DISSERTATION | DOCTORAL THESIS

Titel | Title

Biochemometrics and biopharmaceutical profiling in drug discovery from natural sources

verfasst von | submitted by

Mag.pharm. Sigrid Adelsberger BSc.

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Doktorin der Naturwissenschaften (Dr.rer.nat.)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 796 610 449
Dissertationsgebiet lt. Studienblatt | Field of study as it appears on the student record sheet:	Pharmazie
Betreut von | Supervisor:	Mag. Dr. Ulrike Grienke Privatdoz.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor **PD Mag. pharm. Dr. Ulrike Grienke** (Department of Pharmaceutical Sciences, Division of Pharmacognosy, University of Vienna). Her excellent guidance, support, trust, and motivation throughout the entire duration of my PhD studies were truly inspiring and strongly encouraged me to delve deeply into the fascinating field of natural product research. Her contagious enthusiasm, impressive ability to distill complex topics to their essence, and unwavering commitment to keeping the overall goals in sight helped me remain both curious and focused throughout. Her combination of professional excellence and personal warmth is truly extraordinary, and I am deeply grateful to have had the wonderful opportunity to be the first PhD student of such an inspiring scientist.

I would also like to express my heartfelt thanks to my two mentors **Univ.-Prof. Dr. Lea Ann Dailey** (Deputy Head of the Department of Pharmaceutical Sciences, Head of the Division of Pharmaceutical Technology and Biopharmaceutics, University of Vienna) and **Univ.-Prof. Mag. pharm. Dr. Judith M. Rollinger** (Head of the Research Group Phytochemistry & Biodiscovery at the Department of Pharmaceutical Sciences, Division of Pharmacognosy, University of Vienna, and Member of the Senate of the University of Vienna), and to **Dr. Gabriela Hädrich BSc. MSc.** (Department of Pharmaceutical Sciences, Division of Pharmaceutical Technology and Biopharmaceutics, University of Vienna) for their exceptional expertise, unwavering support, and thoughtful advice. Working with such inspiring leaders and extraordinary role models for women in science has not only been an incredible privilege, but also a deeply meaningful experience that has shaped me both professionally and personally.

I would also like to thank **Univ.-Prof. Dr. Verena M. Dirsch** (Deputy Head of the Department of Pharmaceutical Sciences, Head of the Division of Pharmacognosy, University of Vienna) for granting me the opportunity to conduct my PhD studies within her research group.

My deep gratitude goes to **Mag. pharm. Jacqueline Schwarzinger** (Department of Pharmaceutical Sciences, Division of Pharmaceutical Technology and Biopharmaceutics, University of Vienna), who accompanied me on this exciting journey. Our fruitful discussions and exchange of ideas, along with her support and encouraging words, were incredibly important to me throughout the process. I am truly grateful to have had such an admirable and supportive colleague by my side.

I thank the **Austrian Science Fund** (FWF) for funding my position as a doctoral student as part of the project P35115, 'Inhalation of natural products against lung infections'.

I would further like to thank all collaboration partners of the studies within this thesis. In particular, **Ass. Prof. DI Karin Ortmayr PhD**, **Mag. pharm. Dr. Alexander F. Perhal**, **Mag. pharm. Patrik F. Schwarz**, and **Lorenza Bertaina MSc.** (Department of Pharmaceutical Sciences, Division of Pharmacognosy, University of Vienna), **Assoz. Prof. Mag. Dr. Verena Pichler**, and **Mag. pharm. Sarah Luise Stellnberger BSc.** (Department of Pharmaceutical Sciences, Division of Pharmaceutical Chemistry, University of Vienna), and **Mag. pharm. Dr. Benjamin Kirchweger** (Department of Pharmacy, Pharmaceutical Biology, Ludwig-Maximilians-University Munich).

Additionally, I would like to thank **Mag. pharm. Dr. Julia Langeder** for her valuable help and encouragement during the initial months of my PhD studies, as well as **Dr. Ammar Tahir MSc.** for his support and technical advice. I would also like to thank **Dr. Agnieszka Kowalska** (Department of Pharmaceutical Sciences, Division of Pharmacognosy, University of Vienna), **Dr. Nicolas Coudevylle**, and **Dr. Jiri Mares MSc.** (Department of Pharmaceutical Sciences, Division of Pharmaceutical Chemistry, University of Vienna) for the technical support and **Ing. Andreas Wöber** (Department of Pharmaceutical Sciences, Division of Pharmacognosy, University of Vienna) for the IT assistance.

Furthermore, I would like to thank the master's students **Mag. pharm. Pia Fischer-Windsteig**, **Mag. pharm. Zana Hoti**, **Klaus Leopold Neuhaus BSc.**, **Astrid Eppacher BSc.**, and **Linn Kühnel BSc.** for their dedicated contributions to my projects.

My deepest thanks go to my colleagues **Mag. pharm. Dr. Andreas Wasilewicz**, **Mag. pharm. Martina Redl**, and **Mgr. Zuzana Vaneková D.Pharm. PhD** (Department of Pharmaceutical Sciences, Division of Pharmacognosy, University of Vienna). Their support, help, and camaraderie, along with our scientific discussions and shared humor in the lab and during breaks, made my working days not only productive but also truly enjoyable.

I am also deeply thankful to all my **colleagues from the Division of Pharmacognosy** for the wonderful working atmosphere, the inspiring conversations, and all the creativity and fun we shared.

Finally, I would like to extend my heartfelt gratitude to **my parents, family**, and **friends** for their unwavering support and encouragement throughout this transformative journey. In particular, the presence and motivation of **Hans**, **Ursi**, **Martina**, **Christoph**, **Rosalie**, **Valentina**, **Franziska**, **Stefan**, **Isabell**, **Anja**, **Karim**, **Marie**, **Isa**, and **Kathleen** meant more to me than words can convey.

Above all, I would like to express my deepest and most heartfelt gratitude to my husband **Reinhard Adelsberger**. Without his support, encouragement, and belief in me, none of this would have been possible.

Table of contents

1 Abstract

The aim of this doctoral thesis was to accelerate the search of bioactives from natural sources and to investigate translational aspects regarding their future application. Two different approaches, i.e. biochemometrics and biopharmaceutical profiling (BPP), were applied to unravel multicomponent mixtures and to evaluate their potential for inhalation administration.

Biochemometric methods, such as statistical heterocovariance analysis (HetCA), present valuable tools for the early identification of bioactive constituents in complex extracts prior to their isolation. ^{1}H NMR-based HetCA (1D HetCA) correlates proton resonances with bioactivity data resulting in pseudospectra reflecting calculated correlation coefficients. To overcome limitations in the interpretation of 1D HetCA pseudospectra, e.g. due to overlapped signals, a two-dimensional Heteronuclear Single Quantum Coherence (HSQC)-based HetCA (2D HetCA) approach has been established. Compared to 1D HetCA, the 2D HetCA plots clearly showed an improved differentiation of signals. This biochemometric tool was applied on an *Eriobotrya japonica* Lindl. dichloromethane leaf extract and was able to reveal a minor triterpenoid, i.e. 2α,19α-dihydroxy-3-oxo-12-ursen-28-oic acid (1.98% in the crude extract), with a previously undescribed inverse agonistic activity on the retinoic acid receptor-related orphan receptor gamma (RORγ) (IC_{50} = 2.15 µM), in a targeted, rapid and accurate way (**Manuscript I**).

BPP studies are commonly performed in the development of synthetic small-molecule drug candidates. However, they are not routinely used for evaluations of herbal extracts and constituents from natural sources. Mulberry Diels–Alder-type adducts (MDAAs) from *Morus alba* L. root bark, exhibit both antiviral and antibacterial activity (**Publication I**), but showed low bioavailability and plasma/tissue concentration after oral application. To evaluate their suitability for pulmonary drug delivery, two pure MDAAs, sanggenon C (SGC) and sanggenon D (SGD), and two MDAAs containing extracts (MA21, MA60) were investigated. The profiling included physicochemical properties (pH, pK_a, logP, pH-dependent solubility) and biopharmaceutical parameters (epithelial cytotoxicity, permeability, uptake). Despite the structural similarity of SGC and SGD, the two stereoisomers exhibited distinct behavior in all experiments, but did not demonstrate measurable permeability across Calu-3 monolayers. The results obtained revealed that they are promising candidates for inhalation administration in isolated form rather than as extracts (**Publication II**).

Both approaches, biochemometrics and BPP, not only help to accelerate the search for bioactives and to assess the suitability of the constituents for inhalation administration in an early stage of the drug discovery process, but also contribute to natural product (NP) research in a timesaving and resource-efficient manner.

2 Zusammenfassung

Das Ziel dieser Dissertation war es, einen Beitrag zur Auffindung bioaktiver Stoffe und ihrer Anwendung im Rahmen der Wirkstoffentwicklung basierend auf Naturstoffen zu leisten. Zwei verschiedene Ansätze, Biochemometrie und die Erstellung eines biopharmazeutischen Profils, wurden herangezogen, um die Zusammensetzung von Mehrkomponentengemischen zu entschlüsseln und ihr Potenzial für die inhalative Gabe zu bewerten.

Biochemometrische Methoden, wie etwa die statistische Heterocovarianzanalyse (HetCA), stellen ein präzises Werkzeug zur Identifizierung biogener Wirkstoffe noch vor ihrer Isolierung dar. ^{1}H NMR-basierte HetCA (1D HetCA) korreliert Protonenresonanzen mit Bioaktivitätsdaten und generiert Pseudospektren, die die berechneten Korrelationskoeffizienten widerspiegeln. Um Einschränkungen in der Interpretierbarkeit von 1D HetCA Pseudospektren, wie etwa aufgrund von Signalüberlappungen, zu überwinden, wurde ein zweidimensionaler Heteronuclear Single Quantum Coherence (HSQC)-basierter HetCA Ansatz entwickelt. Im Gegensatz zur 1D HetCA, zeigt das 2D HetCA Pseudospektrum eine deutlich verbesserte Differenzierbarkeit der Signale. Angewendet auf einen Dichlormethan-Extrakt, der aus den Blättern von *Eriobotrya japonica* Lindl. gewonnen wurde, war dieses biochemometrische Werkzeug in der Lage, zielgerichtet, schnell und präzise die terpenoide Nebenverbindung 2α,19α-Dihydroxy-3-oxo-12-ursen-28-säure (1.98% im Extrakt) aufzufinden, deren inverse agonistische Aktivität am retinoic acid receptor-related orphan receptor gamma (RORγ) bisher noch nicht beschrieben war (IC_{50} = 2.15 µM) (**Manuskript I**).

Die Erstellung eines biopharmazeutischen Profils wird häufig bei der Entwicklung synthetischer niedermolekularer Arzneimittelkandidaten durchgeführt, jedoch nicht routinemäßig zur Bewertung von Extrakten und Bestandteilen aus natürlichen Quellen eingesetzt. Addukte vom Mulberry Diels–Alder-Typ (MDAAs) aus der Wurzelrinde von *Morus alba* L. zeigen sowohl antivirale als auch antibakterielle Aktivität (**Publikation I**), weisen jedoch nach oraler Einnahme eine niedrige Bioverfügbarkeit bzw. Plasma-/Gewebekonzentration auf. Um die Eignung von MDAAs für die pulmonale Arzneimittelverabreichung zu bewerten, wurden zwei Reinstoffe, Sanggenon C (SGC) und Sanggenon D (SGD), sowie zwei MDAA-haltige Extrakte (MA21, MA60) untersucht. Das Profil umfasste physikochemische Eigenschaften (pH, pK_a, logP, pH-abhängige Löslichkeit) und biopharmazeutische Parameter (epitheliale Zytotoxizität, Permeabilität, Aufnahme). Trotz der strukturellen Ähnlichkeit von SGC und SGD zeigen die beiden Stereoisomere in allen Experimenten ein unterschiedliches Verhalten, ohne messbare Permeabilität durch Calu-3-Monoschichten. Die Ergebnisse erlauben die Schlussfolgerung, dass SGC und SGD vielversprechende Kandidaten für die Inhalation in isolierter Form sind, nicht jedoch als Bestandteile von Extrakten (**Publikation II**).

Biochemometrie und das Erstellen eines biopharmazeutischen Profils ermöglichen nicht nur die Beschleunigung der Suche nach bioaktiven Substanzen und eine Einschätzung der Eignung dieser Stoffe für die inhalative Gabe in einem sehr frühen Stadium der Wirkstoffentwicklung, sondern erlauben es auch, Naturstoffforschung zeit- und ressourcenschonend durchzuführen.

3 Table of publications

Publications and manuscripts within this doctoral thesis from January 2022 to May 2025:

LABEL	TITLE AND AUTHORS	JOURNAL
Publication I	Identification of Natural Products Inhibiting SARS-CoV-2 by Targeting Viral Proteases: A Combined in Silico and in Vitro Approach Andreas Wasilewicz, Benjamin Kirchweger, Denisa Bojkova, Marie Jose Abi Saad, Julia Langeder, Matthias Bütikofer, **Sigrid Adelsberger**, Ulrike Grienke, Jindrich Cinatl Jr., Olivier Petermann, Leonardo Scapozza, Julien Orts, Johannes Kirchmair, Holger F. Rabenau, Judith M. Rollinger doi: 10.1021/acs.jnatprod.2c00843	*Journal of Natural Products* 2023, 86 (2), 264–275.
Publication II	Biopharmaceutical profiling of anti-infective sanggenons from *Morus alba* root bark for inhalation administration Jacqueline Schwarzinger*, **Sigrid Adelsberger***, Karin Ortmayr, Sarah Luise Stellnberger, Ammar Tahir, Gabriela Hädrich, Verena Pichler, Judith M. Rollinger, Ulrike Grienke and Lea Ann Dailey * equal contribution doi: 10.1016/j.ijpx.2024.100272	*International Journal of Pharmaceutics: X* 2024, 8, 100272.
Manuscript I	Biochemometric 2D NMR-based heterocovariance analysis: A targeted approach for identifying bioactive compounds in complex mixtures **Sigrid Adelsberger**, Alexander F. Perhal, Lorenza Bertaina, Patrik F. Schwarz, Verena M. Dirsch, Judith M. Rollinger, Ulrike Grienke	Submitted 22.04.2025

4 Introduction

4.1 Drug discovery and natural products (NPs)

The human use of natural products (NPs) for the prevention, treatment, and cure of diseases dates back thousands of years, with the earliest written evidence found in Mesopotamia around 2600 B.C.[1,2] NPs are molecules that are biosynthesized by living organisms, e.g., plants, fungi, lichens, algae, microorganisms, etc., and enable adaption to the environment and defense against predators and pathogens.[3,4] While primary metabolites - carbohydrates, proteins and lipids among others - are essential for basal functions such as energy supply, development and growth[1,5], secondary metabolites (termed NPs in a narrow sense) offer a survival advantage through, e.g., interspecies signalling and defense, nutrient-scavenging properties, and protection from ultraviolet radiation.[6-8] Evolutionary pressure over millions of years established biosynthetic pathways with diverse functionality and resulted in efficient NPs with optimal biomolecular compatibility, membrane permeability, and stability.[9] Highly active hormons, toxins, attractants, antimicrobials, etc. interact with biological target macromolecules such as proteins, nucleic acids, and membranes[10-15] and share key properties of drug-likeness.[9]

The vast range of biological activities exhibited by NPs underpins their crucial role as lead structures in drug discovery[16-18] and NPs represent one of the richest sources of inspiration for drug development.[19] A comprehensive study by Newman and Cragg showed that two-thirds of all new drug approvals between 1981 and 2019 are NPs or of NP origin or inspired by NPs[20], particularly with a dominant role in anticancer and anti-infective agents.[21,22] NPs are complex molecules with high structural diversity[23] and not only cover a wider chemical space than synthetic compounds[24-26] but also populate regions within this space that are of great importance for drug discovery.[19] By 2018, over 250,000 NPs had been reported as available[19], and by 2023, a virtual database comprising more than 67 million NP-like molecules had been established[27], highlighting the immense chemical wealth provided by nature.

Compared to small-molecule drug candidates, NPs typically have a higher molecular mass, more H bond acceptors and donors, lower lipophilicity, larger polar surface area, and more stereochemical centers.[16,24,28] Higher numbers of sp^3 carbons, fewer aromatic rings and rotatable bonds indicate higher rigidity with less planar structure and three-dimensional (3D) shape[28] which can be advantageous for interaction with biological macromolecules.[16,29] Moreover, metabolite-like compounds, including many NPs, exhibit enhanced druggability, combining inherent biological activity with efficient intracellular access mediated by endogenous transporters.[9,30,31]

Overall, NPs, with their diverse, complex and unique scaffolds as well as a wide range of functional groups and pharmacophore structures (3D arrangement of chemical features with the potential to

interact with a target)[30,32], have evolved to fit into the binding sites of biological targets with high specificity[33] and reduced off-target effects[34], and are therefore considered ideal starting points for drug discovery.[16]

4.2 The search for bioactive NPs

The chemical structure of a molecule largely determines its biological activity, potential interactions with target molecules and cellular systems, and is therefore the basis for pharmaceutical lead and target identification studies.[28] It has been shown that taxonomically related medicinal sources with similar phytochemical compositions correlate with similar therapeutic uses.[35] However, the search for bioactive NPs spans a broad spectrum of methods, from traditional ethnopharmacological knowledge[36] to modern artificial intelligence (AI) technologies.[8,37]

Since there are no generally accepted, defined classification frameworks, the methods used to search for bioactive NPs are divided into three categories in this work: (i) traditional methods, (ii) modern, computer-assisted methods, and (iii) and biochemometric methods.

5 Traditional methods in the search for bioactive NPs

The body of knowledge accumulated by humanity over millennia through trial and error in the realms of **traditional** and **folk medicine** represents a valuable repository of healthcare wisdom.[36] Throughout this period, medicinal plants were used solely on an empirical basis, with no mechanistic understanding of their pharmacological effects or active constituents.[2] The systematic clinical assessment of NPs began to take shape only in the 18th century, when Anton von Störck investigated aconite and colchicum, and William Withering conducted seminal studies on foxglove.[38]

In the context of modern **ethnopharmacology**, the selection of a plant or other natural source is guided by knowledge of the traditional use, serving as the basis for research aimed at isolating and characterizing its bioactive constituents.[2,36] Ethnopharmacology represents a highly interdisciplinary field that integrates botany, chemistry, biochemistry, and pharmacology, and is further interconnected with anthropology, history, archaeology, and linguistics.[39] The existence of ethnopharmacological records can provide valuable information on therapeutic effects and safety aspects relevant to drug discovery.[2] It is a matter of concern, that traditional knowledge is increasingly losing its significance due to the loss of intergenerational knowledge transfer, the loss of biodiversity, the limited availability of sufficient high-quality raw materials, and other contributing factors.[2,36,40] However, still a substantial proportion of the global population continues to rely primarily on herbal NPs and supplements for their healthcare needs.[41]

In 1805, the application of chemical methods enabled Friedrich Sertürner (1783–1841) to isolate morphine from raw opium, an important medical breakthrough of the 19th century, often considered the first case of **phytochemical investigations**.[42,43] This milestone marked the onset of a new era in pharmacognosy, fostering the systematic investigation of medicinal plants and leading to the isolation of numerous pharmacologically active compounds, primarily alkaloids, from natural sources.[2] Over the course of the 20th century, advances in spectroscopic and chromatographic techniques, including nuclear magnetic resonance (NMR), mass spectrometry (MS), and high-performance liquid chromatography (HPLC), firmly established phytochemistry as a core discipline within pharmacognosy and NP research.[44]

In contemporary NP research, when the mechanism of action and molecular targets are unknown, **bioactivity-guided fractionation**, also referred to as bioassay-guided fractionation[45], is regarded as the gold standard for identifying bioactive compounds within complex mixtures.[46] Starting from a bioactive crude extract, iterative chromatographic fractionation is performed, followed by bioactivity assessment, with activity-based prioritization of fractions for subsequent rounds.[46,47] With each successive step, the complexity of the constituent composition is progressively reduced until the pure bioactive compounds are isolated.[48] Despite its widespread application, this approach has several drawbacks.[16] Highly abundant but inactive compounds may impede the identification of minor bioactive constituents, synergistic effects may distort bioactivity results, and loss or degradation of bioactive compounds during isolation may occur.[49,50] Despite its overall success, bioactivity-guided fractionation remains a tedious, time-consuming, and costly endeavour.[51]

High-throughput screening (HTS) was developed in the 1990s to enable the activity testing of large numbers of samples against various targets in a short time.[52,53] Although HTS for NPs have demonstrated significantly higher hit rates compared to HTS conducted on synthetic compound libraries[54], NPs derived from complex crude extracts are less compatible with HTS than chemically well-defined small-molecule drug candidates and require pre-preparation steps, e.g., generation of suitable sub-fractions.[16,30,55] Several limitations associated with HTS must be carefully considered, particularly in the context of multicomponent mixtures with intrinsic confounding biological activities.[56,57] Extracts containing a variety of constituents carry a high risk of interference with, or precipitation of, assay reagents.[58,59] Bioactive compounds present at concentrations below the limit of detection (LOD) of the corresponding assay may remain undetected.[30,60] Moreover, assay outcomes may be affected by the cytotoxicity of individual compounds or by the presence of compounds exhibiting counteractive effects toward the target.[61,62]

6 Computer-assisted methods in the search for bioactive NPs

Since the intital efforts in the 1970s, computational approaches have profoundly transformed the field of NP research and drug discovery, and have led to the emergence of the concept known as computer-aided drug design (CADD).[23,63] Compared to traditional HTS and combinatorial chemistry, CADD offers a more targeted and mechanism-based approach.[64] It enables the efficient selection of candidate molecules from large libraries for subsequent experimental validation, supports lead optimization with regard to target affinity and absorption, distribution, metabolism, and excretion, and toxicity (ADMET) properties, and facilitates the rational design of novel compounds through structural modifications and fragment recombination.[64] Regarding NP-based drug discovery, CADD methods can assist in prioritizing test material of natural sources and identifying bioactive NPs and are able to support hit discovery, hit-to-lead and lead optimization phases.[23]

A key driver of the progress and success of CADD has been the advancement of structural biology techniques, notably X-ray crystallography (XR) and cryo-electron microscopy (cryo-EM), which have enabled the use of 3D structures of clinically relevant targets in their active states as templates for the screening of suitable ligands and for lead optimization.[65,66] Additional pivotal developments include the expansion of accessible chemical space through ultra-large virtual libraries encompassing millions of drug-like molecules, along with advances in computational methods and the availability of computing power capable of processing and analyzing vast amounts of data.[27,65]

One important and long-established CADD approach, **virtual screening** (VS), first described in 1997[67], refers to the computational search for drug candidates with potential activity against a molecular target in large digital compound databases.[68] Based on molecular similarity, VS enables the accelerated identification of so-called virtual hits (VHs), which can subsequently be tested in biological assays.[23,28,69,70] As predicted non-binding compounds are excluded *in silico*, VS typically provides higher hit rates compared to HTS and helps save time and resources.[71] VS offers particular advantages for NPs, as the isolation of bioactive compounds from natural sources typically yields only limited quantities.[72-74] This restricted availability poses a challenge for large-scale screening campaigns, which require sufficient material for extensive testing.[75] Although certain NPs can be purchased commercially, they are often expensive and may not meet the required purity standards.[76] As summarized in various reviews, VS has proven to be a successful strategy for the identification of promising NPs.[23,77,78] Interestingly, the term '**reverse virtual screening**' has also been introduced to describe the identification of unknown protein targets of bioactive compounds or additional targets for drug repositioning.[79]

However, VS is highly compatible with HTS, ultimately giving rise to the concept of **high-throughput virtual screening** (HTVS), a technique that combines VS of compound libraries with the principles of HTS to identify molecules most likely to exhibit modulatory effects on the

macromolecule of interest.[64,71,80] HTVS significantly reduces time and resource demands by identifying non-binding compounds *in silico*, thereby minimizing the number of compounds that need to be tested *in vitro*.[71]

In recent years, groundbreaking advancements in **AI** have profoundly impacted the field of pharmaceutical drug development.[37] AI denotes the overarching discipline dedicated to the development of methodologies and systems capable of addressing problems that typically require human intelligence, such as pattern recognition and decision making.[81,82] It compromises a broad spectrum of subfields such as **machine learning** (ML).[81] ML focuses on autonomous learning and the improvement of performance through experience, without manual human intervention, whereby the quality of outcomes is contingent upon the availability of large volumes of data.[37,81,82] **Deep learning** (DL), in turn a subset of ML, simulates cognitive functions of the human brain by employing artificial neural networks composed of multiple layers of ML, and also relies on substantial computational resources and large-scale datasets for effective model training.[82,83]

AI has also driven significant progress in NP chemistry and research, transforming the landscape of NP drug discovery and advancing the development of innovative therapies for complex diseases.[8,28,37] As summarized by Chi et al., AI is particularly well suited for the analysis of metabolomics and systems biology data due to its high efficiency in identifying hidden patterns within datasets, detecting signals in noisy backgrounds, and generating accurate predictions.[82] As further elaborated by the research group, AI approaches are also capable of analyzing MS and chromatographic data, identifying biomarkers and metabolites, and predicting metabolic pathways.[82] Additionally, AI is a powerful tool for avoiding the re-identification of already known compounds and focusing on the rapid discovery of new ones ('dereplication').[8,84,85]

However, there are several nontrivial limitations and challenges associated with AI in NP drug discovery.[82,86] Bender et al. have extensively discussed the limitations that AI may encounter when applied to various types of drug discovery-related datasets, including, for example, chemical structures, activity on protein targets, physiological data, and animal and clinical endpoints.[86] These limitations include, among others, difficulties in interpreting data related to chemical structures, such as dose-related properties that cannot be directly inferred from the structures themselves.[86] Additionally, the activity of isolated proteins may have limited relevance in *in vivo* systems, physiological data may be incomplete or insufficient, complicating the prediction of exposure-dependent effects, and *in vivo* endpoint readouts may be biased and context-dependent.[86]

Further key challenges for the application of AI in NP research include the requirement for large sample sizes, difficulties in interpreting of complex results, particularly due to the need for interdisciplinary expertise, and the necessity for improved model transparency, data security, and patient privacy protection.[82] This non-exhaustive overview underscores that, despite the considerable advantages AI brings to NP drug discovery, numerous challenges still remain.[82,86]

Returning to the fundamental principles of CADD, this approach encompasses two core methodological paradigms: (i) **ligand-based approaches**, which derive insights from the structural and physicochemical similarities of known bioactive ligands, and (ii) **target-based approaches**, which rely on the experimentally determined 3D structure of the target macromolecule.[18,64,66,87]

6.1 Ligand-based drug design (LBDD)

Ligand-based drug design (LBDD) comprises CADD techniques that rely exclusively on the structural information of ligands, without requiring the 3D structure of a specific biological target.[64,80] This approach aims to identify structural, property-based, or molecular parameter-based ('descriptor-based') similarities between test molecules and known bioactive ligands.[87,88] Consequently, appropriate mathematical methods combined with molecular descriptors are essential, the latter forming the basis for the quality of subsequent predictions.[89] Established methods include **pharmacophore modeling**[7,32,77], **quantitative structure-activity relationship** (QSAR)[64,90,91], and **ligand-based VS** (LBVS).[64,92]

6.2 Target-based drug design (TBDD)

When high-resolution structural data of a specific macromolecule (e.g., receptor, protein, enzyme, ion channel, structural element) are available, target-based drug design (TBDD) enables the screening and assessment of chemical complementarity between the target and potential active compounds in biological cell-free assays.[64,80,93,94] The fundamental assumption of TBDD is that modulation of a particular binding site on the target elicits a specific biological effect, regardless of the ligand, provided it exhibits the necessary interaction profile.[64] A major goal of this approach is to identify target-specific drug candidates with improved ADMET properties and minimized off-target effects.[64] Moreover, TBDD offers *in silico* access to 3D binding conformations of ligands.[87] Important TBDD techniques include **molecular docking**, **molecular dynamics**, and **target-based VS** (TBVS).[77,92,95] The increasing popularity of the TBDD approach is driven by the rapidly expanding knowledge of biological targets and the integration of these data into publicly accessible databases such as the Protein Data Bank (PDB) (https://www.rcsb.org/) and AlphaFold (https://alphafold.ebi.ac.uk/).[64,87,96]

As Sams-Dodd et al. pointed out, TBDD consists of five steps: (i) identification of the target, (ii) target validation, including consideration of its therapeutic relevance, (iii) development of a suitable assay, (iv) lead identification through compound library screening, and (v) optimization of the lead with respect to target affinity and selectivity.[97] Most critical aspects in this context are the identification and validation of the target.[75,94] On the one hand, it is often uncertain whether a specific target is directly associated with a desired biological outcome; on the other hand, unwanted

modulatory effects may arise as unintended consequences.[75,94,97] Furthermore, the target itself must possess an appropriate binding capacity to enable effective interaction with potential ligands ('ligandability').[75] As TBDD is conducted using cell-free *in vitro* assays, it cannot replicate the complex biological environment captured by phenotypic assays, which are performed in cells, tissues, organs, or living organisms.[94,98] Consequently, TBDD is associated with a higher risk of failure in subsequent stages of drug development.[66]

In **Publication I**, a TBVS approach was employed to identify NP compounds targeting the main protease (M^{pro}) and the papain-like protease (PL^{pro}) of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). High-resolution X-ray crystal structures of both viral proteases, co-crystallized with known non-covalently bound synthetic inhibitors, were retrieved from the PDB. These structural insights enabled molecular docking analyses, facilitating the prediction of potential NP-derived inhibitors against the SARS-CoV-2 proteases.

