 15 eV, 20 eV and 25 eV collision energies are applied to the isolated precursor ion. *= elec. Noise.

6.2.7. Synthesis of 4-(6,7- dihydroxynaphthyl)-2,4-dioxobutanoic acid

6.2.7.1. Attachment of protective groups
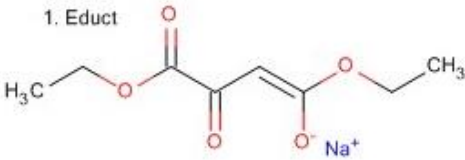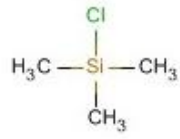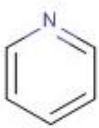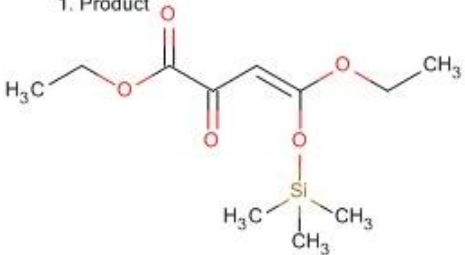
Previous experiments revealed the necessity of a protection group to control the attachment to the 2,3-Dimethoxynaphthalen backbone. Because of that, a trimethylsilylation of the Na-oxaloacetate-dimethylester educt took place.

6.2.7.1.1. Trimethylsilylation of Na-oxaloacetate-dimethylester

A schematic overview of the used conditions is presented in Table 26.

6. Supplementary Information

Table 26: Schematic overview of the trimethylsilylation of Na-oxaloacetate-dimethylester.

1. Educt: Na-oxaloacetate dimethylester (2 g, 9.5 mmol) 2. Educt: Trimethylsilyl chloride (1.41 ml, 11 mmol) 3. Educt: Pyridine (0.77 ml, 9.5 mmol)		1. Product: Na-oxaloacetate-dimethylester sililated (2.69 g, 9.5 mmol)

A well dried 100 ml round bottom flask was needed for this reaction and moister should be avoided at any time of the procedure. 2 g of Na-oxaloacetate dimethylester was added to 11.42 ml trimethylsilyl chloride in tetrahydrofuran (THF) solution (1M) in an ice bath. 21.42 ml THF was added to the mixture under rotation. In a next step 0.77 ml pyridine was added to the mixture under continues rotation to accelerate the trimethylsilylation reaction. Pyridine acts as an acid scavenger to produce pyridinium chloride. The reaction time was 30 min in an ice bath and it was further stirred for additional 3h at 55°C. The mixture cool down at room temperature and the THF solvent was removed under vacuum for 1 h until 10% of the THF was remaining in the flask to prevent the production of possible peroxides.

For the purification of the product 50 ml of 10% sodium hydroxide aqueous solution was added to the round flask and transferred after complete solution to a separating funnel. A schematic overview of the products and educts in the organic and water phase is presented in Figure 76.
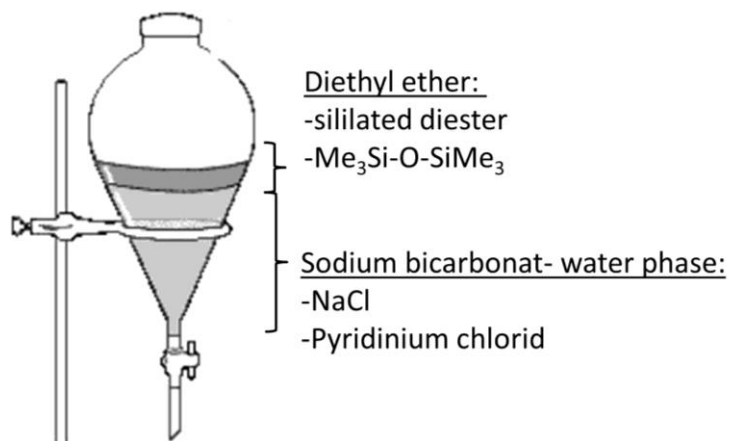
Figure 76: Schemativ overview of the products solved in the organic pase (diethyl ether) and the water phase (sodium bicarbonat).

In the separating funnel water was reacting with the excess of Me_3-SiCl and produced Me_3-Si-O-Si-Me_3 and HCl. Me_3-Si-O-Si-Me_3 has a boiling point of 100°C and is very hydrophobic. It retains in the organic phase. HCl reacts with sodium hydroxide under the production of NaCl. 25 ml diethyl ether is added and the funnel was shacked for 5 min. The water phase, which contains NaCl and pyridinium chloride was discarded and the organic phase was separated and kept for further experiments. It contained silylated diester together with Me_3-Si-O-Si-Me_3 and slightly remaining pyridine. Pyridine boils at 115°C and could like THF (boiling point 66°C) and Me_3-Si-O-Si-Me_3 (boiling point 100°C) be removed through vacuum. The boiling point of the final product is expected to be in the range 200-250°C. and could therefore be cleaned.

6.2.7.1.2.   alkaline hydrolysis of the silylated diester

In a next step the silylated diester need to be hydrolyzed. A schematic overview of the performed hydrolysis is displayed in Table 27.

6. Supplementary Information

Table 27: Schematic overview of the alkaline hydrolysis of 2 g silylated diester.

1. Educt: Diethyl oxalate sodium salt sililated in THF (2g, 7.7 mmol) 2. Educt: Natriumhydroxyde (7.7 mmol)	ethanol	1. + 2. Product: Monoethyl oxalate sodium salt sililated

A 4 ml (7.7 mmol) solution of sodium hydroxide was prepared in water-methanol (50:50). Afterwards a 16 ml solution of 2g silylated diester (7.7 mmol) in THF was transferred in a 50 ml round flask and 4 ml sodium hydroxide solution was added. The reaction mixture was stirring at room temperature for 1h and was left unstirred in a closed flask overnight. As a result, salt precipitation could be observed overnight. The supernatant contains the wanted product (silylated monoester). The supernatant which contains THF was pumped out under N_2 flow and the remaining product yield was determined (1.0601g = 59.61%).

6.2.7.2. Upscaling: Attachment of protective groups for 20 g Na-oxaloacetate-dimethylester

After successfully synthesizing 1.0601 g of the silylated monoester, the synthesis procedure was upscaled from 2 g Na-oxaloacetate-dimethylester to 20 g Na-oxaloacetate-dimethylester as more of the produced silylated monoester was needed for further synthesis steps.

6.2.7.2.1.   Trimethylsilylation of Na-oxaloacetate-dimethylester

A Well dried 250 ml round bottom flask was needed and moister should be avoided at any time of the synthesis. 20 g of Na-oxaloacetate-dimethylester was added to 114.2 ml trimethylsilyl chloride in THF solution (1M) in ice bath. 40 ml THF was added to the mixture under rotation. In a next step 7.7 ml pyridine was added to the mixture under continues rotation to accelerate the trimethylsilylation reaction. Pyridine acts as an acid scavenger to produce pyridinium chloride. The reaction time was 30 min in an ice bath. Afterwards, the mixture was further stirred for additional 3 h at 55°C. The mixture cool down at room temperature and THF was removed under vacuum. In total 73 mL THF could be pumped out from the flask in 2.5 h. A magnetic steer bar and boiling stones were utilized for vacuum pumping. On the next day, the flask was heated to 50°C and vacuum pumping took place for additional one hour. Due to the high viscosity of the product, only 13mL additional THF could be pumped out under vacuum in 1 hour. The remaining THF in the flask was kept there to prevent the production of possible peroxides.

6.2.7.2.2.   alkaline hydrolysis of the silylated diester

To remove possible organic peroxides that may be formed during the previously described silylation reaction, 200 ml (10%) acqeous NaOH solution was added to the viscous product at continuous stirring for one hour, to help decomposing possible organic peroxides. The mixture was then transferred to a 500 ml funnel and was shaken with 100 ml ether. The ether phase was separated. Subsequently, 300 ml NaOH solution was added to wash the flask from the remaining viscous silylated diester, transferred to the funnel and shaked with 100 ml ether. Both ether phases were collected. Mass spectrometric analysis showed, that the NaOH solution did not only served to decompose possibly formed organic peroxides, but also helped to simultaneously hydrolyse the viscous silylated diester to form silylated monoester compound. The ether phase was dried under N_2 flow in a 50°C heated water bath to yield the desired silylated monoester.

6.2.7.2.3.   Further syntheses using silylated monoesters
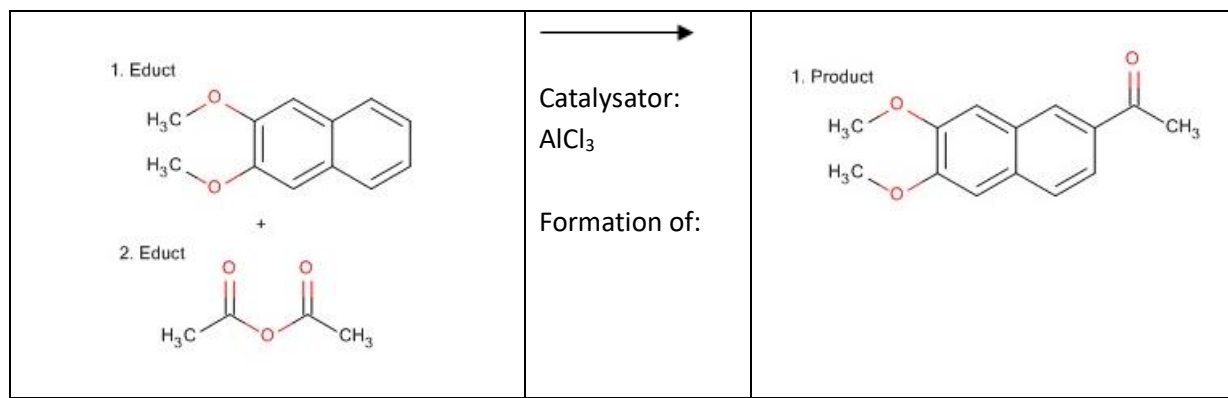
For the subsequent synthesis the silylated monoester was used. Due to time issues only the first steps until 6.2.7.2.4 could be performed. The subsequent synthesis steps were successfully performed prior to the experiment using a training substance as a proof of principle for the designed synthesis pathway. The synthesis was successfully performed with a training substance (educt: Trans-Cinnemanic acid and Na-oxaloacetate-dimethylester (Ianni & Waldvogel, 2006) an therefore was adapted to the 4-(6,7-dihydroxynaphthyl)-2,4-dioxobutanoic acid syntheses. The designed synthesis pathway is summarized from chapter 6.2.7.2.4 to 6.2.7.2.8.

6.2.7.2.4.   Friedel-Crafts Acetylation

Using the protocol described in (Baddeley, 1949; Dowdy, Gore, & Waters, 1991; Gore, Thadani, & Thorburn, 1968; Naito et al., 1991), an acetylation from 2,3-Dimethoxynaphthalen using acetic anhydride was performed. The schematic overview is displayed in Table 28.

6. Supplementary Information

Table 28: Schematic overview of the acetylation from 2,3-Dimethoxynaphthalen.
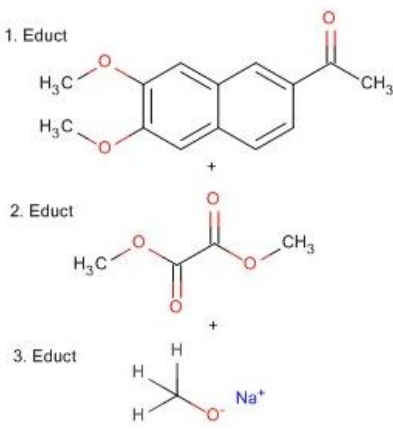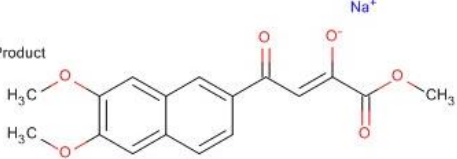
1. Educt: 2,3-Dimethoxynaphthalen (2 g, 10.6 mmol) 2. Educt: acetic anhydride (1.64 ml, 17.4 mmol)	AlCl₃ (5.57 g, 41.8 mmol)	1. Product: 6-Acetyl-2,3-Dimethoxynaphthalene (2.44 g, 10.6 mmol)

AlCl₃ was added in portions to a solution of 2,3-Dimethoxynaphthalen and acetic anhydride in Nitrobenzene at 0°C. The mixture was stirred at 0°C for 1h and poured (infused) into ice water. The mixture was extracted with diethyl ether and the extract was washed with water, aqueous Sodium bicarbonate, brine and dried over magnesium sulfate. After evaporation of the solvent under N_2 flow the product was collected by precipitation from Ethylacetate/hexane

6.2.7.2.5.   Oxalylation

After the successful acetylation of 2,3-Dimethoxynaphthalen a subsequent oxalylation using the protocol described in (Maurin, Bailly, & Cotelle, 2004) need to take place. Two different silylated monoesters are the result of the previous synthesis steps (6.2.7.2.2). They need to be separated before further synthesis steps can take place. It was not possible to separate the mixture in the remaining time. Different HPLC methods were performed but more investigation in the adequate purification of the mixture need to be invested. The following synthesis was validated by processing it with the training substances and therefore is displayed in the next chapters. As for the proof of principle synthesis no protection groups were necessary, they need to be additionally implemented and the intermediate synthesis steps validated when going on with the synthesis and are therefore not shown here. The schematic overview of the theoretically following oxalylation is displayed in Table 29.

Table 29: Schematic overview of the 6-Acetyl-2,3-Dimethoxynaphthalene oxalylation.

	diethyl ether	
1. Educt: 6-Acetyl-2,3-Dimethoxynaphthalene (10 mmol) 2. Educt: Monoethyl oxalate sodium salt sililated (10 mmol) 3. Educt: sodium methoxide (12 mmol)		1. Product: Sodium ketoenolate ester

Sodium (12 mmol) need to be added slowly in methanol (6 ml) at -5°C to give a 2 M solution of sodium methoxide. 6-Acetyl-2,3-Dimethoxynaphthalene (10 mmol) and Monoethyl oxalate sodium salt sililated (10 mmol) need to be dissolved in dried diethyl ether (10 ml) and the freshly prepared solution of sodium methoxide need to be slowly added. The mixture should stir overnight, the solid filtered off, washed with methanol, diethyl ether and dried. Sodium ketoenolate esters should be obtained in yields ranging from 75 to 92%.

