Biosynthesis and Biological Profiling of Collinolactone and Semisynthetic Derivatives

and

MetabolDent, a Novel Tool for Automated Dereplication

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

Zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Julian Christopher Schmid

aus Tübingen

Tübingen

2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	10.02.2023
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Prof. Dr. Stephanie Grond
2. Berichterstatter/-in:	PD Dr. Bertold Gust
3. Berichterstatter/-in:	Prof. Dr. Valentin Wittmann

Die vorliegende Arbeit wurde im Institut für Organische Chemie der Eberhard Karls Universität Tübingen von Mai 2018 bis Juni 2022 unter der Anleitung von Prof. Dr. Stephanie Grond durchgeführt.

SUMMARY

The rising number of neurodegenerative diseases such as Alzheimer, increasing cancer rates and the global spread of antibiotic-resistant bacteria demand novel drugs and therapeutic concepts. Nature has already proven to be an extraordinary rich source of novel structural motifs which can serve as lead structures for drug development.[1]

In this thesis, a novel natural product was isolated from *Streptomyces sp.* Gö 40/10 and studied in detail. The compound collinolactone possess an unprecedented and unique 6-10-7-membered tricyclic system with a cyclodecatriene ring flanked by two lactone rings.[2] Based on stable precursor feeding, a biosynthetic hypothesis was developed, which was then used for the identification of the encoding gene cluster. The fermentation parameters and the isolation procedure were optimized in order to achieve a stable yield of about 25 mg/L.

A representative and divers set of derivatives was synthesized using a variety of different reactions including the use of Burgess reagent, Strykers reagent and Gilman cuprates. All compounds were profiled in a cell-viability assay on L929 cell line, where some compounds showed cytotoxicity in the micromolar range. The mode of action was studied in more detail using fluorescence microscopy on PtK2 cell line and an increased number of monopolar spindles was found. The inhibition of putative molecular targets was investigated in a malachite green ATP-based assay.
Furthermore, collinolactone was found to exhibit neuroprotective properties in a glutamate induced intramolecular oxidative stress assay on HT22 cell line. Only collinolactone showed an effect here, indicating that the activity is highly specific and even the smallest structural changes will lead to a loss of functionality. In addition, the reduction of amyloid beta protein aggregates, one of the hallmarks for the development of Alzheimer disease, was studied. Again, collinolactone showed a reduction of aggregation, making it a very promising lead structure for further development towards a drug for Alzheimer therapy.

Since the 1950s, millions of substances isolated from biological sources have already been described in literature and the chances of reisolating already known structures is higher than ever. Consequently, reliable, and efficient tools for the early annotation of already known substances in complex samples derived from nature can save time and money and are desperately required.

One aim of the research presented here was the prototyping of a standalone application for automated and solid annotation of known compounds based on LC-MS datasets. This led to the development of MetaboIDent, which was especially designed and organized with particular focus on easy installation and usage without deep knowledge of computers or programming skills. The platform includes functionality to import compound libraries from various commercial and freely accessible sources. Retention times for compounds, where no standard is available, can be predicted by training of different models including support vector machine (SVM) and random forest (RF) on a small reference data set. Spectral matching is performed based on reference spectra and is further supported by the popular MetFrag and CFM-ID *in-silico* fragmentation tools.

ZUSAMMENFASSUNG

Die steigende Zahl neurodegenerativer Erkrankungen wie Alzheimer, die wachsende Anzahl an neuen Fällen von Krebs und die weltweite Ausbreitung antibiotikaresistenter Bakterien erfordert neue Arzneimittel und Therapiekonzepte. Die Natur hat sich bereits als unglaublich reiche Quelle für neuartige Strukturmotive erwiesen, die als Leitstrukturen für die Arzneimittelentwicklung dienen können.[1]

Im Rahmen eines klassischen Screenings wurde ein neuartiger Naturstoff aus *Streptomyces sp.* Gö 40/10 isoliert und eingehend untersucht. Die Verbindung Collinolacton besitzt ein beispielloses und einzigartiges 6-10-7-gliedriges trizyklisches System mit einem Cyclodecatrienring, der von zwei Lactonringen flankiert wird.[2] Auf der Grundlage von Fütterungsexperimenten mit isotopenmarkierten biosynthetischen Vorstufen wurde eine Biosynthesehypothese entwickelt, die dann zur Identifizierung des kodierenden Genclusters verwendet wurde. Weiterhin wurden die Fermentationsparameter und das Isolierungsverfahren optimiert, um eine reproduzierbare Ausbeute von etwa 25 mg/L zu erreichen.
Eine Vielzahl repräsentativer Derivate wurden durch unterschiedliche Synthesestrategien wie Gilman-Cupraten und Strykers-Reagenz dargestellt. Alle Verbindungen wurden auf ihre Zytotoxizität an der murinen L929-Zelllinie getestet. Mithilfe der Fluoreszenzmikroskopie an der PtK2-Zelllinie wurde der Wirkmechanismus studiert, wobei eine erhöhte Anzahl von monopolaren Spindeln festgestellt wurde. Die Hemmung mutmaßlicher molekularer Ziele wurde in einem ATP-basierten Malachitgrün-Assay untersucht.

In einem Glutamat-induzierten intramolekularen oxidativen Stress-Assay an der HT22-Zelllinie wurden die neuroprotektive Eigenschaften aufweist. Nur Collinolacton zeigte hier eine Wirkung, was darauf hindeutet, dass die Aktivität sehr spezifisch ist und selbst kleinste strukturelle Veränderungen zu einem Verlust der Funktionalität führen. Darüber hinaus wurde die Fähigkeit zur Verringerung von Amyloid-Beta-Protein-Aggregaten untersucht. Diese stehen im Verdacht, für die Entwicklung der Alzheimer-Krankheit verantwortlich zu sein. Auch hier zeigte Collinolacton eine Verringerung der Aggregation, was es zu einer vielversprechenden Leitstruktur für die weitere Entwicklung eines Arzneimittels zur Alzheimer-Therapie macht.

Seit den 1950er Jahren sind bereits Millionen von aus biologischen Quellen isolierten Substanzen in der Literatur beschrieben worden, und die Wahrscheinlichkeit, dass bereits bekannte Strukturen erneut isoliert werden, ist größer denn je.

Folglich sind zuverlässige und effiziente Werkzeuge für die frühzeitige Annotation bereits bekannter Substanzen in komplexen Proben aus der Natur dringend erforderlich, um Zeit und Geld zu sparen.

Ziel der hier vorgestellten Forschung war die Entwicklung eines Prototyps für eine eigenständige Anwendung zur automatisierten und verlässlichen Annotation bekannter Substanzen in LC-MS-Datensätzen. Dies führte zur Entwicklung von MetaboIDent, welches mit besonderem Augenmerk auf eine einfache Installation und Nutzung ohne tiefgreifende Computerkenntnisse oder Programmierkenntnisse konzipiert wurde. Die Plattform beinhaltet Funktionen zum Import von Substanzbibliotheken aus verschiedenen kommerziellen und frei zugänglichen Datenbanken. Retentionszeiten für Verbindungen, für die kein Standard verfügbar ist, können durch Training verschiedener Modelle, einschließlich Support Vector Machine (SVM) und Random Forest (RF), mit einem kleinen Referenzdatensatz vorhergesagt werden. Der spektrale Abgleich erfolgt auf der Grundlage von Referenzspektren und wird durch die weit verbreiteten *In-silico*-Fragmentierungstools MetFrag und CFM-ID unterstützt.

DECLARATION OF PERSONAL CONTRIBUTIONS

Some of the data presented in this thesis have already been published in: "The Structure of Cyclodecatriene Collinolactone, its Biosynthesis, and Semisynthetic Analogues: Effects of Monoastral Phenotype and Protection from Intracellular Oxidative Stress", **J. C. Schmid**, K. Frey, M. Scheiner, J. F. G. Garzón, L. Stafforst, J.-N. Fricke, M. Schuppe, H. Schiewe, A. Zeeck, T. Weber, I. Usón, R. Kemkemer, M. Decker, S. Grond, *Angew. Chem. Int. Ed.* **2021**, *60*, 23212-23216.

A more detailed description of the sequencing and strain isolation can be found in: "The complete genome of the collinolactone producer *Streptomyces sp.* Gö40/10", E. Sterndorff, T. S. Jørgensen, J. Guerrero-Garzón, **J. C. Schmid**, S. Grond, T. Weber, *Microbiology Resource Announcements,* **2022***,11*.

The genetical engineering and the fermentation process are part of a patent application: **J. C. Schmid**, J. F. G. Garzón, A. Zeeck, T. Weber, S. C. Grond, Germany, 2021, Patent-Number: DE 10 2021 003 944.0.

Personal contribution to data presented in this thesis, which are also part of the above-mentioned publications:
I developed and optimized the production and isolation procedures for collinolactone and its biosynthetic precursor collinoketone. This includes a detailed comparison and evaluation of NMR data from the proposed structure of rhizolutin and collinolactone.
All presented feeding experiments were performed, analyzed, and interpreted by me.
The experimental part of the strain sequencing and engineering was done by Jaime Felipe Guerrero Garzón (DTU Denmark), but the prior selection of gene cluster candidates was done in close consultation with me. I cultivated, analyzed and screened all engineered strains.
I planned and performed the chemical synthesis and purification, including the analytical identity checks by HR-MS and NMR for all derivatives.
Also, I performed cell viability assays and fluorescence microscopy after a general initial training in cell cultivation techniques by Kerstin Frey (Reutlingen University). I generated and recorded all related data, including the analysis and interpretation.
In addition, I developed, validated, and performed all cell free assays presented here. The neuroprotection assay presented in this thesis was designed and performed by Matthias Scheiner (University of Würzburg), who also did the data analysis.
During the whole study, I was under the supervision of Prof. Dr. Stephanie Grond.

TABLE OF CONTENT

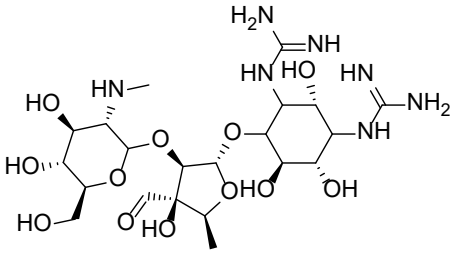
1. INTRODUCTION

1.1 Natural products as promising source for drug discovery

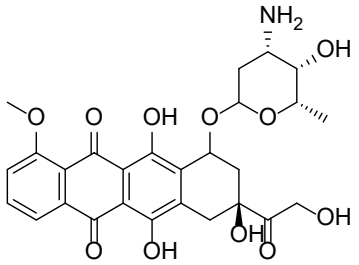
Natural products (NPs) possess an enormous structural and chemical diversity, which is unmatched by any existing synthetic small molecule library.[1] They are evolutionarily optimized over the course of millions of years for their individual and specific purposes in their habitats, even though scientists often can only speculate about the original function of these complex molecules in nature.

Natural products already played a major role in human life for thousands of years, with the first written compendium of 1000 plants and plant-derived materials and their indications dating back to 2600 BC in Mesopotamia.[3] Even today, about 60 % of the world's population trusts in traditional medicine, especially in development countries due to their low cost and availability according to the World Health Organization (WHO).[4,5]

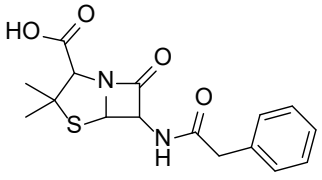
The outstanding importance and relevance of natural products for today's life can be derived from some remarkable discoveries. For instance, the Nobel Prize in Physiology or Medicine was awarded to Sir Alexander Fleming, Ernst Chain and Sir Howard Florey in 1945 for the discovery of penicillin, a β-lactam antibiotic and in 1952, to Selman Waksman for the discovery of streptomycin, another antibiotic which is now classified as critically important for human medicine by the WHO.[6]



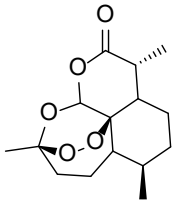
Streptomycin



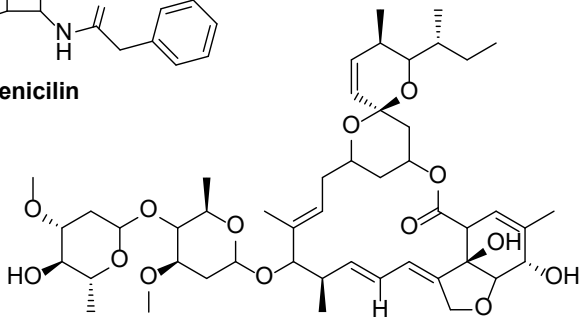
Doxorubicin



Penicilin



Artemisinin



Avermectin

These findings were the beginning of the golden age of natural product discovery, with a high point in the 1960s, when countless new natural substances have been isolated and studied, e.g., antibiotics such as erythromycin[7] and vancomycin[8] or anti-cancer drugs such as doxorubicin.[9,10] Even today, natural products still have a high impact as the awarding of the 2015 Nobel Prize in Physiology or Medicine to William Campbell and Satoshi Ōmura for the discovery of the natural product class of avermectins[11] and Tu Youyou for the discovery of the plant natural product artemisinin proves.[12,13]

1.2 Natural product discovery workflow

Nature has proven to be an extraordinary rich source of unique natural product (NP) scaffolds as lead structure for drug development. The diversity across the discovered structures is also attributed to the various sources that are exploited. For instance, plants and extracts thereof have already been used as a source for thousands of years in “traditional medicine”. With the discovery of penicillin, bacteria and fungi came into focus. More recently, animals such as ants,[14] bees,[15] or snakes[16] were found to be a rich source of novel natural products. In addition, some of the most potent natural products were found in organisms collected in the ocean’s unique habitats.[17-19]

Access to unique and promising sources of natural products is the first step in the workflow of natural product discovery (Figure 1) and probably the most important one for successful isolation of novel chemical structures. The potential of these sources can be estimated prior to cultivation and isolation by genome mining. Genome mining approaches combine the technological improvements in genome sequencing such as PacBio[20] or Nanopore[21] sequencing and bioinformatic algorithms and tools. For example, antiSMASH[22,23] or BLAST[24] can provide insight into the biosynthetic potential of the source. However, the presence of a biosynthetic gene cluster (BGC) encoding a natural product which is likely unknown, does not necessarily mean that this encoded NP is also synthesized by the organism. Most of the BGCs (approximately 90 %) identified in various sources of bacteria are silent under laboratory conditions and the corresponding products cannot be isolated.[25,26]

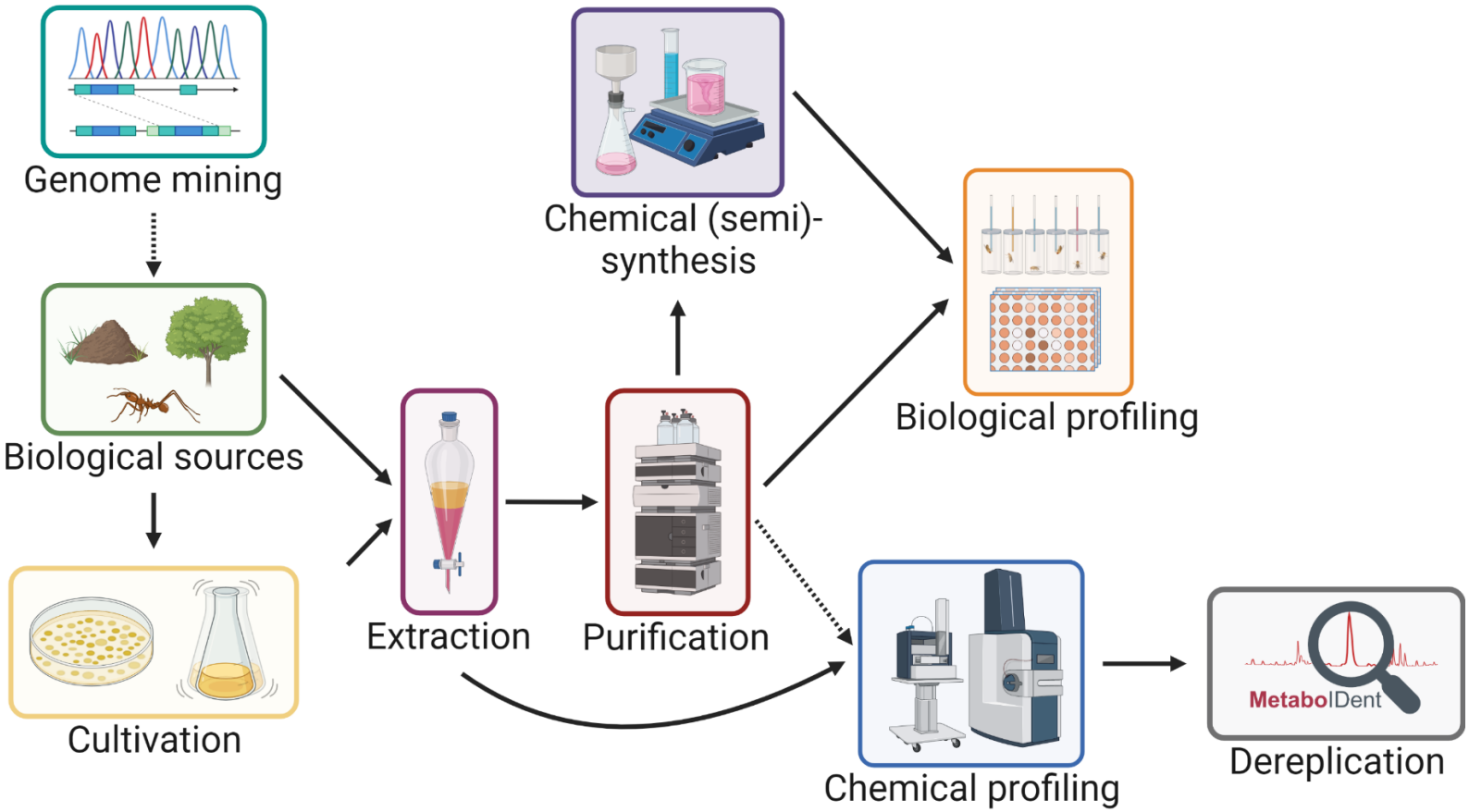


Figure 1: Established natural product discovery workflow. The workflow can differ between labs depending on research interest and available techniques in the lab. Dashed arrows indicate steps which are not always part of the workflow.[1]

[1] Created with BioRender.com.

Once a promising source is identified, sufficient biomaterial must be gathered. It can either be collected directly from nature or produced by cultivation under laboratory conditions. The biomass is then extracted, followed by various purification steps, depending on the complexity of the sample and the properties of interest (e.g., only water-soluble molecules or only lipids).

With the pure compound in hand, the next step involves biological and chemical profiling. Chemical profiling uses state-of-the art technologies such as high-resolution mass spectrometry (HR-MS), nuclear magnetic resonance (NMR) spectroscopy and other analytical and physicochemical methods to determine the compound's 2D and ideally 3D structure as well as other molecular characteristics. Biological profiling is applied to investigate the potential of the isolated substance towards an application in medicinal therapies, for example as an antibiotic or an anti-cancer drug. Semisynthetic modifications and total synthesis of the natural product and derivatives can be used to enable studies towards the mode of action (MoA) and to improve its pharmacological properties previously found during the biological profiling step.

One of the major bottlenecks of this workflow is the time-consuming extraction and purification. Since millions of compounds have already been described in literature, the rediscovery rate of already known substances is continuously increasing. Consequently, a very important and time saving step is the fast identification of already known compounds in extracts prior to extraction and purification. This process is called dereplication.

Mass spectrometry coupled to a high-performance liquid chromatography instrument (HPLC-MS) is the most frequently used setup for dereplication due to its high throughput, the online chromatographic separation, and the high sensitivity. Retained parameters such as exact mass, retention time and multi-stage fragmentation pattern are then used to search databases with reference spectra. One major problem is the small number of available reference spectra and commercially available compounds compared to the huge number of molecules that have already been identified across the world. To overcome this issue, tools for the *in-silico* prediction of fragmentation spectra based on the structure only, have been developed. One prominent example is MetFrag, one of the earliest available tools, that uses a simple bond-disconnection approach where each bond in a candidate molecule is broken and the masses of the obtained fragments are compared to the provided spectrum of the unknown substance by the user.[27,28]

Another example is CFM-ID, which models the fragmentation process based on thousands of recorded training spectra on the same instrument and uses statistical techniques to optimize parameters.[29-31]

The most advanced and most accurate *in-silico* predication is done by SIRUS.[32] This tool uses high-resolution mass spectrometry data and isotope pattern analyses to collect a list of compound candidates. For each entry, a fragmentation tree is predicted based on a trained algorithm, which is then matched with the generated tree from the input spectra.[33]

Besides the spectral prediction, another approach called molecular networking was developed.[34] Here, spectra are clustered based on their similarity. Annotations of reference spectra can then be propagated through the generated network to the unknown, but related input spectra.[35] In principle, this approach does not help to overcome the problem with the lack of available reference spectra, but Wang *et al.* developed an ecosystem called Global Natural Products Social Molecular Networking (GNPS) which provides access for all researches across the world to the voluntarily shared reference spectra of the scientific community.[36,37]

Although most technologies for dereplication focus on the use of mass spectrometry data, recent developments in NMR e.g., improved sensitivity and sophisticated 2D-NMR experiments have led to the development of “Small Molecule Accurate Recognition Technology” (SMART).[38] With SMART, recorded $^{1}H,^{13}C$-HSQC (Heteronuclear Single Quantum Coherence) data of pure or enriched fractions of natural products can be mapped into a chemical space and linked to similar HSQC reference spectra using algorithms similar to facial recognition techniques.[39]

All these advantages in the fields of bioinformatics and cheminformatics in addition to improved analytical technologies and novel tools for data analysis ushered a second golden age of natural product drug discovery.[1] The expected discovery of novel structures thereof are desperately needed as lead structures for drug development with the rise of multiresistant bacteria,[40-42] worldwide virus pandemics[43] and an expected increase in age-related diseases such as dementia[44] and cancer.[45]

1.3 Cancer – a worldwide health issue

Cancer is defined as a large group of diseases that can start in any organ or tissue, but for some organs such as the heart, only very view cases are reported per year.[46] Hallmarks are unregulated growth of tissue with no respond to growth-promoting signaling molecules and the suppression of programmed cell death (immortality). Consequently, a cancerous cell divides almost indefinitely.[47] One of the main problems usually is, that the cancer tissue is usually not confined to its original location but can spread from one part of the body to another organ or region. This process is referred to as metastasizing. It is often discussed that cancer is a consequence of today's extended life expectations and related to the unhealthy lifestyle, but cancer has already been described in a papyrus approximately 3000 BC. It is even hypothesized that cancer is as old as humankind.[48] Several remedies were developed by the Egyptians and other advanced civilizations of this time, such as surgical removal with knives and heated curling irons or treatment with arsenic pastes.[49]

According to the World Health Organization (WHO), around one-third of deaths from cancer are due to tobacco use, high body mass index, alcohol consumption, low fruit and vegetable intake, and lack of physical activity.[50,51]

Even though today's medical treatments are far more sophisticated, the burden of cancer is on a global rise.[52]

1.3.1 Worldwide cancer data

Cancer is the second leading cause of death, accounting for approximately 10 million deaths worldwide in 2020. In countries with a high Human Development Index (HDI), especially in Europe, the United States and Australia, it is even ranked first place (Figure **2**). With the globally increasing HDI over the past decades, it is expected that cancer will advance to the leading cause of death worldwide in the 21st century.[53]

In 2020, 19.3 million people worldwide were diagnosed with cancer and this number is expected to increase to 28.9 million annual cases in 2040 due to aging and growth of the population.[54] Besides the personal impact on families and friends, this burden exerts high financial strain on the national health systems, especially in low-and middle-income countries.[55,56] The national expenditures for cancer care in the United States in 2018 were $150.8 billion and costs are expected to increase with the rise in cancer incidences and prevalence and because novel, even more expensive treatments will enter the market.

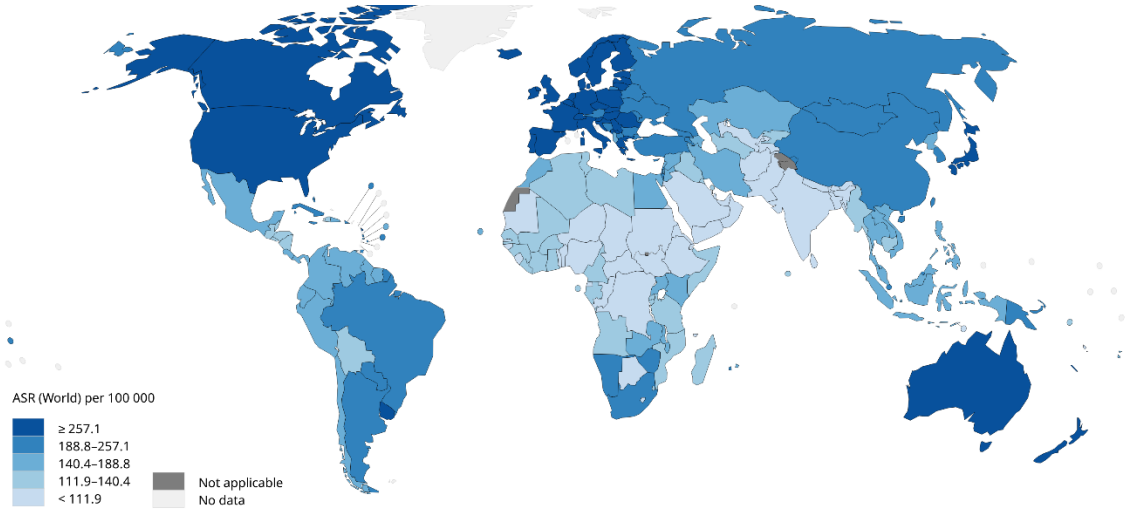


Figure 2: Age-standardized incidence rates (ASR)[2] of all types of cancer across the world in 2020. Presented data include humans of all genders and no age restrictions. Countries with high incidence rates are colored in dark blue, while countries with low incidence rates are displayed in light blue. Countries for which no data were available are displayed in grey.
The graph was generated with IARC (http://www.iarc.fr/), based on the GLOBOCAN 2020 data source, a service by the World Health Organization.

There are significant differences between sexes (male and female) when looking at the incidence and mortality rates of the most common types of cancer (Figure 3). Men are mainly diagnosed with lung cancer (23 %), followed by prostate cancer (22 %) and types of colorectal cancer (17 %). For women, breast cancer is ranked first, responsible for 39 % of all cases alone, followed by types of colorectal cancer (13 %) and lung cancer (12 %).[57,58] Notably, almost 45 % of all cases (both sexes) are related to sexual characteristics. Unaffected by sex, leukemia ranks first as the most common type of cancer for children and adolescents of ages 0 – 24.

Lung cancer still has the highest mortality rates for both men and women. Its five-year survival rate is estimated to 18 % which is significantly lower than many other leading cancer types, such as breast (almost 90 %) and prostate (98 %). More than half of the patients with lung cancer die within one year after diagnosis.[59]

Over the past decade, novel therapeutic approaches led to an increase of survival rates and improved life quality of patients diagnosed with cancer.

[2] Standardization of rates account for the differences in the age structure of populations being compared. It is especially important when comparing several populations that differ in respect to age which has a powerful influence on the risk of dying from cancer. Age structures of populations can differ e.g., due to different geographical areas which are compared.

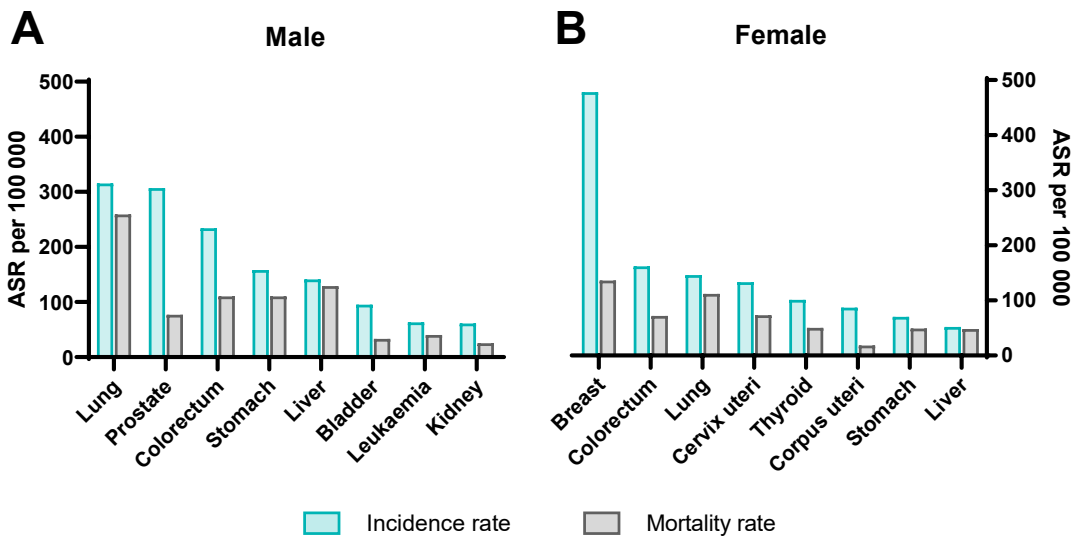


Figure 3: Age-standardized rates (ASR) of selected cancer types across the world in 2020 for males (A) and females (B). Displayed in turquoise is the incidence rate, displayed in grey the mortality rate.
Graph is based on data from GLOBOCAN 2020, Global Cancer Observatory (http://gco.iarc.fr/), International Agency for Research on Cancer 2022.

1.3.2 Therapeutic concepts and molecular targets of cancer therapy

The increasing knowledge and understanding of cancer using state-of-the-art methods such as bioinformatics and genomics[60-62] has led to the development of novel therapeutic concepts such as gene therapy[63,64] and cancer vaccination.[65] However, since these novel concepts are often only applicable to certain cancer types, standard therapies will also continue to play a major role in cancer treatment. These standard therapies typically include surgery, radiotherapy, and chemotherapy. To increase the chances of success, a combination of multiple therapies is often advisable.

Surgery is performed not only to remove the tumors partially or completely, but also for diagnostic reasons, i.e., biopsies. A small piece of tissue is collected and then tested to characterize the type of cancer before a treatment plan is developed. Side effects of this therapy are severe levels of pain and a high risk of wound infection.[66]

Often combined with a surgery, radiotherapy is applied to reduce the size of the tumor and therefore increase the rate of success. Here, high-energy rays (X-ray or gamma-ray) are used to damage the deoxyribonucleic acid (DNA) of cancer cells, so that these can no longer grow and divide, and eventually they will die. There are two kinds of radiotherapy, depending on where the radiation comes from. The radiation can be generated outside and is then directed as a beam to the tumor tissue, or it can be generated inside the body using carrier drugs with radioactive nuclei such as

^{137}Cs or ^{125}I.[67] Patients often feel very tired and weak after radiation, and side effects can appear directly after the treatment but also after several years.

Chemotherapy is a third and orthogonal concept, where drugs are applied either intravenously or orally. These drugs are cytotoxic and eventually kill the malignant cells. This approach makes use of the fact that degenerated cells divide much faster than healthy cells.[68] Most drugs are applied systemically and transported in the blood stream and can therefore not only reach cancer cells anywhere in the patient's body but also affect healthy cells. These cells usually can recover from damage using a wide range of cellular repair mechanisms[69] which are often disrupted in cancer cells.[70,71] Nevertheless, the impact on healthy cells can cause significant side effects such as loss of hair or nausea and vomiting.

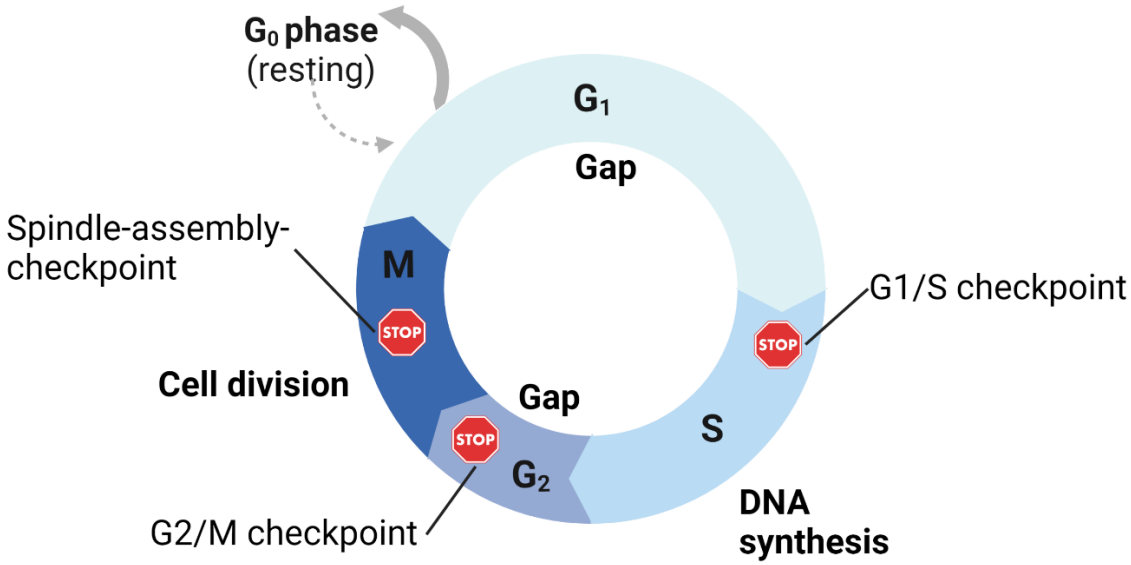


Figure 4: Schematic representation of the cell cycle found in mammalian cells. Stop signs represent the checkpoints which serve as potential termination points.[3]

Almost all chemotherapeutic drugs impact the cell cycle, which is defined as a series of temporally ordered events that leads a cell to divide itself into two daughter cells with an exact copy of the original genetic material. A schematic representation of the cell cycle (Figure 4) highlights the cell cycle with the four phases (or five, depending on how the phases are counted): in G1 (gap), the cells are metabolically active and prepare for the upcoming S phase, but no DNA replication is taking place. The duration of the G1 phase greatly varies depending on the cell type. More specifically, embryonic cells and cancer cells often have no or only a very short G1 phase.[72,73] Fully differentiated cells on the other hand, e.g., neuronal cells responsible for long term memory, have an extended or

[3] Created with BioRender.com.

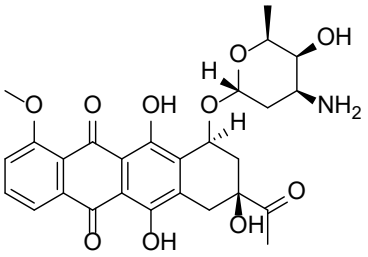
possibly indefinitely G1 phase, which is then often referred as an additional G0 resting phase (quiescence).[74] A representative proliferating human cell requires 24 hours for a full cell cycle, where the G1 phase lasts about 11 hours, the S phase approximately 8 hours, the G2 phase four hours and the M phase only about one hour.[75]

The following phase after G1 is called S phase, where the DNA is replicated while the protein synthesis rates are very low. In addition to the DNA, the centrosome, a cell organelle which is responsible for the assembly of the microtubule spindle apparatus is also duplicated in this phase.[76] Since it is impossible to move backwards in the cell cycle, a cell in S phase is now fully committed to cell division.
After the S phase, there is an additional gap phase called G2, where protein synthesis levels are high, and the cell prepares for the upcoming mitosis phase.
In the final phase called M (mitosis) phase, the cell divides by passing through a heavily regulated and controlled subset of phases which is discussed in more detail in chapter 1.3.3 – Molecular mechanism of mitosis.

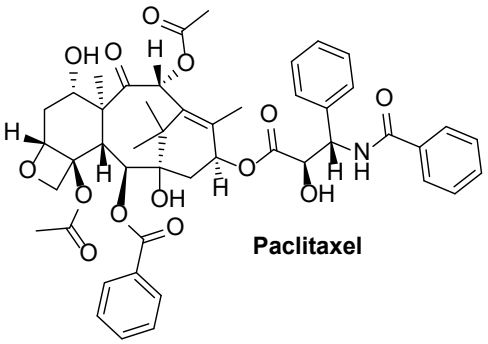
Chemotherapeutic drugs can be classified into five classes based on their mode of action (MoA),: alkylating agents, cytotoxic antibiotics, microtubule-targeting agents, topoisomerase inhibitors and antimetabolites.[77]
Alkylating agents represent the largest and a highly divers class of cytotoxic agents with prominent representatives such as cis-platin or cyclophosphamide. They usually generate a highly reactive intermediate which covalently binds to the nucleotide bases of the DNA.[78,79] This prevents any further transcription of the DNA strand, and therefore, protein biosynthesis is stopped. Alkylating agents deploy their functionality regardless of the cell cycle.
Cytotoxic antibiotics are another class of chemotherapeutics that intercalate to the DNA (for example, daunorubicin). Further transcription is prevented by cross-linking of the double strands or the generation of free radicals, which then cleave the double strands.[80-82] Dactinomycin, for example, binds to a certain DNA conformation within the transcriptional enzyme complex and inhibits the elongation of the RNA chain by the RNA polymerase.[83-85]
Microtubule targeting agents bind to the subunits of microtubules and either prevent the assembly of the tubulin monomers to microtubule polymers, or they prevent the disassembly of already polymerized microtubule structures.[86-89] In both cases, the mitosis cannot be successfully completed, and the cell eventually undergoes apoptosis (cell death). Vinca alkaloids are drugs, that inhibit the assembly of microtubules, with vinblastine as a prominent natural product representative.[90,91] In contrast, taxanes are a class of drugs that prevent disassembly, with paclitaxel as a very prominent example

of a natural product used in chemotherapy.[92,93] Although cells are most vulnerable during mitosis, this class of agents is the only one that interferes with the cell cycle in this phase. Topoisomerases are enzymes responsible for topological changes of DNA prior to replication and transcription. Inhibitors of topoisomerases such as topotecan are most effective in S and G2 phase where an uncoiling of the DNA is required.[94,95]

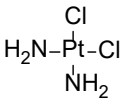
Antimetabolites such as mercaptopurine and fluorouracil often mimic the motifs of nucleotides and are incorporated into the DNA leading to breaking strands or inhibit required enzymes responsible for the synthesis of DNA building blocks.[96,97] Therefore, they terminate DNA duplication during S phase.



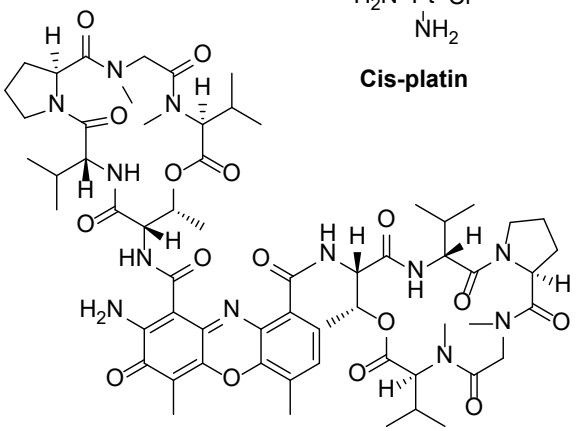
Daunorubicin



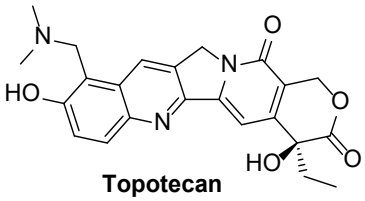
Paclitaxel



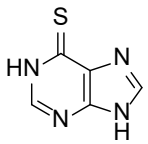
Cis-platin



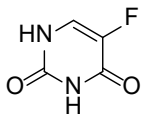
Dactinomycin



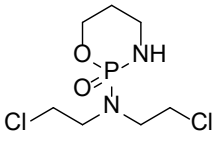
Topotecan



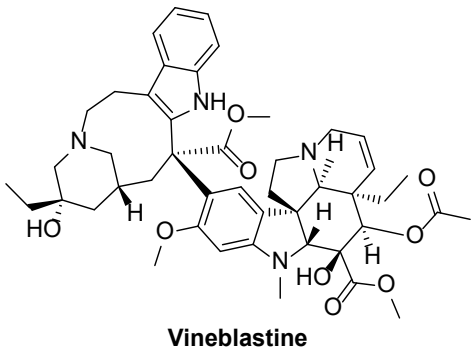
Mercaptopurine



Fluorouracil



Cyclophosphamide



Vineblastine

1.3.3 Molecular mechanism of mitosis

The final M (mitosis) phase consists of four heavily controlled and regulated subphases (Figure 5).

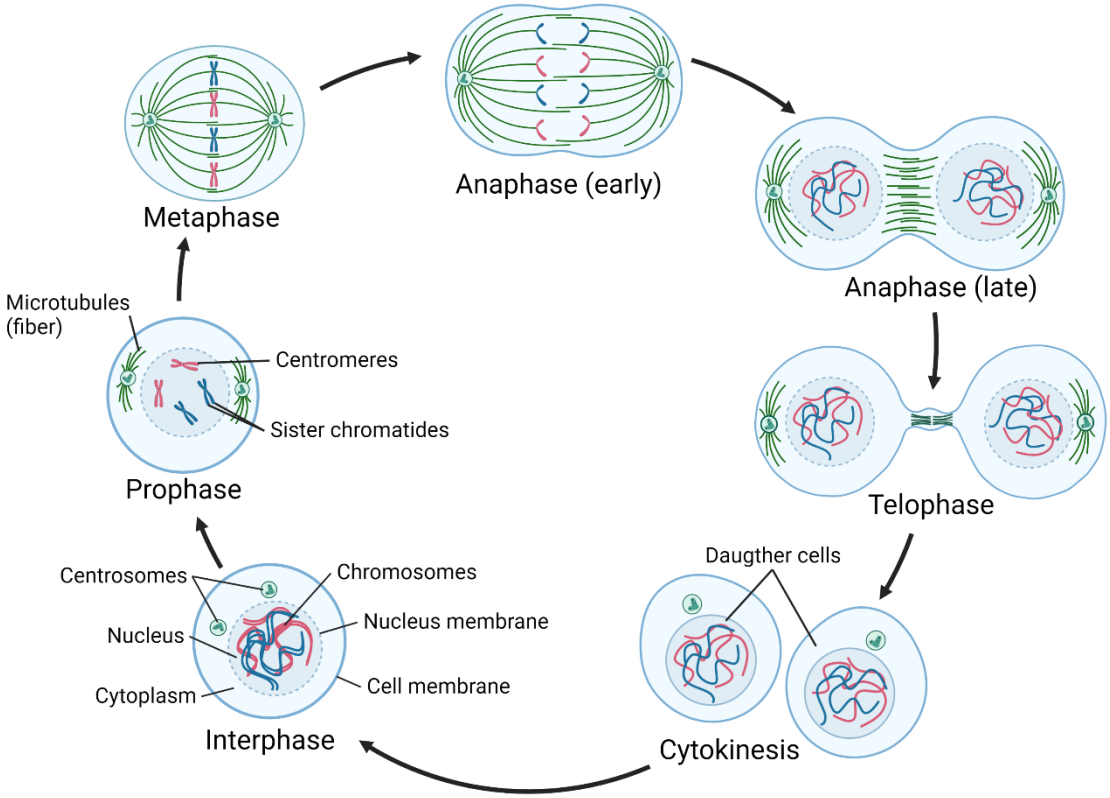


Figure 5: Phases of mitosis: chromosomes/DNA content in red and blue, microtubule spindles in green.[4]

Prophase:

During the initial prophase, chromosomes are condensed, and the formation of the spindle apparatus is initiated by microtubules extending from the centrosomes, which are located at opposite sides of the cell. Microtubules and another polymer called actin, are part of the cytoskeleton and responsible for a variety of properties e.g., cell structure, shape, or movement. They also serve as a platform for intracellular transport. At the end of the prophase, after the chromosomes paired up so that both identical copies are linked together by a protein called cohesin,[98] the nucleus membrane disintegrates and releases the chromosomes.

[4] Created with BioRender.com.

Metaphase:

In the following metaphase, microtubules projecting from the centrosomes, bind to the chromosomes at specific points which are called kinetochores and align them in a so-called equatorial plane.

Anaphase:

During the anaphase, the microtubules pull the chromatids of each chromosome towards the opposite end of the cell after the cohesion was cleaved by the separase enzyme, a thiol protease.[99] In addition, microtubules push against each other leading to an elongation of the cell.

Telophase:

In the last phase called telophase, a new nucleus membrane is formed around the separated chromosomes, before the cell undergoes cytokinesis, the progress where the cytoplasm is divided into two by an actin-myosin contractile ring which eventually separates the cells in two parts.[100]

The progression of the cell cycle is regulated by oscillating levels of two classes of proteins – cyclines and cyclin-dependent kinases (CDKs) identified by Hartwell, Nurse and Hunt who were awarded the Nobel Prize in Physiology or Medicine in 2001.[101,102] In addition, there are several checkpoints along the cell cycle to prevent the transfer of damaged genetic material to the daughter cells. The first major check point is also known as restriction point (G1/S checkpoint, see Figure 5) where the presence of sufficient educts (nucleotide bases, chromatin etc.) for DNA replication is checked. A second important checkpoint is called G2/M checkpoint where cytoplasm and phospholipid amounts are validated.[103] The third major checkpoint is called spindle assembly checkpoint where the alignment of the chromosomes and the formation of the spindle apparatus are checked.[104] The downregulation or complete inactivation of these control mechanisms results in genetic instability associated with cancer or birth defects such as Down's syndrome.[105]

1.3.4 Dynamics of the cytoskeleton

The cytoskeleton is an interconnected and heavily regulated dynamic network of protein filaments which can be found in the cytoplasm of all eukaryotic cells. In recent years it has been shown that also bacteria and fungi possess homologues of cytoskeletal structures.[106,107] The main components of these networks are classified as microfilaments, also referred to as actin filament, intermediate filaments, and microtubules.[108,109] All of them are capable of fast expending and shrinking, depending

on the cells current state and as a response to the cellular environment. In most cells, the actual growth rate of microtubules is about 0.4 µm/s, and the shrink is substantially faster.[110]

The cytoskeleton has several functions e.g., to maintain or adjust the cellular shape or to facilitate the cellular uptake of extracellular materials.[111] On a macroscopic scale, muscle contraction is an example of an action performed by the cytoskeleton.

Especially important for this work is that the cytoskeleton was found to be a key player during cell division (chapter 1.3.2 – Therapeutic concepts and molecular targets of cancer therapy). More specifically, microtubules are responsible for the segregation of the chromosomes in the M phase. This process requires high precision and coordination and is orchestrated by an interaction of mainly three proteins: the kinesins Eg5 and Kif15,[112] and the dynein (Figure 6).[113]

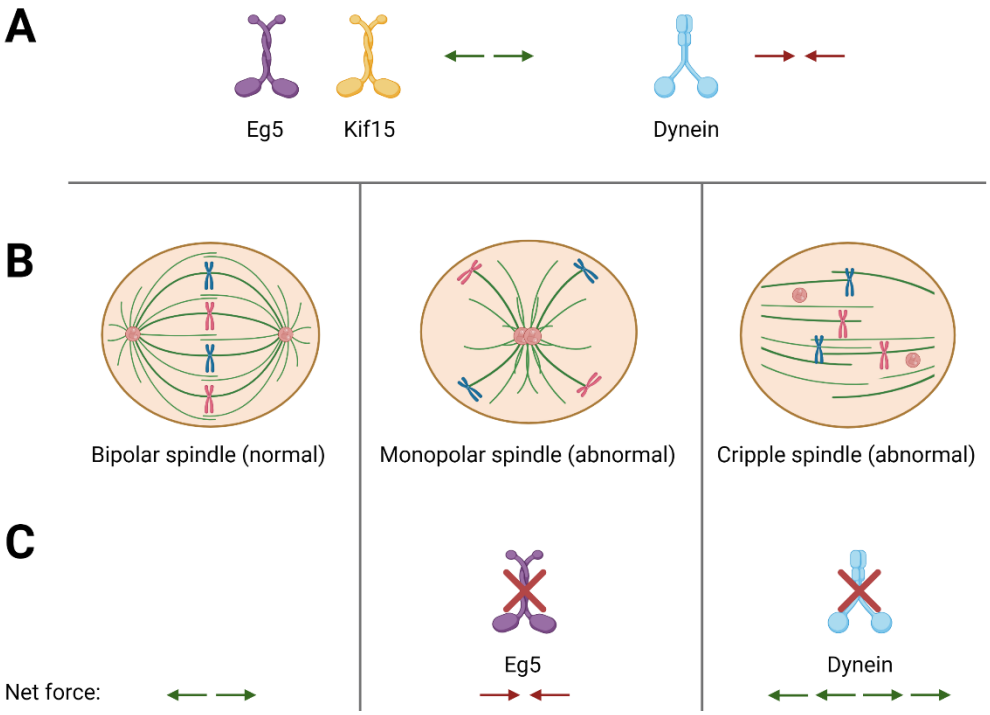


Figure 6: (A) Forces generated by the microtubule associated motor proteins Eg5 (purple) and Kif15 (yellow) and dynein (light blue). Eg5 and Kif15 generate an outwards direct force, while dynein is generating an inward directed force.
(B/C) Bipolar spindle assembly (left) requires a net outward directed force. Inhibition of Eg5 (center) leads to an inward directed force and inhibition of dynein (right) creates an excessive net outward force. Both events lead to the formation of abnormal spindles.[5]

[5] Created with BioRender.com.

All three of them belong to the class of motor proteins due to their ability to convert chemical energy obtained from the hydrolysis of ATP to mechanical work which is used to move along the microtubule polymers.[114] While they move, they can transport various cellular cargos such as vesicles from the endoplasmic reticulum or endosomes and lysosomes.[115,116] The motor proteins are classified depending on the direction of their movement. Kinesins such as Eg5 and Kif15 move anterograde (forwards) while motor proteins such as the dynein move retrograde (backwards) along the microtubules. This movement generates a tiny force. In case of the kinesins Eg5 and Kif15, the generated force is directed outwards while the movement of dynein generates an inwards directed force (Figure 6A). The generated net force is essential for the orientation and organization of the microtubules in the spindle apparatus. In order to obtain a bipolar spindle formation where the spindles poles are at the opposite sides of the cell and the chromosomes are oriented in an equatorial plane ready for successful cell division, the generated net force has to be directed outwards (Figure 6B, left).[117,118]

The inhibition of these motor proteins leads to the formation of abnormal spindles and the cell can no longer divide, ultimately leading to cell death. Mayer *et al.* discovered the formation of so-called monopolar spindles (Figure 6B center) induced by a compound which was named monastrol.[119] Monopolar spindles are characterized by radial microtubules directed from a center point containing the two centrosomes (spindle poles) and the chromosomes arranged in an outer sphere around the centrosomes and microtubules. Later, it was found that monastrol inhibits the Eg5 motor protein, leading to inwards directed net force and thus suppress the bipolar spindle formation.[120-122] Similarly, inhibiting the dynein by ciliobrevin D for example,[123,124] the generated net force is directed outwards, but the sum is larger than required for the bipolar spindle formation, leading to the formation of cripple spindles.[125] Within cripple spindles, the microtubules lose their pole focus and are no longer connected to the centrosomes.

1.4 Neurodegenerative diseases – an emerging global threat

Improved health care combined with discovery of novel treatments for infectious diseases and cancer as well as technological innovation has significantly increased life expectations. Due to the resulted demographic aging process the proportion of older people has changed and therefore, age-related diseases such as cancer, cardiovascular diseases, diabetes, and dementia are on the rise.

Dementia describes a class of symptoms caused by various diseases of the brain. They all result in a progressive loss of neuronal structure and functionality, which ultimately leads to neuronal cell death (referred to as neurodegeneration). Up to now, there is no known way to reverse the neuronal impairment. Thus, these diseases are considered to be incurable at present.

1.4.1 Prevalence, classification and significance

In 2020, over 50 million people worldwide had a diagnosed form of Dementia.[126] This number is estimated to double every 20 years, leading to approximately 131.5 million cases by 2050.[44] Commonly known diseases responsible for the development of Dementia are Multiples Sclerosis, Parkinson disease, Alzheimer disease and Huntington's disease.

Due to the large variety of symptoms and underlying causes for Dementia, different types of Dementia are not necessarily easy to distinguish. The most dominant cause is Alzheimer's disease (AD) with approximately 50 – 75 % of the cases, followed by vascular dementia due to reduced blood flow to the brain. (17 – 30 %).[44]

Dementia is currently the seventh leading cause of death among all diseases and one of the major causes of disability and dependency among older people worldwide. This accounts for the huge impact on the economy with annual costs of one trillion dollars in 2018, expected to rise to two trillion dollars by 2030.

1.4.2 Molecular mechanisms of AD development

In 1906, Dr. Alois Alzheimer found changes in the brain tissue of a woman who had died of an unusual mental illness. He was the first to discover the presence of many abnormal clumps and tangled bundles of fibers in specific areas of the brain, and presented his findings to a professional audience in Tübingen for the very first time. Sadly, the anticipated interest was never achieved.[127,128] The formation of these abnormal clumps and tangles was later found to be neurotoxic, leading to neurons losing their functionality

and connection to other neurons and eventually to neuronal death. Due to the complexity of the disease and the variation of symptoms and biochemical markers across patients, there is still no full scientific consensus about the stages of development and the molecular processes involved. Several hypotheses are discussed as hallmarks for AD, including the discovery of plaques in the brain by Alois Alzheimer.