6.2.1 Publication I

Identification of Natural Products Inhibiting SARS-CoV-2 by Targeting Viral Proteases: A Combined *in Silico* and *in Vitro* Approach

Andreas Wasilewicz, Benjamin Kirchweger, Denisa Bojkova, Marie Jose Abi Saad, Julia Langeder, Matthias Bütikofer, **Sigrid Adelsberger**, Ulrike Grienke, Jindrich Cinatl Jr., Olivier Petermann, Leonardo Scapozza, Julien Orts, Johannes Kirchmair, Holger F. Rabenau, Judith M. Rollinger

Journal of Natural Products 2023, 86 (2), 264-275.
doi: 10.1021/acs.jnatprod.2c00843

Impact Factor (2023): 3.360; Plant Sciences: Q1

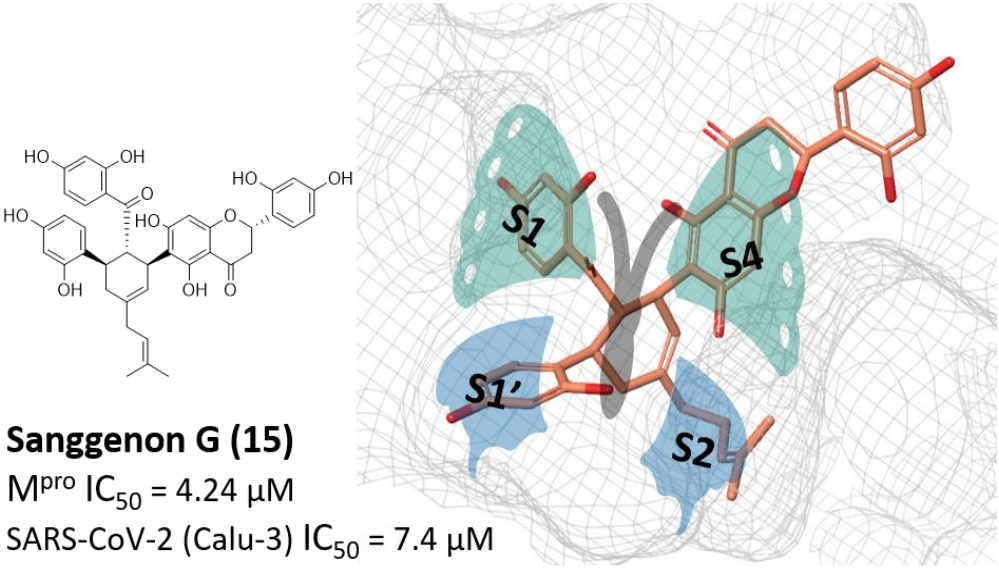


Figure A: Graphical abstract of **Publication I**

In this study, I contributed to the sample preparation and the experimental testing of hits from the TBVS in the PL^{pro} *in vitro* assay and supported processing of corresponding data.

JOURNAL OF NATURAL PRODUCTS

pubs.acs.org/jnp

Article

Identification of Natural Products Inhibiting SARS-CoV-2 by Targeting Viral Proteases: A Combined in Silico and in Vitro Approach

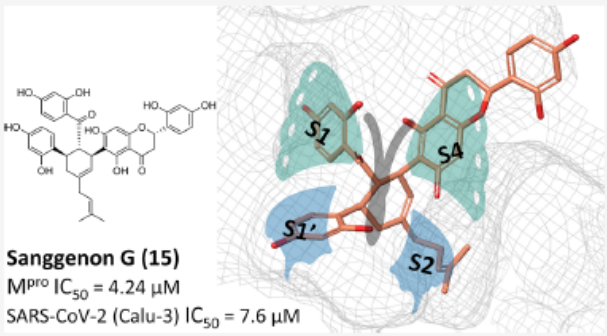
Andreas Wasilewicz, Benjamin Kirchweger, Denisa Bojkova, Marie Jose Abi Saad, Julia Langeder, Matthias Bütikofer, Sigrid Adelsberger, Ulrike Grienke, Jindrich Cinatl Jr., Olivier Petermann, Leonardo Scapozza, Julien Orts, Johannes Kirchmair, Holger F. Rabenau, and Judith M. Rollinger*

Cite This: *J. Nat. Prod.* 2023, 86, 264–275

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: In this study, an integrated in silico—in vitro approach was employed to discover natural products (NPs) active against SARS-CoV-2. The two SARS-CoV-2 viral proteases, i.e., main protease (M^{pro}) and papain-like protease (PL^{pro}), were selected as targets for the in silico study. Virtual hits were obtained by docking more than 140,000 NPs and NP derivatives available in-house and from commercial sources, and 38 virtual hits were experimentally validated in vitro using two enzyme-based assays. Five inhibited the enzyme activity of SARS-CoV-2 M^{pro} by more than 60% at a concentration of 20 μM, and four of them with high potency (IC_{50} < 10 μM). These hit compounds were further evaluated for their antiviral activity against SARS-CoV-2 in Calu-3 cells. The results from the cell-based assay revealed three mulberry Diels—Alder-type adducts (MDAAs) from *Morus alba* with pronounced anti-SARS-CoV-2 activities. Sanggenons C (12), O (13), and G (15) showed IC_{50} values of 4.6, 8.0, and 7.6 μM and selectivity index values of 5.1, 3.1 and 6.5, respectively. The docking poses of MDAAs in SARS-CoV-2 M^{pro} proposed a butterfly-shaped binding conformation, which was supported by the results of saturation transfer difference NMR experiments and competitive ^{1}H relaxation dispersion NMR spectroscopy.

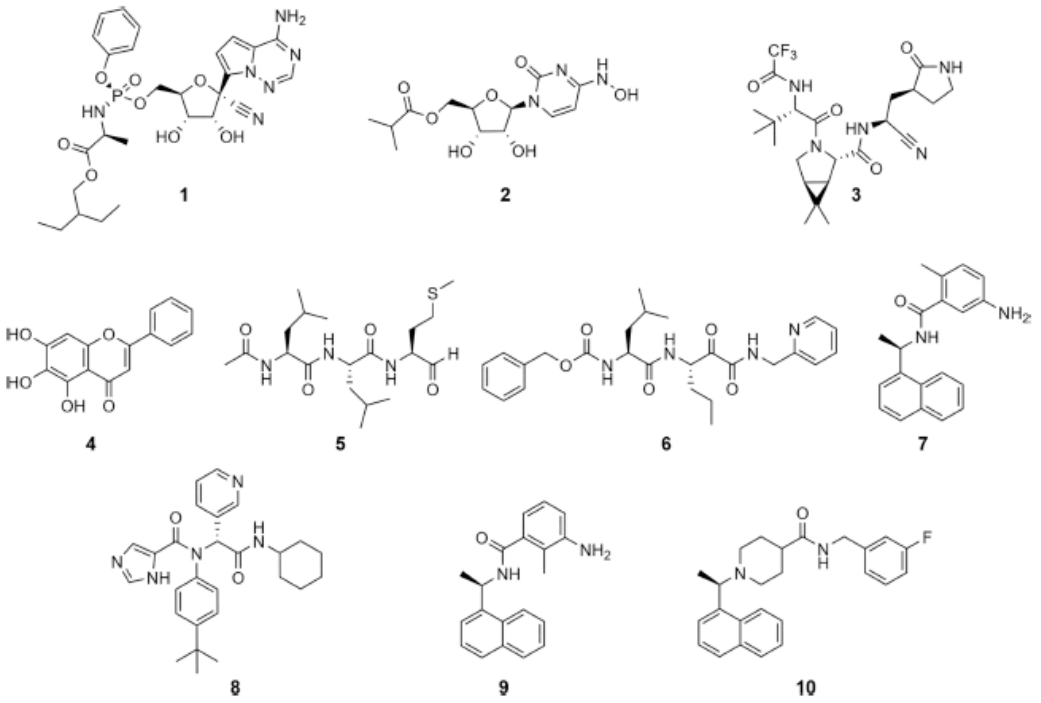


Since the outbreak of the COVID-19 pandemic, more than 6.4 million people worldwide have died by being infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).[1] The rapid development and availability of safe and effective vaccines represented a giant leap toward global immunity and significantly reduced the severity of the pandemic. However, disruptive seasonal waves of COVID-19 outbreaks are expected to contribute to a significant death toll in the future, comparable to seasonal influenza epidemics, which claim roughly 145,000 lives per year.[2] In addition, it is likely that new variants of SARS-CoV-2 will emerge that escape the immune system or increase transmissibility.[3] Hence, an additional cornerstone to combat COVID-19 besides vaccination and hygiene measures are antiviral drugs for the treatment of acute infections.[4,5] Within the last years, several small-molecule drugs and monoclonal antibodies have become available for the treatment of SARS-CoV-2. The nucleoside analogues remdesivir (1) and molnupiravir (2) are two repurposed drugs that have been developed initially for the treatment of Ebola virus and Venezuelan equine encephalitis virus, respectively (Chart 1).[6] The most recently approved drug, nirmatrelvir (3), targets the viral main protease (M^{pro}; 3C-like protease, nsp5) and was developed using a structure-based design approach.[7] Its clinical efficacy substantiates SARS-CoV-2 proteases as promising targets for drug discovery, similar to other viral infections that can be treated by targeting viral proteases.[8] For the present in silico study, the two SARS-CoV-2 proteases, M^{pro} and papain-like protease (PL^{pro}; domain of nsp3), were selected as targets. Both are cysteine proteases and essential for virus replication, as they cleave the viral polyprotein on distinct sites.[9] While M^{pro} leads to the formation of nonstructural proteins (nsp) 4–16,[10] PL^{pro} splits the polyprotein into nsp 1–3.[11] In addition, PL^{pro} also possesses deubiquitination and deISGylation activities that cause the release of K48-Ub2 and ISG15 from host cell proteins. Since both K48-Ub2 and ISG15 result in important cell signals upon infection, PL^{pro} is further able to counteract the host cell immune response.[12]

Received: September 20, 2022
Published: January 18, 2023

https://doi.org/10.1021/acs.jnatprod.2c00843
J. Nat. Prod. 2023, 86, 264–275

Chart 1. Chemical Structures of Approved Drugs against SARS-CoV-2 (1–3), Reported Inhibitors of M^{pro} and PL^{pro} Used as Positive Controls (4–7), and Co-Crystallized Ligands of the Protein Structures Used as Queries (8–10)



For both proteases, several structures including a bound inhibitor derived by X-ray crystallography are available from the Protein Data Bank (PDB). This allowed for the implementation of a structure-based in silico workflow for the identification of promising inhibitors of the SARS-CoV-2 proteases by virtual screening of natural product (NP) databases.

The focus on NPs was set as they are the most prolific source of new anti-infective drugs.[13] Some NPs were already reported to inhibit SARS-CoV-2 in vitro, e.g., oridonin,[14] cannabinoid acids,[15] and gallinamide A.[16] Baicalein (4) and andrographolide are also described as M^{pro} inhibitors with anti-SARS-CoV-2 activity.[17,18]

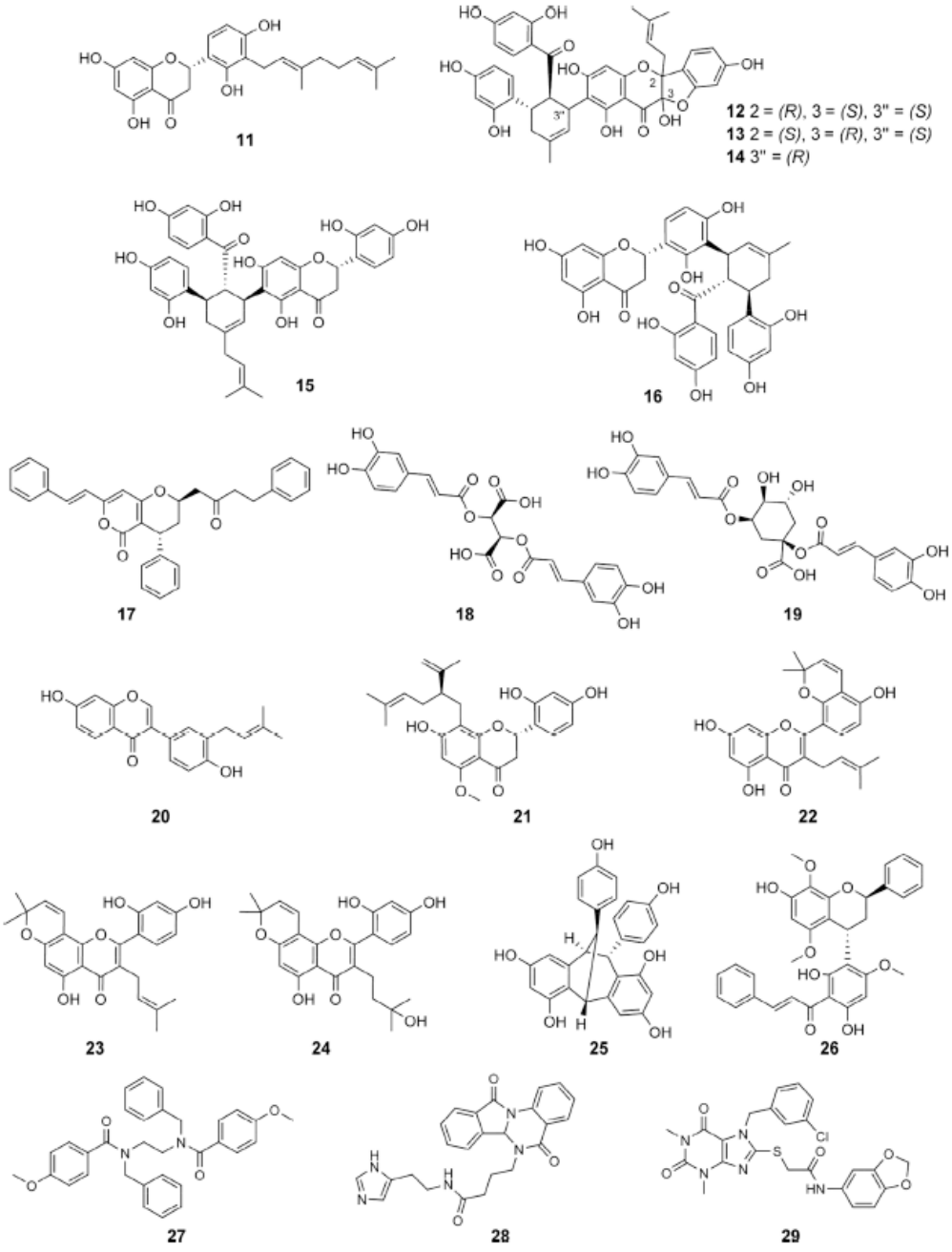
To discover NPs active against SARS-CoV-2, selected virtual hits (VHs) were experimentally validated for their inhibitory activity in enzyme-based assays and further tested for their antiviral effect in Calu-3 cells. Herein, mulberry Diels–Alder-type adducts (MDAAs) were discovered as a new class of M^{pro} inhibitors with potent anti-SARS-CoV-2 activities. Saturation transfer difference (STD) NMR experiments were conducted to corroborate the docking poses predicted in silico. STD amplification factors (STD-AF) gave insights into ligand–target interactions.

RESULTS AND DISCUSSION

Molecular Docking. A structure-based in silico approach was employed to identify NPs and NP derivatives with a high probability to competitively inhibit the SARS-CoV-2 protease M^{pro} or PL^{pro}. Three protein structures were selected for virtual screening: one structure for M^{pro} (PDB code: 6W63) and two structures for PL^{pro} (PDB codes: 7JN2 and 4OW0).[19−21] For PL^{pro}, two different structures were selected to consider the conformational flexibility of Tyr268 and Gln269, which are in proximity to the binding site (Figure S1, Supporting Information). Other major conformational changes were not observed in any of the X-ray structures of M^{pro} and PL^{pro} available at that time (October 2020). The fact that 4OW0 is an X-ray structure of SARS-CoV PL^{pro} and not SARS-CoV-2 PL^{pro} was regarded as negligible given their structurally similar binding site.[22] Docking of two NP databases against the substrate sites of the three enzyme structures was carried out with GOLD.[23] The databases used were (i) an in-house database (IHDB) of 1309 NPs that are physically available at the Department of Pharmaceutical Sciences, Division of Pharmacognosy, University of Vienna,[24] and (ii) the Molport Natural Products Database (MNPDB), comprising 140,164 commercially available NPs and NP derivatives.[25] Docking poses were ranked with the ChemPLP scoring function. Among the top-ranked docking poses, altogether 61 compounds were preselected by considering (i) the predicted protein–ligand interactions, (ii) the predicted binding site occupation, and (iii) the pose reproduction and comparison with that of known inhibitors (X77 (8), Y41 (9), S88 (10)) (Table S1, Supporting Information).[26] For the final selection of VHs to be purchased and experimentally validated, the following criteria were employed: (i) availability of high-quality compounds (purity ≥95%, according to the vendor), (ii) low probability of assay interference as assessed by Hitdexter 2.0[27] and FAFDrugs 4,[28] and (iii) relevance as known constituents of herbal remedies that are used traditionally for the treatment of respiratory infections.[29] Tables reporting all VHs including SMILES notations and information on potential assay interferences are provided in Tables S2–S4 (Supporting Information). Finally, 19 VHs were selected for testing in an M^{pro} enzyme inhibition assay (Chart 2), and a further 19 VHs were selected for evaluation in a PL^{pro} enzyme inhibition assay (Chart 3).

https://doi.org/10.1021/acs.jnatprod.2c00843
J. Nat. Prod. 2023, 86, 264−275

Chart 2

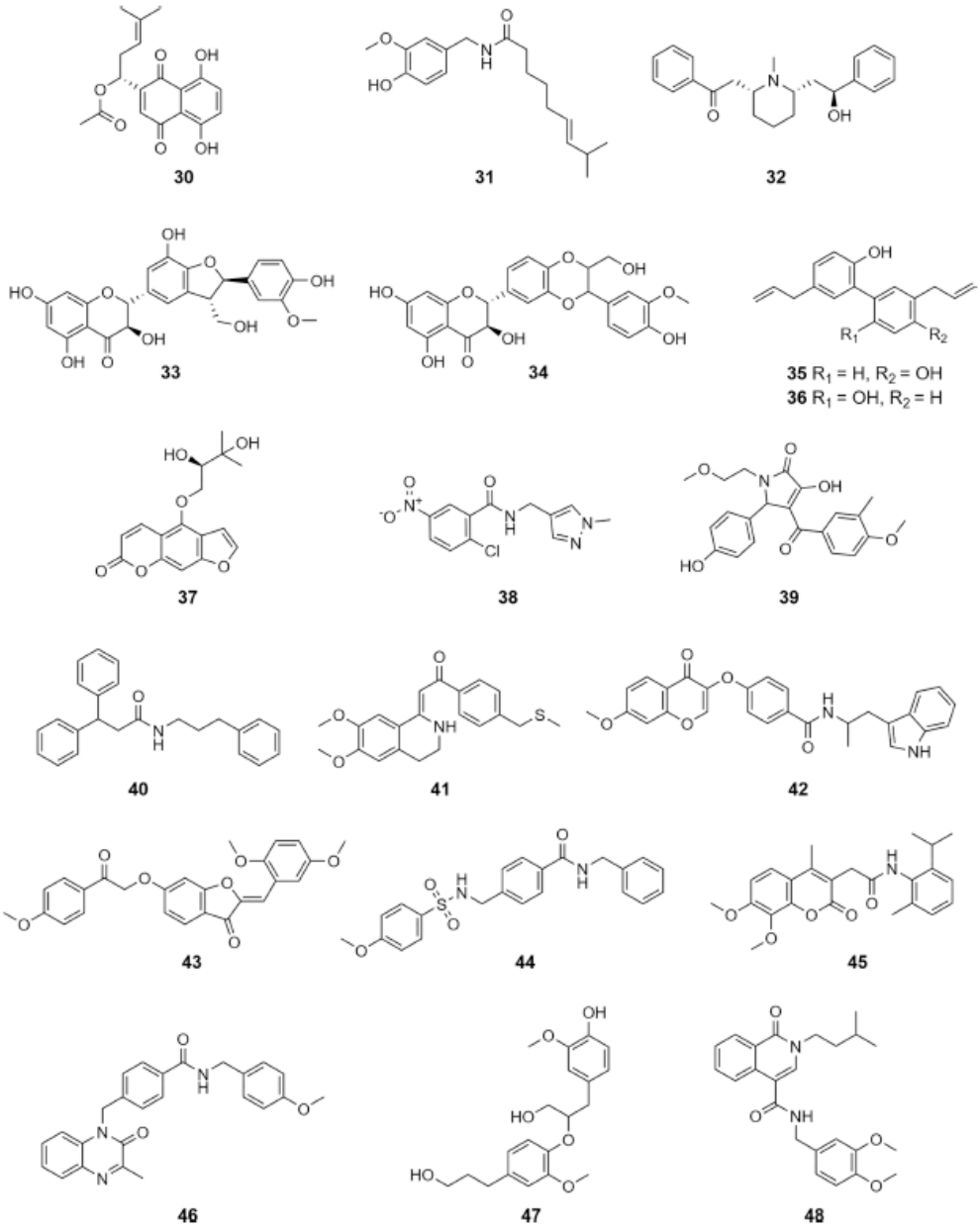


Experimental Validation. For validation of the virtual predictions, the 19 M^{pro} VHs, **11**–**29**, were tested for their potential to inhibit the protease activity of SARS-CoV-2 M^{pro} at 20 and 100 μM using a fluorescence resonance energy transfer quenching assay with 5-FAM-AVLQSGFR-Lys-(Dabcyl)-K-amide as substrate. For validation of PL^{pro} inhibition, compounds **30**–**48** were tested in a fluorogenic peptide cleavage assay at 20 and 100 μM using Z-RLRGG-AMC as substrate. In Tables 1 and 2, the inhibitory activities of the respective VHs against SARS-CoV-2 M^{pro} and PL^{pro} are reported. Considering an efficacy threshold of 60% enzyme inhibition at 20 μM, a hit rate of 13% was achieved.

Five out of the 19 VHs for M^{pro}, namely, **12**–**16**, showed more than 60% M^{pro} inhibition at 20 μM. Compounds **12**–**16** were identified as potent inhibitors, with IC_{50} values of 7.24, 6.70, 13.85, 4.82, and 5.25 μM, respectively (Figure S2, Supporting Information). In addition, compounds 22 and 23 were determined as weak M^{pro} inhibitors with IC_{50} values of 35.2 and 33.5 μM, respectively. It should be noted that baicalein (**4**) was selected as a positive control. It was previously reported as an NP with strong inhibitory activity against M^{pro} (IC_{50} values of 5.16 and 0.98 μM).[17,30] In our assay, however, **4** showed significant enzyme inhibition only at higher concentrations. An IC_{50} of 50.6 μM was determined, which is comparable to that reported by Hengphasatporn et

https://doi.org/10.1021/acs.jnatprod.2c00843
J. Nat. Prod. 2023, 86, 264−275

Chart 3



al.[31] One reason for the previously reported low IC_{50} values might be the absence of a reducing agent such as dithiothreitol (DTT) in the assay buffer. Compound **4** contains a catechol moiety, a common PAINS motif that is known to form unspecific covalent bonds to proteins, e.g., by binding to the thiol of a cysteine.[32] Since M^{pro} is a cysteine protease, it is plausible that the reported strong M^{pro} inhibition by **4** is due to covalent binding to the cysteine in the active center. In the presence of a reducing agent (as performed in the present study), **4** showed only moderate M^{pro} inhibitory activity.

Experimental validation of the PL^{pro} VHs was initially performed with a commercial PL^{pro} kit and revealed four VHs (**30**, **35**, **37**, **42**) with significant enzyme inhibition at 20 μM. These compounds were tested further for their antiviral activity against SARS-CoV-2 in two different cell lines (Caco-2, Calu-3, Table S5, Supporting Information). Herein, only **35** and its congener **36** were identified as moderately acting antiviral compounds. It should be noted that the supplier of the PL^{pro} kit did not provide details on substrate identity and buffer composition in the documentation and neither on later request. Hence, another assay based on the cleavage of Z-RLRGG-AMC was used to validate the results, as shown in Table 1. None of the tested VHs showed an enzyme inhibition >60% at 20 μM in this assay. However, compound **30** was confirmed as a weak PL^{pro} inhibitor with an IC_{50} value of 24.3 μM (Figure S3, Supporting Information). Interestingly, **30** as well as its nonacetylated derivative, shikonin, were previously reported to inhibit M^{pro} with IC_{50} values of 3.8 and 1.6 μM, respectively.[33] Ma and co-workers revealed that the reported M^{pro} inhibition of shikonins is an assay artifact.[34] However,

https://doi.org/10.1021/acs.jnatprod.2c00843
J. Nat. Prod. 2023, 86, 264−275

Table 1. Evaluation of Inhibitory Activities of VHs for SARS-CoV-2 M^{pro}

			SARS-CoV-2 M^{pro}		
			enzyme inhibition (%)		
compound no.	compound name		100 μM	20 μM	IC_{50} (μM)
4	baicalein	pos. control	90.0 ± 1.4	8.6[a]	50.6
5	calpain inhibitor II	pos. control			2.0
6	calpain inhibitor XII	pos. control			1.0
11	sanggenol A	VH	98.3 ± 2.4	31.9[a]	
12	sanggenon C	VH	99.5 ± 0.07	63.0[a]	7.2
13	sanggenon O	VH	99.4 ± 0.9	76.2[a]	6.7
14	sanggenon D	VH	98.3 ± 2.2	76.5[a]	13.9
15	sanggenon G	VH	99.6 ± 0.2	92.6[a]	4.8
16	kuwanon L	VH	97.6 ± 2.5	70.4[a]	5.3
17	katsumadain A	VH	47.7 ± 2.8	5.0[a]	
18	chicoric acid	VH	11.5 ± 0.6	0.7[a]	
19	cynarine	VH	9.8 ± 13.3	ND	
20	neobavaisoflavone	VH	84.9 ± 6.3	18.0[a]	
21	kurarinone	VH	99.2 ± 2.5	30.2[a]	
22	kuwanon A	VH	79.2 ± 2.4	45.5 ± 7.1	35.2
23	morusin	VH	84.0 ± 0.4	42.0 ± 9.6	33.5
24	morusinol	VH	71.2 ± 1.9	23.1 ± 2.8	
25	ampelopsin F	VH	37.5 ± 0.8	18.9 ± 2.2	
26	sarcandrone B	VH	7.5 ± 2.5	1.2 ± 2.5	
27	Molport-000-689-498	VH	28.2 ± 8.3	4.8 ± 4.0	
28	Molport-035-701-904	VH	59.7 ± 1.0	14.1 ± 5.1	
29	Molport-002-206-761	VH	7.3 ± 10.2	ND	

[a]For these values, the relative standard deviation of the fluorescence resonance energy transfer quenching assay of 6% applies; ND, not determinable.

Table 2. Evaluation of Inhibitory Activities of VHs for SARS-CoV-2 PL^{pro}

			SARS-CoV-2 PL^{pro}		
			enzyme inhibition (%)		
compound no.	compound name		100 μM	20 μM	IC_{50} (μM)
7	GRL-0617	pos. control		95.8 ± 1.0	0.8
30	acetylshikonin	VH	84.3 ± 4.6	41.7 ± 3.8	24.3
31	*trans*-capsaicin	VH	5.3 ± 0.1	5.6 ± 9.7	
32	lobelin	VH	14.6 ± 7.7	3.7 ± 2.6	
33	silicristin	VH	19.8 ± 2.0	3.4 ± 7.8	
34	silibinin	VH	16.8 ± 3.0	6.5 ± 3.6	
35	honokiol	VH	5.3 ± 3.2	4.9 ± 4.6	
36	magnolol	VH	23.5 ± 5.7	4.3 ± 6.6	
37	oxypeucedanin hydrate	VH	15.7 ± 3.2	10.9 ± 9.0	
38	MolPort-001-633-809	VH	6.4 ± 4.6	3.0 ± 0.7	
39	Molport-000-428-993	VH	42.0 ± 1.3	10.0 ± 5.4	
40	Molport-001-540-175	VH	5.0 ± 2.6	1.6 ± 8.8	
41	Molport-002-532-107	VH	14.6 ± 13.1	9.6 ± 5.8	
42	Molport-005-910-551	VH	8.8 ± 7.0	4.89 ± 1.3	
43	Molport-002-520-481	VH	14.3 ± 3.4	10.0 ± 5.4	
44	Molport-007-627-328	VH	1.5 ± 1.2	6.4 ± 4.2	
45	Molport-002-535-048	VH	2.6 ± 0.1	9.0 ± 4.3	
46	Molport-007-806-805	VH	14.0 ± 3.4	12.7 ± 6.9	
47	Molport-039-338-319	VH	0.5 ± 0.8	8.8 ± 5.6	
48	Molport-019-909-691	VH	0.3 ± 0.5	8.2 ± 2.3	

herein, it is the first time that **30** is reported as a PL^{pro} inhibitor by ruling out assay interference through the use of controls addressing compound aggregation, background fluorescence, and protein oxidation.