6.2.7.2.6. Acidic hydrolysis

After successfully oxylalation a subsequent acidic hydrolysis (procedure described in (Maurin et al., 2004) need to took place. A schematic overview of the syntheses is given in Table 30.

6. Supplementary Information

Table 30: Schematic overview of the acidic hydrolysis from sodium ketoenolate esters.

1. Educt: Sodium ketoenolate ester (1 g) 2. Educt: acetic acid (until pH 3-4)		1. Product: Ketoenol esters

Sodium ketoenolate ester (1 g) need to be solved in water (about 100 ml) at room temperature for 1 h. Then the solution need to be acidified by adding acetic acid to reach pH 3–4 and kept at 0°C for 1 h. The precipitate should be filtered off, washed with water and dried. Ketoenol esters should be obtained in yields ranging from 70 to 90%.

6.2.7.2.7.   Demethylation

After successfully acidic hydrolysis a demethylation on the naphthalene backbone is necessary. The procedure is adapted from Naito, Sugiura et al. 1991 and a schematic overview is presented in Table 31.

Table 31: Schematic overview of the demethylation at the naphthalene backbone.
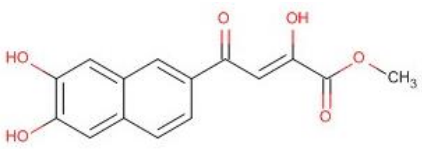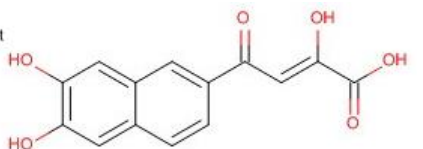
1. Educt: ketoenol ester (10 ml) 2. Educt: BBr$_3$ in CH$_2$Cl$_2$ (2,6 ml, 1M)		1. Product: Ketoenol ester

140

1M BBr$_3$ in CH$_2$Cl$_2$ (2.6 ml) need to be added dropwise to a solution of 200 mg of the ketoenol ester in dry CH$_2$Cl$_2$ (10 ml) at -70°C. After the addition, the reaction mixture need to stir at 0°C – 40°C for 10min and poured into ice water. The resultant mixture need to be extracted with AcOEt and the extract should be washed with water, brine and afterwards dried over MgSO$_4$. After evaporation of the solvent, the residue should be chromatographed on silica gel with AcOEt-hexane and recrystallized from CHCL$_3$-hexane (56% yield).

6.2.7.2.8. Hydrolysis of the ester group

As a last step the ester group of the Ketoenol ester need to be hydrolyzed. The procedure is adapted from (Maurin et al., 2004) and schematically displayed in Table 32.

Table 32: Schematic overview of the ester group hydrolysis.

1. Educt: Ketoenol ester (1 mmol) 2. Educt: HCl (25 ml, 1M)		1. Product: 4-(6,7-dihydroxynaphthyl)-2,4-dioxobutanoic acid

Ketoenol ester (1 mmol) need to be dissolved in freshly distilled 1,4-dioxane (25 ml) and 1M HCl (25 ml) need to be added. The mixture should reflux for 4 h. Removal of the solvents under reduced pressure will give a solid, which should be washed with water, CH$_2$Cl$_2$ and dried. The resulting ketoenol acid will then be recrystallized in the appropriate solvent.

6.2.8. MS/MS of 4-(6,7-dihydroxy naphthyl)- 2,4-dioxobutanoic acid
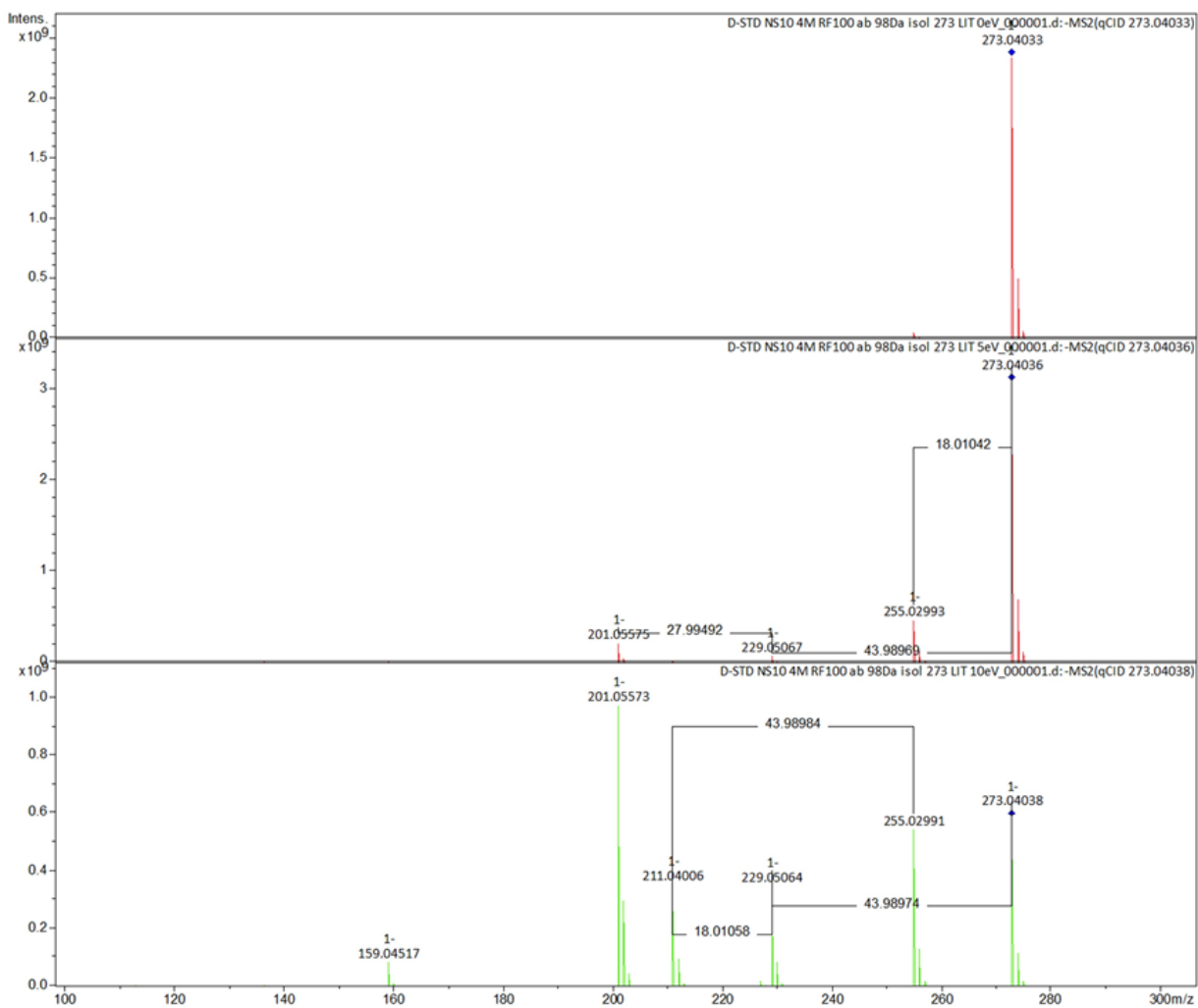
Figure 77: Fragmentation of 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid standard using (-) ESI-FT-ICR-MS/MS ultrahigh resolution mass spectrometry. From top to the bottom 0eV, 5eV and 10eV, collision energies are applied to the isolated precursor ion.
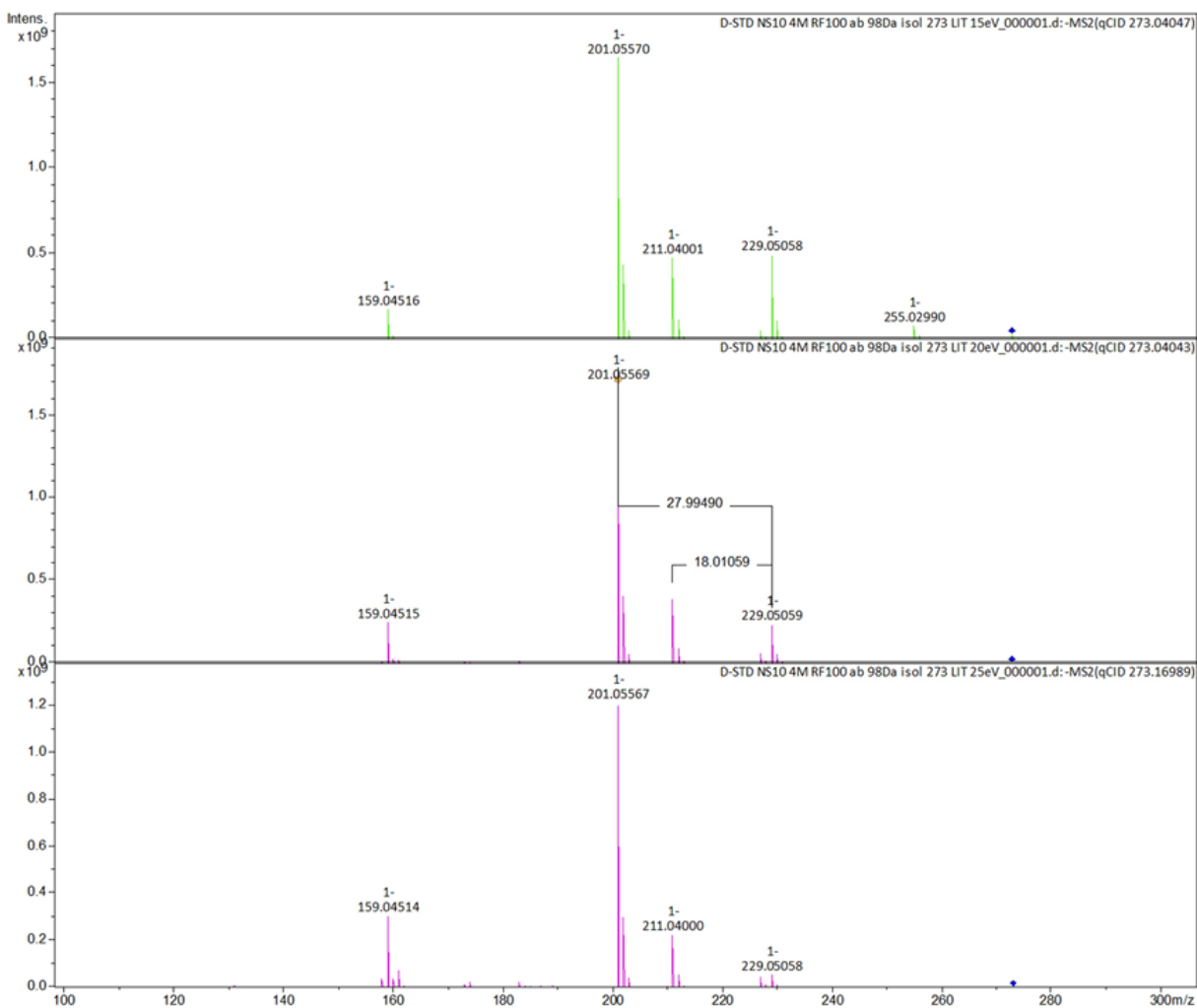
Figure 78: Fragmentation of 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid standard using (-) ESI-FT-ICR-MS/MS ultrahigh resolution mass spectrometry. From top to the bottom 15eV, 20eV and 25eV, collision energies are applied to the isolated precursor ion.