Detailed studies found out that these plaques are mainly composed of a peptide called amyloid-beta (Aβ), which can be of varying size in a range of 37 to 49 amino acids. The most common chain length is 40 (Aβ1-40) or 42 amino acids (Aβ1-42), though it was found that Aβ1-42 possesses a much higher neurotoxicity.[129,130]
In healthy human neuronal cells (Figure 7A), a protein called amyloid-precursor-protein (APP) can be found in the outer cell membrane. It has several proposed functions such as regulation of synapse formation,[131] neural plasticity[132] or iron transport regulation[133]. It is cleaved by another membrane enzyme called α-secretase. This leads to the formation of a fragment called soluble amyloid precursor protein alpha (sAPPα). The role of sAPPα is subject of ongoing investigations but neuroprotective and regulatory effects of neuronal and memory mechanisms have been described.[134,135] The consecutive cleaving of the remaining sAPPα-protein by a multi-subunit complex called γ-secretase releases a small protein which is referred to as p3.[136] Once again, the role of p3 remains unclear, and even though it is capable of forming fibrillar aggregates, it does not necessarily exhibit neurotoxicity.[137,138]

In diseased neuronal cells (Figure 7B), the APP protein is cleaved by the β-secretase complex, a membrane protein often referred to as BACE, leading to a fragment called soluble amyloid precursor protein beta (sAPPβ). The remaining part of the original APP is then cleaved once by the γ-secretase, which releases Aβ. The individual domains of this complex have not been fully characterized yet, but one was identified as presenilin. Mutations in the encoding genes of presenilin were identified as major genetic risk factors for AD.[139]

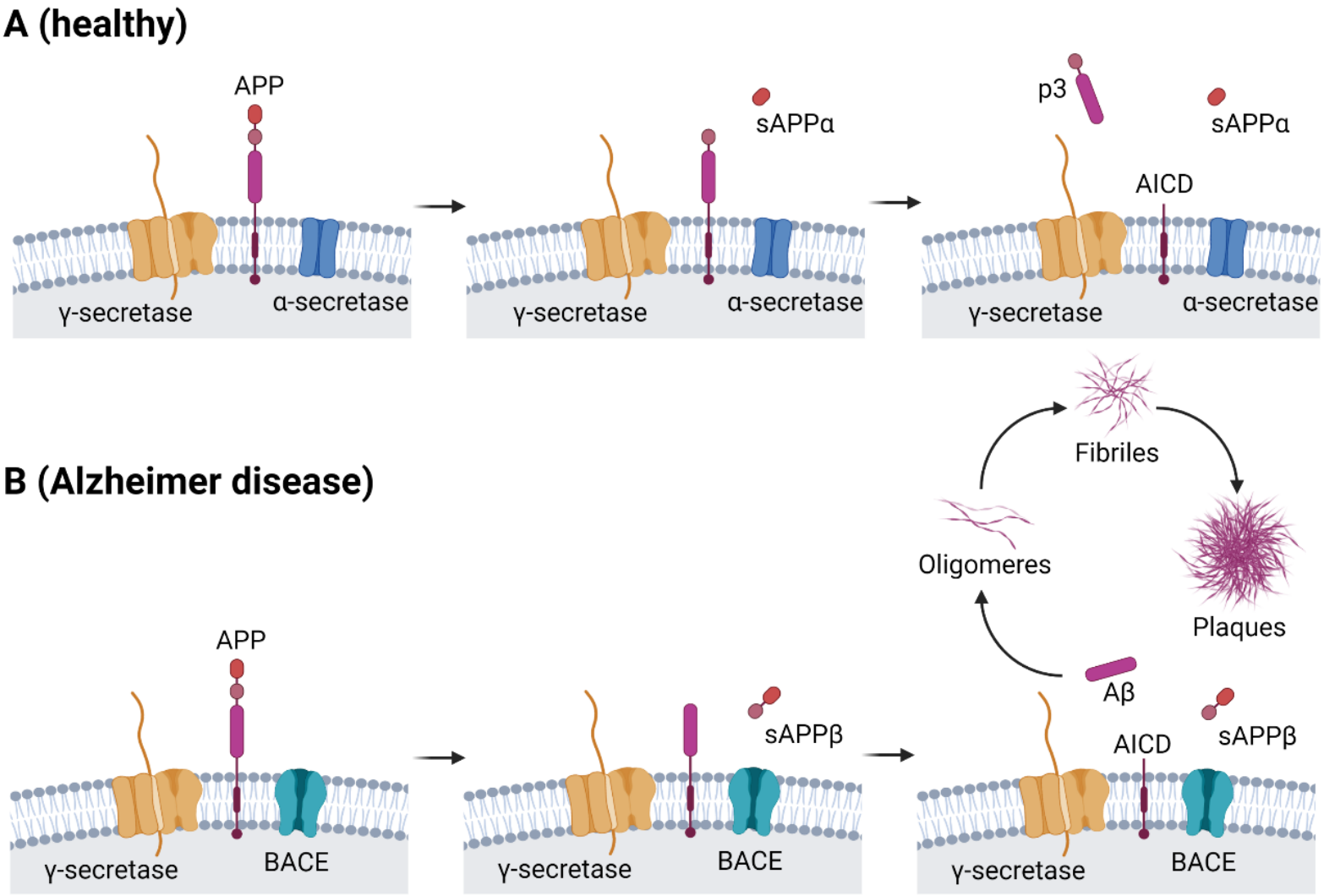


Figure 7: Putative mechanisms for the development of Alzheimer disease. (A) healthy state, where amyloid precursor protein (APP) is cleaved by an α-secretase forming a soluble peptide sAPPα. In a second step, the remaining protein is cleaved by the γ-secretase complex to release a protein called p3. (B) Alzheimer disease development state, where APP is cleaved first by BACE leading to a soluble peptide sAPPβ. The remaining protein is then cleaved by a γ-secretase complex leading to amyloid beta (Aβ). Aβ then aggregates and forms plaques. (Created with BioRender.com).

The released Aβ peptides have a high affinity towards aggregation and several aggregation states, e.g., oligomers and fibrils with different levels of neurotoxicity were observed.[130] The aggregation process ultimately leads to the formation and deposition of extracellular plaques, one of the hallmarks of AD. Interestingly, oligomers of Aβ were found to be most neurotoxic compared to the fibril and plaque structures.[140] A variety of receptors that interact with oligomeric Aβ were identified[141] e.g., a nerve growth receptor (NGF) which can trigger an apoptosis signal cascade,[142] or a glutamate receptor (NMDAR) leading to increased oxidative stress.[143,144] Moreover, binding of oligomeric Aβ to the Frizzled (Fz) receptor ultimately leads to an increased phosphorylation and deposition rate of the protein tau,[145] a second hallmark of AD.

This phenomenon is called tauopathy and related to the tau protein, a microtubule-associated protein which is especially abundant in neuronal cells. It is responsible for promoting the assembly and modulating the stability of the microtubules.[146] More recent studies discovered that tau is also responsible for the regulation of microtubule-regulated axonal transport (more details about transport in cells can be found in chapter 1.3.3 – Molecular mechanism of mitosis).[147] When tau is phosphorylated by kinases, the binding affinity towards microtubules significantly decreases, leading to the disintegration of the microtubules.[148] Normally, there is a precisely controlled equilibrium state between the kinase responsible for the phosphorylation of tau and the phosphatase enzyme, which is responsible for the cleavage of the phosphate groups.[149] The release of hyperphosphorylated tau leads to the formation of intracellular tau tangles due to its increased self-aggregating affinity compared to the non-phosphorylated tau.

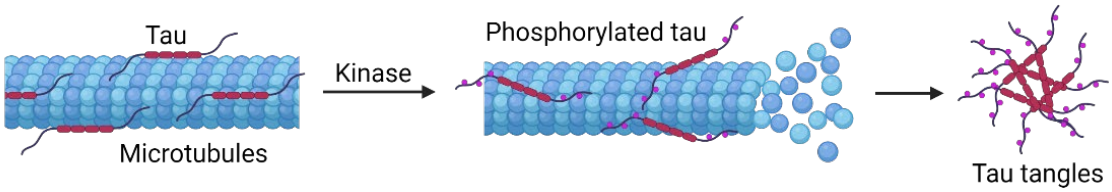
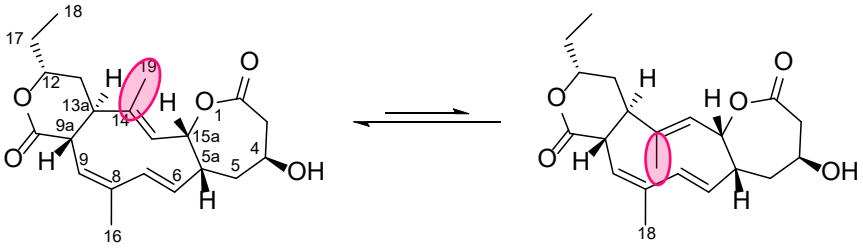


Figure 8: The microtubule associated protein tau is phosphorylated leading to a disintegration of the microtubule stability and releasing the phosphorylated tau which then forms tau tangles. Created with BioRender.com.

There are 121 unique therapies currently in clinical trials for Alzheimer's disease, 21 of which target either amyloid-β or tau protein.[150] Since 2003, only Aducanumab (2021)[151] was approved by the FDA and the series of candidates failing in clinical trials is constantly growing.[152] In addition, the number of repurposed agents in the pipeline (43 %) indicates that not only new therapeutic approaches are required, but also novel and highly potent drugs.[153]

1.5 Aim of the thesis

The strain *Streptomyces sp.* Gö 40/10 was isolated from a Bolivian soil sample in 1989. SCHIEWE studied the rich secondary metabolite profile of the strain and was first to isolate a natural product, which was named collinolactone (**1**).[154] The novel compound **1** consists of an unprecedented and unique 6-10-7-membered tricyclic system with a cyclodecatriene ring flanked by two lactone rings. A characteristic feature is the presence of two isomers in a ratio of 4 to 1, which cannot be separated by high-performance-liquid-chromatography (HPLC) and cause a duplication of signals in nuclear magnetic resonance (NMR) spectra. HOFFMANN could prove that the rotation of the methyl group at position C14 is responsible for the conversion between the rotamers **1a** and **1b** by advanced temperature modulated NMR experiments.[155]

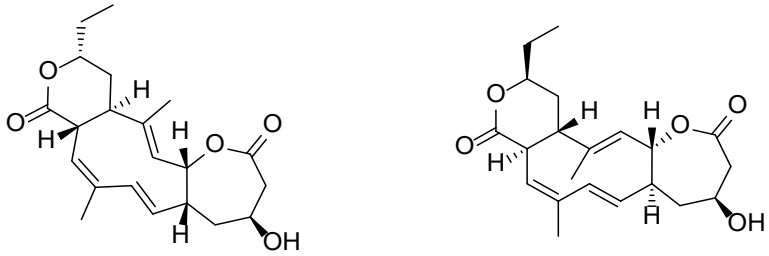


1a (anti) **1b** (syn)

The natural product collinolactone (**1**) itself does not show any activity in antibacterial, antifungal, or cytotoxicity assays. However, semi-synthetically generated esters of **1** showed a cytotoxic effect in the micromolar range against L929, PtK2 and MCF-7 cell lines by disruption of bipolar spindle formation during mitosis.[156] Since all these assays were performed by external cooperation partners on different compound sets and cell lines, only a highly fragmented overall picture of the biological potential of these derivatives was available at this time. One of the aims of the thesis is the chemical synthesis of a representative and diverse set of derivatives and examine their cytotoxic potential in a standardized and validated cell viability assay. The mechanism for cytotoxicity shall be studied in detail by fluorescence microscopy. Furthermore, an ATP-based assay with putative molecular targets should be established, validated and the compounds tested towards their inhibitory effects.

During this PhD research work, *Kwon et al.* described the discovery of a compound which was named rhizolutin (**1R**).[157] Detailed *in vivo* and *in vitro* studies indicate that rhizolutin dissociates misfolded and aggregated proteins, which is one of the hallmarks of Alzheimer Disease (AD). The spectral data reported for **1R** and **1** share significant similarities. Consequently, the same carbon skeleton was proposed for both compounds, but a different stereochemistry. Thus, one of the aims of this thesis is a detailed analysis

and review of the structure elucidation of rhizolutin done by *Kwon et al.* and the comparison to collinolactone (**1**).



Collinolactone 2006 (1) **Rhizolutin 2020 (1R)**

Figure 9: Proposed structure of rhizolutin by *Kwon et al.* in 2020[157] and of collinolactone (**1**).[155]

Semisynthetic derivatization and biological profiling, especially fluorescence microscopy, require a high amount of compound **1** as educt. In addition, planned mouse experiments to investigate the potential of collinolactone (**1**) as an anti-Alzheimer drug can easily consume the compound on the gram scale. Since no total synthesis route is yet established, one can only rely on biotechnological production. Therefore, the biosynthesis should be studied in more detail by feeding experiments and time resolved metabolomic profiling. Notably, the isolation and production procedure of collinolactone (**1**) shall be optimized based on the results. In collaboration with Tilmann Weber from the DTU Denmark, the gene cluster in *Streptomyces sp. Gö 40/10* responsible to produce collinolactone (**1**) shall be identified.

One of the major problems in natural product discovery is the isolation of already described substances. This increasingly occurring event costs money and manpower and can be especially frustrating for researchers. Therefore, the development and implementation of methods to identify already known substances early on in the natural product discovery workflow (a process called dereplication) is of great interest. Even though there are already several informatic tools available for dereplication, most of them are either not convenient for processing of full LC-MS analyses or require in depth knowledge of at least one programming language (such as R or python, which are most frequently used in the scientific community) to set up and perform such analyses. One aim of the thesis is the development and prototyping of a platform which combines and facilitates multiple approaches and tools for dereplication of natural products and can be used by inexperienced users in terms of programming and computer knowledge.

2. RESULTS AND DISCUSSION

2.1 Spectral comparison of collinolactone (**1**) and rhizolutin (**1R**)

In 2020, during the experimental work for this thesis, *Kwon et al.* described the discovery of a compound which was named rhizolutin (**1R**).[157] The spectral data reported for **1R** and **1** share the same spectral properties. For example, both compounds display two conformers in both, ^{1}H and ^{13}C-NMR spectra, caused by the freely rotating methyl group at C14. The ratio of the rotamers is dependent on the solvent and the temperature for both structures: with increasing temperature, the ratio changes in favor of the syn component.[2,157]
Consequently, the same carbon skeleton was proposed for both structures. However, there is a discrepancy of the stereochemistry (highlighted in Figure 10) proposed for rhizolutin (**1R**) and collinolactone (**1**) at positions 5a-H, 9a-H, 12-H and 13a-H.

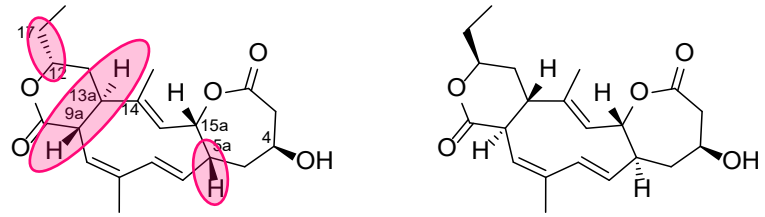


Figure 10: Differences of the stereochemistry of collinolactone (**1**) and rhizolutin (**1R**).

Extensive 2D-NOESY- and 2D-ROESY-NMR experiments were performed to investigate the nature of the ongoing conversion, and to state the relative configuration. In the ^{1}H-^{1}H-NOESY-spectrum of **1**, only one correlation was found which linked the left-hand part of the molecule to the right-hand part (see Figure 11, correlation between hydrogens 15a-H and 3-H, marked with an asterisk). No correlation was observed between hydrogens 4-H and 5a-H, indicating that these hydrogens are not in the same plane and in close proximity (NOESY shows correlations within less than 5 Å).[158] However, several correlations indicate that the hydrogen 5a-H is in the same plane as 15a-H and 3-H.

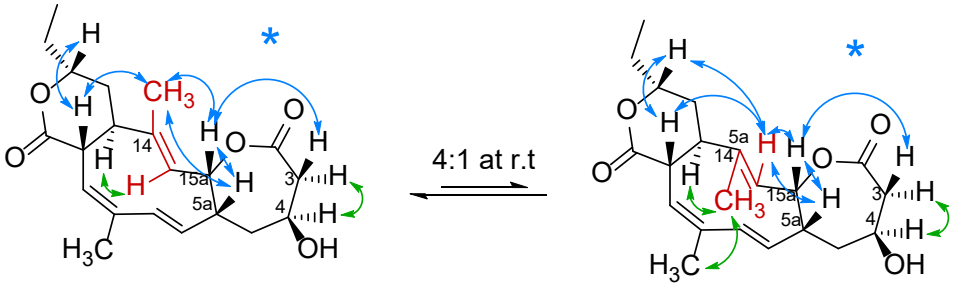


Figure 11: Selected NOE correlations of collinolactone (**1**): NOE above ring plane (green), NOEs below ring plane (blue). The rotating methyl group displayed in red is responsible for the undergoing interconversion leading to the rotamers (**1a** and **1b**).

The source data files (Mnova) of the experimental NMR-data of rhizolutin (**1R**) were kindly provided by Prof. Dr. Oh and coworkers from the Natural Products Research Institute, College of Pharmacy, Seoul National University. With these files in hand, a painstakingly detailed assignment and analysis was performed. One major point of criticism of the performed structure elucidation done by *Kwon et al.* is, that only [D_5]pyridine was used as a solvent for NMR-experiments.

However, a majority of collinolactone's (**1**) and rhizolutin's (**1R**) protons overlap with signals derived from other protons in a highly similar chemical environment. Due to the dynamic nature of the structures, the solvent has a huge impact on the chemical shift and can help to separate overlapping peaks in NMR spectra. More precisely, the signal of 15a-H (δ_H = 6.12 ppm) overlaps with the signal from 7-H and can therefore be misinterpreted when attempting to determine the configuration at 5a-H.

A correlation between 4-H and 5a-H (δ_H = 4.67 ppm and δ_H = 3.31 ppm) was shown in the summary image of ROESY correlations provided by Kwon *et al.* for rhizolutin (**1R**).[157] However, this correlation was not found in the superimposed spectra of **1** and **1R** (see Figure 12, red boxes). Instead, a strong correlation between H-5a and either H-7 or H-15a was found, indicating that these protons are in the same plane.

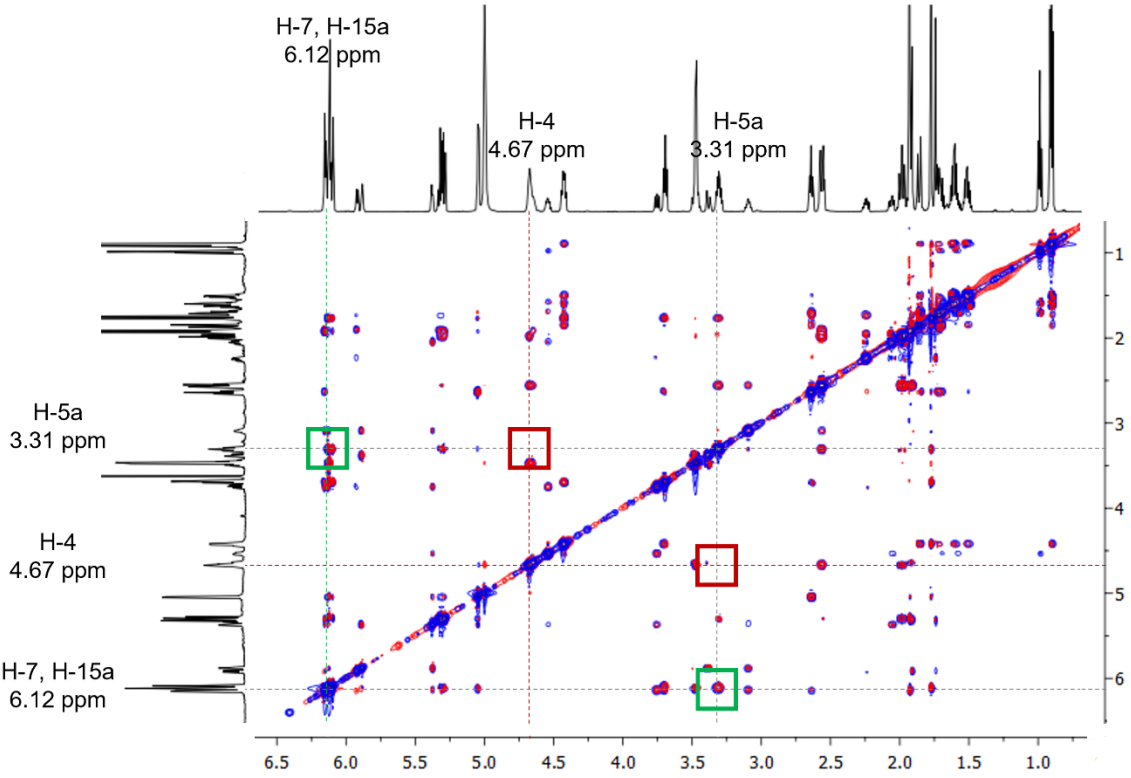


Figure 12: ^{1}H,^{1}H-ROESY spectra of collinolactone (1, red) and rhizolutin (1R, blue). Spectra were recorded in pyridine-d5 at 298 K. Red boxes show the missing correlation between H-4 and H-5a, green boxes show the existing correlation between H-5a and H-7 or H-15a, which are overlapping and cannot be distinguished in pyridine-d5.

In [D_6]benzene, the hydrogens at 15a-H, H-5a, and 7-H were well-separated and not overlapping. A strong correlation between 15a-H and 5a-H was found, proving that these two protons are indeed located in the same plane. This finding is further supported by the fact that no correlation is present between 4-H and 5a-H, and that these hydrogens are actually not in the same plane.

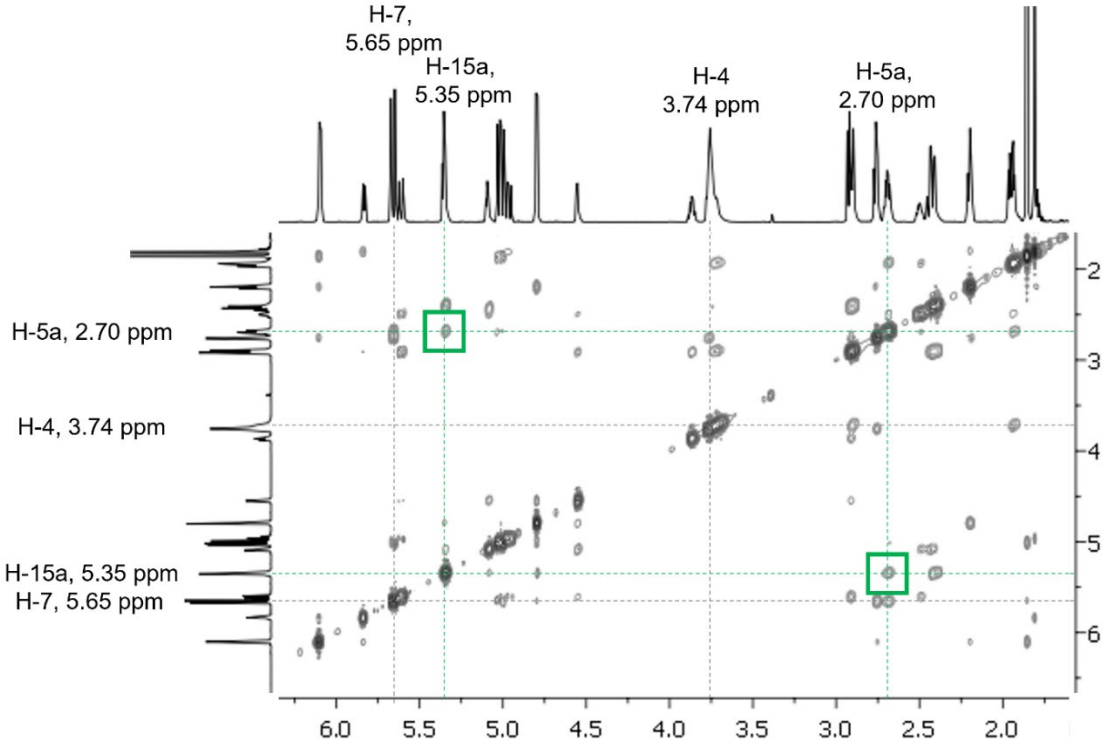


Figure 13: ^{1}H,^{1}H-NOESY spectrum in [D_6]benzene for collinolactone (**1**); green boxes show the correlation between H-15a and H-5a, indicating that these protons are in the same plane. No correlation was observed between H-4 and H-15a.

2.2 Biosynthesis of collinolactone (**1**)

The polyketide synthase (PKS) enzyme complex is responsible for the synthesis of a diverse class of natural products found in fungi and bacteria as well as in plants and animals.[159] Within the complex, the polyketide skeleton is synthesized in a sequence of Claisen condensations. In contrast to the fatty acid synthesis, where each building block is reduced, the building blocks for PKS are not always complete reduced between each elongation step and can therefore possess different oxidation states.[160] Typical starter units are acetyl-CoA or propionyl-CoA while either malonyl-CoA or methylmalonyl-CoA serve as an extender unit.
Depending on the flexibility of the synthesis domains, a variety of unusual and artificial building blocks can be used to modify the resulting products, e.g., to insert additional functional groups or to tune its molecular properties.[161] This can either be achieved by genetic engineering[162,163] or by a so-called precursor-directed biosynthesis approach[164,165], where a large excess of the building block is fed during cultivation.

From the structural skeleton of **1**, it was assumed that it is derived from the polyketide biosynthetic pathway. Therefore, the common building blocks acetate and propionate were fed as ^{13}C-labeled precursors during fermentation.[155] After extraction and purification, the analysis of the recorded ^{13}C-NMR spectra revealed that three propionates via methylmalonyl-CoA and six acetate via malonyl-CoA were incorporated.[2] The insertion of molecular oxygen at position C-2 and C-15a was confirmed by ^{13}C-NMR and the observed oxygen-18 isotope NMR-shift upon ^{13}C carbon after cultivation under $^{18}O_2$-atmosphere.[166]
Based on these results, a biosynthesis route was postulated: firstly, a nonaketide is synthesized via PKS type I pathway with six acetate and three propionate units. When the ketide is cleaved from the CoA-Enzyme, it undergoes cyclization, forming an 18-membered lactone ring. Next, a cyclic rearrangement – presumably a [6+4] cyclic rearrangement or a [4+2] cyclization, occurs. A [3,3]-Cope rearrangement, which has been studied more recently in detail,[167] leads to the biosynthetic intermediate collinoketone (**2**). This intermediate can be isolated in small yields (0.3 mg/L, see chapter 5.6.1 – Production of collinolactone (**1**) for more information). In the last step, molecular oxygen is incorporated via Baeyer-Villiger like oxidation (BVO), leading to the final product collinolactone (**1**). For all biosynthetic intermediates, which were not isolated so far, the E/Z-configuration of the double bonds was estimated based on the final configuration in the natural product **1**.

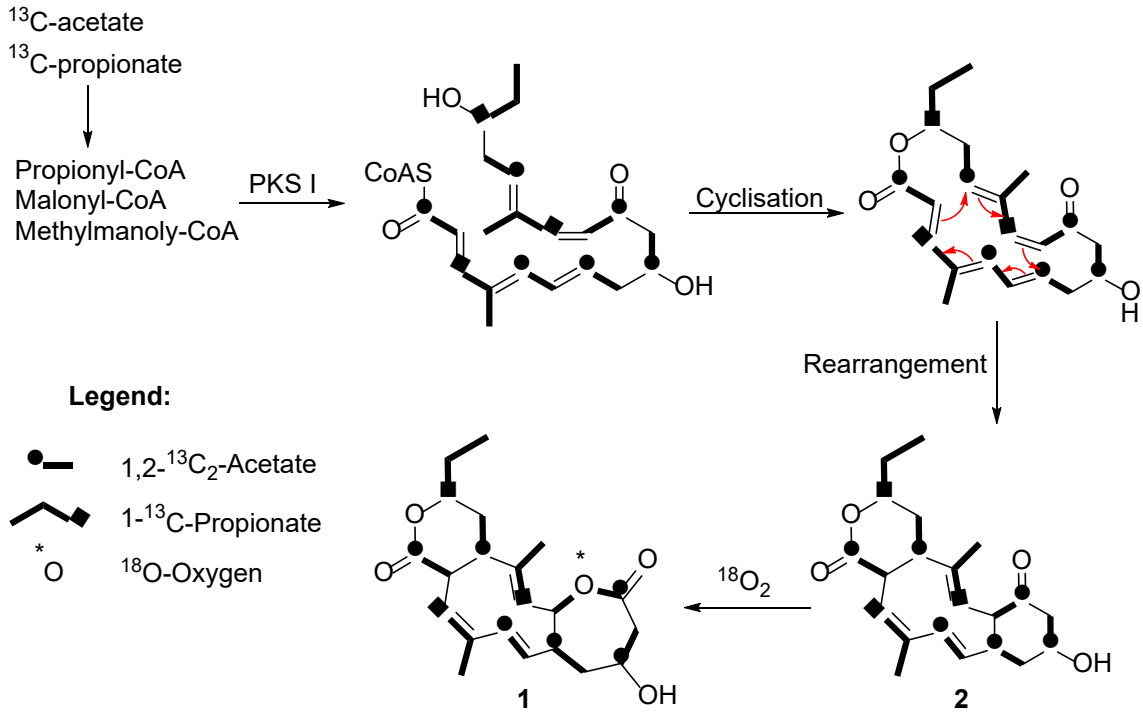


Figure 14: Proposed biosynthesis of **1** via type I PKS pathway based on the results from the isotope labelled precursor feeding experiments.
Adapted and modified from J. C. Schmid *et al.*, *Angew. Chem. Int. Ed.* **2021**, *60*, 23212.

To exclude the presence of methyltransferases, the uniformly labeled 1,2,3-$^{13}C_3$-propionate (sodium salt) was also fed. All 9 labeled carbon atoms were identified by NMR in the skeleton of **1** proving that no methyl transfer has occurred.
The origin of the hydroxyl-group at position C-4 was examined by feeding ^{18}O-labeled propionate. Due to the possible exchange of the labeled hydroxyl-group with water in the culture broth and the presence of a large excess of non-labeled propionate in the media, the incorporation rate was not sufficient to be identified by the reported oxygen-18 isotope NMR-shift.[166] By extending a recently published approach to link biosynthetic gene clusters (BGCs) with mass spectrometry data,[168] the incorporation was examined by HPLC-MS. Even with the use of high-resolution instruments such as the ESI-TOF Bruker MaXis 4G, the separation of the $^{13}C_2$-isotope signal from the ^{18}O-signal is still not satisfactory. Therefore, a sample with only natural occurring ^{13}C sources is required as reference. Also, a chromatographic baseline separation, e.g., by HPLC is required for this approach.
While the compound peak elutes, the intensity of the isotopic signal of interest is monitored while it is monotonically increasing. After normalization and linear fitting, the slope derived from a reference sample with only natural occurring ^{13}C-isotopes and the slope of a sample with the labeled precursor can be compared. The incorporation was successful when the slopes are significantly different. It is important to ensure that the

detector saturation is not reached, otherwise the isotope intensities of the samples cannot be compared.

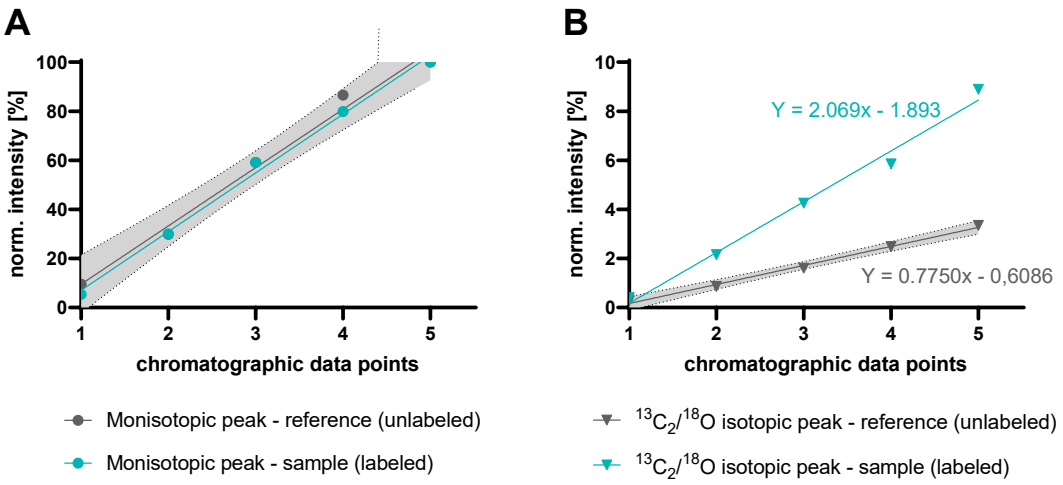


Figure 15: Analysis of the isotopic peak intensity slopes. The reference is displayed in grey, the sample with the enriched isotope is displayed in turquoise. The 95 % confidence interval areas for the references are displayed in grey. (A) slope of the monoisotopic peak of collinolactone (**1**), (B) slope of the $^{13}C_2/^{18}O$ isotopic peak.

The slopes of an unlabeled reference sample (in gray) and the labeled sample with the labeled precursor (in turquoise) were determined by tracing intensity values of selected isotopic peaks during peak elution (Figure 15A). The monoisotopic peak[6] of **1** (m/z = 361.2010) can be used to validate the measurement and the linearity of the method due to the fact, that the feeding of labeled precursors has no effect on the monoisotopic peak intensities. As expected, the monoisotopic peak of the labeled sample is within the 95 % confidence interval of the unlabeled monoisotopic reference peak and is not significantly different. In contrast, the $^{13}C_2/^{18}O$ isotope peak of the labeled sample is outside the 95 % confidence interval of the unlabeled reference peak, and the slope is approximately 2.7 times lower. This indicates that the ^{18}O-labeled propionate is indeed incorporated into compound **1** and that the hydroxyl group is originated from the propionate unit.

[6] The monoisotopic peak is defined as the sum of masses from the most abundant naturally occurring stable isotope of each atom in the compound.

2.3 Genome sequencing and BGC identification of collinolactone (**1**)

The genome sequencing and genetic engineering has been done by Dr. Jaime Felipe Guerrero Garzón at the DTU in Denmark supervised by Prof. Dr. Tilmann Weber.[2,169]

The Baeyer−Villiger oxidation (BVO) is generally carried out on laboratory scale using a peroxide source such as peroxybenzoic acid that oxidizes ketones (linear and cyclic ones) into esters or lactones.[170,171] The corresponding molecular enzymes that facilitates this reaction in living organisms are called Baeyer-Villiger-monooxygenases (BVMOs) and were already discovered in the late 1960s.[172] Up to now, more than a hundred representatives from prokaryotes and eukaryotes have been identified,[173] some of them playing a major role in the synthesis of toxins[174,175] and antibiotics.[176] BVMOs are classified into two main groups based on their cofactor specificity: type I uses flavin adenine dinucleotide (FAD) and nicotinamide adenine dinucleotide phosphate (NADP/H) while type II BVMOs bind flavin mononucleotide (FMN) and use nicotinamide adenine dinucleotide (NAD/H). Compared to type I, type II BVMOs are relatively rare and have mainly been described in *Pseudomonas putida* ATCC 17453.[177] Cytochrome P450 monooxygenases employ heme cofactors instead of flavins and are therefore not directly related to BVMOs, but it has been shown that some are also capable of catalyzing Baeyer–Villiger reactions.[178,179]

The first approach to identify the collinolactone (**1**) encoding gene cluster was to use the fact of molecular oxygen incorporation via a Baeyer-Villiger-like oxygenase (Figure 14) as a starting point for the search. VOLLMAR has identified the MLAGEL motif as a putative BVO encoding conserved fingerprint sequence in multiple streptomyces strains, including the collinolactone producing *Streptomyces sp.* Gö 40/10.[180] The amino acid sequence of all BGCs in the producer strain was then searched for the identified MLAGEL motif. This approach did not lead to any valid gene cluster candidates.

Two strains have been described in the literature that produce collinolactone (**1**): *Streptomyces sp.* Gö 40/10[181], mainly used for the experiments in this thesis and the recently described strain *Streptomyces* sp. WON17.[157] The strain *Streptomyces achromogenes* also produces compound **1**, and was kindly provided by AnalytiCon Discovery.

In a second approach, the gene clusters found in *Streptomyces sp.* Gö 40/10 and *Streptomyces achromogenes* were compared. Since both strains produce collinolactone (**1**), there must be at least one similar gene cluster which both strains have in common.

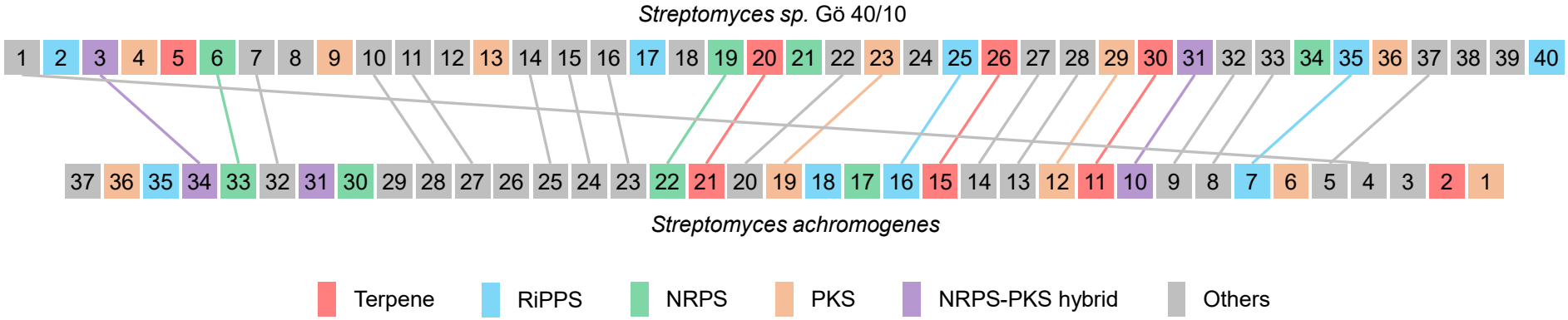


Figure 16: Gene cluster similarity matching of *Streptomyces sp.* Gö 40/10 and *Streptomyces achromogenes*. Clusters are color coded depending on the predicted compound class: terpenes are displayed in red, ribosomal synthesized and post-translationally modified peptides (RiPPS) in light blue, Nonribosomal peptide synthesis (NRPS) in green, PKS (all types) in orange, hybrid clusters of NRPS and PKS in purple and all other types in grey.

The genome sequences of both strains were analyzed using antiSMASH 5.1.2.[22] For *Streptomyces sp.* Gö 40/10, at least 40 BGCs were identified supporting the described biosynthetic potential of the strain.[154,181,182] Only four compound classes (ansatrienins, chineromycins, napthomycins and γ-butyrolactones, see Figure 40) were identified as secondary metabolites produced by the strain so far, indicating that there have to be at least some yet uncovered novel natural products.

For the second producer strain *Streptomyces achromogenes,* at least 37 BGCs were identified. The gene cluster similarity clustering between these two strains was performed using BiG-SCAPE version 1.1.2 with default parameters.[183] 14 BGCs were found to be unique for *Streptomyces sp.* Gö 40/10, compared to 13 unique BGCs for *Streptomyces achromogenes,* leading to 24 BGCs which are present in both strains. Based on the antiSMASH annotation of the BGC types, only cluster 23 (*Streptomyces sp.* Gö 40/10) and cluster 19 (*Streptomyces achromogenes)* respectively, were classified as type I PKS cluster (T1PKS, see Figure 16). This is most likely the wanted collinolactone encoding gene cluster.

Based on the feeding experiments, four additional PKS-like gene clusters were identified as candidates (clusters 1, 3, 31 and 39) in *Streptomyces sp.* Gö 40/10 and deletion mutants of all five candidates were generated using either CRISPR-BEST[184] technology in case of clusters 1, 3, 23 and 31 or suicide vectors in case of clusters 3 and 39.[2] All mutant strains were cultivated under standard conditions in soy flour mannitol media (see chapter 5.4 – Strains and culture conditions), extracted and analyzed by LC-ESI-MS. The wild type strain served as a control and was cultivated and extracted simultaneously under identical conditions.

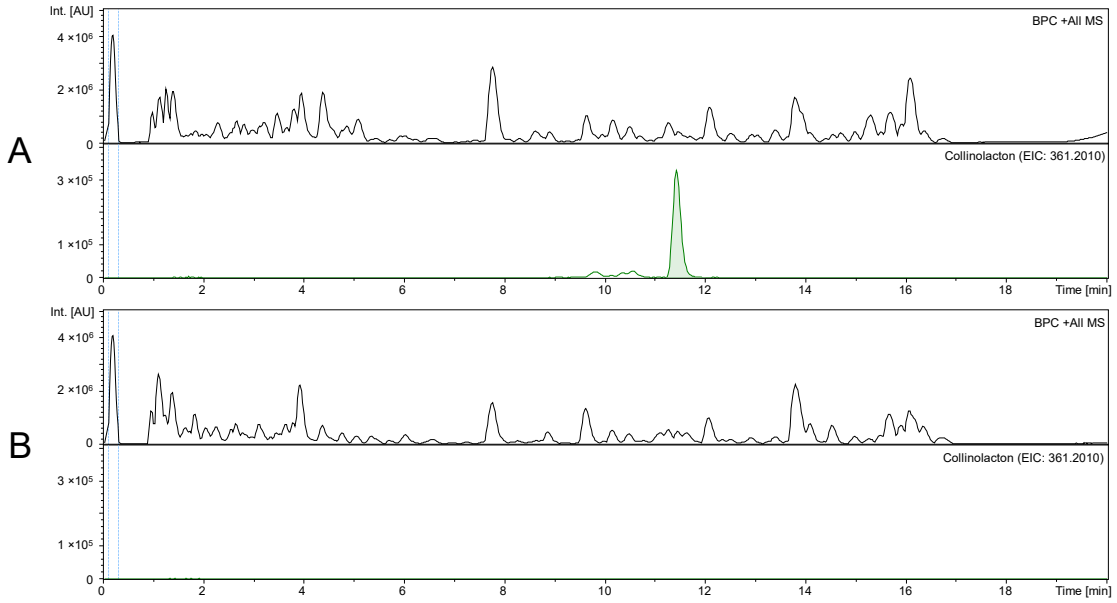


Figure 17: Displayed in black is the base peak chromatogram (BPC), in green the extracted ion chromatogram (EIC) of **1** (361.2010 ± 0.005 Da). Both chromatograms are scaled equally. (A) LC-MS chromatogram of the extracted supernatant from wild type strain *Streptomyces sp.* Gö 40/10, B: LC-MS chromatogram of extracted supernatant from the mutant strain *Streptomyces sp.* Gö 40/10 – *col* (cluster 23).

As already hypothesized from the gene cluster similarity analysis, the collinolactone (**1**) encoding BGC is cluster 23, named *col*. The corresponding mutant strain was the only one where a complete loss of collinolactone (**1**) production was observed. In addition, the BGC responsible for the already described secondary metabolite class of napthomycins[185] is encoded in cluster 31 (named *nat,* see Figure 41) which complies to 71 % with the already described cluster found in *Streptomyces sp.* CS according to antiSMASH.[186] The class of napthomycins currently includes a total of 14 derivatives, with antibacterial, antifungal and antitumor activities[187]

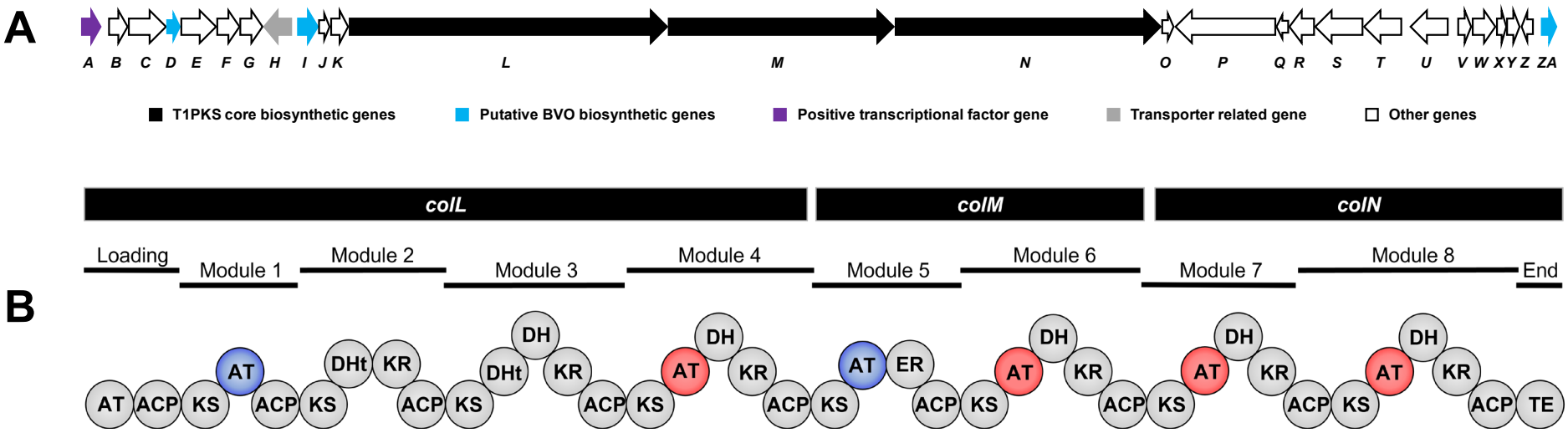


Figure 18: (A) Organization of genes present in the identified gene cluster *col* found in *Streptomyces sp.* Gö 40/10. Positive transcriptional factor genes are displayed in purple, the putative BVO genes in light blue and the core biosynthetic genes in black. All other genes are colorless.
B: Domain organization of the core biosynthetic genes, using the following abbreviations: AT: Acyltransferase domain, ACP: Acyl-carrier protein domain, KS: Keto-synthase domain, DH: Dehydratase domain, ER: Enoylreductase domain, TE: Thioesterase domain. Blue colored AT domains indicate a predicted specificity for methylmalonyl-CoA and red colored domains for malonyl-CoA extender units.
Adapted and modified from J. C. Schmid *et al.*, *Angew. Chem. Int. Ed.* **2021**, *60*, 23212.

The identified gene cluster *col* shown in Figure 18A has a length of 77 kb and consists of 27 genes. A full list of the predicted gene functions is provided in the appendix (Table **7**). Three PKS type I core biosynthesis genes (*colL*, *colM* and *colN*) have been identified in this cluster, with their modules presented in Figure 18B. The module pattern and the predicted acyltransferase domain specificities are not in accordance with the proposed biosynthesis and the chemical structure of **1** (Figure 14). That is mainly the reason why the initial search for the corresponding BGC of compound **1** only based on the sequencing results of *Streptomyces sp.* Gö 40/10, turned out to be very challenging. One must now assume that the cluster *col* is highly non-colinear and some modules are inactive or used multiple times to synthesize the core skeleton of collinolactone (**1**).

No genes, such as the already discussed MLAGEL or other described motifs[188,189] were found, that encode the putative BVO enzyme in this cluster. Additionally, performed feeding experiments with ancymidole, a cytochrome P450 inhibitor,[190] did not lead to a loss in production of compound **1** or an increase in the biosynthetic intermediate collinoketone (**2**).[191] Consequently, the utilization of a cytochrome P450 monooxygenase can be excluded.

More recently, the combination of a luciferase-like monooxygenase and a flavin-reductase pair (AbmE2 and AbmZ) was found to catalyze the rare type II BVO reaction in the neoabyssomicin biosynthesis in *Streptomyces koyangensis* SCSIO 5802.[192] BLAST analysis[24] was used to align the sequences and determine similarity. Similar genes were found in the identified gene cluster *col* (Table 1). Even though the matching was performed using sequences derived from the same genus and are therefore expected to be relatively high, the obtained scores are far from a confident match. But one can surely say that the corresponding genes share at least some motifs.

Table 1: BLAST sequence alignment and similarity scores for selected genes, presumably involved in the BVO type II reaction from *Streptomyces sp.* Gö 40/10 and *Streptomyces koyangensis* SCSIO 5802.[193]

Streptomyces sp. Gö 40/10	Similarity	*Streptomyces koyangensis* SCSIO 5802
colI	34.7	*AbmE2*
colZA	55.1	
colD	63.5	*AbmZ*

Heterologous expression is the ultimate proof that the putative gene cluster encodes a certain natural product. In case of the gene cluster *col*, the heterologous expression using the Cas9-Assisted Targeting of CHromosome segments (acronym CATCH) protocol[194,195] was not successful. To provide further proof, the luxR-type positive transcriptional factor encoded in the gene *colA* (Figure 18 and Table 7) was overexpressed under the control of the strong promotor kasOP*.[196] This led to a 5-fold increase in the production of collinolactone (**1**),[2] which is discussed in more detail in chapter 2.4 – Isolation and purification.

To study the biosynthesis and the involved genes in more detail, stop codons were introduced into genes *colD, colE, colI, colO, colP* and *colZA*. Only the introduction of a stop codon in *colE*, a putative anthranilate 3-monooxygenase was successful, yielding growing colonies. No colonies were obtained for the other five modified strains.
When cultivating the mutant strain with the silenced *colE* gene, the production of collinolactone (**1**) was not affected and no increase of biosynthetic intermediates or a delayed degradation was observed. Consequently, the putative anthranilate 3-monooxygenase encoded in gene *colE* is not required for the biosynthesis of **1**.
When introducing stop codons in the mutant strain *Streptomyces sp.* Gö 40/10 – *col*, which has lost its ability to produce **1**, the introduction of stop codons was successful also for the genes *colD, colI, colO, colP* and *colZA*. These findings indicate that the method used for introducing the stop codons is working as expected. The intolerance of the collinolactone-producing wildtype-strain towards these genetic modifications is presumably due to the formation of toxic intermediates, leading to the strain's death.

2.4 Isolation and purification

When collinolactone (**1**) was first isolated by KIND in 1991 and a first structure was proposed by SCHIEWE, the overall (unstable) isolation yield was only 1.5 mg/L.[154,197] The addition of adsorber resins such as Amberlite XAD resin has been shown to improve the production rate of secondary metabolites in a variety of organisms[198,199], especially when the produced metabolite is also toxic for the producer strain.[200] By adding Amberlite XAD-16 resin during cultivation in an airlift fermenter, HOFFMANN was able to boost the production to a constant level of 11 mg/L.[155] Airlift fermenter require trained technical laboratory employees and constant maintenance and not all strains are suitable for fermentation in airlift fermenters as it is the case for *Streptomyces sp.* Gö 40/10.
For chemical semi-synthesis, biological profiling and planned animal experiments in the future, the required amount of collinolactone (**1**) is very high. For example, an experiment with 15 rats, each with a body weight of 300 g and a daily application dose of 8 mg/kg body weight over a course of 3 weeks, devours over 750 mg of collinolactone (**1**). Therefore, a higher overall production rate is urgently required. It is also important to optimize the procedure for the isolation of collinolactone (**1**), especially in terms of required working time and consumables.

A detailed description of each purification step (Figure 19) is given in chapter 5.4 – Strains and culture conditions. Shaking flasks were found to result in a more stable production rather than using a classical fermenter, and an airlift fermenter was not available during the thesis. It turned out to be most efficient to scale the production towards the use of 30 shaking flasks with a total volume of 6 L, which could be simultaneously autoclaved and incubated. After fermentation was stopped, the supernatant and the mycelia were separated by centrifugation instead of using large funnels, which often clogged after a short time. The added XAD-16 resin binds the produced collinolactone (**1**) almost quantitatively, and only minor traces of compound **1** were found in the supernatant. Consequently, it was discarded after testing.

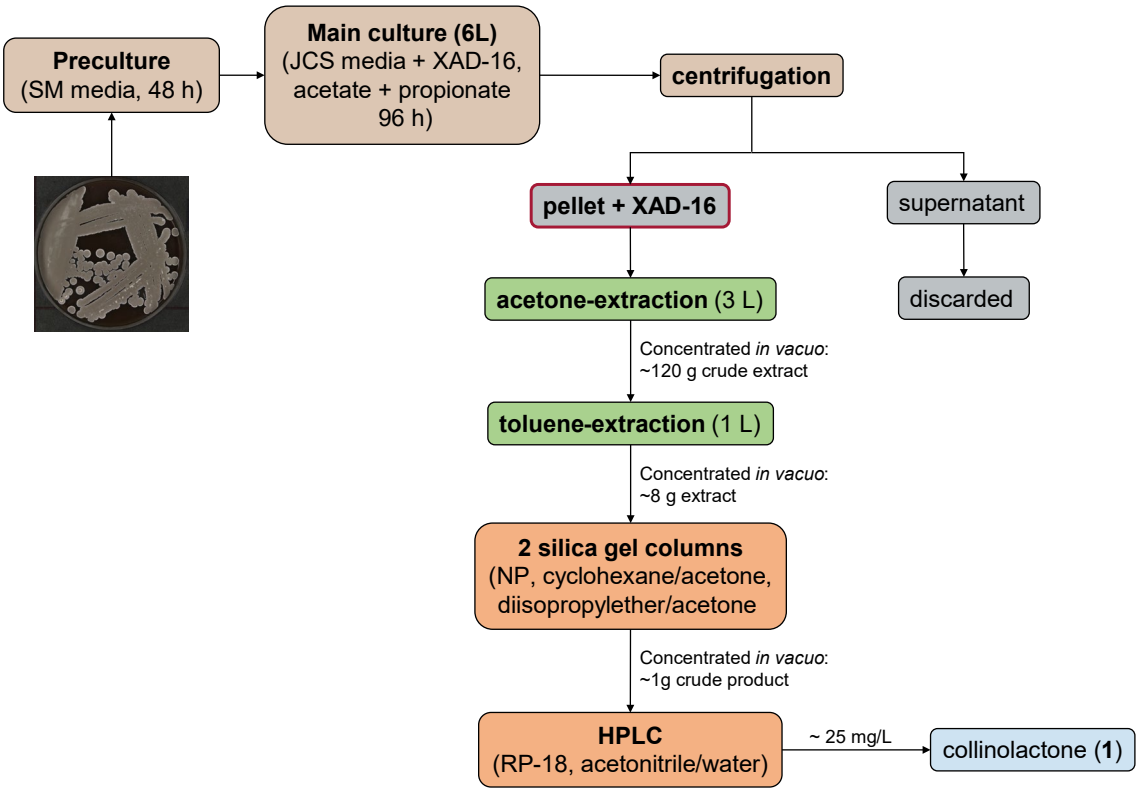


Figure 19: Purification strategy for collinolactone (**1**) purification including several extraction steps (green), followed by chromatographic purification steps (orange).

The resulting XAD-16/mycelia residue was extracted rigorously with acetone while stirring in a sonicator bath. In an earlier procedure, a mixture of methanol and acetone was used for the extraction of secondary metabolites. In case of collinolactone (**1**), methanolysis by ring opening of the 7-membered lactone ring was observed when sonication is performed in the presence of methanol. Consequently, the use of methanol was avoided in all steps during purification. After extraction and evaporation of the acetone, the remaining tar-like residue still contained lots of water-soluble components. Column chromatography with such high amounts of extract require unhandy and large columns with lots of silica gel and a high solvent consumption. It was found to be much more effective to include a second extraction step with toluene against water. Almost all water-soluble components were removed in this step and the overall weight was reduced to about 8 g, which was a suitable amount for column chromatography. After three more chromatographic steps including medium-pressure liquid chromatography (MPLC) and high-performance liquid chromatography (HPLC), the pure compound collinolactone (**1**) was obtained in overall stable yields of 15-25 mg/L.

2.5 Production optimization of collinolactone (**1**)

Based on the findings of a media screening for optimized collinolactone (**1**) production, the JCS media was developed.[191] It contains the same ingredients as the soy flour mannitol media (SM), with additional glucose, starch, yeast extract and calcium carbonate. Furthermore, the addition of the biosynthetic precursors propionate and malonate led to a stable production of about 25 mg/L. The procedure was adopted without changes for the genetically engineered overproducer *Streptomyces sp.* Gö 40/10 KasOP* LuxR, which will be referred to as *Streptomyces sp.* Gö 40/10^{HP} in the following. The large-scale fermentations with the overproducer strain revealed several problems, indicating that both strains do not behave identical, as initially assumed. Therefore, a systematic investigation and comparison of collinolactone (**1**) production rates under different conditions was necessary. The results are presented in the following part of this chapter.

For the systematic investigation, both strains were cultivated simultaneously under the same conditions in two different media: the standard soy flower mannitol media (SM) and the optimized JCS production media. In addition, the influence of the absorber resin XAD-16 was investigated. All experiments were performed in triplicate.
To make the results more comparable, both strains were initially grown on agar plates and spore suspensions were prepared following a well-established protocol.[201] Subsequently, a colony forming unit test was performed for the prepared spore suspensions. The same number of spores was then used to inoculate the flasks, and samples of 1 mL volume were taken every 24 hours.
The strain *Streptomyces sp.* Gö 40/10^{HP} was engineered using a hygromycin resistance gene, therefore the addition of hygromycin (50 µg/mL) is required for fermentation. To ensure that the overproducing capability was not lost during fermentation, the flasks were used to inoculate agar plates with molded hygromycin. All flasks with Gö 40/10^{HP} maintained their overproduction capabilities, as indicated by growth across all plates.

A direct injection of the supernatant to LC-MS is not feasible, since the addition of the adsorber resin requires an extraction step. All samples were freeze-dried and then extracted for 24 hours with the same amount of acetonitrile/water (9:1). After centrifugation, the supernatants were separated and analyzed by LC-MS. The data were automatically processed in Bruker DataAnalysis using a custom developed method script in the course of this thesis, which identified, integrated, and exported the detected compounds.

The compounds were then imported into MATLAB and further analyzed. Since the signal of the monoisotopic mass of collinolactone (**1**) and other metabolites of interest in these samples might lead to reach the detector saturation limit, the ^{13}C-isotopic signal masses were used to generate the EIC traces.