To corroborate the enzymatic inhibition results, the five virtually identified and experimentally confirmed protease inhibitors **12**–**16**, as well as the determined weak inhibitors **22**, **23**, and **30**, were tested for antiviral activities against SARS-CoV-2 in Calu-3 cells. Compounds **12**, **13**, and **15** inhibited the SARS-CoV-2 replication significantly with IC_{50} values of 4.6, 8.0, and 7.6 μM, respectively (Table 3, Figure 1A–C). Although compounds **12**, **13**, and **15** showed some cytotoxicity

https://doi.org/10.1021/acs.jnatprod.2c00843
J. Nat. Prod. 2023, 86, 264−275

Table 3. Evaluation of Anti-SARS-CoV-2 Activity and Cytotoxicity in Calu-3 Cells of Determined M^{pro} and PL^{pro} Inhibitors

compound	IC_{50} (μM)	CC_{50} (μM)	compound	IC_{50} (μM)	CC_{50} (μM)
remdesivir (**1**)	0.00125	>0.1	kuwanon L (**16**)	>20	97.7
sanggenon C (**12**)	4.6	23.6	kuwanon A (**22**)	35.6	47.7
sanggenon O (**13**)	8.0	24.9	morusin (**23**)	24.9	25.2
sanggenon D (**14**)	18.5	46.2	acetylshikonin (**30**)	>100	>100
sanggenon G (**15**)	7.6	49.1			

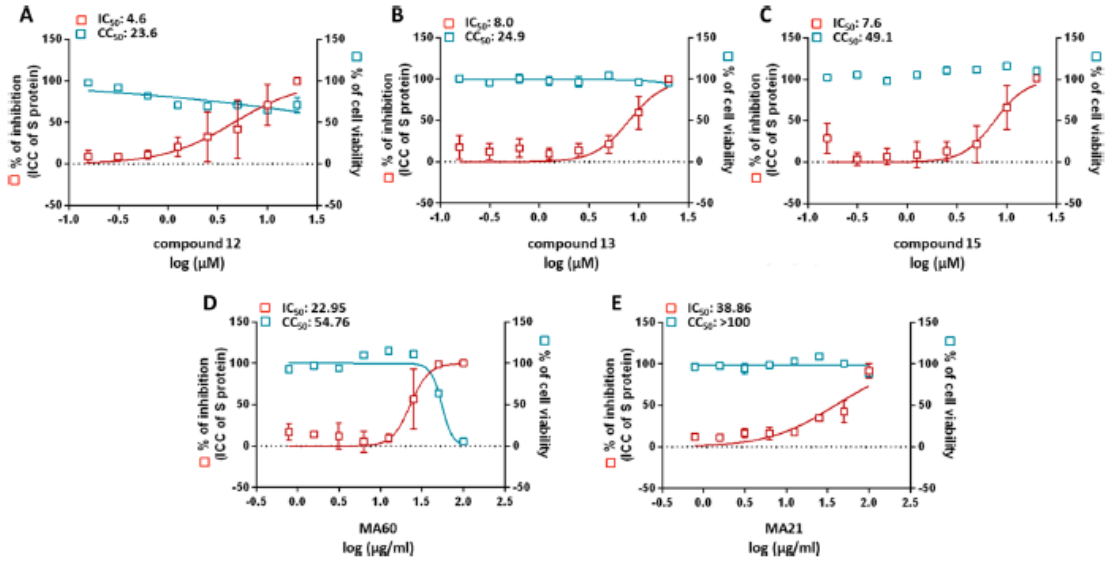


Figure 1. Dose–response curves showing anti-SARS-CoV-2 activity of compounds **12** (A), **13** (B), and **15** (C) and of the *Morus alba* root bark extracts MA60 (D) and MA21 (E) in Calu-3 cells, as determined by a spike staining assay.

at higher concentrations, they still exerted a decent selectivity toward SARS-CoV-2 with SIs of 5.1, 3.1, and 6.5, respectively.

Compounds **12** to **16** are MDAAs, which are derived from the cycloaddition of a chalcone and a dehydroprenylphenol such as prenylated flavonoids.[35] They are characteristic constituents of the root bark of the medicinal plant *Morus alba* L. (the white mulberry tree), traditionally used in Chinese medicine "sāng bái pí". In addition to testing the virtually predicted and isolated MDAAs, two extracts of the underlying herbal drug were prepared and also tested for their potential to inhibit the SARS-CoV-2 replication in Calu-3 cells. Thus, MA21 is a hydro-ethanolic extract with a content of 5.4% MDAAs, and MA60 represents an extract optimized toward a high yield of this bioactive compound class (29% MDAAs).[36] For MA60 and MA21, IC_{50} values of 23.0 and 38.9 μg/mL were determined, respectively (Figure 1D,E).

Docking Analysis and NMR Experiments. The substrate binding site of M^{pro} consists of four subsites, S1, S1′, S2, and S4 (Figure 2).[37] These subsites form a "butterfly-shaped" binding pocket. Well-known M^{pro} inhibitors such as **8** (Figure 2A) occupy all of these subsites. Molecular docking suggests a similar occupation of all four subsites by **12**–**16** (Figure 2B–F). The conserved benzoyl and cyclohexene moieties of these compounds are placed in the S1 and the lipophilic S2 pocket, respectively, and align well with the pyridine and phenyl ring of **8**. For **12**, **13**, **15**, and **16**, the phenyl ring of the chalcone is embedded in S1′, while the flavonoid part occupies S3. In contrast, **14** shows an inverted binding orientation regarding the S1′ and S4 subsites in comparison to the other compounds. One reason could be the stereochemical (*R*)-configuration of **14** at position C-3″ (the connection between chalcone and flavonoid), which is opposite to the (*S*)-configuration of **12** and **13**. Similar observations were reported on cocrystallized human immunodeficiency virus (HIV) protease inhibitors.[38] Regarding the suggested molecular interactions, all five MDAAs identified by docking could be predicted to form hydrogen bonds in S1 with Leu141, Asn142, and His163 and lipophilic interactions with residues forming S2. Additionally, all compounds except **14** were predicted to form hydrogen bonds with Thr26 and Glu166 in S1′ and S4, respectively. The hydrogen bonds to His163 and Glu166 and the lipophilic moiety in S2 were predicted to be also present for compound **8**. These interactions are considered to be essential for ligand binding.[39]

Validation of the ligand–target interactions for **12**–**16** was performed with STD NMR spectroscopy. The resulting STD NMR spectrum generated by the difference of recorded NMR spectra of ligand and protein with and without protein saturation (on- and off-resonance) indicates ligand protons likely to contact the protein. Hence, ligand moieties that are not involved in protein binding are likely to not show an STD signal. STD NMR experiments were conducted for each of the selected ligands in the presence of uniformly ^{15}N-labeled recombinant M^{pro}.[40] An (^{1}H, ^{15}N)-HSQC of the purified protein was recorded for quality control and comparison with the previously reported protocol to ensure that the protein adopts the fully active dimeric form in vitro (Figure S4, Supporting Information).[41] Only compounds **14** (Figure S5, Supporting Information) and **16** (Figure 3A) exhibited STD signals, while no STD signals were detected for **12**, **13**, and **15**,

https://doi.org/10.1021/acs.jnatprod.2c00843
J. Nat. Prod. 2023, 86, 264–275

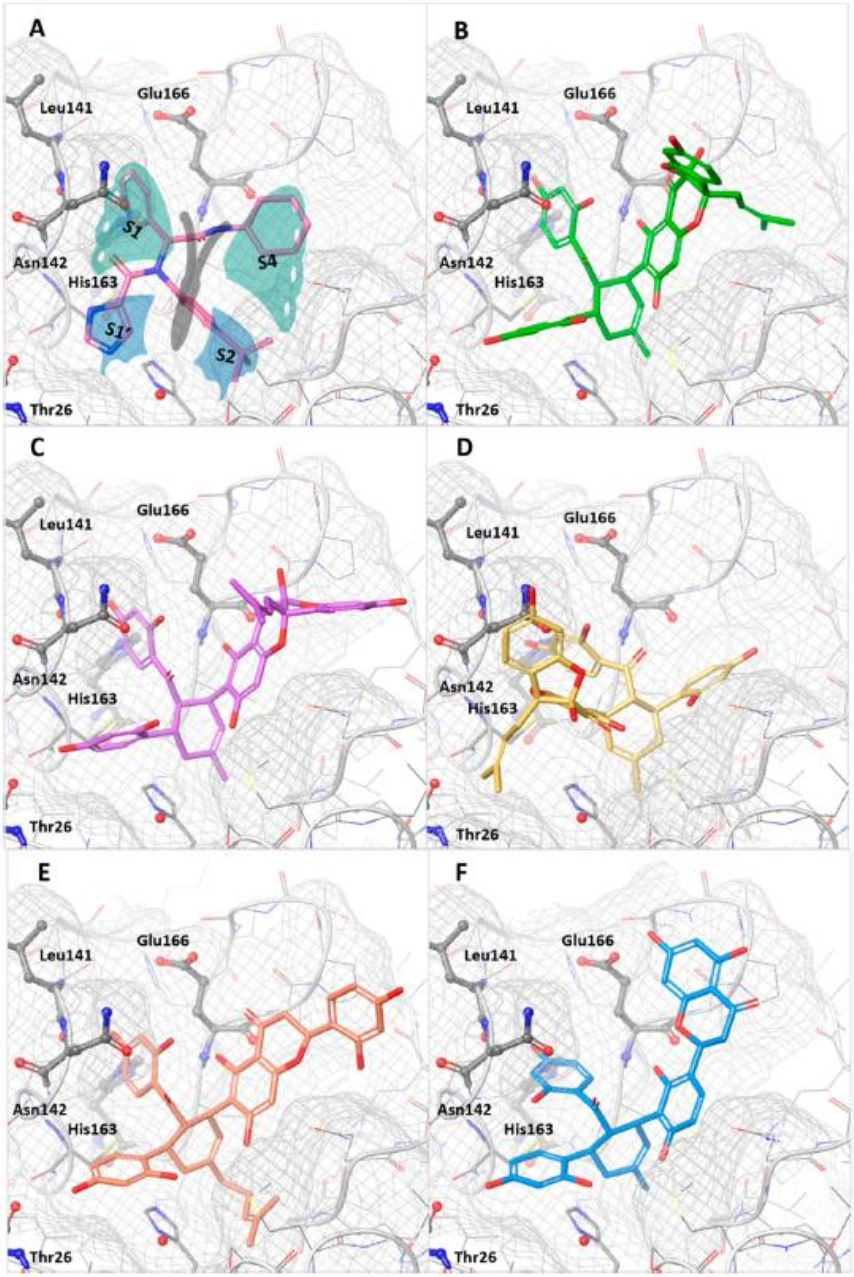


Figure 2. (A) Compound **8** (X77) bound to the butterfly-shaped substrate binding site of M^{pro} (PDB code: 6W63). (B–F) Docking poses of compounds **12** (green), **13** (purple), **14** (yellow), **15** (orange), and **16** (blue) in the binding site of 6W63. The substrate binding site is indicated as gray mesh. The water molecule is shown as a red sphere.

suggesting that these compounds bind to M^{pro} with a longer residence time in the binding site.[42] The STD-AF is used for the quantitation of the response of the ligand in interaction with the receptor of the protein.[43]

The STD-AF of **14** and **16**, indicating how buried molecular moieties are in the binding site, were compared with the proposed binding poses from molecular docking (Figure 3B, Figure S5, Supporting Information). The relatively uniform STD-AF data indicate that compounds **14** and **16** exhibit contact with the protein involving all moieties, supporting the large interaction surface proposed by the docking studies (Figure 3C).

To confirm the binding of compounds **12**, **13**, and **15** to M^{pro}, competitive relaxation dispersion experiments were carried out using compound **4** as a reporter ligand (Figures S6 and S7, Supporting Information). In the ^{1}H Carr–Purcell–Meiboom–Gill relaxation dispersion (CPMG-RD) experiments, transverse relaxation rates of the ligand resonances are enhanced upon protein binding and therefore report on the bound population of the ligand.[44] The addition of **12** or **15** to the mixture of **4** with M^{pro} reduced the relaxation rates of **4**, indicating its displacement from the binding site by **12** and **15** via competition (Figures S8 and S9, Supporting Information). Compound **13** however did not significantly displace **4**,

https://doi.org/10.1021/acs.jnatprod.2c00843
J. Nat. Prod. 2023, 86, 264–275

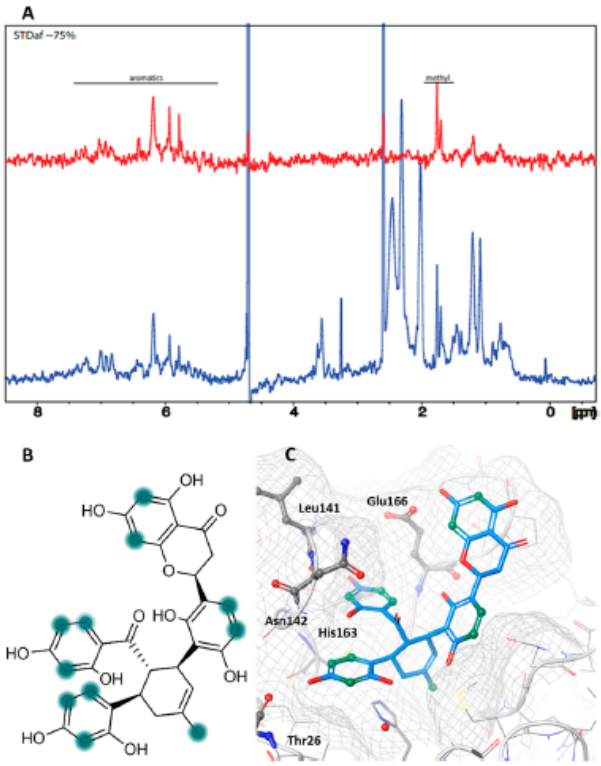


Figure 3. (A) STD NMR of M^{pro} and compound **16**. The STD difference spectrum is shown in red, with the STD off spectrum (1D) superposed in blue. (B) Proposed binding epitope map of **16**. (C) Combined presentation of the docking pose of **16** in 6W63. Determined interactions from STD NMR are highlighted in green (6B, 6C).

suggesting a noncompetition, while both showed an enhanced transverse relaxation rate (Figure S9, Supporting Information).

In conclusion, MDAAs were discovered as a unique NP structural class with anti-SARS-CoV-2 activity. The results presented combine computational, biochemical, and biophysical approaches. The VHs sanggenon C (**12**), sanggenon O (**13**), and sanggenon G (**15**) were confirmed not only to inhibit the predicted target but also to exert potent phenotypic antiviral activity in Calu-3 cells with IC_{50} values of 4.6, 8.0, and 7.6 μM, respectively. Despite showing a slight cytotoxicity, compounds **12** and **15** have reasonable SI values of >5. In agreement with the applied virtual screening, compounds **12**–**16** act as ligands of M^{pro} and inhibit its activity at low micromolar concentrations. M^{pro} has been shown to be essential for viral replication. Hence, its inhibition by the MDAAs identified is a plausible antiviral mode of action. According to the predicted binding mode and the STD NMR data, MDAAs occupy four subsites in the substrate binding pocket. In addition, the results from the CPMG-RD experiments further support the assumed competitive M^{pro} inhibition of MDAAs by binding to the substrate binding site. The two resorcyl residues flank the catalytic Cys134 in the S1 and S1′ pockets, while the cyclohexene ring with the methyl (**12**, **13**) or prenyl (**15**) groups is accommodated in the lipophilic S2 pocket. The benzopyran moiety loosely covers the solvent-accessible S4 pocket. Since **12**, **13**, and **15** all share a similar docking pose with **16**, we propose the chalcone moiety as a promising starting point for improved potent M^{pro} inhibitors.

These new findings further enrich the anti-infective and anti-inflammatory profile of the MDAAs investigated in this study. Previously, it has been shown that they exert (i) a pronounced in vitro anti-inflammatory activity,[36] (ii) an antibacterial activity against *Staphylococcus aureus* and *Streptococcus pneumoniae* accompanied by a distinct inhibition of pneumococcal biofilm formation,[45] (iii) an antiviral activity against different influenza virus strains, while (iv) being well tolerated by lung cells (A549, Calu-3).[36] This was also confirmed for an MDAA-enriched extract (MA60) from mulberry root bark.[36]

From an unbiased chemical perspective, which was the starting point for the present in silico-driven, target-based drug discovery study, the obtained results underline the utility of NP virtual screening for the discovery of genuine NPs with high to moderate drug-likeness as novel anti-SARS-CoV-2 scaffolds. From a multitarget and multicomponent perspective, the rich bioactivity profile of the MDAAs derived from previous findings and from this study provide a rationale for the long-standing use of *M. alba* root bark "sāng bái pí" in traditional Chinese medicine for the treatment of respiratory infections. It is worth mentioning that the root bark is the only plant part from white mulberry enriched with MDAAs. Commonly consumed mulberry fruits and other plant parts monographed in the Chinese pharmacopoeia such as leaves and twigs do not contain MDAAs studied in this work in detectable amounts.[36] MDAAs and MDAA-enriched multicomponent extracts of *M. alba* root bark warrant further studies to evaluate their potential for the treatment of Covid-19 and other acute respiratory infections.

EXPERIMENTAL SECTION

Preparation of Molecular Databases for Molecular Docking. Two molecular databases were employed for docking: the MNPDB, which is a database of commercially available NPs and derivatives listed in catalogues of Molport chemical suppliers (https://www.molport.com),[25] and the IHDB, which is a manually curated library of physically available compounds at the Division of Pharmacognosy of the Department of Pharmaceutical Sciences, University of Vienna.[24] The databases were prepared for docking using RDKit and CDK nodes in KNIME.[46−48] Entries with unconnected compounds were split with the RDKit salt stripper into individual entries.[46] Next, the element filter of CDK was applied to filter any molecules containing C, H, N, O, P, S, I, Br, Cl, or F.[47] The duplicate remover of LigandScout was employed to remove duplicates.[49,50] Only molecules of high to moderate drug-likeness (MW of 130−900 Da; cLogP of −1.3−7.2; calculated with LigandScout) were included. Molecules with more than three undefined chiral centers were removed; those with undefined chiral centers were flagged. In the final step, one 3D conformation for each molecule was generated using Icon from LigandScout.[50,51] For all molecules with an undefined stereocenter, all possible enantiomers were generated using LigPrep from Schrödinger.[52] The prepared databases with 1,309 and 140,164 molecules were stored in MOL2 file format.

Molecular Docking. For the selection of suitable target structures for docking, X-ray structures of M^{pro} and PL^{pro} from SARS-CoV and SARS-CoV-2 cocrystallized with noncovalent inhibitors bound to the substrate binding site and available from the Protein Data Bank (PDB, https://www.rcsb.org/) in October 2020 with a resolution <2.5 Å were considered. The electron density support of the key residues and ligands was verified in each structure, and structures with a concerning goodness of fit were rejected.[53,54] Furthermore, the binding site alignment feature of the Schrödinger modeling suite was applied to select potentially different conceivable protein conformations and ligand−target interactions.[55] Finally, one structure of M^{pro} (6W63) and two structures of PL^{pro} (7JN2, 4OW0) were selected for virtual screening. The PDB M^{pro} structure 6W63 contained two homodimer

https://doi.org/10.1021/acs.jnatprod.2c00843
J. Nat. Prod. 2023, 86, 264−275

chains from which chain A was selected, as its ligand better fits the electron density map. The structures were prepared with the Protein Preparation Wizard of the Schrödinger modeling suite with default settings. The addition of hydrogens and the generation of heterostates were performed using Epik followed by the refinement of H-bond assignment.[56,57] Water molecules beyond 5 Å from heteroatoms and fewer than two H-bonds to nonwaters were removed. In addition, all other solvents and ions were deleted. Finally, restrained minimization was performed using the OPLS3e force field.[58]

For virtual screening, the docking software GOLD (version 5.7.3) was employed.[23] The docking runs were configured with the Docking Wizard of the GOLD software suite. For each ligand, the number of docking runs per molecule was set to 15. The fitness function ChemPLP was selected as the scoring function using default parameters. For the genetic algorithm parameters that determine docking speed, default settings were applied ("slow/automatic"). The reliability of the molecular docking procedure was first assessed by self-docking, which successfully retrieved docking poses of the corresponding ligands from the protein structures at RMSD values lower than 1 Å (Figures S8−S10, Table S6, Supporting Information). The docking results from both databases for all three protein structures were evaluated independently according to the following criteria: the top 100 ranked molecules from the IHDB, according to the ChemPLP score, were considered for visual inspection. For the analysis of the MNPDB results, an arbitrary threshold was set for each protein structure with regard to the ChemPLP score of the redocked ligands (80.00 for 7JN2 and 4OW0, 70.00 for 6W63). All docking poses above this threshold were then evaluated by visual inspection.

Compounds and Chemicals. Compound **1** was obtained from Selleckchem (Houston, TX, USA). Compounds **4** (catalogue ID 465119), **7** (catalogue ID SML2961), **18** (catalogue ID C7243), **35** (catalogue ID H4914), and **36** (catalogue ID M3445) were obtained from Sigma-Aldrich (St. Louis, MO, USA). Compounds **5** (catalogue ID 15494869) and **6** (catalogue ID 15486559) were obtained from ThermoFisher. Compounds **11**−**16** were isolated from *Morus alba*.[45] Compound **17** was isolated from the seeds of *Alpinia katsumadai*.[59] Compound **19** (catalogue ID SRP04671c) was obtained from Sequoia Research Products, Ltd. (Pangbourne, UK). Compound **20** (catalogue ID PS020342) was obtained from Chengdu Push (Chengdu, People's Republic of China). Compound **21** was previously isolated from *Sophora flavescens*. Compounds **22** (catalogue ID CFN90834), **23** (catalogue ID CFN97083), **24** (catalogue ID CFN97086), **25** (catalogue ID CFN92340), **26** (catalogue ID CFN99300), and **47** (catalogue ID CFN97728) were obtained from ChemFaces (Wuhan, People's Republic of China). Compounds **27** (catalogue ID STK024437), **28** (catalogue ID STL521306), **29** (catalogue ID STK743870), **39** (catalogue ID STK798803), **40** (catalogue ID STK439583), **41** (catalogue ID STL515531), **42** (catalogue ID STL099426), **43** (catalogue ID STK923950), **45** (catalogue ID STL517522), and **48** (catalogue ID STL521306) were obtained from Vitas-M Laboratory (Wuhan, People's Republic of China). Compound **30** (catalogue ID 018-19191) was obtained from Wako Pure Chemical Industries (Osaka, Japan). Compound **31** was obtained from Merck (Darmstadt, Germany). Compound **32** was obtained from Boehringer-Ingelheim (Ingelheim am Rhein, Germany). Compounds **33** and **34** were obtained from Madaus (Köln, Germany). Compound **37** was isolated from roots of *Peucedanum ostruthium*,[60] and **38** (catalogue ID AK-968/41026577) was obtained from SPECS (Delft, Netherlands). **44** (catalogue ID C382-0112) and **46** (catalogue ID F671-0036) were obtained from ChemDiv (San Diego, CA, USA). The purity of compounds **12**−**16** was checked with UPLC-ELSD (Supporting Information). All compounds were prepared as 100 mM stock solutions in DMSO. Working stocks were diluted accordingly to achieve a final DMSO concentration of 0.2% in the experiments. For STD experiments, a 10 mM stock solution of compounds in d_6-DMSO was prepared. All compound stocks were stored at −20 °C until use.

Protein Production. The DNA sequence encoding M^{pro} SARS-CoV-2 (residues 1−306) was synthesized by GenScript (The Netherlands) with an N-terminal solubility tag and C-terminal affinity tag and cloned into a Pet-28a+ expression vector. The fusion protein encoded for glutathione *S*-transferase (GST) tag-tobacco etch virus (TEV) cleavage site−M^{pro}−6xHis-tag was expressed in *E. coli* RIL cells and purified. The bacteria were grown at 37 °C in labeled ^{15}N-M9 medium supplemented with kanamycin and chloramphenicol. When $A_{600\ nm}$ reached ∼0.8, the temperature was lowered to 30 °C and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and left overnight. The solubility tag was self-cleaved after expression, and cells were harvested in Linx6000 at 4 °C with a speed of 4000*g*. The cell lysate was passed through a HisTrap HP column (Cytivia). Then, TEV protease, produced in-house, was added to the fractions containing this protein, and the mixture was dialyzed overnight at 4 °C, followed by another pass through the HisTrap HP column. The relevant fractions were concentrated and applied to a Sephadex G75 gel filtration column (Cytivia), as reported.[41] The purified SARS-CoV-2 M^{pro} protein quality check was assessed from the 2D-[^{1}H, ^{15}N]-TROSY-HSQC spectrum (Figure S5). The protein was stable at room temperature for the time of the NMR experiments. Unlabeled M^{pro} was produced following the same protocol using Lysogeny broth medium instead of M9 medium.

M^{pro} Assay. Unlabeled M^{pro} was buffer exchanged to 50 mM Tris-HCl (pH 7.55), 1 mM EDTA, and 5 mM DTT. VHs that showed inhibitory activity at 20 μM were selected for further characterization. Enzyme activity measurements were performed in buffer (50 mM Tris-HCl (pH 7.55), 1 mM EDTA, and 5 mM DTT) at 30 °C. Mixtures of 1 μM M^{pro} with a series of increasing compound concentrations (0−300 μM) were preincubated for 30 min at 30 °C in black flat-bottomed 96-well plates (Costar #3915). Then, 10 μM of the fluorescently labeled substrate (crb1101508j, Discovery Peptides, Billingham, UK), containing an N-terminal 5-carboxyfluorescein (5-FAM) and the fluorescence quencher dabcyl, was added. Upon cleavage of the peptide by SARS-CoV-2 M^{pro}, the 5-FAM group and the dabcyl quencher are separated, and a fluorescence signal can be detected. Fluorescence signals were obtained by exciting the dye at 483 nm and reading at 530 nm using a Tecan M200 plate reader. Data were measured for 60 cycles every 30 s at 30 °C and analyzed using Rstudio.[61] For each data point, measurements were replicated five times.

Recording of NMR Data. Isotopically ^{15}N-labeled M^{pro} was buffer exchanged to 10 mM sodium phosphate buffer, pH 7.5 in D_2O using a concentrator. A 10 mM stock solution in d_6-DMSO was prepared for each compound. The samples were prepared with a ratio of 1:25 protein:ligand. Each tube contained 2 μM of protein in deuterated buffer containing 10 mM sodium phosphate pH 7.5, with 50 μM of ligand for a final volume of 500 μL in a 5 mm tube.

All experiments were recorded at 298 K on a 500 MHz Bruker Advance spectrometer equipped with a prodigy 5 mm TCI-cryoprobe and using TopSpin 3.1. All spectra were analyzed in Bruker TopSpin software versions 3.1 and 4.1.1. Compound structures were drawn by Chemdraw.[63]

The ^{1}H 1D proton experiments for the ligands were obtained with 1024 scans, 8k points, and an acquisition time of 0.51 s for a total experimental time of approximately 20 min. The STD experiments were acquired with 1024 scans, 8k points, and an interscan delay of 2.5 s, with an acquisition time of 0.59 s. The protein saturation was performed with a train of 50 ms Gaussian-shaped pulses, centered at −1 ppm, for a total of 2.5 s saturation time. The protein signals were filtered out using a spin-lock period of 30 ms and field strength of 10 kHz. The total duration of each STD experiment was approximately 90 min. An STD hit was considered when the STD NMR signals had a signal-to-noise ratio of min 6.5. Control experiments were repeated without the protein. Epitope mapping was determined by calculating the STD amplification factors.[43] The ^{1}H CPMG-RD experiments were conducted with relaxation delays set to 0, 8, 120, 200, and 400 ms.[64] The samples were prepared with same conditions as in the STD NMR experiments. The CPMG-RD experiments used an interscan relaxation delay of 2.5 s, 256 scans, and an acquisition time of 2.04 s with 8k points. The intensities were fitted to a monoexpansional decay using Rstudio.[61]

https://doi.org/10.1021/acs.jnatprod.2c00843
J. Nat. Prod. 2023, 86, 264−275

PL^{pro} Assay. VHs were tested for PL^{pro} inhibition using a fluorescence assay based on cleavage of a Z-RLRGG-AMC substrate (abcr; catalogue ID AB478009). The assay buffer contained 40 mM Tris-HCl (pH 7.8), 100 mM NaCl, 5 mM DTT, 0.01% (w/v) Triton X 100, and 0.1 mg/mL BSA. Compounds were preincubated with His-PL^{pro} (BPS-100735-1, BPS Biosciences) for 60 min. The substrate was added, and fluorescence at Exc/Emi 360/460 nm was measured every 3 min for 72 min using a Tecan Sparks plate reader (Männedorf, Switzerland). The final concentrations of PL^{pro} enzyme and substrate were 50 nM and 50 μM, respectively. The assay volume was 100 μL. Final test concentrations of VHs were 20 and 100 μM with 0.2% DMSO in assay buffer. Positive control 7 was tested at a final concentration of 10 μM and 0.2% DMSO in assay buffer. Experiments were performed in two well replicates. Two background control wells without enzymes were measured in parallel to correct for background fluorescence. Means of three independent experiments are presented. PL^{pro} inhibition curves were obtained by measuring the enzymatic activity at six compound concentrations. The IC_{50} values of the compounds were determined by plotting enzyme inhibition against various concentrations of the test inhibitor by using the dose–response curve in GraphPad Prism 6.[65]

Cells and Viruses. Calu-3 cells (ATCC, Manassas, VA, USA) were grown at 37 °C in minimal essential medium (MEM) supplemented with 10% fetal bovine serum (FBS), 100 IU/mL of penicillin, and 100 μg/mL of streptomycin. All culture reagents were purchased from Sigma-Aldrich. SARS-CoV-2 was isolated by using the Caco-2 clonal subline as previously described.[62] SARS-CoV-2 isolate (SARS-CoV-2/FFM7, MT358643) used in the experiment had undergone a maximum of three passages, and stocks were stored at −80 °C.