6.2.9. MS/MS of m/z = 273 in D1 20%
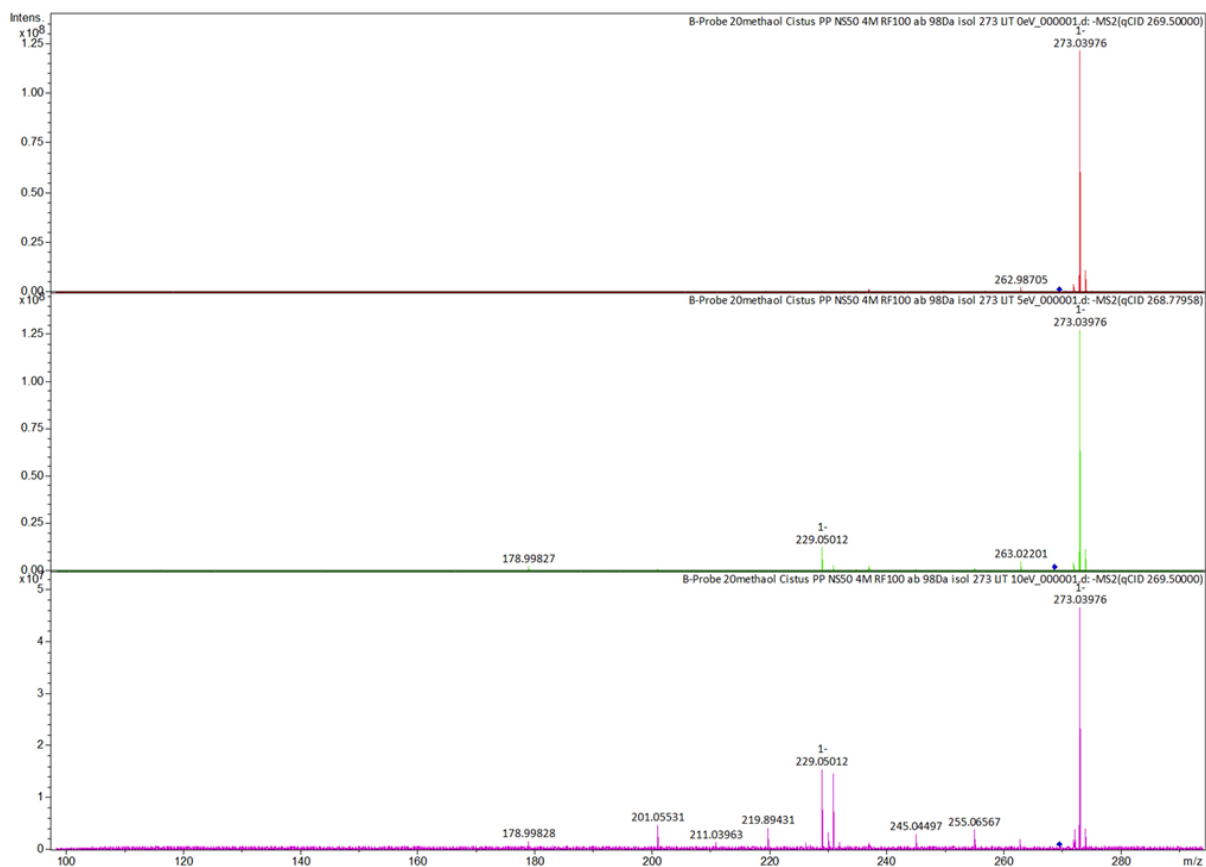
Figure 79: Fragmentation of 273 in D1 20% using (-) ESI-FT-ICR-MS/MS ultrahigh resolution mass spectrometry. From top to the bottom 0eV, 5eV and 10eV, collision energies are applied to the isolated precursor ion.
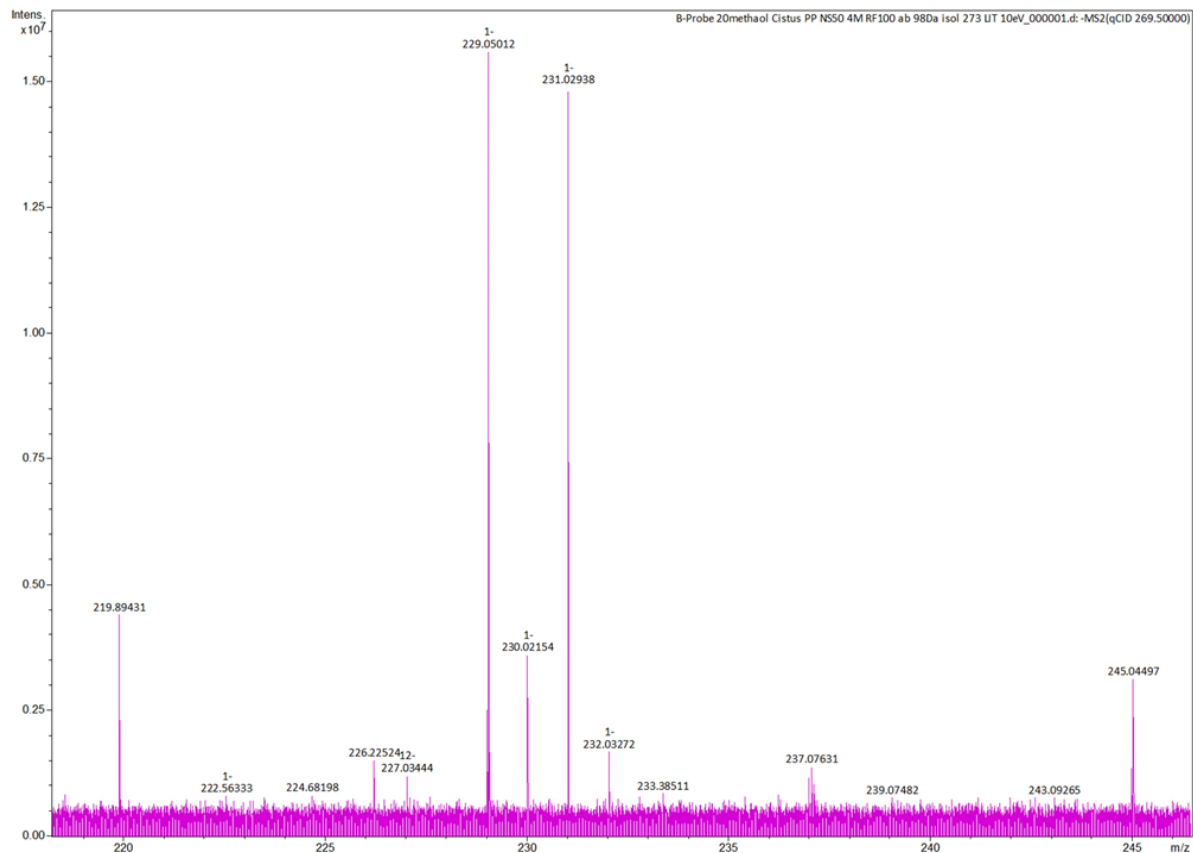
Figure 80: Fragmentation of 273 in D1 20% using (-) ESI-FT-ICR-MS/MS ultrahigh resolution mass spectrometry. Enlargement of 220 Da to 245 Da from 15eV collision energy applied to the isolated precursor ion.
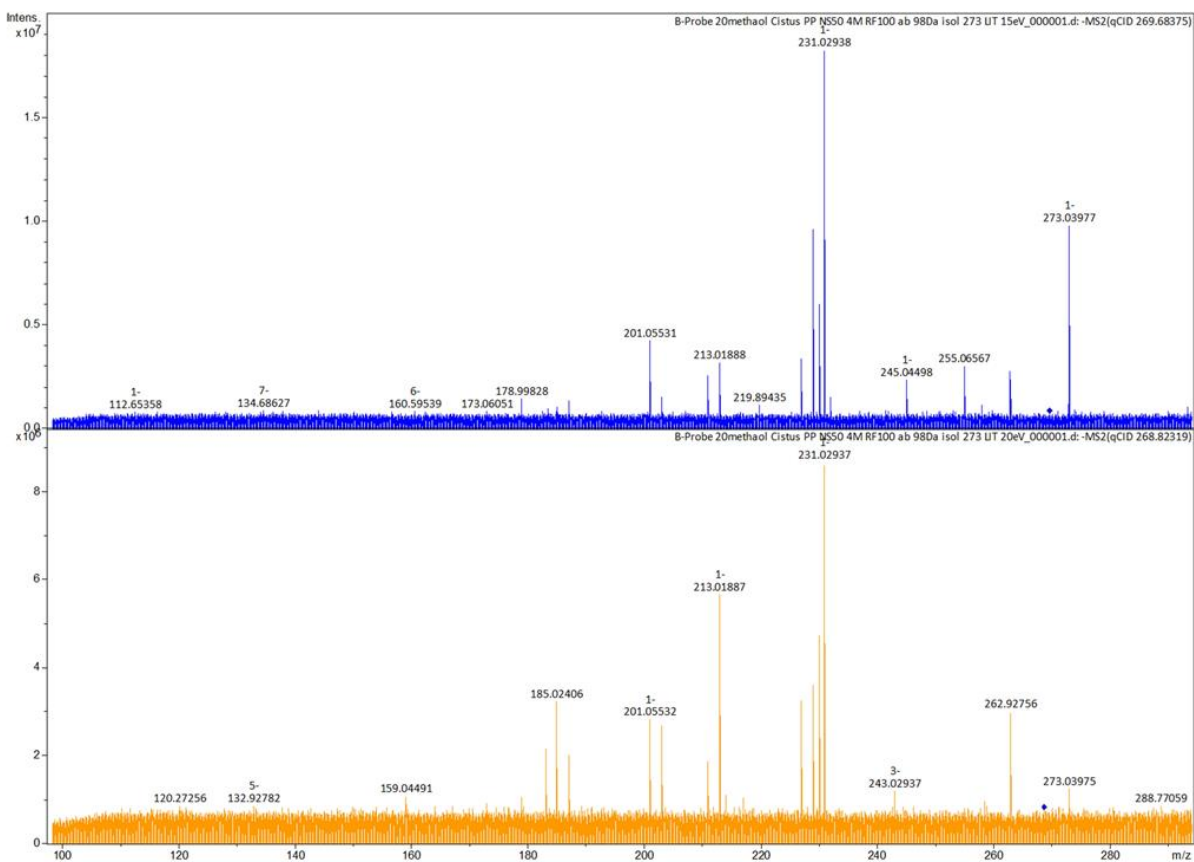
Figure 81: Fragmentation of 273 in D1 20% using (-) ESI-FT-ICR-MS/MS ultrahigh resolution mass spectrometry. From top to the bottom 15eV, and 20eV, collision energies are applied to the isolated precursor ion.

6.2.10.   Biological characterization
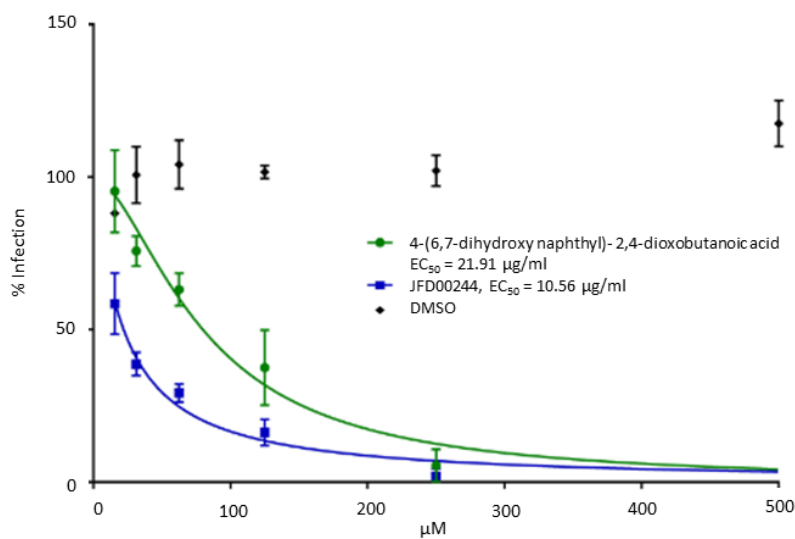
Figure 82: Overview of the EC$_{50}$ results for 4-(6,7-dihydroxy naphthyl)-2,4-dioxobutanoic acid, JFD00244 and DMSO (n(experiments)=3, mean±SD). The concentration of the extracts is shown in µg dry weight per ml cell culture medium. The experiment was performed in the lab of Prof. Dr. Ruth-Brack-Werner.

6.2.11.   Calibration lists

Table 33: *Cistus incanus* specific substances used for calibration.

Substances	Molecular formula [M]	Calculated [M-H]$^-$
Gallic acid	$C_7H_6O_5$	169.014249
Methyl gallic acid	$C_8H_8O_5$	183.029899
Uralenneoside (1-O-(3,4-Dihydroxybenzoyl)-β-D-ribopyranose)	$C_{12}H_{14}O_8$	285.061594
Catechin	$C_{15}H_{14}O_6$	289.071764
Ellagic acid	$C_{14}H_6O_8$	300.998994
Gallocatechin	$C_{15}H_{14}O_7$	305.066679
Hydroxyferuloyl-rhamnose	$C_{16}H_{20}O_9$	355.103459
Rosmarinic acid	$C_{18}H_{16}O_8$	359.077244
Quercetin-O-xylosid	$C_{20}H_{18}O_{11}$	433.077639
Quercitrin	$C_{21}H_{20}O_{11}$	447.093289
Myricetin-O-xylosid	$C_{21}H_{20}O_{11}$	449.072554
Epigallocatechin-gallat	$C_{22}H_{18}O_{11}$	457.077639

6. Supplementary Information

Myricitrin	$C_{21}H_{20}O_{12}$	463.088204
Quercetin-3-O-galactosid	$C_{21}H_{20}O_{12}$	463.088204
Myricetin-3-O-galactosid	$C_{21}H_{20}O_{13}$	479.083119
Hexahydroxydiphenoyl-glucose	$C_{20}H_{18}O_{14}$	481.062384
Procyanidin B1	$C_{30}H_{26}O_{12}$	577.135154
Cis-Tilirosid	$C_{30}H_{26}O_{13}$	593.130069
Vicenin-2	$C_{27}H_{30}O_{15}$	593.151199
Rutin	$C_{27}H_{30}O_{16}$	609.146114

Table 34: *Ocimum tenuiflorum* **specific substances used for calibration.**

Substances	Molecular formula [M]	Calculated [M-H]$^-$
(Z)-3-Hexenol	$C_6H_{10}O$	99.081539
Camphene	$C_{10}H_{16}$	135.117924
Methylchavicol	$C_{10}H_{12}O$	147.081539
(-)-champhor	$C_{10}H_{16}O$	151.112839
Linalool	$C_{10}H_{18}O$	153.128489
Eugenol	$C_{10}H_{12}O_2$	163.076454
Methyleugenol	$C_{11}H_{14}O_2$	177.092104
Alpha-Humulene	$C_{15}H_{24}$	203.180524
Nerolido	$C_{15}H_{26}O$	221.191089

Table 35: Fatty acids and MeOH specific compounds for general calibration.

Molecular formula [M]	Calculated [M-H]$^-$
$C_5H_9O_2$	102.060803
$C_7H_{15}O_2$	129.092103
$C_9H_{19}O_2$	157.123403
$C_{11}H_{21}O_2$	185.154703
$C_{13}H_{25}O_2$	213.186004
$C_{15}H_{29}O_2$	241.217304
$C_{16}H_{31}O_2$	255.232954
$C_{17}H_{33}O_2$	269.248604

C$_{18}$H$_{35}$O$_2$	283.264254
C$_{19}$H$_{37}$O$_2$	297.279904
C$_{17}$H$_{27}$O$_3$S	311.168639
C$_{21}$H$_{41}$O$_2$	325.311204
C$_{18}$H$_{29}$O$_3$S	325.184289
C$_{19}$H$_{31}$O$_3$S	339.199939
C$_{23}$H$_{45}$O$_2$	353.342504
C$_{25}$H$_{49}$O$_2$	381.373804
C$_{27}$H$_{53}$O$_2$	409.405105
C$_{29}$H$_{57}$O$_2$	437.436405
C$_{31}$H$_{61}$O$_2$	465.467705

6.3. Chapter 4

6.3.1. HCPP Tulsi
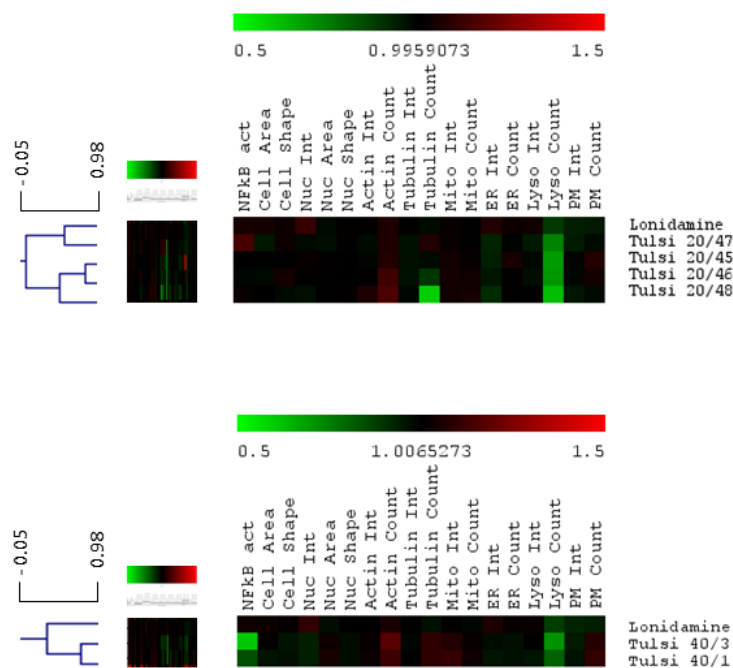
Figure 83: Visualization of one peak that elute in end of D2 20% MeOH and beginning of D2 40% MeOH. Visible in Spearman Rank and Kendall's tau clustering.