The best and most reliable way for quantification is the use of a stable isotope labeled (SIL) internal standard.[202] This kind of standard requires an isotopic enrichment of at least 95 %. Besides the fact that no total synthesis for collinolactone (**1**) is currently available, such a standard would be still very expensive and is not an option. Another way is the standard addition method by spiking multiple concentrations of **1** as non-labeled internal standard,[203] but this requires that each sample is split into at least three aliquots: the non-spiked, original sample and two samples spiked with high and low concentrations of standard, respectively. This would lead to a tripling of samples and blanks, and therefore to an impractical number in terms of analysis run time and data processing time. External standards were not available for all metabolites of interest, so external calibration was not an option, too. Consequently, only relative quantification was possible for this dataset, and the results are depicted in Figure 20.

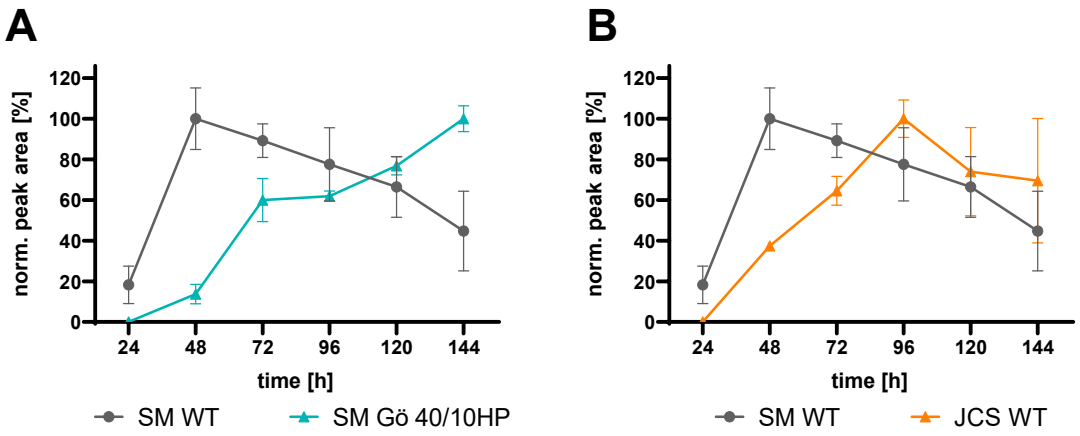


Figure 20: Relative amounts of collinolactone (**1**) of producer strains cultivated either in SM or JCS media over 6 days. The relative peak area of collinolactone in the wild type strain grown in SM media (grey) serves as a reference.
(A) peak area of collinolactone of the overproducer strain *Streptomyces sp.* Gö 40/10^{HP} in JCS media (turquoise).
(B) peak area of the wild type strain grown in JCS media (orange). All experiments were performed in triplicate; the data points represent the mean ± SEM.

As evident from Figure 20A, the onset of collinolactone (**1**) production in *Streptomyces sp.* Gö 40/10^{HP} in SM media is shifted by 24h compared to the wild type. Likewise, a production delay was observed when using JCS media instead of the standard media (Figure 20B) for the wildtype whereas the addition of XAD-16 does not show any effect on the start of production but on production saturation This indicates, that production rates of the wild type strain grown in SM media cannot be directly transferred to the overproducer strain or to fermentation in JCS media and an adaption of the overall incubation time will be necessary for future cultivation.

By taking the aforementioned changes in incubation time into consideration, the experimental set-up was repeated. The obtained maximum concentrations of collinolactone (**1**) under different growth conditions (Figure 21) point to the fact that there was no significant increase of collinolactone (**1**) in JCS media compared to SM media. However, the addition of XAD-16 resin triggered a 5-fold production boost of collinolactone (**1**) concentration for the wild type strain (displayed in gray; $p \leq 0.01$).. While no significant difference between wild type and overproducer strain was observed in SM, there was a significant 6.5-fold increase in the production for the overproducer Gö 40/10^{HP} (displayed in turquoise) in JCS indicating that production capacity of collinolactone (**1**) is limited in SM probably due to negative feedback mechanisms and hindered precursor supply.

The addition of XAD-16 resin to Gö 40/10^{HP} (in SM as well JCS media) only led to an additional 1.5-fold boost compared to the production of the without XAD-16. This indicates that some regulatory effects or nutrient deficiency limit the overall production rate.

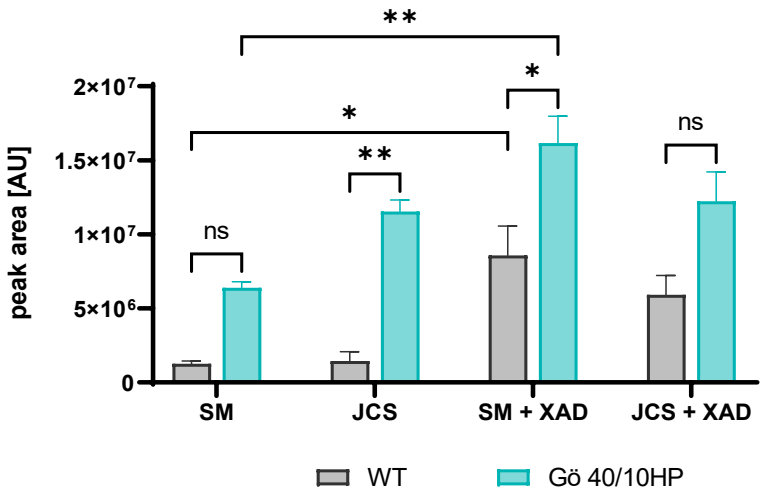


Figure 21: Maximum concentrations of collinolactone (**1**) under different growth conditions. Data obtained for the wild type strain are displayed in grey, data from the strain *Streptomyces sp.* Gö 40/10^{HP} are presented in turquoise. All experiments were performed in triplicates. The data points represent the mean ± SEM. Significance was determined using two-way ANOVA followed by Tukey's multiple comparison test. Significance levels are: *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001, ****p ≤ 0.0001, ns: not significant.

2.5.1 Isolation of a novel metabolite derived from collinolactone (**1**)

Within the reference cultivation (Figure 20, gray), the concentration of collinolactone (**1**) reached a maximum after 48 hours, before it began to decline until the fermentation was aborted after 144 hours. By comparing the metabolite profile of multiple samples after 48 hours, a new compound was identified.

Using the advantage of high-resolution mass spectrometry, which allows to calculate possible sum formulas for a compound based on exact mass and isotopic pattern matching, $C_{21}H_{30}O_6$ was identified as the most likely sum formula candidate. Based on the difference between the sum formula of **1** and the formula candidate, it was assumed, that **1** was hydrolyzed by the addition of water, either at the 7-membered or 6-membered lactone rings (Figure 22).

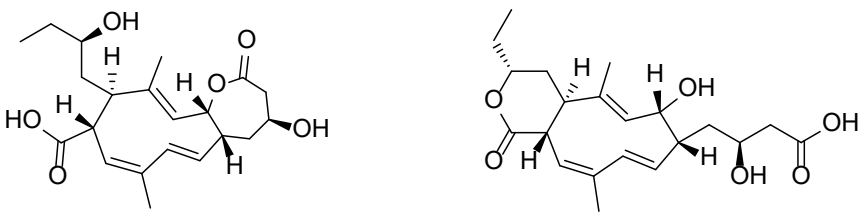


Figure 22: Proposed structures of the side-product hydroxy-collinolactone with hydrolysis of either the 6-membered or 7-membered ring system.

The preparative isolation of this compound turned out to be very challenging, presumably due to the formation of a free carboxylic acid and the yield (< 0.1 mg/L) was not sufficient for a detailed characterization by NMR. The metabolite has not been described before, because it is almost insoluble in ethyl acetate, which was exclusively used for all extractions of *Streptomyces sp.* Gö 40/10 in the past.[154-156]

When comparing the amounts of hydroxylated collinolactone under different growth conditions (Figure 23A), almost none was found in samples where XAD-16 resin was added during the fermentation. This indicates that the addition of XAD-16 adsorber resin suppresses the degradation of **1**, by protecting the product from hydrolysis. The ratio of hydroxylated product to collinolactone (**1**) was significantly higher in the wild type strain than the overproducer Gö 40/10^{HP} in both, SM and JCS media. This is due to the fact, that the production of collinolactone (**1**) declines after 48 hours for the wild type (Figure 20B, gray curve) while it is still ongoing in the overproducer Gö 40/10^{HP} after 144 hours (Figure 20B, turquoise curve). Only when the timely increase in hydroxylated collinolactone outweighs the increase in collinolactone (**1**) production, the ratio of both starts to increase, i.e., after 24h for wild type in SM media and after 96h for wild type in JCS media and the overproduction strain (Figure 23B).

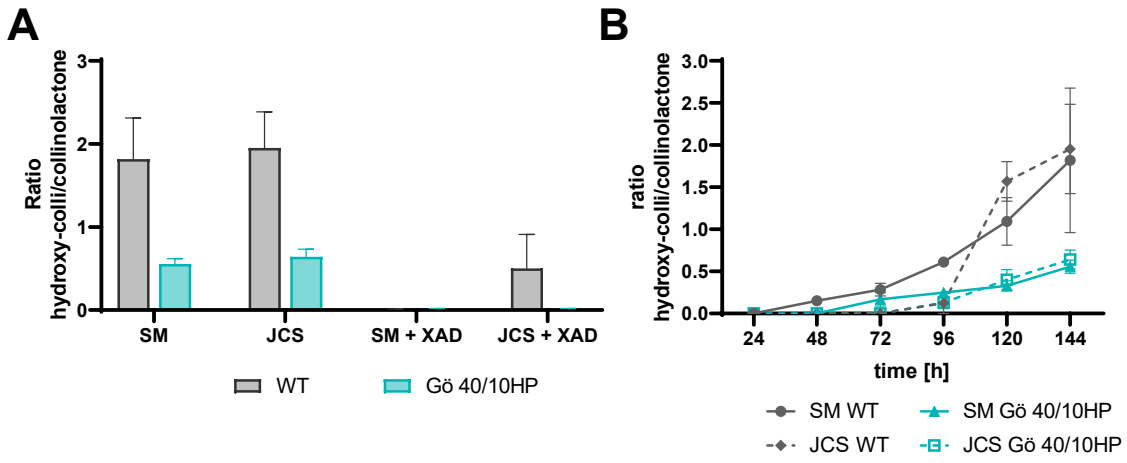
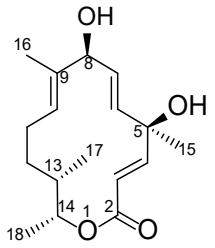


Figure 23: Ratio of hydroxylated collinolactone (hydroxy-collino) and collinolactone under different growth conditions and over time, (SM media and JCS media with and without the addition of XAD-16). Wild type strain (grey) and strain *Streptomyces sp.* Gö 40/10^{HP} (turquoise).
(A) Ratio of hydroxylated collinolactone (hydroxy-collino) and collinolactone under different growth conditions (SM media and JCS media with and without the addition of XAD-16) after 144h of incubation. (B) Ratio of hydroxylated collinolactone (hydroxy-colli) and collinolactone under different growth conditions over time. All experiments were performed in triplicates. The data points represent the mean ± SEM.

2.5.2 Isolation and identification of cineromycin (**3**)

During the final purification step of crude products of **1** by HPLC obtained from fermentation of *Streptomyces sp.* Gö 40/10^{HP} in JCS media supplemented with XAD-16 resin, a yet unidentified impurity was found. All attempts to optimize the parameters towards a base peak separation on preparative HPLC column (diameter of 20 mm) failed. Hence, a change to a semi-preparative column with a smaller diameter (8 mm) and also smaller particle size resulting in a better peak separation, was required. This led to a significant increase of the time required to purify **1** due to the reduced capacity of the semi-preparative column compared to the preparative one.
Initial structure elucidation of the impurity by LC-ESI-HR-MS failed, therefore, the impurity was purified and characterized by NMR. By comparing the recorded spectra with metabolites already known to be produced by this strain, the impurity was identified as cineromycin B (**3**).



3

In this particular case, the LC-MS analysis was not successful due to a quantitative loss of water during ionization in the ESI source, leading to an incorrect assignment of the compound's $[M+H]^+$ peak. Consequently, a search based on the wrong exact mass was performed and returned no candidates.

It was hypothesized that the production of cineromycin B (**3**) and collinolactone (**1**) are linked as products of the same gene cluster. Single gene clusters which are responsible for different secondary metabolites have already been described.[204,205] By insertion of the KasOP* promotor, not only the production of **1** was increased but also of compound **3**. This hypothesis was refuted by the fact, that compound **3** is still produced in the deletion strain -*col*.
The ratio of the amounts of cineromycin (**3**) and collinolactone (**1**) (Figure 24A) are significantly higher in the wildtype strain (gray) than in the overproducer *Streptomyces sp.* Gö 40/10^{HP} (turquoise). An effect from the media was not observed. Also, the ratios were found to be much lower for cultivation with XAD-16 in the case of the wildtype strain as well as the overproducer strain. When considering the fact, that

the amount of **1** found in cultivation with XAD-16 is 5-fold increased, the production of **3** is also heavily boosted by the addition of XAD-16 to compensate for the increased amount of collinolactone (**1**).

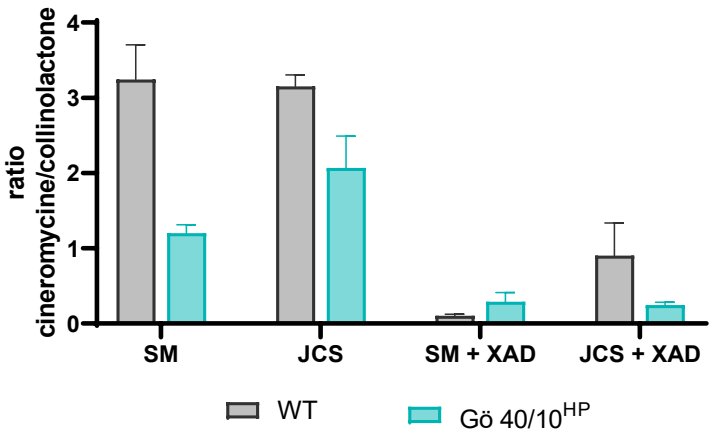


Figure 24: (A) Maximum ratio of cineromycine B (**3**) to collinolactone (**1**) under different growth conditions.
Data obtained for the wild type strain are displayed in grey, data from *Streptomyces sp.* Gö 40/10^{HP}are presented in turquoise. All experiments were performed in triplicates. The data points represent the mean ± SEM.

2.6 Chemical derivatization

Structure-Activity relationship (SAR) studies are a common and well-established way to investigate the potential mode of action by semisynthetic modification of the original scaffold to tune the molecular properties.[206] Collinolactone's (**1**) deficiency of functional groups only allows modifications of the hydroxyl group at C-4 besides lactone ring opening attempts (chapter 2.4 – Isolation and purification) and hydration of the cyclodecatriene system. Initial experiments performed by HOFFMANN indicated that a hydration of **1** is far from trivial. The overall yield was only 15 % and detailed NMR studies showed that the obtained product was not completely hydrogenated with the double bond between C-14/C-15 still in place.[155]

Epoxidation:

Single crystal X-ray crystallography is the gold standard for structure elucidation, especially to determine the absolute configuration.[207,208] Due to the flexible scaffold of collinolactone (**1**), all crystallization attempts of the natural product itself failed. HOFFMANN was able to obtain a crystal suitable for X-ray crystallography after fixation of the flexible methyl-group at C-14 using meta-chloroperoxybenzoic acid (mCPBA).[209] It was expected that all double bonds are epoxidized at once. Indeed, the NMR spectrum of the purified product only had one distinct set of signals, indicating that flexible scaffold was fixed, and interconversion is disabled. Nevertheless, a full structure elucidation revealed that a transannular cyclization occurred and collazulene (**4**) was obtained as product. A plausible mechanism for this intramolecular reaction is presented in Figure **25**. In a first step, the electrophilic oxygen atom of the mCPBA reacts with the nucleophilic carbon-carbon double bond in a concerted reaction resulting in an epoxide. In a second step, the oxygen is protonated (by the generated *m*-chlorobenzoic acid) which weakens the oxygen-carbon bond, before a nucleophilic attack of the negatively polarized double bond is taking place. In a last step, the generated carbenium ion undergoes elimination reaction, leading to the final product **4**.

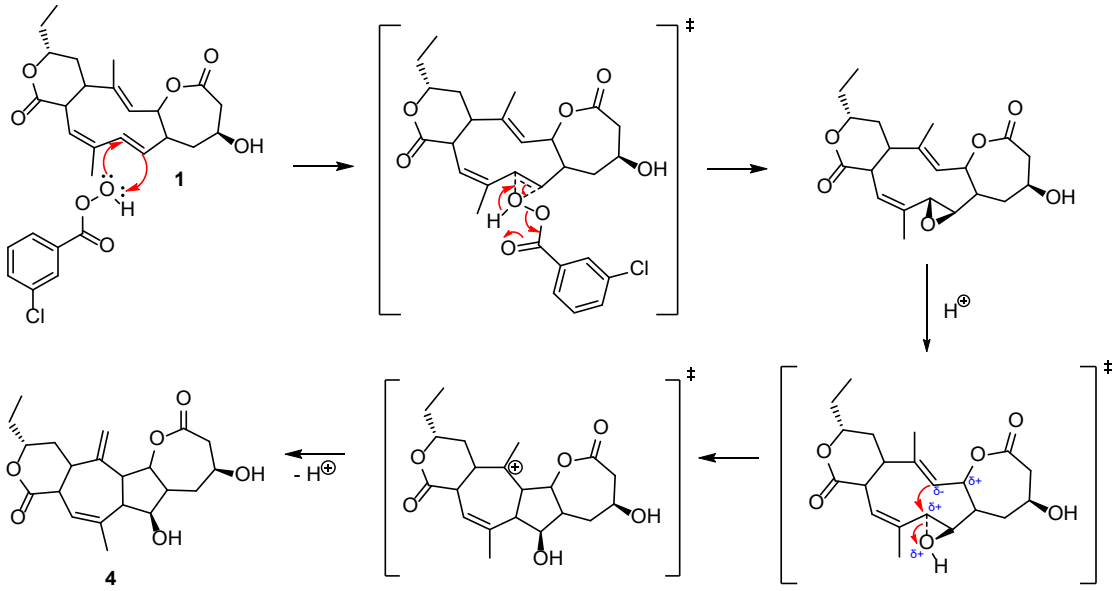


Figure 25: Postulated mechanism for the intramolecular transannular cyclization observed for the reaction of collinolactone (**1**) with mCPBA to yield the tetracyclic compound **4**.

Acetylation:

The poor cellular uptake of polar components which contain free acid or hydroxyl moieties is one major problem to address. Acylation of these moieties with a variety of different acids has shown to increase the cellular uptake and eventually the biological activity.[210] It is assumed that acylated compounds can cross lipid bilayers more easily and are then unmasked by endogenous esterase activity prior to their utilization.[211] Collinolactone (**1**) can be acylated with acetic acid anhydride (to **5**) in very good yields (> 80 %)[155] and by benzoic acid (to **6**) using a classical Steglich approach for stereo chemically hindered secondary alcohols.[212]

Dehydration (Burgess reaction):

During synthesis of derivatives **5** and **6**, the formation of small amounts of collinolactenone (**7**) was observed as a side product.[156] The current hypothesis states that the reaction occurs by forming a six-membered transition state which is heavily influenced by the leaving group ability of the linked acid (Figure 26). Collinolactenone (**7**) features an α,β-unsaturated michael system and is therefore favored to be targeted by nucleophilic attacks, e.g., of free thiol group containing compounds, e.g., glutathione or cysteine containing proteins in cells. The use of burgess dehydration gave direct synthetic access to **7**. The obtained yields fluctuated between 30 and 70 %, highly dependent on the quality of the commercially available Burgess reagent. Reactive impurities found in the reagent led to the formation of undesired side products.

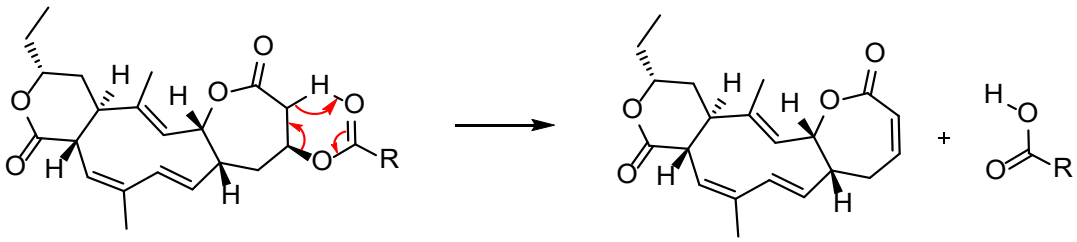


Figure 26: Proposed reaction for the formation of collinolactenone (**7**) over an energetically favored 6-membered transition state.

.

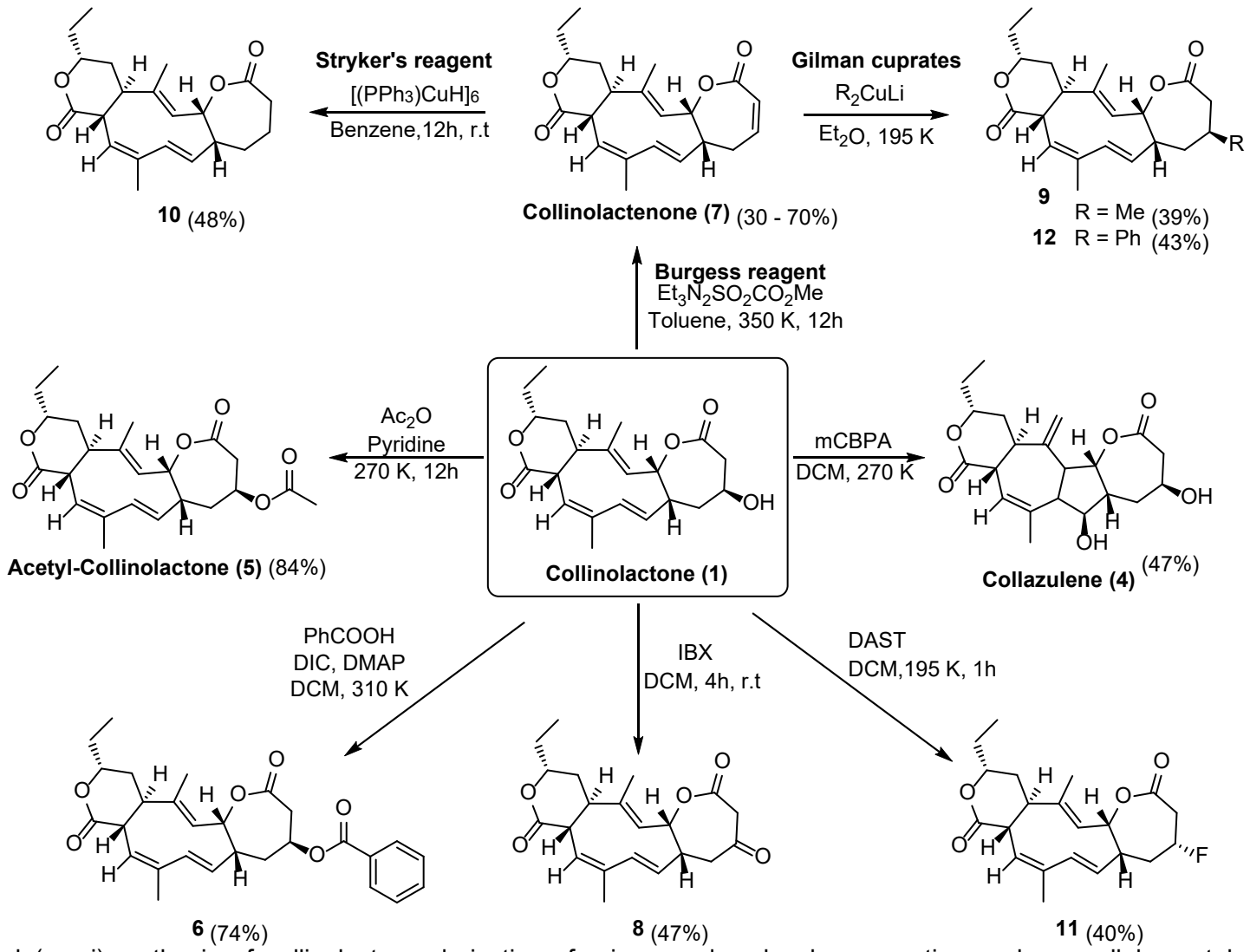


Figure 27: Chemical (semi)-synthesis of collinolactone derivatives for improved molecular properties such as cellular uptake and improved crystallographic characteristics.
Adapted and modified from J. C. Schmid *et al.*, *Angew. Chem. Int. Ed.* **2021**, *60*, 23212.

Oxidation (Dess-Martin-Oxidation):

It was hypothesized that collinolactenone (**7**) is the bioactive species and other modifications such as esters only serve as a prodrug. Metabolic studies of lysed cells incubated with compounds **5** and **6** were not conclusive, as small amounts of compound **7** were already present in the stock solutions. Therefore, synthetic ways to inhibit the *in-vitro* formation of **7** and to improve cellular uptake at the same time were examined.

Oxidizing the hydroxyl group of **1** with 2-Iodoxybenzoic acid (IBX) to yield compound **8**, the affinity towards forming collinolactenone (**7**) is supposed to be decreased and indeed, no formation was observed. The polarity of the molecule remained comparable to collinolactone (**1**) based on their retention time profiles.

Alkylation (Gilman cuprates):

To present a non-hydroxylated analogue of the natural product **1**, Gilmann cuprates were used to replace the hydroxyl group by a methyl group at position C-4 yielding compounds **9** and **12**. Gilman cuprates are a well characterized class of lithium containing diorganocopper reagents.[213] They favor 1,4 additions on enones rather than 1,2 additions due to their low nucleophile. Furthermore, they tolerate a range of functional groups such as lactones and alkenes making them useful for synthesis of complex natural products.[214] One disadvantage is the missing stereoselectivity, usually resulting in mixtures of isomers where the observed ratio of the isomers is dependent on steric effects of the educt. The stereo configuration has been determined afterwards based on NOESY correlations and was found to be the same as in the initial natural product **1**. Insertion of a methyl-group (**9**) and a benzyl-group (**12**) successfully suppressed the formation of **7** and decreased polarity at the same time.

Hydration (Strykers reagent):

Stryker's reagent, another organocopper reagent, was used for the selective hydration of the α,β-unsaturated carbonyl group of compound **7** yielding compound **10**. All other double bonds in the cyclodecatriene system remained unaffected. It was not possible to synthesize the desired product using commercially available Stryker's reagent. Instead, a fresh batch following a previously described protocols had to prepared.[215]

Hydration of the double bond at position C3-C4 suppressed the formation of **7** and decreased polarity at the same time.

Fluorination (DAST):

One commonly used approach in medicinal and pharmaceutical chemistry is the replacement of hydroxyl groups by fluorine to improve their pharmacological properties.[216] Collinolactone (**1**) was fluorinated (yielding compound **11**) using diethylaminosulfur trifluoride (DAST) which facilitates the replacement of primary and secondary alcohols by fluorine. The reaction mainly follows a SN_2 mechanism, but elimination and SN_1 products have also been reported.[217] For this reaction, the elimination product collinolactenone (**7**) was observed as a side product in overall yields up to 45 %.

All synthesized compounds except collazulene (**4**) showed two distinct sets of signals in NMR. Interestingly, the biosynthetic intermediate **2** with its 6-10-6 membered ring system does not show an additional set of signals. This indicates that the scaffold of collinoketone (**2**) is more rigid due to a higher rotational energy barrier and becomes flexible as soon as the molecular oxygen was incorporated forming collinolactone (**1**).

2.7 Biological profiling of collinolactone

For more than a decade, the natural product collinolactone (**1**) itself remained inactive in a variety of assays including tests for antibacterial or antifungal activity.[181] This changed when *Kwon et al.* discovered that **1** dissociates misfolded protein aggregates and reduces apoptosis, both phenomena related with Alzheimer's Disease.[157]
For some derivatives of collinolactone (**1**), cytotoxic effects on eukaryotic cells were found and further studies discovered the formation of monopolar spindles.[155] However, insufficient amounts of the isolated natural product **1**, non-systematic experimental design, inconsistent data availability and a lack of reproducibility required further and extended biological profiling of collinolactone (**1**) and selected derivatives.

2.7.1 Cell viability screening on murine L929 cell model

Cell viability is an important parameter in various medical fields, e.g., therapeutic treatment development or in biomaterials research for implants. These experiments can either be used solely to determine the cytotoxicity of a potential drug candidate, but also correlated to cell behavior or other parameters such as optimized growth conditions. Regardless of the type of cell-based assay being used, it is important to know the precise number of viable cells in response to extracellular stimuli, chemical agents, or therapeutic treatments. To address this question, several methods have been developed using a variety of different parameters such as adenosine triphosphate (ATP) activity,[218,219] protease activity,[220] reduction activity,[221,222] lactate dehydrogenase activity[223] and deoxyribonucleic acid (DNA) binding dyes.[224] Based on the type of research question one must decide which readout is most suitable depending on the experimental design and the technical equipment available in the laboratory. Possible methods for readout are flow cytometry, fluorescence microscopy or microplate readers.[225] Also, it is critical to decide in advance, whether real time monitoring is required, or single endpoint readout is sufficient.[220]
In general, there are two ways how cells can die. The most common pathway is called apoptosis, a highly regulated and controlled process where several biomarkers are expressed which can be correlated to cell viability.[226,227] The second pathway is called necrosis, where cells typically undergo rapid swelling after acute cellular injury. They lose membrane integrity, shut down metabolism and release their cytoplasmic contents into the surrounding culture medium and usually no apoptotic markers are expressed.[227,228] When choosing a suitable assay readout, the cell death signaling pathway has to be included into the considerations.

MTS tetrazolium assay was found to be the most suitable readout for cell viability assays performed for collinolactone (**1**) and its derivatives. It is a quantitative colorimetric assay which is based on the reduction of tetrazolium salts to generate formazan. In contrast to the widely used MTT assay, a water-soluble formazan product is obtained. Therefore, the additional washing step required for MTT assays is omitted, increasing reproducibility and reducing time consumption per assay.[220] The assay can be performed in well plates using a plate reader for the readout and multiple assays can be done simultaneously. Since this assay has no endpoint and conversion of tetrazolium continues until it is depleted, the readout time must be optimized (1 - 4 hours) and controls must be included for every well plate to precisely track differences in processing and incubation time.

The initial cell viability screening concentrations on the L929 cell line ranged from 0.13 µM to 250 µM and were extended to 1000 µM when required for improved fitting purposes (Figure 28 and Figure 42). Dimethyl sulfoxide (DMSO) was used to dissolve the compound stocks, which were then further diluted into culture media. Due to its cell toxic effects,[229] the maximum DMSO concentration was set to 0.1 %. Due to solubility issues of acetyl-collinolactone (**5**) at higher compound concentration levels, the DMSO concentration had to be increased to 0.2 %, leading to a decrease of cell viability of the control (Figure 28B). Only acetyl-collinolactone (**5**) and collinolactenone (**7**) were found to reduce cell viability with EC_{50} values of 106.5 µM and 35.6 µM, respectively. The natural product collinolactone (**1**) and all other derivatives were inactive.

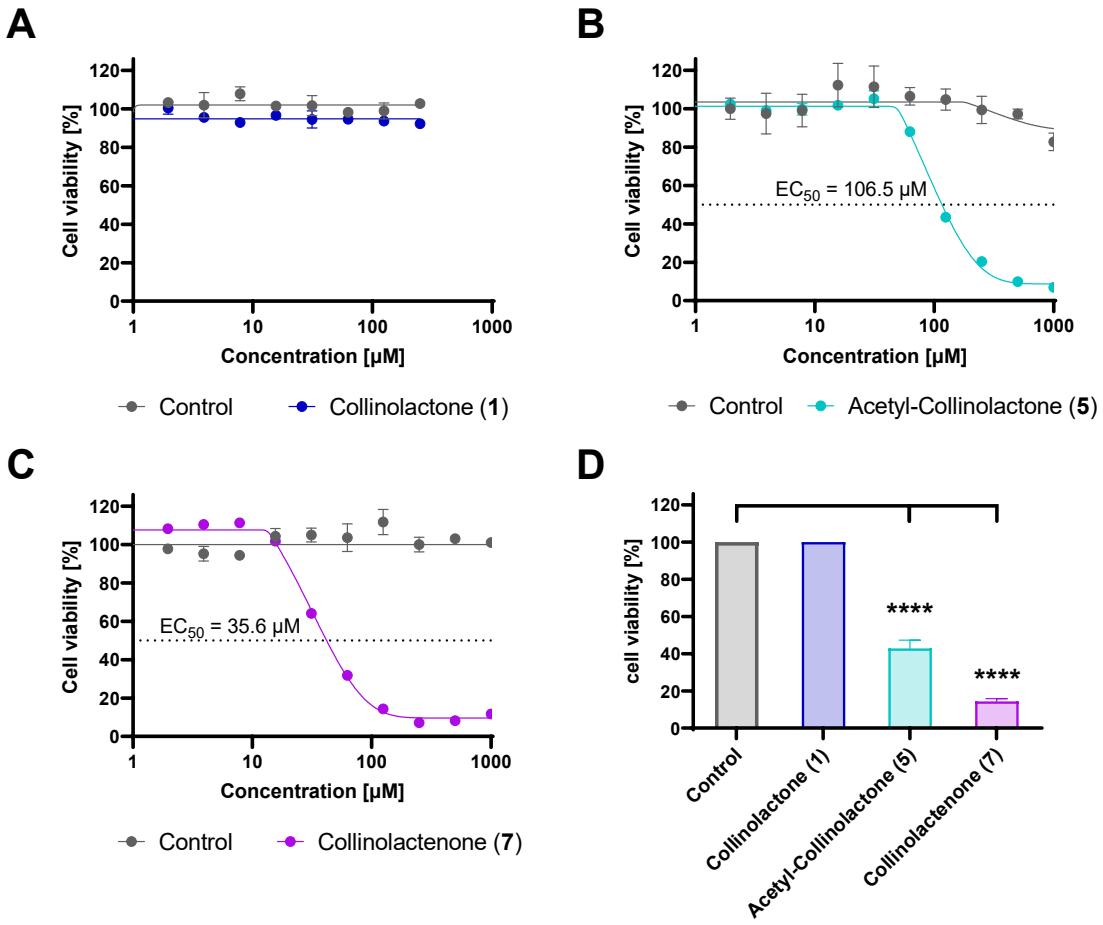


Figure 28: Results of cell viability screening on L929 cell line. The control is displayed in black, the response to treatment is colored. (A) collinolactone (**1**), (B) Acetyl-collinolactone (**5**) and C: collinolactenone (**7**). D: overview of the obtained EC_{50} values. All experiments were performed in triplicates. The data points represent the mean ± SEM. Significance was determined using one-way ANOVA followed by Dunnet's multiple comparison test. Significance levels are: *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ****P ≤ 0.0001, ns: not significant. All other comparisons were also not significant.

2.7.2 Fluorescence microscopy on Ptk2

Fluorescence microscopy is a highly sensitive and reliable technology used by scientists to observe the localization of molecules within cells. In general, it is a non-destructive and gentle method and allows to study dynamic processes in living cells. For example, it can be used to visualize the mode of action of natural products in eukaryotic cells and bacteria.[230-232] One major requirement is that the molecule of interest must have fluorescence properties suitable for microscopy. For example, excitation and emission wavelengths have to be within available filters on the instrument and the compounds

must exhibit a certain photo-stability. Another option to visualize the cellular effects of a molecule of interest is to stain potentially affected cell organelles with a fluorescence dye. This is often achieved by using antibodies linked to a fluorophore.[233]

To investigate the mode of action for collinolactone (**1**) and its derivatives **5-7** that exhibit effects on cell viability, fluorescence microscopy on PtK2 cells were performed. The PtK2 cell line, established from a rat kangaroo,[234] is a popular model for the study of mitoses. They only have 14 chromosomes, which are large compared to the chromosomes of humans so that they can be easily visualized with standard fluorescence microscopes. In addition, the cells itself are relatively large and grow as an adherent monolayer on the surface and remain flat throughout the whole cell cycle.

To increase the number of cells which are currently in a specific cell cycle state during fixation prior to fluorescence microscopy, a method called cell synchronization was applied.[235] The overall goal is to synchronize cells at different stages, so that all cells are then in the same cell cycle phase. This can be achieved by depriving FBS (fetal bovine serum) which contains essential growth factors for the cells. Synchronization turned out to be challenging in initial starving experiments. Even after 48 hours of FBS depletion, the cells were still dividing, and confluence was almost reached at this time. The highest division rates were found after 30 hours of incubation after seeding. Therefore, two sets were prepared – the first set was incubated for 30 hours (set A) and the second one for 72 hours (set B) before the compounds were added and further incubated for another 12 hours prior to fixation. The first set is expected to have fewer cells but higher division rates, while the second set has a higher cell density but an already decreased division rate due to the reached confluence. The images of the dividing cells (Figure 29) were all collected after 30 hours of initial cultivation (set A).

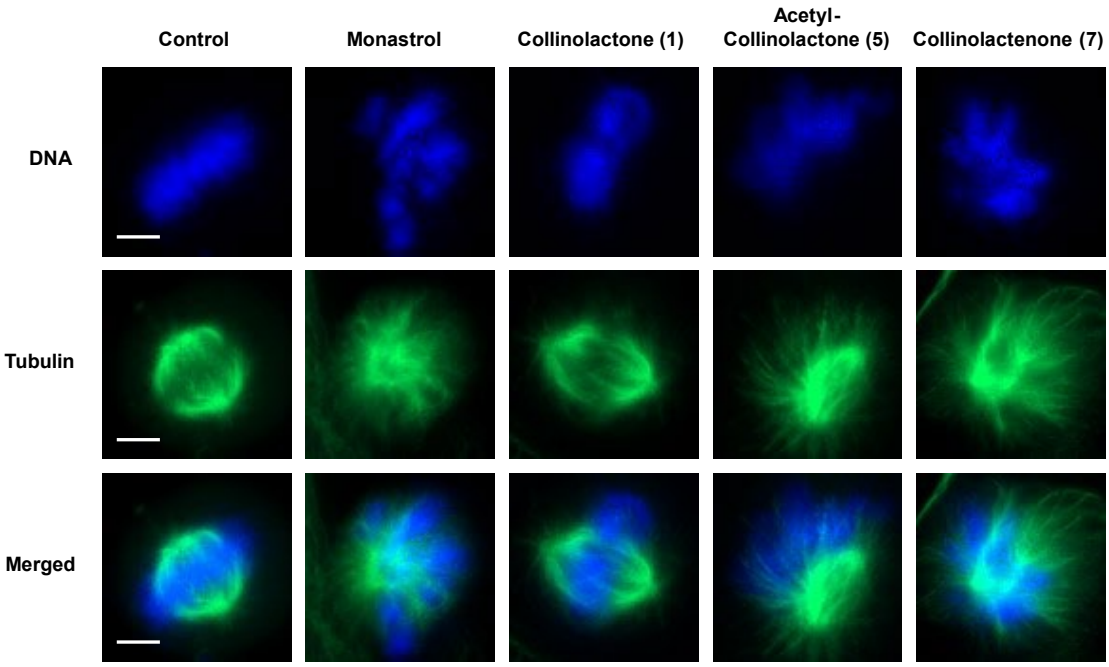


Figure 29: Fluorescence microscopy images of selected PtK2 cells during mitosis (scale bar 5 µm). Blue (top): DNA, condensed to chromosomes; Green (middle): microtubules form the mitotic spindles; Merged (bottom): both channels are superimposed. Cells were incubated with 0.2% DMSO (control), monastrol (a positive control for monoastral phenotype), collinolactone (**1**), acetyl-collinolactone (**5**) and collinolactenone (**7**). Cells were cultivated for 30 hours prior incubation with compounds for 12 hours before fixation and staining (set A).

Normally, before a cell can divide, the genetic information is duplicated into two identical copies. This is done during mitosis, where the chromosomes are oriented in an equatorial plane and microtubules pull the chromosomes towards the two spindle poles (Figure 29 – control) before the cell divides. For the derivatives **5** and **7**, an increase in the presence of abnormal spindles by 40 % (Figure 43) was found where the spindle poles are centered, and the chromosomes spherically arranged around. The observed type of abnormal spindles is called monoastral spindles and was reported also for Monastrol, which served as a control in this experiment.[120] No abnormal spindles were observed for collinolactone (**1**).

Due to laboratory and time limitations[7], the experiments could not be performed in independent triplicates and the number of cells undergoing mitosis during microscopy is

[7] Prof. Dr. Ralf Kemkemer from the Reutlingen University kindly provided his laboratory and equipment for the experiments. I thank Kerstin Frey for the initial training in cell culture techniques and fluorescence microscopy. All experiments and data were independently recorded and analyzed afterwards by me.

low compared to the total number of cells (less than 0.1 %). Thus, these results can at best serve as an initial trend. Nevertheless, the formation of abnormal spindles was observed in different laboratories over several years and repetitions.[155,156]

Following a procedure which describes the utilization of fluorescence microscopy images to determine the cell stage distribution,[236] CellProfiler[237] was used to process the raw data images, including nuclei detection and segmentation. The obtained intensity distribution plot (Figure 30) indicates, that the derivatives acetyl-collinolactone (**5**) and collinolactenone (**7**) had no effect on the cell cycle state distribution. However, for collinolactone (**1**), significantly more cells were found to be in S and G2 phase (see Figure **4**). These results should be considered as preliminary, as this method has some drawbacks compared to the widely applied flow cytometry method. For example, although cells in G2/M phase are correctly classified, cells in anaphase and telophase are sometimes recognized by the algorithm as separate objects and counted to the G1 phase, leading to an overall smaller G2/M population compared to standard flow cytometry.[236]

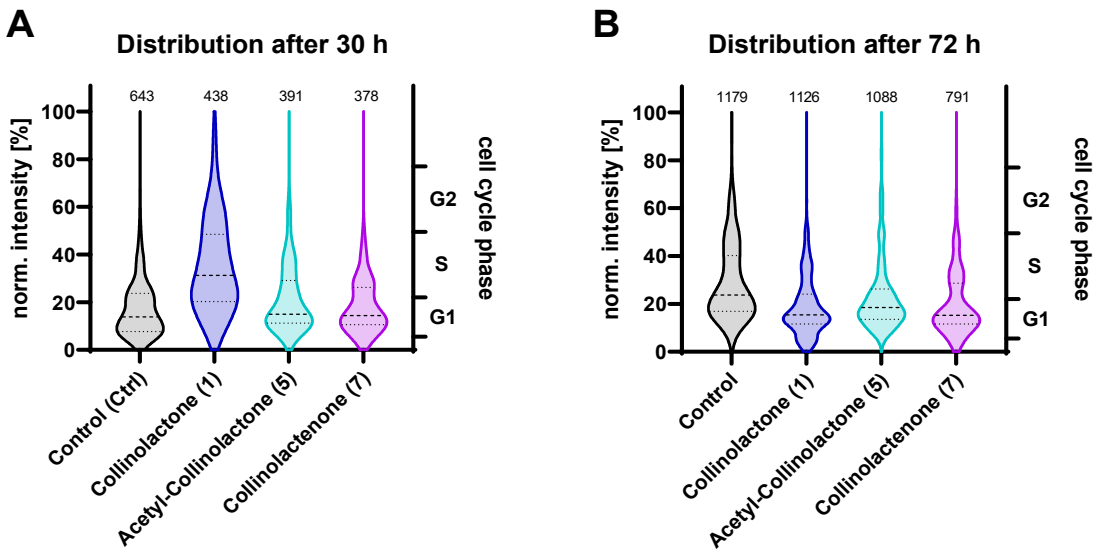


Figure 30: Nuclei intensity distribution based on fluorescence microscopy images from PtK2 cell line after 30 hours (set A) and 72 hours (set B). The intensity correlates with the cell cycle state in which the cell is currently in and is displayed on the right-hand axis. The control (0.2 % DMSO) is displayed in grey, collinolactone (**1**) in blue, acetyl-collinolactone (**5**) in turquoise and collinolactenone (**7**) in purple. The number of cells detected during the analysis is displayed above each violine.

In general, cell-cycle modulating compounds are of therapeutic interest for anti-cancer therapy as phase-specific cytotoxic drugs are more potent when cells pass through their sensitive phases more frequently.[238] Progression of cell cycle is primarily controlled by

the ubiquitin-proteasome system (proteolytic degradation) and phosphorylation of specific proteins by cyclin-dependent kinases (CDKs) and their dephosphorylation by phosphatases.[239] The latter case could be linked to the described reduction of apoptosis and clearance of misfolded hyperphosphorylated tau protein by collinolactone (**1**).[157]

2.7.3 Eg5 Kinesin and Dynein motor protein assay

Inhibitors or modulators of the cell cycle can serve as potential therapeutic drug in cancer therapy. Therefore, the formation of the monopolar spindles is of particular interest and was studied in more detail. Monastrol, used as a control for the monopolar spindle formation, has been found to inhibit the Eg5 kinesin motor protein[122] making the Eg5 a putative target for compounds **5** and **7**, which were also found to exhibit monopolar spindles.

Kinesin motor proteins utilize the energy of ATP hydrolysis to move along microtubules. The resulting ADP and, more specifically, the inorganic phosphate can be correlated with the motor protein activity using the malachite green method to quantify the phosphate.[240] The initial assay was developed by Kodama et al.,[241] and a kit containing the required reagents and proteins is commercially available from Cytoskeleton, Inc.[8] However, the assay requires a significant amount of preparation and optimization for each type of motor protein used in this assay. Since each motor protein batch has a different activity, constant adjustments are essential.

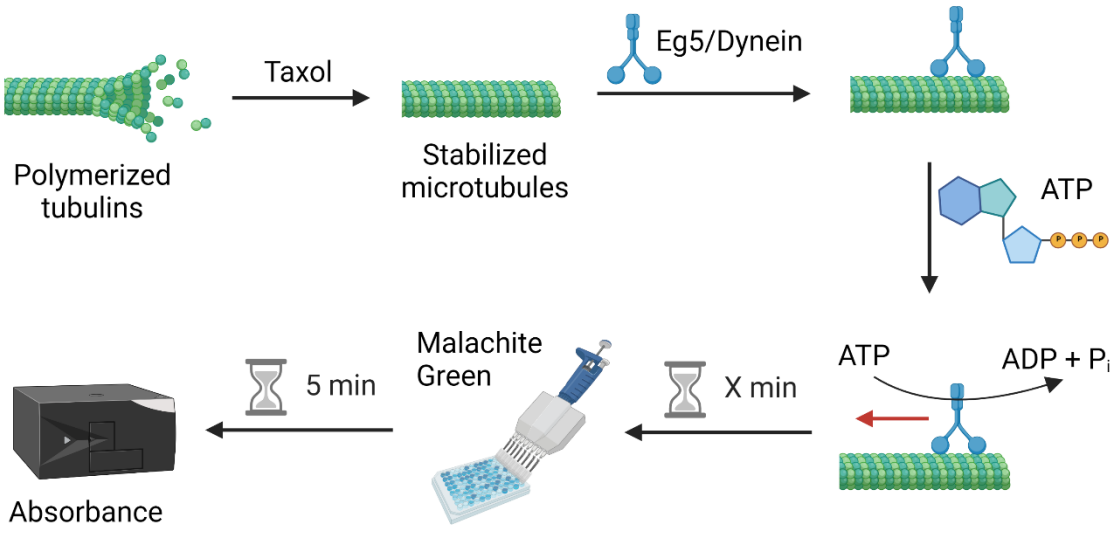


Figure 31: Schematic representation of the motor protein assay based on the ATP hydrolysis rate.

[8] https://www.cytoskeleton.com/

Microtubules are a highly dynamic protein system which spontaneously polymerizes at temperatures above 30 °C, and depolymerizes below 12 °C.[242] Therefore, it was important to control the temperature of the microtubules stock solution across the whole experiment (Figure **31**). To prevent depolymerization, taxol,[93] a compound used in cancer therapy and a known depolymerization inhibitor, was added.[243]

In a second step, the motor protein of interest is added which then binds to the polymerized microtubules. The required amount of motor protein is dependent on the specific activity and can vary from batch to batch. A titration of the kinesin is essential to obtain optimal signal to noise ratios. The linearity of the assay ranges from 0.2 to 1.5 nmol. Therefore, an ideal turnover rate is between 0.5 and 1 nmol. For Eg5, an amount of 0.6 µg was found to result in an optimal absorbance increase, while for dynein 1 µg was used. The results of titration for the Eg5 and the dynein motor protein batches are presented in Figure 44 C and D.

As soon as ATP is added, the motor protein starts to hydrolyze the ATP to ADP and inorganic phosphate. The resulting energy from the hydrolysis is consumed to move along the microtubules track. The execution time of the assay is another important parameter which must be optimized. It is related to the specific activity of the motor protein and the amount of the applied motor protein. The optimization of three dependent parameters without consuming a high amount of expensive protein is challenging. For example, the Eg5 protein has a minimum conversion rate of 600 nmol/min/mg while dynein has a rate of only 13 nmols/min/mg according to their datasheets.[9] Therefore, either much higher amounts of dynein are required, or the running time must be extended. In this experiment, Eg5 was incubated for 5 minutes and dynein for 90 minutes to obtain ideal results.

By adding the malachite green solution which contains high concentrations of sulfuric acid, immediate deactivation by denaturation of the proteins and a sudden stop of the ATP conversion is accomplished. The remaining non-converted ATP tends to hydrolyze under acidic conditions, thus the strict adherence of a 5-minute reaction time window is crucial, before the readout is performed in a plate reader.

[9] https://www.cytoskeleton.com/pdf-storage/datasheets/cs-dn01.pdf, https://www.cytoskeleton.com/pdf-storage/datasheets/eg01.pdf (access: 28.06.2022, 20:25)

It is mandatory to avoid any sources containing inorganic phosphate when buffers and other reagents are prepared. More precisely, only triple-dest water and high-purity sodium hydroxide for pH adjustment of buffers were used.

Monastrol, a known inhibitor of the Eg5, served as a positive control for the Eg5 (Figure 32A) while ciliobrevin D served as a positive control for the dynein (B). Collinolactone (**1**) did not show any inhibition effect on Eg5 nor on dynein. Both, acetyl-collinolactone (**5**) and collinolactenone (**7**) showed an inhibition effect on Eg5, but not on dynein.
More interestingly, also benzoyl-collinolactone (**6**) showed an effect on the Eg5, but also not on dynein. This was unexpected as **6** did not show any activity in the cell viability screening (chapter 2.7.1 – Cell viability screening on murine L929). A reason might be, that the cellular uptake of **6** is significantly reduced compared to compounds **5** and **7** or that there are differences in the binding pocket of Eg5 from the murine derived L929 cell line and the commercially available homo sapiens recombinant Eg5.
Many of the known antimitotic compounds interfere with the assembly and disassembly of tubulin into microtubules. Therefore, motor proteins can no longer functionalize, and a decrease of activity would be observed. Since compounds **5-7** only showed an effect on the Eg5 activity, but not on the dynein, any disruptive effect caused by interaction with the tubulin can be excluded as the dynein activity was not effected at all.[120]

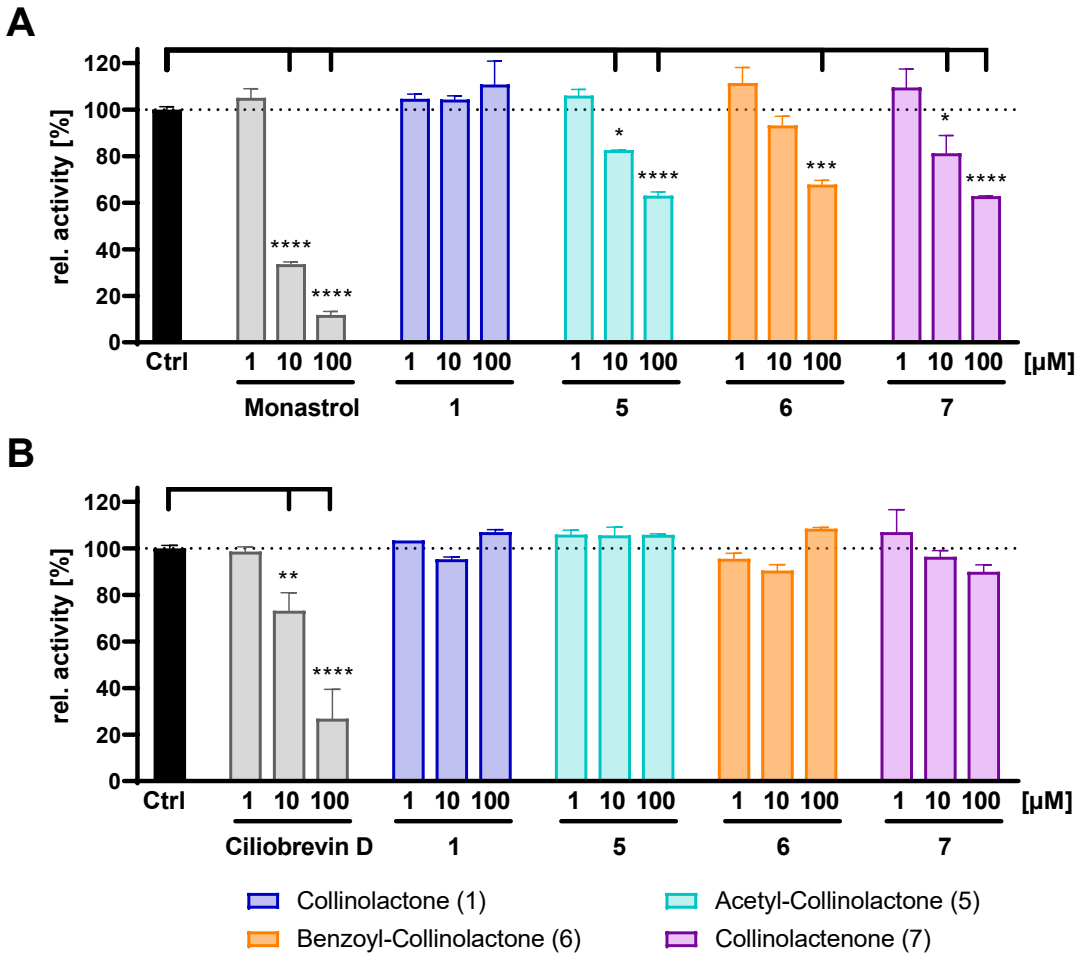


Figure 32: Inhibition of Eg5 (A) and Dynein (B). The obtained values are scaled to the activity of the control. Displayed in grey are the controls, in case of the Eg5 monastrol was used, for the dynein ciliobrevin D served as a control. Collinolactone (**1**) is displayed in blue, acetyl-collinolactone (**5**) in turquoise, benzoyl-collinolactone (**6**) in orange and collinolactenone (**7**) in purple. Each experiment was performed in duplicates. The data points represent the mean ± SEM. Significance was determined using two-way ANOVA followed by Dunnett's multiple comparison test. Significance levels are: *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ****P ≤ 0.0001, ns: not significant.

2.7.4 Glutamate induced oxidative stress assay

The here presented neurotoxicity and neuroprotection assays have been done by Dr. Matthias Scheiner at the Julius Maximilians University of Würzburg, supervised by Prof. Dr. Michael Decker.[2]

In 1989, it was reported that glutamate could induce a calcium-dependent form of delayed cell death by increased intracellular oxidative stress that was associated with depletion of glutathione (GSH).[244] Further studies revealed that extracellular glutamate acts as a competitive inhibitor of the cystine-glutamate antiporter. Increased concentrations of extracellular glutamate lead to a reduction of the available intracellular amount of cystine that is required for the synthesis of GSH. However, GSH is essential to prevent damage to important cellular components caused by reactive oxygen species.[245]
This new glutamate-dependent form of delayed cell death distinct from apoptosis was later referred to as oxytosis.[246] The key steps of oxytosis include glutathione depletion, lipoxygenase activation, reactive oxygen species accumulation and calcium influx.[247]
The neuronal HT22 cell line was originally isolated from the hippocampus of a mouse. It was found to be especially sensitive to glutamate and therefore established itself as a model for glutamate-cytotoxicity.[248]

Collinolactone's (**1**) potent neuroprotective properties against Alzheimer were discovered by Kwon *et al.*, using the exact same cell line. In this assay, cytotoxicity was induced by the addition of preformed amyloid-beta aggregates and recovery rates of the cells treated with testing compounds were measured.[157]
Based on these findings, collinolactone (**1**) and derivatives were tested in an oxytosis assay to assess their neuroprotective properties. In contrast to the assay performed by Kwon *et al.*, the activity of compounds in this assay is solely based on an intracellular effect, which was not triggered by amyloid-beta.