Antiviral Assay. Confluent layers of cells in 96-well plates were treated with decreasing concentrations of each test compound and subsequently infected with SARS-CoV-2 at a MOI of 0.01. Evaluation of inhibitory rate was performed by immunohistochemistry of viral spike protein 48 h postinfection. Briefly, cells were fixed with acetone–methanol (40:60) solution, and immunostaining was performed using a monoclonal antibody directed against the spike protein of SARS-CoV-2 (1:1500, Sinobiological), which was detected with a peroxidase-conjugated anti-rabbit secondary antibody (1:1,000, Dianova), followed by addition of AEC substrate. The spike positive area was scanned and quantified using a Bioreader 7000-F-Z-I microplate reader (Biosys). The results are expressed as percentage of inhibition relative to a virus control that received no test compound.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jnatprod.2c00843.

> Additional figures showing structural differences in the active site between PDB structures 7JN2 and 4OW0, concentration–response curves from M^{pro} and PL^{pro} assay, STD NMR data of compound **14**, results from CPMG-RD experiments, results from redocking studies; supporting tables containing virtual hits and additional in vitro antiviral activities for VHs of PL^{pro}; UPLC analyses of compounds **12**–**16** (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Judith M. Rollinger – *Department of Pharmaceutical Sciences, University of Vienna, 1090 Vienna, Austria;* orcid.org/0000-0001-6581-0774; Phone: +43-1-4277-55255; Email: Judith.Rollinger@univie.ac.at; Fax: +43-1-4277-855255

Authors

Andreas Wasilewicz – *Department of Pharmaceutical Sciences and Vienna Doctoral School of Pharmaceutical, Nutritional and Sport Sciences, University of Vienna, 1090 Vienna, Austria*

Benjamin Kirchweger – *Department of Pharmaceutical Sciences, University of Vienna, 1090 Vienna, Austria;* orcid.org/0000-0001-8948-1012

Denisa Bojkova – *Institute of Medical Virology, University Hospital Frankfurt, 60596 Frankfurt am Main, Germany*

Marie Jose Abi Saad – *Department of Pharmaceutical Sciences and Vienna Doctoral School of Pharmaceutical, Nutritional and Sport Sciences, University of Vienna, 1090 Vienna, Austria*

Julia Langeder – *Department of Pharmaceutical Sciences and Vienna Doctoral School of Pharmaceutical, Nutritional and Sport Sciences, University of Vienna, 1090 Vienna, Austria*

Matthias Bütikofer – *Swiss Federal Institute of Technology, Laboratory of Physical Chemistry, ETH Zurich, 8093 Zurich, Switzerland*

Sigrid Adelsberger – *Department of Pharmaceutical Sciences and Vienna Doctoral School of Pharmaceutical, Nutritional and Sport Sciences, University of Vienna, 1090 Vienna, Austria;* orcid.org/0000-0001-7310-3199

Ulrike Grienke – *Department of Pharmaceutical Sciences, University of Vienna, 1090 Vienna, Austria;* orcid.org/0000-0003-0305-9270

Jindrich Cinatl Jr. – *Institute of Medical Virology, University Hospital Frankfurt, 60596 Frankfurt am Main, Germany*

Olivier Petermann – *Pharmaceutical Biochemistry Group, School of Pharmaceutical Sciences and Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, 1205 Geneva, Switzerland*

Leonardo Scapozza – *Pharmaceutical Biochemistry Group, School of Pharmaceutical Sciences and Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, 1205 Geneva, Switzerland*

Julien Orts – *Department of Pharmaceutical Sciences, University of Vienna, 1090 Vienna, Austria;* orcid.org/0000-0003-3287-1532

Johannes Kirchmair – *Department of Pharmaceutical Sciences, University of Vienna, 1090 Vienna, Austria;* orcid.org/0000-0003-2667-5877

Holger F. Rabenau – *Institute of Medical Virology, University Hospital Frankfurt, 60596 Frankfurt am Main, Germany*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jnatprod.2c00843

Author Contributions

A.W. and B.K. contributed equally to this paper.

Funding

Open Access is funded by the Austrian Science Fund (FWF).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank R. Blažević for technical assistance and J. Zwirchmayr for providing compound 37 for experimental testing. This work was funded by the Austrian Science Fund (FWF, P34028 and P35115), the Swiss National Science Foundation (SNF-310030_192646), and Stiftung Krebsforschung Schweiz (KFS-4903-08-2019).

REFERENCES

(1) World Health Organisation. *WHO Coronavirus (COVID-19) Dashboard* https://covid19.who.int/ (accessed August 18, 2022).
(2) GBD 2017 Influenza Collaborators.. *Lancet Respir. Med.* **2019**, *7*, 69−89.
(3) Hu, J.; Peng, P.; Cao, X.; Wu, K.; Chen, J.; Wang, K.; Tang, N.; Huang, A. L. *Cell. Mol. Immunol.* **2022**, *19*, 293−295.
(4) Gao, K.; Wang, R.; Chen, J.; Tepe, J. J.; Huang, F.; Wei, G. W. *J. Med. Chem.* **2021**, *64*, 16922−16955.
(5) Zhou, S.; Hill, C. S.; Sarkar, S.; Tse, L. V.; Woodburn, B. M. D.; Schinazi, R. F.; Sheahan, T. P.; Baric, R. S.; Heise, M. T.; Swanstrom, R. *J. Infect. Dis.* **2021**, *224*, 415−419.
(6) Cully, M. *Nat. Rev. Drug Discovery* **2022**, *21*, 3−5.
(7) Halford, B. *ACS Cent. Sci.* **2022**, *8*, 405−407.
(8) Li, G.; De Clercq, E. In *Antiviral Discovery for Highly Pathogenic Emerging Viruses*; Muñoz-Fontela, C., Delgado, R., Eds.; Royal Society of Chemistry, 2022; pp 1−27.
(9) Gil, C.; Ginex, T.; Maestro, I.; Nozal, V.; Barrado-Gil, L.; Cuesta-Geijo, M.; Urquiza, J.; Ramírez, D.; Alonso, C.; Campillo, N. E.; Martinez, A. *J. Med. Chem.* **2020**, *63*, 12359−12386.
(10) Fan, K.; Wei, P.; Feng, Q.; Chen, S.; Huang, C.; Ma, L.; Lai, B.; Pei, J.; Liu, Y.; Chen, J.; Lai, L. *J. Biol. Chem.* **2004**, *279*, 1637−1642.
(11) Harcourt, B. H.; Jukneliene, D.; Kanjanahaluethai, A.; Bechill, J.; Severson, K. M.; Smith, C. M.; Rota, P. A.; Baker, S. C. *J. Virol.* **2004**, *78*, 13600−13612.
(12) Shin, D.; Mukherjee, R.; Grewe, D.; Bojkova, D.; Baek, K.; Bhattacharya, A.; Schulz, L.; Widera, M.; Mehdipour, A. R.; Tascher, G.; Geurink, P. P.; Wilhelm, A.; van der Heden van Noort, G. J.; Ovaa, H.; Müller, S.; Knobeloch, K. P.; Rajalingam, K.; Schulman, B. A.; Cinatl, J.; Hummer, G.; Ciesek, S.; Dikic, I. *Nature* **2020**, *587*, 657−662.
(13) Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2020**, *83*, 770−803.
(14) Littler, D. R.; Liu, M.; McAuley, J. L.; Lowery, S. A.; Illing, P. T.; Gully, B. S.; Purcell, A. W.; Chandrashekaran, I. R.; Perlman, S.; Purcell, D. F. J.; Quinn, R. J.; Rossjohn, J. *J. Biol. Chem.* **2021**, *297*, 101362.
(15) van Breemen, R. B.; Muchiri, R. N.; Bates, T. A.; Weinstein, J. B.; Leier, H. C.; Farley, S.; Tafesse, F. G. *J. Nat. Prod.* **2022**, *85*, 176−184.
(16) Ashhurst, A. S.; Tang, A. H.; Fajtová, P.; Yoon, M. C.; Aggarwal, A.; Bedding, M. J.; Stoye, A.; Beretta, L.; Pwee, D.; Drelich, A.; Skinner, D.; Li, L.; Meek, T. D.; McKerrow, J. H.; Hook, V.; Tseng, C. T.; Larance, M.; Turville, S.; Gerwick, W. H.; O'Donoghue, A. J.; Payne, R. J. *J. Med. Chem.* **2022**, *65*, 2956−2970.
(17) Su, H. X.; Yao, S.; Zhao, W. F.; Li, M. J.; Liu, J.; Shang, W. J.; Xie, H.; Ke, C. Q.; Hu, H. C.; Gao, M. N.; Yu, K. Q.; Liu, H.; Shen, J. S.; Tang, W.; Zhang, L. K.; Xiao, G. F.; Ni, L.; Wang, D. W.; Zuo, J. P.; Jiang, H. L.; Bai, F.; Wu, Y.; Ye, Y.; Xu, Y. C. *Acta Pharmacol. Sin.* **2020**, *41*, 1167−1177.
(18) Shi, T. H.; Huang, Y. L.; Chen, C. C.; Pi, W. C.; Hsu, Y. L.; Lo, L. C.; Chen, W. Y.; Fu, S. L.; Lin, C. H. *Biochem. Biophys. Res. Commun.* **2020**, *533*, 467−473.
(19) Mesecar, A. D. Structure of COVID-19 main protease bound to potent broad-spectrum non-covalent inhibitor X77. 2020, DOI: 10.2210/pdb6w63/pdb.
(20) Osipiuk, J.; Tesar, C.; Endres, M.; Lisnyak, V.; Maki, S.; Taylor, C.; Zhang, Y.; Zhou, Z.; Azizi, S. A.; Jones, K.; Kathayat, R.; Snyder, S. A.; Dickinson, B. C.; Joachimiak, A. The crystal structure of Papain-Like Protease of SARS CoV-2 in complex with PLP_Snyder441 inhibitor. 2020, DOI: 10.2210/pdb7jn2/pdb.
(21) Báez-Santos, Y. M.; Barraza, S. J.; Wilson, M. W.; Agius, M. P.; Mielech, A. M.; Davis, N. M.; Baker, S. C.; Larsen, S. D.; Mesecar, A. D. *J. Med. Chem.* **2014**, *57*, 2393−2412.
(22) Klemm, T.; Ebert, G.; Calleja, D. J.; Allison, C. C.; Richardson, L. W.; Bernardini, J. P.; Lu, B. G.; Kuchel, N. W.; Grohmann, C.; Shibata, Y.; Gan, Z. Y.; Cooney, J. P.; Doerflinger, M.; Au, A. E.; Blackmore, T. R.; van der Heden van Noort, G. J.; Geurink, P. P.; Ovaa, H.; Newman, J.; Riboldi-Tunnicliffe, A.; Czabotar, P. E.; Mitchell, J. P.; Feltham, R.; Lechtenberg, B. C.; Lowes, K. N.; Dewson, G.; Pellegrini, M.; Lessene, G.; Komander, D. *EMBO J.* **2020**, *39*, e106275.
(23) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727−748.
(24) Kirchweger, B.; Wasilewicz, A.; Fischhuber, K.; Tahir, A.; Chen, Y.; Heiss, E. H.; Langer, T.; Kirchmair, J.; Rollinger, J. M. *Planta Med.* **2022**, *88*, 794−804.
(25) Molport. *Natural Compound and Natural-Like Compound Database*. 2020 https://www.molport.com/shop/natural-compound-database (accessed on October 21, 2020).
(26) Fischer, A.; Smieško, M.; Sellner, M.; Lill, M. A. *J. Med. Chem.* **2021**, *64*, 2489−2500.
(27) Stork, C.; Chen, Y.; Šícho, M.; Kirchmair, J. *J. Chem. Inf. Model.* **2019**, *59*, 1030−1043.
(28) Lagorce, D.; Bouslama, L.; Becot, J.; Miteva, M. A.; Villoutreix, B. O. *Bioinformatics* **2017**, *33*, 3658−3660.
(29) Langeder, J.; Grienke, U.; Chen, Y.; Kirchmair, J.; Schmidtke, M.; Rollinger, J. M. *J. Ethnopharmacol.* **2020**, *248*, 112298.
(30) Xiao, T.; Cui, M.; Zheng, C.; Zhang, P.; Ren, S.; Bao, J.; Gao, D.; Sun, R.; Wang, M.; Lin, J.; Zhang, L.; Li, M.; Li, D.; Zhou, H.; Yang, C. *Inflammation* **2022**, *45*, 1076−1088.
(31) Hengphasatporn, K.; Wilasluck, P.; Deetanya, P.; Wangkanont, K.; Chavasiri, W.; Visitchanakun, P.; Leelahavanichkul, A.; Paunrat, W.; Boonyasuppayakorn, S.; Rungrotmongkol, T.; Hannongbua, S.; Shigeta, Y. *J. Chem. Inf. Model.* **2022**, *62*, 1498−1509.
(32) Baell, J. B. *J. Nat. Prod.* **2016**, *79*, 616−628.
(33) Zhang, Y.; Gao, H.; Hu, X.; Wang, Q.; Zhong, F.; Zhou, X.; Lin, C.; Yang, Y.; Wei, J.; Du, W.; Huang, H.; Zhou, H.; He, W.; Zhang, H.; Zhang, Y.; McCormick, P. J.; Fu, J.; Wang, D.; Fu, Y.; Lu, X.; Zhang, T.; Duan, J.; Qin, B.; Jiang, H.; Luo, J.; Zhang, Y.; Chen, Q.; Luo, Q.; Cheng, L.; Zhang, Z.; Zhang, J.; Li, J. *J. Virol.* **2022**, *96*, e0125321.
(34) Ma, C.; Tan, H.; Choza, J.; Wang, Y.; Wang, J. *Acta Pharmacol. Sin. B* **2022**, *12*, 1636−1651.
(35) Luo, S. Y.; Tang, Z. Y.; Li, Q.; Weng, J.; Yin, S.; Tang, G. H. T. *J. Org. Chem.* **2021**, *86*, 4786−4793.
(36) Langeder, J.; Döring, K.; Schmietendorf, H.; Grienke, U.; Schmidtke, M.; Rollinger, J. M. *J. Nat. Prod.* **2023**, DOI: 10.1021/acs.jnatprod.2c00481.
(37) Luttens, A.; Gullberg, H.; Abdurakhmanov, E.; Vo, D. D.; Akaberi, D.; Talibov, V. O.; Nekhotiaeva, N.; Vangeel, L.; De Jonghe, S.; Jochmans, D.; Krambrich, J.; Tas, A.; Lundgren, B.; Gravenfors, Y.; Craig, A. J.; Atilaw, Y.; Sandström, A.; Moodie, L. W. K.; Lundkvist, Å.; van Hemert, M. J.; Neyts, J.; Lennerstrand, J.; Kihlberg, J.; Sandberg, K.; Danielson, U. H.; Carlsson, J. *J. Am. Chem. Soc.* **2022**, *144*, 2905−2920.
(38) Meyer, E. F.; Botos, I.; Scapozza, L.; Zhang, D. *Drug Discovery Des.* **1995**, *3*, 168−195.
(39) Unoh, Y.; Uehara, S.; Nakahara, K.; Nobori, H.; Yamatsu, Y.; Yamamoto, S.; Maruyama, Y.; Taoda, Y.; Kasamatsu, K.; Suto, T.; Kouki, K.; Nakahashi, A.; Kawashima, S.; Sanaki, T.; Toba, S.; Uemura, K.; Mizutare, T.; Ando, S.; Sasaki, M.; Orba, Y.; Sawa, H.; Sato, A.; Sato, T.; Kato, T.; Tachibana, Y. *J. Med. Chem.* **2022**, *65*, 6499−6512.
(40) Mayer, M.; Meyer, B. *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 1784−1788.
(41) Altincekic, N.; Korn, S. M.; Qureshi, N. S.; Dujardin, M.; Ninot-Pedrosa, M.; Abele, R.; Abi Saad, M. J.; Alfano, C.; Almeida, F. C. L.; Alshamleh, I.; de Amorim, G. C.; Anderson, T. K.; Anobom, C. D.; Anorma, C.; Bains, J. K.; Bax, A.; Blackledge, M.; Blechar, J.; Böckmann, A.; Brigandat, L.; Bula, A.; Bütikofer, M.; Camacho-Zarco, A. R.; Carlomagno, T.; Caruso, I. P.; Ceylan, B.; Chaikuad, A.; Chu, F.; Cole, L.; Crosby, M. G.; de Jesus, V.; Dhamotharan, K.; Felli, I. C.; Ferner, J.; Fleischmann, Y.; Fogeron, M. L.; Fourkiotis, N. K.; Fuks, C.; Fürtig, B.; Gallo, A.; Gande, S. L.; Gerez, J. A.; Ghosh, D.; Gomes-Neto, F.; Gorbatyuk, O.; Guseva, S.; Hacker, C.; Häfner, S.; Hao, B.; Hargittay, B.; Henzler-Wildman, K.; Hoch, J. C.; Hohmann, K. F.; Hutchison, M. T.; Jaudzems, K.; Jović, K.; Kaderli, J.; Kalniņš, G.; Kaņepe, I.; Kirchdoerfer, R. N.; Kirkpatrick, J.; Knapp, S.;

https://doi.org/10.1021/acs.jnatprod.2c00843
J. Nat. Prod. 2023, 86, 264−275

Krishnathas, R.; Kutz, F.; Zur Lage, S.; Lambertz, R.; Lang, A.; Laurents, D.; Lecoq, L.; Linhard, V.; Löhr, F.; Malki, A.; Bessa, L. M.; Martin, R. W.; Matzel, T.; Maurin, D.; McNutt, S. W.; Mebus-Antunes, N. C.; Meier, B. H.; Meiser, N.; Mompeán, M.; Monaca, E.; Montserret, R.; Mariño Perez, L.; Moser, C.; Muhle-Goll, C.; Neves-Martins, T. C.; Ni, X.; Norton-Baker, B.; Pierattelli, R.; Pontoriero, L.; Pustovalova, Y.; Ohlenschläger, O.; Orts, J.; Da Poian, A. T.; Pyper, D. J.; Richter, C.; Riek, R.; Rienstra, C. M.; Robertson, A.; Pinheiro, A. S.; Sabbatella, R.; Salvi, N.; Saxena, K.; Schulte, L.; Schiavina, M.; Schwalbe, H.; Silber, M.; Almeida, M. D. S.; Sprague-Piercy, M. A.; Spyroulias, G. A.; Sreeramulu, S.; Tants, J. N.; Tārs, K.; Torres, F.; Töws, S.; Treviño, M.; Trucks, S.; Tsika, A. C.; Varga, K.; Wang, Y.; Weber, M. E.; Weigand, J. E.; Wiedemann, C.; Wirmer-Bartoschek, J.; Wirtz Martin, M. A.; Zehnder, J.; Hengesbach, M.; Schlundt, A. *Front. Mol. Biosci.* **2021**, *8*, 653148.
(42) Hall, J. L.; Sohail, A.; Cabrita, E. J.; Macdonald, C.; Stockner, T.; Sitte, H. H.; Angulo, J.; MacMillan, F. *Sci. Rep.* **2020**, *10*, 16483.
(43) Mayer, M.; Meyer, B. *J. Am. Chem. Soc.* **2001**, *123*, 6108−6117.
(44) Moschen, T.; Grutsch, S.; Juen, M. A.; Wunderlich, C. H.; Kreutz, C.; Tollinger, M. *J. Med. Chem.* **2016**, *59*, 10788−10793.
(45) Grienke, U.; Richter, M.; Walther, E.; Hoffmann, A.; Kirchmair, J.; Makarov, V.; Nietzsche, S.; Schmidtke, M.; Rollinger, J. M. *Sci. Rep.* **2016**, *6*, 27156.
(46) RDKit: Open-source cheminformatics. https://www.rdkit.org.
(47) Steinbeck, C.; Han, Y.; Kuhn, S.; Horlacher, O.; Luttmann, E.; Willighagen, E. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 493−500.
(48) Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. In *Data Analysis, Machine Learning and Applications*; Preisach, C.; Burkhardt, H.; Schmidt-Thieme, L.; Decker, R., Eds.; Springer: Berlin, 2008; pp 319−326.
(49) Wolber, G.; Langer, T. *J. Chem. Inf. Model.* **2005**, *45*, 160−169.
(50) *LigandScout* version 4.4.5; Inte:Ligand GmbH: Vienna, Austria, 2020.
(51) Poli, G.; Seidel, T.; Langer, T. *Front. Chem.* **2018**, *6*, 229.
(52) Schrödinger release 2020-3: *LigPrep*; Schrödinger, LLC: New York, 2020.
(53) Chakraborti, S.; Hatti, K.; Srinivasan, N. *Int. J. Mol. Sci.* **2021**, *22*, 6830.
(54) Yang, H.; Peisach, E.; Westbrook, J. D.; Young, J.; Berman, H. M.; Burley, S. K. *J. Appl. Crystallogr.* **2016**, *49*, 1081−1084.
(55) Schödinger release 2020-3: *Maestro 12.5*; Schrödinger LLC: New York, 2020.
(56) Schrödinger release 2020-3: *Epik*; Schrödinger LLC: New York, 2020.
(57) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. *J. Comput. Aided Mol. Des.* **2007**, *21*, 681−691.
(58) Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D. *J. Chem. Theory Comput.* **2019**, *15*, 1863−1874.
(59) Grienke, U.; Schmidtke, M.; Kirchmair, J.; Pfarr, K.; Wutzler, P.; Dürrwald, R.; Wolber, G.; Liedl, K. R.; Stuppner, H.; Rollinger, J. M. *J. Med. Chem.* **2010**, *53*, 778−786.
(60) Zwirchmayr, J.; Grienke, U.; Hummelbrunner, S.; Seigner, J.; de Martin, R.; Dirsch, V. M.; Rollinger, J. M. *Biomolecules.* **2020**, *10*, 679.
(61) RStudio Team. *RStudio: Integrated Development for R*; RStudio, PBC: Boston, MA, 2020; http://www.rstudio.com.
(62) Evans, D. A. *Angew. Chem., Int. Ed. Engl.* **2014**, *53*, 11140−11145.
(63) Aguilar, J. A.; Nilsson, M.; Bodenhausen, G.; Morris, G. A. *Chem. Commun.* **2012**, *48*, 811−813.
(64) Swift, M. L. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 411−412.
(65) Bojkova, D.; Klann, K.; Koch, B.; Widera, M.; Krause, D.; Ciesek, S.; Cinatl, J.; Münch, C. *Nature* **2020**, *583*, 469−472.

https://doi.org/10.1021/acs.jnatprod.2c00843
J. Nat. Prod. 2023, 86, 264−275

Supporting information

Identification of Natural Products Inhibiting SARS-CoV-2 by Targeting the Viral Proteases: A Combined in Silico and in Vitro Approach

Andreas Wasilewicz,[†,‡,#] Benjamin Kirchweger,[†,#] Denisa Bojkova,[§] Marie Jose Abi Saad,[†,‡] Julia Langeder,[†,‡] Matthias Bütikofer,[⊥] Sigrid Adelsberger,[†,‡] Ulrike Grienke,[†] Jindrich Cinatl Jr,[§] Olivier Petermann,[I,○] Leonardo Scapozza,[I,○] Julien Orts,[†] Johannes Kirchmair,[†] Holger F. Rabenau,[§] Judith M. Rollinger,[†,*]

[†] Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria

[‡] Vienna Doctoral School of Pharmaceutical, Nutritional, Sport Sciences, University of Vienna, Josef-Holaubek-Platz-2, 1090 Vienna, Austria

[§] Institute of Medical Virology, University Hospital Frankfurt, Paul-Ehrlich-Straße 40, 60596 Frankfurt am Main, Germany

[⊥] Swiss Federal Institute of Technology, Laboratory of Physical Chemistry, ETH Zurich, Vladimir-Prelog-Weg 1-5/10 8093 Zurich, Switzerland

[I] Pharmaceutical Biochemistry Group, School of Pharmaceutical Sciences, University of Geneva, Switzerland

[○] Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, Switzerland

Table of content

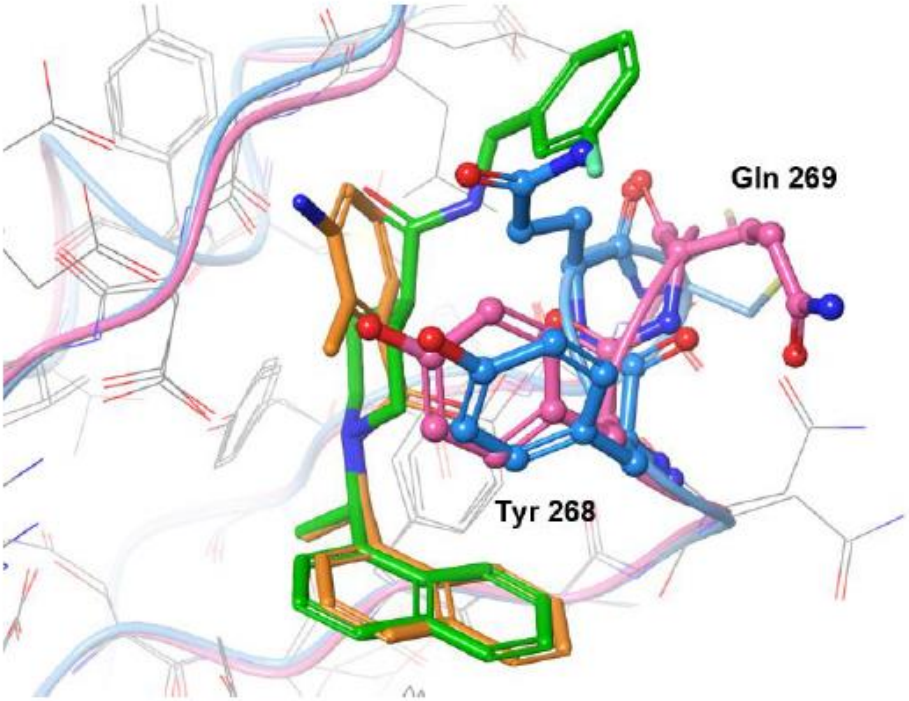


Figure S1. Superimposition of the subtrate-binding sites of PL^{pro} structures 7JN2 (blue) and 4OW0 (pink) and their respective co-crystallized ligands Y41 (orange) and S88 (green).