6. Supplementary Information

6.3.2. Support trees
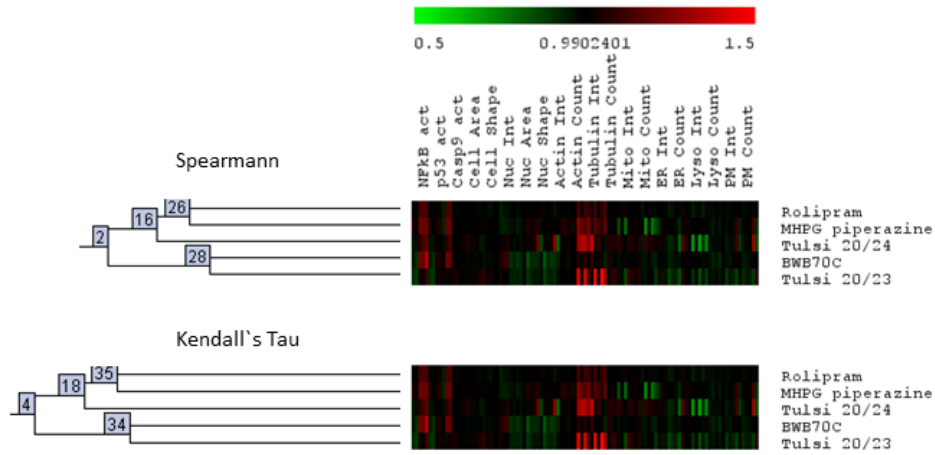
Figure 84: Support trees for the Tulsi D2 20% 23/24 cluster using Spearman Rank (up) and Kendall`s Tau (down) clustering.
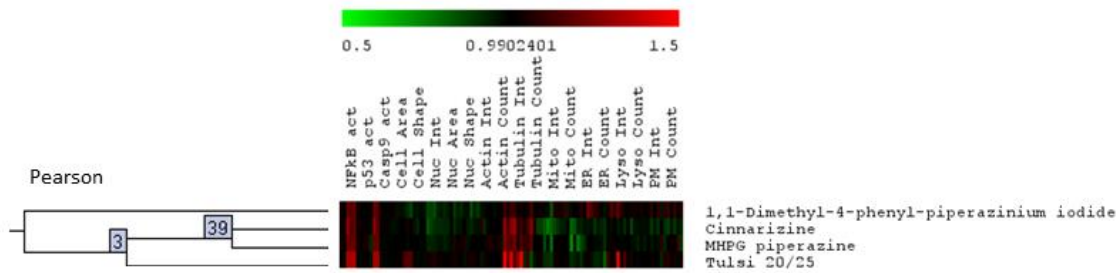
Figure 85: Support tree for the Tulsi D2 20% 25 cluster using Pearson clustering.
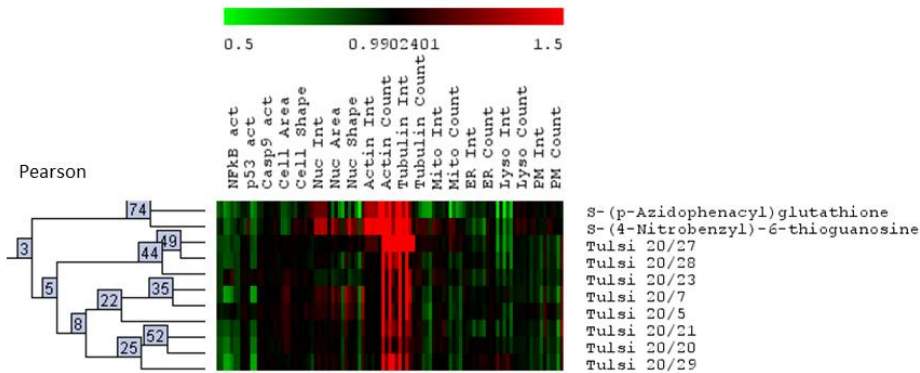
Figure 86: Support trees for the Tulsi D2 20% 5,7,20-23,27-29 cluster using Pearson clustering.

6.3.3. EASY-HIT and MTT
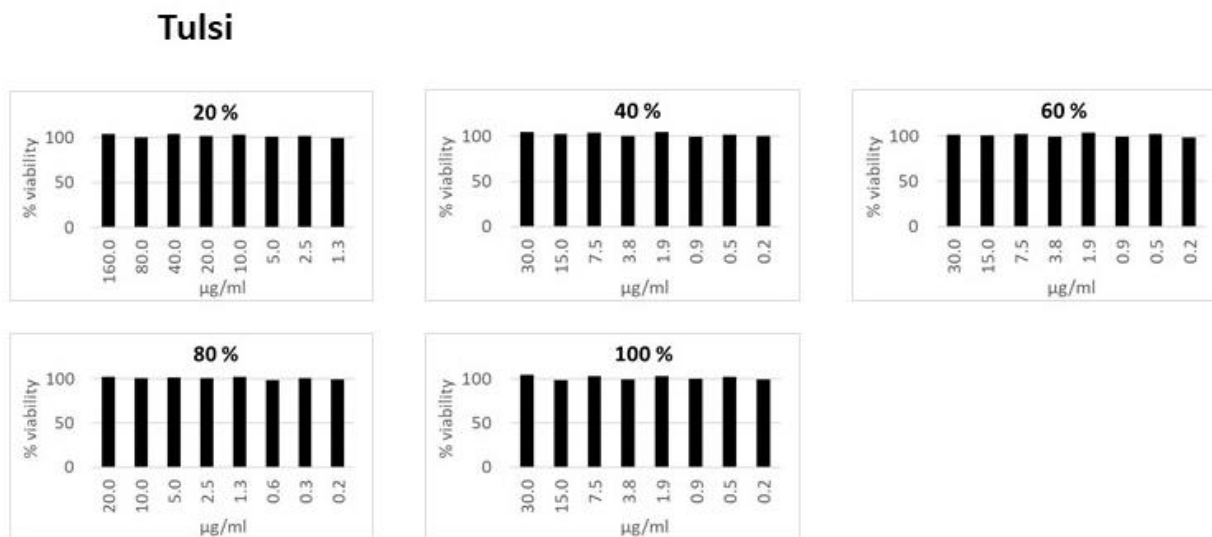
Tulsi

Figure 87: Overview of the cell viability tests (MTT tests) for the D1 fractions of Tulsi. No reduction of the metabolic process under 80% appeared.

7. Literature

Petrauskas, A., & Kolovanov, E. (2000). *ACD/Log P method description* (Vol. 19).

Allison, A. J., Christophe, M., & Yves, P. (2004). HIV-1 Integrase Inhibitors: A Decade of Research and Two Drugs in Clinical Trial. *Current Topics in Medicinal Chemistry, 4*(10), 1059-1077.

Allwood, J. W., De Vos, R. C., Moing, A., Deborde, C., Erban, A., Kopka, J., Goodacre, R., & Hall, R. D. (2011). Plant metabolomics and its potential for systems biology research background concepts, technology, and methodology. *Methods Enzymol, 500*, 299-336.

Allwood, J. W., & Goodacre, R. (2010). An introduction to liquid chromatography-mass spectrometry instrumentation applied in plant metabolomic analyses. *Phytochem Anal, 21*(1), 33-47.

Ammaranond, P., & Sanguansittianan, S. (2012). Mechanism of HIV antiretroviral drugs progress toward drug resistance. *Fundam Clin Pharmacol, 26*(1), 146-161.

Amster, I. J. (1996). Fourier Transform Mass Spectrometry. *Journal of Mass Spectrometry, 31*(12), 1325-1337.

Andrae-Marobela, K., Ghislain, F. W., Okatch, H., & Majinda, R. R. (2013). Polyphenols: a diverse class of multi-target anti-HIV-1 agents. *Curr Drug Metab, 14*(4), 392-413.

Annesley, T. M. (2003). Ion suppression in mass spectrometry. *Clin Chem, 49*(7), 1041-1044.

Anthony, N. J. (2004). HIV-1 integrase: a target for new AIDS chemotherapeutics. *Curr Top Med Chem, 4*(9), 979-990.

Appleton, D. R., Buss, A. D., & Butler, M. S. (2007). A Simple Method for High-Throughput Extract Prefractionation for Biological Screening. *CHIMIA International Journal for Chemistry, 61*(6), 327-331.

Araujo, P., Tilahun, E., Breivik, J. F., Abdulkader, B. M., Frøyland, L., & Zeng, Y. (2016). A simple liquid extraction protocol for overcoming the ion suppression of triacylglycerols by phospholipids in liquid chromatography mass spectrometry studies. *Talanta, 148*, 463-471.

Aturki, Z., Brandi, V., & Sinibaldi, M. (2004). Separation of Flavanone-7-O-glycoside Diastereomers and Analysis in Citrus Juices by Multidimensional Liquid Chromatography Coupled with Mass Spectrometry. *Journal of Agricultural and Food Chemistry, 52*(17), 5303-5308.

Baddeley, G. (1949). S 20. The acylation of naphthalene by the friedel-crafts reaction. *Journal of the Chemical Society (Resumed)*(0), S99-S103.

Banerjee, S., Prashar, R., Kumar, A., & Rao, A. R. (1996). Modulatory influence of alcoholic extract of Ocimum leaves on carcinogen-metabolizing enzyme activities and reduced glutathione levels in mouse. *Nutr Cancer, 25*(2), 205-217.

Barrajon-Catalan, E., Fernandez-Arroyo, S., Roldan, C., Guillen, E., Saura, D., Segura-Carretero, A., & Micol, V. (2011). A systematic study of the polyphenolic composition of aqueous extracts deriving from several Cistus genus species: evolutionary relationship. *Phytochem Anal, 22*(4), 303-312.

Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W., & Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science, 220*(4599), 868-871.

Bartholomew, C. F., & Jones, A. M. (2006). *Human bites: a rare risk factor for HIV transmission*: AIDS. 2006 Feb 28;20(4):631-2.

Bast, F., Rani, P., & Meena, D. (2014). Chloroplast DNA Phylogeography of Holy Basil (Ocimum tenuiflorum) in Indian Subcontinent. *The Scientific World Journal, 2014*, 847482.

Bender, A., & Glen, R. C. (2004). Molecular similarity: a key technique in molecular informatics. *Org Biomol Chem, 2*(22), 3204-3218.

Berridge, M. V., Tan, A. S., McCoy, K. D., & Wang, R. (1996). The biochemical and cellular basis of cell proliferation assays that use tetrazolium salts. *Biochemica, 4*(1), 14-19.

Blankson, J. N., Persaud, D., & Siliciano, R. F. (2002). The challenge of viral reservoirs in HIV-1 infection. *Annu Rev Med, 53*, 557-593.

Board, N. (2005). Compendium of Medicinal Plants. *National Institute of Industrial Research*, 444.

Bondia-Pons, I., Savolainen, O., Törrönen, R., Martinez, J. A., Poutanen, K., & Hanhineva, K. (2014). Metabolic profiling of Goji berry extracts for discrimination of geographical origin by non-targeted liquid chromatography coupled to quadrupole time-of-flight mass spectrometry. *Food Research International, 63, Part B*, 132-138.

Bowen, B. P., & Northen, T. R. (2010). Dealing with the unknown: metabolomics and metabolite atlases. *J Am Soc Mass Spectrom, 21*(9), 1471-1476.

Bray, M. A., Singh, S., Han, H., Davis, C. T., Borgeson, B., Hartland, C., Kost-Alimova, M., Gustafsdottir, S. M., Gibson, C. C., & Carpenter, A. E. (2016). Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat Protoc, 11*(9), 1757-1774.

Broeckling, C. D., Huhman, D. V., Farag, M. A., Smith, J. T., May, G. D., Mendes, P., Dixon, R. A., & Sumner, L. W. (2005). Metabolic profiling of Medicago truncatula cell cultures reveals the effects of biotic and abiotic elicitors on metabolism. *J Exp Bot, 56*(410), 323-336.

Buhrman, D. L., Price, P. I., & Rudewiczcor, P. J. (1996). Quantitation of SR 27417 in human plasma using electrospray liquid chromatography-tandem mass spectrometry: A study of ion suppression. *J Am Soc Mass Spectrom, 7*(11), 1099-1105.

Burt, S. M., Carter, T. J., & Kricka, L. J. (1979). Thermal characteristics of microtitre plates used in immunological assays. *J Immunol Methods, 31*(3-4), 231-236.

Butler, M. S., Fontaine, F., & Cooper, M. A. (2014). Natural Product Libraries: Assembly, Maintenance, and Screening. *Planta Med, 80*(14), 1161-1170.

Buzon, M. J., Sun, H., Li, C., Shaw, A., Seiss, K., Ouyang, Z., Martin-Gayo, E., Leng, J., Henrich, T. J., Li, J. Z., Pereyra, F., Zurakowski, R., Walker, B. D., Rosenberg, E. S., Yu, X. G., & Lichterfeld, M. (2014). HIV-1 persistence in CD4+ T cells with stem cell–like properties. *Nature Medicine, 20*, 139.

Caicedo, J. C., Singh, S., & Carpenter, A. E. (2016). Applications in image-based profiling of perturbations. *Curr Opin Biotechnol, 39*, 134-142.

Chattopadhyay, R. R. S., S. K. / Ganguly, S. / Basu, T. K. . (1994). A comparative evaluation of some anti-inflammatory agents of plant origin *FITOTERAPIA -MILANO-, 65*(2), 146.

Cheers. (2000). *Botanica* (Vol. 1). Potsdam: Edition Könemann, Tandem Verlag GmbH.

Choudhury, S. S., Bashyam, L., Manthapuram, N., Bitla, P., Kollipara, P., & Tetali, S. D. (2014). Ocimum sanctum leaf extracts attenuate human monocytic (THP-1) cell activation. *J Ethnopharmacol, 154*(1), 148-155.