In a first experiment, the cytotoxicity of the compounds **1-11** on HT22 cells was investigated. In contrast to the results from the cell viability screening (chapter 2.7.1 – Cell viability screening on murine L929), where acetyl-collinolactone (**5**) and collinolactenone (**7**) reduced cell viability with IC_{50} values of 106 and 36 µM, respectively, benzoyl-collinolactone (**6**) and collinolactenone (**7**) showed a decrease on cell viability of HT22 cells (Figure 45A) at concentrations of 25 µM. All other derivatives, including the natural product **1** (Figure 33A) did not show any effect on HT22 cells.

In a second experiment, the neuroprotective properties were investigated. As expected, the addition of glutamate induced toxicity leading to cell death, whereas co-treatment with quercetin protected the cells from the toxic effect of glutamate (Figure 33B).[249] Also collinolactone (**1**) showed neuroprotective properties against glutamate-induced oxidative stress in a dose-dependent manner. Remarkably, none of the other derivatives **2-11** showed any effect of neuroprotection at concentrations up to 25 µM (Figure 33B). Therefore, the activity of compound **1** is highly specific and even small structural changes are associated with a loss of activity. Due to the structural diversity of the tested compounds, the lack of activity could also be related to phenomena such as cellular uptake or different metabolism compared to the active compound **1**.

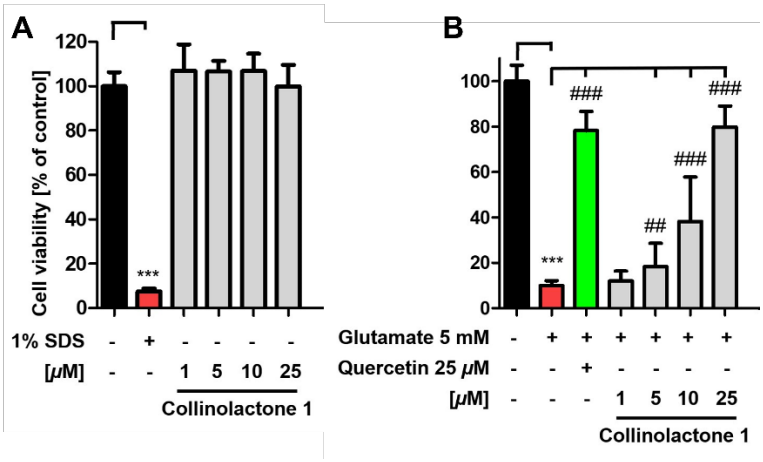


Figure 33: Neurotoxicity (A) and neuroprotective properties (B) of collinolactone (**1**) against glutamate-induced oxidative stress at 1–25 µM using HT22 cell line. 1% SDS served as positive control (red) for neurotoxic effects whereas quercetin (25 µM) served as positive control for neuroprotection (green). Results are presented as means ± SEM of three independent experiments, each performed in sextuplicate, and refer to untreated control cells (100 %). Levels of significance: ***p < 0.001; $^{\#\#}$p<0.01; $^{\#\#\#}$p < 0.001.
Adapted from J. C. Schmid *et al.*, *Angew. Chem. Int. Ed.* **2021**, *60*, 23212.

2.7.5 Amyloid aggregation and disaggregation assay

Aggregation of misfolded proteins, either intra- or extracellular are hallmarks of several neurodegenerative diseases. In fact, accumulation of aggregated amyloid-beta (Aβ) in human brains is believed to be one of the key drivers for the development of Alzheimer disease (Figure 7).[152] Consequently, drugs which disaggregate or prevent the formation of aggregates are of high interest.[250]

Several assays have been developed to assess the potential of small molecules towards

Aβ aggregation. The most frequently used assay is based on Thioflavin T (ThT), which binds the β-sheets of Aβ leading to a significant increase of fluorescence intensity compared to the unbound state.[251] Advantages of this assay are its simple design, high sensitivity and low costs.

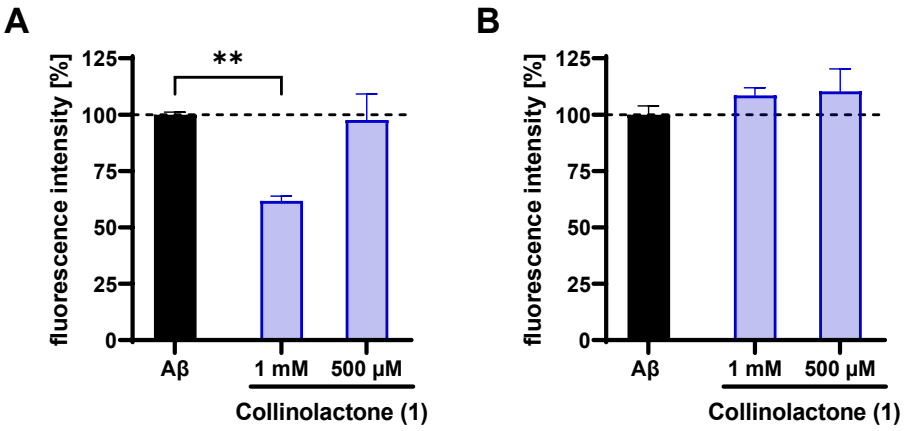


Figure 34: Fluorescence intensity of thioflavin T based amyloid-beta aggregation assay (A) and disaggregation assay (B) for collinolactone (**1**) in concentrations of 1 mM and 500 µM. Each experiment was performed in triplicates. The data points represent the mean ± SEM. Significance was determined using one-way ANOVA followed by Dunnett's multiple comparison test. Significance levels are: *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ****P ≤ 0.0001, ns: not significant. All other comparisons were not significant.

Similar to the assay described by Kwon *et al.*,[157] collinolactone (**1**) and derivatives **3-11** were profiled towards their effects on Aβ. A reduction of Aβ aggregate formation by about 35 % percent for a 1 mM concentration of the natural product **1** was found (Figure 34A), similar to the recently described effect.[157] No effect was observed for a concentration below 1 mM. This raises the question, whether the mode of action of collinolactone (**1**) really is related to the reduction of aggregates as claimed by Kwon *et al.*,[157] as these concentrations are far from any physiologically relevant concentrations.[252,253] One must assume that this is just a side effect and not the true molecular target, especially given the fact, that collinolactone (**1**) has neuroprotective properties in the micromolar range against intracellular oxidative stress (discussed in chapter 2.7.4 – Glutamate induced oxidative stress assay). Another reasonable explanation for the observed reduction of aggregates could be related to competitive binding of tested compounds with ThT to the fibrils as it has already been proven to be the case for some natural products.[254]

None of the tested derivatives showed a decrease in the formation of aggregates (Figure 34B). In contrast, some derivatives were found to increase the formation of aggregates. This might also be due to interference of these compounds with the fluorescence of ThT

itself,[254] when they are spectroscopically active as it is the case for compound **6.** Also, some previously described compounds were found to accelerate fibrilization. Since these fibrillar structures are considered less toxic compared to the intermediate oligomers, a decrease of toxicity was observed for these compounds.[255,256] This might be the case for the derivatives with increased aggregation.

In addition, the controlled formation of Aβ turned out to be very challenging. Several methods have been described, but often with contradictory outcomes.[251,257,258] Apparently, the formation can vary between vendors and even within batches from the same supplier. Therefore, the ability to form aggregates must be verified prior to each experiment.

3. METABOIDENT, A NOVEL PLATFORM FOR AUTOMATED DEREPLICATION

The characterization and quantification of various biological molecules in temporal and spatial dimensions of organisms is referred to as a discipline simply called omics. It can be further divided into several subareas e.g., genomics, proteomics, and metabolomics based on the biological molecule of interest. Proteomics, for example, investigates protein structure, their functionality, and relation to other proteins of an organism or a system,[259,260] while metabolomics uses small molecules and their abundance levels in organisms to obtain insight in cellular chemical processes and to assess the organisms' physiological state.[261,262] Combining and linking collected data from these fields will help to increase our understanding of diseases, hopefully leading to the discovery of novel biomarkers and hence affecting future drug development and therapy strategies.[263]

For omics studies, mass spectrometers are used the most due to their high sensitivity and high throughput capabilities compared to other methods such as nuclear magnetic spectroscopy (NMR). Another advantage of mass spectrometry is its interface to separation techniques such as capillary electrophoresis (CE), gas chromatography (GC) or high-performance liquid chromatography (HPLC) which is required for analysis of complex sample matrices derived from living organisms.

3.1 Organization and structure of MetaboIDent

Even with state-of-the-art separation techniques such as ultra-high-performance-liquid chromatography (UHPLC), one will never obtain a fully resolved chromatogram for matrices such as blood or urine due to the tens of thousands of known and (yet) unknown compounds. The complicated nature of overlapping and unresolved peaks requires the support of automated software for analysis. Multiple tools have been developed for various stages of processing and analysis, e.g., Mzmine,[264] MS-DIAL,[265] CFM-ID[29-31] and MetFrag[27,28] or MetaboAnalyst.[266,267] The application scope of these tools is very different: Mzmine was developed for initial data import and processing and filtering while MS-DIAL was designed for identification of compounds using databases of reference spectra. CFM-ID and MetFrag can predict spectra based on a compound structure and match it to an unknown fragmentation spectrum, which is especially useful for substances where no reference standard is available. MetaboAnalyst was designed for quantitative comparison and statistical analysis. The interconversion capabilities of these tools are very limited and required proficient knowledge of programming to convert data file formats and integrate these tools in a metabolomics workflow. To overcome these

problems, a new platform called MetaboIDent was designed, and a prototype was developed. During the development process of MetaboIDent, several improvements and extensions of functionality were also made to the above-mentioned tools, making transfer of data between these much easier now.

A schematic representation of the design and organization of the software components is shown in Figure 35. MetaboIDent was developed as a standalone platform for reliable automated structure annotation of LC-MS data using both reference spectra from various data sources and *in-silico* predicted spectra. To use the platform, no knowledge of programming is required.

The platform was developed using MATLAB (2022a, Natick, Massachusetts: The MathWorks Inc.), which is a numerical programming language designed for engineers and scientists. One of the goals during designing was the utilization of already existing and highly specialized tools into the platform, which are developed and validated by experts in their respective fields. For some of these tools, no standalone application was available, or the installation was not straightforward, especially when additional third-party libraries and frameworks were required. Since Linux is often used as the underlying operating system for the development of open-source software due to it's free of charge availability, some tools are only compiled for Linux and cannot be executed under Windows, the most common operating system worldwide.

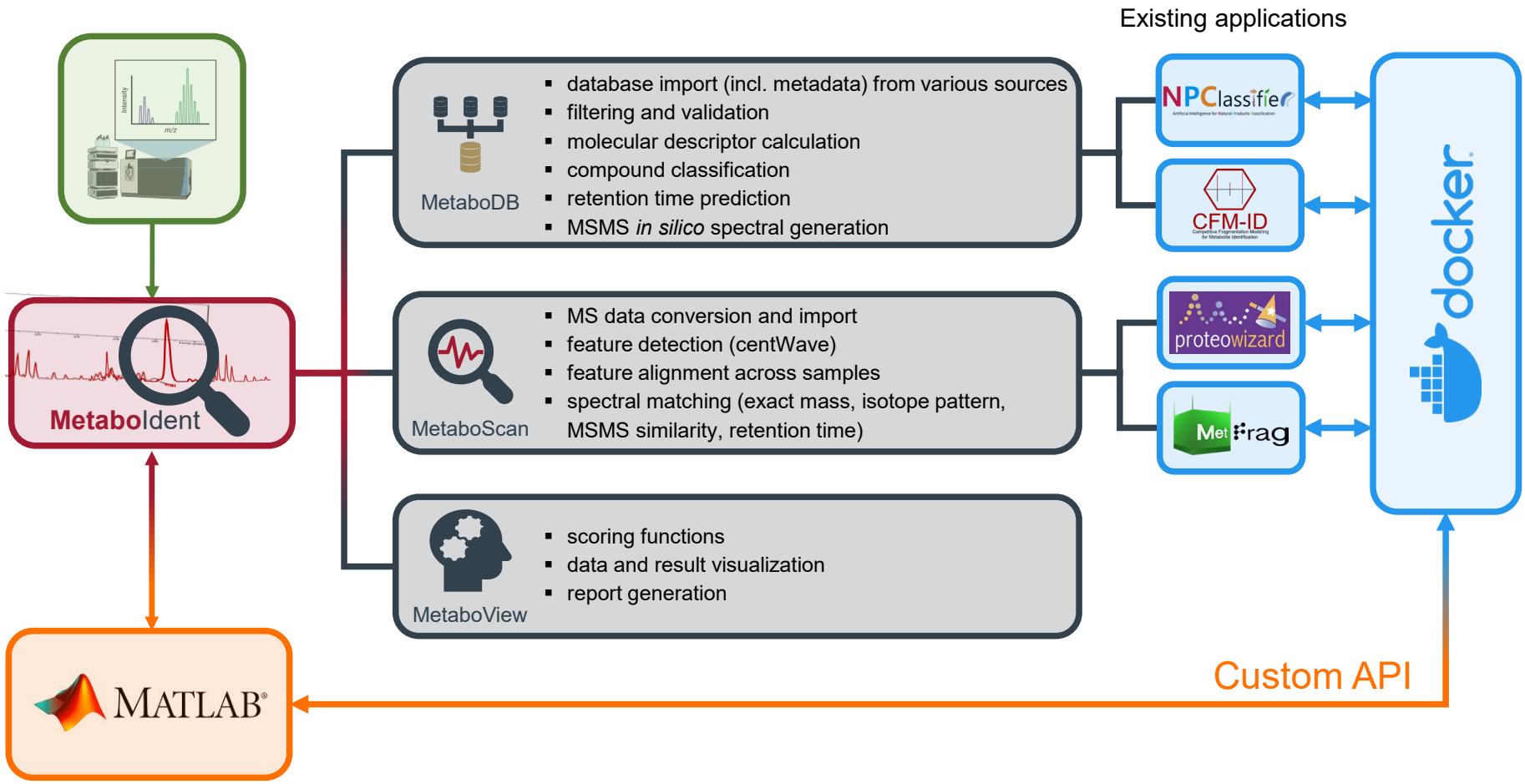


Figure 35: Components and module organization of MetaboIDent, which is divided into three submodules for database creation (MetaboDB), mass spectrometry data processing and matching (MetaboScan) and visualization of results (MetaboView).

To address these problems, the containerizing software Docker[268] was used to integrate the required tools into the platform. With Docker, applications and their dependencies can be packaged into containers by the developers of the tools and shared with the community. Docker must be manually installed once on the computer where MetaboIDent will be executed. This installation is straightforward, well documented and can also be carried out by inexperienced users.

Since MATLAB does not provide a fully operational inbuilt application programming interface (API) to Docker, a custom API was developed. Using this API, MetaboIDent is taking care of the automated downloading, installation, and execution of these tools without any additional user input.

MetaboIDent itself is organized in three independent parts: MetaboDB, MetaboScan and MetaboView. MetaboDB is used to build a custom database which contains compound information such as structural information, molecular properties, compound classification and predicted retention times as well as reference spectra and predicted spectra. MetaboScan is used for LC-MS data conversion, import, feature detection and matching to entries from the MetaboDB database. MetaboView is then used to visualize the results and for a manual inspection and evaluation. The individual components are discussed in the following chapters.

3.2 Data integration into MetaboDB

Within the last decade, several attempts were made to establish a public database source for molecular structure. Clearly, the most prominent example is the PubChem collection[10], a service provided by the United States National Institutes of Health (NIH). It contains information of more than 110 million compounds and metadata such as bioactivities, collected from 867 data sources (state July 2022).[269] Since PubChem stores natural products and molecules of synthetic origin, it is only of limited value when used for natural product research. Dereplication based on PubChem leads to a huge list of candidate molecules. This increases computation time to process and evaluate all candidates and significantly increases false positive annotations.[32,270] Several approaches were developed to identify natural products in these databases based on their structure,[271,272] but the predicted results are heavily dependent on their initial training data sets. Also, since therapeutic drug molecules often mimic natural products, they share a high structural similarity and are often incorrectly classified.[273]

[10] https://pubchem.ncbi.nlm.nih.gov/

A more feasible approach is the utilization of dedicated natural product libraries and databases.[274] Several sources are available, each using a different approach of data collection and curation. A detailed list of the available data sources would go beyond the scope of this thesis; therefore, the focus is maintained on some of the most frequently used databases. All databases listed below can be used as source for the construction of the MetaboDB database, one of the key components of the MetaboIDent platform.

- The freely available COlleCtion of Open Natural ProdUcTs (COCONUT) combines different small natural product collections, some of which are no longer available and applies various filters to remove invalid structures.[275]
- ChEBI is the acronym for Chemical Entities of Biological Interest, an open-access database actively curated by the European Bioinformatics Institute (EBI).[276,277]
- More recently, the freely available naturaL prOducTs occUrrence database (LOTUS) was established with a special focus on structure-organism pairs automatically derived from various other data sources with heavy data harmonization, processing and validation.[278]
- The manually curated and freely available Natural Product Atlas (NPA) contains natural products and taxonomy data for structures isolated either from bacteria or fungi, with crosslinks to primary literature.[279,280]
- The Dictionary of Natural Products (DNP) is a manually curated commercial product, with rich metadata containing UV maxima, source organisms, bioactivity and links to all relevant literature publications.[281]

Chemical structures are stored as textual representations in these databases, which can be read and processed by computers and humans as well. One common format is called International Chemical Identifier (InChI) and consists of multiple layers to store different properties such as atom types and their connectivity and stereochemistry (Figure 36).[282] During the generation of such a string, multiple standardization and normalization rules are applied. Thus, multiple representations of the same molecule will result in the same InChI string, e.g., acetic acid and acetate will both have the same representation (of acetic acid).

The InChIKey is a condensed (hashed) version of the original InChI string, and also consists of a connectivity layer and a stereochemistry layer. Chemical structures can be converted from InChI to InChIKey, but this process cannot be reversed. One of the advantages of InChIKeys is that they can be used to efficiently search for structures in large databases. Since mass spectrometry cannot differentiate between different stereoisomers, a search request to collect potential structural candidates is sufficient based on only the first part of the InChIKey (connectivity layer).

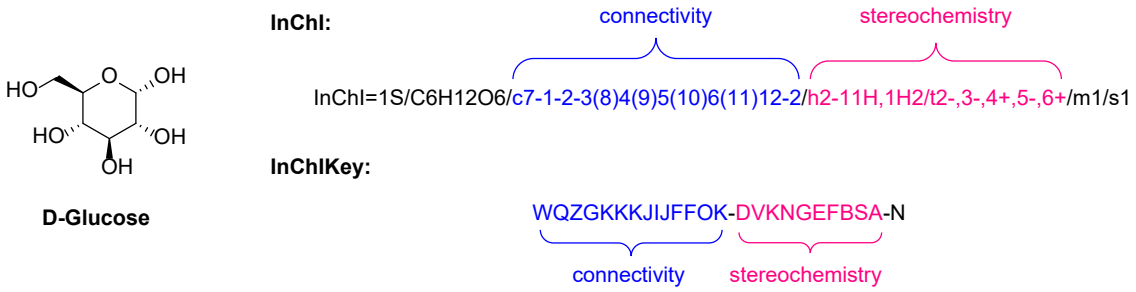


Figure 36: Textual representation of D-glucose as International Chemical Identifier (InChI) and its hashed InChIKey version with the connectivity layer, which contains information about atom types and their connection (blue) and the stereochemistry layer (pink), which adds additional information about the relative position in the three-dimensional space.

The decision to include multiple data sources for the generation of the MetaboDB was based on a detailed analysis of the above-mentioned databases. The results are presented in Figure 37. In total, the top five most frequently used databases for natural product research contain information on almost 1 million molecules (based on unique InChIkeys, Figure 37A) with more than 40 % imported from the COCONUT database, followed by LOTUS with over 27 % and the commercially available DNP with almost 24 %. When solely looking on the connectivity, a reduction to about 830 000 unique compounds was observed. This indicates that about 170 000 entries do not possess a stereochemistry at all. It is highly unlikely for natural products due to their usual structural complexity, so it is more likely that the stereochemistry is yet undetermined and not stored in any database.

When looking on the distribution of unique compounds (Figure 37C), almost 570 000 entries were found to exist only in a single database (in contrast to about 430 000 entries present in multiple databases). This highlights the importance of integrating and accessing multiple data sources for the generation of MetaboDB.

No standardized data format was yet established in the community, so each database comes in its own format and requires the development of custom import routines. Also, small changes in the underlying database structure can easily lead to a loss of functionality and adjustments have to be made in these cases by people with knowledge of programming and software development, which is especially challenging for university with frequent personal changes.

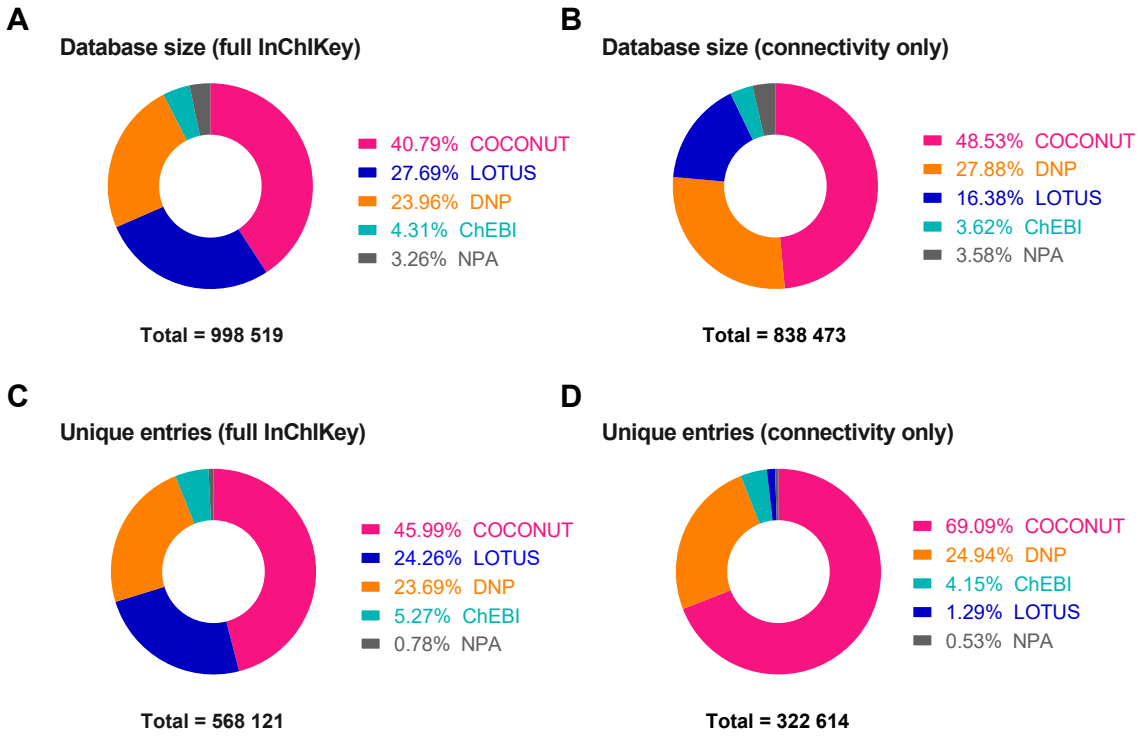


Figure 37: Database comparison based on full InChIKey (left hand side) and the connectivity layer without any encoded stereochemistry (right hand side).

When comparing the proportions of unique entries in LOTUS with (Figure 37C) and without (Figure 37D) stereochemistry, a significant drop from about 24 % to 1 % was observed. This indicates that a significantly higher proportion of the structures stored in LOTUS database have an encoded stereochemistry. This is especially relevant when taking bioactivity into consideration, where different stereochemistry can have a huge impact on biological activity.[283,284]

3.2.1 Processing and classification of metadata

Metadata is often referred to as additional data stored in databases containing chemical structures such as information of origin and bioactivity and can help to improve the annotation quality.[285,286] The Dictionary of Natural Products (DNP) is considered to be one of the richest sources of metadata.[274] These data are provided in a human-readable text format with sentence structure, which cannot be efficiently processed and interpreted as a filter by automated analysis tools such as the MetaboIDent platform. Therefore, the implementation of a prior conversion and classification step was required. Initial attempts using a decision tree to classify the results into six categories, namely cytotoxic, inhibitor, pesticide, antimicrobial, antifungal, and antibacterial failed due to inconsistent data quality. The DNP stands out for the manual curation by experts on the field of natural products. However, humans make inconsistent errors, leading to problems when using such a rule-based approach. For example, some entries contained phrases with repeated words or additional sentence marks. Also, different phrases were used to express the same meaning.

The constructed rule-based decision tree became far too complex to address all these specific cases. As a solution, a deep learning model for automated text classification (long short-term memory network, LSTM) was trained based on one thousand manually classified strings. The performance was investigated using a different testing data set with another one thousand manually classified strings. The training and testing phases were repeated 6 times, each phase was independent of the performance of the prior phases. The averaged results are presented in Figure 38.

No correction was done by transfer of prediction results to other classes. More precisely, antifungal, and antibacterial compounds are also antimicrobial, but compounds which were correctly classified as antifungal or antibacterial were not automatically classified as antimicrobial.

Overall, an average prediction accuracy of more than 90 % was achieved, with the highest true positive rates (about 97 %) for the classification of strings containing information about cytotoxicity (Figure 38A). No correlation of performance was observed when profiling the true positive rates of all categories across multiple repetitions (Figure 38B). This indicates that prediction outcomes are independent of each other. The false positive rates are significantly higher (1 – 10 %) compared to the false negative rates (1 – 3 %). This result is satisfactory, especially when the classification should be used as a filter for compound candidates. Here, low false negative rates mean that only a very small proportion of compounds are missed when applying filters based on the predicted bioactivity.

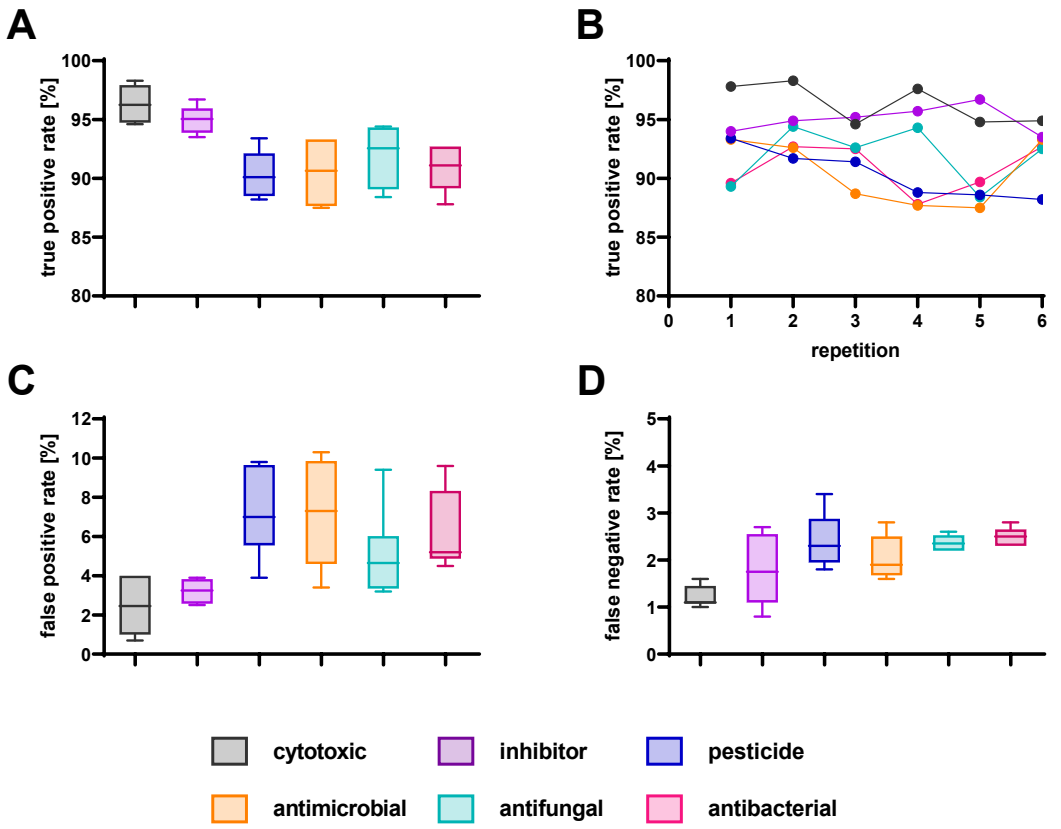


Figure 38: Classification of bioactivity metadata from the Dictionary of Natural Products (DNP) using machine-learning. Displayed are the prediction results of 6 independent training sessions. (A) true positive rates, (B) correlation of categories across repetitions, (C) false positive rates, (D) false negative rates. Classification categories are color-coded as follows: cytotoxic is presented in grey, inhibitors in purple, pesticides in blue, antimicrobial in orange, antifungal in turquoise and antibacterial in pink. Error bars present the min/max values.

3.2.2 Retention time prediction modelling

Metabolite identification can be performed at different levels e.g., accurate mass, predicted sum formula matching based on isotope pattern analysis or spectral matching of tandem mass spectra. However, identification of compounds solely based on accurate mass, even if high resolution, is not feasible. A search in MetaboDB (compiled as discussed in 3.2 – Data integration into MetaboDB) with an accurate mass of 318.2175 ± 0.001 Da leads to more than 1500 hits. Similarly, searching for the sum formula $C_6H_5N_3O$ will return over 1700 hits.

Even though spectral matching is ranked second by the Metabolomics Standard Initiative (MSI) in their confidence annotation level system,[287] different metabolites can show

similar or almost identical fragmentation pattern, leading to many incorrect identifications. To boost the accuracy of identification, retention time (RT) was identified as a key parameter to filter candidate lists for spectral matching. Multiple problems arise when trying to utilize retention time (RT). Firstly, the measured retention time of a compound is highly dependent on the instrument type and setup used for the analysis. Solvent, temperature, column type and flow rate, just to name a few, have a huge impact on retention times and comparable retention times can only be obtained when using the exact same setup with identical parameters. In addition, the number of commercially available standards for metabolites is very limited, and they are often very expensive.
To address this issue, several methods have been developed to predict the retention time based on molecular descriptors or chemical fingerprints.[288] Molecular descriptors are related to structural properties such as the number of hydrogen bonds in a molecule, or the number of oxygen atoms and can be calculated based on the chemical structure alone, without the need to synthesize or buy this compound. Nevertheless, the retention time prediction model must be calibrated, which requires at least a couple of hundred representative reference compounds and their retention times measured under identical conditions.

To implement retention time prediction for the MetaboIDent platform, a library containing retention times from more than 500 unique compounds was compiled. Most of the reference substances were kindly provided by Prof. Dr. Fiedler from the Interfaculty Institute of Microbiology and Infection Medicine Tübingen (IMIT). In addition, a natural product compound set was purchased from AdipoGen Life Science/Bioviotica. The initial lack of very unipolar compounds was resolved by including data from a fatty acid methyl ester (FAME) standard mixture, which was purchased from Sigma-Aldrich. Some natural products which were isolated during the time of this thesis in the research group of Prof. Dr. Stephanie Grond were also included. A list of proportions can be found in Figure 47A. All compounds were measured by HPLC-MS under identical conditions, with standard runs in between to monitor column performance over time. In case of significantly reduced performance, the column was either thoroughly purged or replaced. Data were manually curated and information such as structural data (InChI and InChIKey) added. A graphical representation of the retention time distribution for the compiled reference set is presented in Figure 47B.

For a robust and reliable prediction of retention times, it is highly important to construct a training set which covers a broad spectrum of different structures. Lots of very similar structures in such a training set will lead to an overfitting of the models, causing a significant drop in prediction accuracy for compounds which have dissimilar structures which are not represented in the training data set. To measure the structural similarity

between two structures, the most frequently applied method is the Tanimoto similarity score.[289,290] A low similarity score indicates that the two compounds have less structural motifs in common, while compounds with a score above 0.85 are usually considered similar. It is important to point out that a high similarity score does not automatically imply related bioactivities.[291] The Tanimoto similarity score was calculated using the Chemical Development Kit (CDK), an open-source modular Java libraries for Cheminformatics.[292,293] The obtained score density plot is displayed in Figure 47C.

A total of 159 distinctive molecular descriptors were calculated for each compound of the reference set, using the CDK. Noisy data such as parameters with insignificant standard deviations or low correlation coefficients can reduce the performance of prediction. Therefore, a reduced subset of 117 molecular descriptors was used in a second training round and performance compared to the full set was investigated. In addition, all compounds were classified using NPClassifier, a deep neural classification tool for natural products.[294] It is common knowledge, that compounds that belong to a particular class have retention times different from other classes, e.g., sugars are usually considered to be highly polar and elute very early on reversed-phase columns, while fatty acids are very unpolar and will elute late. Consequently, the incorporation of such information into the list of parameters for retention time prediction should increase the prediction performance. The predicted compound classes are displayed in Figure 47D.

Five different models, namely random forest (RF), neural network regression (NNR), multiple linear regression (MLR) and support vector machine (SVM) have been shown to return satisfying results for retention time prediction[295-298] and were evaluated towards their prediction accuracy for MetaboIDent. Machine learning models often show excellent performance when predicting retention times of compounds which were part of the initial training set. To address this issue and to validate the true prediction performance of unknown compounds, the mentioned models were trained on only 80 % of the full reference dataset (427) using 10-fold cross validation, while the other 20 % served as testing dataset (107). The splitting of training and testing dataset was random.

The prediction outcomes for the 5 models with full and reduced parameter set are displayed in Figure 39. For comparison, a reference value was calculated based on the difference of the measured retention time and center of the retention time window of all compounds (see chapter 5.11 – Database access and processing for detailed information). For all models, the predicted retention time is significantly lower than the reference retention time. This is clear evidence, that there is indeed a correlation between retention times and molecular structure. This connection was questioned by ZUBEIL in his thesis, when looking at correlations between single properties such as the

octanol-water partition coefficient (logP) and retention times.[299] Clearly, the underlying complex physical processes in column chromatography cannot be reduced to a handful of physical parameters, but rather require a large set of parameters with complex correlation pattern.

The training on reduced parameter set does not have a significant impact on the prediction outcome. For the neural network regression (NNR) model, the subset reduces the overall prediction error range, in contrast to the multiple linear regression (MLR) model, where the prediction error range is increases for the reduced subset.

Looking at the prediction error distribution over the full span of the retention time (Figure 48), the MLR and the SVM model have a significant performance drop for highly polar compounds which are not retained by the column used for data collection. This has been reported in literature before,[300] and exclusion of such compounds is recommended. The RF and NMR models do not overfit these compounds, and a splitting into smaller training datasets would most likely result in higher prediction errors. Therefore, the polar compounds were not excluded from the training datasets.

Overall, the best performance was achieved by the random forest (RF) model with a root-mean-square-error (RSME) of 114 ± 5 seconds, while the average RSME error of all models was 118 ± 6 seconds.

In addition, the enhancement of a consensus value derived from the predicted values of all models was investigated. The consensus value was calculated using a k-means clustering algorithm based on the Euclidean distance between all predicted values. It aims to minimize the variance of data points within a cluster and requires a maximum number of clusters as input. There is currently no general mathematical procedure available to determine the ideal number of clusters. Thus, this number has to be manually determined by investigation of datasets. It was found that a maximum cluster size of three seemed to be most appropriate, so that outliers above and below the center can be classified into two separate clusters, while one remains for the consensus values of the models. The returned RSME error for the consensus model was 129 ± 8 seconds. This indicates that a systemic prediction error is present and that either all models perform well, or all models fail to predict accurate values for this particular compound. This is most likely due to the limited number of reference compounds compared to the wide range of structural natural product diversity which has to be covered.

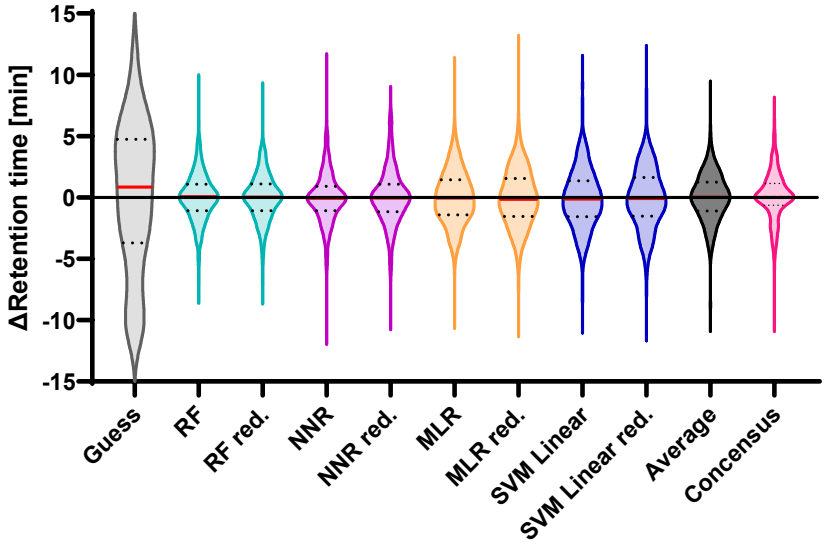


Figure 39: Violine plots for retention time prediction errors by four machine learning models. For training and testing, a total of 534 unique compounds was used, which were divided 80/20 into training and testing set. Red lines indicate the median value, dashed black lines the quartiles. The random guess value (displayed in grey) serves as a reference. The average of all models is displayed in dark grey, and a consensus model based on the prediction outcomes using a clustering algorithm is displayed in pink.

3.3 Data import and processing with MetaboScan

MetaboScan is used to perform the core analysis of the LC-MS data. The conversion of various data formats from different vendors, including Thermo Scientific, Bruker or Agilent is automatically done by msConvert,[301,302] which is integrated via Docker. In addition, common open data formats as mzXML[303] and mzML[304] can be directly imported. After import was completed, fragmentation spectra were filtered by their spectral entropy,[305] a measurement, to identify spectra with low information count due to low fragmentation of the selected precursor ion.
Since compounds can only be identified when a fragmentation spectrum is available, mass spectra were further filtered and signals which were not selected as precursor for fragmentation by the MS instrument during acquisition are discarded.

Features are defined as a selection of intensity values from similar m/z values within a certain retention time window. Consequently, one chromatographic peak can be composed of multiple features related to many co-eluting or just a single compound. Even when a chromatographic peak is derived only from a single compound, the formation of adducts and the presence of multiple charged signals in addition to their corresponding isotopic signals will result in multiple features.[306] In MetaboScan, these features are detected using a modified version of the powerful centWave algorithm.[307]

This algorithm was developed for high resolution mass spectrometry and applies continues wavelet transformation for feature detection.[308] In consequence, MetaboScan performs best for high resolution data from orbitraps or time-of-flight (TOF) instruments while data from quadrupole or ion trap instruments are usually considered low resolution and are not suitable for analysis by MetaboScan. Since chromatographic peaks are detected on wavelet curves by the algorithm, the resulting boundaries are often not correctly identified. This problem has been addressed and solved by an additionally added local minima search using the original boundary value as starting point in MetaboScan. At the end of the feature detection step, each feature is linked to a fragmentation spectrum and features without a corresponding spectrum are discarded.

In the next step, if multiple replicates of samples and references are selected for analysis, features are aligned across all analyses of the same group (sample or reference) using retention time windows, precursor mass tolerance and fragmentation spectra similarity. Spectral similarity can be calculated using either cosine-similarity[309] or spectral entropy similarity, depending on users choice.[305] Features below a defined occurrence level within the replicates are discarded. This significantly reduces the number of artifacts and fluctuation due to biological variance in a dataset. Features, which are also present in the reference group (for example in media blanks or wild types strains when compared to mutants) are discarded based on the same parameters described above.

Depending on the structure, not only singly charged, but also multiply charged signals can be detected by mass spectrometry. These signals can cause significant problems when remained unidentified, as generation of candidate lists using the incorrect exact masses of an adduct will result in increased false positive annotation rates. Also, the fragmentation spectra acquired from multiple charged ions can significantly differ from ones with single charged precursors and both, MetFrag and CFM-ID, are only able to predict and annotate single charged ions. The charge state of features can be determined by measuring the mass difference between isotopic peaks, which is 1.003 Da for singly charged ions and changes by 1/z with z being the charge of the ion. Even with a chromatograph separation system, mixed mass spectra of multiple compounds are obtained with overlapping and interfering signals. This makes an annotation based on only a single mass spectrum highly error-prone. Several algorithms are available, but they are usually developed for proteomics and work best for peptides with predictable fragmentation pattern.[310-313]

Therefore, the charge state was determined using a custom developed sophisticated algorithm. In a first step, a list with exact masses of all potential ^{13}C isotopic signals derived from a parent signal for charge states ranging from 1 to 3 were generated. Based

on the created list, putative isotopic signals were then searched within a certain m/z-window. Intensities of all putative signals were collected within the retention time range of the parent signal and independently normalized to the maximum observed intensity. Saturated peak regions were cropped, and the area for all putative isotopic peaks were calculated. In addition, a procruste analysis was performed to determine the similarity between the isotopic peak intensity curves and the parent intensity curve.[314] If putative isotopic peaks are indeed related to the parent peak, the areas, and the shape of the curves have to be almost identical.
To avoid the loss of features which were incorrectly identified as multiply charged, it was decided that charge annotation is performed, but no features are automatically removed based on the results. This has to be done by the user during the result inspection phase.

Ions can also generate different types of adducts. The most frequently observed adduct type is $[M+H]^+$ for positive mode and $[M-H]^-$ in negative ionization mode. Other adduct types are generated by the agglomeration of metal ions such as sodium or potassium and by the addition or cleavage of small stable molecules such as carbon dioxide, water, or ammonia. Once again, the fragmentation spectra can be significantly different between adduct ions. In case of CFM-ID, only sufficient reference fragmentation spectra for $[M+H]^+$ and $[M-H]^-$ were available for model training. Consequently, only the adducts $[M+H]^+$ and $[M-H]^-$ are supported by CFM-ID. Intensity profiles of adducts derived from the same compound are significantly different and can therefore not be used in an approach similar to the one described above.[310]
For annotation of adducts in MetaboScan, a graph-based approach has been implemented, similar to a recently described method.[310]
In a first step, an averaged spectrum including all scans within the retention time boundaries of the detected feature was generated, and a mass difference matrix of all peaks is calculated. A list of adduct masses was generated for each signal within the averaged spectrum, and the corresponding adducts were identified based on a mass difference matrix between all signals. Identified masses were represented as nodes, while edges encoded the adduct type (more precisely, the mass difference between the two nodes). Edge weights represented the explained intensity by the adduct. The $[M+H]^+$ or the $[M-H]^-$ peak was then identified as the peak with the most outgoing edges. In case of a draw, the sum of explained intensities was used as a second criterion.
Once charge state and adduct type for each feature were annotated, the feature was matched against a list of candidates from the MetaboDB database with exact mass as search criteria. Retention time (measured or predicted), isotope pattern matching, and spectral similarity scores were used to score the candidates. Results are then stored for further manual analysis and interpretation with MetaboView.

4. CONCLUSION

The presented work aimed at the investigation and profiling of biological activities for the natural product collinolactone, precursors and metabolites thereof and semisynthetic derivatives. Derivatization and modification of natural products to conduct structure relationship studies and mode of action investigations require a significant amount (several hundred milligrams) of the natural product as an educt. Since no total synthesis was yet established, the educt had to be biotechnologically generated. Therefore, the biosynthesis has been investigated by stable-isotope precursor feeding of common polyketide synthase building blocks. These findings were then used to select putative gene cluster candidates in *Streptomyces sp.* Gö 40/10 and deletion mutants were prepared for each gene cluster candidate using CRISPR-BEST or classical suicide vectors in collaboration with Tilmann Weber from the DTU Denmark. The collinolactone (**1**) encoding gene cluster *col* was identified and a high-production mutant strain was generated by cloning a strong promotor KasOp* into the wild type strain, resulting in a 5-fold increased production rate compared to the wild type strain.
In addition, cultivation and fermentation conditions have been investigated and optimized for the wild type strain. The fermentation is now conducted within a newly designed media called JCS with the addition of amberlite XAD-16 resin and by biosynthetic precursor feeding. A straightforward, robust, and efficient workflow was developed with a significant reduction of manual and work intense steps. Applying optimized fermentation and cultivation conditions and following the new workflow increased and stabilized the production rate of collinolactone (**1**) to more than 25 mg/L.
The genetically engineered overproducer strain *Streptomyces sp.* Gö $40/10^{HP}$ and the optimized fermentation and isolation procedure are currently part of a patent application process (J. C. Schmid, J. F. G. Garzón, A. Zeeck, T. Weber, S. C. Grond, Germany, 2021, Patent-Number: DE 10 2021 003 944.0.).

During the course of this research, a compound called rhizolutin (**1R**) was published by *Kwon et al*. and a structure was proposed that shared significant similarities with compound **1**.[157] The structure elucidation of rhizolutin was painstakingly investigated and compared to collinolactone (**1**). Based on a detailed NMR analysis, strong evidence was found that the elucidation of rhizolutin was not unambiguous and led to the wrong stereo-configuration and that collinolactone (**1**) is indeed the correct structure.[2]

A structural diverse and comprehensive set of derivatives was synthesized by a variety of different approaches, including reactions such as esterification via Steglich reaction, dehydrogenation by the facilitation of burgess reagent, hydrogenation via Strykers reagent or methylation by Gilman cuprates in scales of 2-10 milligrams.

All compounds have been investigated towards their cytotoxicity in a cell-viability assay on L929 cell line. No toxicity was found for collinolactone (**1**) but the derivatives collinolactenone (**7**) and acetyl-collinolactone (**5**) showed an effect on cell viability with EC_{50} values of 36 µM and 106 µM, respectively. The mode of action of the natural product and derivatives was further investigated by fluorescence microscopy on PtK2 cells. Compounds **5** and **7** showed a significant increase of cells with monopolar spindle formation, which inhibits further cell division and ultimately leads to apoptosis (cell death). Molecules, that interfere with the cell cycle are of high interest for anti-cancer drugs and therapies. Two molecular targets, namely the motor proteins kinesin Eg5 and dynein, were identified as putative molecular targets which cause the formation of monopolar spindles when inhibited. Consequently, the inhibitory effects of the derivatives on these putative molecular targets were investigated in a malachite green ATP based assay. Collinolactone (**1**) did not show any effect while compounds **5** and **7** as well as compound **6** showed a significant decrease for 10 and 100 µM of the Eg5 kinesin activity. No effect was observed for the dynein motor protein.
The selective inhibitory effect makes them highly attractive lead structures for further development and optimization towards an anti-cancer drug. Until now, no drug was approved that targets the Eg5 motor protein.

Based on the initial findings by Kwon *et al.*, the neuroprotective properties of the natural product **1** and derivatives **2**-**12** were studied in a glutamate-induced intramolecular oxidative stress assay on murine hippocampal neuronal cell line HT22 cell in collaboration with Michael Decker from the university of Würzburg. Interestingly, only the natural product collinolactone (**1**) showed a neuroprotective effect, starting at about 5 µM concentration, while all other derivatives had no effect at all. This indicates that the activity of compound **1** is highly specific, and even the smallest structural changes lead to a loss of activity.
Since neurodegenerative diseases are on the rise, novel therapeutic drugs to effectively treat these diseases are urgently needed. The positive effect of collinolactone (**1**) in Alzheimer mouse models, combined with its discovered neuroprotective properties as part of this thesis, makes it a potential compound for further development towards a drug candidate.
In addition, all compounds were profiled in an amyloid beta (Aβ) aggregation and disaggregation assay, similar to the procedure described by Kwon *et al.*[157] Only collinolactone (**1**) showed a decrease of aggregated protein aggregates, but only at high concentrations of 1 mM which is usually far from any biologically relevant concentrations.

The presented work also aimed at the development of a standalone application for automated annotation of LC-MS analyses, with special focus on easy installation and usage without deep knowledge of computers or any programming skills. As a result, a prototype called MetaboIDent was developed using MATLAB (2022a, Natick, Massachusetts: The MathWorks Inc.). The platform is structured into three major segments: MetaboDB, MetaboScan and MetaboView. MetaboDB is used to import, filter and store compounds and measured reference spectra while MetaboScan is used to import LC-MS analyses, perform feature detection using an optimized version of the centWave algorithm and spectral matching. MetaboView is then used to visually and modify the results.

Due to the fragmented availability of natural product database and collections, an integration routine was developed for multiple of the most popular databases, including the COlleCtion of Open Natural ProdUcTs (COCONUT) and the commercially available Dictionary of Natural Products (DNP).

A key feature is the retention time prediction of compounds based on a small set of reference substances, using a variety of deep learning models which have been shown to be suitable for modelling of such complex physical processes, including support vector machines (SVM) or random forest (RF) models.

The platform facilitates the use of the *in-silico* fragmentation tools MetFrag[27,28] and CFM-ID[29-31] by simple integration via the containerizing software Docker, which automatically collects all required additional libraries for execution of external applications. No additional time and resources must be invested by the user to integrate these tools into the platform. This will hopefully contribute to and boost the outcome of the laboratories research.

5. MATERIALS AND METHODS

5.1 Instrumental analytics

Ultrapure water was generated by a Millipore Milli-Q (Reference A+) processing plant. Solvents were degassed in line. Solvents (methanol and acetonitrile) were of HPLC-grade.

5.1.1 (Semi)-preparative HPLC

Thermo Scientific Ultimate 3000 with HPG 3200 BX pump, Knauer online degasser and DAD-3000RS detector with a semipreparative detection cell, operated by Chromeleon® version 6.80. A manual Rheodyne valve with 200 – 1000 µL loop for sample injection was used.

5.1.2 Medium-Performance-Liquid chromatography (MPLC)

Purification of crude extracts was performed on a Varian IntelliFlash 971-FP (Flash Purification System) with control software version 3.1.524. Prepacked Chromabond® Flash RS 25 SiOH (15 - 40 µm) cartridges from Machery-Nagel GmbH & Co. KG were used. Technical grade or analytical grade solvents were used when available.
For injection, samples dissolved and mixed with small amounts of silica gel and then evaporated to dryness. The residue was transferred into an empty cartridge replacing the proprietary DASi module and automatically injected onto the column. Column equilibration and system flushing were performed manually before installing the sample cartridge.

5.1.3 HPLC-ESI-MS

High resolution mass spectrometry data were acquired on a Bruker MaXis 4G ESI-QTOF instrument coupled to a Dionex Ultimate 3000 HPLC system (Thermo Fisher Scientific). Each analysis was calibrated using sodium formiate as internal calibrant.
The nebulizer pressure of the ESI source was set to 2.0 bar and dry gas was set to 8.0 L min^{-1} at an operating temperature of 200 °C. For MS/MS spectra acquisition, auto MS/MS mode with collision energy stepping was used. The routine gradient was 10% to 100% methanol in 20 min with a flow rate of 0.3 mL/min on a Nucleoshell® EC RP-C_{18} (150 x 2 mm, 2.7 µm) from Macherey-Nagel.

Sample preparation was performed as follows: for pure compounds, a concentration of 0.05 mg/mL in methanol or acetonitrile was prepared by serial dilution. For extracts, the concentration was adjusted to of 3 mg/mL with methanol or acetonitrile. Samples were then centrifuged and 50 µL of the particle free solution transferred into a HPLC vial.

5.1.4 NMR

NMR spectra were recorded on a Bruker Avance III HDX 700 instrument at 298 K, if not stated otherwise. Chemical shifts are given as δ-values (ppm) relative to the solvent as internal standard. Coupling constants (J) are given in Hertz (Hz). Abbreviations for multiplicity description are as follows: s: singlet, d: duplet, t: triplet, q: quartet, p: quintet, dd: duplet of duplet, m: multiplet.

5.2 Thin-layer chromatography

TLC silica gel 60 F254 (normal phase, 0,2 mm layer thickness) and TLC silica gel 60 RP-18 (reversed phase, 0,2 mm layer thickness) chromatography plates were purchased from Merck KGaA.

For staining, a freshly prepared solution of Anisaldehyde was used: Stock solution (500 mL) contained the following: 50 mL acetic acid, 25 mL sulfuric acid and 425 mL Methanol. 1 mL 4-methoxybenzaldehyde (from Sigma-Aldrich) was added to 50 mL of stock solution for staining.

5.3 Other laboratory equipment

Autoclave:	Thermo Scientific Varioklav 75S.
Balance:	Sartorius CP26P, Sartorius Entris, Sartorius BP2215.
Shaker:	Eppendorf Thermomixer Compact.
Centrifuge:	Eppendorf Centrifuge 5424 and Sigma 4-16k.
SpeedVac:	Eppendorf Concentrator Plus.
Microplate reader:	Tecan Life Sciences Infinite M Plex 200 Pro.
Laminar flow box:	Thermo Scientific MSC-ADVANTAGE.
Incubator:	Braun Biotec International Certomat BS-1
Sonicator:	Branson Sonifier 250.
Lyophilizer:	Christ Beta 2-8 LD Plus with Vacuubrand Chemistry Hybrid Pump RC6.

5.4 Strains and culture conditions

All microbiological preparations were carried out under sterile conditions using a laminar flow box. Sterilization of media and culture flasks was carried out in an autoclave at 121 °C and 1 bar overpressure for 20 min.
50 µg/mL of antibiotics to media was added when required (for *Streptomyces sp.* Gö 40/10 KasOp* LuxR and deletion mutants for BGC identification).

5.4.1 Soy flour mannitol media (SM media)

20 g/L of fat reduced soy fluor meal (Henselwerk GmbH) and 20 g/L of D-mannitol (AppliChem GmbH) were dissolved in 1 L of hot tap water. After complete dissolution, pH value was adjusted to pH = 7.0 by addition of 1 N sodium hydroxide solution.
For the preparation of agar plates with SM media, 20 g/L Agar-Agar (Carl Roth GmbH) was added prior to sterilization.

5.4.2 Modified SM media (JCS media)

20 g/L of fat reduced soy fluor meal (Henselwerk GmbH), 20 g/L of D-mannitol (AppliChem GmbH), 20 g/L soluble starch from potato (Carl Roth GmbH), 10 g/L glucose (Sigma Aldrich), 5 g/L yeast extract (Carl Roth GmbH) and 1 g/L calcium carbonate (Carl Roth GmbH) were dissolved in 1 L of hot tap water. After complete dissolution, pH value

was adjusted to pH = 7.0 by addition of 1 N sodium hydroxide solution.
For the preparation of agar plates with JCS media, 20 g/L Agar-Agar (Carl Roth GmbH) was added prior to sterilization.

5.4.3 Preparation of spore suspensions

For long term storage, cryo stocks of spores were prepared for the strains, following a procedure described in literature.[201] The strains were grown on agar plates (SM media) until the whole plate was covered by mycelium. (5 to 7 days at 28 °C depending on the type of strain). 3 mL of sterile water was added to the plate and a sterile loop was used to carefully scrape the plate surface. To remove any tiny agar plates, the spore containing liquid was transferred into a syringe with sterilized cotton at the bottom. Using the syringe plunge, the solution was filtered through the cotton and the filtrate was collected in a 50 mL falcon tube. This process can be repeated multiple times for the same or multiple agar plates of the same strain depending on the required number of stocks.
To break the spore chains, the tube was vigorously vortexed for 3 minutes, and then centrifuged at 2000 × g for 10 minutes. The supernatant was discarded, and the spore pellet was dissolved in 25 % sterile glycerol (1 mL per harvested agar plate; depending on the desired spore concentration the volume of glycerol solution can be adjusted accordingly). The solution is vortexed once again to homogenize which is then transferred into 1.5 mL crew top vials and immediately frozen at -80 °C.