Table S1. Distribution of virtual hits across protein structures

		Virtual Hits		
Protease	PDB code	IHDB	MNPDB	In Total
$M^{pro}/3CL^{pro}$	6W63	12	13	25
PL^{pro}	4OW0	5	6	11
	7JN2	13	12	25
				61

3

Table S2. Predicted virtual hits from M^{pro} structure 6W63 ranked by ChemPLP score. Tested hits are presented in bold.

compound name	CAS number	SMILES code	ChemPLP score	database origin	HitDexter 2.0	FAF Drugs 4
15, Sanggenon G	**85698-31-3**	**CC(C)=CCCC1=C[C@@H]([C@H]([C@@H](C1)C1=C(O)C=C(O)C=C1)C(=O)C1=CC=C(O)C=C1O)C1=C(O)C2=C(O[C@@H](CC2=O)C2=C(O)C=C(O)C=C2)C=C1O**	**89,47**	**INDB**	**Non-promiscous in CDRA, promiscous in PSA**	**Accepted**
MolPort-005-952-021		C[C@@]1(CC[C@]2([C@@H](C1)C(C[C@@H]2NC(=O)C3=CC=CC=C3)(C)C)CCC(=O)NCC4=CN=CC=C4)O	86,13	MNPDB	Non-promiscous in PSA and CDRA	Accepted
5,7,2',4'-tetrahydroxy-3-geranylflavone	376361-87-4	CC(=CCC/C(=C/CC1=C(OC2=CC(=CC(=C2C1=O)O)O)C3=C(C=C(C=C3)O)O)/C)C	81,4742	MNPDB	Promiscous in PSA and CDRA	Accepted
29, MolPort-002-206-761		**CN1C2=C(C(=O)N(C1=O)C)N(C(=N2)SCC(=O)NC3=CC4=C(C=C3)OCO4)CC5=CC(=CC=C5)Cl**	**80,98**	**MNPDB**	**Non-promiscous in PSA and CDRA**	**Intermediate (halogenure)**
25, Ampelopsin F	**151487-08-0**	**C1=CC(=CC=C1[C@@H]2[C@H]3[C@@H](C4=C([C@@H]2C5=C3C=C(C=C5O)O)O)C=C(C=C4O)O)C6=CC=C(C=C6)O)O**	**80,84**	**MNPDB**	**Promiscous in PSA and CDRA**	**Intermediate (phenol)**
11, Sanggenol A	**174423-30-4**	**CC(=CCC/C(=C/CC1=C(C=CC(=C1O)[C@@H]2CC(=O)C3=C(C=C(C=C3O2)O)O)O)/C)C**	**80,74**	**INDB**	**Promiscous in PSA and CDRA**	**Accepted**
28, MolPort-035-701-904		**C1=CC=C2C(=C1)C3N(C(=O)C4=CC=CC=C4N3C2=O)CCCC(=O)NCCC5=CN=CN5**	**80,08**	**MNPDB**	**Non-promiscous in PSA and CDRA**	**Accepted**
27, MolPort-000-689-498		**COC1=CC=C(C=C1)C(=O)N(CCN(CC2=CC=CC=C2)C(=O)C3=CC=C(C=C3)OC)CC4=CC=CC=C4**	**77,31**	**MNPDB**	**Non-promiscous in PSA and CDRA**	**Accepted**
26, Sarcandrone B	**1190225-48-9**	**COC1=C(C(=C(C(=C1)O)C(=O)/C=C/C2=CC=CC=C2)O)[C@H]3C[C@@H](OC4=C(C(=CC(=C34)OC)O)OC)C5=CC=CC=C5**	**77,23**	**MNPDB**	**Non-promiscous in PSA and CDRA**	**Rejected (covalent binder)**
12, Sanggenon C	**80651-76-9**	**CC1=C[C@@H]([C@@H]([C@H](C1)C2=C(C=C(C=C2)O)O)C(=O)C3=C(C=C(C=C3)O)O)C4=C(C5=C(C=C4O)O[C@@]6(C7=C(C=C(C=C7)O)O[C@@]6(C5=O)O)CC=C(C)C)O**	**76,61**	**INDB**	**Non-promiscous in CDRA, promiscous in PSA**	**Accepted**
14, Sanggenon D		**CC(C)=CCC12OC3=C(C(O)=C([C@@H]4C=C(C)C[C@@H]([C@H]4C(=O)C4=CC=C(O)C=C4O)C4=CC=C(O)C=C4O)C(O)=C3)C(=O)C1(O)OC1=C2C=CC(O)=C1**	**76,26**	**INDB**	**Non-promiscous in CDRA, promiscous in PSA**	**Accepted**
MolPort-000-717-985	672272-01-4	CN1C=C(C2=CC=CC=C21)/C=C(/C(=O)NCCCO)\NC(=O)C3=CC(=C(C=C3)OC)OC	75,77	MNPDB	Non-promiscous in PSA and CDRA	Rejected (covalent binder)
Lupulone	468-28-0	CC(C)CC(=O)C1=C(C(=C(C(C1=O)(CC=C(C)C)CC=C(C)C)O)CC=C(C)C)O	75,35	INDB	Non-promiscous in PSA and CDRA	Rejected (covalent binder)
24, Morusinol	**62949-93-3**	**CC1(C=CC2=C(O1)C=C(C3=C2OC(=C(C3=O)CCC(C)(C)O)C4=C(C=C(C=C4)O)O)O)C**	**75,03**	**MNPDB**	**Non-promiscous in PSA and CDRA**	**Accepted**
8-Isomulberrin hydrate	1432063-35-8	CC1(CCC2=C(O1)C=C(C3=C2OC(=C(C3=O)CCC(C)(C)O)C4=C(C=C(C=C4)O)O)O)C	74,25	MNPDB	Non-promiscous in PSA and CDRA	Accepted
22, Kuwanon A	**62949-77-3**	**CC(=CCC1=C(OC2=CC(=CC(=C2C1=O)O)O)C3=C4C(=C(C=C3)O)C=CC(O4)(C)C)C**	**74,05**	**MNPDB**	**Promiscous in PSA and CDRA**	**Accepted**
Acrovestone	24177-16-0	CC(C)CC(C1=C(C(=C(C(=C1O)CC=C(C)C)O)C(=O)C)O)C2=C(C(=C(C=C2O)CC=C(C)C)OC)C(=O)C)O	74,02	INDB	Non-promiscous in PSA and CDRA	Accepted
18, Chicoric acid	**6537-80-0**	**C1=CC(=C(C=C1/C=C/C(=O)O[C@@H](C(=O)O)[C@@H](OC(=O)/C=C/C2=CC(=C(C=C2)O)O)C(=O)O)O)O**	**72,68**	**INDB**	**Non-promiscous in PSA and CDRA**	**Rejected (1 PAINS, covalent binder)**
23, Morusin	**62596-29-6**	**CC(=CCC1=C(OC2=C(C1=O)C(=CC3=C2C=CC(O3)(C)C)O)C4=C(C=C(C=C4)O)O)C**	**72,62**	**MNPDB**	**Promiscous in PSA and CDRA**	**Accepted**
21, Kurarinone	**34981-26-5**	**CC(=CC[C@H](CC1=C2C(=C(C=C1O)OC)C(=O)C[C@H](O2)C3=CC=C(C=C3)O)O)C(=C)C)C**	**73,32**	**INDB**	**Non-promiscous in PSA and CDRA**	**Intermediate (halo alkene)**

16, Kuwanon L	88524-65-6	CC1=C[C@@H]([C@H]([C@@H](C1)C2=C(C=C(C=C2)O)O)C(=O)C3=C(C=C(C=C3)O)O)C4=C(C=CC(=C4O)[C@@H]5CC(=O)C6=C(C=C(C=C6O5)O)O)O	72,28	INDB	Non-promiscous in PSA and CDRA	Accepted
13, Sanggenon O	101664-32-8	CC1=C[C@@H]([C@@H]([C@H](C1)C2=C(C=C(C=C2)O)O)C(=O)C3=C(C=C(C=C3)O)O)C4=C(C5=C(C=C4O)O[C@]6(C7=C(C=C(C=C7)O)O[C@]6(C5=O)O)CC=C(C)C)O	69,51	INDB	Non-promiscous in CDRA, promiscous in PSA	Accepted
19, Cynarine	30964-13-7	C1[C@H]([C@@H]([C@@H]([C@@H](C[C@]1(C(=O)O)OC(=O)/C=C/C2=CC(=C(C=C2)O)O)OC(=O)/C=C/C3=CC(=C(C=C3)O)O)O)O)O	67,73	INDB	Non-promiscous in PSA, promiscous in CDRA	Rejected (1 PAINS, covalent binder)
17, Katsumadain A	252239-83-1	C1[C@H](OC2=C([C@@H]1C3=CC=CC=C3)C(=O)OC(=C2)/C=C/C4=CC=CC=C4)CC(=O)CCC5=CC=CC=C5	63,64	INDB	Non-promiscous in PSA and CDRA	Rejected (aliphatic ketone)
20, Neobavaiso-flavone	41060-15-5	CC(=CCC1=C(C=CC(=C1)C2=COC3=C(C2=O)C=CC(=C3)O)O)C	58,68	INDB	Promiscous in PSA and CDRA	Accepted

INDB, in-house database; MNPDB, Molport Natural Product Database; PSA, primary screening assay; CDRA, confirmatory dose-response asssay

Table S3. Predicted virtual hits from PL^{pro} structure 7JN2 ranked by ChemPLP score. Tested hits are presented in bold.

compound name	CAS number	SMILES code	ChemPLP score	database origin	HitDexter 2.0	FAF Drugs 4
MolPort-000-851-002		CCOC(=O)CCC1=C(C2=C(C(=C(C=C2)O)C[NH2+]CC3=CC(=C(C=C3)OC)OC)OC1=O)C.[Cl-]	118,84	MNPDB	Non-promiscous in PSA and CDRA	Intermediate (1 PAINS)
48, MolPort-019-909-691	**1324072-44-7**	**CC(C)CCN1C=C(C2=CC=CC=C2C1=O)C(=O)NCC3=CC(=C(C=C3)OC)OC**	**102,79**	**MNPDB**	**Non-promiscous in PSA and CDRA**	**Accepted**
39, MolPort-000-428-993		**COCCN1C(C(C(=O)C2=CC=C(OC)C(C)=C2)=C(O)C1=O)C1=CC=C(O)C=C1**	**97,68**	**MNPDB**	**Non-promiscous in PSA and CDRA**	**Rejected (covalent binder)**
Molport-001-013-844		C1=CC=C(C=C1)OC2=CC=C(C=C2)C(=O)NCC3=CC=CC4=CC=CC=C43	97,68	MNPDB	Non-promiscous in PSA and CDRA	Accepted
44, MolPort-007-627-328		**COC1=CC=C(C=C1)S(=O)(=O)NCC2=CC=C(C=C2)C(=O)NCC3=CC=CC=C3**	**97,00**	**MNPDB**	**Non-promiscous in PSA and CDRA**	**Accepted**
Molport-002-826-944		CC(C1=CC(=CC=C1)C(=O)C2=CC=CC=C2)C(=O)OC3=CC=CC4=CC=CC=C43	96,46	MNPDB	Non-promiscous in PSA and CDRA	Accepted
47, MolPort-039-338-319	**53505-68-3**	**COC1=C(C=CC(=C1)CCCO)OC(CC2=CC(=C(C=C2)O)OC)CO**	**96,28**	**MNPDB**	**Non-promiscous in PSA and CDRA**	**Accepted**
MolPort-000-665-142		CCCCCCC1=CC(=C(C=C1O)O)C(=O)COC2=CC(=CC(=C2)C)C	94,98	MNPDB	Non-promiscous in PSA and CDRA	Rejected (conecutive alkyl chains)
42, MolPort-005-910-551	**951970-62-0**	**CC(CC1=CNC2=CC=CC=C21)NC(=O)C3=CC=C(C=C3)OC4=COC5=C(C4=O)C=CC(=C5)OC**	**93,61**	**MNPDB**	**Non-promiscous in PSA and CDRA**	**Accepted**
40, MolPort-001-540-175		**C1=CC=C(C=C1)CCCNC(=O)CC(C2=CC=CC=C2)C3=CC=CC=C3**	**93,49**	**MNPDB**	**Non-promiscous in PSA and CDRA**	**Accepted**
45, Molport-002-535-048	**865281-26-1**	**CC1=C(C(=CC=C1)C(C)C)NC(=O)CC2=C(C3=C(C(=C(C=C3)OC)OC)OC2=O)C**	**91,58**	**MNPDB**	**Non-promiscous in PSA and CDRA**	**Intermediate (Coumarines)**
31, Capsaicin	**404-86-4**	**CC(C)/C=C/CCCCC(=O)NCC1=CC(=C(C=C1)O)OC**	**86,70**	**IHDB**	**Non-promiscous in PSA, promiscous in CDRA**	**Accepted**
30, Acetylshikonin	**24502-78-1**	**CC(=CC[C@H](C1=CC(=O)C2=C(C=CC(=C2C1=O)O)O)OC(=O)C)C**	**78,92**	**IHDB**	**Non-promiscous in PSA and CDRA**	**Rejected (3 PAINS, covalent binder**
34, Silibinin		**COC1=C(O)C=CC(=C1)C1OC2=C(OC1CO)C=CC(=C2)[C@H]1OC2=CC(O)=CC(O)=C2C(=O)[C@@H]1O**	**78,50**	**IHDB**	**Promiscous in PSA and CDRA**	**Accepted**
33, Silicristin	**33889-69-9**	**COC1=C(C=CC(=C1)[C@H]2[C@@H](C3=C(O2)C(=CC(=C3)[C@@H]4[C@H](C(=O)C5=C(C=C(C=C5O4)O)O)O)O)CO)O**	**77,85**	**IHDB**	**Promiscous in PSA and CDRA**	**Accepted**
38, 001-633-809		**CN1C=C(C=N1)CNC(=O)C2=C(C=CC(=C2)[N+](=O)[O-])Cl**	**77,15**	**IHDB**	**Non-promiscous in PSA and CDRA**	**Intermediate (nitro)**
32, Lobeline	**90-69-7**	**CN1[C@@H](CCC[C@@H]1CC(=O)C2=CC=CC=C2)C[C@@H](C3=CC=CC=C3)O**	**72,77**	**IHDB**	**Non-promiscous in PSA and CDRA**	**Accepted**
MolPort-003-292-157		COC1=CC2=C(C=C1)C(=CC(=O)O2)COC(=O)C3=CC=CC=C3NCCO	93,62	MNPDB	Non-promiscous in PSA and CDRA	Intermediate (coumarines)
MolPort-009-375-739	1110928-66-9	CC1=CC(=C(C=C1)OC)C(=O)OC(C)C(=O)C2=C(NC3=CC=CC=C32)C	91,20	MNPDB	Non-promiscous in PSA and CDRA	Accepted
Isovalerylshikonin	52387-14-1	CC(C)CC(=O)O[C@H](CC=C(C)C)C1=CC(=O)C2=C(C=CC(=C2C1=O)O)O	85,57	IHDB	Non-promiscous in PSA and CDRA	Rejected (3 PAINS, covalent binder)
Arctiin	20362-31-6	COC1=C(C=C(C=C1)C[C@H]2COC(=O)[C@@H]2CC3=CC(=C(C=C3)O[C@H]4[C@@H]([C@H]([C@@H]([C@H](O4)CO)O)O)O)OC)OC	84,03	IHDB	Non-promiscous in PSA and CDRA	Accepted
β-β-Dimethylacryl-shikonin	24502-79-2	CC(=CC[C@H](C1=CC(=O)C2=C(C=CC(=C2C1=O)O)O)OC(=O)C=C(C)C)C	83,47	IHDB	Non-promiscous in PSA and CDRA	Rejected (3 PAINS, covalent binder)

5-Geranyloxy-7-methoxycoumarin	7380-39-4	CC(=CCC/C(=C/COC1=CC(=CC2=C1C=CC(=O)O2)OC)/C)C	81,90	IHDB	Non-promiscous in PSA and CDRA	Intermediate (coumarines)
Bergamottin	7380-40-7	CC(=CCC/C(=C/COC1=C2C=CC(=O)OC2=CC3=C1C=CO3)/C)C	75,33	IHDB	Non-promiscous in PSA and CDRA	Intermediate (coumarines, furanocoumarines)
Ostruthol	642-08-0	C/C=C(/C)\C(=O)O[C@H](COC1=C2C=CC(=O)OC2=CC3=C1C=CO3)C(C)(C)O	73,90	IHDB	Non-promiscous in PSA and CDRA	Rejected (covalent binder)

INDB, in-house database; MNPDB, Molport Natural Product Database; PSA, primary screening assay; CDRA, confirmatory dose-response assay

Table S4. Predicted virtual hits from PL^{pro} structure 4OW0 ranked by ChemPLP score. Tested hits are presented in bold.

compound name	CAS number	SMILES code	ChemPLP score	database origin	HitDexter 2.0	FAF Drugs 4
MolPort-047-538-149		C1CC2=C(C3=CC=CC=C3N=C2C1)NC(=O)C[NH2+]CCC4=CC=C(C=C4)O	95,80	MNPDB	Non-promiscous in PSA and CDRA	Intermediate (phenol)
MolPort-046-848-556	1030021-10-3	Cl.COC(=O)Cc1c(C)c2cc(CNCCc3ccccc3)c(O)c(C)c2oc1=O	95,48	MNPDB	Non-promiscous in PSA and CDRA	Intermediate (1 PAINS, coumarines)
Molport-010-696-387		CCN1C2=CC=CC=C2N(C(=O)C1=O)CC3=CC=C(C=C3)C(=O)NCC4=CC=C(C=C4)OC	94,46	MNPDB	Non-promiscous in PSA and CDRA	Intermediate (1 PAINS)
46, MolPort-007-806-805		CC1=NC2=CC=CC=C2N(C1=O)CC3=CC=C(C=C3)C(=O)NCC4=CC=C(C=C4)OC	**91,27**	**MNPDB**	**Non-promiscous in PSA and CDRA**	**Accepted**
41, MolPort-002-532-107	**864760-73-6**	**COC1=C(C=C\2C(=C1)CCN/C2=C\C(=O)C3=CC=C(C=C3)CSC)OC**	**88,93**	**MNPDB**	**Non-promiscous in PSA and CDRA**	**Rejected (frequent hitter dopamine, covalent binder)**
43, MolPort-002-520-481	**858762-81-9**	**COC1=CC=C(C=C1)C(=O)COC2=CC3=C(C=C2)C(=O)/C(=C/C4=C(C=CC(=C4)OC)OC)/O3**	**88,69**	**MNPDB**	**Promiscous in PSA and CDRA**	**Rejected (covalent binder)**
Byakangelicin	482-25-7	CC(C)([C@@H](COC1=C2C(=C(C3=C1OC(=O)C=C3)OC)C=CO2)O)O	84,10	IHDB	Non-promiscous in PSA and CDRA	Intermediate (coumarines, furanocoumarines)
37, Oxypeucedanin hydrate	**2643-85-8**	**CC(C)([C@@H](COC1=C2C=CC(=O)OC2=CC3=C1C=CO3)O)O**	**81,78**	**IHDB**	**Non-promiscous in PSA and CDRA**	**Intermediate (coumarines, furanocoumarines)**
Molport-000-699-830	5667-50-5	C1CN(CCN1C2=C(C=C(C=C2)[N+](=O)[O-])Cl)C(=O)C3=CC(=C(C=C3Cl)F)F	78,86	IHDB	Non-promiscous in PSA and CDRA	Intermediate (1 PAINS, nitro)
36, Magnolol	**528-43-8**	**C=CCC1=CC(=C(C=C1)O)C2=C(C=CC(=C2)CC=C)O**	**78,03**	**IHDB**	**Promiscous in PSA and CDRA**	**Intermediate (Halo alkene)**
35, Honokiol	**35354-74-6**	**C=CCC1=CC(=C(C=C1)O)C2=CC(=C(C=C2)O)CC=C**	**70,49**	**IHDB**	**Promiscous in PSA and CDRA**	**Intermediate (Halo alkene)**

INDB, in-house database; MNPDB, Molport Natural Product Database; PSA, primary screening assay; CDRA, confirmatory dose-response assay

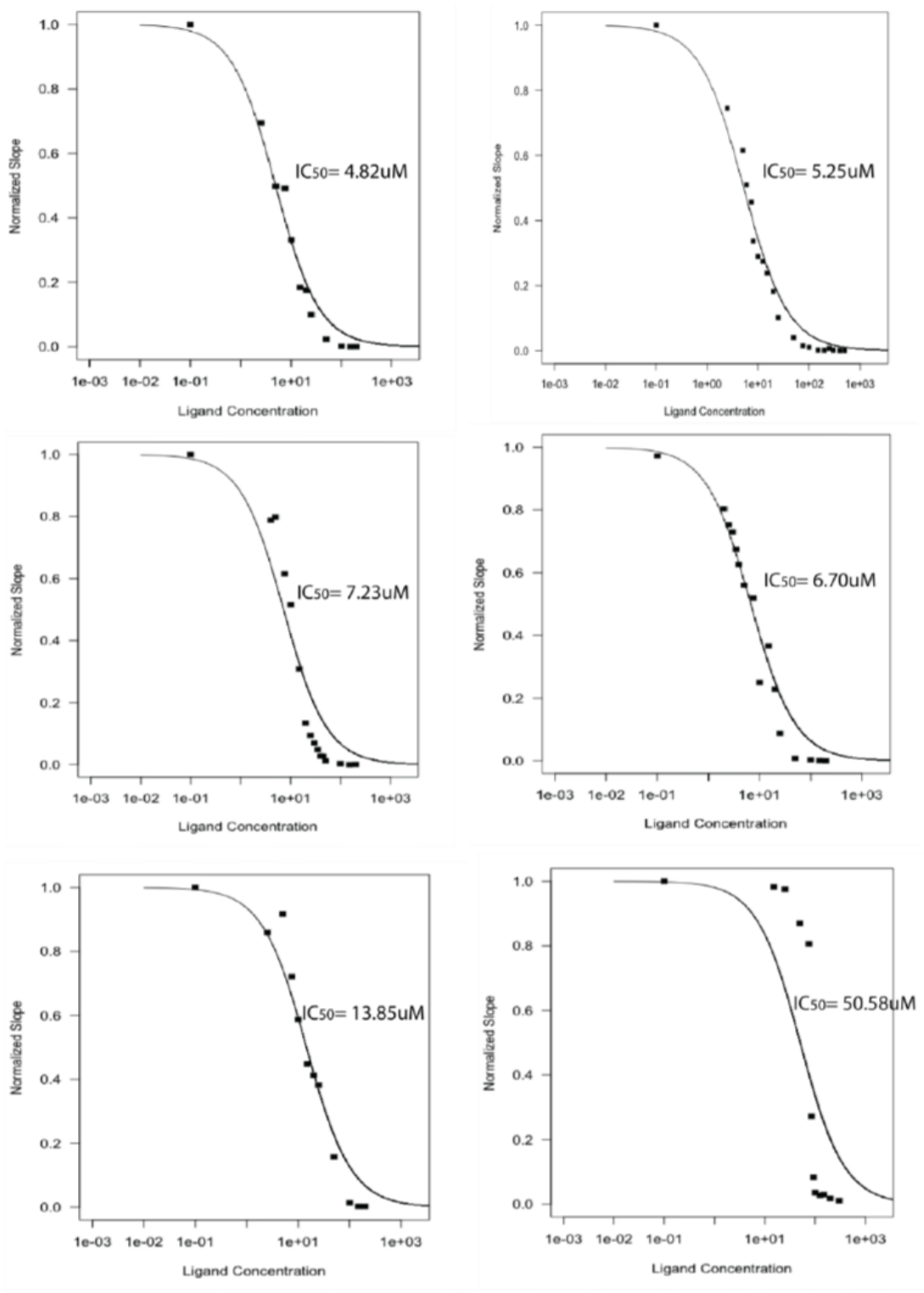


Figure S2. IC_{50} curves of the different VHs tested against M^{pro}. Dose-response curves of compounds (A) **15**, (B) **16**, (C) **12**, (D) **13**, (E) **14**, and (F) **4**. The data and IC_{50} were fitted and calculated, respectively, based on the hill equation using Rstudio.

Table S5. Inhibitory activities of virtual hits for PL^{pro}

Method

Virtual hits were tested for PL^{Pro} inhibition using a commercial assay kit from Anaspec (Fremont, USA; catalogue ID AS-72263) according to the manufacturer's protocols. Final test concentrations of virtual hits were 20 µM with 0.2% DMSO in assay buffer. Positive control 7 was tested at a final concentration of 10 µM and 0.2% DMSO in assay buffer. Compounds and vehicle control were incubated for 15 minutes with the PL^{pro} enzyme in two well replicates before the deubiqitination substrate was added. Final concentrations of protease and substrate were 100 ng/ml and 2 µM. Two background fluorescence control wells without protease were included for each sample. After 60 minutes of incubation at room temperature, the enhanced fluorescence due to the protease-catalyzed substrate cleavage was measured at Exc/Emi 490/520 nm using a Tecan Sparks plate reader (Männedorf, Switzerland). Means of three independent experiments are presented. Statistical significance was calculated by one-way ANOVA and Dunnett's multiple comparison test.

Table S5. Inhibitory activities of virtual hits for SARS-CoV-2 PL^{pro} measured using the commercial Anaspec assay kit including the anti SARS-CoV-2 activity and cytotoxicity of putatively active hits in two different cell lines. Significant inhibition of virtual hits was calculated with one-way ANOVA and Dunnett's multiple comparisons test (****, $p < 0.0001$; **, $p < 0.01$; *, $p < 0.05$).

Compound	SARS-CoV-2 PL^{pro} inhibition % enzyme inhibition (20 µM)	Anti SARS-CoV-2 activity		Cytotoxicity	
		Caco-2	Calu-3	Caco-2	Calu-3
		IC_{50} (µM)	IC_{50} (µM)	CC_{50} (µM)	CC_{50} (µM)
1, remdesivir			0.00125		>0.1
7, GRL-0617	96.6 ± 10.3****				
30, acetylshikonin	100.3 ± 9.5****	>100	>100	65.8	>100
31, trans-Capsaicin	32.4 ± 13.7				
32, lobelin	26.6 ± 11.0				
34, silibinin	50.3 ± 14.3*		>100		>100
35, honokiol	80.6 ± 9.9****	16.2	28.4	91.2	>20
36, magnolol	58.2 ± 11.5**	17.7	26.8	47.3	>20
37, oxypeucedanin hydrate	97.5 ± 14.6****		>100		>100
38, MolPort-001-633-809	26.1 ± 15.2				
39, Molport-000-428-993	26.0 ± 34.7				
40, Molport-001-540-175	41.5 ± 16.5				
42, Molport-005-910-551	82.3 ± 25.3****	7.1	>100	>100	>100
43, Molport-002-520-481	13.5 ± 25.5				
44, Molport-007-627-328	18.5 ± 16.6				
45, Molport-002-535-048	-2.6 ± 36.1				
47, Molport-039-338-319	-55.8 ± 32.4				
48, Molport-019-909-691	32.0 ± 4.2				

n. t. – not tested

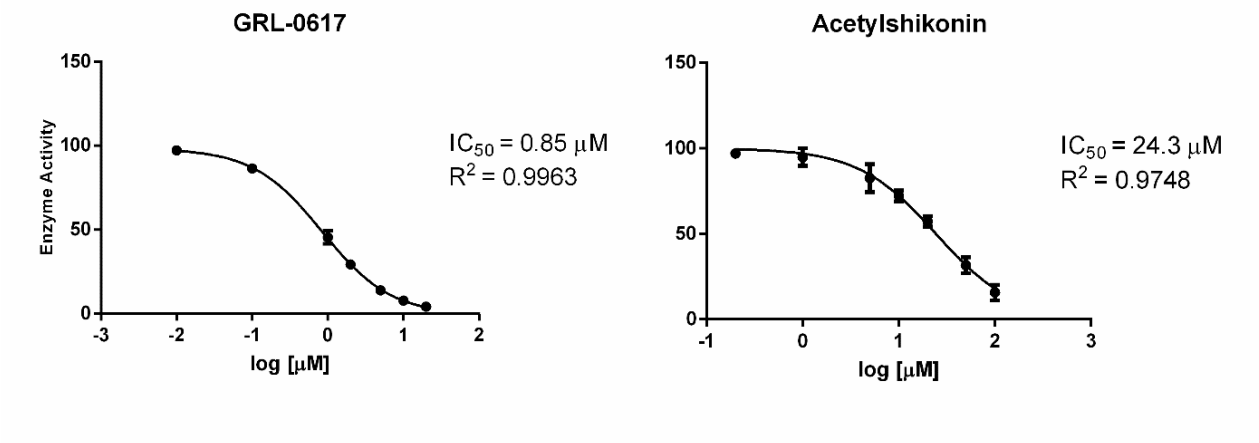


Figure S3. Concentration-response curves of positive control GRL-0617 (**7**) and acetylshikonin,(**30**) against SARS-CoV-2 PL^{pro} measured using the peptide Z-RLRGG-AMC as substrate. Fluorescence signal at 360 nm (excitation)/460 nm (emission) was immediately measured by continuous 24 points for 72 min. IC_{50}s of the compounds were determined by plotting enzyme inhibition of at least three independent experiments against seven concentrations of the test inhibitor by using the dose-response curve in GraphPad Prism.