Chowdhury, T., Mandal, A., Roy, S. C., & De Sarker, D. (2017). Diversity of the genus Ocimum (Lamiaceae) through morpho-molecular (RAPD) and chemical (GC–MS) analysis. *Journal of Genetic Engineering and Biotechnology, 15*(1), 275-286.

Clavel, F., Guetard, D., Brun-Vezinet, F., Chamaret, S., Rey, M. A., Santos-Ferreira, M. O., Laurent, A. G., Dauguet, C., Katlama, C., Rouzioux, C., & et al. (1986). Isolation of a new human retrovirus from West African patients with AIDS. *Science, 233*(4761), 343-346.

Coffin, J., Haase, A., Levy, J. A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., Vogt, P., & et al. (1986). *Human immunodeficiency viruses*: Science. 1986 May 9;232(4751):697.

Collins, K. L., Chen, B. K., Kalams, S. A., Walker, B. D., & Baltimore, D. (1998). HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature, 391*(6665), 397-401.

Comisarow, M. B., & Marshall, A. G. (1974). Fourier transform ion cyclotron resonance spectroscopy. *Chemical Physics Letters, 25*(2), 282-283.

7. Literature

Cramer, R. D., Patterson, D. E., & Bunce, J. D. (1988). Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society, 110*(18), 5959-5967.

Crisman, T. J., Parker, C. N., Jenkins, J. L., Scheiber, J., Thoma, M., Kang, Z. B., Kim, R., Bender, A., Nettles, J. H., Davies, J. W., & Glick, M. (2007). Understanding False Positives in Reporter Gene Assays: in Silico Chemogenomics Approaches To Prioritize Cell-Based HTS Data. *Journal of Chemical Information and Modeling, 47*(4), 1319-1327.

Cubbon, S., Bradbury, T., Wilson, J., & Thomas-Oates, J. (2007). Hydrophilic interaction chromatography for mass spectrometric metabonomic studies of urine. *Anal Chem, 79*(23), 8911-8918.

Daelemans, D., Pauwels, R., De Clercq, E., & Pannecouque, C. (2011). A time-of-drug addition approach to target identification of antiviral compounds. *Nat Protoc, 6*(6), 925-933.

Dalgleish, A. G., Beverley, P. C., Clapham, P. R., Crawford, D. H., Greaves, M. F., & Weiss, R. A. (1984). The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature, 312*(5996), 763-767.

Danne, A., Petereit, F., & Nahrstedt, A. (1993). Proanthocyanidins from Cistus incanus. *Phytochemistry, 34*(4), 1129-1133.

Das, R., Raman, R. P., Saha, H., & Singh, R. (2015). Effect of Ocimum sanctum Linn. (Tulsi) extract on the immunity and survival of Labeo rohita (Hamilton) infected with Aeromonas hydrophila. *Aquaculture Research, 46*(5), 1111-1121.

De Bruyne, T., Pieters, L., Witvrouw, M., De Clercq, E., Vanden Berghe, D., & Vlietinck, A. J. (1999). Biological evaluation of proanthocyanidin dimers and related polyphenols. *J Nat Prod, 62*(7), 954-958.

de Hoffmann, E. S., V.: (2013). *Mass spectrometry: Principles and applications* New York: John Wiley & Sons.

de Villiers, A., Venter, P., & Pasch, H. (2016). Recent advances and trends in the liquid-chromatography–mass spectrometry analysis of flavonoids. *Journal of Chromatography A, 1430*, 16-78.

Dezzutti, C. S., & Hladik, F. (2013). Use of Human Mucosal Tissue to Study HIV-1 Pathogenesis and Evaluate HIV-1 Prevention Modalities. *Current HIV/AIDS reports, 10*(1), 12-20.

Dieffenbach, C. W., & Fauci, A. S. (2011). Thirty years of HIV and AIDS: future challenges and opportunities. *Ann Intern Med, 154*(11), 766-771.

Dixon, R. A., & Strack, D. (2003). *Phytochemistry meets genome analysis, and beyond*: Phytochemistry. 2003 Mar;62(6):815-6.

Dole, M., Mack, L. L., Hines, R. L., Mobley, R. C., Ferguson, L. D., & Alice, M. B. (1968). Molecular Beams of Macroions. *The Journal of Chemical Physics, 49*(5), 2240-2249.

Dowdy, D., Gore, P. H., & Waters, D. N. (1991). The Friedel-Crafts acetylation of naphthalene in 1,2-dichloroethane solution. Kinetics and mechanism. *Journal of the Chemical Society, Perkin Transactions 2*(8), 1149-1159.

Droebner, K., Ehrhardt, C., Poetter, A., Ludwig, S., & Planz, O. (2007). CYSTUS052, a polyphenol-rich plant extract, exerts anti-influenza virus activity in mice. *Antiviral Res, 76*(1), 1-10.

Dugo, P., Kumm, T., Crupi, M. L., Cotroneo, A., & Mondello, L. (2006). Comprehensive two-dimensional liquid chromatography combined with mass spectrometric detection in the analyses of triacylglycerols in natural lipidic matrixes. *J Chromatogr A, 1112*(1-2), 269-275.

Ehrhardt, C., Hrincius, E. R., Korte, V., Mazur, I., Droebner, K., Poetter, A., Dreschers, S., Schmolke, M., Planz, O., & Ludwig, S. (2007). A polyphenol rich plant extract, CYSTUS052, exerts anti influenza virus activity in cell culture without toxic side effects or the tendency to induce viral resistance. *Antiviral Res, 76*(1), 38-47.

Ernst, M., Silva, D. B., Silva, R. R., Vencio, R. Z., & Lopes, N. P. (2014). Mass spectrometry in plant metabolomics strategies: from analytical platforms to data acquisition and processing. *Nat Prod Rep, 31*(6), 784-806.

Faber, J. H., Malmodin, D., Toft, H., Maher, A. D., Crockford, D., Holmes, E., Nicholson, J. K., Dumas, M. E., & Baunsgaard, D. (2007). Metabonomics in diabetes research. *J Diabetes Sci Technol, 1*(4), 549-557.

Farag, M. A., Huhman, D. V., Lei, Z., & Sumner, L. W. (2007). Metabolic profiling and systematic identification of flavonoids and isoflavonoids in roots and cell suspension cultures of Medicago truncatula using HPLC-UV-ESI-MS and GC-MS. *Phytochemistry, 68*(3), 342-354.

Farhin Inam, S. D., Neha Narkhede. (2014). HPLC –UV Method Development and Quantification of Eugenol from Methanolic Extracts of Some Spices. *International Journal of Chemical and Physical Sciences 3*(6), 96-102.

Feng, B. Y., & Shoichet, B. K. (2006). Synergy and Antagonism of Promiscuous Inhibition in Multiple-Compound Mixtures. *Journal of medicinal chemistry, 49*(7), 2151-2154.

Feng, Y., Mitchison, T. J., Bender, A., Young, D. W., & Tallarico, J. A. (2009). Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds. *Nat Rev Drug Discov, 8*(7), 567-578.

Fenn, J. B. (2003). Electrospray wings for molecular elephants (Nobel lecture). *Angew Chem Int Ed Engl, 42*(33), 3871-3894.

Fiehn, O. (2001). Combining Genomics, Metabolome Analysis, and Biochemical Modelling to Understand Metabolic Networks. *Comparative and Functional Genomics, 2*(3), 155-168.

Fiehn, O., Sumner, L. W., Rhee, S. Y., Ward, J., Dickerson, J., Lange, B. M., Lane, G., Roessner, U., Last, R., & Nikolau, B. (2007). Minimum reporting standards for plant biology context information in metabolomic studies. *Metabolomics, 3*(3), 195-201.

Fischl, M. A., Richman, D. D., Hansen, N., Collier, A. C., Carey, J. T., Para, M. F., Hardy, W. D., Dolin, R., Powderly, W. G., Allan, J. D., & et al. (1990). The safety and efficacy of zidovudine (AZT) in the treatment of subjects with mildly symptomatic human immunodeficiency virus type 1 (HIV) infection. A double-blind, placebo-controlled trial. The AIDS Clinical Trials Group. *Ann Intern Med, 112*(10), 727-737.

Fishman, M. C., & Porter, J. A. (2005). *Pharmaceuticals: a new grammar for drug discovery*: Nature. 2005 Sep 22;437(7058):491-3.

Forcisi, S., Moritz, F., Kanawati, B., Tziotis, D., Lehmann, R., & Schmitt-Kopplin, P. (2013). Liquid chromatography–mass spectrometry in metabolomics research: Mass analyzers in ultra high pressure liquid chromatography coupling. *Journal of Chromatography A, 1292*, 51-65.

Freedberg, K. A., Losina, E., Weinstein, M. C., Paltiel, A. D., Cohen, C. J., Seage, G. R., Craven, D. E., Zhang, H., Kimmel, A. D., & Goldie, S. J. (2001). The cost effectiveness of combination antiretroviral therapy for HIV disease. *N Engl J Med, 344*(11), 824-831.

Gallo, R. C., Sarin, P. S., Gelmann, E. P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V. S., Mann, D., Sidhu, G. D., Stahl, R. E., Zolla-Pazner, S., Leibowitch, J., & Popovic, M. (1983). Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science, 220*(4599), 865-867.

Ganzera, M., & Sturm, S. (2018). Recent advances on HPLC/MS in medicinal plant analysis—An update covering 2011–2016. *Journal of Pharmaceutical and Biomedical Analysis, 147*, 211-233.

Gebo, K. A., Fleishman, J. A., Conviser, R., Hellinger, J., Hellinger, F. J., Josephs, J. S., Keiser, P., Gaist, P., & Moore, R. D. (2010). Contemporary Costs of HIV Health Care in the HAART Era. *AIDS (London, England), 24*(17), 2705-2715.

Geller, O., & Lifshitz, C. (2003). Gas Phase H/D Exchange of Protonated Arginine Monomers and Dimers. *The Journal of Physical Chemistry A, 107*(30), 5654-5659.

7. Literature

Ghani, A. C., de Wolf, F., Ferguson, N. M., Donnelly, C. A., Coutinho, R., Miedema, F., Goudsmit, J., & Anderson, R. M. (2001). Surrogate markers for disease progression in treated HIV infection. *J Acquir Immune Defic Syndr, 28*(3), 226-231.

Ghaste, M., Mistrik, R., & Shulaev, V. (2016). Applications of Fourier Transform Ion Cyclotron Resonance (FT-ICR) and Orbitrap Based High Resolution Mass Spectrometry in Metabolomics and Lipidomics. *International Journal of Molecular Sciences, 17*(6), 816.

Ghose, A. K., & Crippen, G. M. (1985). Use of physicochemical parameters in distance geometry and related three-dimensional quantitative structure-activity relationships: a demonstration using Escherichia coli dihydrofolate reductase inhibitors. *J Med Chem, 28*(3), 333-346.

Ghose, A. K., & Crippen, G. M. (1987). Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions. *Journal of Chemical Information and Computer Sciences, 27*(1), 21-35.

Ghose, A. K., Viswanadhan, V. N., & Wendoloski, J. J. (1999). A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J Comb Chem, 1*(1), 55-68.

Giddings, J. C. (1967). Maximum number of components resolvable by gel filtration and other elution chromatographic methods. *Analytical Chemistry, 39*(8), 1027-1028.

Giddings, J. C. (1984). Two-dimensional separations: concept and promise. *Anal Chem, 56*(12).

Giuliano, K. A., DeBiasio, R. L., Dunlay, R. T., Gough, A., Volosky, J. M., Zock, J., Pavlakis, G. N., & Taylor, D. L. (1997). High-Content Screening: A New Approach to Easing Key Bottlenecks in the Drug Discovery Process. *J Biomol Screen, 2*(4), 249-259.

Goble, C., & Stevens, R. (2008). State of the nation in data integration for bioinformatics. *Journal of Biomedical Informatics, 41*(5), 687-693.

Godhwani, S., Godhwani, J. L., & Vyas, D. S. (1987). Ocimum sanctum: an experimental study evaluating its anti-inflammatory, analgesic and antipyretic activity in animals. *J Ethnopharmacol, 21*(2), 153-163.

Gomez-Cabrero, D., Abugessaisa, I., Maier, D., Teschendorff, A., Merkenschlager, M., Gisel, A., Ballestar, E., Bongcam-Rudloff, E., Conesa, A., & Tegnér, J. (2014). Data integration in the era of omics: current and future challenges. *BMC Systems Biology, 8*(2)

González-Rodríguez, J., Pérez-Juan, P., & Luque de Castro, M. D. (2002). Method for the simultaneous determination of total polyphenol and anthocyan indexes in red wines using a flow injection approach. *Talanta, 56*(1), 53-59.

Goodacre, R., Vaidyanathan, S., Dunn, W. B., Harrigan, G. G., & Kell, D. B. (2004). Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol, 22*(5), 245-252.

Gore, P. H., Thadani, C. K., & Thorburn, S. (1968). Friedel-Crafts acylations of aromatic hydrocarbons. Part VII. The acetylation and benzoylation of 2,3-dimethylnaphthalene. *Journal of the Chemical Society C: Organic*(0), 2502-2508.

Gori, A., Ferrini, F., Marzano, M. C., Tattini, M., Centritto, M., Baratto, M. C., Pogni, R., & Brunetti, C. (2016). Characterisation and Antioxidant Activity of Crude Extract and Polyphenolic Rich Fractions from C. incanus Leaves. *International Journal of Molecular Sciences, 17*(8), 1344.

Gottlieb, M. S., Schroff, R., Schanker, H. M., Weisman, J. D., Fan, P. T., Wolf, R. A., & Saxon, A. (1981). Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N Engl J Med, 305*(24), 1425-1431.

Grabley, S., Thiericke, R., & Sattler, I. (2000). *Tools for Drug Discovery: Natural Product-based Libraries.* Paper presented at the The Role of Natural Products in Drug Discovery, Berlin, Heidelberg.

Grayer, R. J., Bryan, S. E., Veitch, N. C., Goldstone, F. J., Paton, A., & Wollenweber, E. (1996). External flavones in sweet basil, Ocimum basilicum, and related taxa. *Phytochemistry, 43*(5), 1041-1047.

Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics, 11*(1), 1-21.

Guo, C.-C., Tong, R.-B., & Li, K.-L. (2004). Chloroalkyl piperazine and nitrogen mustard porphyrins: synthesis and anticancer activity. *Bioorganic & Medicinal Chemistry, 12*(9), 2469-2475.