5.4.4 Colony-forming-unit testing

In order to obtain reproducible results and to compare the production and growth rates of various strains, an equal number of spores should be used for inoculation. Therefore, the number of spores per milliliter of the cryo stocks has to be determined. Following the described procedure in chapter 5.4.3 – Preparation of spore suspensions, the estimated number of spores per milliliter lies within a magnitude of 10^6 to 10^9. Therefore, multiple serial dilution is required for the colony-forming unit testing: 10 µL of spore solution was dissolved in 1 mL of sterilized water and diluted 5 more times by 1:10. 100 µL of the last serial dilution step was used to inoculate an agar plate which was then incubated for 3 to 4 days at 28 °C. The number of formed colonies is then determined by manual counting. For a more reliable result, this experiment was performed three times for each cryo stock.

5.4.5 Preparation of Amberlite XAD-16

Before addition to the media, Amberlite XAD-16 was thoroughly washed with 1 N HCl in methanol, followed by acetone, 1 N aqueous sodium hydroxide and finally deionized water until the pH value of the washing solution became neutral again. The clean Amberlite XAD-16 was kept damp throughout the whole process and was used immediately after preparation to prevent microbial growth.

5.5 Production curves of collinolactone (**1**) and cineromycine (**3**)

The experiment was conducted in 300 mL flasks with baffled bottom. Each flask contained 100 mL of either SM-Media or JCS-Media (prepared as described in chapters 5.4.1 and 5.4.2). In addition, 15 g per flask of Amberlite XAD-16 was added to one half of the flasks with SM and JCS- media.
Inoculation with the wild type strain *Streptomyces sp.* Gö 40/10 or the overproducer strain *Streptomyes sp.* Gö 40/10 KasOp* LuxR was done from spore cryo stocks with 10.000 spores per flask. Incubation was performed at 28 °C and 180 rpm for 6 days. Samples (1 mL) were taken under sterile conditions every 24 hours and directly stored at -80 °C until further batch processing.
To check whether the overproducer has lost its capabilities, samples of *Streptomyes sp.* Gö 40/10 KasOp* LuxR containing flasks were plated on agar plates containing hygromycin B, which addresses the resistance gene used to engineer the strain. The strain maintained its resistance across the whole experiment.

Collected samples were dried using a lyophilizer, followed by extraction with 1 mL of acetonitrile/water 9:1 overnight. Samples were centrifuged and 50 µL of the supernatant transferred into HPLC vials for LC-MS analysis.
A custom method was programmed in Brukers DataAnalysis 4.2 for automated data processing. First, samples were checked for their calibration status, and recalibration was performed when required. In a second step, EICs were generated using a user provided list of retention time ranges and exact masses of interest. Third, all EICs were automatically integrated using Bruker's inbuilt integration tool and the results exported to csv.
MATLAB (2022a, Natick, Massachusetts: The MathWorks Inc.) was used to import the obtained csv files and to match the collected data across all samples.

5.6 Fermentation of *Streptomyces sp.* Gö 40/10

The fermentation procedure as well as the stable isotope feeding experiments are already described in J. C. Schmid *et al.*, *Angew. Chem. Int. Ed.* **2021**, *60*, 23212.

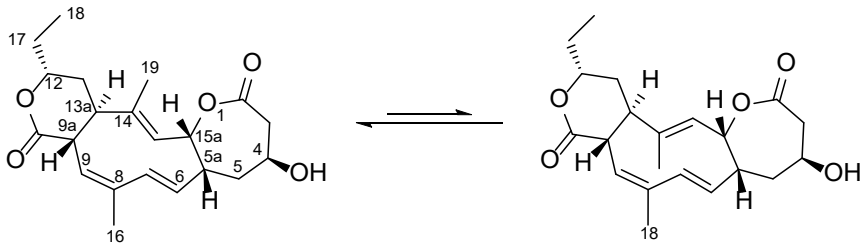
5.6.1 Production of collinolactone (**1**)

The production of collinolactone (**1**) was done using 1 L shaking flasks with baffled bottom. 150 mL of SM or JCS-media was added to each flask, followed by 15 g per flask of Amberlite XAD-16 resin before sterilization. The production culture was inoculated with 10 mL per shaking flask of a 48-hour old culture grown in soy flour mannitol media (SM). The fermentation was run for 72 hours at 28 °C and 200 rpm. To further increase the production of 1, a sterile-filtered solution of the biosynthetic precursors sodium propionate and malonic acid (each 500 mg/L culture broth) with the pH adjusted by addition of sodium hydroxide to 7.0 was added to the culture 8 hours after the fermentation started.

After 3 days, the viridescent colored culture broth was centrifuged to separate the bacteria and the Amberlite XAD-16 from the supernatant. Due to the quantitative absorbing of 1 to the absorbing resin, the supernatant did not contain any product with was confirmed by TLC) and was discarded.
The pellet, together with the resin was then extracted three times with 2 L of acetone while stirring. The solid material was separated by filtration through a Büchner funnel, and the solvent of the combined organic extracts was removed *in vacuo* using a rotary evaporator with water bath temperature set to 40 °C. The obtained viscous black residue was dissolved in 300 mL of water and extracted three times with 200 mL of toluene. The solvent was removed *in vacuo* again. The oily residue was purified by chromatography over silica gel (2 L of cyclohexane/acetone 3:2, 200 g silica gel, 6 cm column diameter). The fractions containing 1 were identified by TLC (RP-18, methanol/water 7:3, Rf = 0.4) and concentrated *in vacuo*. The enriched extract was further purified by flash chromatography (A: diisopropylether, B: acetone, 25 g silica gel, column dimensions: 13 x 2.5 cm, 20 mL/min, fraction size 20 mL) using the following gradient: 5 min 2% B, 10 min 8 % B, 20 min 8 % B, 25 min 16 % B, 35 min 16 % B, 40 min 20 % B, 45 min 20 % B. The crude product was finally purified by preparative HPLC (Kromasil 100 $C_{18,}$ 250 x 21.2 mm, 7 µM particle size, 100 Å pore size) with 34 % acetonitrile/water at 17.5 mL/min.

The pure compound 1 eluted at R_t = 25 - 27 min. Fractions containing the product were lyophilized to give a white, fluffy solid. The obtained yields varied between 15 - 30 mg/L. The biosynthetic intermediate collinoketone (**2**) was isolated using the same purification workflow (R_t = 19.5 - 22 min) in yields of 0.3 mg/L.

5.6.2 Characterization of collinolactone (**1**)



1 (anti) **1** (syn)

HR-ESI-MS (m/z): $[M+H]^+$ calculated for $C_{21}H_{28}O_5$: 361.2010; found 361.2010 (1.6 ppm err, 25.6 mSigma)

anti-form:

^{1}H NMR (700 MHz, CD_2Cl_2) δ = 1.01 (t, *J* = 7.6, 4H, 18-H_3), 1.60 – 1.66 (m, 1H, 17-H_b), 1.69 (d, *J* = 1.2, 3H, 19-H_3), 1.72 (q, *J* = 6.9, 1H, 17-H_a), 1.76 (dt, *J* = 3.3, 11.7, 1H, 5-H_b), 1.77 – 1.83 (m, 1H, 13-H_a), 1.91 (s, 3H, 16-H_3), 1.89 – 1.94 (m, 2H, 13-H_2), 2.33 (dt, *J* = 4.9, 15.1, 1H, 5-H_a), 2.63 (td, *J* = 3.0, 8.5, 1H, 13a-H), 2.72 (br, 1H, 20-OH), 2.85 – 2.93 (m, 1H, 3-H_b), 2.92 (dd, *J* = 4.7, 14.7, 1H, 5a-H), 3.14 (dd, *J* = 9.8, 15.0, 1H, 3-H_a), 3.28 (t, *J* = 8.0, 1H, 9a-H), 4.37 (d, *J* = 1030.7, 1H, 12-H), 4.41 – 4.50 (m, 1H, 4-H), 4.83 (d, *J* = 6.8, 1H, 15-H), 5.14 (dd, *J* = 10.2, 16.4, 1H, 6-H), 5.63 (t, *J* = 6.3, 1H, 15a-H), 5.74 (dt, *J* = 1.5, 7.6, 1H, 9-H), 5.89 (d, *J* = 16.4, 1H, 7-H).

^{13}C NMR (176 MHz, CD_2Cl_2) δ = 9.7 (C-18), 13.2 (C-19), 19.9 (C-16), 28.4 (C-17), 32.5 (C-13), 35.1 (C-5), 40.7 (C-9a), 42.9 (C-3), 43.1 (C-5a), 46.8 (C-13a), 62.8 (C-4), 78.6 (C-12), 82.4 (C-15a), 123.5 (C-9), 127.2 (C-6), 133.9 (C-15), 136.0 (C-8), 138.1 (C-14), 138.8 (C-7), 171.9 (C-2), 174.2 (C-10).

^{1}H NMR (700 MHz, Pyr) δ = 0.90 (t, *J* = 7.4, 3H, 18-H_3), 1.51 (dqd, *J* = 5.3, 7.4, 14.7, 1H, 17-H_a), 1.56 – 1.65 (m, 1H, 17-H_b), 1.66 – 1.73 (m, 1H, 13-H_b), 1.77 (d, *J* = 1.1, 3H, 19-H_3), 1.86 (dq, *J* = 2.8, 14.8, 1H, 13-H_a), 1.93 (t, *J* = 1.2, 3H, 16-H_3), 1.95 – 2.02 (m, 1H, 5-H_a), 2.56 (dt, *J* = 5.0, 14.6, 1H, 5-H_b), 2.64 (td, *J* = 2.2, 8.7, 1H, 13a-H), 3.30 (tt, *J* = 5.4, 10.9, 1H, 5a-H), 3.44 – 3.47 (m, 1H, 3-H_b), 3.48 (dd, *J* = 5.6, 13.3, 1H, 3-H_a), 3.69 (t, *J* = 8.0, 1H, 9a-H), 4.42 (dddd, *J* = 1.8, 5.2, 7.2, 10.8, 1H, 12-H), 4.67 (d, *J* = 6.2, 1H, 4-H), 5.04 (d, *J* = 6.5, 1H, 15-H), 5.30 (dt, *J* = 10.0, 16.4, 1H, 6-H), 6.10 (d, *J* = 10.4, 1H, 7-H), 6.12 (d, *J* = 6.2, 1H, 15a-H), 6.13 – 6.16 (m, 1H, 9-H).

^{13}C NMR (176 MHz, Pyr) δ = 9.9 (C-18), 13.2 (C-19), 20.3 (C-16), 28.7 (C-17), 33.0 (C-13), 36.0 (C-5), 40.9 (C-9a), 43.8 (C-5a), 44.1 (C-3), 47.2 (C-13a), 62.3 (C-4), 78.1 (C-12), 82.8 (C-15a), 124.7 (C-9), 128.0 (C-6), 134.4 (C-15), 135.7 (C-8), 136.2 (C-14), 139.2 (C-7), 172.6 (C-2), 174.1 (C-10).

syn-form:

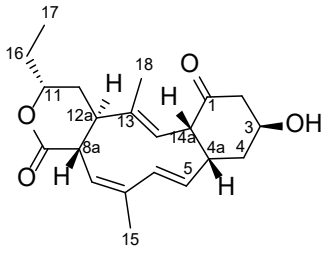
^{1}H NMR (700 MHz, CD_2Cl_2) δ = 1.00 (t, *J* = 7.6, 3H, 18-H_3), 1.59 (s, 3H, 19-H_3), 1.60 – 1.67 (m, 1H, 17-H_b), 1.70 – 1.73 (m, 1H, 5-H_a), 1.73 – 1.76 (m, 1H, 17-H_a), 1.92 (s, 3H, 16-H_3), 2.03 (dt, *J* = 9.3, 14.5, 1H, 13-H_b), 2.08 (ddd, *J* = 4.7, 7.0, 14.6, 1H, 13-H_a), 2.23 (ddd, *J* = 7.0, 9.2, 11.9, 1H, 13a-H), 2.34 – 2.38 (m, 1H, 5-H_b), 2.77 (hept, *J* = 5.1, 1H, 5a-H), 2.87 – 2.92 (m, 1H, 3-H_a), 3.14 (dd, *J* = 9.5, 15.3, 1H, 3-H_b), 3.37 (dd, *J* = 8.4, 11.9, 1H, 9a-H), 4.39 – 4.43 (m, 1H, 12-H), 4.43 – 4.47 (m, 1H, 4-H), 5.18 (dd, *J* = 7.1, 16.4, 1H, 6-H), 5.27 (dd, *J* = 1.5, 5.4, 1H, 15-H), 5.48 (t, *J* = 4.8, 1H, 15a-H), 5.57 (dt, *J* = 1.6, 8.4, 1H, 9-H), 5.94 (d, *J* = 16.3, 1H, 7-H).

^{13}C NMR (176 MHz, CD_2Cl_2) δ = 10.0 (C-18), 20.4 (C-16), 23.6 (C-19), 28.6 (C-17), 35.0 (C-13), 35.7 (C-5), 43.1 (C-3), 44.6 (C-5a), 45.4 (C-13a), 46.7 (C-9a), 62.6 (C-4), 78.8 (C-12), 80.3 (C-15a), 120.5 (C-9), 122.0 (C-15), 124.5 (C-6), 133.8 (C-8), 138.4 (C-7), 148.1 (C-14), 171.9 (C-2), 174.2 (C-10).

^{1}H NMR (700 MHz, Pyr) δ = 0.99 (t, *J* = 7.4, 3H, 18-H_3), 1.55 – 1.60 (m, 1H, 17-H_b), 1.65 – 1.74 (m, 1H, 17-H_a), 1.74 (d, *J* = 1.3, 3H, 19-H_3), 1.91 (d, *J* = 1.5, 4H, 5-H_a, 16-H_3), 1.95 – 2.02 (m, 1H, 13-H_b), 2.06 (ddd, *J* = 4.5, 6.6, 14.4, 1H, 13-H_a), 2.24 (ddd, *J* = 6.6, 9.7, 11.9, 1H, 13a-H), 2.56 (dt, *J* = 5.0, 14.6, 1H, 5-H_b), 3.09 (dp, *J* = 5.0, 10.1, 1H, 5a-H), 3.38 (dd, *J* = 4.2, 15.1, 1H, 3-H_a), 3.43 – 3.48 (m, 1H, 3-H_b), 3.75 (dd, *J* = 8.3, 12.0, 1H, 9a-H), 4.54 (ddt, *J* = 4.8, 7.5, 9.8, 1H, 12-H), 4.64 (dt, *J* = 3.6, 8.5, 1H, 4-H), 5.28 – 5.35 (m, 1H, 6-H), 5.37 (dd, *J* = 1.4, 5.3, 1H, 15-H), 5.88 (t, *J* = 4.8, 1H, 15a-H), 5.91 (dt, *J* = 1.5, 8.4, 1H, 9-H), 6.12 (d, *J* = 6.2, 1H, 7-H).

^{13}C NMR (176 MHz, Pyr) δ = 10.3 (C-18), 20.6 (C-16), 24.0 (C-19), 28.9 (C-17), 35.6 (C-13), 36.5 (C-5), 44.4 (C-3), 45.2 (C-5a), 45.9 (C-13a), 46.6 (C-9a), 62.1 (C-4), 78.3 (C-12), 80.6 (C-15a), 121.6 (C-9), 123.2 (C-15), 125.1 (C-6), 134.5, 137.9 (C-8), 138.6 (C-7), 147.7 (C-14), 172.5 (C-2), 174.2 (C-10).

5.6.3 Characterization of collinoketone (**2**)



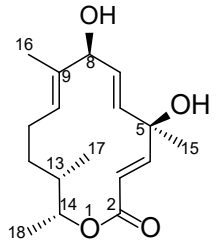
2

HR-ESI-MS (m/z): $[M+H]^+$ calculated for $C_{21}H_{28}O_4$: 345.2060; found 345.2058 (0.6 ppm err, 4.0 mSigma)

^{1}H NMR (700 MHz, CD_2Cl_2) δ = 1.00 (t, *J* = 7.4, 3H, 17-H_3), 1.49 (d, *J* = 1.3, 3H, 18-H_3), 1.56 – 1.65 (m, 1H, 16-H_a), 1.67 – 1.73 (m, 2H, 4-H_a, 16-H_b), 1.73 – 1.80 (m, 1H, 12-H_a), 1.85 (s, 3H, 16-H_3), 1.99 (dt, *J* = 2.4, 14.9, 1H, 12-H_b), 2.09 – 2.18 (m, 1H, 4a-H), 2.28 (dtd, *J* = 2.2, 4.0, 12.9, 1H, 4-H_b), 2.36 (td, *J* = 2.3, 10.0, 1H, 12a-H), 2.45 (ddd, *J* = 1.1, 11.6, 12.7, 1H, 2-H_a), 2.74 (ddd, *J* = 2.2, 4.8, 12.7, 1H, 2-H_b), 2.81 – 2.89 (m, 2H, 14a-H, 8a-H), 3.93 (tt, *J* = 4.6, 11.3, 1H, 3-H), 4.25 (dddd, *J* = 2.2, 5.4, 7.4, 11.3, 1H, 11-H), 5.01 (d, *J* = 10.1, 1H, 14-H), 5.27 (dd, *J* = 9.5, 11.3, 1H, 5-H), 5.39 – 5.53 (m, 1H, 8-H), 6.01 (ddt, *J* = 1.0, 2.1, 11.2, 1H, 6-H).

^{13}C NMR (176 MHz, CD_2Cl_2) δ = 9.9 (C-17), 13.3 (C-18), 24.5 (C-15), 28.5 (C-16), 31.6 (C-12), 40.7 (C-4), 41.1 (C-4a), 43.3 (C-8a), 45.9 (C-12a), 52.0 (C-2), 54.2 (C-15), 69.3 (C-3), 78.7 (C-11), 124.3 (C-8, 14), 132.3 (C-6), 134.3 (C-5), 136.9 (C-13), 138.1 (C-7), 174.1 (C-9), 207.6 (C-1).

5.6.4 Characterization of cineromycin B (**3**)



3

HR-ESI-MS (m/z): $[M-H_2O+H]^+$ calculated for $C_{17}H_{26}O_4$: 277.1798; found 277. 1806 (-0.5 ppm err, 11.8 mSigma)

^{1}H NMR (700 MHz, CD_2Cl_2) δ = 0.88 (d, *J* = 6.9, 3H, 17-H_3), 1.15 – 1.19 (m, 1H, 12-H_a), 1.20 (d, *J* = 6.4, 3H, 18-H_3), 1.22 – 1.28 (m, 1H, 12-H_b), 1.37 – 1.46 (m, 1H, 13-H), 1.53 (s, 3H, 15-H_3), 1.70 (q, *J* = 1.3, 3H, 16-H_3), 1.84 – 1.90 (m, 1H, 11-H_b), 2.08 – 2.14 (m, 1H, 11-H_a), 4.50 – 4.56 (m, 2H, 8-H, 14-H), 5.23 (ddp, *J* = 1.3, 5.4, 6.7, 1H, 10-H), 5.73 (dd, *J* = 4.9, 16.3, 1H, 7-H), 5.77 – 5.83 (m, 2H, 3-H, 6-H), 6.93 (d, *J* = 15.5, 1H, 4-H).

^{13}C NMR (176 MHz, CD_2Cl_2) δ = 14.9 (C-16), 16.1 (C-17), 18.3 (C-18), 25.4 (C-11), 27.4 (C-15), 34.6 (C-12), 39.8 (C-13), 73.7 (C-5), 75.0 (C-14), 75.7 (C-8), 115.3 (C-3), 129.5 (C-10), 134.2 (C-7), 135.3 (C-6), 138.4 (C-9), 155.6 (C-4), 166.2 (C-2).

5.6.5 Feeding experiments with stable isotope precursors

Feeding experiments were performed in a total volume of 1 L of SM media, equally distributed to five 1L baffled flasks. Equilibrated Amberlite XAD-16 was added before autoclaving.

The fed cultures were inoculated with 10 mL per shaking flask of a 48-hour old culture grown in SM media. The fermentation was run for 72 hours at 28 °C and 200 rpm. After 6 hours, 2 mM of ^{13}C-labeled sodium acetate and non-labeled sodium propionate (2 mM) or ^{13}C-labeled sodium propionate with non-labeled malonic acid (2 mM, pH-value adjusted by adding sodium hydroxide), respectively, was added. The addition was repeated after 12 and 24 hours with 1 mM of each labeled and non-labeled component. Purification procedure is the same for labeled and non-labeled collinolactone (**1**) and is described in chapter 5.6.1 – Production of collinolactone (**1**). At the end, about 19 mg of each labeled pure compound were isolated.

5.7 Cell culture methods

The cell viability assay and the fluorescence microscopy procedure are already described in J. C. Schmid *et al.*, *Angew. Chem. Int. Ed.* **2021**, *60*, 23212.

RPMI 1640 GlutaMAX (Gibco by life technologies) containing 110 mg/L sodium pyruvate and supplemented with 10 % (v/v) fetal calf serum (FCS, Gibco by life technologies) and 1 % (v/v) penicillin/streptomycin (Gibco by life technologies) was used for experiments.

L929 cells (DSMZ Leibniz) were cultivated at 37 °C with 5 % CO_2 in a humidified incubator and subcultivated every two to three days (approx. 80 % confluency) using trypsine/EDTA (Gibco by life technologies) for detaching.
Ptk2 cells (CLS Cell Line Services) were cultivated at 37 °C with 5 % CO_2 in a humidified incubator. Growth medium was replaced every two to three days and cells were subcultured once a week at approx. 80 % confluency using trypsine/EDTA for detaching.

5.7.1 Cell viability assay on L929 cell line

The assay was performed according to ISO 10993-5:2009 (Biological evaluation of medical devices - Part 5: Tests for in vitro cytotoxicity): a reduction of cell viability by more than 30 % was considered as a cytotoxic effect.
Cells (10.000 cells/well, 100 µL total volume) were seeded to a sterile, clear 96-well plate (Ependorf). After cultivation for 24 hours, cells were incubated with the compounds. Compounds were dissolved in pure DMSO and diluted with medium to get the final test concentrations. Tested concentrations ranged from 0.12 µM to 1000 µM. The maximum DMSO concentration was set to 0.1 % to avoid cytotoxic effects caused by the DMSO. The growth medium was discarded and replaced by the medium containing the desired compound concentrations. DMSO (10 %) served as a positive control for reduced viability, cells without compound were used as a negative control. Cells were incubated for 24 hours with the compound dilutions, then cell viability was determined using MTS (CellTiter96 AQueous One Solution Cell Proliferation Assay from Promega).
The growth medium was discarded, and MTS diluted in RPMI cultivation medium without phenol red (20 µL in 100 µL/well) was added (total volume of 120 µL per well) and incubated for 1.5 hours. Absorbance at 490 nm was determined with a multiwell plate photometer (BMG, Plate Reader Omega), values were blank corrected (only MTS in RPMI) and normalized by negative control (cells without compound and without DMSO). Concentration dependent values were plotted, and the half maximal effective concentration (EC_{50}) values were determined.

5.7.2 Fluorescence microscopy on of PtK2 cell line

PtK2 cells were seeded into a sterile µ-slide 8 Well Glass Bottom chamber (ibidi) (10.000 cells/well) and used for experiments after 30 hours of cultivation. Compounds were dissolved in pure DMSO and diluted with medium to get the final test concentrations with a maximum DMSO concentration of 0.2 %. The growth medium was discarded and replaced by the medium containing the desired compound concentrations. DMSO (0.2 %) served as a negative control. After incubation for 12 hours, cells were fixed and stained for fluorescence microscopy.

Cells were fixed using 4 % para-formaldehyde solution for 15 min at room temperature. After washing with phosphate-buffered saline (PBS, Gibco by life technologies), cells were incubated for 10 minutes at room temperature with 0.1 % (v/v) Triton X-100 in PBS for permeabilization of the cell membrane.
Cells were washed 3 times and blocked with 3 % (w/v) bovine serum albumin (BSA) in PBS-Tween20 0.05 % (v/v) for 30 min at room temperature before washing once with PBS.
For the immunostaining, Mouse Anti-beta-tubulin (Tub 41 mono T4026, sigma-aldrich T4026) was diluted 1:300 in PBS, added and incubated for 1 hour. The antibody solution was removed and cells were washed 3 times with 0.05 % Tween20 in PBS (v/v) containing 1 % Bovine serum albumin (BSA, w/v). Goat anti mouse Alexa 488 (ab150113, abcam) secondary antibody was used at 1:400 in PBS and incubated for 1 hour in the dark. Antibody solution was removed, and cells were washed 3 times with PBS. NucBlue LiveReadyProbes Reagent (Hoechst 33342, Molecular Probes) and Actin Red 555 Ready Probes Reagent (Rhodamine phalloidin, Molecular Probes) were used for staining DNA and F-Actin. The applied concentration was two droplets of each solution per 1 mL of PBS and the incubation time was set to 30 min at room temperature in the dark. After a final washing step, PBS was replaced by mounting media (ibidi).
Slides were stored at 8 °C in the dark until used for fluorescence imaging.

Fluorescence microscopy was used to visualize DNA and tubulin structures (microtubule, spindle apparatus and monoastral spindle apparatus) using an inverted microscope (Axio Observer Z1, Carl Zeiss, Germany with LED-Colibri illumination unit) with a 20× objective (Plan-Apochromat 20x/0.80 Ph2 M27). For each compound, 8 randomly selected regions across the well were selected and images were acquired using z-Stack mode.
For image processing, software ZEN 2 Blue was used. Z-stacks were merged using the wavelet method. Histogram intensities were optimized using the Min/Max function.

Images were acquired and processed according to the published protocol which allows to determine the cell stage distribution using fluorescence images.[236] DAPI channel was extracted from images of 8 randomly selected regions across the well and exported using ZEN 2 Blue export function with no compression and intensity scaling. CellProfiler[237] (version 4.1.3) was used for image analysis, using default primary object detection settings. For object detection, image rescaling was performed before, but integrated DAPI intensity was calculated from raw data. Object diameter was adjusted to range from 40 - 90, thresholding method was set to manual and a threshold of 0.15 was used. The integrated DAPI intensity of each cell was automatically exported to Excel, normalized, and visualized using GraphPadPrism 9. Smoothing parameter was set to high.

5.8 Amyloid beta aggregation and disaggregation assay

Commercially available amyloid-beta (Aβ) peptide 1-42 from Bachem was used for the experiments. 0.5 mg Aβ was dissolved in 400 µL hexafluoro-2-propanol (HFIP, Carl Roth GmbH) and aliquoted into 4 PCR tubes. HFIP was then evaporated over night at room temperature in a fume hood. Residual HFIP and moisture were removed by additional evaporation for 2 hours in a lyophilizer. Aliquots were stored at -20 °C until usage.
Prior usage, one aliquot was allowed to warm up to room temperature. DMSO (Carl Roth GmbH) was added to prepare 1 mM stock which was vortexed for 30 seconds and sonicated for 10 min before diluted with deionized water to obtain 50 µM concentration. Compounds were dissolved in DMSO to obtain 25 mM stocks which were further diluted with deionized water to 2 mM.

For the disaggregation assay, Aβ stock (50 µM) was incubated for 48 hours at 37 °C. For the aggregation assay, freshly prepared Aβ stock (50 µM) was immediately used after monomerization.

12.5 µL of Aβ stock (50 µM) and 12.5 µL of compounds stocks (2 mM) were mixed in a 96 well PCR Plate (Biodeal Thermo-Fast). Final concentrations were 25 µM for Aβ and 1 mM for compounds.
25 µM for Aβ with similar concentrations of DMSO served as negative control. The sealed PCR plate was incubated for 48 hours in an Eppendorf Thermocycler at 37 °C, before 75 µL of freshly prepared thioflavin T (ThT, Sigma Aldrich) solution (5 µM ThT in 50 mM glycine buffer, pH 8.8) was added. Samples were transferred into a 96 well plate (Sarstedt, 96 well plate, flat bottom, black) and fluorescence intensity was determined at 450 nm (excitation) and 485 nm (emission) using a microplate reader. All experiments were performed in triplicates.

5.9 Motor protein assay

The procedure was adapted and modified from Cytoskeleton Inc. Kinesin ATPase END-POINT BIOCHEM KIT (BK053). To obtain high signal to noise data, special care must be taken with any phosphate containing chemicals e.g., sodium hydroxide used for buffers or water used to prepare these buffers. Laboratory glassware which was cleaned in a laboratory dishwasher (using phosphoric acid to protonate the glass surface at the end of the program) must be avoided at all costs.

5.9.1 Preparation of ATP and microtubule stocks

ATP (disodium hydrate salt) was purchased from Sigma Aldrich (A3377, Grade II, ≥97% (HPLC) and stocks were prepared as follows:
ATP (55.1 mg) was dissolved in 800 µL of PIPES buffer (100 mM, pH = 7.0) and the pH was adjusted to pH = 7 by addition of 1 M sodium hydroxide solution. The final volume of 1 mL was made up by addition of PIPES buffer. ATP stock solution was aliquoted in 9 × 100 µL and stored at -20 °C.

Microtubule stocks were prepared as follows:
500 µL of kinesin reaction buffer in an Eppendorf tube was temperated at 37 °C in a water bath for 10 minutes. 5 µL of a 2 mM stock of paclitaxel (taxol) in DMSO was then added to the reaction buffer and stored at room temperature. 200 µL of taxol supplemented kinesin reaction buffer was added to 1 mg of microtubules and incubated at room temperature for 10 minutes. The completely resuspended microtubules were aliquoted into 10 × 20 µL, snap frozen in liquid nitrogen and stored at -78 °C.

5.9.2 Assay preparation

Diluted protein stocks (microtubules and motor proteins) must be freshly prepared prior to the experiment and cannot be refrozen as they will lose their biological activity.

Preparation of motor proteins:
Eg5 motor protein was reconstituted at a concentration of 5 mg/mL with water and diluted with kinesin reaction buffer.
Dynein motor protein was resuspended by adding 50 µL of ice-cold water to prepare a 1 mg/mL solution.

Preparation of kinesin reaction buffer with taxol:
1 mL of kinesin reaction buffer was supplemented with 10 µL of taxol stock (2 mM in DMSO) and stored at room temperature.

Preparation of microtubule solution:
One aliquot of microtubules was thawn in a water bath at room temperature for 3 minutes, mixed with 240 µL of the kinesin reaction buffer supplemented with taxol to obtain a concentration of 0.2 mg/mL and then stored at room temperature.

Preparation of the ATP solution:
One aliquot of ATP was diluted with 5 mL of ice-cold water and placed on ice until use.

5.9.3 Assay setup and execution

For the experiments, transparent 96 half area microplates from Greiner were used.

The assay can be divided into three major steps. During the first step, the experiment is set up by pipetting the required protein solutions and buffer into the wells. This step is not time-critical and should be performed very accurately. Small droplets on the wall of the wells will have a significant impact on the results.
In the second phase, ATP solution is added to each well. The reaction starts as soon as the ATP was added, so the use of a microchannel pipette is required to obtain optimal results and pipetting has to be done as quick as possible.
After the time is up, the reactions are terminated by the addition of ready-to-use CytoPhos reagent (malachite green reagent). Again, the pipetting has to be done as quick as possible to avoid variations due to significantly different reaction times. The incubation of the CytoPhos reagent must be exactly 5 minutes followed by an immediate readout. Otherwise, too much of the remaining ATP is undergoing hydrolysis in the acidic environment of the CytoPhos reagent and will make the readout results invalid.

The first cycle of experiments should be used to optimize the parameters of the assay. See chapter 5.9.4 for detailed information.

The total volume of the reactions was 30 µL. 10 µL of the 0.2 mg/mL microtubule stock was added to each well, except for the blanks where no microtubules were required. The motor protein solution was added (0.2 – 1 µg of protein, depending on the motor protein batch activity). The compound of interest was dissolved in DMSO (1 mM) stock and diluted with kinesin reaction buffer supplemented with taxol. Pure DMSO, which was diluted similar to the compounds served as a control.

The final volume of 30 µL was made up by the addition of kinesin reaction buffer supplemented with taxol.

5 µL of ATP solution were added simultaneously to all wells, followed by an incubation at room temperature of 5 minutes for Eg5 and 90 for dynein.
The assay was terminated by the addition of 70 µL CytoPhos reagent to each well, followed by an additional incubation time of exactly 5 minutes.
Absorbance readout was done at 650 nm in a microplate reader.

5.9.4 Assay validation and protein titration

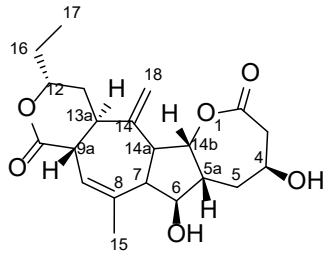
A phosphate standard (0.1 mM) was used to validate assay linearity in the range of 1 µM – 20 µM and to determine optimal absorbance wavelength readout (set to 650 nm, Figure 44B).

Since the activity of the motor protein is significantly different between types and batches of the same motor proteins can also have varying activities due to impurities, a titration of the ideal motor protein amount is required for each batch and type of motor protein. The concentration of the motor proteins should be optimized towards 10 µM of generated inorganic phosphate at the readout point.
For the Eg5 protein batch, a working concentration of 0.6 µg per reaction with a run time of 5 minutes was found to be best, while for the dynein protein batch, a concentration of 0.8 µg per reaction with a run time of 90 minutes was used.

5.10 Chemical synthesis

All reactions, unless otherwise stated, were performed under argon inert gas. Solvents for synthesis were dried over molecular sieves and degassed using the freeze-thaw method before use. Chemicals were purchased from Sigma-Aldrich and Acros Organics and used without further purification.

5.10.1 Synthesis of collazulen (**3**)



4

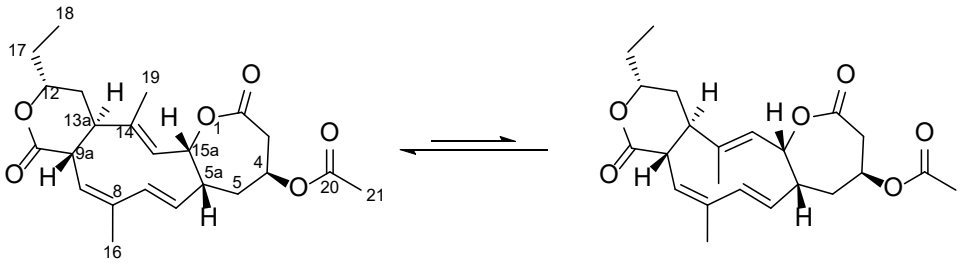
Collinolactone (**1**) (30 mg, 83.2 µmol, 1 Eq.) was dissolved in 20 mL of dry dichloromethane and cooled to 0 °C. 3-chloroperbenzoic acid (17.67 mg, 83.2 µmol, 1 Eq.) was dissolved in 20 mL of dry DCM and slowly added over 1 hour. The reaction mixture was allowed to warm up to room temperature and was stirred for another 4 hours. Saturated sodium bicarbonate solution was added to the reaction which was then extracted three times with dichloromethane (20 mL each). The combined organic fractions were evaporated *in vacuo* and the raw product was purified by preparative HPLC on a Kromasil 100 C_{18} (250 × 21.2 mm, 7 µM particle size) with 30 % acetonitrile/water at 17.5 mL/min. The pure compound 3 eluted at R_t = 14.0 – 15.0 min. Lyophilizing the product containing fractions yielded 14.9 mg (47 %) of a colorless, fluffy powder.

HR-ESI-MS (m/z): $[M+H]^+$ calculated for $C_{21}H_{29}O_6$: 377.1959; found 377.1956 (0.8 ppm err, 6.3 mSigma)

^{1}H NMR (700 MHz, CD_2Cl_2) δ = 1.03 (t, J = 7.4, 3H, 18-H_3), 1.62 – 1.67 (m, 1H, 5-H_a), 1.67 – 1.72 (m, 1H, 17-H_b), 1.73 – 1.76 (m, 1H, 17-H_a), 1.75 – 1.80 (m, 1H, 12-H_a), 1.83 (br, 1H, 15-OH), 2.01 (s, 3H, 16-H_3), 2.16 (dt, J = 3.4, 14.8, 1H, 5-H_b), 2.20 (ddd, J = 2.0, 3.5, 14.9, 1H, 12-H_b), 2.29 (td, J = 3.5, 8.4, 1H, 12a-H), 2.34 – 2.39 (m, 1H, 5a-H), 2.39 – 2.42 (m, 2H, 6a, 13a-H), 2.65 (br, 1H, 14-OH), 2.75 (dd, J = 7.9, 12.9, 1H, 3-H_a), 2.88 – 2.92 (m, 1H, 8a-H), 2.95 (dd, J = 8.2, 13.1, 1H, 3-H_b), 3.99 (t, J = 9.1, 1H, 6-H), 4.35 – 4.41 (m, 2H, 4, 11-H), 5.07 (d, J = 71.6, 2H, 19-H_a), 5.16 (dd, J = 4.2, 8.6, 1H, 13b-H), 6.23 (dt, J = 1.7, 5.5, 1H, 8-H).

^{13}C NMR (176 MHz, CD_2Cl_2) δ = 9.7 (C-18), 21.4 (C-16), 28.3 (C-17), 32.9 (C-12), 34.4 (C-5), 41.9 (C-12a), 42.0 (C-3), 43.2 (C-5a), 45.5 (C-8a), 53.0 (C-13a), 54.6 (C-6a), 64.2 (C-4), 78.9 (C-11), 79.1 (C-6), 79.7 (C-13), 105.2 (C-19), 127.9 (C-8), 138.9 (C-7), 153.4 (C-13), 171.3 (C-2), 173.7 (C-9).

5.10.2 Synthesis of acetyl-collinolactone (**5**)



5 (anti) **5** (syn)

Collinolactone (10 mg, 27.7 µmol, 1 Eq.) was dissolved in 5 mL of dry pyridine. At 0 °C acetic acid anhydride (2 mL) was slowly added over 10 min. The reaction mixture was stirred for 20 hours at room temperature before 5 mL of water was added to quench the reaction. The solvent was removed *in vacuo* and the raw product was purified by preparative HPLC on a Kromasil 100 C_{18} (250 × 21.2 mm, 7 µM particle size) with 50 % acetonitrile/water at 17.5 mL/min (R_t = 16.8 min) to afford 9.4 mg (84 %) of compound **5** as white, fluffy powder after lyophilization.

HR-ESI-MS (m/z): $[M+H]^+$ calculated for $C_{23}H_{31}O_6$: 403.2115; found 403.2118 (0.6 ppm err, 1.2 mSigma)

anti-form:

^{1}H NMR (700 MHz, CD_2Cl_2) δ = 1.01 (t, *J* = 7.4, 3H, 18-H_3), 1.58 – 1.67 (m, 1H, 17-H_b), 1.69 (s, 3H, 19-H_3), 1.70 – 1.77 (m, 1H, 17-H_a), 1.79 – 1.85 (m, 2H, 5-H_a, 13-H_b), 1.91 – 1.92 (m, 3H, 16-H_3), 1.92 – 1.94 (m, 1H, 13-H_a), 2.11 (s, 3H, 21-H_3), 2.48 (dt, *J* = 5.1, 15.3, 1H, 5-H_b), 2.64 (td, *J* = 3.0, 8.5, 1H, 13a-H), 2.73 – 2.80 (m, 1H, 5a-H), 2.92 (dt, *J* = 4.6, 15.8, 1H, 3-H_b), 3.25 – 3.33 (m, 2H, 3-H_a, 9a-H), 4.33 – 4.39 (m, 1H, 12-H), 4.83 – 4.86 (m, 1H, 15-H), 5.10 – 5.16 (m, 1H, 6-H), 5.26 – 5.31 (m, 1H, 4-H), 5.56 (t, *J* = 6.1, 1H, 15a-H), 5.76 (dp, *J* = 1.4, 7.5, 1H, 9-H), 5.90 (dt, *J* = 1.1, 16.5, 1H, 7-H).

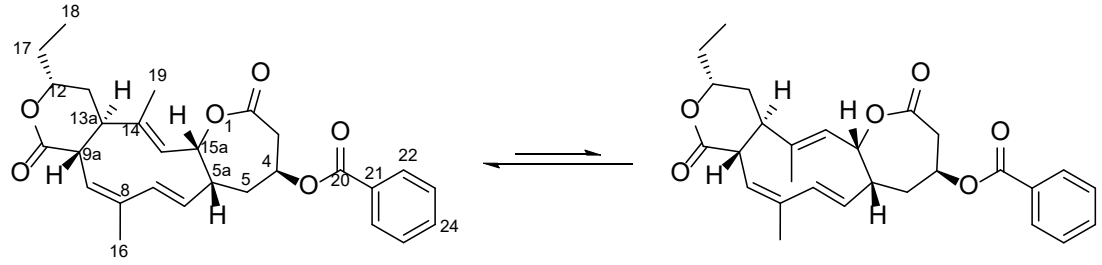
^{13}C NMR (176 MHz, CD_2Cl_2) δ = 9.7 (C-18), 13.3 (C-19), 19.8 (C-16), 21.2 (C-21), 28.5 (C-17), 31.9 (C-5), 32.4 (C-13), 40.5 (C-3), 40.6 (C-9a), 43.9 (C-5a), 46.7 (C-13a), 65.4 (C-4), 78.5 (C-12), 82.2 (C-15a), 123.8 (C-9), 126.3 (C-6), 133.6 (C-15), 133.9 (C-14), 135.9 (C-8), 139.2 (C-7), 170.3 (C-20), 170.6 (C-2), 173.7 (C-10).

syn-form:

^{1}H NMR (700 MHz, CD_2Cl_2) δ = 1.01 (td, J = 1.2, 7.4, 3H, 18 H_3), 1.60 (d, J = 1.5, 3H, 19 H_3), 1.58 – 1.67 (m, 1H, 17-H_a), 1.75 – 1.79 (m, 2H, 5-H_b, 17 H_b), 1.91 – 1.92 (m, 3H, 16-H_3), 2.02 – 2.08 (m, 2H, 13-H_2), 2.09 (s, 3H, 21-H_3), 2.24 (ddd, J = 7.0, 9.2, 11.9, 1H, 13a-H), 2.52 (dd, J = 4.8, 10.4, 1H, 5-H_a), 2.66 – 2.70 (m, 1H, 5a-H), 2.92 (dt, J = 4.6, 15.8, 4H, 3-H_b), 3.26 – 3.30 (m, 7H, 3-H_a), 3.36 (dd, J = 8.4, 11.9, 1H, 9a-H), 4.41 (ddt, J = 5.2, 7.4, 9.1, 1H, 12-H), 5.15 (d, J = 5.8, 1H, 6-H), 5.27 (dt, J = 1.8, 4.9, 1H, 15-H), 5.28 – 5.29 (m, 1H, 4-H), 5.39 (t, J = 4.9, 1H, 15a-H), 5.60 (dp, J = 1.6, 8.4, 1H, 9-H), 5.95 (dt, J = 1.3, 16.4, 1H, 7 H).

^{13}C NMR (176 MHz, CD_2Cl_2) δ = 10.0 (C-18), 20.4 (C-16), 21.2 (C-21), 23.7 (C-19), 28.6 (C-17), 32.6 (C-5), 35.1 (C-13), 40.6 (C-3), 44.9 (C-5a), 45.4 (C-13a), 46.7 (C-9a), 65.2 (C-4), 78.7 (C-12), 80.2 (C-15a), 120.6 (C-9), 121.5 (C-15), 124.1 (C-6), 138.0 (C-8), 138.7 (C-7), 148.9 (C-14), 170.3 (C-20), 170.5 (C-2), 173.8 (C-10).

5.10.3 Synthesis of benzoyl-collinolactone (**6**)



6 (anti) **6** (syn)

Collinolactone (**1**) (10 mg, 27.7 µmol, 1 Eq.) was dissolved in 20 mL of dry dichloromethane. DIC (N,N'-Diisopropylcarbodiimide, 10.7 µL,69.3 µmol, 2.5 Eq.) and DMAP (4-Dimethylaminopyridine, 0.68 mg, 5.5 µmol, 0.2 Eq.) were added, followed by benzoic acid (8.47 mg, 69.3 µmol, 2.5 Eq.). The reaction mixture was allowed to stir for 30 minutes at room temperature before heating up to 40 °C for additional 20 hours. The solvent was then removed *in vacuo* and the raw product was purified by preparative HPLC on a Kromasil 100 C_{18} (250 × 21.2 mm, 7 µM particle size) with 65 % acetonitrile/water at 17.5 mL/min (Rt = 12.9 min) to afford 9.6 mg (74 %) of compound **6** as white, fluffy powder after lyophilization.

HR-ESI-MS (m/z): $[M+H]^+$ calculated for $C_{28}H_{33}O_6$: 465.2272; found 465.2280 (1.8 ppm err, 3.2 mSigma)

anti-form:

^{1}H NMR (700 MHz, CD_2Cl_2) δ = 1.01 (t, J = 7.5, 3H, 18-H_3), 1.59 – 1.66 (m, 1H, 17-H_b), 1.71 (d, J = 1.2, 3H, 19-H_3), 1.74 (dd, J = 7.2, 14.3, 1H, 17-H_a), 1.81 (ddd, J = 8.6, 10.2, 15.0, 1H, 13-H_b), 1.91 – 1.95 (m, 3H, 16-H_3), 1.96 (d, J = 2.0, 2H, 5-H_a, 13-H_a), 2.63 – 2.69 (m, 2H, 5-H_b, 13a-H), 2.86 – 2.93 (m, 1H, 5a-H), 3.09 (dd, J = 4.4, 15.7, 1H, 3-H_b), 3.28 (t, J = 8.0, 1H, 9a-H), 3.44 – 3.49 (m, 1H, 3-H_a), 4.33 – 4.39 (m, 1H, 12-H), 4.88 (dd, J = 1.4, 6.2, 1H, 15-H), 5.19 (dd, J = 10.1, 16.4, 1H, 6-H), 5.56 (dtd, J = 1.9, 4.6, 9.7, 1H, 4-H), 5.68 (t, J = 6.0, 1H, 15a-H), 5.78 (dt, J = 1.4, 7.5, 1H, 9-H), 5.93 (dt, J = 1.1, 16.4, 1H, 7H), 7.48 – 7.53 (m, 2H, 23-H, 25-H), 7.64 (tdd, J = 1.0, 2.3, 7.6, 1H, 24 H), 8.07 (ddd, J = 1.4, 5.0, 8.4, 2H, 22-H, 26-H).

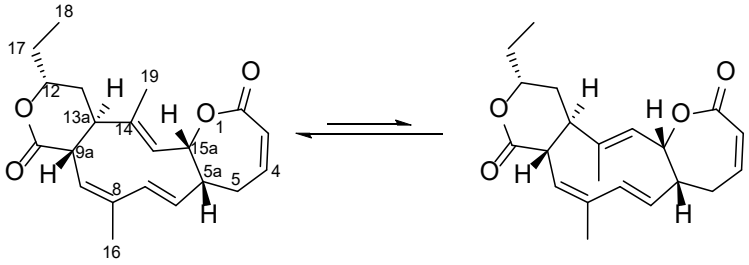
^{13}C NMR (176 MHz, CD_2Cl_2) δ = 9.7 (C-18), 13.3 (C-19), 19.8 (C-16), 28.5 (C-17), 32.1 (C-5), 32.5 (C-13), 40.6 (C-9a), 40.8 (C-3), 44.2 (C-5a), 46.7 (C-13a), 66.1 (C-4), 78.4 (C-12), 82.3 (C-15a), 123.8 (C-9), 126.1 (C-6), 129.0 (C-23, C-25), 129.8 (C-22, C-26), 130.2 (C-14), 133.7 (C-15), 133.8 (C-21), 135.9 (C-8), 139.4 (C-7), 165.7 (C-20), 170.6 (C-2), 173.7 (C-10).

syn-form:

^{1}H NMR (700 MHz, CD_2Cl_2) δ = 1.01 (t, *J* = 7.5, 3H, 18-H_3), 1.62 – 1.64 (m, 3H, 19-H_3), 1.69 – 1.85 (m, 1H, 17-H_a), 1.88 (ddd, *J* = 1.4, 3.5, 14.2, 1H, 5-H_a), 1.89 – 1.98 (m, 1H-17-H_b), 1.93 – 1.95 (m, 3H, 16-H_3), 2.02 – 2.11 (m, 2H, 13-H_2), 2.26 (ddd, *J* = 7.0, 9.2, 12.0, 1H, 13a-H), 2.67 (t, *J* = 5.1, 2H, 5-H_b), 2.76 – 2.82 (m, 1H, 5a-H), 3.09 (dd, *J* = 4.4, 15.7, 1H, 3-H_b), 3.37 (dd, *J* = 8.4, 11.9, 1H, 9a-H), 3.41 – 3.46 (m, 1H, 3-H_a), 4.40 – 4.44 (m, 1H, 12-H), 5.21 (s, 1H, 6-H), 5.30 (dd, *J* = 1.5, 5.4, 1H, 15-H), 5.52 (dt, *J* = 4.4, 9.5, 2H, 4, 15a-H), 5.61 (dp, *J* = 1.5, 8.4, 1H, 9-H), 5.98 (dt, *J* = 1.3, 16.2, 1H, 7-H), 7.48 – 7.53 (m, 2H, 23-H, 25-H), 7.64 (tdd, *J* = 1.0, 2.3, 7.6, 1H, 24-H), 8.07 (ddd, *J* = 1.4, 5.0, 8.4, 2H, 22-H, 26-H).

^{13}C NMR (176 MHz, CD_2Cl_2) δ = 10.0 (C-18), 20.4 (C-16), 23.7 (C-19), 28.6 (C-17), 32.8 (C-5), 35.1 (C-13), 40.7 (C-3), 45.1 (C-5a), 45.4 (C-13a), 46.6 (C-9a), 66.1 (C-4), 78.6 (C-12), 80.3 (C-15a), 120.7 (C-9), 121.5 (C-15), 124.0 (C-6), 128.9 (C-23, C-25), 129.9 (C-22, C-26), 130.2 (C-21), 133.8 (C-24), 133.9, 138.0 (C-8), 138.9 (C-7), 149.0 (C-14), 165.9 (C-20), 170.4 (C-2), 173.8 (C-10).

5.10.4 Synthesis of collinolactenone (**7**)



7 (anti) **7** (syn)

Collinolactone (**1**) (25 mg, 69.4 µmol, 1 Eq.) was dissolved in 50 mL of dry acetonitrile. Burgess reagent (1-Methoxy-N-triethylammoniosulfonyl-methanimidate, 66.1 mg, 277.4 µmol, 4 Eq.) was added slowly over 1 hour before the reaction solution was refluxed for 20 hours. After removal of the solvent *in vacuo*, the product was purified by preparative HPLC on a Kromasil 100 C_{18} (250 × 21.2 mm, 7 µM particle size) with 50 % acetonitrile/water at 17.5 mL/min. The pure compound collinolactenone (**7**) eluted at R_t = 16.0 min. Lyophilizing the product containing fractions yielded 17.1 mg (72 %) of compound **7** as a white, fluffy powder.

HR-ESI-MS (m/z): $[M+H]^+$ calculated for $C_{21}H_{27}O_4$: 343.1904; found 343.1904 (0 ppm err, 10.5 mSigma)

anti-form:

^{1}H-NMR (700 MHz, CD_2Cl_2): δ (ppm) = 1.00 (t, J = 7.5, 2.4 Hz, 3H, 18-H_3), 1.60 (s, 3H, 19-H_3), 1.63 - 1.76 (m, 1H), 1.80 - 1.86 (m, 1H, 13-H_b), 1.82 – 1.86 (m, 1H, 17-H), 1.90 (dd, J = 3.0, 2.2 Hz, 1H, 13-H_a), 1.94 (s, 3H, 16-H_3), 2.56 – 2.48 (m, 1H, 5-H_b), 2.63 (td, J = 8.5, 3.0 Hz, 1H, 13a-H), 2.78 – 2.71 (m, 2H, 5-H, 5-H_a), 3.25 (ddd, J = 8.6, 7.5, 1.1 Hz, 1H, 12-H), 4.35 (tdd, J = 7.5, 6.6, 5.7, 2.1 Hz, 1H, 15-H), 4.89 – 4.85 (m, 1H, 6-H), 5.38 (dd, J = 7.1, 5.6 Hz, 1H, 15a-H), 5.76 (dp, J = 7.5, 1.5 Hz, 1H, 9-H), 5.86 (dt, J = 16.4, 1.0 Hz, 1H, 7-H), 6.07 (ddd, J = 10.8, 2.4, 1.1 Hz, 1H, 3-H), 6.67 – 6.62 (m, 1H, 4-H).

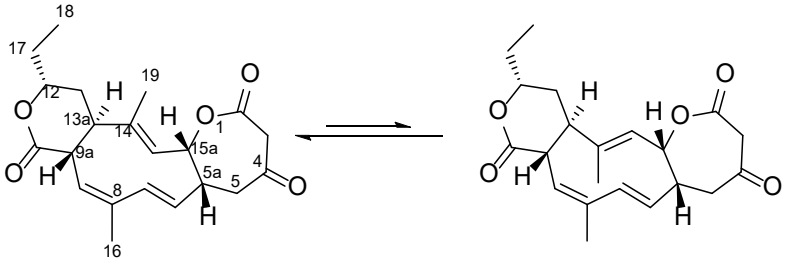
^{13}C-NMR (176 MHz, CD_2Cl_2): δ (ppm) = 9.7 (C-18), 13.2 (C-19), 19.9 (C-16), 28.5 (C-17), 29.2 (C-13), 32.5 (C-15), 40.6 (C-9a), 49.5 (C-5a), 54.2 (C-13a), 78.5 (C-12), 83.0 (C-15a), 123.8 (C-9), 125.45 (C-6), 126.8 (C-4), 133.3 (C-5), 133.9 (C-14), 135.9 (C-3), 139.3 (C-8), 140.93 (C-7), 169.5 (C-2), 173.8 (C-10).

syn-form:

^{1}H-NMR (700 MHz, CD_2Cl_2): δ (ppm) = 1.00 (td, J = 7.5, 2.4 Hz, 3H, 18-H_3), 1.58 (s, 3H, 19-H_3), 1.65 – 1.86 (m, 1H, 17-H_3), 1.92 (s, 3H, 16-H_3), 2.10 – 2.06 (m, 2H, 13-H), 2.24 – 2.28 (m, 1H, 13a-H), 2.45 – 2.39 (m, 1H, 5-H_b), 2.70 – 2.64 (m, 1H, 5a-H), 2.82 – 2.78 (m, 1H, 5-H_a), 3.33 (dd, J = 11.9, 8.4 Hz, 1H, 9a- H), 4.36 – 4.43 (m, 1H, 12-H), 5.20 (m, 3H, 6-H, 14-H, 15a-H), 5.60 (dp, J = 8.4, 1.5 Hz, 1H, 9-H), 5.92 (dt, J = 16.4, 1.3 Hz, 1H, 7-H), 6.03 (ddd, J = 10.9, 2.5, 1.2 Hz, 1H, 3-H), 6.62 – 6.58 (m, 1H, 4-H).

^{13}C-NMR (176 MHz, CD_2Cl_2): δ (ppm) = 10.0 (C-18), 20.5 (C-19), 23.7 (C-16), 28.6 (C-17), 29.7 (C-13), 35.1 (C-15), 44.9 (C-9a), 49.7 (C-5a), 54.2 (C-13a), 78.8 (C-12), 81.1 (C-15a), 120.6 (C-9), 121.7 (C-6), 124.8 (C-4), 125.3 (C-3), 133.3 (C-8), 140.6 (C-7), 149.4 (C-2), 173.8 (C-10).