11

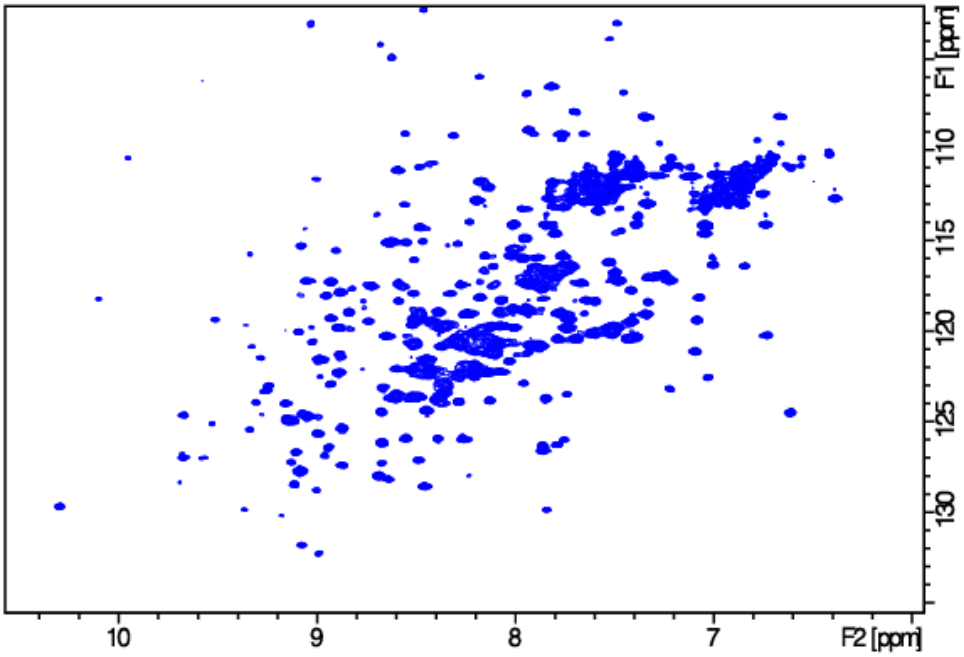


Figure S4: 2D-[^{1}H, ^{15}N]-TROSY-HSQC of 450 **µ**M of M^{pro} of SARS-CoV-2 at pH 7.6 in 50 mM phosphate buffer, 50 mM NaCl. The HSQC was acquired on a 700MHz NMR Spectrometer Bruker Advance III at 30°C.

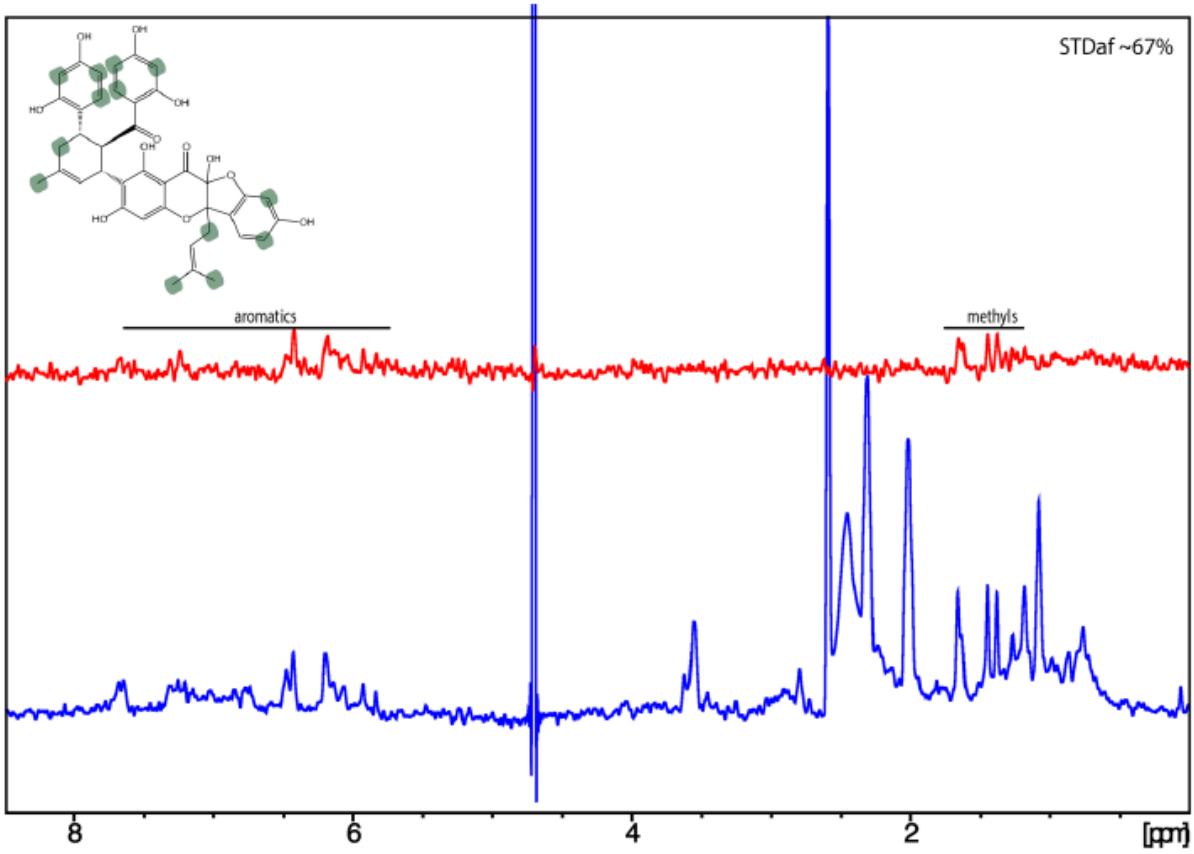


Figure S5. STD NMR of M^{pro} with compound **14.** The measurements were acquired on a 500MHz NMR Spectrometer Bruker Advance III with TCI-cryoprobe at 25°C.

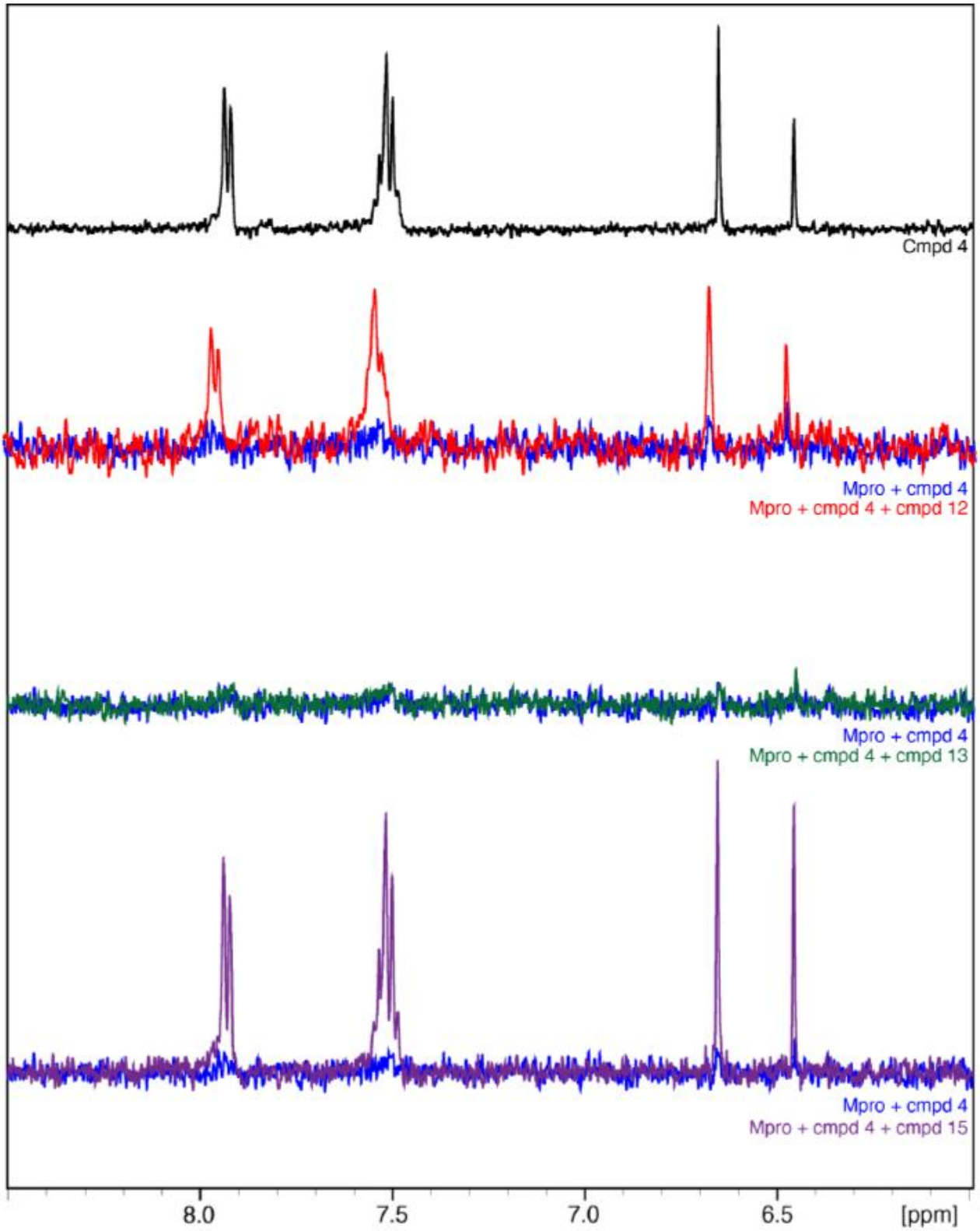


Figure S6. CPMG experiments measuread for different mixtures: compound **4**, compound **4** and M^{pro}, compound **4** with compound **12** and M^{pro}, compound **4** with compound **13** and M^{pro}, compound **4** with compound **15** and M^{pro}, depicted in black, blue, red, green and purple respectively. The sample compositions are reported in the experimental section of the main text.

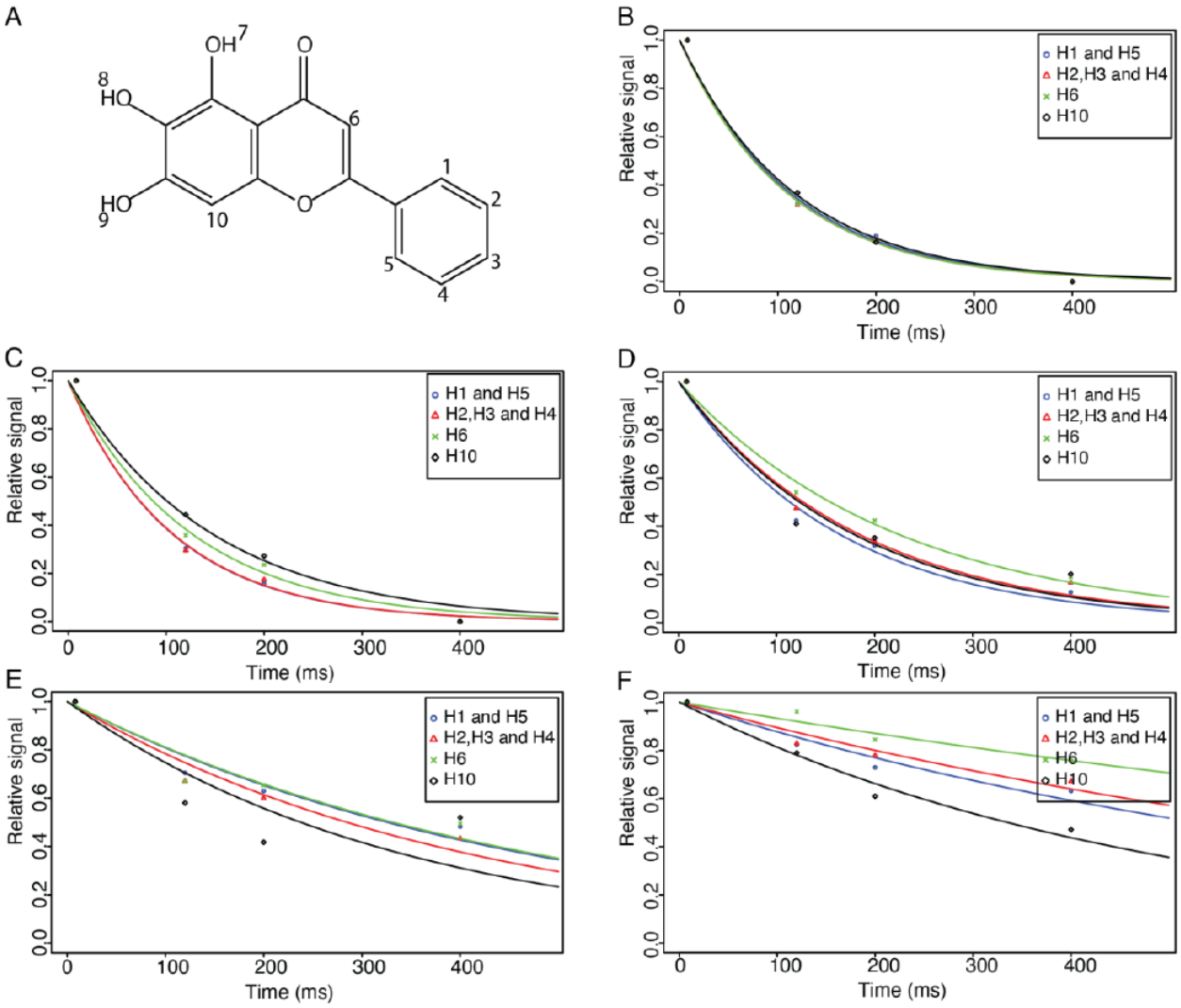


Figure S7. CPMG decay curves for each of the reporter compound **4** proton, depicted in (A), measured with the samples composed of: (B) compound **4** and M^{pro}, (C) compound **4** with compound **13** and M^{pro}, (D) compound **4** with compound **12** and M^{pro}, (E) compound **4** with compound **15** and M^{pro}, (F) compound **4**. The sample compositions are reported in the experimental section of the main text.

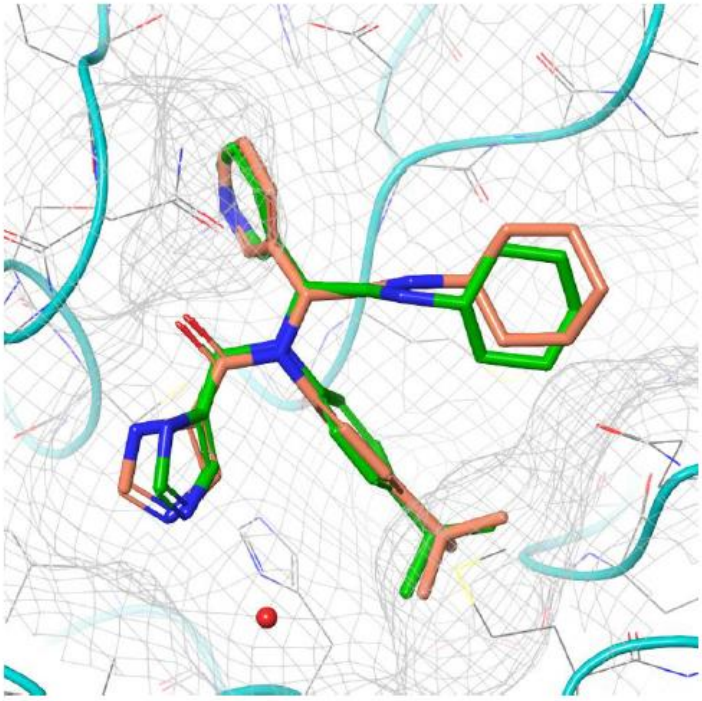


Figure S8. Comparison of the binding pose of co-crystallized ligand of 6W63 (orange) and the re-docked binding pose (green)

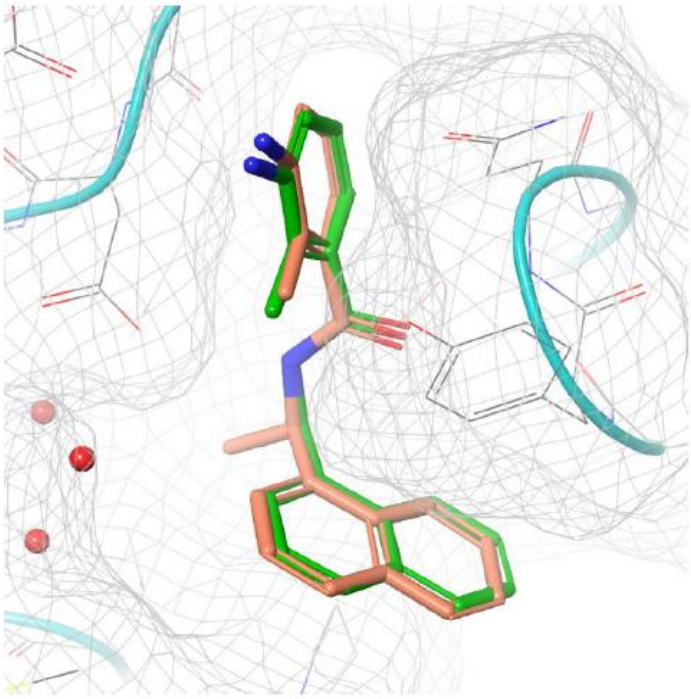


Figure S9. Comparison of the binding pose of co-crystallised ligand of 7JN2 (orange) and the re-docked binding pose (green)

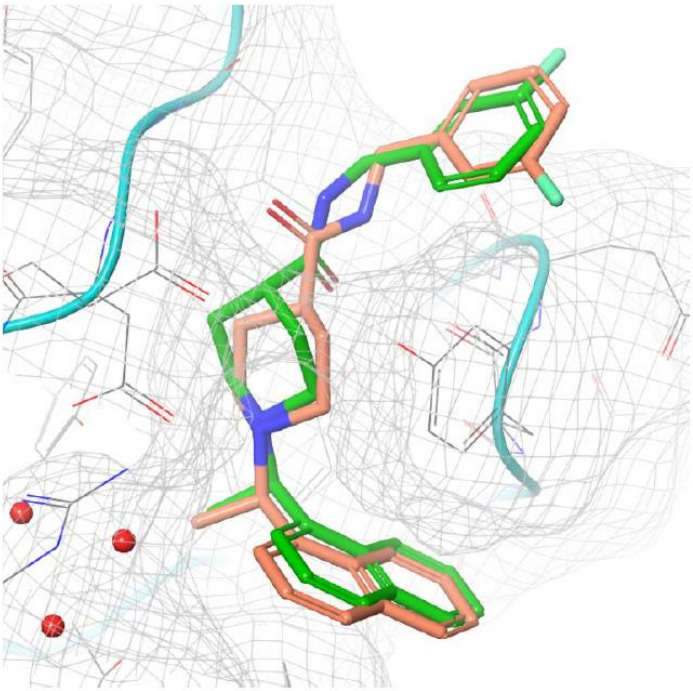


Figure S10. Comparison of the binding pose of co-crystallised ligand of 4OW0 (orange) and the re-docked binding pose (green)

Table S5. RMSD values and ChemPLP score of re-docked ligands

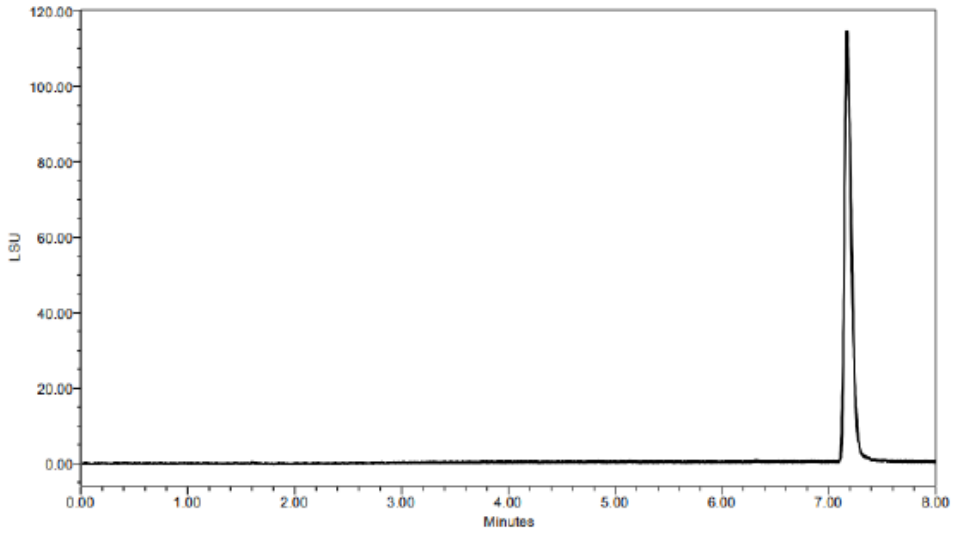
Protein structure (PDB)	RMSD (Å)	ChemPLP score
6W63	0,86	76,73
7JN2	0,31	91,12
4OW0	0,74	89,10

UPLC analyses of compounds **12-16**

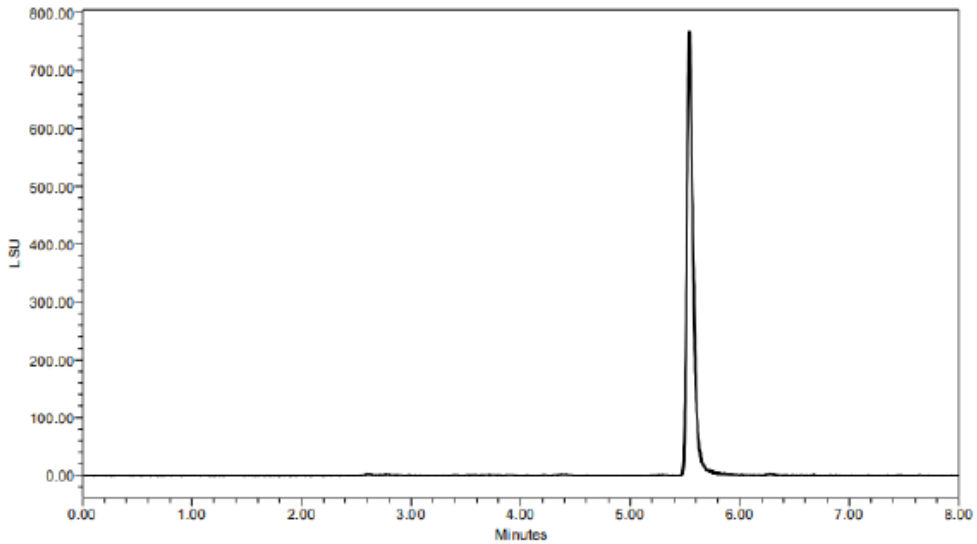
Method

Ultra performance liquid chromatography/evaporative light scattering detector (UPLC-ELSD) was performed on Waters Acquity UPLC H-Class system equipped with a column manager, a sample manager, a quartenary solvent manager, an isocratic solvent manager, a fraction collector, a photodiode array (PDA) detector, a mass detector (QDa) and an ELSD. For analysis a Waters Acquity UPLC BEH C18 column (1.7 µm, 2.1 x 100 mm) was used. The mobile phase consits of water (A) and acetonitril (B) using gradient elution starting with 95% of B. Gradient in detail: 5% B at 0 min, from 0% to 98% B in 5 min, isocratic 98% B for 3 min. The flowrate was set at 0.3 ml/min. The software Waters Empower 3 was used for data acquistion.

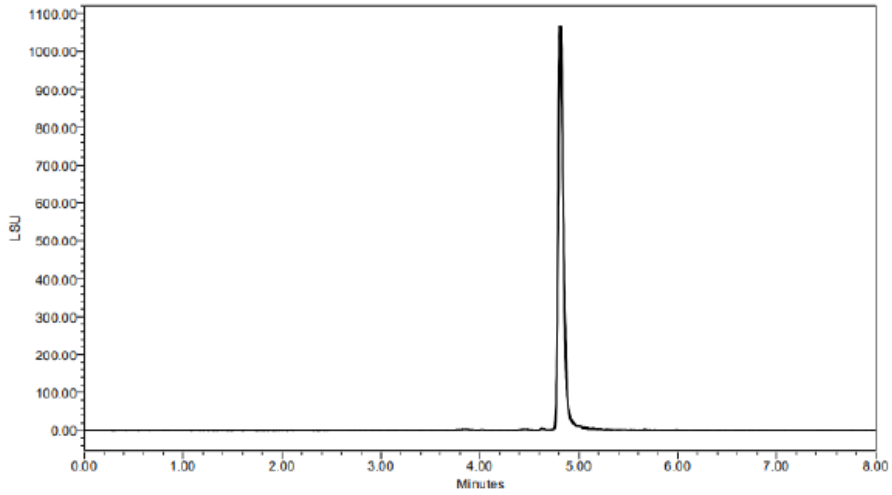
UPLC-ELSD chromatogram of compound **12**



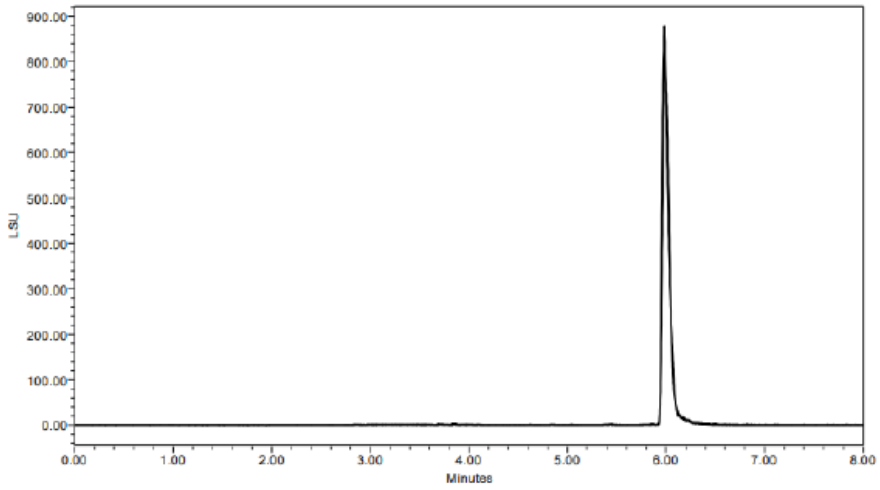
UPLC-ELSD chromatogram of compound **13**



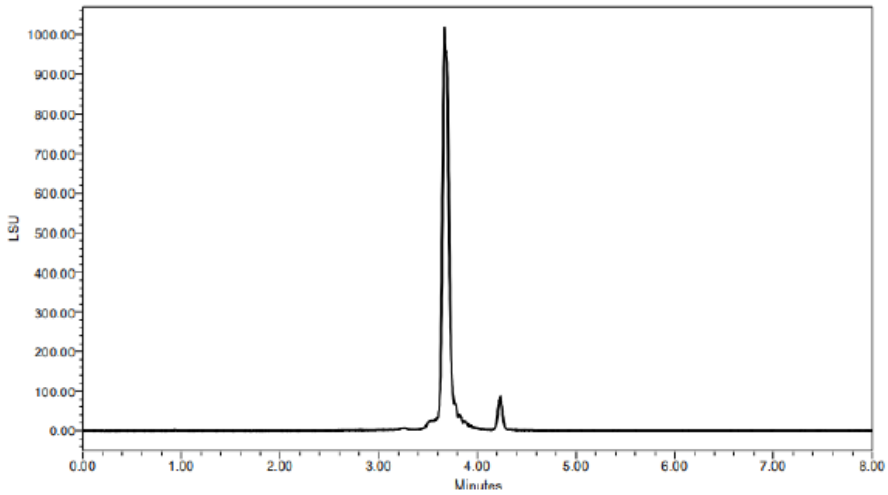
UPLC-ELSD chromatogram of compound **14**



UPLC-ELSD chromatogram of compound **15**



UPLC-ELSD chromatogram of compound **16**



18

6.2.2 Review of Publication I

A crucial step in TBDD is the identification and validation of a suitable target.[75,94] Antiviral compounds may act on proteins originating from either the host or the virus, but these targets must play an essential role in the viral replication cycle.[99] In **Publication I**, the viral proteases M^{pro} (non-structural protein 5 (nsp5), also referred to as 3C-like protease or $3CL^{pro}$)[100,101] and PL^{pro} (domain of nsp3)[102] were selected as targets for TBVS to search for NPs with inhibitory effects against SARS-CoV-2.