Gustafsdottir, S. M., Ljosa, V., Sokolnicki, K. L., Anthony Wilson, J., Walpita, D., Kemp, M. M., Petri Seiler, K., Carrel, H. A., Golub, T. R., Schreiber, S. L., Clemons, P. A., Carpenter, A. E., & Shamji, A. F. (2013). Multiplex cytological profiling assay to measure diverse cellular states. *PLoS ONE, 8*(12).

H. Nörr, H. W. (1992). *New Constituents from Ocimum sanctum* (Vol. 58). New York: Georg Thieme Verlag Stuttgart

Hajjar, D., Kremb, S., Sioud, S., Emwas, A. H., Voolstra, C. R., & Ravasi, T. (2017). Anti-cancer agents in Saudi Arabian herbals revealed by automated high-content imaging. *PLoS ONE, 12*(6).

Hansch, C. (1993). Quantitative structure-activity relationships and the unnamed science. *Accounts of Chemical Research, 26*(4), 147-153.

Hauber, I., Hofmann-Sieber, H., Chemnitz, J., Dubrau, D., Chusainow, J., Stucka, R., Hartjen, P., Schambach, A., Ziegler, P., Hackmann, K., Schrock, E., Schumacher, U., Lindner, C., Grundhoff, A., Baum, C., Manz, M. G., Buchholz, F., & Hauber, J. (2013). Highly significant antiviral activity of HIV-1 LTR-specific tre-recombinase in humanized mice. *PLoS Pathog, 9*(9), 26.

Hazuda, D. J., Anthony, N. J., Gomez, R. P., Jolly, S. M., Wai, J. S., Zhuang, L., Fisher, T. E., Embrey, M., Guare, J. P., Egbertson, M. S., Vacca, J. P., Huff, J. R., Felock, P. J., Witmer, M. V., Stillmock, K. A., Danovich, R., Grobler, J., Miller, M. D., Espeseth, A. S., Jin, L., Chen, I.-W., Lin, J. H., Kassahun, K., Ellis, J. D., Wong, B. K., Xu, W., Pearson, P. G., Schleif, W. A., Cortese, R., Emini, E., Summa, V., Holloway, M. K., & Young, S. D. (2004). A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase. *Proceedings of the National Academy of Sciences of the United States of America, 101*(31), 11233-11238.

Hazuda, D. J., Felock, P., Witmer, M., Wolfe, A., Stillmock, K., Grobler, J. A., Espeseth, A., Gabryelski, L., Schleif, W., Blau, C., & Miller, M. D. (2000). Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science, 287*(5453), 646-650.

Helfer, M., Koppensteiner, H., Schneider, M., Rebensburg, S., Forcisi, S., Müller, C., Schmitt-Kopplin, P., Schindler, M., & Brack-Werner, R. (2014). The Root Extract of the Medicinal Plant Pelargonium sidoides a Potent HIV-1 Attachment Inhibitor. *PLoS ONE, 9*(1), e87487.

Henderson, D. K., Fahey, B. J., Willy, M., Schmitt, J. M., Carey, K., Koziol, D. E., Lane, H. C., Fedio, J., & Saah, A. J. (1990). Risk for occupational transmission of human immunodeficiency virus type 1 (HIV-1) associated with clinical exposures. A prospective evaluation. *Ann Intern Med, 113*(10), 740-746.

Hertkorn, N., Frommberger, M., Witt, M., Koch, B. P., Schmitt-Kopplin, P., & Perdue, E. M. (2008). Natural Organic Matter and the Event Horizon of Mass Spectrometry. *Analytical Chemistry, 80*(23), 8908-8919.

Hoffman, A. F., & Garippa, R. J. (2007). A pharmaceutical company user's perspective on the potential of high content screening in drug discovery. *Methods Mol Biol, 356*, 19-31.

Hommerson, P., Khan, A. M., de Jong, G. J., & Somsen, G. W. (2011). Ionization techniques in capillary electrophoresis-mass spectrometry: principles, design, and application. *Mass Spectrom Rev, 30*(6), 1096-1120.

Hughey, C. A., Hendrickson, C. L., Rodgers, R. P., Marshall, A. G., & Qian, K. (2001). Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra. *Analytical Chemistry, 73*(19), 4676-4681.

Huhman, D. V., & Sumner, L. W. (2002). Metabolic profiling of saponins in Medicago sativa and Medicago truncatula using HPLC coupled to an electrospray ion-trap mass spectrometer. *Phytochemistry, 59*(3), 347-360.

7. Literature

Ianni, A., & Waldvogel, S. R. (2006). Reliable and Versatile Synthesis of 2-Aryl-Substituted Cinnamic Acid Esters. *Synthesis, 2006*(13), 2103-2112.

Ivanova-Petropulos, V., Ricci, A., Nedelkovski, D., Dimovska, V., Parpinello, G. P., & Versari, A. (2015). Targeted analysis of bioactive phenolic compounds and antioxidant activity of Macedonian red wines. *Food Chem, 171*, 412-420.

Jandera, P. (2006). Column selectivity for two-dimensional liquid chromatography. *Journal of Separation Science, 29*(12), 1763-1783.

Jandera, P. (2012). Programmed elution in comprehensive two-dimensional liquid chromatography. *Journal of Chromatography A, 1255*, 112-129.

Jiang, L., Tao, Y., Wang, D., Tang, C., Shao, Y., Wang, Q., Zhao, X., Zhang, Y., & Mei, L. (2014). A novel two-dimensional preparative chromatography method designed for the separation of traditional animal Tibetan medicine Osteon Myospalacem Baileyi. *Journal of Separation Science, 37*(21), 3060-3066.

Kalus, U., Kiesewetter, H., & Radtke, H. (2010). Effect of CYSTUS052 and green tea on subjective symptoms in patients with infection of the upper respiratory tract. *Phytother Res, 24*(1), 96-100.

Kanawati, B., Bader, T. M., Wanczek, K.-P., Li, Y., & Schmitt-Kopplin, P. (2017). Fourier transform (FT)-artifacts and power-function resolution filter in Fourier transform mass spectrometry. *Rapid Communications in Mass Spectrometry, 31*(19), 1607-1615.

Kang, J., Hsu, C. H., Wu, Q., Liu, S., Coster, A. D., Posner, B. A., Altschuler, S. J., & Wu, L. F. (2016). Improving drug discovery with high-content phenotypic screens by systematic selection of reporter cell lines. *Nat Biotechnol, 34*(1), 70-77.

Karpas, A., Fleet, G. W., Dwek, R. A., Petursson, S., Namgoong, S. K., Ramsden, N. G., Jacob, G. S., & Rademacher, T. W. (1988). Aminosugar derivatives as potential anti-human immunodeficiency virus agents. *Proceedings of the National Academy of Sciences of the United States of America, 85*(23), 9229-9233.

Kashiwada, Y., Hashimoto, F., Cosentino, L. M., Chen, C. H., Garrett, P. E., & Lee, K. H. (1996). Betulinic acid and dihydrobetulinic acid derivatives as potent anti-HIV agents. *J Med Chem, 39*(5), 1016-1017.

Kashman, Y., Gustafson, K. R., Fuller, R. W., Cardellina, J. H., McMahon, J. B., Currens, M. J., Buckheit, R. W., Hughes, S. H., Cragg, G. M., & Boyd, M. R. (1992). HIV inhibitory natural products. Part 7. The calanolides, a novel HIV-inhibitory class of coumarin derivatives from the tropical rainforest tree, Calophyllum lanigerum. *Journal of Medicinal Chemistry, 35*(15), 2735-2743.

Kebarle, P. (2000). A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom, 35*(7), 804-817.

Kim, H. K., Choi, Y. H., & Verpoorte, R. (2010). NMR-based metabolomic analysis of plants. *Nat Protoc, 5*(3), 536-549.

Kim, S., Kramer, R. W., & Hatcher, P. G. (2003). Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram. *Analytical Chemistry, 75*(20), 5336-5344.

Kind, T., Wohlgemuth, G., Lee, D. Y., Lu, Y., Palazoglu, M., Shahbaz, S., & Fiehn, O. (2009). FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal Chem, 81*(24), 10038-10048.

King, R., Bonfiglio, R., Fernandez-Metzler, C., Miller-Stein, C., & Olah, T. (2000). Mechanistic investigation of ionization suppression in electrospray ionization. *J Am Soc Mass Spectrom, 11*(11), 942-950.

Koop, H. S., de Freitas, R. A., de Souza, L. M., Martinez, G. R., & Silveira, J. L. M. (2013). Development and Validation of a RP-HPLC–PDA Method for Determination of Curcuminoids in Microemulsions. *Chromatographia, 76*(15), 1041-1048.

Kostiainen, R., & Kauppila, T. J. (2009). Effect of eluent on the ionization process in liquid chromatography-mass spectrometry. *J Chromatogr A, 23*(4), 685-699.

Kothari, S. K., Bhattacharya, A. K., & Ramesh, S. (2004). Essential oil yield and quality of methyl eugenol rich Ocimum tenuiflorum L.f. (syn. O. sanctum L.) grown in south India as influenced by method of harvest. *J Chromatogr A, 29*, 1-2.

Kremb, S., Helfer, M., Wolff, H., Kleinschmidt, A., Durner, J., Schmitt-Kopplin, P., & Brack-Werner, R. (2011). HIV Full-Replication Technology for Identification of Novel HIV Inhibitors from Multiple Sources. *Antiviral Research, 90*(2), A36.

Kremb, S., Müller, C., Schmitt-Kopplin, P., & Voolstra, C. (2017). Bioactive Potential of Marine Macroalgae from the Central Red Sea (Saudi Arabia) Assessed by High-Throughput Imaging-Based Phenotypic Profiling. *Marine Drugs, 15*(3), 80.

Kremb, S., & Voolstra, C. R. (2017). High-resolution phenotypic profiling of natural products-induced effects on the single-cell level. *Scientific Reports, 7*, 44472.

Krevelen, V. (1950). Graphical statistical method for the study of structure and reaction processes of coal. *Fuel, 29*, 269-284.

Kumar, S., & Pandey, A. K. (2013). Chemistry and Biological Activities of Flavonoids: An Overview. *The Scientific World Journal, 2013*, 16.

Kuntz, I. D. (1992). Structure-based strategies for drug design and discovery. *Science, 257*(5073), 1078-1082.

Kurita, K. L., Glassey, E., & Linington, R. G. (2015). Integration of high-content screening and untargeted metabolomics for comprehensive functional annotation of natural product libraries. *Proc Natl Acad Sci U S A, 112*(39), 11999-12004.

Kusano, M., Redestig, H., Hirai, T., Oikawa, A., Matsuda, F., Fukushima, A., Arita, M., Watanabe, S., Yano, M., Hiwasa-Tanase, K., Ezura, H., & Saito, K. (2011). Covering chemical diversity of genetically-modified tomatoes using metabolomics for objective substantial equivalence assessment. *PLoS ONE, 6*(2)

Lei, Z., Huhman, D. V., & Sumner, L. W. (2011). Mass spectrometry strategies in metabolomics. *J Biol Chem, 286*(29), 25435-25442.

Leung, A. Y. a. F., S. (1996). *Encyclopedia of Common Natural Ingredients Used in Food,Drugs, and Cosmetics*. New York, USA.: John Wiley & Sons,.

Li, X., & Vince, R. (2006). Conformationally restrained carbazolone-containing α,γ-diketo acids as inhibitors of HIV integrase. *Bioorganic & Medicinal Chemistry, 14*(9), 2942-2955.

Liao, C., Marchand, C., Burke, T. R., Pommier, Y., & Nicklaus, M. C. (2010). Authentic HIV-1 integrase inhibitors. *Future medicinal chemistry, 2*(7), 1107-1122.

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997) 3–25.1. *Advanced Drug Delivery Reviews, 46*(1), 3-26.

Little S, D. G., Schooley R. (2005). Antiretroviral effect of L-000870810, a novel HIV-1 integrase inhibitor, in HIV-1-infected patients. *Program and abstracts of the 12th Conference on Retroviruses and Opportunistic Infections, Mass. Abstract 161.*

Lucci, P., Saurina, J., & Núñez, O. (2017). Trends in LC-MS and LC-HRMS analysis and characterization of polyphenols in food. *TrAC Trends in Analytical Chemistry, 88*, 1-24.

Lundholt, B. K., Scudder, K. M., & Pagliaro, L. (2003). A simple technique for reducing edge effect in cell-based assays. *J Biomol Screen, 8*(5), 566-570.

MacDonald, M. L., Lamerdin, J., Owens, S., Keon, B. H., Bilter, G. K., Shang, Z., Huang, Z., Yu, H., Dias, J., Minami, T., Michnick, S. W., & Westwick, J. K. (2006). Identifying off-target effects and hidden phenotypes of drugs in human cells. *Nat Chem Biol, 2*(6), 329-337.

7. Literature

Magalhaes, P. J., Vieira, J. S., Goncalves, L. M., Pacheco, J. G., Guido, L. F., & Barros, A. A. (2010). Isolation of phenolic compounds from hop extracts using polyvinylpolypyrrolidone: characterization by high-performance liquid chromatography-diode array detection-electrospray tandem mass spectrometry. *J Chromatogr A, 1217*(19), 3258-3268.

Mandal, S., Das, D. N., De, K., Ray, K., Roy, G., Chaudhuri, S. B., Sahana, C. C., & Chowdhuri, M. K. (1993). Ocimum sanctum Linn--a study on gastric ulceration and gastric secretion in rats. *Indian J Physiol Pharmacol, 37*(1), 91-92.

Manfredi, K. P., Blunt, J. W., Cardellina, J. H., McMahon, J. B., Pannell, L. L., Cragg, G. M., & Boyd, M. R. (1991). Novel alkaloids from the tropical plant Ancistrocladus abbreviatus inhibit cell killing by HIV-1 and HIV-2. *Journal of Medicinal Chemistry, 34*(12), 3402-3405.

Marshall, A. G., Hendrickson, C. L., & Jackson, G. S. (1998). Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev, 17*(1), 1-35.

Martin, A. C., Pawlus, A. D., Jewett, E. M., Wyse, D. L., Angerhofer, C. K., & Hegeman, A. D. (2014). Evaluating solvent extraction systems using metabolomics approaches. *RSC Advances, 4*(50), 26325-26334.