5.10.5 Synthesis of keto-collinolactone (**8**)



8 (anti) **8** (syn)

Collinolactone (**1**) (10 mg, 27.7 µmol, 1 Eq.) was dissolved in 10 mL of dry dichloromethane and Dess–Martin periodinane (3-oxo-1,3-dihydro-11λ5,2-benziodoxole-1,1,1-triyl triacetate, 12.94 mg, 30.5 µmol, 1.1 Eq.) were added. The reaction mixture was stirred for 4 hours at room temperature before saturated sodium bicarbonate solution was added and extracted three times with dichloromethane (20 mL each). The combined organic fractions were evaporated *in vacuo* and the raw product was purified by preparative HPLC on a Kromasil 100 C_{18} (250 × 21.2 mm, 7 µM particle size) with 34 % acetonitrile/water at 17.5 mL/min (R_t = 29.3 min) to afford 14.9 mg (47 %) of **8** as white, fluffy powder after lyophilization.

HR-ESI-MS (m/z): $[M+H]^+$ calculated for $C_{21}H_{27}O_5$: 359.1853; found 377.1851 (0.4 ppm err, 14.7 mSigma)

anti-form:

^{1}H NMR (700 MHz, CD_2Cl_2): δ (ppm) = 1.01 (t, J = 7.4 Hz, 3H, 18-H_3), 1.58 – 1.64 (m, 1H, 17-H_b), 1.63 (s, 3H, 19-H_3), 1.67 – 1.79 (m, 1H, 17-H_a), 1.79 – 1.84 (m, 1H, 13- H_b), 1.87 – 1.91 (m, 1H, 13- H_a), 1.92 – 1.93 (m, 3H, 15-H), 2.64 (m, 1H, 13a-H), 2.77 (m, 1H, 5a-H), 2.81 – 3.01 (m, 1H, 3-H), 3.25 (t, J = 8.1 Hz, 1H, 9a-H), 3.56 (d, J = 19.5 Hz, 1H, 5-H_b), 3.83 (d, J = 19.5 Hz, 1H, 5-H_a), 4.30 – 4.36 (m, 1H, 12-H), 4.86 (d, J = 6.4 Hz, 1H, 15-H), 5.15 – 5.24 (m, 1H, 6-H), 5.52 (t, J = 5.7 Hz, 1H, 15a-H), 5.80 (dt, J = 7.5, 1.1, 1.1 Hz, 1H, 9-H), 5.91 (d, J = 16.4 Hz, 1H, 7-H).

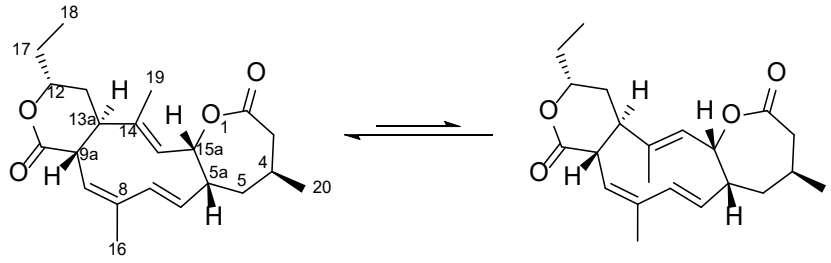
^{13}C-NMR (176 MHz, CD_2Cl_2): δ (ppm) = 9.9 (C-18), 13.6 (C-19), 20.0 (C-16), 28.7 (C-17), 32.6 (C-13), 40.8 (C-9a), 45.0 (C-3), 46.9 (C-13a), 47.4 (C-5a), 49.7 (C-5), 78.5 (C-12), 82.6 (C-15a), 124.4 (C-9), 124.5 (C-6), 132.9 (C-15), 134.9 (C-14), 135.6 (C-8), 140.3 (C-7), 168.3 (C-2), 173.4 (C-10), 200.8 (C-4).

syn-form:

^{1}H NMR (700 MHz, CD_2Cl_2): δ (ppm) = 1.01 (t, J = 7.5 Hz, 3H, 18-H_3), 1.59 – 1.65 (m, 1H, 17-H_b), 1.69 – 1.78 (m, 1H, 17-H_a), 1.64 (s, 3H, 19), 1.95 (d, J = 1.3 Hz, 3H, 16-H_3), 2.01 – 2.06 (m, 1H, 13), 2.29 (m, 1H, 13a-H), 2.66 – 2.71 (m, 1H, 5a-H), 2.82 – 2.91 (m, 1H, 3-H), 3.34 (dd, J = 11.8, 8.3 Hz, 1H, 9a-H), 3.53 (d, J = 19.9 Hz, 1H, 5-H_b), 3.76 (d, J = 20.3 Hz, 1H, 5-H_a), 4.40 (m, 1H, 12), 5.15 – 5.24 (m, 2H, 6-H, 15-H), 5.31 (dd, J = 10.0, 5.1 Hz, 1H, 15a-H), 5.63 (dt, J = 8.4, 1.4, 1.4 Hz, 1H, 9-H), 5.96 (d, J = 16.2 Hz, 1H, 7-H).

^{13}C-NMR (176 MHz, CD_2Cl_2): δ (ppm) = 10.2 (C-18), 20.5 (C-16), 23.9 (C-19), 28.8 (C-17), 35.1 (C-13), 45.4 (C-3), 45.8 (C-13a), 46.9 (C-9a), 48.3 (C-5a), 49.3 (C-5), 78.6 (C-12), 80.9 (C-15a), 120.5 (C-14), 121.2 (C-9), 122.5 (C-6), 137.8 (C-8), 139.7 (C-7), 150.0 (C-15), 167.9 (C-2), 173.5 (C-10), 201.1 (C-4).

5.10.6 Synthesis of methylcollinolactone (**9**)



9 (anti) **9** (syn)

To heat-dried Copper(I)-iodine (16.6 mg, 87.6 µmol, 4 Eq.) a solution of methyllithium (0.08 M) in diethyl ether (1 mL, 87.6 µmol, 4 Eq.) was added at -30 °C and stirred for 1 hour before adding compound **7** (7.5 mg, 21.9 µmol, 1 Eq.) at -78 °C. The reaction mixture was allowed to heat to -10 °C over the course of 4 hours.

Saturated ammonium chloride solution was added and stirred at room temperature for another hour for quenching. The organic phase was extracted three times with DCM. The combined organic phases were dried with sodium sulfate and the solvent was removed *in vacuo*. The raw product was purified by preparative HPLC on a Kromasil 100 C_{18} (250 × 21.2 mm, 7 µM particle size) with 55 % acetonitrile/water at 17.5 mL/min (R_t = 16.9 min) to afford 3.1 mg (39 %) of product **9** after lyophilization.

HR-ESI-MS (m/z): $[M+H]^+$ calculated for $C_{22}H_{31}O_4$: 359.2217; found 359.2225 (2.2 ppm err, 3.7 mSigma)

anti-form:

^{1}H NMR (700 MHz, C_6D_6) δ = 0.59 – 0.62 (m, 3H, 20-H_3), 0.81 (t, *J* = 7.4, 3H, 18-H_3), 1.03 (s, 3H, 19-H_3), 1.18 – 1.26 (m, 2H, 13-H_b, 17-H_a), 1.30 – 1.36 (m, 1H, 13-H_a), 1.40 – 1.48 (m, 1H, 17-H_b), 1.56 – 1.64 (m, 1H, 5-H_b), 1.70 (ddd, *J* = 4.1, 11.9, 14.6, 1H, 5-H_a), 1.75 – 1.83 (m, 1H, 3-H_a), 1.84 (s, 3H, 16-H_3), 1.81 – 1.88 (m, 1H, 4-H), 2.20 (td, *J* = 2.5, 8.7, 1H, 13a-H), 2.29 (tt, *J* = 5.4, 10.9, 1H, 5a-H), 2.50 – 2.54 (m, 1H, 3-H_b), 2.78 (t, *J* = 7.5, 1H, 9a-H), 3.76 – 3.82 (m, 1H, 12-H), 4.79 (d, *J* = 6.6, 1H, 15-H), 4.92 – 4.95 (m, 1H, 15a-H), 4.95 – 4.99 (m, 1H, 6-H), 5.59 (dd, *J* = 1.9, 16.5, 1H, 7-H), 6.13 (dq, *J* = 1.5, 7.4, 1H, 9-H).

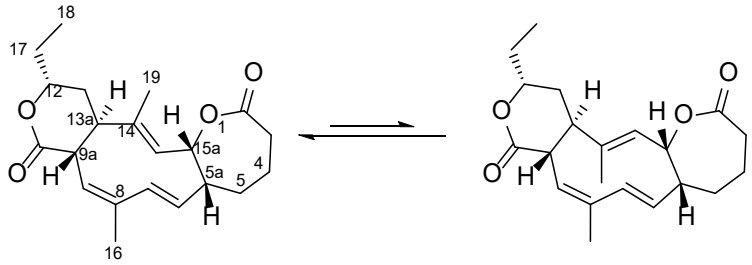
^{13}C NMR (176 MHz, C_6D_6) δ = 9.6 (C-18), 12.5 (C-19), 20.1 (C-16), 20.2 (C-20), 24.8 (C-4), 28.4 (C-17), 32.5 (C-13), 33.5 (C-5), 39.8 (C-3), 40.4 (C-9a), 43.6 (C-5a), 46.4 (C-13a), 77.1 (C-12), 81.1 (C-15a), 124.4 (C-9), 127.5 (C-6), 132.7 (C-14), 134.2 (C-15), 135.1 (C-8), 138.5 (C-7), 171.4 (C-2), 172.1 (C-10).

syn-form:

^{1}H NMR (700 MHz, C_6D_6) δ = 0.62 (dd, *J* = 6.8, 9.0, 3H, 20-H_3), 0.88 (t, *J* = 7.4, 3H, 18-H_3), 1.25 – 1.30 (m, 1H, 17-H_b), 1.40 – 1.45 (m, 1H, 5-H_b), 1.49 – 1.55 (m, 2H, 13-H_b, 17-H_a), 1.56 – 1.58 (m, 1H, 13-H_a), 1.58 (s, 3H, 19-H_3), 1.61 – 1.64 (m, 1H, 5-H_a), 1.74 – 1.79 (m, 1H, 4-H), 1.80 (s, 3H, 16-H_3), 1.81 – 1.85 (m, 1H, 3-H_a), 1.84 – 1.88 (m, 1H, 13a-H), 2.00 (dp, *J* = 4.4, 9.4, 1H, 5a-H), 2.47 (dd, *J* = 7.5, 14.1, 1H, 3-H_b), 2.90 – 2.97 (m, 1H, 9a-H), 3.86 – 3.92 (m, 1H, 12-H), 4.50 – 4.55 (m, 1H, 15-H), 4.55 – 4.59 (m, 1H, 15a-H), 4.96 – 5.01 (m, 3H, 6-H), 5.54 (d, *J* = 16.3, 1H, 7-H), 5.87 (dq, *J* = 1.5, 8.3, 1H, 9-H).

^{13}C NMR (176 MHz, C_6D_6) δ = 10.0 (C-18), 20.2 (C-16), 21.1 (C-20), 23.9 (C-19), 24.9 (C-4) 28.6 (C-17), 34.7 (C-13), 36.0 (C-5), 40.2 (C-3), 45.4 (C-13a), 46.3 (C-5a), 46.4 (C-9a), 77.2 (C-12), 79.8 (C-15a), 121.7 (C-9), 121.9 (C-15), 123.2 (C-6), 137.0 (C-8), 138.2 (C-7), 147.5 (C-14), 171.4 (C-2), 172.3 (C-10).

5.10.7 Synthesis of dehydro-collinolactone (**10**)



10 (anti) **10** (syn)

Commercially obtained Stryker reagent (from Sigma-Aldrich) did not lead to the desired product, so fresh Strykers had to be prepared according to a previously described procedure.[215]

Copper(I)-chloride (397.2 mg, 4 mmol, 1 Eq.), Potassium *tert*-butoxide (449.6 mg, 4 mmol, 1 Eq.) and triphenylphosphine (2.09 g, 8 mmol, 2 Eq.) were dissolved in benzene and stirred for 1 hour at room temperature. Dimethylphenylsilane (1.2 mL, 7.8 mmol, 1.95 Eq.) was added to the solution which turned dark red and was stirred for another 3 hours at room temperature. The reaction solution was filtered under inert conditions through celite. Absolute acetonitrile was added to the filtrate to induce crystallization. Crystals were filtered, washed with acetonitrile, and dried *in vacuo*.

Compound **7** (7 mg, 20.4 µmol, 1 Eq.) and the synthesized stryker reagent (13.3 mg, 6.78 µmol, 0.4 Eq.) were dissolved in 10 mL of abs. benzene. After 12 h of stirring at room temperature the reaction was quenched by addition of saturated ammonium chloride solution and extracted three times with toluene. The combined organic phases were dried with sodium sulfate and purified by preparative HPLC on a Kromasil 100 C_{18} (250 × 21.2 mm, 7 µM particle size) with 50 % acetonitrile/water at 17.5 mL/min (R_t = 16.2 min) to afford 3.4 mg (48 %) of **10** after lyophilization.

HR-ESI-MS (m/z): $[M+H]^+$ calculated for $C_{21}H_{29}O_4$: 345.2060; found 345.2067 (1.9 ppm err, 18.6 mSigma)

anti-form:

^{1}H NMR (700 MHz, CD_2Cl_2) δ = 1.01 (t, *J* = 7.5, 3H, 18-H_3), 1.60 – 1.65 (3H, 5-H_b, 17-H_a), 1.67 (d, *J* = 1.1, 3H, 19-H_3), 1.69 – 1.76 (m, 2H, 4-H_a, 5-H_a), 1.77 – 1.83 (m, 1H, 13-H_a), 1.90 (dd, *J* = 2.2, 3.0, 1H, 13-H_b), 1.91 – 1.92 (s, 3H, 16-H_3), 1.95 – 2.03 (m, 1H, 4-H_b), 2.16 – 2.22 (m, 1H, 17-H_b), 2.48 (ddt, *J* = 5.6, 9.2, 10.8, 1H, 5a-H), 2.62 (td, *J* = 3.0, 8.6, 1H, 13a-H), 2.65 – 2.73 (m, 1H, 3-H_b), 2.94 – 3.00 (m, 1H, 3-H_a), 3.23 – 3.30 (m, 1H, 9a-H), 4.33 – 4.38 (m, 1H, 12-H), 4.78 – 4.83 (m, 1H, 15-H), 5.15 (dd, *J* = 10.0, 16.5, 1H, 6-H), 5.44 (t, *J* = 6.3, 1H, 15a-H), 5.74 (dp, *J* = 1.4, 7.5, 1H, 9), 5.85 (dt, *J* = 1.0, 16.5, 1H, 7-H).

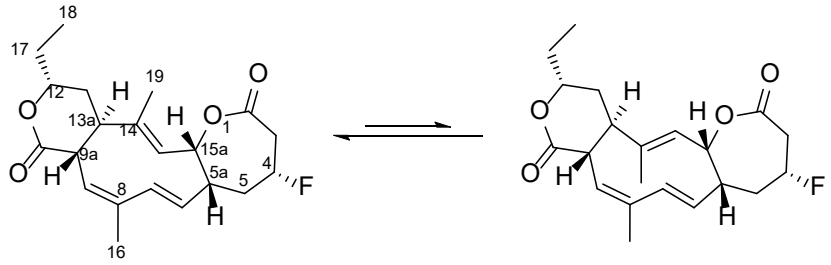
^{13}C NMR (176 MHz, CD_2Cl_2) δ = 9.8 (C-18), 13.4 (C-19), 18.6 (C-4), 20.1 (C-16), 27.6 (C-17), 28.6 (C-5), 32.6 (C-3), 32.7 (C-13), 40.8 (C-9a), 47.0 (C-13a), 49.3 (C-5a), 78.6 (C-12), 82.4 (C-15a), 123.7 (C-9), 127.9 (C-6), 133.7 (C-14), 134.1 (C-15), 136.2 (C-8), 138.6 (C-7), 173.5 (C-2), 174.0 (C-10).

syn-form:

^{1}H NMR (700 MHz, CD_2Cl_2) δ = 1.01 (t, *J* = 7.5, 3H, 18-H_3), 1.59 (d, *J* = 1.3, 3H, 19-H_3), 1.62 – 1.67 (m, 2H, 5-H_a, 17-H_a), 1.69 – 1.75 (m, 1H, 17-H_b), 1.77 – 1.82 (m, 1H, 4-H_b), 1.92 (dd, *J* = 1.9, 3.5, 3H, 16-H_3), 1.93 – 1.98 (m, 1H, 4-H_a), 2.03 – 2.09 (m, 1H, 13-H_b), 2.16 – 2.21 (m, 1H, 5-H_b), 2.22 – 2.2 (m, 1H, 9a-H), 2.35 – 2.41 (m, 1H, 5a-H), 2.65 – 2.73 (m, 1H, 3-H_b), 2.85 – 2.94 (m, 1H, 3-H_a), 3.35 (dd, *J* = 8.4, 11.9, 1H, 13a-H), 4.36 – 4.42 (m, 1H, 12-H), 5.17 – 5.20 (m, 1H, 6-H), 5.21 (tt, *J* = 1.4, 2.8, 1H, 15-H), 5.26 (t, *J* = 4.7, 1H, 15a-H), 5.58 (dp, *J* = 1.5, 8.3, 1H, 9-H), 5.89 (dt, *J* = 1.2, 16.3, 1H, 7-H).

^{13}C NMR (176 MHz, CD_2Cl_2) δ = 10.2 (C-18), 18.4 (C-4), 20.6 (C-16), 23.8 (C-19), 28.8 (C-5, 17), 32.8 (C-3), 35.1 (C-13), 45.6 (C-9a), 46.8 (C-13a), 50.3 (C-5a), 78.8 (C-12), 80.6 (C-15a), 120.6 (C-9), 122.0 (C-15), 124.8 (C-6), 138.3 (C-7), 138.3 (C-8), 148.4 (C-14), 173.4 (C-2), 174.1 (C-10).

5.10.8 Synthesis of epi-fluorocollinolactone (**11**)



11 (anti) **11** (syn)

N,N-Diethylaminosuflur trifluoride (DAST, 16.5 µL, 124.82 µmol, 3 Eq.) was added to a solution of compound **1** (15 mg, 41.6 µmol, 1 Eq.) in 2 mL of abs. DCM at -78 °C. After 3 hours of stirring at -78 °C the reaction was quenched by addition of water. Removal of solvent and purification by preparative HPLC on a Kromasil 100 C_{18} (250 × 21.2 mm, 7 µM particle size) with 45 % acetonitrile/water at 17.5 mL/min (R_t = 27.5 min) to afford 6.1 mg (40 %) of product **11** after lyophilization.

HR-ESI-MS (m/z): $[M+H]^+$ calculated for $C_{21}H_{28}O_4F$: 363.1966; found 363.1967 (2.8 ppm err, 5.5 mSigma)

anti-form:

^{1}H NMR (700 MHz, CD_2Cl_2) δ = 1.01 (td, *J* = 1.9, 7.4, 3H, 18-H_3), 1.60 – 1.64 (m, 1H, 17-H_a), 1.65 (s, 3H, 19-H_3), 1.68 – 1.76 (m, 1H, 17-H_b), 1.76 – 1.83 (m, 1H, 13-H_a), 1.88 – 1.92 (m, 1H, 13-H_b), 1.92 (s, 3H, 16-H_3), 1.94 – 2.00 (m, 1H, 5-H_a), 2.55 – 2.60 (m, 1H, 5a-H), 2.63 (qt, *J* = 3.3, 7.7, 2H, 5-H_b, 13a-H), 2.98 (dd, *J* = 16.1, 30.2, 1H, 3-H_b), 3.25 (t, *J* = 8.0, 1H, 9a-H), 3.42 (dddd, *J* = 2.3, 8.1, 16.3, 24.6, 1H, 3-H_a), 4.35 (tdd, *J* = 2.1, 6.5, 7.4, 1H, 12-H), 4.82 (dd, *J* = 1.4, 6.9, 1H, 15-H), 4.95 – 5.08 (m, 1H, 4-H), 5.14 (dd, *J* = 9.7, 16.4, 1H, 6-H), 5.24 (t, *J* = 6.3, 1H, 15a-H), 5.77 (dt, *J* = 1.5, 7.5, 1H, 9-H), 5.88 (d, *J* = 16.4, 1H, 7-H).

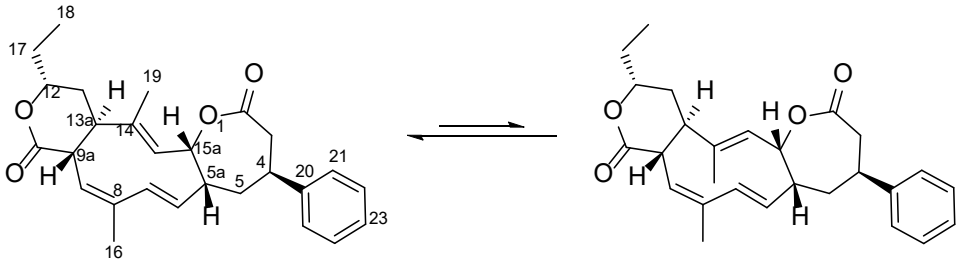
^{13}C NMR (176 MHz, CD_2Cl_2) δ = 9.7 (C-18), 13.3 (C-19), 20.0 (C-16), 28.5 (C-17), 32.4 (C-13), 34.1 (C-5), 34.3, 34.9, 40.6 (C-9a), 41.3 (C-3), 41.5, 44.0 (C-5a), 44.0, 46.8 (C-13a), 78.5 (C-12), 81.8 (C-15a), 85.0, 86.0 (C-4), 124.0 (C-9), 126.2 (C-6), 133.1 (C-15), 134.3 (C-14), 135.8 (C-8), 139.4 (C-7), 170.0 (C-2), 173.6 (C-10).

Syn-form:

^{1}H NMR (700 MHz, CD_2Cl_2) δ = 1.01 (td, *J* = 1.9, 7.4, 3H, 18-H_3), 1.60 (d, *J* = 1.3, 3H, 19-H_3), 1.60 – 1.63 (m, 1H, 17-H_b), 1.75 – 1.79 (m, 1H, 17-H_a), 1.92 (s, 3H, 16-H_3), 1.94 – 1.99 (m, 1H, 5-H_b), 2.00 – 2.08 (m, 2H, 13-H), 2.25 (ddd, *J* = 7.3, 9.0, 11.9, 1H, 13a-H), 2.49 – 2.54 (m, 1H, 5a-H), 2.63 (qt, *J* = 3.3, 7.7, 1H, 5-H_a), 2.98 – 3.05 (m, 1H, 3-H_b), 3.33 (dd, *J* = 8.4, 12.0, 1H, 9a), 3.42 – 3.45 (m, 1H, 3-H_a), 4.37 – 4.41 (m, 1H, 12-H), 5.00 – 5.03 (m, 1H, 15a-H), 4.95 – 5.08 (m, 1H, 4-H), 5.17 – 5.22 (m, 2H, 6-H, 15-H), 5.61 (dt, *J* = 1.6, 8.4, 1H, 9-H), 5.92 (d, *J* = 16.4, 1H, 7-H).

^{13}C NMR (176 MHz, CD_2Cl_2) δ = 10.0 (C-18), 20.3 (C-16), 23.7 (C-19), 28.6 (C-17), 34.3 (C-5), 34.9 (C-13), 34.9, 35.1, 41.2 (C-3), 41.4, 45.5 (C-13a), 45.8 (C-5a), 45.9, 46.7 (C-9a), 78.7 (C-12), 80.2 (C-15a), 85.0 (C-4), 86.0, 120.8 (C-9), 121.0 (C-15), 123.2 (C-6), 137.9 (C-8), 139.0 (C-7), 149.2 (C-14), 169.8 (C-2), 173.7 (C-10).

5.10.9 Synthesis of phenyl-collinolactone (**12**)



11 (anti) **11** (syn)

To heat-dried Copper(I)-iodine (17.9 mg, 94.3 µmol, 4 Eq.) a solution of phenyllithium (0.095 M) in dibutyl ether (1.98 mL, 89 µmol, 3.8 Eq.) was added at -30 °C and stirred for 1h before adding collinolactenone (**7**, 8 mg, 23.6 µmol, 1 Eq.) at -78 °C. The reaction mixture was allowed to heat to 0 °C over the course of 4 hours. To quench the reaction, saturated ammonium chloride solution was added and stirred at room temperature for another hour. The organic phase was extracted three times with DCM. The combined organic phases were then dried with sodium sulfate and the solvent was removed *in vacuo*. The raw product was purified by preparative HPLC on a Kromasil 100 C_{18} (250 × 21.2 mm, 7 µM particle size) with 65 % acetonitrile/water at 17.5 mL/min (R_t = 14.5 min) to afford 2.9 mg (29 %) of the pure product **12** after lyophilization.

HR-ESI-MS (m/z): $[M+H]^+$ calculated for $C_{27}H_{32}O_4$: 421.2373; found 421.2380 (1.5 ppm err, 9.0 mSigma)

anti-form:

^{1}H NMR (600 MHz, CD_2Cl_2) δ = 1.01 (t, *J* = 7.5, 3H, 18-H_3), 1.58 – 1.66 (m, 2H, 17-H_b), 1.63 (s, 3H, 16-H_3), 1.69 – 1.77 (m, 1H, 17-H_a), 1.77 – 1.84 (m, 1H, 13-H_b), 1.87 – 1.91 (m, 4H, 13-H_a, 19-H_3), 2.10 – 2.18 (m, 1H, 5-H_a), 2.32 – 2.38 (m, 1H, 5-H_b), 2.56 (tt, *J* = 5.3, 10.7, 1H, 5a-H), 2.64 (td, *J* = 3.1, 8.5, 1H, 13a-H), 2.97 (dd, *J* = 9.3, 13.9, 1H, 3-H_a), 3.20 – 3.25 (m, 1H, 9a-H), 3.31 (dd, *J* = 9.9, 13.8, 1H, 3-H_b), 3.56 (tdd, *J* = 3.1, 5.2, 9.2, 1H, 4-H), 4.31 – 4.37 (m, 1H, 12-H), 4.88 (d, *J* = 1.4, 1H, 15-H), 5.14 (dd, *J* = 9.9, 16.4, 1H, 6-H), 5.61 – 5.65 (m, 1H, 15a-H), 5.74 (dt, *J* = 1.4, 7.4, 1H, 9-H), 5.77 (dt, *J* = 1.1, 16.4, 1H, 7- H), 7.29 – 7.32 (m, 1H, 23-H), 7.34 – 7.38 (m, 2H, 22-H, 24-H), 7.41 (dd, *J* = 7.2, 8.2, 2H, 21-H, 25-H).

^{13}C NMR (151 MHz, CD_2Cl_2) δ = 9.7 (C-18), 13.3 (C-16), 20.1 (C-19), 28.5 (C-17), 32.5 (C-13), 34.8 (C-5), 36.1 (C-4), 37.6 (C-3), 40.6 (C-9a), 44.1 (C-5a), 46.9 (C-13a), 78.5 (C-12), 82.4 (C-15a), 123.6 (C-9), 127.0 (C-23), 127.3 (C-22, 24), 127.8 (C-6), 129.4 (C-21, 25), 133.8 (C-15), 133.8 (C-14), 135.9 (C-8), 138.5 (C-7), 142.9 (C-20), 172.7 (C-2), 173.7 (C-10).

syn-form:

^{1}H NMR (600 MHz, CD_2Cl_2) δ = 1.01 (t, *J* = 7.5, 3H, 18-H_3), 1.58 – 1.63 (m, 6H, 17-H_a), 1.64 (s, 3H, 19-H_3), 1.69 – 1.76 (m, 1H, 17-H_b), 1.94 (s, 3H, 16-H_3), 1.97 – 2.04 (m, 1H, 13-H_b), 2.04 – 2.10 (m, 1H, 13-H_a), 2.21 – 2.26 (m, 1H, 5-H_a), 2.26 – 2.29 (m, 1H, 13a-H), 2.30 – 2.33 (m, 1H, 5-H_b), 2.44 – 2.52 (m, 1H, 5a-H), 2.89 – 2.94 (m, 1H, 3-H_a), 3.18 – 3.23 (m, 1H, 3-H_b), 3.33 – 3.44 (m, 2H, 4, 9a-H), 4.36 – 4.41 (m, 1H, 12-H), 5.16 – 5.20 (m, 1H, 15-H), 5.30 – 5.37 (m, 1H, 6-H), 5.41 (t, *J* = 4.1, 1H, 15a-H), 5.59 – 5.64 (m, 3H, 9-H), 5.91 (d, *J* = 16.2, 1H, 7-H), 7.25 – 7.32 (m, 2H, 22-H, 24-H), 7.35 – 7.38 (m, 2H, 21-H, 25-H), 7.41 (dd, *J* = 7.2, 8.2, 1H, 23-H).

^{13}C NMR (151 MHz, CD_2Cl_2) δ = 10.0 (C-18), 20.1 (C-16), 23.9 (C-19), 28.7 (C-17), 34.6 (C-13), 36.2 (C-4), 38.1 (C-5), 39.8 (C-3), 46.1 (C-13a), 46.7 (C-9a), 47.9 (C-5a), 78.6 (C-12), 81.4 (C-15a), 120.9 (C-9), 122.0 (C-6), 122.3 (C-15), 127.0 (C-22, C-24), 127.2 (C-21, C-25), 129.2 (C-23), 138.0 (C-8), 138.6 (C-7), 144.4 (C-20), 147.3 (C-14), 172.6 (C-2), 173.9 (C-10).

5.11 Database access and processing

Table 2: Access times and versions of databases used for the analysis. ChEBI is the acronym for Chemical Entities of Biological Interest, COCONUT stands for COlleCtion of Open Natural ProdUcTs, DNP for Dictionary of Natural Products (commercial product), LOTUS for naturaL prOducTs occUrrence database and NPA for The Natural Product Atlas.

Name	Version
ChEBI	downloaded from website on July 13, 2022
COCONUT	January 2022; downloaded from website on July 13, 2022
DNP	30.1 (2021)
LOTUS	March 2021; downloaded from website on July 13, 2022
NPA	August 2021; downloaded from website on July 13, 2022

Data files were imported into MatLab (2022a, Natick, Massachusetts: The MathWorks Inc.) and missing structural information (InChI, InChIKey and SMILES) as well as accurate mass and molecular formula calculated based on provided information using the Chemical Development Kit (CDK) version 2.7.1. Existing information was validated and standardized by converting into IAtomContainer and back.
Entries with missing or invalid values were removed afterwards.

5.11.1 LSTM network setup and training parameters

For classification of biological activity strings from the DNP, the 1000 most frequently observed strings were manually classified as a training set. Strings were tokenized, punctuation was removed, and all letters converted to lower cases. Tokens were then converted to sequence with a sequence length of 8.
Training set was split into 80 % for initial training and 20 % for cross-validation. A LSTM was constructed with a sequenceInputLayer, wordEmbeddingLayer, lstmLayer, fullyConnectedLayer, softmaxLayer and classificationLayer. NumHiddenUnits parameter was set to 80, and embeddingDimension to 50.
Training options are summarized in Table 3. All other settings were set to default.

Table 3: Training options and values for LSTM network setup. All other settings were set to default.

Parameter name	Value
SolverName	adam
MaxEpochs	150
GradientThreshold	1
Shuffle	every-epoch
InitialLearnRate	0.01

Testing was performed on another 1000 manually classified strings, which were not part of the initial training and validation set. Strings identically prior to classification as mentioned above.

5.11.2 Retention time model setup and parameters

For the retention time prediction, molecular descriptors were calculated based on InChI, using the Chemical Development Kit (CDK) version 2.7.1. Invalid and empty descriptors were removed, and descriptors standardized based on z-score. Descriptors were further filtered on standard deviation (less than 0.01) and Pearson correlation score (less than 0.96).

A reduced subset of descriptors was generated using the F-test algorithm, with 25 % percentile removed.

80 % of the initial compound set served as initial training data, and 20 % were used for performance testing after training.

The learners parameters for the random forest (RF) model are displayed in Table 4. Learner for multiple linear regression (MLR) was set to leastsquare, learner for linear support vector machine (SVM) to svm. The selected hyperparameters to optimize for each model are presented in Table 5, and the corresponding HyperparameterOptimizationOptions settings, which are the same for all models, are summarized Table 6.

To proof, that molecular descriptors can be correlated to retention time and it is not only random guessing, a global reference value was calculated. For this purpose, the difference between the latest observed retention time of all compounds and the earliest observed retention time was calculated a divided by two. The resulted value marks the center of the retention time window for all compounds. The difference of this centered value and the measured retention time is used as a reference value.

Table 5: Selected hyperparameters for each model to optimize prior to training.

Model	Value
Random Forest (RF)	all
Multiple Linear Regression (MLR)	Regularization
	lambda
Linear Support Vector Machine (SVM)	Regularization
	lambda
Neural Network Regression (NNR)	activations
	lambda
	LayerSizes
	LayerBiasesInitializer
	LayerWeightsInitializer

Table 4: Learners parameters for the random forest (RF) model.

Parameter name	Value
NumVariablesToSample	all
PredictorSelection	curvature
Surrogate	on

Table 5: Selected hyperparameters for each model to optimize prior to training.

Model	Value
Random Forest (RF)	all
Multiple Linear Regression (MLR)	Regularization
	lambda
Linear Support Vector Machine (SVM)	Regularization
	lambda
Neural Network Regression (NNR)	activations
	lambda
	LayerSizes
	LayerBiasesInitializer
	LayerWeightsInitializer

Table 6: HyperparameterOptimizationOptions settings used for training of all models. Other settings were set to default values.

Parameter name	Value
AcquisitionFunctionName	expected-improvement-plus
MaxObjectiveEvaluations	60
Repartition	true
Kfold	10

REFERENCES

[1] B. Shen, *Cell* **2015**, *163*, 1297-1300.
[2] J. C. Schmid, K. Frey, M. Scheiner, J. F. G. Garzon, L. Stafforst, J. N. Fricke, M. Schuppe, H. Schiewe, A. Zeeck, T. Weber*, et al.*, *Angew Chem Int Ed Engl* **2021**, *60*, 23212-23216.
[3] D. J. Newman, G. M. Cragg, K. M. Snader, *Nat Prod Rep* **2000**, *17*, 215-234.
[4] A. S. Abdel-Razek, M. E. El-Naggar, A. Allam, O. M. Morsy, S. I. Othman, *Processes* **2020**, *8*, 470.
[5] W. H. Organization, *WHO global report on traditional and complementary medicine 2019*, World Health Organization, Geneva, **2019**.
[6] W. H. Organization, *Critically important antimicrobials for human medicine*, 6th rev. ed., World Health Organization, Geneva, **2019**.
[7] R. Bunch, J. Mcguire, Eli Lilly and Co, **1952**.
[8] D. P. Levine, *Clin. Infect. Dis.* **2006**, *42 Suppl 1*, S5-12.
[9] K. Johnson-Arbor, R. Dubey, in *StatPearls*, StatPearls Publishing, Treasure Island (FL), **2022**.
[10] O. Tacar, P. Sriamornsak, C. R. Dass, *Journal of Pharmacy and Pharmacology* **2013**, *65*, 157-170.
[11] W. C. Campbell, M. H. Fisher, E. O. Stapley, G. Albers-Schonberg, T. A. Jacob, *Science* **1983**, *221*, 823-828.
[12] Y. Tu, *Angew Chem Int Ed Engl* **2016**, *55*, 10210-10226.
[13] W. R. Burns, *Endeavour* **2008**, *32*, 101-106.
[14] E. A. Silva-Junior, A. C. Ruzzini, C. R. Paludo, F. S. Nascimento, C. R. Currie, J. Clardy, M. T. Pupo, *Sci. Rep.* **2018**, *8*, 2595.
[15] L. Seabrooks, L. Hu, *Acta Pharm Sin B* **2017**, *7*, 409-426.
[16] S. H. Ferreira, D. C. Bartelt, L. J. Greene, *Biochemistry* **1970**, *9*, 2583-2593.
[17] B. K. Carté, *Bioscience* **1996**, *46*, 271-286.
[18] R. Montaser, H. Luesch, *Future Med. Chem.* **2011**, *3*, 1475-1489.
[19] M. A. Ghareeb, M. A. Tammam, A. El-Demerdash, A. G. Atanasov, *Current Research in Biotechnology* **2020**, *2*, 88-102.
[20] A. Rhoads, K. F. Au, *Genomics Proteomics Bioinformatics* **2015**, *13*, 278-289.
[21] Y. Wang, Y. Zhao, A. Bollas, Y. Wang, K. F. Au, *Nat. Biotechnol.* **2021**, *39*, 1348-1365.
[22] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema, T. Weber, *Nucleic Acids Res* **2019**, *47*, W81-W87.
[23] T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Bruccoleri, S. Y. Lee, M. A. Fischbach, R. Müller, W. Wohlleben*, et al.*, *Nucleic Acids Research* **2015**, *43*, W237-W243.
[24] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, *Nucleic Acids Res* **1997**, *25*, 3389-3402.
[25] C. T. Walsh, M. A. Fischbach, *J. Am. Chem. Soc.* **2010**, *132*, 2469-2493.
[26] P. J. Rutledge, G. L. Challis, *Nat. Rev. Microbiol.* **2015**, *13*, 509-523.
[27] C. Ruttkies, S. Neumann, S. Posch, *BMC Bioinformatics* **2019**, *20*, 376.
[28] C. Ruttkies, E. L. Schymanski, S. Wolf, J. Hollender, S. Neumann, *Journal of Cheminformatics* **2016**, *8*, 3.
[29] F. Wang, J. Liigand, S. Tian, D. Arndt, R. Greiner, D. S. Wishart, *Anal. Chem.* **2021**, *93*, 11692-11700.
[30] Y. Djoumbou-Feunang, A. Pon, N. Karu, J. Zheng, C. Li, D. Arndt, M. Gautam, F. Allen, D. S. Wishart, *Metabolites* **2019**, *9*, 72.
[31] F. Allen, A. Pon, M. Wilson, R. Greiner, D. Wishart, *Nucleic Acids Research* **2014**, *42*, W94-W99.
[32] K. Duhrkop, M. Fleischauer, M. Ludwig, A. A. Aksenov, A. V. Melnik, M. Meusel, P. C. Dorrestein, J. Rousu, S. Bocker, *Nat. Methods* **2019**, *16*, 299-302.

[33] S. Bocker, M. C. Letzel, Z. Liptak, A. Pervukhin, *Bioinformatics* **2009**, *25*, 218-224.
[34] J. Watrous, P. Roach, T. Alexandrov, B. S. Heath, J. Y. Yang, R. D. Kersten, M. van der Voort, K. Pogliano, H. Gross, J. M. Raaijmakers*, et al.*, *Proc Natl Acad Sci U S A* **2012**, *109*, E1743-1752.
[35] F. Vincenti, C. Montesano, F. Di Ottavio, A. Gregori, D. Compagnone, M. Sergi, P. Dorrestein, *Front Chem* **2020**, *8*, 572952.
[36] L. F. Nothias, D. Petras, R. Schmid, K. Duhrkop, J. Rainer, A. Sarvepalli, I. Protsyuk, M. Ernst, H. Tsugawa, M. Fleischauer*, et al.*, *Nat. Methods* **2020**, *17*, 905-908.
[37] M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan*, et al.*, *Nat. Biotechnol.* **2016**, *34*, 828-837.
[38] R. Reher, H. W. Kim, C. Zhang, H. H. Mao, M. Wang, L. F. Nothias, A. M. Caraballo-Rodriguez, E. Glukhov, B. Teke, T. Leao*, et al.*, *J. Am. Chem. Soc.* **2020**, *142*, 4114-4120.
[39] C. Zhang, Y. Idelbayev, N. Roberts, Y. Tao, Y. Nannapaneni, B. M. Duggan, J. Min, E. C. Lin, E. C. Gerwick, G. W. Cottrell*, et al.*, *Sci. Rep.* **2017**, *7*, 14243.
[40] D. van Duin, D. L. Paterson, *Infect. Dis. Clin. North Am.* **2016**, *30*, 377-390.
[41] D. G. J. Larsson, C. F. Flach, *Nat. Rev. Microbiol.* **2022**, *20*, 257-269.
[42] C. J. L. Murray, K. S. Ikuta, F. Sharara, L. Swetschinski, G. Robles Aguilar, A. Gray, C. Han, C. Bisignano, P. Rao, E. Wool*, et al.*, *The Lancet* **2022**, *399*, 629-655.
[43] B. Hu, H. Guo, P. Zhou, Z. L. Shi, *Nat. Rev. Microbiol.* **2021**, *19*, 141-154.
[44] M. Prince, A. Wimo, M. Guerchet, G.-C. Ali, Y.-T. Wu, M. Prina, *World Alzheimer Report 2015*, **2015**.
[45] W. E. Wild C.P., Stewart B.W., **2020**.
[46] K. Y. Lam, P. Dickens, A. C. Chan, *Arch. Pathol. Lab. Med.* **1993**, *117*, 1027-1031.
[47] D. Hanahan, R. A. Weinberg, *Cell* **2011**, *144*, 646-674.
[48] S. I. Hajdu, *Cancer* **2011**, *117*, 1097-1102.
[49] S. I. Hajdu, *Cancer* **2004**, *100*, 2048-2051.
[50] P. Bandi, A. K. Minihan, R. L. Siegel, F. Islami, N. Nargis, A. Jemal, S. A. Fedewa, *Cancer Epidemiol Biomarkers Prev* **2021**, *30*, 1287-1299.
[51] L. M.-I. Institute of Medicine Committee on Cancer Control in, Countries, in *Cancer Control Opportunities in Low- and Middle-Income Countries* (Eds.: F. A. Sloan, H. Gelband), National Academies Press (US), Washington (DC), **2007**.
[52] S. H. Hassanpour, M. Dehghani, *Journal of Cancer Research and Practice* **2017**, *4*, 127-129.
[53] F. Bray, A. Jemal, N. Grey, J. Ferlay, D. Forman, *Lancet Oncol.* **2012**, *13*, 790-801.
[54] R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, *CA Cancer J. Clin.* **2022**, *72*, 7-33.
[55] H. Gelband, R. Sankaranarayanan, C. L. Gauvreau, S. Horton, B. O. Anderson, F. Bray, J. Cleary, A. J. Dare, L. Denny, M. K. Gospodarowicz*, et al.*, *Lancet* **2016**, *387*, 2133-2144.
[56] P. Farmer, J. Frenk, F. M. Knaul, L. N. Shulman, G. Alleyne, L. Armstrong, R. Atun, D. Blayney, L. Chen, R. Feachem*, et al.*, *Lancet* **2010**, *376*, 1186-1193.
[57] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, *CA Cancer J. Clin.* **2021**, *71*, 209-249.
[58] J. Ferlay, M. Colombet, I. Soerjomataram, D. M. Parkin, M. Pineros, A. Znaor, F. Bray, *Int. J. Cancer* **2021**, *149*, 778-789.
[59] A. Noone, N. Howlader, M. Krapcho, D. Miller, A. Brest, M. Yu, J. Ruhl, M. R. Yu, J. , Z. Tatalovich, A. Mariotto*, et al.*, National Cancer Institute, Bethesda, MD, **2018**.

[60] K. Kasaian, Y. Y. Li, S. J. M. Jones, in *Cancer Genomics* (Eds.: G. Dellaire, J. N. Berman, R. J. Arceci), Academic Press, Boston, **2014**, pp. 133-152.
[61] B. Nograd*y, Vol. 579*, Nature, **2020**, pp. 11-12.
[62] M. F. Berger, E. R. Mardis, *Nat. Rev. Clin. Oncol.* **2018**, *15*, 353-365.
[63] D. Cross, J. K. Burmester, *Clin Med Res* **2006**, *4*, 218-227.
[64] S. K. Das, M. E. Menezes, S. Bhatia, X. Y. Wang, L. Emdad, D. Sarkar, P. B. Fisher, *J. Cell. Physiol.* **2015**, *230*, 259-271.
[65] J. Liu, M. Fu, M. Wang, D. Wan, Y. Wei, X. Wei, *J. Hematol. Oncol.* **2022**, *15*, 28.
[66] S. M. Beecher, D. P. O'Leary, R. McLaughlin, M. J. Kerin, *Oncol Res Treat* **2018**, *41*, 478-482.
[67] C. Deville, G. Shukla, R. Rengan, C. R. Thomas, in *The American Cancer Society's Principles of Oncology*, **2017**, pp. 204-219.
[68] L. H. Hartwell, M. B. Kastan, *Science* **1994**, *266*, 1821-1828.
[69] A. D. N. J. de Grey, M. J. Rae, in *Encyclopedia of Gerontology and Population Aging* (Eds.: D. Gu, M. E. Dupre), Springer International Publishing, Cham, **2021**, pp. 877-887.
[70] L. H. Pearl, A. C. Schierz, S. E. Ward, B. Al-Lazikani, F. M. Pearl, *Nat. Rev. Cancer* **2015**, *15*, 166-180.
[71] A. Cerrato, F. Morra, A. Celetti, *J. Exp. Clin. Cancer Res.* **2016**, *35*, 179.
[72] G. H. Williams, K. Stoeber, *J. Pathol.* **2012**, *226*, 352-364.
[73] M. Ter Huurne, H. G. Stunnenberg, *Cell Mol Life Sci* **2021**, *78*, 4507-4519.
[74] J. Massague, *Nature* **2004**, *432*, 298-306.
[75] G. M. Cooper, *The Cell: A Molecular Approach. 2nd edition*, Sinauer Associates 2000, **2000**.
[76] B. A. Sullivan, in *Brenner's Encyclopedia of Genetics* (Eds.: S. Maloy, K. Hughes), Academic Press, San Diego, **2013**, pp. 90-93.
[77] M. T. Amjad, A. Chidharla, A. Kasi, in *StatPearls*, StatPearls Publishing, Treasure Island (FL), **2022**.
[78] B. T. Oronsky, T. Reid, S. J. Knox, J. J. Scicinski, *Transl. Oncol.* **2012**, *5*, 226-229.
[79] D. P. Kufe, Raphael, R. Weichselbaum, *Holland-Frei Cancer Medicine. 6th edition*, BC Decker 2003, **2003**.
[80] R. T. Dorr, *Semin. Oncol.* **1992**, *19*, 3-8.
[81] S. M. Hecht, *J. Nat. Prod.* **2000**, *63*, 158-168.
[82] R. S. Coleman, R. J. Perez, C. H. Burk, A. Navarro, *J. Am. Chem. Soc.* **2002**, *124*, 13008-13017.
[83] in *LiverTox: Clinical and Research Information on Drug-Induced Liver Injury*, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda (MD), **2012**.
[84] J. Humeau, A. Sauvat, G. Cerrato, W. Xie, F. Loos, F. Iannantuoni, L. Bezu, S. Levesque, J. Paillet, J. Pol*, et al.*, *EMBO Mol. Med.* **2020**, *12*, e11622.
[85] H. M. Sobell, *Proc Natl Acad Sci U S A* **1985**, *82*, 5328-5331.
[86] V. Cermak, V. Dostal, M. Jelinek, L. Libusova, J. Kovar, D. Rosel, J. Brabek, *Eur. J. Cell Biol.* **2020**, *99*, 151075.
[87] M. O. Steinmetz, A. E. Prota, *Trends Cell Biol.* **2018**, *28*, 776-792.
[88] L. Wordeman, J. J. Vicente, *Cancers (Basel)* **2021**, *13*.
[89] M. A. Jordan, *Current Medicinal Chemistry - Anti-Cancer Agents* **2002**, *2*, 1-17.
[90] P. Dhyani, C. Quispe, E. Sharma, A. Bahukhandi, P. Sati, D. C. Attri, A. Szopa, J. Sharifi-Rad, A. O. Docea, I. Mardare*, et al.*, *Cancer Cell Int.* **2022**, *22*, 206.
[91] A. Haque, M. A. Rahman, M. S. H. Faizi, M. S. Khan, *Curr. Med. Chem.* **2018**, *25*, 1650-1662.
[92] J. Gallego-Jara, G. Lozano-Terol, R. A. Sola-Martinez, M. Canovas-Diaz, T. de Diego Puente, *Molecules* **2020**, *25*.
[93] M. C. Wani, H. L. Taylor, M. E. Wall, P. Coggon, A. T. McPhail, *J. Am. Chem. Soc.* **1971**, *93*, 2325-2327.
[94] B. K. Sinha, *Drugs* **1995**, *49*, 11-19.

[95] X. Liang, Q. Wu, S. Luan, Z. Yin, C. He, L. Yin, Y. Zou, Z. Yuan, L. Li, X. Song*, et al.*, *Eur J Med Chem* **2019**, *171*, 129-168.
[96] M. Tiwari, *J. Cancer Res. Ther.* **2012**, *8*, 510-519.
[97] G. J. Peters, *Nucleosides Nucleotides Nucleic Acids* **2014**, *33*, 358-374.
[98] C. Michaelis, R. Ciosk, K. Nasmyth, *Cell* **1997**, *91*, 35-45.
[99] R. A. Oliveira, R. S. Hamilton, A. Pauli, I. Davis, K. Nasmyth, *Nat. Cell Biol.* **2010**, *12*, 185-192.
[100] T. H. Cheffings, N. J. Burroughs, M. K. Balasubramanian, *Curr. Biol.* **2016**, *26*, R719-R737.
[101] Y. Sun, Y. Liu, X. Ma, H. Hu, *Int. J. Mol. Sci.* **2021**, *22*.
[102] H. Hochegger, S. Takeda, T. Hunt, *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 910-916.
[103] W. Lin, G. Arthur, *Int J Biochem Cell Biol* **2007**, *39*, 597-605.
[104] A. Musacchio, E. D. Salmon, *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 379-393.
[105] J. Fulka, Jr., N. L. First, J. Fulka, R. M. Moor, *Hum. Reprod.* **1999**, *14*, 1582-1587.
[106] Y. L. Shih, L. Rothfield, *Microbiol Mol Biol Rev* **2006**, *70*, 729-754.
[107] A. Berepiki, A. Lichius, N. D. Read, *Nat. Rev. Microbiol.* **2011**, *9*, 876-887.
[108] D. A. Fletcher, R. D. Mullins, *Nature* **2010**, *463*, 485-492.
[109] S. L. Kline-Smith, C. E. Walczak, *Mol. Cell* **2004**, *15*, 317-327.
[110] A. J. Zwetsloot, G. Tut, A. Straube, *Essays Biochem.* **2018**, *62*, 725-735.
[111] T. D. Pollard, R. D. Goldman, *Cold Spring Harb. Perspect. Biol.* **2018**, *10*.
[112] D. W. Buster, D. H. Baird, W. Yu, J. M. Solowska, M. Chauviere, A. Mazurek, M. Kress, P. W. Baas, *J. Neurocytol.* **2003**, *32*, 79-96.
[113] H. Schmidt, R. Zalyte, L. Urnavicius, A. P. Carter, *Nature* **2015**, *518*, 435-438.
[114] W. Wang, L. Cao, C. Wang, B. Gigant, M. Knossow, *Protein Sci.* **2015**, *24*, 1047-1056.
[115] N. Hirokawa, Y. Noda, Y. Tanaka, S. Niwa, *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 682-696.
[116] S. Karki, E. L. Holzbaur, *Curr. Opin. Cell Biol.* **1999**, *11*, 45-53.
[117] M. E. Tanenbaum, L. Macůrek, A. Janssen, E. F. Geers, M. Alvarez-Fernández, R. H. Medema, *Current Biology* **2009**, *19*, 1703-1711.
[118] R. G. van Heesbeen, M. E. Tanenbaum, R. H. Medema, *Cell Rep.* **2014**, *8*, 948-956.
[119] T. U. Mayer, T. M. Kapoor, *Trends in Cell Biology* **2000**, *10*, 88.
[120] T. U. Mayer, T. M. Kapoor, S. J. Haggarty, R. W. King, S. L. Schreiber, T. J. Mitchison, *Science* **1999**, *286*, 971-974.
[121] T. M. Kapoor, T. U. Mayer, M. L. Coughlin, T. J. Mitchison, *The Journal of Cell Biology* **2000**, *150*, 975-988.
[122] Z. Maliga, T. M. Kapoor, T. J. Mitchison, *Chem. Biol.* **2002**, *9*, 989-996.
[123] R. Sainath, G. Gallo, *Dev. Neurobiol.* **2015**, *75*, 757-777.
[124] D. H. Roossien, K. E. Miller, G. Gallo, *Front. Cell. Neurosci.* **2015**, *9*, 252.
[125] M. E. Tanenbaum, R. H. Medema, *Developmental Cell* **2010**, *19*, 797-806.
[126] G. K. Gouras, in *Reference Module in Biomedical Sciences*, Elsevier, **2014**.
[127] G. Cipriani, C. Dolciotti, L. Picchi, U. Bonuccelli, *Neurol. Sci.* **2011**, *32*, 275-279.
[128] S. Benjamin, L. MacGillivray, B. Schildkrout, A. Cohen-Oram, M. D. Lauterbach, L. L. Levin, *J. Neuropsychiatry Clin. Neurosci.* **2018**, *30*, 279-290.
[129] P. E. Fraser, J. T. Nguyen, W. K. Surewicz, D. A. Kirschner, *Biophys. J.* **1991**, *60*, 1190-1201.
[130] C. J. Pike, D. Burdick, A. J. Walencewicz, C. G. Glabe, C. W. Cotman, *J. Neurosci.* **1993**, *13*, 1676-1687.
[131] C. Priller, T. Bauer, G. Mitteregger, B. Krebs, H. A. Kretzschmar, J. Herms, *J. Neurosci.* **2006**, *26*, 7212-7221.
[132] P. R. Turner, K. O'Connor, W. P. Tate, W. C. Abraham, *Prog. Neurobiol.* **2003**, *70*, 1-32.
[133] J. A. Duce, A. Tsatsanis, M. A. Cater, S. A. James, E. Robb, K. Wikhe, S. L. Leong, K. Perez, T. Johanssen, M. A. Greenough*, et al.*, *Cell* **2010**, *142*, 857-867.