This decision was based not only on the availability of X-ray crystal structures in complex with non-covalently bound synthetic inhibitors, but also on the fact that both M^{pro} and PL^{pro} exhibit ligandability for small molecules and, due to their conserved structures, show a low propensity for mutations.[99,103] Furthermore, the functional roles of these cysteine proteases, which have distinct substrate specificities, were well known and characterized.[99] The genome of SARS-CoV-2 encodes four structural proteins and two polyproteins, which are translated from open reading frames (ORFs).[99] After cell-uptake of the virus and polypeptide translation, M^{pro} is able to distinctly cleave the polyproteins at eleven positions, PL^{pro} is responsible for cleavage at three positions.[99] Subsequently, this leads, among other outcomes, to the release of the RNA-dependent RNA polymerase (RdRp) subunit nsp12, a key enzyme involved in the viral replication/transcription complex.[99,102,104]

However, in October 2020, when the study was initiated, only a limited number of non-covalent inhibitors were available for the selected targets. Consequently, molecular docking appeared more suitable for VS than pharmacophore modeling. Although M^{pro} and PL^{pro} inhibitors were already known at the time[105,106], comprehensive repurposing screenings, such as those involving 3,727 clinically approved drugs targeting PL^{pro}, failed to yield any consistently validated protease inhibitors.[105] Prada Gori et al. identified two VHs in a ligand-based drug repurposing campaign targeting M^{pro}, but neither was active in cell-based assays.[107]

It was only in 2022, that M^{pro} inhibiting drugs discovered through TBDD received market approval: Paxlovid™ (containing nirmatrelvir in combination with ritonavir, which slowes down the metabolism of nirmatrelvir) and Xocova™ (containing the active substance ensitrelvir).[108-114] Further M^{pro} inhibitors in the pipeline included simnotrelvir (in combination with ritonavir, Xiannuoxin™) which has already been approved in China[115,116], olgotrelvir, which showed promising results in a phase 3 study[117], and pomotrelvir (PBI-0451) which, however, failed in a Phase 2 study.[118]

For molecular docking in **Publication I**, one M^{pro} and two PL^{pro} protein structures were employed to screen > 140,000 natural NPs and NP-like compounds. Of the 61 VHs identified, 38 were selected for subsequent *in vitro* validation. In enzyme-based assays, five active inhibitors of M^{pro} (sanggenon G (SGG): IC_{50} = 4.8 µM; kuwanon L: IC_{50} = 5.3 µM; sanggenon O (SGO): IC_{50} = 6.7 µM; sanggenon C (SGC): IC_{50} = 7.2 µM; sanggenon D (SGD): IC_{50} = 13.9 µM), two moderately

active M^{pro} inhibitors (morusin: IC_{50} = 33.5 µM; kuwanon A: IC_{50} = 35.2 µM), and one moderately active PL^{pro} inhibitor (acetylshikonin: IC_{50} = 4.8 µM) could be confirmed.[119]

These eight candidates were subsequently evaluated using a spike immunostaining assay in Calu-3 cells to corroborate the enzymatic inhibition results. Three NPs exhibited anti-SARS-CoV-2 activity in the low micromolar range (SGC: IC_{50} = 4.6 µM; SGG: IC_{50} = 7.6 µM; SGO: IC_{50} = 8.0 µM), with only moderate cytotoxicity observed at high concentrations (SGC: CC_{50} = 23.6 µM; SGG: CC_{50} = 49.1 µM; SGO: CC_{50} = 24.9 µM). The selectivity index (SI; CC_{50}/IC_{50}) was calculated as 5.1 for SGC, 6.5 for SGG, and 3.1 for SGO. An extract, enriched in mulberry Diels–Alder-type adducts (MDAAs), with a total MDAA content of 29.0%, also inhibited SARS-CoV-2 replication (IC_{50} = 23.0 µg/mL). Finally, NMR experiments were conducted to confirm the predicted 'butterfly-shaped' binding mode of the ligands to the target, as well as their competitive inhibition of M^{pro}.[119]

MDAAs such as SGC, SGD, and SGG, which have been isolated from the root bark of *Morus alba* L., are prenylated flavonoids formed via an intermolecular [4 + 2]-cycloaddition between a chalcone unit and a dehydroprenylphenol moiety.[120,121] In addition to their anti-SARS-CoV-2 activity confirmed in **Publication I**, MDAA-enriched extracts and pure MDAAs such as SGC also exhibit antiviral activity against Influenza A and were well tolerated by lung cells (A549, Calu-3).[119,122] Moreover, they showed pronounced anti-inflammatory activity *in vitro*[122] and have been reported to disrupt the lethal synergism between influenza viruses and *Streptococcus pneumoniae*.[123] The dual activity of these NPs against viral and bacterial pathogens makes them particularly promising candidates for the treatment of diseases involving complications from bacterial secondary infections following viral infections.[123,124] However, studies in mice revealed poor bioavailability following oral administration.[125] To investigate whether this limitation could be overcome, selected anti-infective sanggenons and MDAA-containing extracts were subjected to biopharmaceutical profiling (BPP), with a focus on inhalation administration as an alternative route of delivery (**Publication II**, Chapter 8).

During this study, an important consideration was that both M^{pro} and PL^{pro} possess a highly reactive cysteine residue in their catalytic center, which, due to its thiol group, can form nonspecific covalent bonds.[126] Therefore, the reducing agent dithiothreitol (DTT) was added during the assays on PL^{pro} and M^{pro} to prevent nonspecific binding.[127] Considering such effects is crucial, particularly for NPs featuring pan-assay interference compound (PAINS) motifs.[127,128]

In summary, **Publication I** demonstrates the effectiveness of integrating the *in silico* TBVS approach with experimental *in vitro* validation, including both enzyme-based and cell-based assays, for the discovery of NPs active against SARS-CoV-2. Based on 60% enzyme inhibition at 20 µM as efficacy threshold, a hit rate of 13% was achieved.[119] The most promising compound was SGG (IC_{50} = 4.8 µM, SI 6.5), which exhibited the strongest M^{pro} inhibitory activity among all tested compounds.[119]

In a review by Guerra et al., the authors reported that only 45 out of 648 NP-based potential inhibitors of the SARS-CoV-2 proteases M^{pro} and PL^{pro} had been experimentally tested and validated according to the literature.[126] Notably, all 45 tested NPs shared a flavone scaffold.[126] Together with the findings from **Publication I**, this points to the potential of experimentally validating entire compound families, a process that can realistically only be carried out following prior CADD approaches such as TBDD, which aim to reduce the number of candidates and ensure that only promising NPs proceed to *in vitro* testing.[70]

7 Biochemometric methods in the search for bioactive NPs

Recent years have seen not only a substantial rise in the significance of AI-based approaches, but also noteworthy progress in the development of diverse methods for the targeted identification of bioactive compounds in multicomponent mixtures through combination strategies of heterogeneous datasets.[8,28,129,130]

Biochemometrics is a method that combines chemical and bioactivity data.[131-133] It is a discipline that evolved from chemometrics, which applies mathematical methods to chemical data analysis.[132] The term was independently introduced in 2006 by both Kvalheim and Martens, with the name intended to reflect its conceptual relationship to bioinformatics and biostatistics.[133] In 2011, Kvalheim and his group used this approach to create a cross-validated partial least squares (PLS) regression model for statistically analyzing the relationship between chromatographic profiles of synthetic compound mixtures and their corresponding bioactivity data.[134]

According to a 2016 study by Kellogg et al., biochemometrics integrates biological and chemical datasets to construct statistical models that enable the early identification of active compounds during the fractionation process.[131] In essence, biochemometrics correlates chemical or metabolite profiles of multicomponent mixtures, such as extracts and fractions, with bioactivity data using statistical methods.[130,135] The bioactivity data are not restricted to specific pharmacological targets and may be obtained from cell-free, *in vitro*, or *in vivo* assays.[130] NMR and LC-MS data sets are frequently used in biochemometric analyses.[130]

Overall, the key objective of biochemometrics in NP research is the targeted, precise, and rapid identification of bioactive NPs, ideally at an early stage of sample processing.[131] In this context, the approach, like traditional methods and computer-assisted techniques, operates within the broader framework of bioactive compound discovery in NP drug development.

As of 8 May 2025, a SciFinder (https://scifinder-n.cas.org/) search for the term 'biochemometric' returned only 63 articles and five reviews, covering topics related to medicinal and botanical chemistry. This underscores that the approach remains in its infancy with regard to NP drug discovery, even though biochemometrics has already been applied with considerable success, as the following studies will demonstrate.

Beginning in the early 2000s, a research group led by Pauli published several biochemometrics-related studies focused on the identification of active compounds in extracts of NP origin. For example, they applied this approach as a complement to bioactivity-guided fractionation, identifying 29 anti-tuberculosis (anti-TB) active compounds from *Oplopanax horridus* (Sm.) Miq., based on a dataset of extract metabolites obtained through high-speed countercurrent chromatography (HSCCC) and gas chromatography (GC) hyphenated to MS detection (GC-MS).[136] In 2013, they

successfully employed biochemometrics to assess the contribution of individual compounds to anti-TB activity in a complex extract of the same Alaskan plant, using quantitative proton NMR (qHNMR) data.[137] Ramos et al., also affiliated with the Pauli group, aimed to identify the optimal composition of airborne anti-TB bioactives from the essential oil of *Eucalyptus citriodora* Hook., using a biochemometrics approach based on HSCCC, GC-MS, and biological screening data.[138]

Another group of researchers, including Kellogg and Cech, has also made significant contributions to the development of biochemometric approaches. For example, in 2016, Kellogg et al. combined bioactivity-guided fractionation and biochemometrics, focusing on the comparison of three statistical methods (PLS loading vectors, selectivity ratio, and S-plots) for evaluating the antimicrobial activity of constituents from *Alternaria* sp. and *Pyrenochaeta* sp.[131] Britton et al. investigated additive, synergistic, and antagonistic effects exemplified by combinations of *Hydrastis canadensis* L. extract fractions and the alkaloid berberine on antimicrobial activity against *Staphylococcus aureus*.[135] To this end, selectivity ratio plots were generated based on untargeted MS-based metabolomics and bioassay data.[135] Another notable example is a study by Caesar et al. from 2018, in which bioactivity-guided fractionation and biochemometrics, based on selectivity ratio analysis, were combined with molecular networking to assess the antimicrobial activity of compounds from the root of *Angelica keiskei* (Miq.) Koidz. against *S. aureus*.[139]

7.1 MS-based biochemometrics

A strategy that has gained prominence in the context of MS-based NP research is **metabolomics**.[140] This approach was developed in the late 1990s to enable the comprehensive, universally applicable, nonselective and unbiased identification and quantification of small molecule metabolites within a biological system (cell, tissue, or organism) in its current metabolic state.[85,141] It provides insights into the phenotype shaped by the underlying genomic, transcriptomic, and proteomic networks.[142] Metabolomic datasets are based on the integrated use of spectroscopic or spectrometric techniques and multivariate statistical approaches.[129,143] The putative identification of metabolites, based on the assignment of spectral data, is referred to as 'annotation'.[85] Since multiple assignments to the same dataset are often possible, interpretation based on these annotations remains a critical challenge in metabolomics studies.[85,129] However, in NP research, metabolomics is used for metabolite profiling of multicomponent mixtures to identify bioactive constituents and support NP prioritization in drug discovery.[85]

Another prominent and now widely adopted subfield of MS-based methods is **molecular networking** (MN).[50,129,132] Introduced by Watrous et al. in 2012, this method was developed to visualize chemical similarities among molecules as spectral networks.[144] MN operates on the

premise that molecules with similar structures tend to exhibit comparable MS/MS (MS^2) fragmentation patterns, provided that the fragmentation is performed under consistent experimental conditions.[145]

To calculate MS^2 spectra similarity, each spectrum must first be represented as a multidimensional vector, with the individual coordinates determined by the relative peak intensities.[146,147] In principle, the cosine similarity cos δ (as referred to as cosine score) between two MS^2 spectra is computed as the dot-product of the two corresponding vectors divided by the product of the lengths of the vectors.[129,147] This score ranges from 0 (indicating no similarity, with vectors 90 degrees to each other) to 1 (identical spectra with vectors poiting in the same direction).[146,147] Peaks from one spectrum are aligned with those from another, taking into account not only the original *m/z* values but also potential shifts due to fragment loss.[129] These calculated similarity scores for the dataset can then be visualized as molecular network, where each consensus MS^2 spectrum corresponding to a given precursor mass is depicted as node.[144,145] Similar nodes (based on a user-defined cosine score threshold) are connected via edges, whose thickness reflects the cosine score, thereby forming so-called spectral families.[144,145]

One of the major advantages of MN is the existence of a knowledge base known as the Global Natural Product Social Molecular Networking (GNPS), established in 2013, which hosts open-access raw, processed, and annotated MS data.[148] Additionally, the GNPS platform integrates numerous tools that enable a wide range of approaches, from data visualization to interpretation.[132] MN not only allows for the dereplication of known compounds and the identification of new ones,[132,145] but also provides the possibility to integrate bioactivity data into the molecular network ('bioactivity-based MN'), therefore combining chemical and bioactivity data.[49] This combined information can be used for the targeted isolation of putatively active molecules from fractionated bioactive extracts.[49,50,132,149,150]

Recent years have seen the development of MN-based metabolomics strategies that leverage the strengths of both MN and metabolomics for NP drug discovery ('MN-based metabolomics').[151] For example, the insight that molecules within the same spectral family in a molecular network likely share structural similarities can be leveraged for metabolomics data analysis.[129] A comprehensive overview of the combination of MN with other techniques in NP drug discovery is provided by Fox Ramos et al.[132]

7.2 NMR-based biochemometrics

Within the framework of NMR-based biochemometrics, **statistical heterocovariance analysis** (HetCA) is a method that facilitates the early detection of even low-abundance bioactive compounds in complex mixtures prior to isolation.[143,152] HetCA was introduced in 2016 by researchers from the National and Kapodistrian University of Athens with the aim of accelerating the identification of bioactive NPs.[143] In their study, among other findings, the authors demonstrated that the compound kaempferol 3-O-(4-O-acetylrhamnosyl)-7-O-rhamnoside, present in extracts of *Dorycnium hirsutum* L., was correlated with tyrosinase inhibitory activity, as revealed by statistical covariance analysis between proton (^{1}H) NMR and activity data.[143] The study demonstrated that HetCA can uncover structural information of compounds in extracts or fractions without the need for prior purification and that variations in compound concentration can correlate with dose-dependent bioactivity.[143]

HetCA is based on a modified statistical heterospectroscopy (SHY) approach for the statistical correlation of ^{1}H NMR spectra with activity data.[143,152] SHY was initially developed to enable the co-analysis of data across datasets from different techniques (e.g., NMR and MS), based on the fact that the same or associated molecules exhibit inherent covariance among signal intensities.[152,153] For HetCA, the *crosscov* and *corr* functions are used to calculate the covariance and Pearson's correlation coefficients of the datasets and to identify specific chemical features that are positively or negatively correlated with bioactivity.[152] The results can be visualized in so-called pseudospectra, which resemble ^{1}H NMR spectra and display covariance and correlation coefficients ranging from -1.0 to 1.0.[143,152] Positively correlated NMR signals appear in dark red and point upwards, while negatively correlated features are shown in dark blue and point downwards.[143]

Additonally, when multiple NMR signals in the spectrum of a complex mixture originate from the same molecule, statistical total correlation spectroscopy (STOCSY) can highlight related signals, as their peak intensities exhibit multi-colinearity.[50,143] This can be used for the structural identification of compounds by comparing their STOCSY results, i.e. pseudo NMR spectra, with ^{1}H NMR spectra of reference substances.[152,154]

HetCA has been successfully applied to identify bioactive secondary metabolites in plant extracts.[122,152,155-157] Boka et al. employed HetCA and STOCSY to statistically correlate datasets obtained from ^{1}H NMR measurements and tyrosinase inhibition assays of *Paeonia mascula* (L.) Mill. ssp. *hellenica* Tzanoud. extracts.[156] In addition, they integrated these results with a high-performance thin-layer chromatography (HPTLC)-based approach combined with multivariate data analysis to identify bioactive compounds.[156] Michalea et al. identified bioactive compounds in acetylcholinesterase and hyaluronidase inhibition assays based on HetCA evaluations, although the investigated extracts of *Paeoniae parnassica* Tzanoud. themselves were not remarkably active.[157]

7.2.1 Eliciting Nature's Activities (ELINA)

In 2019, Grienke et al. developed the MS- and NMR-based biochemometric workflow **'Eliciting Nature's Activities'** (ELINA), which regarding NMR, primarily relies on HetCA of ^{1}H NMR-based data with the aim of identifying proton resonances associated with bioactive compounds in a complex mixture.[158] ELINA requires two preconditions: (i) the fractionation of a crude extract into microfractions (MFs) with overlapping quantitative variance of constituents, and (ii) a corresponding ascending or descending trend in bioactivity, across at least three consecutive MFs.[158] The workflow comprises microfractionation of a crude extract, ^{1}H NMR measurements and bioactivity testing of all MFs, the selection of MF packages that fulfill the aforementioned criteria, and HetCA and liquid chromatography-MS (LC-MS)/evaporative light scattering detector (ELSD) data analysis for the identification of bioactives[158] (Figure B).

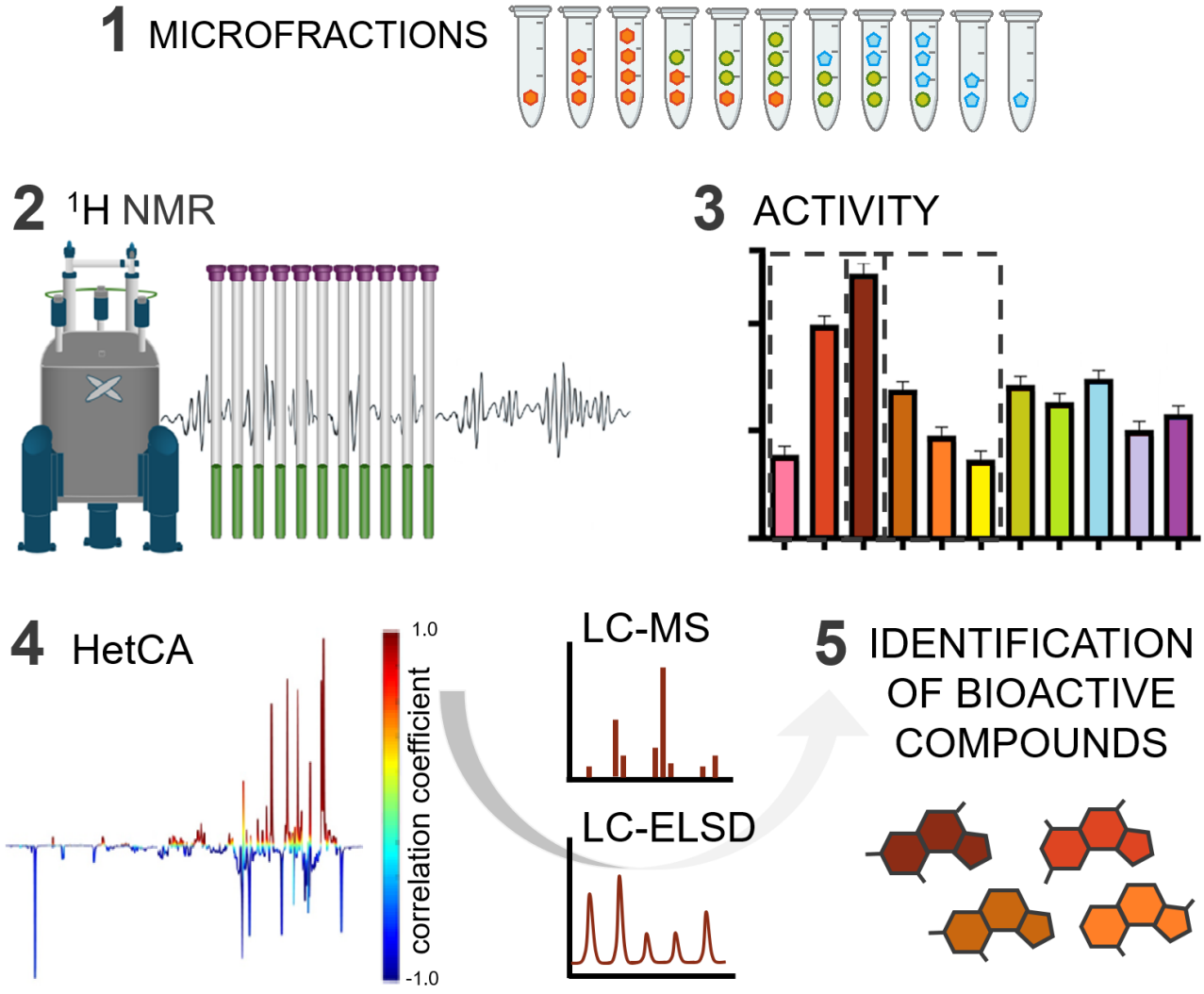


Figure B: ELINA workflow

The ELINA approach demonstrated its effectiveness on various examples, including combinatorial effects between constituents of NP-derived multicomponent mixtures.[48,159-161] Using the ELINA workflow, extracts of *Peucedanum ostruthium* (L.) Koch (masterwort) were investigated for heterocovariance between *in vitro* anti-inflammatory activity results and ^{1}H NMR data.[159] The study revealed four furanocoumarins (imperatorin, ostruthol, saxalin, and 2'-O-acetyloxypeucedanin), one coumarin (ostruthin), and one chromone (peucenin) with nuclear factor-κB (NF-κB) inhibiting activity.[159] In addition, synergistic effects between the metabolites imperatorin and peucenin were detetected.[159]

In 2023, Zwirchmayr et al. identified anthelmintic activity resulting from a combinatorial effect of ostruthin and isoimperatorin, also derived from masterwort, and confirmed this in larval studies of *Caenorhabditis elegans*.[48] Moreover, HetCA pinpointed ostruthin as active against *Enterococcus sp*. and *S. aureus*.[48]

Another example is the HetCA-based investigation of fruits from *Azadirachta indica* A. Juss. for compounds with *in vitro* osteoclast formation-inhibiting activity.[160] Two phenylpropanoids, coniferyl alcohol and sinapyl alcohol, exhibited a dose-dependent inhibition of osteoclast differentiation with negligible cytotoxic effects.[160]

Four flavonoids derived from a *Pterocarpus santalinus* L. extract were identified using the ELINA approach in the search for *in vitro* anti-inflammatory effects related to C-X3-C Motif Chemokine Ligand 1 (CX3CL1) gene expression in human umbilical vein endothelial cells (HUVECs).[161] As the authors pointed out, ELINA provides a holistic perspective on potential pleiotropic effects of bioactive compounds within a multicomponent extract, since this approach simultaneously screens various constituents for statistical covariances.[161]

Another noteworthy study was the ^{1}H NMR-based HetCA investigation by Langeder et al., which identified MDAAs as the most bioactive compounds in the extract of *M. alba* L. root bark, with SGC being the primary contributor to anti-influenza virus activity.[122]

Recently, the concept of ELINA was expanded to integrate both NMR- and MS^{2}-based datasets.[130] To this end, Wasilewicz et al. investigated a bioactive crude flower extract from *Buddleja officinalis* Maxim. and generated both ^{1}H NMR-based HetCA pseudospectra and a bioactivity-based molecular network related to antioxidant effects in the context of dry eye disease pathology.[130] The data were subsequently integrated to facilitate the targeted discovery of bioactive compounds.[130] This method, on the one hand, leverages the advantages of the quantitative and universal information provided by NMR experiments, which is independent of the ionization ability of compounds.[130] On the other hand, it benefits from the high sensitivity of MS measurements and the potential to gain insights into structure−activity relationships through the clustering in molecular families within the molecular network.[130]

By May 2025, all publications referenced in the context of HetCA had employed ^{1}H NMR data, but not two-dimensional (2D) NMR data, for biochemometric analysis. However, the interpretation of ^{1}H NMR-based HetCA pseudospectra derived from multicomponent NP mixtures containing structurally similar molecules can be quite challenging due to extensive signal overlap.[162-164]

For example, the class of triterpenes (TTs) includes many structural analogs and presents several analytical challenges, including overlapping ^{1}H NMR resonances, especially in the aliphatic region.[165,166] The scarcity or absence of chromophores leads to reduced sensitivity for UV absorption[167], while low volatility combined with relatively high molecular weight limits the use of GC without derivatization.[168] Lipophilic TTs with few polar functional groups exhibit poor ionizability in MS-based experiments[169] and may also be affected by ion suppression effects, depending on the respective matrix.[168] Furthermore, MS fragmentation often fails to provide comprehensive structural information for TTs due to identical fragmentation patterns.[170] Consequently, 2D NMR experiments are particularly essential for the unambiguous structural elucidation of TTs.[158]

2D NMR techniques have also found their way into dereplication workflows.[171] The exploitation of Heteronuclear Multiple Bond Correlation spectroscopy (HMBC) and Heteronuclear Single Quantum Coherence spectroscopy (HSQC) data helps to accelerate compound identification within mixtures.[172] Enhancements in 2D NMR barcoding, executed on clusters of fingerprint signals and spatial patterns of TT-enriched fractions, together with software-based recognition, allows for the rapid dereplication of already known compounds in plant derived extracts.[171]

However, as frequently emphasized in the literature, the primary challenge in NP research is to identify and distinguish bioactive compounds from inactive ones in complex mixtures at an early stage of the phytochemical investigation.[57,131,135,158] By May 2025, only a few studies had addressed this issue, utilizing 2D NMR spectra combined with bioactivity data in a statistical manner to guide the phytochemical workup, as demonstrated in the following publications.

An NMR-based annotation platform named MADByTE, which stands for Metabolomics and Dereplication by Two-Dimensional Experiments, combines HSQC and total correlation spectroscopy (TOCSY), where cross-peaks in the 2D spectrum represent correlations between protons of the same spin system.[164,173,174] These data are used to build a chemical similarity network for dereplication and prioritization of bioactive fractions for isolation work.[164]

The so-called Data Fusion-based Discovery (DAFdiscovery) pipeline was developed with the aim of combining MS, NMR, and bioactivity data by applying STOCSY and SHY.[175] Following a GNPS-feature-based MN (GNPS-FBMN) principle, this approach allows highlighting NMR signals and MS nodes that are positively or negatively correlated with bioactivity.[175] So far, DAFdiscovery has been demonstrated for one-dimensional (1D) NMR data sets obtained from ^{1}H NMR and ^{13}C NMR measurements.[175]

'Plasmodesma' is a tool for ML-based automatic acquisition of NMR experiments to uncover spectral fingerprints of bioactive compounds in extracts and perform pharmacophoric deconvolution.[176] The developers make this approach open-access and offer interactive visualization of the statistical analysis.[176] For example, Plasmodesma was successfully used to identify active compounds in a Cinchona bark extract.[176]

While the statistical correlation of 2D NMR and bioactivity data offers promising possibilities, it is still in its early stages. Therefore, it is essential to move beyond dereplication and focus on utilizing these methods to identify bioactive compounds in complex mixtures.

In **Manuscript I**, the biochemometric 2D NMR-based HetCA approach was developed to identify bioactive compounds in multicomponent mixtures. In the first part of the study, a proof-of-concept was established using artificially mixed TTs. Subsequently, as a real-world example, a TT-rich dichloromethane leaf extract of *Eriobotrya japonica* Lindl. was selected. This subtropical, evergreen tree of the Rosaceae family, also known as loquat, is native to southeastern China and has a long history of use in traditional herbal medicine.[177,178] Its extracts, mostly derived from the leaves, have been used to treat respiratory tract diseases such as cough and bronchitis, as well as inflammation, diabetes, and cancer[177,179-185], due to their antitussive, anti-inflammatory, antidiabetic, and anti-tumor properties.[177,184-186]

Given their high structural similarity, the wide variety of pentacyclic TTs in the plant results in many overlapping signals in the ^{1}H NMR spectrum when present as a mixture.[177] Therefore, this plant material serves as a suitable application example for 2D HetCA, as the underlying hypothesis suggests that the method offers improved evaluability, particularly in the case of overlapping signals.

7.2.1.1 Manuscript I

Biochemometric 2D NMR-based heterocovariance analysis: A targeted approach for identifying bioactive compounds in complex mixtures

Sigrid Adelsberger, Alexander F. Perhal, Lorenza Bertaina, Patrik F. Schwarz, Verena M. Dirsch, Judith M. Rollinger, Ulrike Grienke

Submitted 22.04.2025

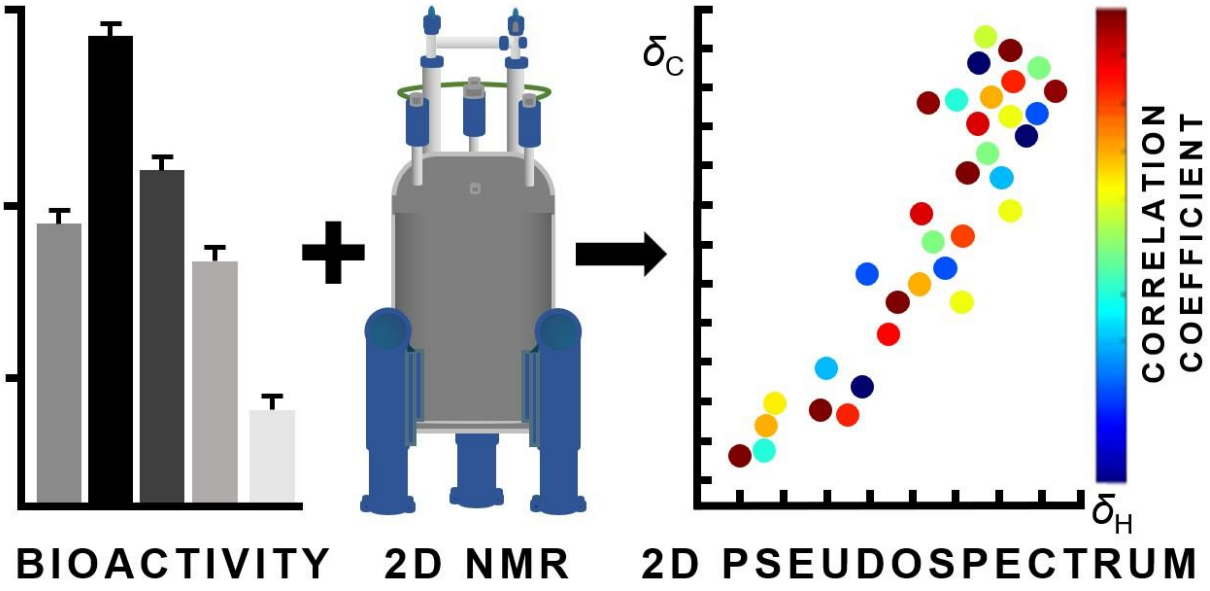


Figure C: Graphical abstract of **Manuscript I**

My contribution to this work included the phytochemical work-up, specifically microfractionation and fractionation of the dichloromethane extract of *E. japonica* leaves, as well as sample preparation, measurements and data processing steps for flash chromatography, semi-preparative supercritical fluid chromatography (SFC), ultra-high performance supercritical fluid chromatography (UHPSFC), thin-layer chromatography (TLC) and NMR experiments. I pre-processed the NMR data and calculated and analyzed 1D and 2D HetCA pseudospectra. In addition, I contributed to structure elucidation. I drafted and wrote the manuscript and prepared all figures and tables.