Maurin, C., Bailly, F., & Cotelle, P. (2004). Improved preparation and structural investigation of 4-aryl-4-oxo-2-hydroxy-2-butenoic acids and methyl esters. *Tetrahedron, 60*(31), 6479-6486.

Maurin, C., Bailly, F., Mbemba, G., Mouscadet, J. F., & Cotelle, P. (2006). Design, synthesis, and anti-integrase activity of catechol–DKA hybrids. *Bioorganic & Medicinal Chemistry, 14*(9), 2978-2984.

Mediani, A., Abas, F., Tan, C., & Khatib, A. (2014). Effects of Different Drying Methods and Storage Time on Free Radical Scavenging Activity and Total Phenolic Content of Cosmos Caudatus. *Antioxidants, 3*(2), 358.

Mehrotra, B., & Mendes, P. (2006). Bioinformatics Approaches to Integrate Metabolomics and Other Systems Biology Data. In K. Saito, R. A. Dixon & L. Willmitzer (Eds.), *Plant Metabolomics* (pp. 105-115). Berlin, Heidelberg: Springer Berlin Heidelberg.

Miroslaw, A. H., Renata, N., Monika, W.-H., Ryszard, S., & Monika, R. (2012). Two-dimensional Thin Layer Chromatographic Separation of Phenolic Compounds from Eupatorium cannabinum Extracts and their Antioxidant Activity. *Medicinal Chemistry, 8*(1), 118-131.

Moco, S., Vervoort, J., Moco, S., Bino, R. J., De Vos, R. C. H., & Bino, R. (2007). Metabolomics technologies and metabolite identification. *TrAC Trends in Analytical Chemistry, 26*(9), 855-866.

Mohammed Al-Ghorbani, B. B. A., Zabiulla, Mamatha S.V., Shaukath Ara Khanum. (2015). iperazine and Morpholine: Synthetic Preview and Pharmaceutical Applications. *Research J. Pharm. and Tech.*(8(5)), Page 611-628.

Moing, A., Aharoni, A., Biais, B., Rogachev, I., Meir, S., Brodsky, L., Allwood, J. W., Erban, A., Dunn, W. B., Kay, L., de Koning, S., de Vos, R. C., Jonker, H., Mumm, R., Deborde, C., Maucourt, M., Bernillon, S., Gibon, Y., Hansen, T. H., Husted, S., Goodacre, R., Kopka, J., Schjoerring, J. K., Rolin, D., & Hall, R. D. (2011). Extensive metabolic cross-talk in melon fruit revealed by spatial and developmental combinatorial metabolomics. *New Phytol, 190*(3), 683-696.

Mondal, S., Mirdha, B. R., & Mahapatra, S. C. (2009). The science behind sacredness of Tulsi (Ocimum sanctum Linn.). *Indian J Physiol Pharmacol, 53*(4), 291-306.

Moritz, T. J., A. I.,. (2007). *Metabolomics, Metabonomics and Metabolite Profiling*: Royal Society of Chemistry.

Mosmann, T. (1983). Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J Immunol Methods, 65*(1-2), 55-63.

Mushtaq, M. Y., Choi, Y. H., Verpoorte, R., & Wilson, E. G. (2014). Extraction for metabolomics: access to the metabolome. *Phytochem Anal, 25*(4), 291-306.

Naito, Y., Sugiura, M., Yamaura, Y., Fukaya, C., Yokoyama, K., Nakagawa, Y., Ikeda, T., Senda, M., & Fujita, T. (1991). Quantitative structure-activity relationship of catechol derivatives inhibiting 5-lipoxygenase. *Chem Pharm Bull (Tokyo), 39*(7), 1736-1745.

Nakayama, M., Suzuki, K., Toda, M., Okubo, S., Hara, Y., & Shimamura, T. (1993). Inhibition of the infectivity of influenza virus by tea polyphenols. *Antiviral Res, 21*(4), 289-299.

Nemes, P., Schlosser, G., & Vekey, K. (2005). Amino acid cluster formation studied by electrospray ionization mass spectrometry. *J Mass Spectrom, 40*(1), 43-49.

Newman, D. J., & Cragg, G. M. (2016). Natural Products as Sources of New Drugs from 1981 to 2014. *Journal of Natural Products, 79*(3), 629-661.

Newman, D. J., Cragg, G. M., & Snader, K. M. (2000). The influence of natural products upon drug discovery. *Nat Prod Rep, 17*(3), 215-234.

Nguyen, H., Lemberkovics, É., Tarr, K., Máthé Jr., I. and Petri, G. (1993). A comparative study on formation of flavonoid, tannin, and polyphenol contents in ontogenesis of Ocimum basilicum L Part I. II. *Acta Agronomica Hungarica, 42*, 31 - 39.

Nguyen, S., & Fenn, J. B. (2007). Gas-phase ions of solute species from charged droplets of solutions. *Proc Natl Acad Sci U S A, 104*(4), 1111-1117.

Nichols, A. (2007). High content screening as a screening tool in drug discovery. *Methods Mol Biol, 356*, 379-387.

Nicholson, J. K., & Lindon, J. C. (2008). *Systems biology: Metabonomics*: Nature. Oct 23;455(7216):1054-6.

Nikolaev, E. N., Kostyukevich, Y. I., & Vladimirov, G. N. (2014). Fourier transform ion cyclotron resonance (FT ICR) mass spectrometry: Theory and simulations. *Mass Spectrom Rev, 35*(2), 219-258.

Nishizawa, M., Emura, M., Kan, Y., Yamada, H., Ogawa, K., & Hamanaka, N. (1992). Macrocarpals: HIV-RTase inhibitors of Eucalyptus globulus. *Tetrahedron Letters, 33*(21), 2983-2986.

O'Rourke, A., Kremb, S., Bader, T. M., Helfer, M., Schmitt-Kopplin, P., Gerwick, W. H., Brack-Werner, R., & Voolstra, C. R. (2016). Alkaloids from the Sponge Stylissa carteri Present Prospective Scaffolds for the Inhibition of Human Immunodeficiency Virus 1 (HIV-1). *Marine Drugs, 14*(2), 28.

Ogata, T., Higuchi, H., Mochida, S., Matsumoto, H., Kato, A., Endo, T., Kaji, A., & Kaji, H. (1992). HIV-1 reverse transcriptase inhibitor from Phyllanthus niruri. *AIDS Res Hum Retroviruses, 8*(11), 1937-1944.

Oms-Oliu, G., Odriozola-Serrano, I., & Martín-Belloso, O. (2013). Metabolomics for assessing safety and quality of plant-derived food. *Food Research International, 54*(1), 1172-1183.

Pan, M. H., Lai, C. S., & Ho, C. T. (2010). Anti-inflammatory activity of natural dietary flavonoids. *Food Funct, 1*(1), 15-31.

Pang, H., Wu, L., Tang, Y., Zhou, G., Qu, C., & Duan, J.-a. (2016). Chemical Analysis of the Herbal Medicine Salviae miltiorrhizae Radix et Rhizoma (Danshen). *Molecules, 21*(1), 51.

Parcerisa, J., Codony, R., Boatella, J., & Rafecas, M. (1999). Triacylglycerol and phospholipid composition of hazelnut (Corylus avellana L.) lipid fraction during fruit development. *J Agric Food Chem, 47*(4), 1410-1415.

Pattanayak, P., Behera, P., Das, D., & Panda, S. (2010). Ocimum sanctum Linn. A reservoir plant for therapeutic applications: An overview. *Pharmacognosy Reviews, 4*(7), 95-105.

Petereit, F., Kolodziej, H., & Nahrstedt, A. (1991). Flavan-3-ols and proanthocyanidins from Cistus incanus. *Phytochemistry, 30*(3), 981-985.

Phillips, A., & Pezzotti, P. (2004). Short-term risk of AIDS according to current CD4 cell count and viral load in antiretroviral drug-naive individuals and those treated in the monotherapy era. *Aids, 18*(1), 51-58.

7. Literature

Polyakov, A. Y., Lebedev, V. A., Shirshin, E. A., Rumyantsev, A. M., Volikov, A. B., Zherebker, A., Garshev, A. V., Goodilin, E. A., & Perminova, I. V. (2017). Non-classical growth of water-redispersible spheroidal gold nanoparticles assisted by leonardite humate. *CrystEngComm, 19*(5), 876-886.

Pomponio, R., Gotti, R., Santagati, N. A., & Cavrini, V. (2003). Analysis of catechins in extracts of Cistus species by microemulsion electrokinetic chromatography. *Journal of Chromatography A, 990*(1), 215-223.

Prokofjeva, M. M., Riecken, K., Spirin, P. V., Yanvarév, D. V., Düsedau, A., Ellinger, B., Fehse, B., Stocking, C., & Prassolov, V. S. (2013). A new system for parallel drug screening against multiple-resistant HIV mutants based on lentiviral self-inactivating (SIN) vectors and multi-colour analyses. *AIDS Research and Therapy, 10*(1), 1.

Queiroz, O., & Morel, C. (1974). Photoperiodism and Enzyme Activity: Towards a Model for the Control of Circadian Metabolic Rhythms in the Crassulacean Acid Metabolism. *Plant Physiology, 53*(4), 596-602.

Rausch, O. (2006). High content cellular screening. *Current Opinion in Chemical Biology, 10*(4), 316-320.

Rebensburg, S. (2015). *Neue Ansätze zur Hemmung der HIV-Replikation*. Dissertation. München.

Rebensburg, S., Helfer, M., Schneider, M., Koppensteiner, H., Eberle, J., Schindler, M., Gurtler, L., & Brack-Werner, R. (2016). Potent in vitro antiviral activity of Cistus incanus extract against HIV and Filoviruses targets viral envelope proteins. *Sci Rep, 6*(20394).

Rege, A., Ambaye, R., & Deshmukh, R. (2010). In-vitro testing of anti-HIV activity of some medicinal plants. *Indian Journal of Natural Products and Resources (IJNPR), 193-199*, 193-199.

Reisen, F., Zhang, X., Gabriel, D., & Selzer, P. (2013). Benchmarking of multivariate similarity measures for high-content screening fingerprints in phenotypic drug discovery. *J Biomol Screen, 18*(10), 1284-1297.

Richard, P. (2010). Naturwirkstoffe aus Pflanzen. Biodiversitätsforschung. *Chemie in unserer Zeit, 44*(4), 260-274.

Riehle, P., Rusche, N., Saake, B., & Rohn, S. (2014). Influence of the Leaf Content and Herbal Particle Size on the Presence and Extractability of Quantitated Phenolic Compounds in Cistus incanus Herbal Teas. *Journal of Agricultural and Food Chemistry, 62*(45), 10978-10988.

Riehle, P., Vollmer, M., & Rohn, S. (2013). Phenolic compounds in Cistus incanus herbal infusions — Antioxidant capacity and thermal stability during the brewing process. *Food Research International, 53*(2), 891-899.

Robertson, D. G., Watkins, P. B., & Reily, M. D. (2011). Metabolomics in Toxicology: Preclinical and Clinical Applications. *Toxicological Sciences, 120*(suppl_1), S146-S170.

Rosemond, M. J., St John-Williams, L., Yamaguchi, T., Fujishita, T., & Walsh, J. S. (2004). Enzymology of a carbonyl reduction clearance pathway for the HIV integrase inhibitor, S-1360: role of human liver cytosolic aldo-keto reductases. *Chem Biol Interact, 147*(2), 129-139.

Roškar, R., & Lušin, T. T. (2012). Analytical Methods for Quantification of Drug Metabolites in Biological Samples. In L. d. A. Calderon (Ed.), *Chromatography - The Most Versatile Method of Chemical Analysis* (pp. Ch. 04). Rijeka: InTech.

Roullier-Gall, C., Witting, M., Gougeon, R. D., & Schmitt-Kopplin, P. (2014). High precision mass measurements for wine metabolomics. *Frontiers in Chemistry, 2*(102).

Ryan, D., & Robards, K. (2006). Analytical Chemistry Considerations in Plant Metabolomics. *Separation & Purification Reviews, 35*(4), 319-356.

Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V., & Quackenbush, J. (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques, 34*(2), 374-378.

Saharkhiz, M. J., Kamyab, A. A., Kazerani, N. K., Zomorodian, K., Pakshir, K., & Rahimi, M. J. (2015). Chemical Compositions and Antimicrobial Activities of Ocimum sanctum L. Essential Oils at Different Harvest Stages. *Jundishapur J Microbiol, 8*(1), e13720.

Saito, K., & Matsuda, F. (2010). Metabolomics for functional genomics, systems biology, and biotechnology. *Annu Rev Plant Biol, 61*, 463-489.

Sakagami, H., Sakagami, T. and Takeda, M. . (1995). Antiviral properties of polyphenols. *1995, 12*, 30-32.

Santagati, N. A., Salerno, L., Attaguile, G., Savoca, F., & Ronsisvalle, G. (2008). Simultaneous determination of catechins, rutin, and gallic acid in Cistus species extracts by HPLC with diode array detection. *J Chromatogr Sci, 46*(2), 150-156.

Sasidharan, S., Chen, Y., Saravanan, D., Sundram, K. M., & Yoga Latha, L. (2011). Extraction, Isolation and Characterization of Bioactive Compounds from Plants' Extracts. *African Journal of Traditional, Complementary, and Alternative Medicines, 8*(1), 1-10.

Scalbert, A., Manach, C., Morand, C., Remesy, C., & Jimenez, L. (2005). Dietary polyphenols and the prevention of diseases. *Crit Rev Food Sci Nutr, 45*(4), 287-306.

Schieber, A., Keller, P., & Carle, R. (2001). Determination of phenolic acids and flavonoids of apple and pear by high-performance liquid chromatography. *J Chromatogr A, 910*(2), 265-273.

Schmitt-Kopplin, P., Gabelica, Z., Gougeon, R. D., Fekete, A., Kanawati, B., Harir, M., Gebefuegi, I., Eckel, G., & Hertkorn, N. (2010). High molecular diversity of extraterrestrial organic matter in Murchison meteorite revealed 40 years after its fall. *Proceedings of the National Academy of Sciences, 107*(7), 2763-2768.

Schmitt-Kopplin, P., Gelencsér, A., Dabek-Zlotorzynska, E., Kiss, G., Hertkorn, N., Harir, M., Hong, Y., & Gebefügi, I. (2010). Analysis of the Unresolved Organic Fraction in Atmospheric Aerosols with Ultrahigh-Resolution Mass Spectrometry and Nuclear Magnetic Resonance Spectroscopy: Organosulfates As Photochemical Smog Constituents. *Analytical Chemistry, 82*(19), 8017-8026.