[134] S. M. Ohline, C. Chan, L. Schoderboeck, H. E. Wicky, W. P. Tate, S. M. Hughes, W. C. Abraham, *Mol. Brain* **2022**, *15*, 5.
[135] N. J. Corbett, N. M. Hooper, in *Biochemical and Biophysical Roles of Cell Surface Molecules* (Eds.: K. Chattopadhyay, S. C. Basu), Springer Singapore, Singapore, **2018**, pp. 177-183.
[136] H. Zheng, E. H. Koo, *Mol. Neurodegener.* **2006**, *1*, 5.
[137] F. Dulin, F. Leveille, J. B. Ortega, J. P. Mornon, A. Buisson, I. Callebaut, N. Colloc'h, *FEBS Lett.* **2008**, *582*, 1865-1870.
[138] A. J. Kuhn, B. S. Abrams, S. Knowlton, J. A. Raskatov, *ACS Chem. Neurosci.* **2020**, *11*, 1539-1544.
[139] F. Chen, H. Hasegawa, G. Schmitt-Ulms, T. Kawarai, C. Bohm, T. Katayama, Y. Gu, N. Sanjo, M. Glista, E. Rogaeva*, et al.*, *Nature* **2006**, *440*, 1208-1212.
[140] R. Kayed, E. Head, J. L. Thompson, T. M. McIntire, S. C. Milton, C. W. Cotman, C. G. Glabe, *Science* **2003**, *300*, 486-489.
[141] M. Sakono, T. Zako, *FEBS J.* **2010**, *277*, 1348-1358.
[142] N. Yamamoto, E. Matsubara, S. Maeda, H. Minagawa, A. Takashima, W. Maruyama, M. Michikawa, K. Yanagisawa, *J. Biol. Chem.* **2007**, *282*, 2646-2655.
[143] F. G. De Felice, P. T. Velasco, M. P. Lambert, K. Viola, S. J. Fernandez, S. T. Ferreira, W. L. Klein, *J. Biol. Chem.* **2007**, *282*, 11590-11601.
[144] G. M. Shankar, B. L. Bloodgood, M. Townsend, D. M. Walsh, D. J. Selkoe, B. L. Sabatini, *J. Neurosci.* **2007**, *27*, 2866-2875.
[145] M. H. Magdesian, M. M. Carvalho, F. A. Mendes, L. M. Saraiva, M. A. Juliano, L. Juliano, J. Garcia-Abreu, S. T. Ferreira, *J. Biol. Chem.* **2008**, *283*, 9359-9368.
[146] D. W. Cleveland, S.-Y. Hwo, M. W. Kirschner, *Journal of Molecular Biology* **1977**, *116*, 207-225.
[147] L. Dehmelt, S. Halpain, *Genome Biol.* **2005**, *6*, 204.
[148] K. Iqbal, C. Alonso Adel, S. Chen, M. O. Chohan, E. El-Akkad, C. X. Gong, S. Khatoon, B. Li, F. Liu, A. Rahman*, et al.*, *Biochim Biophys Acta* **2005**, *1739*, 198-210.
[149] J. Z. Wang, I. Grundke-Iqbal, K. Iqbal, *Eur. J. Neurosci.* **2007**, *25*, 59-68.
[150] J. Cummings, G. Lee, A. Ritter, M. Sabbagh, K. Zhong, *Alzheimers Dement (N Y)* **2020**, *6*, e12050.
[151] G. C. Alexander, S. Emerson, A. S. Kesselheim, *JAMA* **2021**, *325*, 1717-1718.
[152] F. Panza, M. Lozupone, G. Logroscino, B. P. Imbimbo, *Nat. Rev. Neurol.* **2019**, *15*, 73-88.
[153] K. G. Yiannopoulou, S. G. Papageorgiou, *J Cent Nerv Syst Dis* **2020**, *12*, 1179573520907397.
[154] H.-J. Schiewe, Dissertation thesis, Georg-August-Universität zu Göttingen **1997**.
[155] L. Hoffmann, Dissertation thesis, Georg-August-Universität zu Göttingen **2006**.
[156] J.-N. Fricke, Dissertation thesis, Eberhard Karls Universität Tübingen (Tübingen), **2012**.
[157] Y. Kwon, J. Shin, K. Nam, J. S. An, S. H. Yang, S. H. Hong, M. Bae, K. Moon, Y. Cho, J. Woo*, et al.*, *Angew Chem Int Ed Engl* **2020**, *59*, 22994-22998.
[158] Z. Gáspári, A. Perczel, in *Annu. Rep. NMR Spectrosc., Vol. 71* (Ed.: G. A. Webb), Academic Press, **2010**, pp. 35-75.
[159] J. P. Torres, Z. Lin, J. M. Winter, P. J. Krug, E. W. Schmidt, *Nat Commun* **2020**, *11*, 2882.
[160] J. Staunton, K. J. Weissman, *Nat Prod Rep* **2001**, *18*, 380-416.
[161] J. Moschny, W. Lorenzen, A. Hilfer, R. Eckenstaler, S. Jahns, H. Enke, D. Enke, P. Schneider, R. A. Benndorf, T. H. J. Niedermeyer, *J. Nat. Prod.* **2020**, *83*, 1960-1970.
[162] C. R. Huitt-Roehl, E. A. Hill, M. M. Adams, A. L. Vagstad, J. W. Li, C. A. Townsend, *ACS Chem. Biol.* **2015**, *10*, 1443-1449.
[163] J. F. Barajas, J. M. Blake-Hedges, C. B. Bailey, S. Curran, J. D. Keasling, *Synthetic and Systems Biotechnology* **2017**, *2*, 147-166.
[164] in *Modern Tools for the Synthesis of Complex Bioactive Molecules*.

[165] D. E. Cane, F. Kudo, K. Kinoshita, C. Khosla, *Chem. Biol.* **2002**, *9*, 131-142.
[166] J. M. Risley, R. L. Van Etten, *Journal of the American Chemical Society* **2002**, *101*, 252-253.
[167] B. Zhang, K. B. Wang, W. Wang, X. Wang, F. Liu, J. Zhu, J. Shi, L. Y. Li, H. Han, K. Xu*, et al.*, *Nature* **2019**, *568*, 122-126.
[168] C. S. McCaughey, J. A. van Santen, J. J. J. van der Hooft, M. H. Medema, R. G. Linington, *Nat. Chem. Biol.* **2022**, *18*, 295-304.
[169] E. B. Sterndorff, T. S. Jorgensen, J. F. G. Garzon, J. Schmid, S. Grond, T. Weber, *Microbiol Resour Announc* **2022**, *11*, e0022022.
[170] P. S. Starcher, B. Phillips, *Journal of the American Chemical Society* **2002**, *80*, 4079-4082.
[171] G. R. Krow, in *Organic Reactions*, **2004**, pp. 251-798.
[172] F. W. Forney, A. J. Markovetz, R. E. Kallio, *J. Bacteriol.* **1967**, *93*, 649-655.
[173] M. J. L. J. Fürst, A. Gran-Scheuch, F. S. Aalbers, M. W. Fraaije, *ACS Catalysis* **2019**, *9*, 11207-11241.
[174] S. M. McGuire, C. A. Townsend, *Bioorg. Med. Chem. Lett.* **1993**, *3*, 653-656.
[175] H. Zeng, G. Yin, Q. Wei, D. Li, Y. Wang, Y. Hu, C. Hu, Y. Zou, *Angew Chem Int Ed Engl* **2019**, *58*, 6569-6573.
[176] J. Jiang, C. N. Tetzlaff, S. Takamatsu, M. Iwatsuki, M. Komatsu, H. Ikeda, D. E. Cane, *Biochemistry* **2009**, *48*, 6431-6440.
[177] K. H. Jones, R. T. Smith, P. W. Trudgill, *J. Gen. Microbiol.* **1993**, *139*, 797-805.
[178] T. Nomura, T. Kushiro, T. Yokota, Y. Kamiya, G. J. Bishop, S. Yamaguchi, *J. Biol. Chem.* **2005**, *280*, 17873-17879.
[179] T. W. Kim, J. Y. Hwang, Y. S. Kim, S. H. Joo, S. C. Chang, J. S. Lee, S. Takatsuto, S. K. Kim, *Plant Cell* **2005**, *17*, 2397-2412.
[180] D. Vollmar, PhD thesis, Georg-August-Universität Göttingen **2010**.
[181] H. B. Bode, B. Bethe, R. Hofs, A. Zeeck, *Chembiochem* **2002**, *3*, 619-627.
[182] H. J. Schiewe, A. Zeeck, *J Antibiot (Tokyo)* **1999**, *52*, 635-642.
[183] J. C. Navarro-Munoz, N. Selem-Mojica, M. W. Mullowney, S. A. Kautsar, J. H. Tryon, E. I. Parkinson, E. L. C. De Los Santos, M. Yeong, P. Cruz-Morales, S. Abubucker*, et al.*, *Nat. Chem. Biol.* **2020**, *16*, 60-68.
[184] Y. Tong, C. M. Whitford, H. L. Robertsen, K. Blin, T. S. Jorgensen, A. K. Klitgaard, T. Gren, X. Jiang, T. Weber, S. Y. Lee, *Proc Natl Acad Sci U S A* **2019**, *116*, 20366-20375.
[185] M. Balerna, W. Keller-Schierlein, C. Martius, H. Wolf, H. Z**?**hner, *Archiv f**?**r Mikrobiologie* **1969**, *65*, 303-317.
[186] Y. Wu, Q. Kang, Y. Shen, W. Su, L. Bai, *Mol. Biosyst.* **2011**, *7*, 2459-2469.
[187] Q. Kang, Y. Shen, L. Bai, *Nat Prod Rep* **2012**, *29*, 243-263.
[188] M. W. Fraaije, N. M. Kamerbeek, W. J. van Berkel, D. B. Janssen, *FEBS Lett.* **2002**, *518*, 43-47.
[189] A. Riebel, H. M. Dudek, G. de Gonzalo, P. Stepniak, L. Rychlewski, M. W. Fraaije, *Appl Microbiol Biotechnol* **2012**, *95*, 1479-1489.
[190] F. Surup, O. Wagner, J. von Frieling, M. Schleicher, S. Oess, P. Müller, S. Grond, *The Journal of Organic Chemistry* **2007**, *72*, 5085-5090.
[191] J. C. Schmid, Master thesis, Eberhard Karls Universität Tübingen **2018**.
[192] X. Ji, J. Tu, Y. Song, C. Zhang, L. Wang, Q. Li, J. Ju, *ACS Catalysis* **2020**, *10*, 2591-2595.
[193] J. Tu, S. Li, J. Chen, Y. Song, S. Fu, J. Ju, Q. Li, *Microb Cell Fact* **2018**, *17*, 28.
[194] W. Jiang, X. Zhao, T. Gabrieli, C. Lou, Y. Ebenstein, T. F. Zhu, *Nat Commun* **2015**, *6*, 8101.
[195] W. Jiang, T. F. Zhu, *Nat. Protoc.* **2016**, *11*, 960-975.
[196] W. Wang, X. Li, J. Wang, S. Xiang, X. Feng, K. Yang, *Appl Environ Microbiol* **2013**, *79*, 4484-4492.
[197] R. Kind, Dissertation thesis, Georg-August-Universität zu Göttingen **1991**.
[198] P. L. Wong, A. J. Royce, C. W. T. Lee-Parsons, *Biochem. Eng. J.* **2004**, *21*, 253-258.

[199] J. C. Lee, H. R. Park, D. J. Park, H. B. Lee, Y. B. Kim, C. J. Kim, *Lett. Appl. Microbiol.* **2003**, *37*, 196-200.
[200] M. P. Singh, M. M. Leighton, L. R. Barbieri, D. M. Roll, S. E. Urbance, L. Hoshan, L. A. McDonald, *J Ind Microbiol Biotechnol* **2010**, *37*, 335-340.
[201] M. D. Shepherd, M. K. Kharel, M. A. Bosserman, J. Rohr, *Curr. Protoc. Microbiol.* **2010**, *Chapter 10*, Unit 10E 11.
[202] G. Ruth, **2013**.
[203] G. Loos, A. Van Schepdael, D. Cabooter, *Philos Trans A Math Phys Eng Sci* **2016**, *374*, 20150366.
[204] J. J. Zhang, X. Tang, T. Huan, A. C. Ross, B. S. Moore, *Nat. Chem. Biol.* **2020**, *16*, 42-49.
[205] L. Martinet, A. Naome, B. Deflandre, M. Maciejewska, D. Tellatin, E. Tenconi, N. Smargiasso, E. de Pauw, G. P. van Wezel, S. Rigali, *mBio* **2019**, *10*.
[206] H. Itoh, M. Inoue, *Chem Rev* **2019**, *119*, 10002-10031.
[207] A. D. Bond, *Resonance* **2015**, *19*, 1087-1092.
[208] J. M. Bijvoet, A. F. Peerdeman, A. J. van Bommel, *Nature* **1951**, *168*, 271-272.
[209] N. Prileschajew, *Berichte der deutschen chemischen Gesellschaft* **2006**, *42*, 4811-4815.
[210] R. Zhang, X. Qin, F. Kong, P. Chen, G. Pan, *Drug Deliv.* **2019**, *26*, 328-342.
[211] L. D. Lavis, *ACS Chem. Biol.* **2008**, *3*, 203-206.
[212] B. Neises, W. Steglich, *Angewandte Chemie International Edition in English* **1978**, *17*, 522-524.
[213] H. Gilman, R. G. Jones, L. A. Woods, *The Journal of Organic Chemistry* **2002**, *17*, 1630-1634.
[214] L. E. Loffler, C. Wirtz, A. Furstner, *Angew Chem Int Ed Engl* **2021**, *60*, 5316-5322.
[215] P. Chiu, Z. Li, K. C. M. Fung, *Tetrahedron Letters* **2003**, *44*, 455-457.
[216] B. M. Johnson, Y. Z. Shu, X. Zhuo, N. A. Meanwell, *J. Med. Chem.* **2020**, *63*, 6315-6386.
[217] M. Baumann, I. R. Baxendale, S. V. Ley, *Synlett* **2008**, *2008*, 2111-2114.
[218] P. Thangaraj, *Pharmacological Assays of Plant-Based Natural Products*, Springer Cham, **2016**.
[219] A. A. Bulich, D. Isenberg, *ISA Trans.* **1981**, *20*, 29-33.
[220] T. L. Riss, R. A. Moravec, A. L. Niles, S. Duellman, H. A. Benink, T. J. Worzella, L. Minor, in *Assay Guidance Manual* (Eds.: S. Markossian, A. Grossman, K. Brimacombe, M. Arkin, D. Auld, C. Austin, J. Baell, T. D. Y. Chung, N. P. Coussens, J. L. Dahlin, V. Devanarayan, T. L. Foley, M. Glicksman, J. V. Haas, M. D. Hall, S. Hoare, J. Inglese, P. W. Iversen, S. C. Kales, M. Lal-Nag, Z. Li, J. McGee, O. McManus, T. Riss, P. Saradjian, G. S. Sittampalam, M. Tarselli, O. J. Trask, Jr., Y. Wang, J. R. Weidner, M. J. Wildey, K. Wilson, M. Xia, X. Xu), Eli Lilly & Company and the National Center for Advancing Translational Sciences, Bethesda (MD), **2004**.
[221] S. Cole, *Cancer chemotherapy and pharmacology* **1986**, *17*, 259-263.
[222] T. Mosmann, *J. Immunol. Methods* **1983**, *65*, 55-63.
[223] T. Decker, M. L. Lohmann-Matthes, *J. Immunol. Methods* **1988**, *115*, 61-69.
[224] L. Sliwka, K. Wiktorska, P. Suchocki, M. Milczarek, S. Mielczarek, K. Lubelska, T. Cierpial, P. Lyzwa, P. Kielbasinski, A. Jaromin*, et al.*, *PLoS One* **2016**, *11*, e0155772.
[225] S. Kamiloglu, G. Sari, T. Ozdal, E. Capanoglu, *Food Frontiers* **2020**, *1*, 332-349.
[226] Z. Jin, W. S. El-Deiry, *Cancer Biol. Ther.* **2005**, *4*, 139-163.
[227] J. G. Nirmala, M. Lopus, *Cell Biol Toxicol* **2020**, *36*, 145-164.
[228] L. Galluzzi, I. Vitale, S. A. Aaronson, J. M. Abrams, D. Adam, P. Agostinis, E. S. Alnemri, L. Altucci, I. Amelio, D. W. Andrews*, et al.*, *Cell Death Differ* **2018**, *25*, 486-541.

[229] L. de Abreu Costa, M. Henrique Fernandes Ottoni, M. G. Dos Santos, A. B. Meireles, V. Gomes de Almeida, W. de Fatima Pereira, B. Alves de Avelar-Freitas, G. Eustaquio Alvim Brito-Melo, *Molecules* **2017**, *22*, 1789.
[230] A. K. Buck, D. E. Elmore, L. E. Darling, *Future Med. Chem.* **2019**, *11*, 2445-2458.
[231] M. A. Morris, A. Vallmitjana, F. Grein, T. Schneider, M. Arts, C. R. Jones, B. T. Nguyen, M. H. Hashemian, M. Malek, E. Gratton*, et al.*, *Chem. Sci.* **2022**, *13*, 7747-7754.
[232] A. B. Schafer, M. Wenzel, *Front Cell Infect Microbiol* **2020**, *10*, 540898.
[233] S. Vira, E. Mekhedov, G. Humphrey, P. S. Blank, *Anal. Biochem.* **2010**, *402*, 146-150.
[234] K. H. Walen, S. W. Brown, *Nature* **1962**, *194*, 406.
[235] G. Banfalvi, *Methods Mol. Biol.* **2011**, *761*, 1-23.
[236] V. Roukos, G. Pegoraro, T. C. Voss, T. Misteli, *Nat. Protoc.* **2015**, *10*, 334-348.
[237] C. McQuin, A. Goodman, V. Chernyshev, L. Kamentsky, B. A. Cimini, K. W. Karhohs, M. Doan, L. Ding, S. M. Rafelski, D. Thirstrup*, et al.*, *PLoS Biol.* **2018**, *16*, e2005970.
[238] A. E. Granada, A. Jimenez, J. Stewart-Ornstein, N. Bluthgen, S. Reber, A. Jambhekar, G. Lahav, *Mol. Biol. Cell* **2020**, *31*, 845-857.
[239] J. Bai, Y. Li, G. Zhang, *Cancer Biol Med* **2017**, *14*, 348-362.
[240] K. Itaya, M. Ui, *Clin. Chim. Acta* **1966**, *14*, 361-366.
[241] T. Kodama, K. Fukui, K. Kometani, *J. Biochem.* **1986**, *99*, 1465-1472.
[242] T. Horio, T. Murata, *Front Plant Sci* **2014**, *5*, 511.
[243] P. Q. Farshori, D. Goode, *J. Submicrosc. Cytol. Pathol.* **1994**, *26*, 137-146.
[244] P. Froissard, D. Duval, *Neurochem. Int.* **1994**, *24*, 485-493.
[245] A. Pompella, A. Visvikis, A. Paolicchi, V. De Tata, A. F. Casini, *Biochem. Pharmacol.* **2003**, *66*, 1499-1503.
[246] S. Tan, D. Schubert, P. Maher, *Curr Top Med Chem* **2001**, *1*, 497-506.
[247] P. Maher, A. Currais, D. Schubert, *Cell Chem Biol* **2020**, *27*, 1456-1471.
[248] J. B. Davis, P. Maher, *Brain Res.* **1994**, *652*, 169-173.
[249] H. Khan, H. Ullah, M. Aschner, W. S. Cheang, E. K. Akkol, *Biomolecules* **2019**, *10*.
[250] J. Cummings, P. S. Aisen, B. DuBois, L. Frolich, C. R. Jack, Jr., R. W. Jones, J. C. Morris, J. Raskin, S. A. Dowsett, P. Scheltens, *Alzheimers Res Ther* **2016**, *8*, 39.
[251] C. Xue, T. Y. Lin, D. Chang, Z. Guo, *R Soc Open Sci* **2017**, *4*, 160696.
[252] A. Ugarte, D. Corbacho, M. S. Aymerich, A. Garcia-Osta, M. Cuadrado-Tejedor, J. Oyarzabal, *Neurotherapeutics* **2018**, *15*, 742-750.
[253] D. R. Liston, M. Davis, *Clin Cancer Res* **2017**, *23*, 3489-3498.
[254] S. A. Hudson, H. Ecroyd, T. W. Kee, J. A. Carver, *FEBS J.* **2009**, *276*, 5960-5972.
[255] Y. Zhang, D. Zhang, Y. Tang, B. Ren, F. Liu, L. Xu, Y. Chang, J. Zheng, *Materials Advances* **2020**, *1*, 1241-1252.
[256] D. Ghosh, S. Samanta, T. Govindaraju, *Bull. Mater. Sci.* **2020**, *43*, 309.
[257] W. B. Stine, L. Jungbauer, C. Yu, M. J. LaDu, in *Alzheimer's Disease and Frontotemporal Dementia: Methods and Protocols* (Ed.: E. D. Roberson), Humana Press, Totowa, NJ, **2011**, pp. 13-32.
[258] P. N. Nirmalraj, J. List, S. Battacharya, G. Howe, L. Xu, D. Thompson, M. Mayer, *Sci Adv* **2020**, *6*, eaaz6014.
[259] N. L. Anderson, N. G. Anderson, *Electrophoresis* **1998**, *19*, 1853-1861.
[260] W. P. Blackstock, M. P. Weir, *Trends Biotechnol.* **1999**, *17*, 121-127.
[261] C. H. Johnson, J. Ivanisevic, G. Siuzdak, *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 451-459.
[262] D. S. Wishart, *Physiol. Rev.* **2019**, *99*, 1819-1875.
[263] J. Paananen, V. Fortino, *Brief. Bioinform.* **2020**, *21*, 1937-1953.
[264] T. Pluskal, S. Castillo, A. Villar-Briones, M. Oresic, *BMC Bioinformatics* **2010**, *11*, 395.

[265] H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. VanderGheynst, O. Fiehn, M. Arita, *Nat. Methods* **2015**, *12*, 523-526.
[266] J. Xia, N. Psychogios, N. Young, D. S. Wishart, *Nucleic Acids Res* **2009**, *37*, W652-660.
[267] Z. Pang, G. Zhou, J. Ewald, L. Chang, O. Hacariz, N. Basu, J. Xia, *Nat. Protoc.* **2022**, *17*, 1735-1761.
[268] D. Merkel, *Linux J.* **2014**, *2014*, Article 2.
[269] S. Kim, *Curr Protoc* **2021**, *1*, e217.
[270] M. Ludwig, M. Fleischauer, K. Dührkop, M. A. Hoffmann, S. Böcker, in *Computational Methods and Data Analysis for Metabolomics* (Ed.: S. Li), Springer US, New York, NY, **2020**, pp. 185-207.
[271] M. Sorokina, C. Steinbeck, *J. Cheminform.* **2019**, *11*, 55.
[272] K. V. Jayaseelan, P. Moreno, A. Truszkowski, P. Ertl, C. Steinbeck, *BMC Bioinformatics* **2012**, *13*, 106.
[273] P. D. Dobson, Y. Patel, D. B. Kell, *Drug Discov. Today* **2009**, *14*, 31-40.
[274] M. Sorokina, C. Steinbeck, *J. Cheminform.* **2020**, *12*, 20.
[275] M. Sorokina, P. Merseburger, K. Rajan, M. A. Yirik, C. Steinbeck, *J. Cheminform.* **2021**, *13*, 2.
[276] P. de Matos, R. Alcantara, A. Dekker, M. Ennis, J. Hastings, K. Haug, I. Spiteri, S. Turner, C. Steinbeck, *Nucleic Acids Res* **2010**, *38*, D249-254.
[277] K. Degtyarenko, P. de Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcantara, M. Darsow, M. Guedj, M. Ashburner, *Nucleic Acids Res* **2008**, *36*, D344-350.
[278] A. Rutz, M. Sorokina, J. Galgonek, D. Mietchen, E. Willighagen, A. Gaudry, J. G. Graham, R. Stephan, R. Page, J. Vondrášek*, et al.*, *bioRxiv* **2021**, 2021.2002.2028.433265.
[279] J. A. van Santen, E. F. Poynton, D. Iskakova, E. McMann, T. A. Alsup, T. N. Clark, C. H. Fergusson, D. P. Fewer, A. H. Hughes, C. A. McCadden*, et al.*, *Nucleic Acids Res* **2022**, *50*, D1317-D1323.
[280] J. A. van Santen, G. Jacob, A. L. Singh, V. Aniebok, M. J. Balunas, D. Bunsko, F. C. Neto, L. Castano-Espriu, C. Chang, T. N. Clark*, et al.*, *ACS Cent Sci* **2019**, *5*, 1824-1833.
[281] CRC, *Dictionary of Natural Products, Vol. 30.1*, Chapman & Hall/CRC, **2022**.
[282] S. Heller, A. McNaught, S. Stein, D. Tchekhovskoi, I. Pletnev, *J. Cheminform.* **2013**, *5*, 7.
[283] J. McConathy, M. J. Owens, *Prim. Care Companion J. Clin. Psychiatry* **2003**, *5*, 70-73.
[284] J. Caldwell, *J. Clin. Pharmacol.* **1992**, *32*, 925-929.
[285] N. B. Cech, M. H. Medema, J. Clardy, *Nat Prod Rep* **2021**, *38*, 1947-1953.
[286] S. A. Jarmusch, J. J. J. van der Hooft, P. C. Dorrestein, A. K. Jarmusch, *Nat Prod Rep* **2021**, *38*, 2066-2082.
[287] A. C. Schrimpe-Rutledge, S. G. Codreanu, S. D. Sherrod, J. A. McLean, *J Am Soc Mass Spectrom* **2016**, *27*, 1897-1905.
[288] X. Domingo-Almenara, C. Guijas, E. Billings, J. R. Montenegro-Burke, W. Uritboonthai, A. E. Aisporna, E. Chen, H. P. Benton, G. Siuzdak, *Nat Commun* **2019**, *10*, 5811.
[289] D. Bajusz, A. Racz, K. Heberger, *J. Cheminform.* **2015**, *7*, 20.
[290] T. Tanimoto, in *IBM Internal Report*, IBM, **1958**.
[291] G. Maggiora, M. Vogt, D. Stumpfe, J. Bajorath, *J. Med. Chem.* **2014**, *57*, 3186-3204.
[292] E. L. Willighagen, J. W. Mayfield, J. Alvarsson, A. Berg, L. Carlsson, N. Jeliazkova, S. Kuhn, T. Pluskal, M. Rojas-Cherto, O. Spjuth*, et al.*, *J. Cheminform.* **2017**, *9*, 33.
[293] C. Steinbeck, Y. Han, S. Kuhn, O. Horlacher, E. Luttmann, E. Willighagen, *J Chem Inf Comput Sci* **2003**, *43*, 493-500.

[294] H. W. Kim, M. Wang, C. A. Leber, L. F. Nothias, R. Reher, K. B. Kang, J. J. J. van der Hooft, P. C. Dorrestein, W. H. Gerwick, G. W. Cottrell, *J. Nat. Prod.* **2021**, *84*, 2795-2807.
[295] P. Bonini, T. Kind, H. Tsugawa, D. K. Barupal, O. Fiehn, *Anal. Chem.* **2020**, *92*, 7515-7522.
[296] R. Bouwmeester, L. Martens, S. Degroeve, *Anal. Chem.* **2019**, *91*, 3694-3703.
[297] A. M. Wolfer, S. Lozano, T. Umbdenstock, V. Croixmarie, A. Arrault, P. Vayer, *Metabolomics* **2015**, *12*, 8.
[298] E. Bach, S. Szedmak, C. Brouard, S. Bocker, J. Rousu, *Bioinformatics* **2018**, *34*, i875-i883.
[299] F. Zubeil, Dissertation thesis, Eberhard Karls Universität Tübingen **2017**.
[300] C. A. Garcia, A. Gil-de-la-Fuente, C. Barbas, A. Otero, *J. Cheminform.* **2022**, *14*, 33.
[301] R. Adusumilli, P. Mallick, *Methods Mol. Biol.* **2017**, *1550*, 339-368.
[302] J. D. Holman, D. L. Tabb, P. Mallick, *Curr Protoc Bioinformatics* **2014**, *46*, 13 24 11-19.
[303] P. G. Pedrioli, J. K. Eng, R. Hubley, M. Vogelzang, E. W. Deutsch, B. Raught, B. Pratt, E. Nilsson, R. H. Angeletti, R. Apweiler*, et al.*, *Nat. Biotechnol.* **2004**, *22*, 1459-1466.
[304] L. Martens, M. Chambers, M. Sturm, D. Kessner, F. Levander, J. Shofstahl, W. H. Tang, A. Rompp, S. Neumann, A. D. Pizarro*, et al.*, *Mol Cell Proteomics* **2011**, *10*, R110 000133.
[305] Y. Li, T. Kind, J. Folz, A. Vaniya, S. S. Mehta, O. Fiehn, *Nat. Methods* **2021**, *18*, 1524-1531.
[306] N. G. Mahieu, G. J. Patti, *Anal. Chem.* **2017**, *89*, 10397-10406.
[307] R. Tautenhahn, C. Bottcher, S. Neumann, *BMC Bioinformatics* **2008**, *9*, 504.
[308] A. Wee, D. B. Grayden, Y. Zhu, K. Petkovic-Duran, D. Smith, *Electrophoresis* **2008**, *29*, 4215-4225.
[309] K. X. Wan, I. Vidavsky, M. L. Gross, *Journal of the American Society for Mass Spectrometry* **2002**, *13*, 85-88.
[310] T. Stricker, R. Bonner, F. Lisacek, G. Hopfgartner, *Anal Bioanal Chem* **2021**, *413*, 503-517.
[311] A. P. Tay, A. Liang, J. J. Hamey, G. Hart-Smith, M. R. Wilkins, *Proteomics* **2019**, *19*, e1800444.
[312] J. Shi, F. X. Wu, *Proteome Sci* **2011**, *9 Suppl 1*, S3.
[313] W. Marco, H. Ralf, *Journal of Mass Spectrometry* **2002**, *37*, 223-229.
[314] D. G. Kendall, *Statistical Science* **1989**, *4*, 87-99.
[315] H. Wei, Z. Lin, D. Li, Q. Gu, T. Zhu, *Wei Sheng Wu Xue Bao* **2010**, *50*, 701-709.
[316] J. Fuchser, A. Zeeck, *Liebigs Annalen* **1997**, *1997*, 87-95.

APPENDIX

Abbreviations

Aβ	Amyloid-beta	GNPS	Global Natural Products Social Molecular Networking
AD	Alzheimer disease	HDI	Human Development Index
ATP	adenosine triphosphate	HPLC	High performance liquid chromatography
AU	Arbitrary units	HSQC	Heteronuclear Single Quantum Coherence
BC	Before Christ	mCPBA	Meta-chloroperoxybenzoic acid
BGC	Biosynthetic gene cluster	MPLC	Medium performance liquid chromatography
BPC	Basepeak chromatogram	NP	Natural Product
BSA	Bovine serum albumin	NMR	Nuclear magnetic resonance
BVO	Baeyer-Villiger oxidation	PBS	Phosphate-buffered saline
DNA	Deoxyribonucleic acid	PKS	Polyketide synthase
DMSO	Dimethyl sulfoxide	SM	Soy flour mannitol media
EIC	Extracted ion chromatogram	ThT	Thioflavin T
FBS	Fetal bovine serum	wt	Wild type
GSH	Glutathione		

Supplementary data to Chapter 2.3 – Genome sequencing and BGC identification

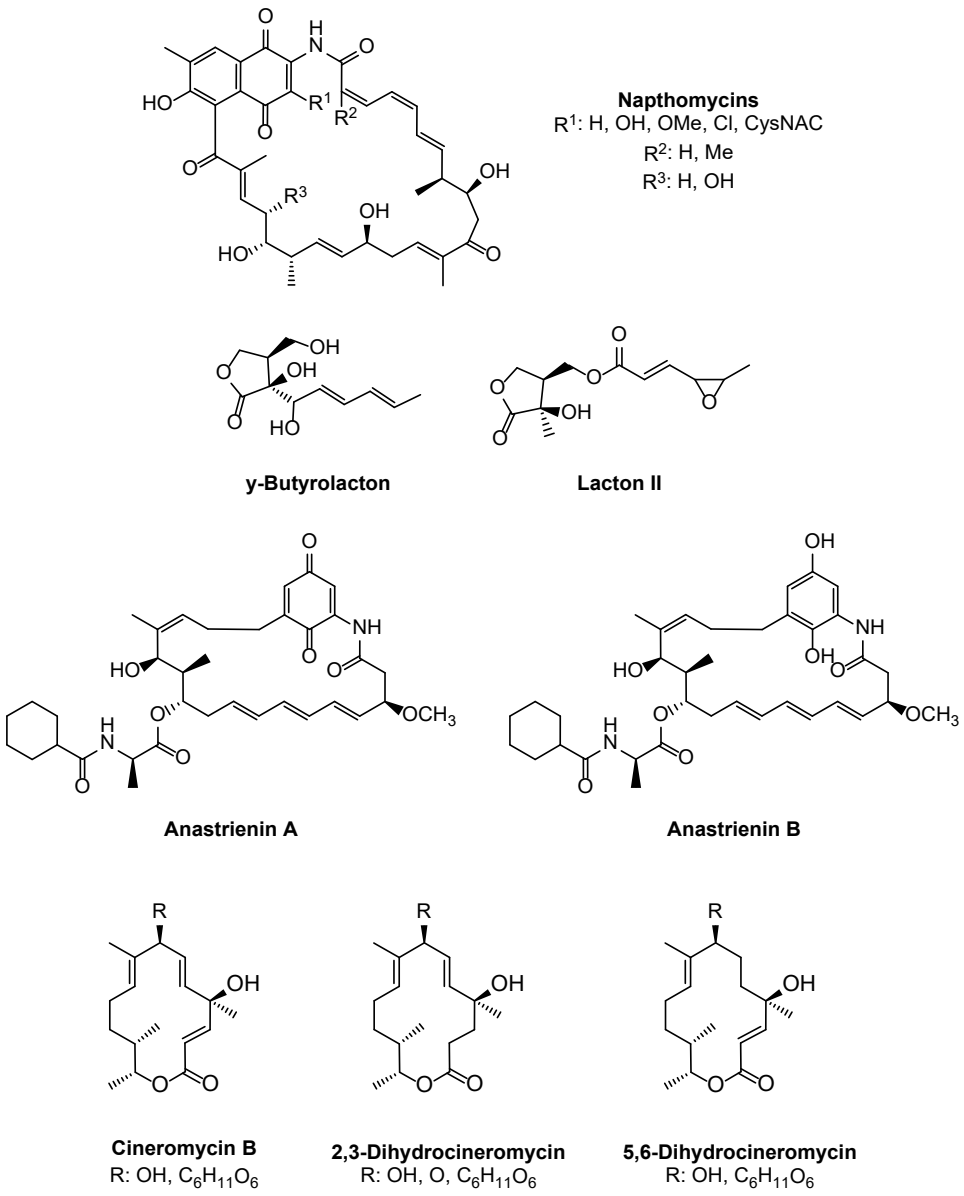


Figure 40: Secondary metabolites isolated from *Streptomyces sp.* Gö 40/10 using the One Strain Many Compounds (OSMAC) approach.[181,315,316]

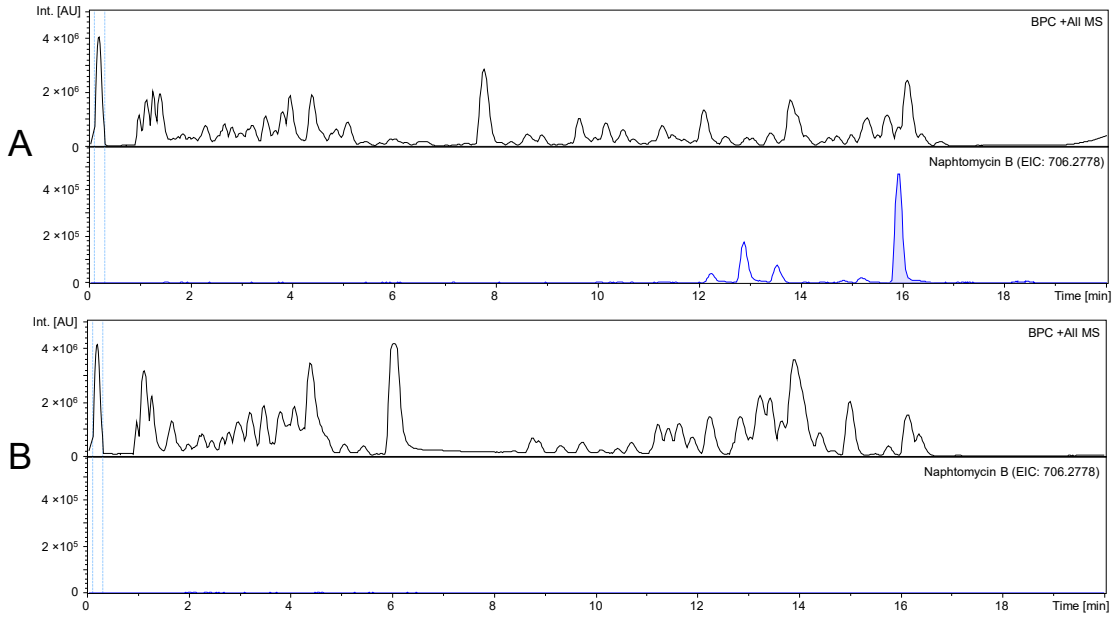


Figure 41: Displayed in black is the base peak chromatogram (BPC), in green the extracted ion chromatogram (EIC) of **1** (361.2010 ± 0.005 Da). Both chromatograms are scaled equally. (A) LC-MS chromatogram of the extracted supernatant from wild type strain *Streptomyces sp.* Gö 40/10, (B) LC-MS chromatogram of extracted supernatant from the mutant strain *Streptomyces sp.* Gö 40/10 -*nat*.

Table 7: Predicted functions of the genes present in the col gene cluster responsible for the biosynthesis of collinolactone (**1**)

Gene	Predicted gene product
colA	Positive transcriptional regulator (luxR family)
colB	Unknown protein
colC	Asparagine synthase
colD	Flavin reductase
colE	Anthranilate 3-monooxygenase (4-hydroxyphenylacetate 3-monooxygenase)
colF	Methyltransferase
colG	DNA-binding protein
colH	Purine efflux pump PbuE
colI	Luciferase like monooxygenase
colJ	Class-II DAHP synthetase family protein
colK	Methylenetetrahydrofolate reductase
colL	T1PKS subunit
colM	T1PKS subunit
colN	T1PKS subunit
colO	Polyketide cyclase (SnoaL family)
colP	Chaperone protein HtpG
colQ	Unknown protein
colR	Methylmalonyl Co-A mutase (associated GTPase MeaB)
colS	Methylmalonyl Co-A mutase
colT	Methylmalonyl Co-A mutase (subunit beta)
colU	Membrane protein
colV	Cell division protein SepF
colW	bldA-regulated nucleotide-binding protein
colX	Nucleotide pyrophosphohydrolase
colY	Unknown protein
colZ	Unknown protein
colZA	Luciferase like monooxygenase

Amino acid sequences for BLAST alignment

***Streptomyces koyangensis* SCSIO 5802 (GenBank: MG243704):**

AbmE2 (Luciferase-like monooxygenase alpha-subunit):
MDIGIAYLGASSSSGRAELISTYDDFFGEVRWAEGRGFSGIWITEHHFSTYSASSSPLL
LLARAAMLAPSLRVGTSVLVLPLWDPARLVADVSTLDVLTGGRFDLGIGRGYQPHEFL
GWGRDVAGSRGRFEESVELMLRLLTEEDVTFSGDHHRVAAPVTVLPRPVQSPHPPV
WMAAVSAESVRYAARHGFHLLTPAAQTVGQLAVQSQLMATALRESGRDPAEHRVST
TRYVYCGADARGRDTARREAARQLQFSSLLAGGGTPYGGNHPAVDEVDPVLVEKAD
ELLLAGEPDAVLEQLEQLAGAGVTQVIAGFRYGDMPRELAQRSMELFAGQVLPYTKDI
QAEGFWRGGGGSP

AbmZ (NADPH-dependent flavin reductase):
MPPLPSTNRIHADEFRTALGRFATGVVAITAYEHATGRPTGMAANSFASVSVDPPLVS
FCVGRTSSTWPLLRASERLCVAILGAHQVDTCGRLARRGGDKFRDVAWQPATPGGL
PVPDGALGWLECSVEAEHPAGDHDIVVARVHHLALGEDRAPLVFFRSQYGAFRLLKE
DGTRDGE

***Streptomyces sp.* Gö 40/10 (NCBI BioProject: SAMN18970196):**

colD (Flavin reductase):
MTAPAGQPRAIHPDALRKAARGCATGVAVLTTRQGEDLFAKTVSSFVTLSMDPPLISV
AVGRHSPLVQAARDTGRLAVCVLRVGQEQLSQRFATPGAGRATGAFAEVRTRTEVT
GAPVLDDCLTWFDCRLHTVLPGGDHSILIGMPVAVACSEGEPLLYHEGGYHAPAPLT
APSAPRPHRAPTASGVRT

colI (Luciferase-like monooxygenase):
MPSSDSSPLAQPNWGVTLPLPGITIGHHRHIVERLPEIGYSDVWTAEGGGTDAFTPLA
ATAAWSPTLRLGTGIVPVFTRGPAVIAQTAATLAQLAPGRLLLGLGSSVPAHVSDINGI
PFEEPFKRTRDVLRFVTKALRGEYIGDEFETFSVTGYRLPHPPAEPAKVILGALRPNML
RLAFGEGDGAITNLLFPEDVPKILQAVGPQPVGKELVVKVFVCPTQDKHYARANARPF
LAWILNREPYRKFHEWLGNGELLREVHEAWEAADPAGSQKALPDDVVDGLFISGSPE
ECREQIMRYHLPGVTSIQLFVSLPPEAFYSQETVLDVLRRLGPAAM

colZA (Luciferase-like monooxygenase):
MDLRIFTEPQQGATYDTLLTVAKATEDLGFDAFFRSDHYLRMGSVDGLPGPTDAWITL
AGLARETKRIRLGTLMTAGTFRLPGVLAIQVAQVDQMSGGRVELGLGAGWFEEEHKA
YGIPFPKEKFARLEEQLAIVTGLWATEAGKTFDFHGTYYDLTDSPALPKPAQKKVPVLI
GGHGATRTPRLAARYADEFNIPFASVEDSARQFGRVRDAAVEAGRDADELTYSNALV
VCVGKDDQEVARRAAAIGREVDELKANGLAGSPAEVVDKIGRYAEIGSRRIYLQILDLG
DLDHLELISAQVQSQLS

Supplementary data for chapter 2.7.1 – Cell viability screening on murine L929

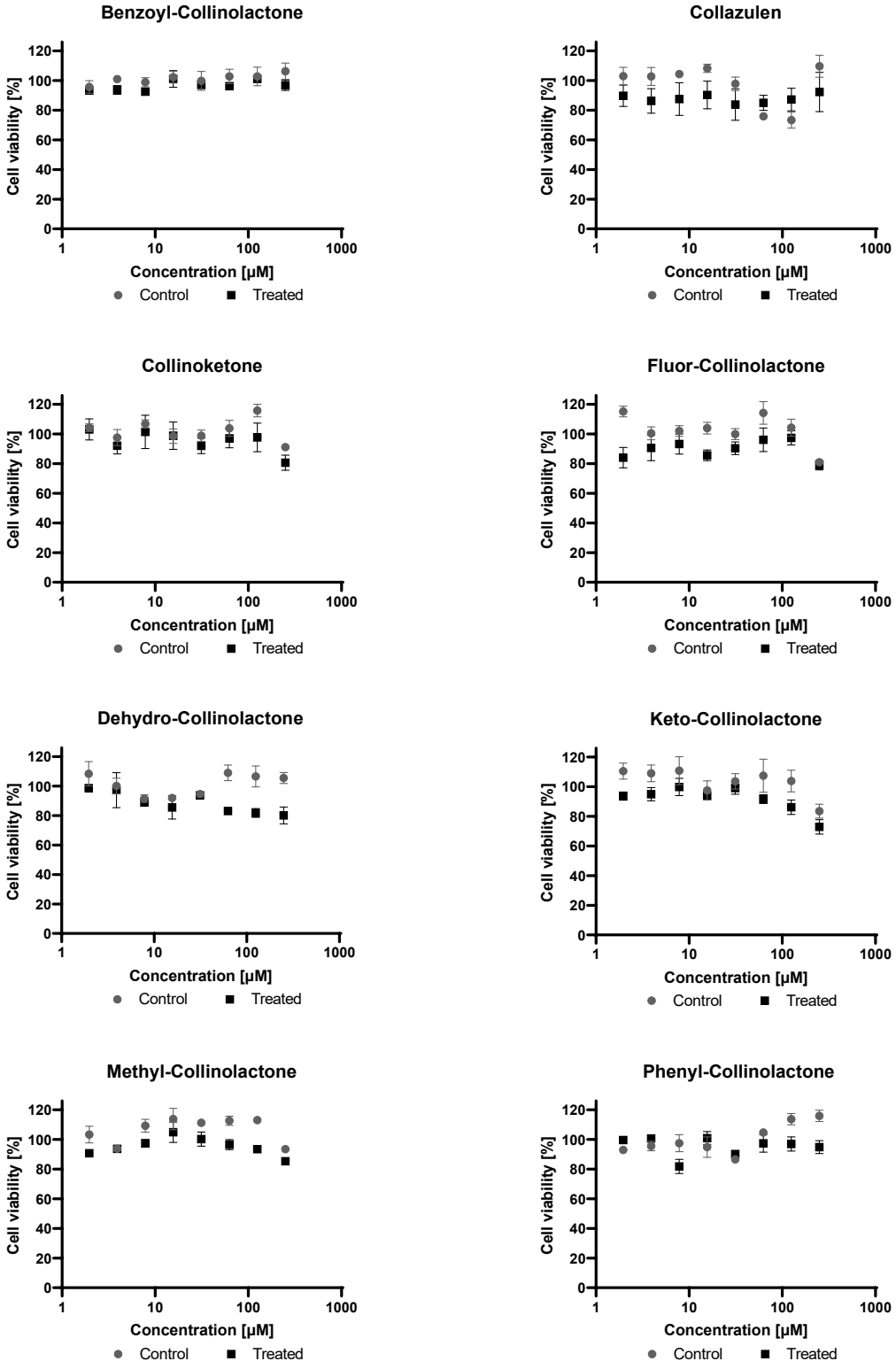


Figure 42: Results of cell viability screening on L929 cell line. The control is displayed in black, the response to treatment is colored. All experiments were performed in triplicates. The data points represent the mean ± SEM.

Supplementary data to chapter 2.7.2 – Fluorescence microscopy on Ptk2

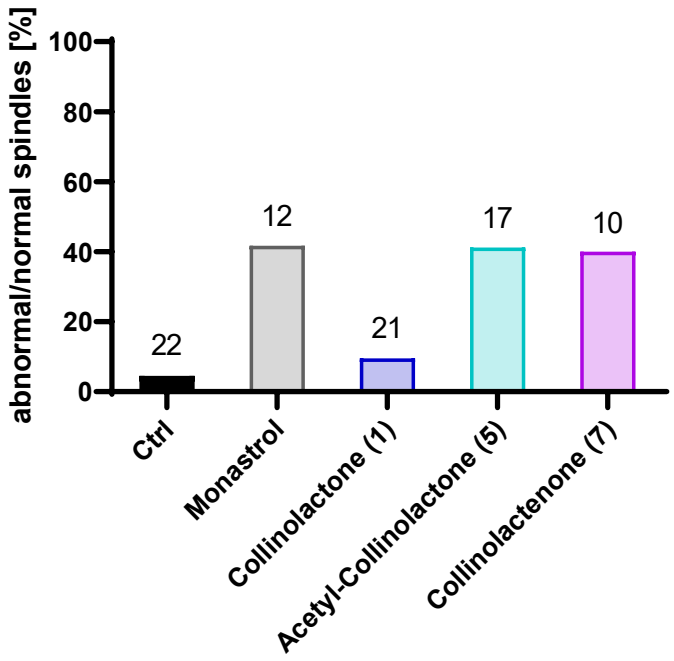


Figure 43: Abnormal spindle formation count, based on randomly selected areas for microscopy imaging. Numbers above bars indicate the total number of cells found in mitosis.

Supplementary data to chapter Eg5 Kinesin and Dynein motor protein assay2.7.3 – Eg5 Kinesin and Dynein motor protein assay

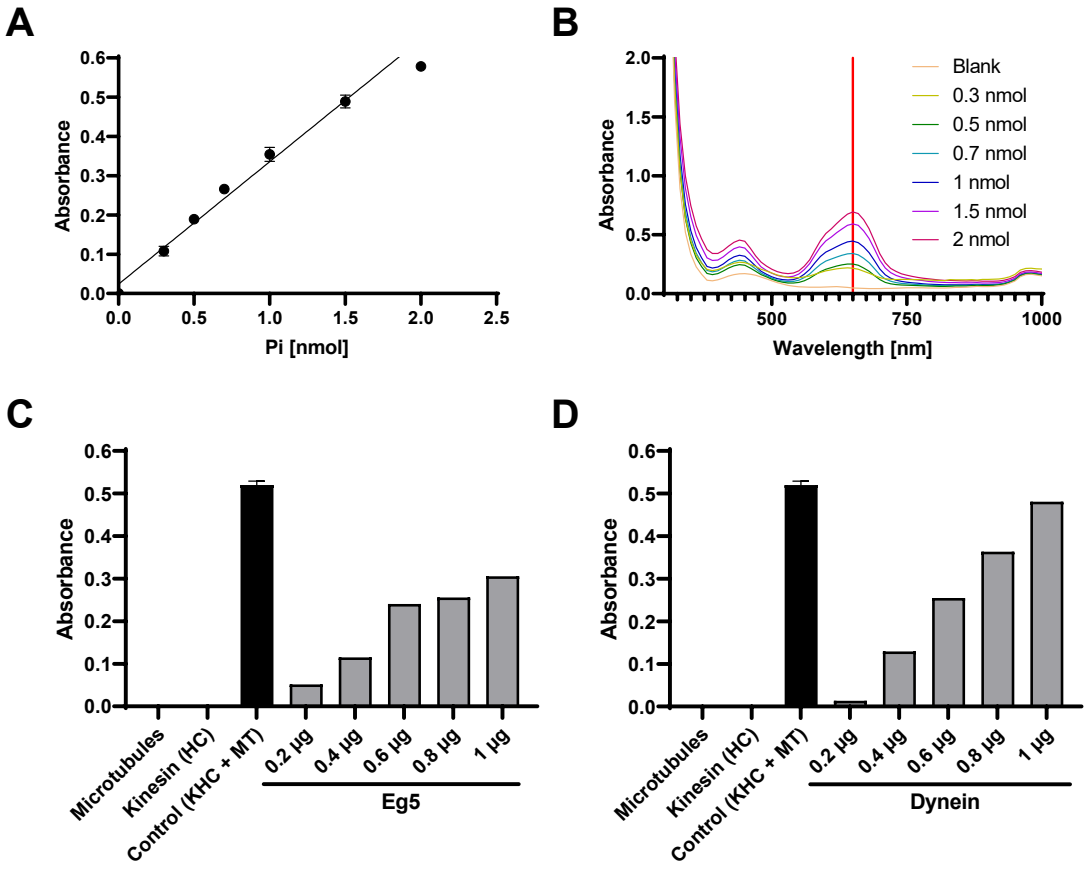


Figure 44: Optimization of parameters for the motor protein assay. (A) phosphate standard curve recorded to quantify the amount generated inorganic phosphate, (B) Absorbance scan of different inorganic phosphate concentrations. The red line indicates the wavelength used for all follow up experiments, (C) Eg5 kinesin motor protein titration, (D) Dynein motor protein titration.

Supplementary data for 2.7.4 – Glutamate induced oxidative stress assay

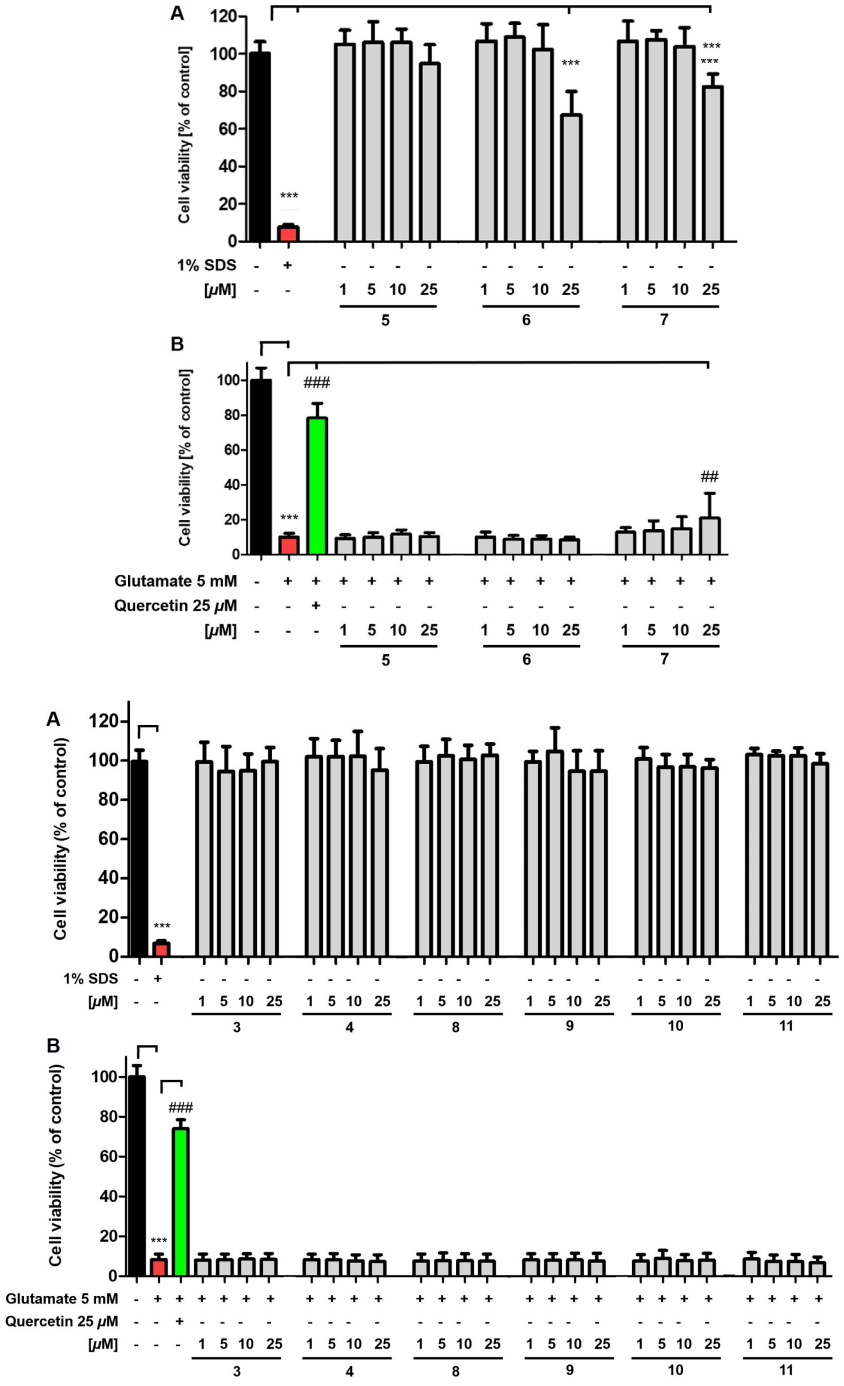


Figure 45: Neurotoxicity (A) and neuroprotective properties (B) of derivatives against glutamate-induced oxidative stress at 1–25 µM using HT22 cell line. 1 % SDS served as positive control for neurotoxic effects whereas quercetin (25 µM) served as positive control for neuroprotection. Results are presented as means ± SEM of three independent experiments, each performed in sextuplicate, and refer to untreated control cells (100 %). Levels of significance: ***p < 0.001; ##p<0.01; ###p < 0.001.

Supplementary data for chapter 2.7.5 – Amyloid aggregation and disaggregation assay

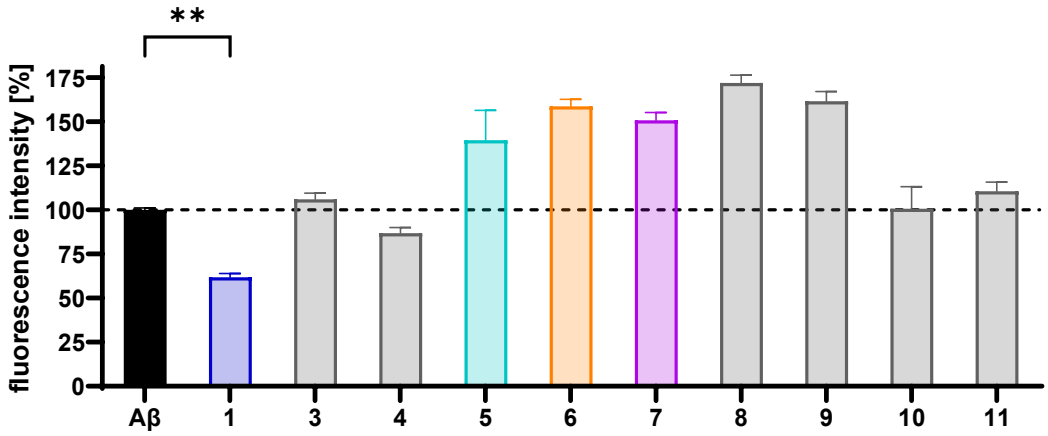


Figure 46: Thioflavin T based amyloid-beta aggregation assay. Applied compound concentrations were 1 mM. Each experiment was performed in triplicates. The data points represent the mean ± SEM. Significance was determined using one-way ANOVA followed by Dunnett's multiple comparison test. Significance levels are: *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ****P ≤ 0.0001, ns: not significant. All other comparisons were not significant.