Schoenmakers, P., Marriott, P and Beens,. (2003). Nomenclature and conventions in comprehensive multidimensional chromatography. *LCGC Europe, 25*(5).

Schulze, C. J., Bray, W. M., Woerhmann, M. H., Stuart, J., Lokey, R. S., & Linington, R. G. (2013). 'Function-first' Lead Discovery: Mode of Action Profiling of Natural Product Libraries Using Image-Based Screening. *Chemistry & biology, 20*(2), 285-295.

Sechi, M., Derudas, M., Dallocchio, R., Dessì, A., Bacchi, A., Sannia, L., Carta, F., Palomba, M., Ragab, O., Chan, C., Shoemaker, R., Sei, S., Dayam, R., & Neamati, N. (2004). Design and Synthesis of Novel Indole β-Diketo Acid Derivatives as HIV-1 Integrase Inhibitors. *Journal of Medicinal Chemistry, 47*(21), 5298-5310.

SHABIR, G. A. (2004). A Practical Approach to Validation of HPLC Methods Under Current Good Manufacturing Practices. *Journal of Validation Technology, 10*.

Shahat, A. A., Cos, P., De Bruyne, T., Apers, S., Hammouda, F. M., Ismail, S. I., Azzam, S., Claeys, M., Goovaerts, E., Pieters, L., Vanden Berghe, D., & Vlietinck, A. J. (2002). *Antiviral and antioxidant activity of flavonoids and proanthocyanidins from Crataegus sinaica*: Planta Med. 2002 Jun;68(6):539-41.

Shubha, P., Namratha, K., Aparna, H. S., Ashok, N. R., Mustak, M. S., Chatterjee, J., & Byrappa, K. (2017). Facile green reduction of graphene oxide using Ocimum sanctum hydroalcoholic extract and evaluation of its cellular toxicity. *Materials Chemistry and Physics, 198*, 66-72.

Singh, S., Majumdar, D. K., & Yadav, M. R. (1996). Chemical and pharmacological studies on fixed oil of Ocimum sanctum. *Indian J Exp Biol, 34*(12), 1212-1215.

Smyth, W. F., Smyth, T. J. P., Ramachandran, V. N., O'Donnell, F., & Brooks, P. (2012). Dereplication of phytochemicals in plants by LC-ESI-MS and ESI-MSn. *TrAC Trends in Analytical Chemistry, 33*, 46-54.

Stefansky, W. (1972). Rejecting Outliers in Factorial Designs. *Technometrics, 14*(2), 469-479.

7. Literature

Stephen, B., Gary, K., & Jean-Louis, D. (2005). Metabolomics applications of FT-ICR mass spectrometry. *Mass Spectrom Rev, 24*(2), 223-231.

Sumiya, E., Shimogawa, H., Sasaki, H., Tsutsumi, M., Yoshita, K., Ojika, M., Suenaga, K., & Uesugi, M. (2011). *Cell-morphology profiling of a natural product library identifies bisebromoamide and miuraenamide A as actin filament stabilizers*: ACS Chem Biol. 2011 May 20;6(5):425-31. Epub 2011 Feb 24.

Sumner, L. W. (2010). Recent advances in plant metabolomics and greener pastures. *F1000 Biology Reports, 2*, 7.

Suzgec-Selcuk, S., Mericli, A. H., Guven, K. C., Kaiser, M., Casey, R., Hingley-Wilson, S., Lalvani, A., & Tasdemir, D. (2011). Evaluation of Turkish seaweeds for antiprotozoal, antimycobacterial and cytotoxic activities. *Phytother Res, 25*(5), 778-783.

Swinney, D. C., & Anthony, J. (2011). *How were new medicines discovered?* : Nat Rev Drug Discov. 2011 Jun 24;10(7):507-19.

Taiwo, B., Hicks, C., & Eron, J. (2010). Unmet therapeutic needs in the new era of combination antiretroviral therapy for HIV-1. *J Antimicrob Chemother, 65*(6), 1100-1107.

Tan, G. T., Pezzuto, J. M., Kinghorn, A. D., & Hughes, S. H. (1991). Evaluation of Natural Products as Inhibitors of Human Immunodeficiency Virus Type 1 (HIV-1) Reverse Transcriptase. *Journal of Natural Products, 54*(1), 143-154.

Taylor, P. J. (2005). Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography–electrospray–tandem mass spectrometry. *Clinical Biochemistry, 38*(4), 328-334.

Tervo, A. J., Suuronen, T., Kyrylenko, S., Kuusisto, E., Kiviranta, P. H., Salminen, A., Leppänen, J., & Poso, A. (2006). Discovering Inhibitors of Human Sirtuin Type 2: Novel Structural Scaffolds. *Journal of Medicinal Chemistry, 49*(24), 7239-7241.

Thomas, M., Badr, A., Desjardins, Y., Gosselin, A., & Angers, P. (2018). Characterization of industrial broccoli discards (Brassica oleracea var. italica) for their glucosinolate, polyphenol and flavonoid contents using UPLC MS/MS and spectrophotometric methods. *Food Chemistry, 245*, 1204-1211.

Trethewey, R. N. (2004). Metabolite profiling as an aid to metabolic engineering in plants. *Curr Opin Plant Biol, 7*(2), 196-201.

Tu, Y., Jeffries, C., Ruan, H., Nelson, C., Smithson, D., Shelat, A. A., Brown, K. M., Li, X.-C., Hester, J. P., Smillie, T., Khan, I. A., Walker, L., Guy, K., & Yan, B. (2010). Automated High-Throughput System to Fractionate Plant Natural Products for Drug Discovery. *Journal of Natural Products, 73*(4), 751-754.

Tziotis, D., Hertkorn, N., & Schmitt-Kopplin, P. (2011). *Kendrick-analogous network visualisation of ion cyclotron resonance Fourier transform mass spectra: improved options for the assignment of elemental compositions and the classification of organic molecular complexity* (Vol. 17): Eur J Mass Spectrom (Chichester).

Usha, P. R., Naidu, M. U. R., & Raju, Y. S. N. (2003). Evaluation of the Antiretroviral Activity of a New Polyherbal Drug (Immu-25) in Patients with HIV Infection. *Drugs in R & D, 4*(2), 103-109.

van der Burgt, Y. E. M., Taban, I. M., Konijnenburg, M., Biskup, M., Duursma, M. C., Heeren, R. M. A., Römpp, A., van Nieuwpoort, R. V., & Bal, H. E. (2007). Parallel Processing of Large Datasets from NanoLC-FTICR-MS Measurements. *Journal of the American Society for Mass Spectrometry, 18*(1), 152-161.

Vasudevan P, K. S., Sharma S. (1999). Bioactive botanicals from basil (Ocimum sp.) *J Sci Ind Res, (C)*(58), 332–338.

Villas-Boas, S. G., Mas, S., Akesson, M., Smedsgaard, J., & Nielsen, J. (2005). Mass spectrometry in metabolome analysis. *Mass Spectrom Rev, 24*(5), 613-646.

Volberding, P. A., Lagakos, S. W., Koch, M. A., Pettinelli, C., Myers, M. W., Booth, D. K., Balfour, H. H., Jr., Reichman, R. C., Bartlett, J. A., Hirsch, M. S., & et al. (1990). Zidovudine in asymptomatic human

immunodeficiency virus infection. A controlled trial in persons with fewer than 500 CD4-positive cells per cubic millimeter. The AIDS Clinical Trials Group of the National Institute of Allergy and Infectious Diseases. *N Engl J Med, 322*(14), 941-949.

Wagenaar, M. M. (2008). Pre-fractionated microbial samples--the second generation natural products library at Wyeth. *Molecules, 13*(6), 1406-1426.

Watson, J. T. S. (2013). *Introduction to Mass Spectrometry: Instrumentation, Applications, and Strategies for Data Interpretation*. New York: John Wiley & Sons,.

Watzl, L. (2005). *Bioaktive Substanzen in Lebensmitteln.* Stuttgart: Hippokrates Verlag,.

Weckwerth, W. (2007). *Methods and Protocols*. Berlin Heidelberg,: Springer Science & Business Media.

Willemse, C. M., Stander, M. A., Tredoux, A. G. J., & de Villiers, A. (2014). Comprehensive two-dimensional liquid chromatographic analysis of anthocyanins. *Journal of Chromatography A, 1359*, 189-201.

Witt. (1964). *Knaurs Pflanzenreich in Farben*: Droemersche Verlagsanstalt A. G., Zürich.

Wolfe, M. H. D. A. (1973). Nonparametric Statistical Methods (pp. 185--194). New York: John Wiley & Sons.

Wolfender, J. L., Rudaz, S., Choi, Y. H., & Kim, H. K. (2013). Plant metabolomics: from holistic data to relevant biomarkers. *Curr Med Chem, 20*(8), 1056-1090.

Xu, Z., Chang, F. R., Wang, H. K., Kashiwada, Y., McPhail, A. T., Bastow, K. F., Tachibana, Y., Cosentino, M., & Lee, K. H. (2000). Anti-HIV agents 45(1) and antitumor agents 205.(2) two new sesquiterpenes, leitneridanins A and B, and the cytotoxic and anti-HIV principles from Leitneria floridana. *J Nat Prod, 63*(12), 1712-1715.

Yasmina, L., Chao, M. M., T., K. I., Hirotsugu, M., Masao, H., P., G. M., & Mireye, C. (1997). HIV-1 reverse transcriptase inhibitory principles from Chamaesyce hyssopifolia. *Phytotherapy Research, 11*(1), 22-27.

Young, D. W., Bender, A., Hoyt, J., McWhinnie, E., Chirn, G.-W., Tao, C. Y., Tallarico, J. A., Labow, M., Jenkins, J. L., Mitchison, T. J., & Feng, Y. (2007). Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nature Chemical Biology, 4*, 59.

Young, D. W., Bender, A., Hoyt, J., McWhinnie, E., Chirn, G. W., Tao, C. Y., Tallarico, J. A., Labow, M., Jenkins, J. L., Mitchison, T. J., & Feng, Y. (2008). Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nat Chem Biol, 4*(1), 59-68.

Zheljazkov, V. D., Cantrell, C. L., Tekwani, B., & Khan, S. I. (2008). Content, composition, and bioactivity of the essential oils of three basil genotypes as a function of harvesting. *J Agric Food Chem, 56*(2), 380-385.

Zuo, Y., Wang, C., & Zhan, J. (2002). Separation, characterization, and quantitation of benzoic and phenolic antioxidants in American cranberry fruit by GC-MS. *J Agric Food Chem, 50*(13), 3789-3794.

8. Scientific communications

Paper

- B. Kanawati, <u>T. Bader</u>, K.-P. Wanczek, Y. Li, P. Schmitt-Kopplin, **FT-Artifacts and Power-function Resolution Filter in Fourier Transform Mass Spectrometry**. Rapid Commun. Mass Spectrom, 31, 1607-1615 (2017)

- O'Rourke, A. ; Kremb, S. ; <u>Bader, T.M</u>. ; Helfer, M. ; Schmitt-Kopplin, P. ; Gerwick, W.H. ; Brack-Werner, R. ; Voolstra, C.R., **Alkaloids from the sponge Stylissa carteri present prospective scaffolds for the inhibition of Human Immunodeficiency Virus 1 (HIV-1)**. Mar. Drugs 14:28 (2016)

- Velikova, V.B. ; Müller, C. ; Ghirardo, A. ; <u>Rock, T.</u> ; Aichler, M. ; Walch, A.K. ; Schmitt-Kopplin, P. ; Schnitzler, J.-P. **Knocking down of isoprene emission modifies the lipid matrix of thylakoid membranes and influences the chloroplast ultrastructure in poplar**. Plant Physiol. 168, 859-870 (2015)

- Müller, C.* ; Fonseca, J.R.* ; <u>Rock, T.*</u> ; Krauss-Etschmann, S. ; Schmitt-Kopplin, P. **Enantioseparation and selective detection of D-amino acids by ultra-high-performance liquid chromatography/mass spectrometry in analysis of complex biological samples**. J. Chromatogr. A 1324, 109-114 (2014)
  (*Authors contributed equally)

- Müller, C.* ; Kanawati, B. *; <u>Rock, T.</u>* ; Forcisi, S. ; Moritz, F. ; Schmitt-Kopplin, P. **Dimer ion formation and intermolecular fragmentation of 1,2-diacylglycerols revealed by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for more comprehensive lipid analysis**. Rapid Commun. Mass Spectrom. 28, 1735-1744 (2014)
  (*Authors contributed equally)

Conference – Poster

08/2016   „International Mass Spectrometry Conference (IMSC)" in Toronto, Canada: „Characterization of anti-HIV-1 active compound classes from *Cistus incanus* via LC-fractionation and ultrahigh resolution mass spectrometry (FT-ICR-MS)"

09/2014   "Metabomeeting" in London, Great Britain: "dimer ion formation and inter-molecular fragmentationof 1,2-diacylglycerols revealed by electrospray FT-ICR-MS for more comprehensive lipids analysis"

9. Curriculum vitae

Personal information

Name: Theresa Maria Bader
Date of Birth: 02.02.1988

Scientific work experience

01/2015 – 02/2015	**Guest scientist**, King Abdullah University for Science and Technology (KAUST), Red Sea Research Center, Saudi Arabia
03/2013 – 08/2013	**Guest scientist**, HelmholtzZentrum Munich, Research Unit Analytical BioGeoChemistry (BGC), Technical University of Munich (TUM),

Education

08/2013 – 10/2016	**PhD student**, HelmholtzZentrum Munich, Research Unit Analytical BioGeoChemistry (BGC), Technical University of Munich (TUM), Chair of Analytical Food Chemistry *Thesis: Isolation and characterization of antivirals against HIV-1 from plant extracts*
10/2010 – 03/2013	**Master of Education in Science in chemistry and biology (M.Ed.),** Technische Universität München *Thesis: "Occurrence of catecholamines and other biomarkers in stressed and non-stressed amphipod"*
10/2007 – 10/2010	**Bachelor of Education in Science in chemistry and biology (B.Ed.)**, Technische Universität München *Thesis: Synthese und Charakterisierung eines Hybrid-Additives für Portlandzement*
10/2007 – 07/2011	**First examination in educational science** psychology, general pedagogy and school pedagogic
09/1999 – 06/2007	**Abitur** (University entrance qualification), Johann-Sebastian-Bach-Gymnasium, Windsbach