Supplementary data for chapter 3 – MetaboIDent, a novel platform for automated dereplication

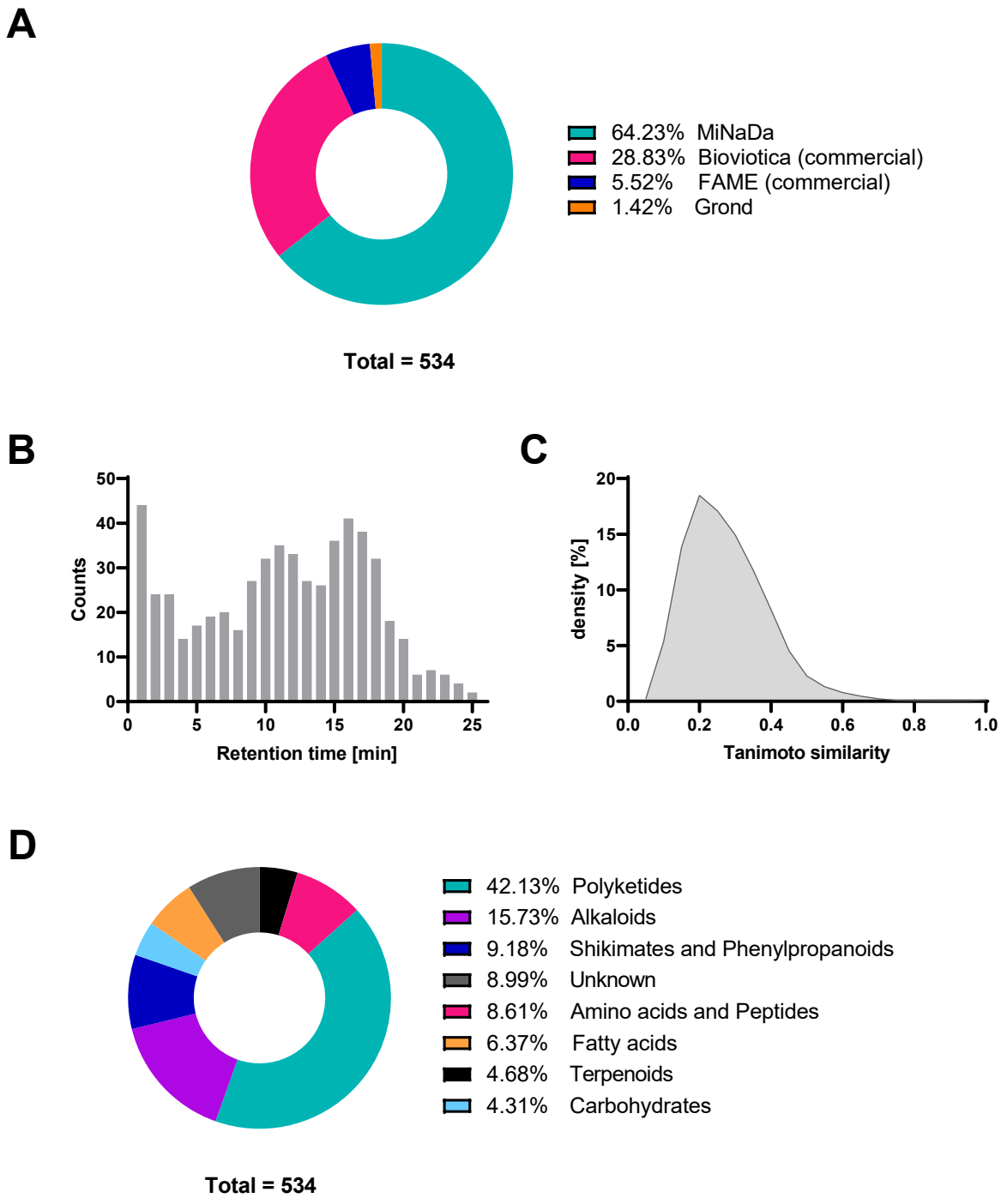


Figure 47: (A) Origin of reference substances. B: Retention times of reference compounds in database used for training and validation (binning 60 s). C: Tanimoto similarity score distribution of compounds in database. D: Classification of reference compounds (pathways) using NPClassifier.[294]

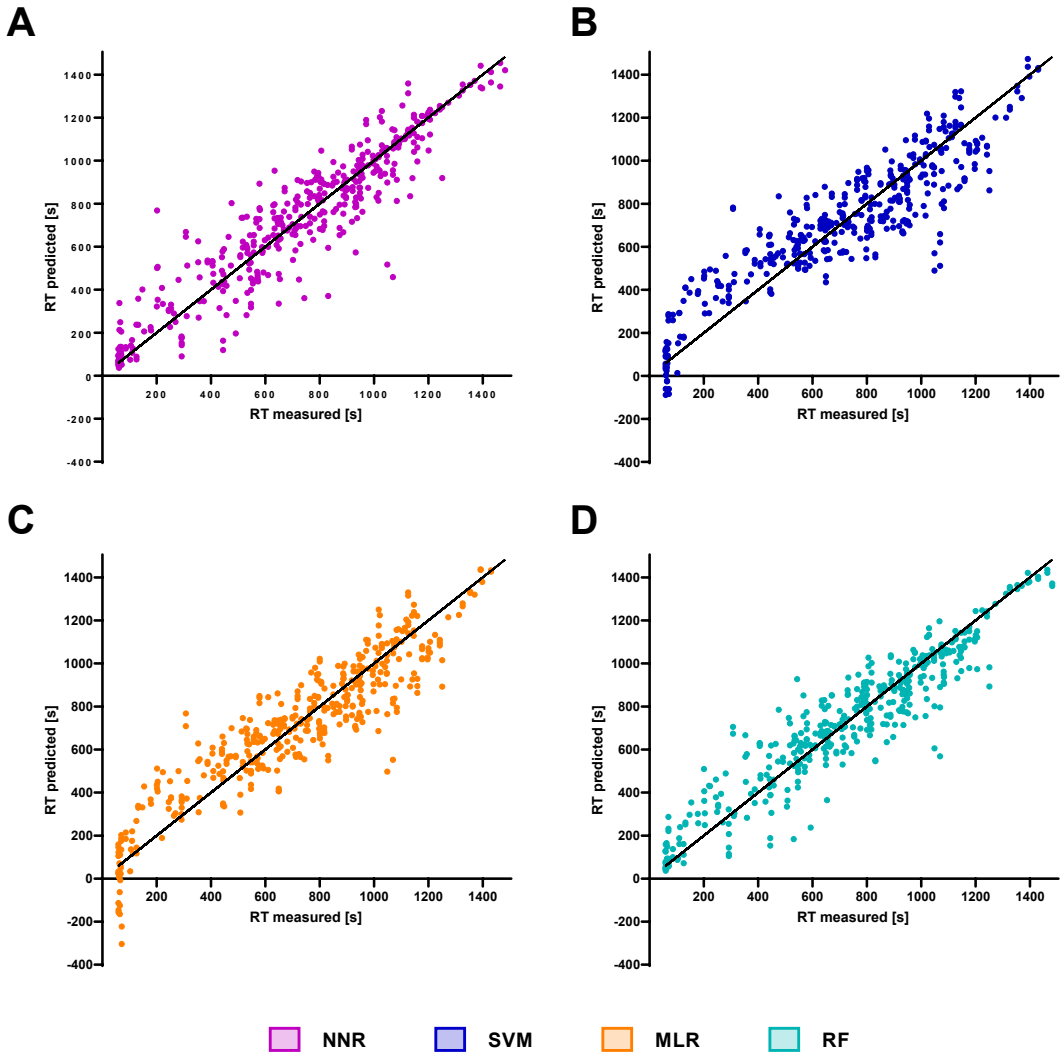
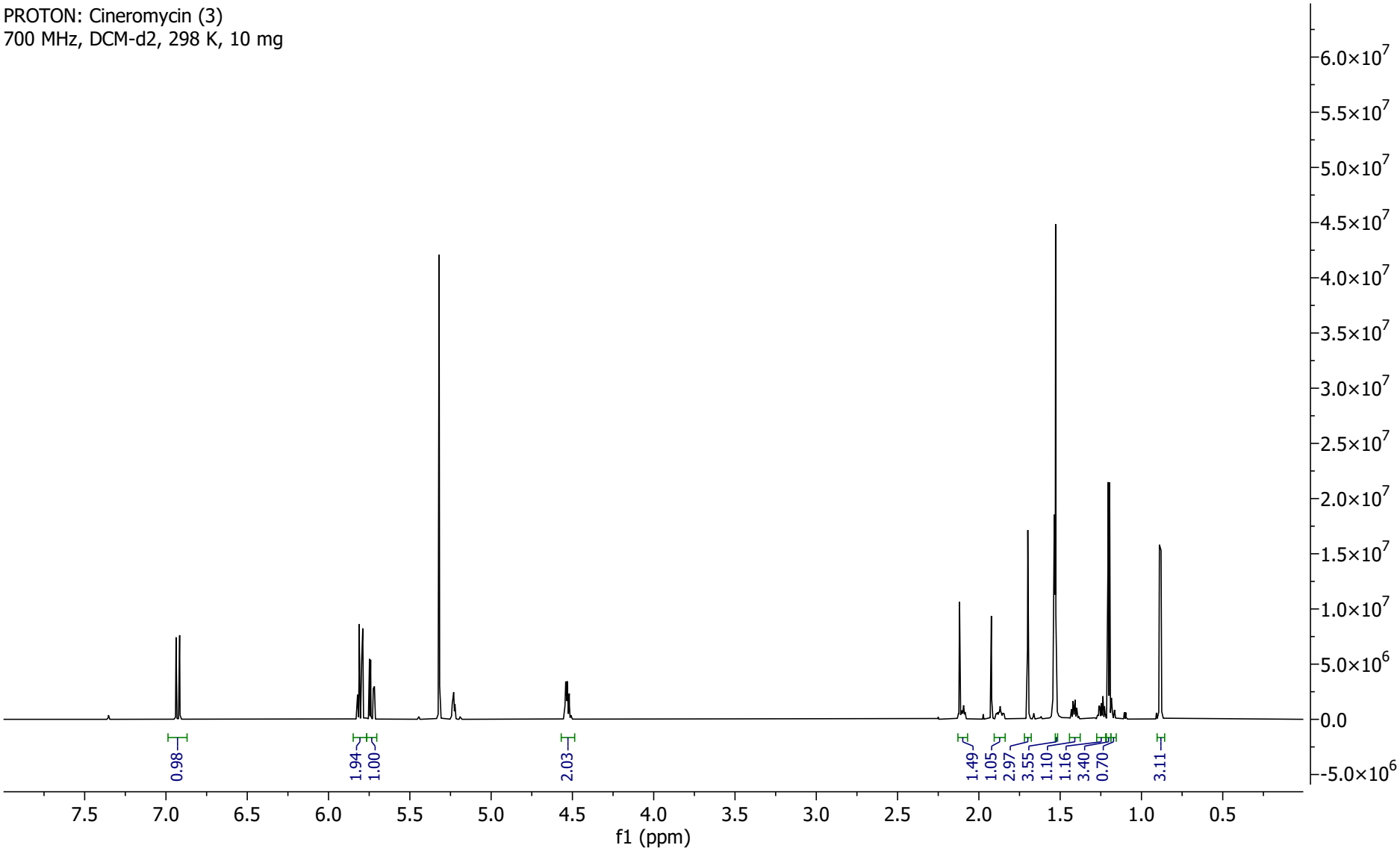
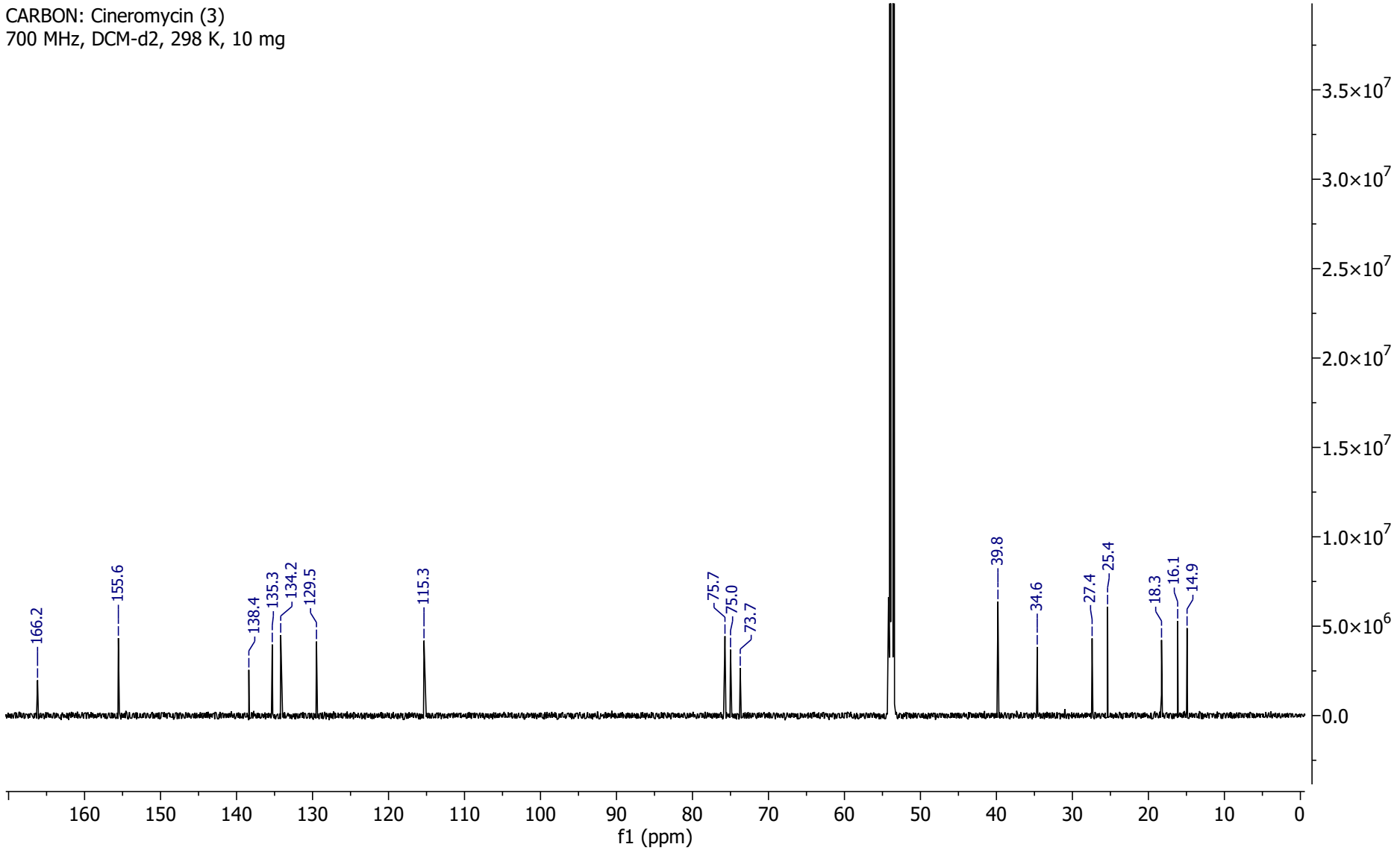
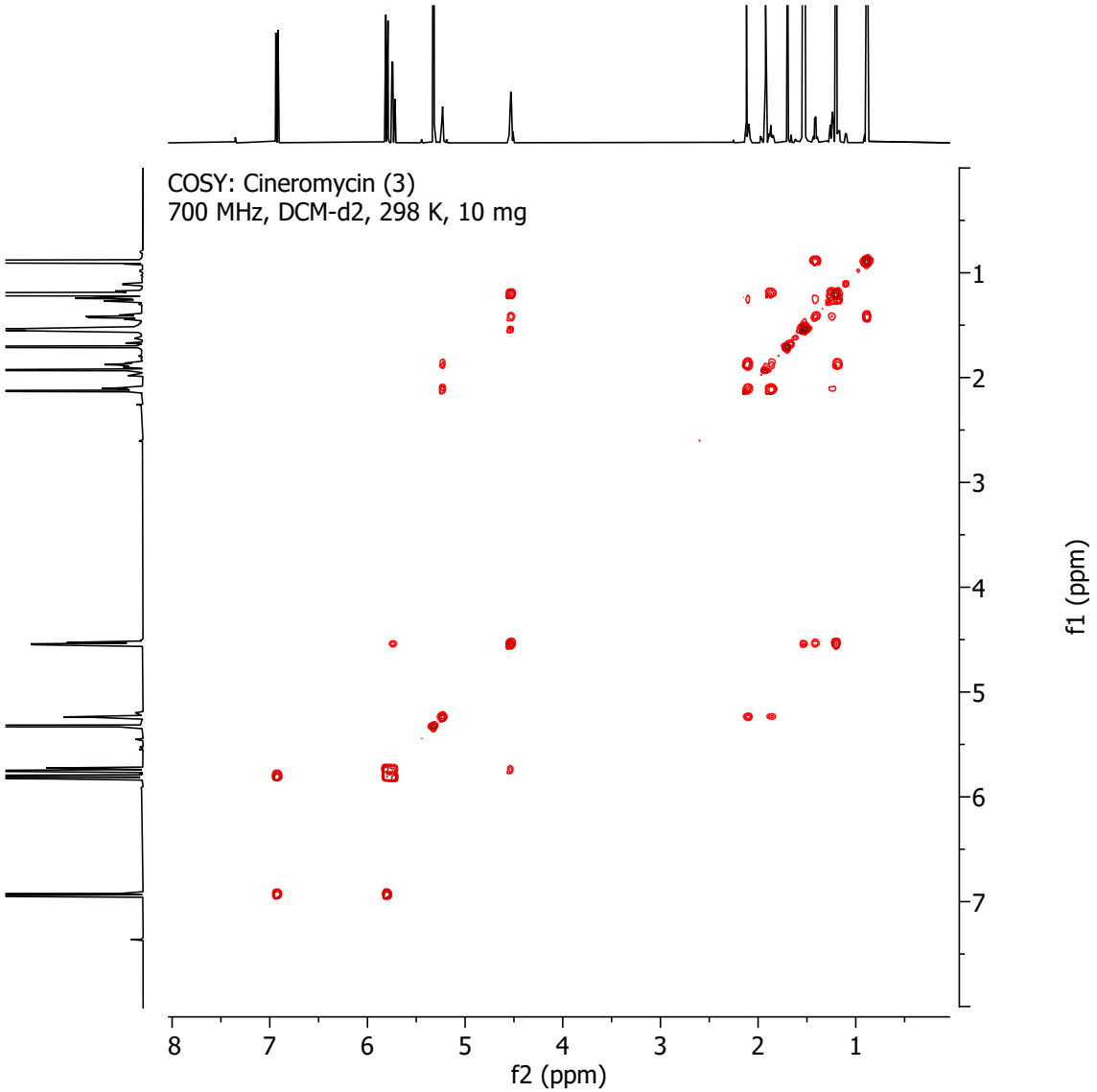
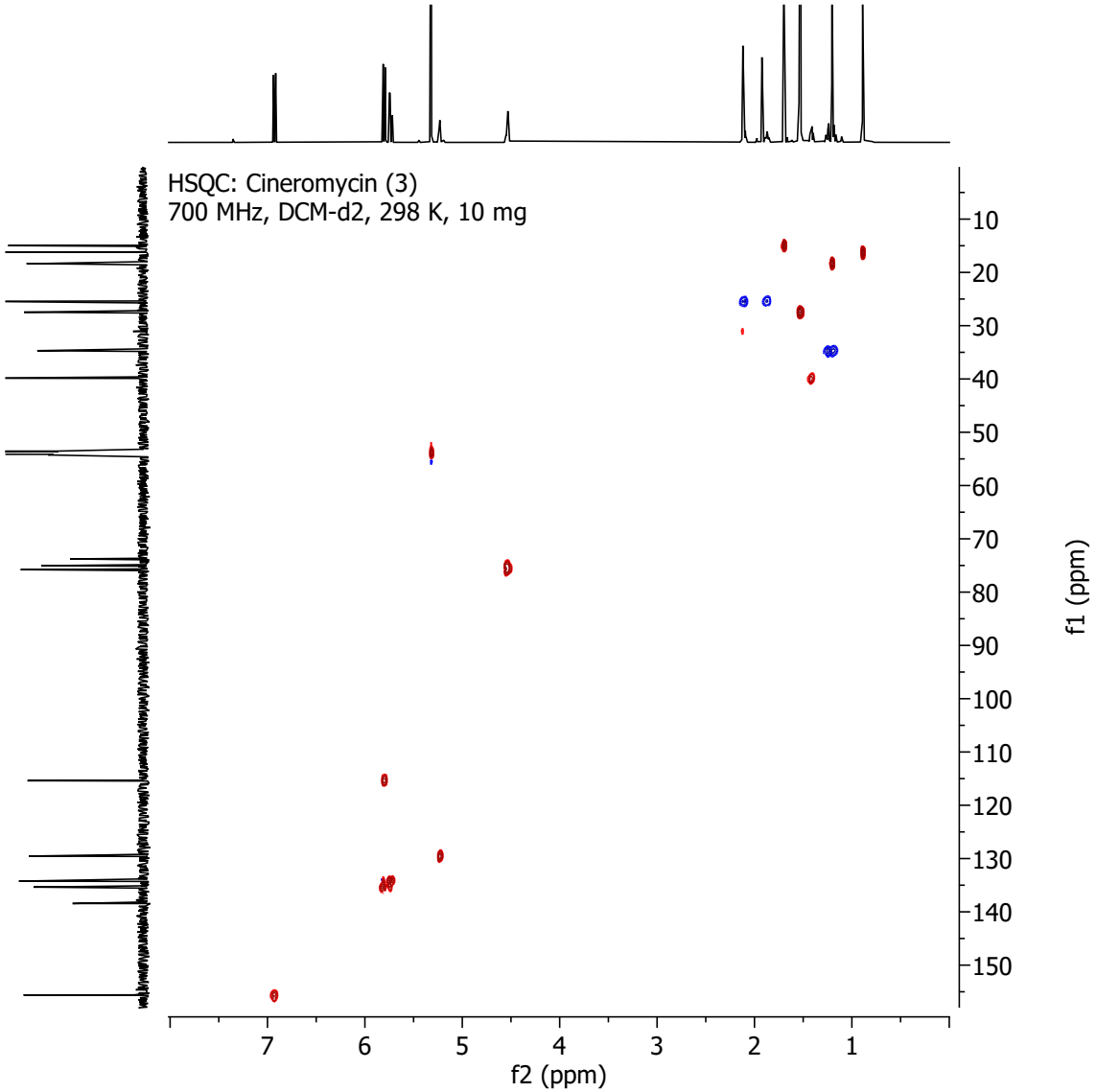


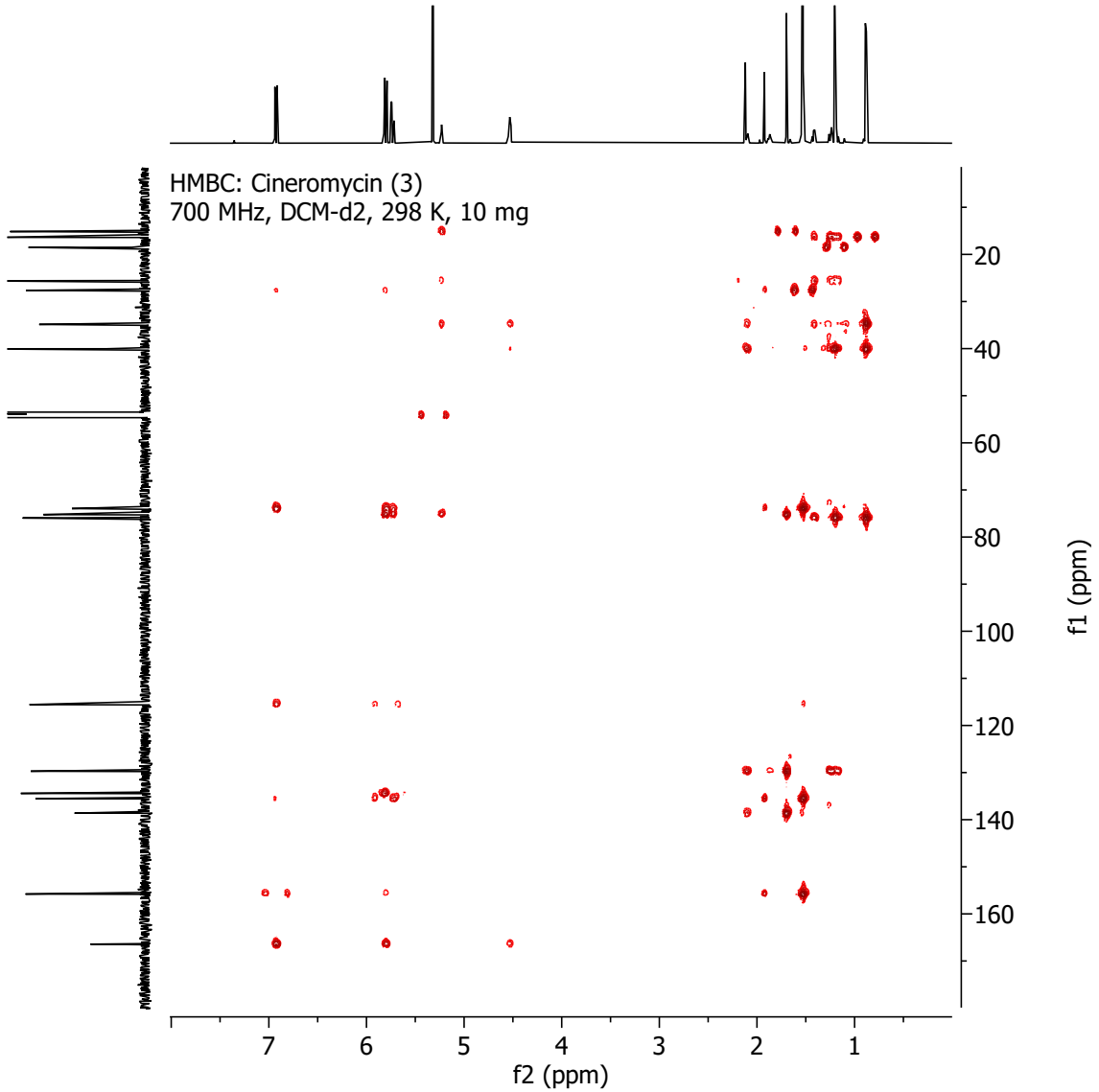
Figure 48: Retention time prediction outcome of four models. (A) Neural Network Regression (NNR) displayed in purple, (B) Support Vector Machine (SVM) in blue, (C) Multiple Linear Regression (MLR) in orange and (D) Random Forest (RF) in turquoise. Each analysis consists of 6 complete training sessions with randomized training and testing set split (80/20).

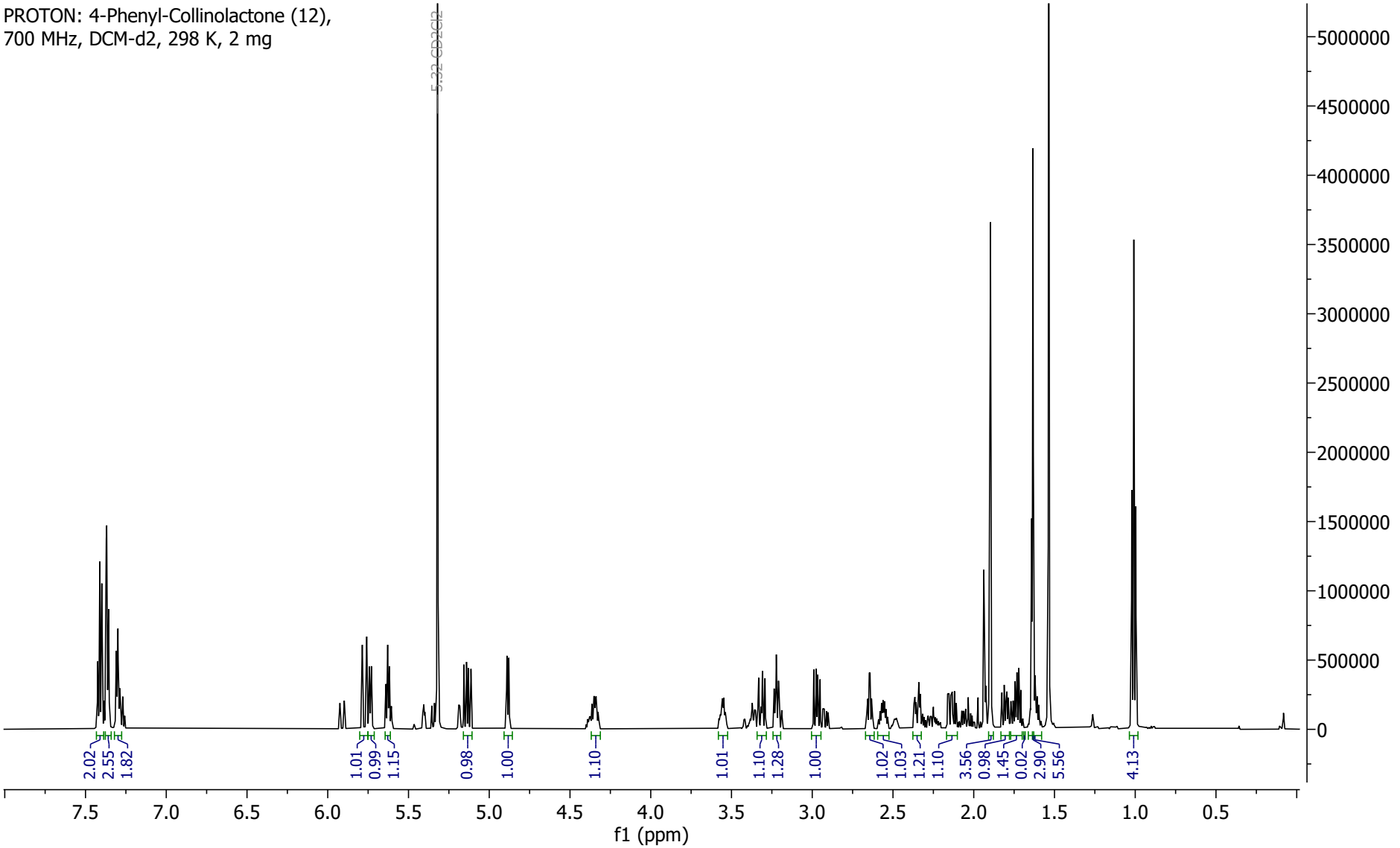

PROTON: Cineromycin (3)
700 MHz, DCM-d2, 298 K, 10 mg
6.0×10^7
5.5×10^7
5.0×10^7
4.5×10^7
4.0×10^7
3.5×10^7
3.0×10^7
2.5×10^7
2.0×10^7
1.5×10^7
1.0×10^7
5.0×10^6
0.0
-5.0×10^6
0.98
1.94
1.00
2.03
1.49
1.05
2.97
3.55
1.10
1.16
3.40
0.70
3.11
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
f1 (ppm)

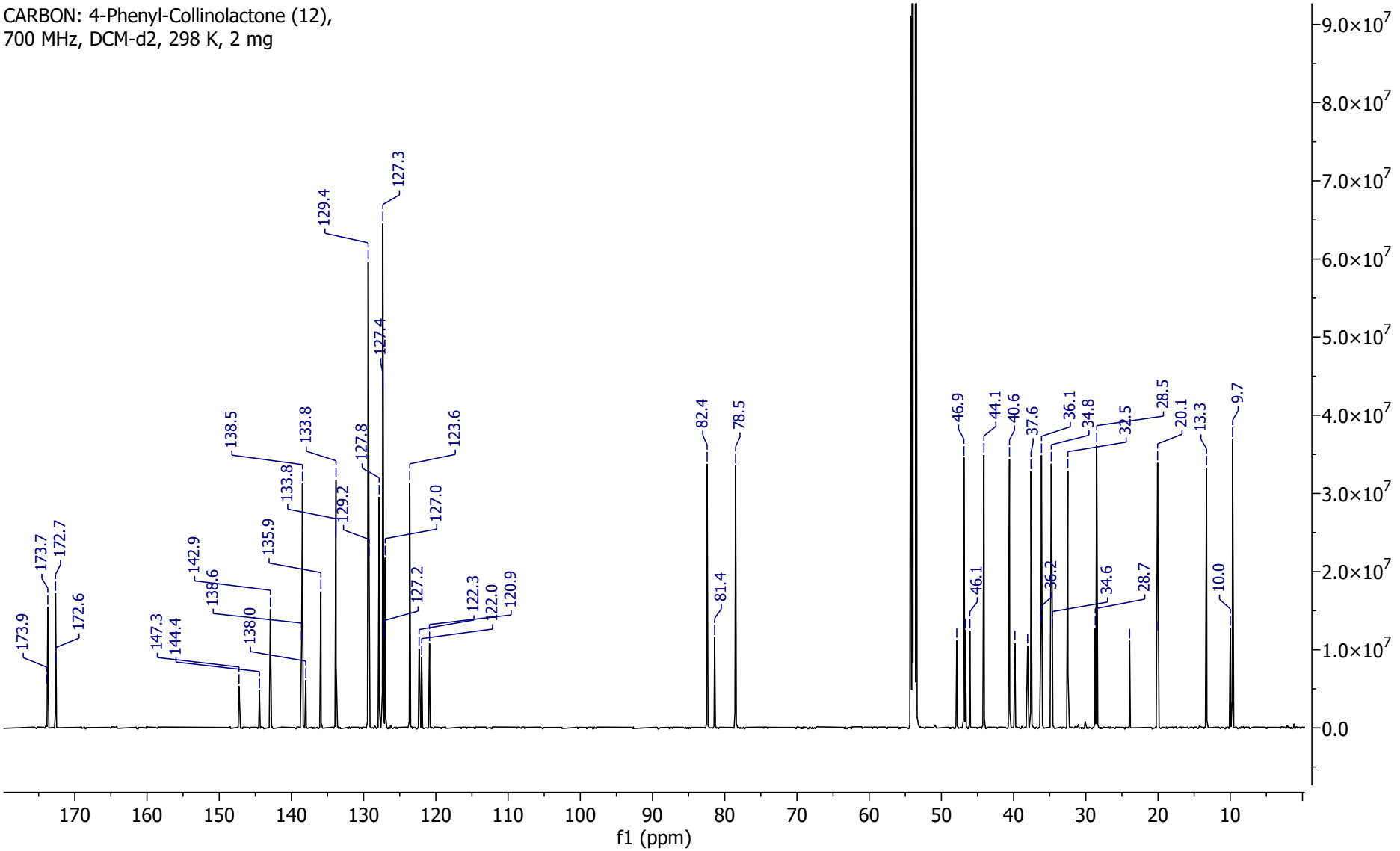

CARBON: Cineromycin (3)
700 MHz, DCM-d2, 298 K, 10 mg
166.2
155.6
138.4
135.3
134.2
129.5
115.3
75.7
75.0
73.7
39.8
34.6
27.4
25.4
18.3
16.1
14.9
3.5×10^7
3.0×10^7
2.5×10^7
2.0×10^7
1.5×10^7
1.0×10^7
5.0×10^6
0.0
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
f1 (ppm)

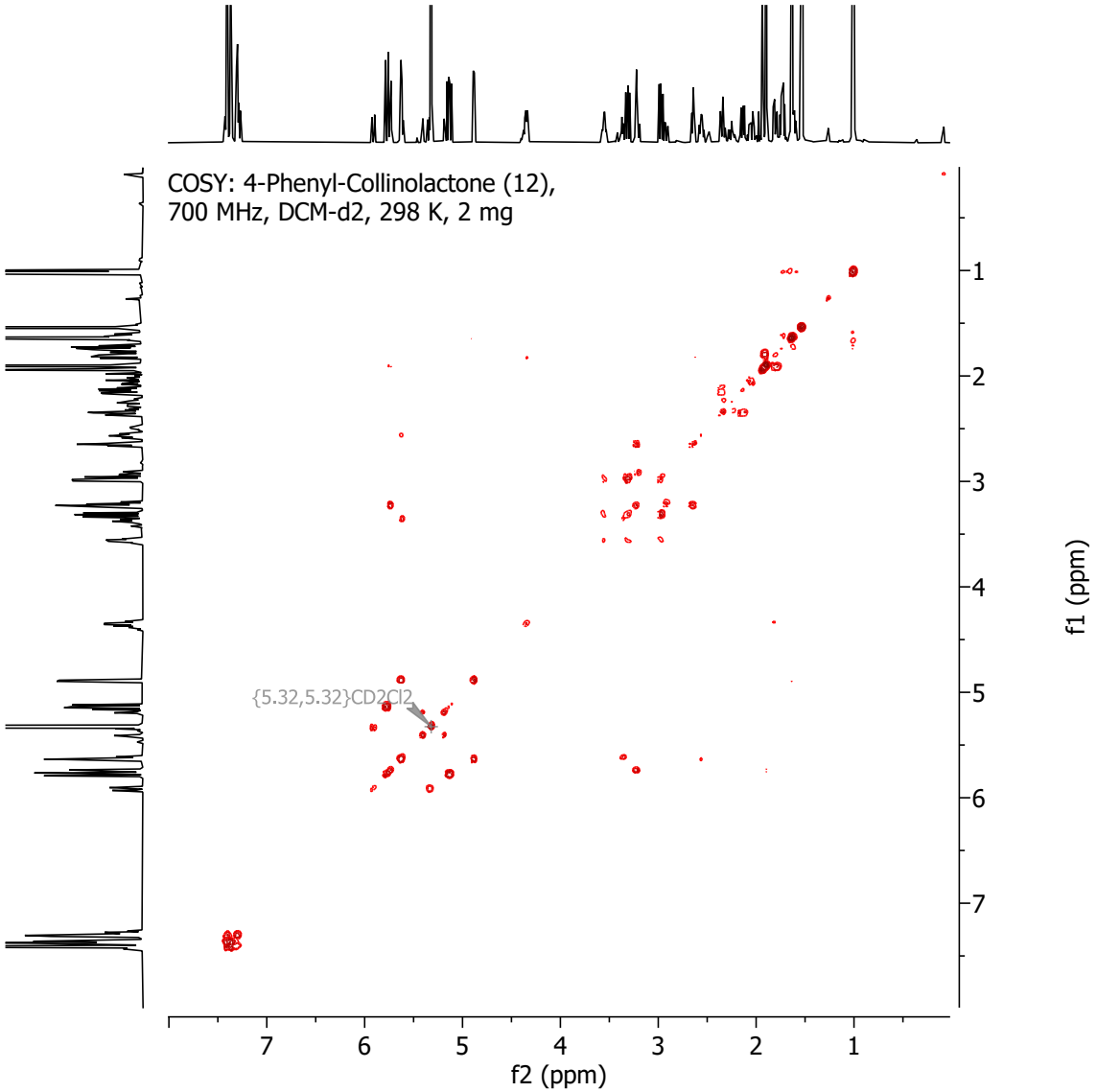

COSY: Cineromycin (3)
700 MHz, DCM-d2, 298 K, 10 mg
8
7
6
5
4
3
2
1
f2 (ppm)
1
2
3
4
5
6
7
f1 (ppm)

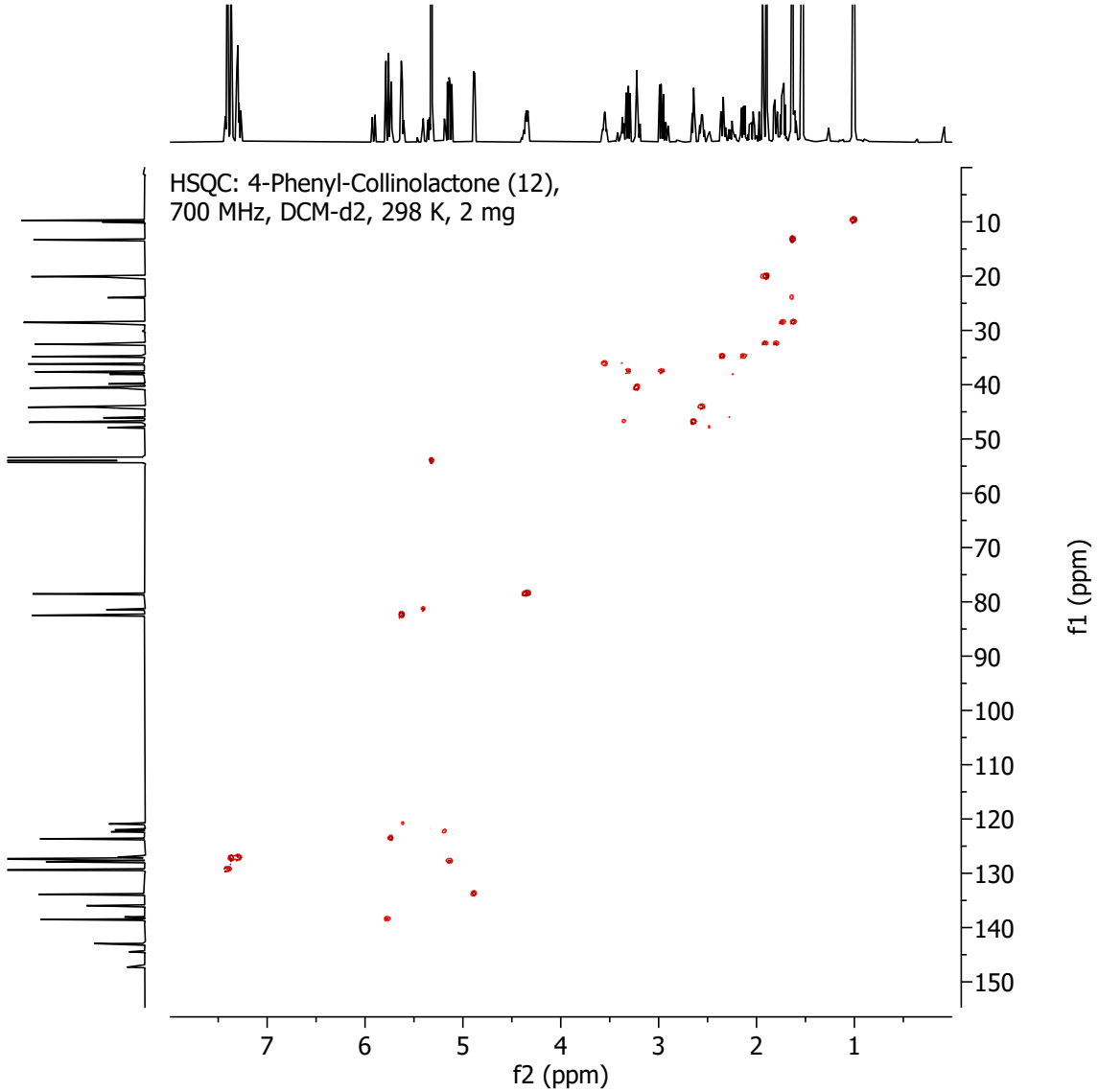

HSQC: Cineromycin (3)
700 MHz, DCM-d2, 298 K, 10 mg
f2 (ppm)
f1 (ppm)

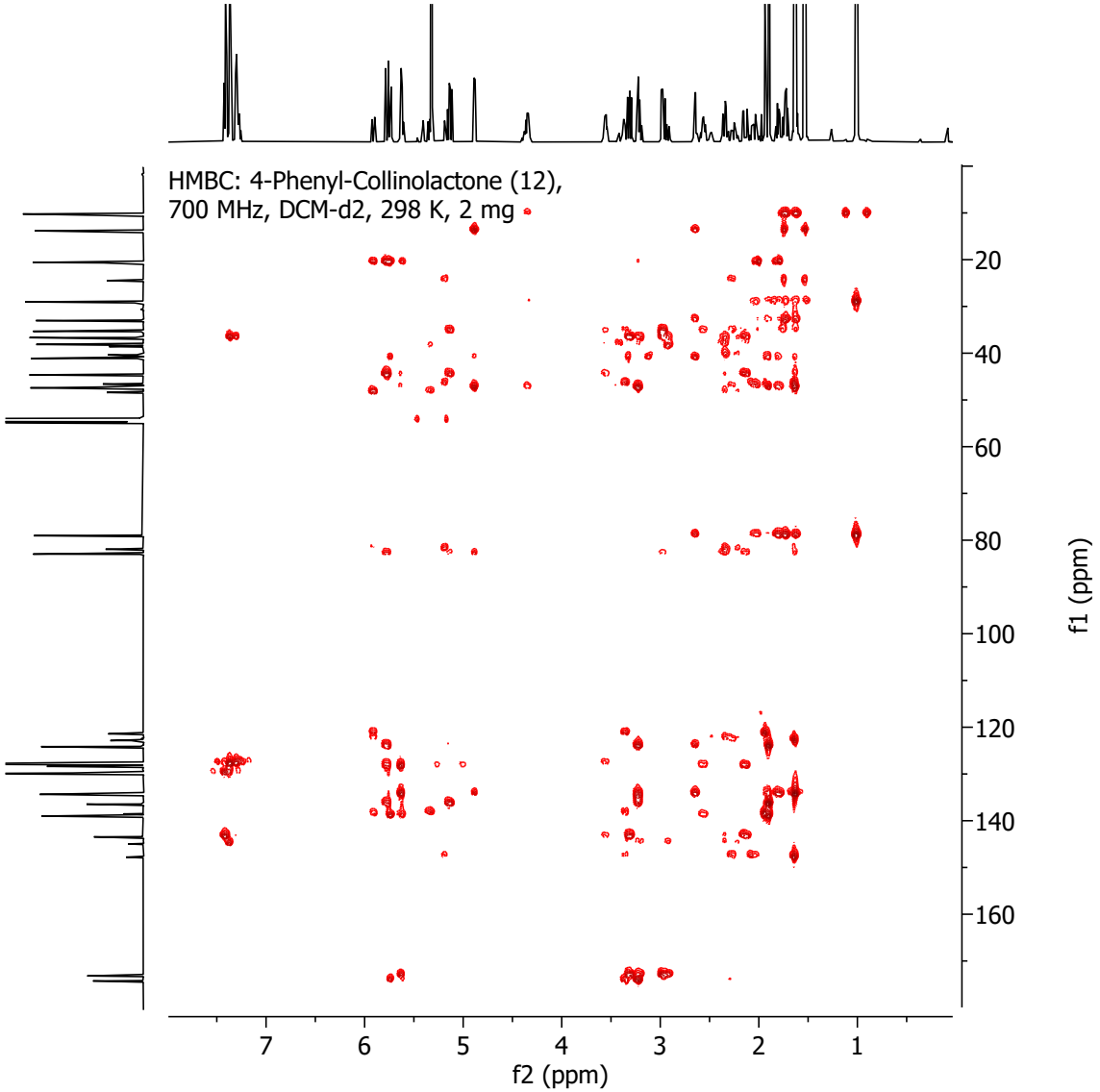

HMBC: Cineromycin (3)
700 MHz, DCM-d2, 298 K, 10 mg
20
40
60
80
100
120
140
160
f1 (ppm)
7
6
5
4
3
2
1
f2 (ppm)

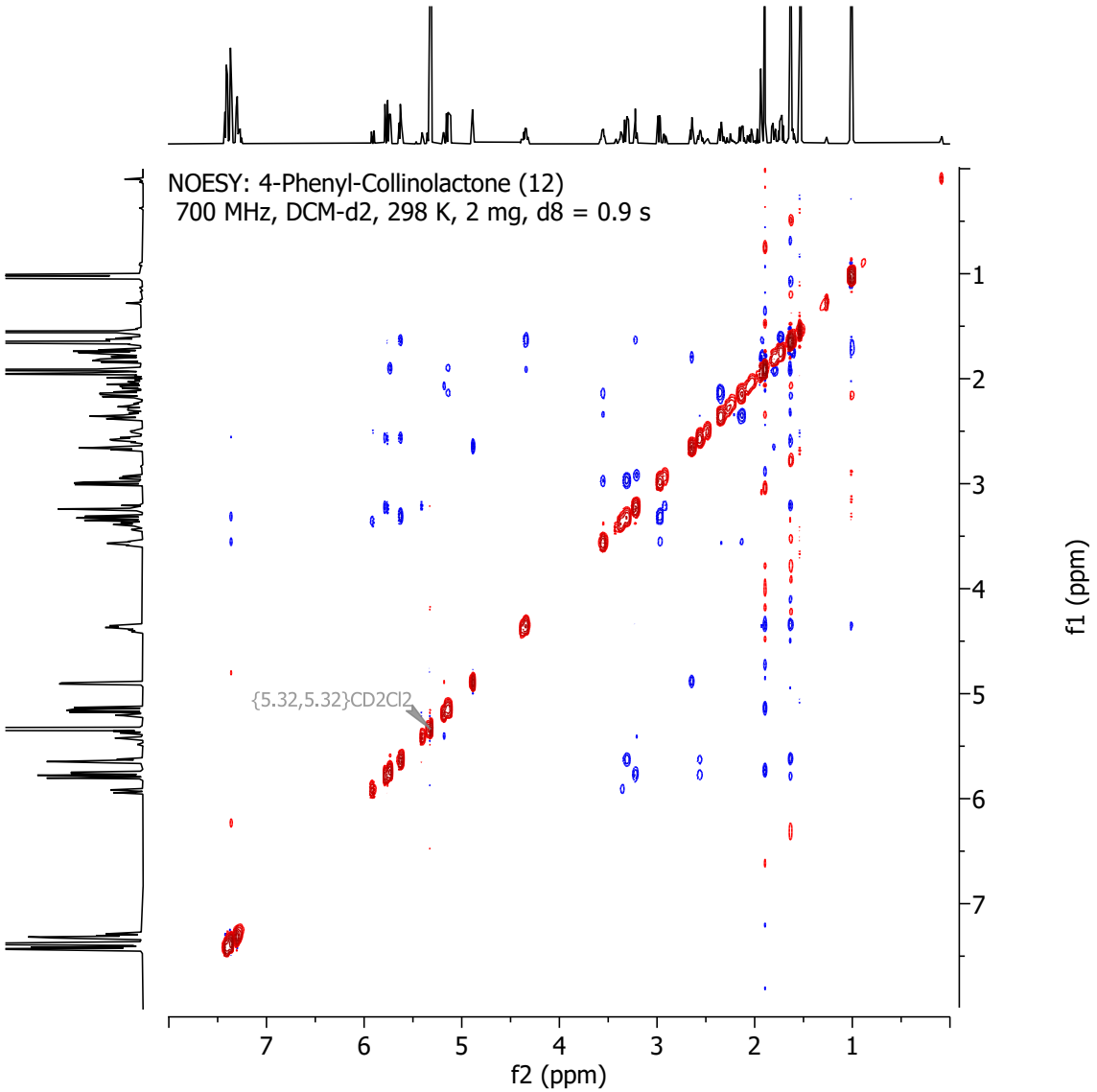

PROTON: 4-Phenyl-Collinolactone (12),
700 MHz, DCM-d2, 298 K, 2 mg
5.32 CD2Cl2
5000000
4500000
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
2.02
2.55
1.82
1.01
0.99
1.15
0.98
1.00
1.10
1.01
1.10
1.28
1.00
1.02
1.03
1.21
1.10
3.56
0.98
1.45
0.02
2.90
5.56
4.13
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
f1 (ppm)


CARBON: 4-Phenyl-Collinolactone (12),
700 MHz, DCM-d2, 298 K, 2 mg
173.9
173.7
172.7
172.6
147.3
144.4
142.9
138.6
138.5
138.0
135.9
133.8
133.8
129.4
129.2
127.8
127.4
127.3
127.2
123.6
127.0
122.3
122.0
120.9
82.4
81.4
78.5
46.9
46.1
44.1
40.6
37.6
36.2
36.1
34.8
34.6
32.5
28.7
28.5
20.1
13.3
10.0
9.7
9.0×10^7
8.0×10^7
7.0×10^7
6.0×10^7
5.0×10^7
4.0×10^7
3.0×10^7
2.0×10^7
1.0×10^7
0.0
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
f1 (ppm)


COSY: 4-Phenyl-Collinolactone (12),
700 MHz, DCM-d2, 298 K, 2 mg
{5.32,5.32}CD2Cl2
f2 (ppm)
f1 (ppm)
7
6
5
4
3
2
1


HSQC: 4-Phenyl-Collinolactone (12),
700 MHz, DCM-d2, 298 K, 2 mg
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
f1 (ppm)
7
6
5
4
3
2
1
f2 (ppm)


HMBC: 4-Phenyl-Collinolactone (12),
700 MHz, DCM-d2, 298 K, 2 mg
20
40
60
80
100
120
140
160
f1 (ppm)
7
6
5
4
3
2
1
f2 (ppm)


NOESY: 4-Phenyl-Collinolactone (12)
700 MHz, DCM-d2, 298 K, 2 mg, d8 = 0.9 s
{5.32,5.32}CD2Cl2
f2 (ppm)
f1 (ppm)
7
6
5
4
3
2
1

ACKNOWLEDGEMENT

Mein besonderer Dank gilt Stephanie Grond für die Möglichkeit, meine Promotion in ihrem Arbeitskreis anzufertigen. Vielen Dank für das überaus spannende und interdisziplinäre Thema, durch welches sich vielfältige, teils auch internationale Kooperationen entwickelt haben. Ganz besonders danken möchte ich Dir für deine durchgehende Unterstützung und das entgegengebrachte Vertrauen, sowie die eingeräumten Freiheiten, um die Projekte nach meinen Vorstellungen voranzutreiben.

Danken möchte ich auch Bertolt Gust dafür, dass er die Zweitbetreuung dieser Arbeit übernommen hat.

Den Mitgliedern des Arbeitskreises danke ich für den Zusammenhalt und die gute Stimmung, auch außerhalb der regulären Arbeitsaktivitäten. Ganz besonders danken möchte ich Manuel Beltran und Michaela Schuppe für ihre durchgehende Unterstützung bei dieser Arbeit, sei es durch das Übernehmen organisatorischer Aufgaben, der Instanthaltung und Wartung der Geräte und Instrumente oder einfach bei den alltäglichen Laborarbeiten.

Ebenfalls danken möchte ich auch meinen PraktikantInnen Jens Mößner, Heiko Pfeffer, Niklas Würtenberg und Sehee Jang, sowie meiner Masterandin Rhena Decker. Es hat mir immer viel Freude bereitet, mit euch zusammen an Collinolactone zu arbeiten und neue Dinge auszuprobieren.

Die Geheimnisse des Collinolacton ließen sich nur in einer Vielzahl an Kooperationen entlocken. Deshalb bin ich ganz besonders meinen Kooperationspartnern Kerstin Frey, Matthias Scheiner und Jaime Felipe Guerrero Garzón zu Dank verpflichtet.

Danken möchte ich auch Priska Kolb und Dominik Brzecki für das Aufsetzen der NMR-Proben. Ebenso gilt mein Dank Dorothee Wistuba, die mit Rat und Tat bei der Betreuung der HPLC-MS zur Seite stand.

Ein riesiger Dank gilt meiner Familie - meinen Eltern, für die Unterstützung über all die Jahre und natürlich meinen beiden Geschwistern, die zwar von Chemie reichlich wenig verstanden haben, sich trotzdem größte Mühe gaben, um den Anschein zu erwecken, sie wüssten wovon ich gerade spreche 😊

Einen ganz lieben Dank auch an Dich Julia, dafür dass Du immer für mich da warst, in den guten wie auch den schlechten Phasen, die eine solche Arbeit unweigerlich mit sich bringt.