Biochemometric 2D NMR-based heterocovariance analysis: A targeted approach for identifying bioactive compounds in complex mixtures

Sigrid Adelsberger,[†,§] Alexander F. Perhal,[†] Lorenza Bertaina,[†,§] Patrik F. Schwarz,[†,§] Verena M. Dirsch,[†] Judith M. Rollinger,[†] Ulrike Grienke[,†]*

[†] Division of Pharmacognosy, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria

[§] Vienna Doctoral School of Pharmaceutical, Nutritional and Sport Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria

* Corresponding Author

Ulrike Grienke – Division of Pharmacognosy, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria; orcid.org/0000-0003-0305-9270; E-mail: ulrike.grienke@univie.ac.at

ABSTRACT

Biochemometric approaches, which integrate bioactivity data with spectroscopic or spectrometric data, offer significant potential to streamline the discovery of bioactive compounds in targeted isolation strategies. However, the complexity of natural extracts and the presence of structurally similar analogs make this process time-consuming and resource intensive. This study introduces a 2D Nuclear Magnetic Resonance (NMR)-based heterocovariance analysis (HetCA) workflow to identify chemical features that correlate positively or negatively with bioactivity in complex mixtures. As a proof-of-concept, the workflow was established using artificially mixed samples of pentacyclic triterpenes which were screened for modulatory activities of the retinoic acid receptor-related orphan receptor gamma (RORγ) and the G protein-coupled bile acid receptor (TGR5). The validated concept was then exemplified using a triterpene-rich *Eriobotrya japonica* leaf extract. The applied workflow enabled the rapid, targeted, and accurate identification of bioactive constituents from *E. japonica* that modulate RORγ and/or TGR5 using this newly developed biochemometric 2D NMR HetCA approach.

INTRODUCTION

In search of new drug leads, natural products (NPs) are an important source[1] with the disadvantage of having to deal with complex multicomponent crude extracts as starting materials. [2] Biochemometric approaches allow the precise and targeted detection of bioactive constituents. [3-7] The previously developed workflow ELINA (Eliciting Nature's Activities) [8] primarily relies on ^{1}H NMR-based statistical heterocovariance analysis (HetCA) [9], which detects specific chemical features, i.e. resonances, that are positively or negatively correlated with bioactivity prior to isolation. [10-13]

The interpretation of ^{1}H NMR HetCA pseudospectra derived from mixtures of structural analogs with overlapping resonances pushes biochemometric methods to their limits. [14-16] In particular, triterpenes (TTs) represent a class of compounds that includes many structural analogs. [17] In addition to overlapping resonances, especially in the aliphatic region, TTs pose several analytical challenges, e.g. reduced sensitivity for UV absorption[18], low volatility[19], and poor ionizability in MS-based experiments. [20] Hence, 2D NMR experiments are crucial for unambiguous structure determination of TTs. [8]

The retinoic acid receptor-related orphan receptor gamma isoform RORγt and the Takeda G protein-coupled receptor 5 (TGR5) were selected as suitable targets since their endogenous ligands show a structural similarity to the investigated compound class of tetracyclic and pentacyclic TTs. [21, 22] RORγ promotes the differentiation of naïve T cells into interleukin (IL)-17 producing T helper cells (Th17), which release pro-inflammatory cytokines such as IL-17 and IL-22. [21] Small molecules can act as inverse agonists of RORγ thereby reducing cytokine production and inflammation. [23] TGR5, also known as GPBAR-1 or M-BAR[22], is involved in the regulation of metabolic diseases and inflammation. [24] When stimulated by agonists, TGR5 exerts anti-inflammatory effects via inhibition of transcriptional nuclear factor-κB (NF-κB) as well as modulating other signaling pathways. [25-27]

2D NMR techniques such as Heteronuclear Single Quantum Coherence (HSQC) have found their way into dereplication workflows. [28] However, statistical correlation of 2D NMR with bioactivity data, while promising, is still in its infancy. The aim of this study was to move beyond dereplication and to unravel RORγ and TGR5 modulating TTs in mixtures in a fast, targeted, and unambiguous manner. Therefore, a biochemometric 2D NMR-based HetCA approach was developed. As a proof-of-concept, artificially mixed TTs underwent biochemometric analyses and as an authentic application sample, a dichloromethane extract of TT-rich *Eriobotrya japonica* leaves was selected.

EXPERIMENTAL SECTION

Solvents and Reagents. All solvents (analytical grade), ammonium formate (HiPerSolv CHROMANORM®, ≥99%), and ethanol (>99.7%) were purchased from VWR Chemicals (Radnor, USA). Chloroform-D1 (MagniSolv^{TM}, 99.8%) was purchased from Merck (Darmstadt, Germany), carbon dioxide (CO_2) 4.5 from Messer (Vienna, Austria), sulfuric acid (95.0-98.0%), vanillin (ReagentPlus®, 99%), and Dulbecco`s Modified Eagle`s Medium (DMEM) (D6546) from Sigma-Aldrich (St. Luois, USA). Fetal bovine serum (FBS) (S1810) was purchased from Biowest (Nuaille, France), glutamine (BE17-605E), penicillin and streptomycin (DE17-602E) from Lonza Group AG (Basel, Switzerland), 5X reporter lysis buffer (E397A) from Promega (Fitchburg, USA), trypsin/EDTA from Thermo Fisher Scientific (Waltham, USA). Double-distilled deionized water (dd H_2O) was produced by a Milli-Q-Plus ultrapure water device (Millipore Corporation, Bedford, MA, USA) and by an Arium® Pro Ultrapure Lab Water System from Sartorius (Göttingen, Germany).

Plant Material. Dried leaves of *E. japonica* (batch number 030724) were provided by Plantasia GmbH (Oberndorf/Salzburg, Austria). A voucher specimen (JR-20250218-A1) is stored at the Division of Pharmacognosy, Department of Pharmaceutical Sciences, University of Vienna, Austria.

Sample Preparation. *Extraction.* An *E. japonica* dichloromethane extract (EJD) was prepared from 512.31 g dried and milled leaves with 4.5 L dichloromethane at room temperature on an orbital

shaker GFL-3005 (GFL Gesellschaft für Labortechnik, Burgwedel, Germany). Therefore, the plant material was divided into two equal parts and put into two 3000 mL Erlenmeyer flasks with 2250 mL dichloromethane each. The extraction was executed four times (extraction duration 4 days, 2 days, 1 day, and 5 days). After filtering, the solvent was evaporated on a rotary evaporator (Rotavapor R II, Büchi, Flawil, Switzerland) at 40 °C. The material was dried in a desiccator, 16.46 g EJD were obtained.

Flash chromatography. EJD was fractionated by flash chromatography on an Interchim puriFlash 4250 instrument (Interchim, Montluçon, France) controlled by Interchim Software into 45 microfractions (MFs), named MF1–MF45. The device was equipped with a photo diode array (PDA) detector, an evaporative light scattering detector (ELSD), and a Silica HP 220 Gramm 25 μm 220 bar puriFlash column (PF-25SIHC-F0220, Interchim, Montluçon, France). 13.75 g of EJD were prepared for dry load application by mixing it with silica gel 60 (0.040-0.063 mm, CAS-No 7631-86-9, 60.08 g/mol, Merck KGaA, Darmstadt, Germany) in a ratio of 1:1 . The gradient for n-hexane (A) and acetone (B), with a flow rate of 50.0 mL/min, was: 2% B for 3 column volumes (cv), 2% B to 5% B in 7 cv, 5% B for 5 cv, 5% B to 10% B in 7 cv, 10% B for 5 cv, 10% B to 15% B in 7 cv, 15% B for 5 cv, 15% B to 20% B in 7 cv, 20% B to 40% B in 5 cv, 40% B for 7 cv, 40% B to 100% B in 10 cv, 100% B for 7 cv. To combine in total 2121 test tubes with 20 mL each into 45 MFs, results of Ultra-high performance supercritical fluid chromatography (UHPSFC) and thin layer chromatography (TLC) were used.

The fractionation of MF33 with the same device and detectors via a C_{18} HQ 12 Gramm 15 μm 22 bar column (PF-15C18HQ-F0012) was conducted with a flow rate of 15.0 mL/min with dd H_2O (A) and acetonitrile (B): 5% B for 3 min, 5% B to 45% B in 4 min, 45% B to 98% B in 80 min, 98% B for 17 min, 98% to 5% B in 0 min, 5% B for 15 min. 124.20 mg of MF33 were homogenized 1:1 with 125.0 mg Silica gel 60 for dry load application. UHPSFC and TLC analyses were used to guide the pooling of 359 test tubes into eight fractions A1–A8.

Semi-preparative SFC. Fractionation of 30 mg A3 in methanol (25.29 mg/mL) was performed on a SFC Prep-15 device (Waters, Milford, MA, USA) equipped with an ELSD and PDA detector. A Torus™ 1-aminoanthracene column (OBD™, 130 Å, 10 x 250 mm, 5 μm) (Waters, Milford, MA, USA) was kept at 45 °C. The mobile phase consisted of supercritical CO_2 and ethanol absolute >99.7% with a gradient [time(min)/% B]: 0.0/10, 1.5/10, 16.5/11, 18.5/50, 20.5/50, 21.5/10, 22.5/10. Three fractions, named B1–B3, were collected according to the ELSD chromatogram.

UHPSFC. UHPSFC analysis of MF1–MF45 was performed on an Acquity UPC² device (Waters, Milford, MA, USA) equipped with a PDA detector and ELSD. A Torus™ 1-aminoanthracene column (130 Å, 3.0 x 100 mm, 1.7 μm) (Waters, Milford, MA, USA) was kept at 45 °C. Supercritical CO_2 (A) and methanol (B), flow rate 1.0 mL/min, were used for the gradient in [time(min)/% B]: 0/10, 6.5/19, 7.5/50, 8.5/50, 9/10, 10/10. Samples were prepared with 5 mg/mL in n-hexane and 2-propanol 7:3. The measurements for A1–A8 and B1–B3 (5 mg/mL in acetonitrile and methanol) were performed in the same way with some changes. The gradient in [time(min)/% B] was: 0/10, 15/13, 15.2/50, 16.2/50, 16.50/10, 17/10. Additionally, a single quadrupole mass spectrometry detector (QDa by Waters (Milford, MA, USA)) was used with a flow rate of 0.6 mL/min of 10 µM ammonium formate (in dd H_2O : methanol 1:9) as make-up solvent. Both positive (15 V cone voltage) and negative mode (30 V) were recorded with a scan range from 100.00–1000.00 Da.

TLC. TLC was performed using silica gel 60 F_{254} plates (Merck, Darmstadt, Germany). The mobile phase consisted of methanol, dichloromethane and ethyl acetate with 10:1.2:0.5 v/v. The analysis was performed at visible light after derivatization with vanillin (1% in methanol) and sulfuric acid (5% in methanol) and temperature increase to 105 °C for 2 min.

Cell culture. HEK293 cells (CRL-1573, ATCC) were grown in DMEM supplemented with 10% FBS, 2 mM glutamine, 100 U/mL penicillin, and 100 µg/mL streptomycin. Cells were cultured at 37°C and 5% CO_2 and passaged every 2-3 days, with a maximum use up to passage number 35.

RORγ-Gal4 luciferase reporter gene assay. A hybrid RORγ-Gal4 construct was employed, in which the RORγ ligand-binding domain (LBD) is fused to the Gal4 DNA-binding domain (DBD),

a yeast transcription factor that drives luciferase expression through a yeast upstream activating sequence, as previously described. [29] Briefly, 8 x 10^6 HEK293 cells were transfected with the calcium phosphate method using 5 μg of Gal4-receptor (RORγ-Gal4: Dr. Fabio R. Santori, Center for Molecular Medicine, University of Georgia, Athens, Georgia, USA), 5 μg of response element (tk(MH1000)4xLuc: Prof. Ronald Evans, Salk Institute for Biological Studies, La Jolla, California, USA), and 3 μg of eGFP plasmid (pEGFP-N1: Clontech, Mountain View, CA, USA), and incubated overnight. Medium was changed, and cells were further incubated for 5 h before resuspension in 5% charcoal-stripped FBS-DMEM. Cells were then reseeded into a 96-well plate (5 x 10^4 cells/well) and treated with compound solutions prepared in the same medium for 18 h. After cell lysis in 5X reporter lysis buffer, luminescence and fluorescence values were measured on a TECAN Spark Microplate Reader (TECAN Group AG, Männedorf, Switzerland). Luminescence signals were normalized to fluorescence emissions and expressed as fold activations normalized to the vehicle control dimethyl sulfoxide (DMSO) 0.1%.

CRE-luciferase reporter gene assay. As an indirect measurement of TGR5 receptor activation by compound treatment, CRE-luciferase reporter gene assays were performed as previously described. [30] Briefly, HEK293 cells stably expressing an EPAC-based FRET biosensor as well as the TGR5 receptor (TGR5 HEK EPAC) were seeded at a concentration of 8 × 10^6 cells on 150 mm cell culture dishes 4–5 h prior to transfection. Cells were then transfected with 10 μg CRE-luciferase reporter plasmid overnight. On the next day, cells were allowed to recover from transfection for 4–5 h in fresh medium before detachment from dishes using trypsin/EDTA. Cell suspensions were diluted in stripped (5% charcoal-stripped FBS) DMEM to a concentration of 5 × 10^4 cells per well onto a 96-well plate and treated with vehicle control (0.1% DMSO), positive control LCA (10 μM) or compounds at the indicated concentrations for 18 h. Following treatment, the cells were lysed in 5X reporter lysis buffer. The basal fluorescence emission of the stably transfected EPAC FRET biosensor was used for cell number normalization and measured at an emission wavelength of 520 nm (excitation wavelength at 485 nm) using a TECAN Spark Microplate Reader (TECAN Group

AG, Männedorf, Switzerland). Luminescence values (RLU) were measured and normalized to the respective fluorescence values (RFU) to account for differences in cell numbers (RLU/RFU) and normalized to the vehicle control (0.1% DMSO). Results are expressed as fold activations relative to vehicle control.

Statistics. Data obtained in the RORγ-Gal4 luciferase reporter gene assays and the CRE-luciferase reporter gene assays were checked for normal distribution via D`Agostino & Pearson calculations. For nonparametric data (as was the case for MF13 in Figure S8 only), Mann-Whitney-U-test was performed. Statistical analysis included one-way ANOVA and Dunnett`s post-hoc test (****$p \leq 0.0001$, ***$p \leq 0.001$, **$p \leq 0.01$, *$p \leq 0.05$, n. s. $p > 0.05$) versus vehicle control DMSO 0.1%. Data analysis was performed with Microsoft Excel version 2404 (Microsoft Corporation, Redmond, USA) and GraphPad Prism version 10.2.3 (403) (Dotmatics, Boston, USA). Results are expressed as mean ± standard deviation measured in n biological replicates. Nonlinear regression was used for IC_{50} value calculation. NMR data processing was performed with TopSpin (Version 4.1.4, Bruker BioSpin GmbH, Rheinstetten, Germany, 2022). Statistical correlation of NMR data with bioactivity data for the calculation of 1D and 2D HetCA pseudospectra was performed with MATLAB (R2020a Update 4, 9.8.0.1417392, The MathWorks GmbH., Munich, Germany).

NMR Spectroscopy and Data Pretreatment. Compounds **1**, **7**, **8**, **12**, **13**, AMF1–AMF8, MF33–MF35, and B1–B3, were dissolved in deuterated chloroform-*d* with 3 mg/mL. 750 μL of each sample were centrifuged at 3000 rpm for 5 min. 650 μL of the supernatant were transferred to an NMR tube (LabScape, Bruker, Germany). For all samples both ^{1}H NMR spectra and HSQC spectra were recorded, for structure elucidation of B1– B3, additionally APT, HMBC, and ^{1}H-^{1}H COSY experiments were conducted, as supplied by the manufacturer. NMR experiments were performed at 298 K on a Bruker Ascend 500 MHz NMR spectrometer (Bruker, Billerica, MA, USA), equipped with a SampleJet, a 5 mm TCI Prodigy CryoProbe, and an Avance NEO console, and referenced to the residual non-deuterated solvent signals (δ_H 7.26 ppm; δ_C 77.16 ppm). The resonance frequency

for ^{1}H NMR was 500.19 MHz and for ^{13}C NMR 125.77 MHz. HSQC spectra were collected with 2028 points in F2 and 256 increments in F1, with 32 dummy scans and 4 scans.

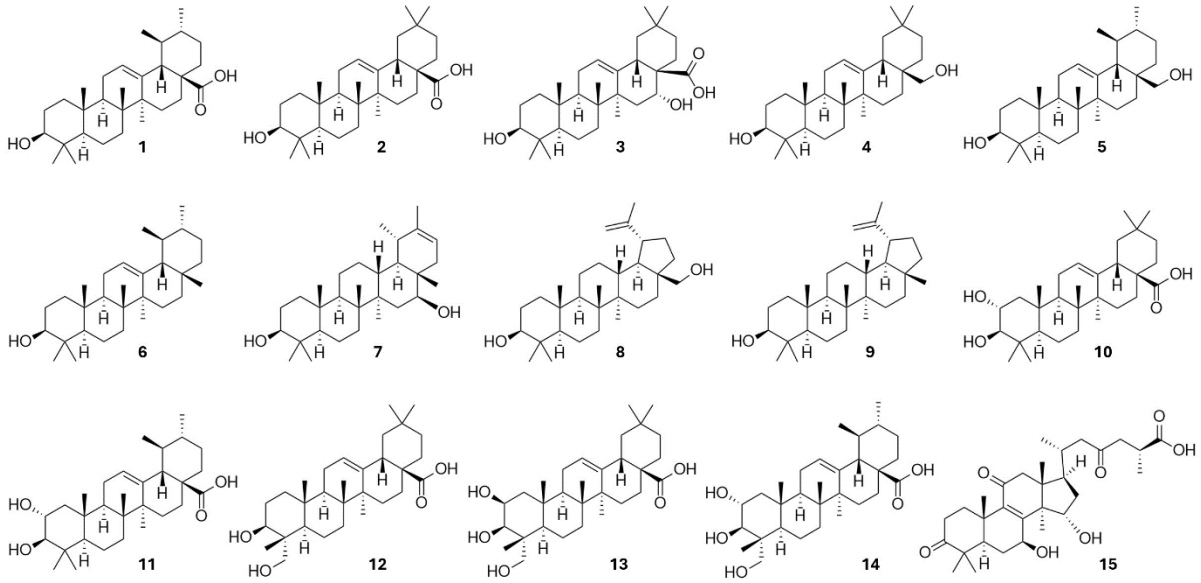
NMR data processing and HetCA. As pre-processing steps, the ^{1}H NMR spectral data underwent the protocol as described before[8,9] including phase correction, a manually processed fifth order polynomial baseline correction based on the mathematical equation $A + Bx + Cx^2 + Dx^3 + Ex^4$ and an axis calibration for the solvent used. Spectral features from ^{1}H NMR and HSQC data sets were then statistically correlated with the bioactivity data on RORγ and TGR5 via HetCA by MATLAB. For 1D HetCA of AMF1–AMF4 and MF33–MF35, the proton NMR spectra were bucketed with a width of 0.0005 ppm, in the range δ_H 0.5–10.0. Peak picking of HSQC cross-peaks for 2D HetCA of AMF1–AMF4 was performed in the range of δ_H 0.0–10.0 and δ_C 0–150 with a maximum of 100 picked peaks, and for MF33–35 in the range of δ_H 0.0–8.0 and δ_C 0–150 with a maximum of 500 picked peaks with at least 2% of the highest intensity. Some signals, that were not ideally captured during automatic processing, were corrected manually. Additionally, signals of artefacts, known from the 1D HetCA results, were excluded from further processing. The size of the generated cross-peaks in the 2D HetCA pseudospectra were unified. The HetCA plots visualize the calculated covariance, with positively correlated signals pointing upward and negatively correlated signals pointing downward. The gradation of the correlation coefficient (cc) between -1.0 and 1.0 for all HetCA experiments was depicted by MATLAB default color-code type jet from dark blue to dark red.

RESULTS AND DISCUSSION

Selection of compounds and preparation of artificial TT mixtures for a proof-of-concept study. The first aim of this work was to set up a proof-of-concept study to evaluate whether a 2D NMR-based biochemometric approach is superior to 1D techniques and suitable for the differentiation of active and inactive structural analogs with overlapping ^{1}H NMR resonances. For this purpose, 15 known active as well as known inactive TTs were selected deliberately: ursolic acid (**1**), oleanolic acid (**2**), echinocystic acid (**3**), erythrodiol (**4**), uvaol (**5**), α-amyrin (**6**), faradiol (**7**),

betulin (**8**), lupeol (**9**), maslinic acid (**10**), corosolic acid (**11**), hederagenin (**12**), bayogenin (**13**), asiatic acid (**14**), and ganoderic acid A (**15**) (Chart 1, Table S1).

Chart 1. Chemical structures of 15 selected triterpenes (TTs). Ursolic acid (**1**), oleanolic acid (**2**), echinocystic acid (**3**), erythrodiol (**4**), uvaol (**5**), *α*-amyrin (**6**), faradiol (**7**), betulin (**8**), lupeol (**9**), maslinic acid (**10**), corosolic acid (**11**), hederagenin (**12**), bayogenin (**13**), asiatic acid (**14**), and ganoderic acid A (**15**).



To obtain consistent bioactivity data independent of already described activities, all 15 TTs were tested in the same assay systems. For evaluating the activity on ROR*γ*, a Gal4 luciferase assay was employed, and for testing the activity on TGR5, a stable TGR5-expressing cell line with a CRE luciferase reporter gene were used. Based on these results, the 15 selected TTs were classified into three groups covering active, moderately active, and inactive TTs (Figure S1). In general, the results for TGR5 were in accordance with the literature and a structure-activity relationship study proposed by Genet et al. [31], concluding that TTs with the chemical features of a 3*α*-hydroxyl group, a carboxyl group at C-17 combined with a pentacyclic scaffold, have a high potential for activity on TGR5. Only for compound **12**, the result obtained for TGR5 was contradicting. For ROR*γ*, the results were consistent with literature data except for the lower activity of **4**[32] (Table S2).

To keep the proof-of-concept study simple, five out of the 15 TTs were selected to cover the three preassigned bioactivity groups on both targets. One active compound (**1**), one moderately active compound (**12**), and three inactive compounds (**7**, **8**, and **13**) were used to prepare mixtures in different quantitative ratios to obtain eight artificial microfractions (AMF1–AMF8) simulating chromatographic fractionation (Table S3). All AMFs were subjected to bioactivity assessment on RORγ and TGR5 in luciferase reporter gene assays as well as 1D and 2D NMR measurement. Based on the bioactivity results, it was possible to define fraction packages with increasing or decreasing bioactivity, e.g. AMF1–AMF4 for the target TGR5 showing a distinct sequentially decreasing bioactivity (Figure 1).

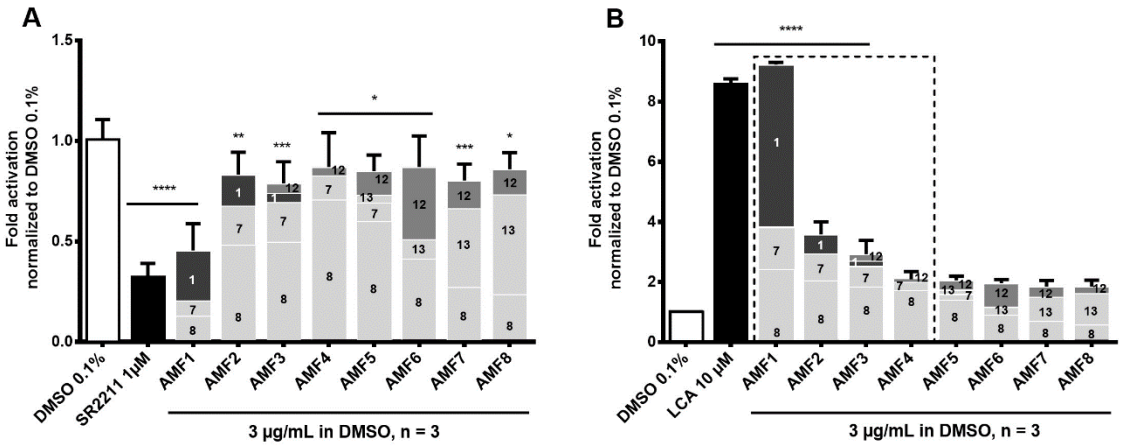


Figure 1. (A) Bioactivities of artificial TT mixtures AMF1–AMF8 (3 µg/mL) on the targets RORγ in RORγ-Gal4 luciferase assays and (B) TGR5 in CRE-luciferase assays. The color coding within the individual bars reflects the bioactivity of the respective compound (dark grey for active, medium grey for moderately active, and light grey for inactive) of the mixed TTs, and each colored section is labeled with the respective compound number. The size of a section within a bar represents the relative amount of a compound in the corresponding AMF. Package AMF1–AMF4 was used for HetCA (marked with a dashed rectangle).

1D and 2D NMR measurements including data analysis. For 1D HetCA, proton NMR experiments were performed, for 2D HetCA HSQC data sets were recorded. To ensure consistent

signal-to-noise ratios, samples were prepared in the same way and measured under the same conditions (Figures S2, S3, S4).

The compounds selected possess no aromatic structures, only a few alkenyl hydrogens and many aliphatic protons. Thus, in the chemical shift region δ_H 2.5–6.0 only a few resonances are present while the upfield region (δ_H 0.0–2.5) is rather crowded with overlapping signals. Due to signal overlap, unambiguous assignment of all five compounds in the mixture is challenging. Comparison with superimposed ^{1}H NMR spectra allowed only to assign **1** and **8** unambiguously (Figure S5), whereas evaluation of the superimposed 2D HSQC spectra made assignments for **1**, **7**, **6**, **12**, and **13** possible (Figure S6). In general, for the identification of bioactives in mixtures of unknown compounds, the aromatic region (approximately δ_H 6.5–8.5) can also show overlapped signals and would benefit from available HSQC data.

1D and 2D NMR data processing for HetCA. To perform HetCA, two prerequisites must be met: (i) a bioactive crude extract must be fractionated into microfractions (MFs) that contain an intersecting quantitative variance of constituents, and (ii), this quantitative variance must be reflected in a correlated ascending or descending variance of bioactivity. [8] For the statistical correlation between chemical features and bioactivity data, raw ^{1}H NMR spectral data are processed according to a standard protocol. [8, 9] MATLAB is then used for HetCA of selected fraction packages consisting of at least 3 MFs. To reduce the complexity of spectra, each spectrum is divided into buckets of defined width. The data points of each spectroscopic bucket are summed up and the corresponding ^{1}H NMR resonance intensities are calculated. After implementing bioactivity data, the HetCA process correlates them with the ^{1}H NMR resonance intensities. To obtain color-coded pseudospectra with positively (red, upward signals) or negatively (blue, downward signals) correlated resonances, the correlation coefficient (cc), calculated via normalization of the covariance, is plotted for each bucket.

To calculate HetCA with 2D NMR spectra, the same preprocessing steps were performed as for 1D spectra. However, after phase correction and calibration of the spectra, the first step consisted of

peak picking for each individual spectrum of the selected fraction package (AMF1-AMF4). All cross-peak intensities together with their δ_H and δ_C values were collected in one file and imported into MATLAB. In accordance with 1D processing, the spectral data were divided into buckets of a certain ppm width. The next step was to import the bioactivity values for each fraction package. HetCA was then performed in the same way as for the 1D spectra. [8, 9] Finally, the 2D HetCA pseudospectrum was visualized, again using a color-code to show the cc.

1D and 2D HetCA pseudospectra for proof-of-concept study. Chemical features of the 1D HetCA pseudospectrum for TGR5 activity of AMF1–AMF4 appear as dark red signals when positively correlated (upward), and dark blue when correlated negatively with bioactivity (downward) (Figure 2).

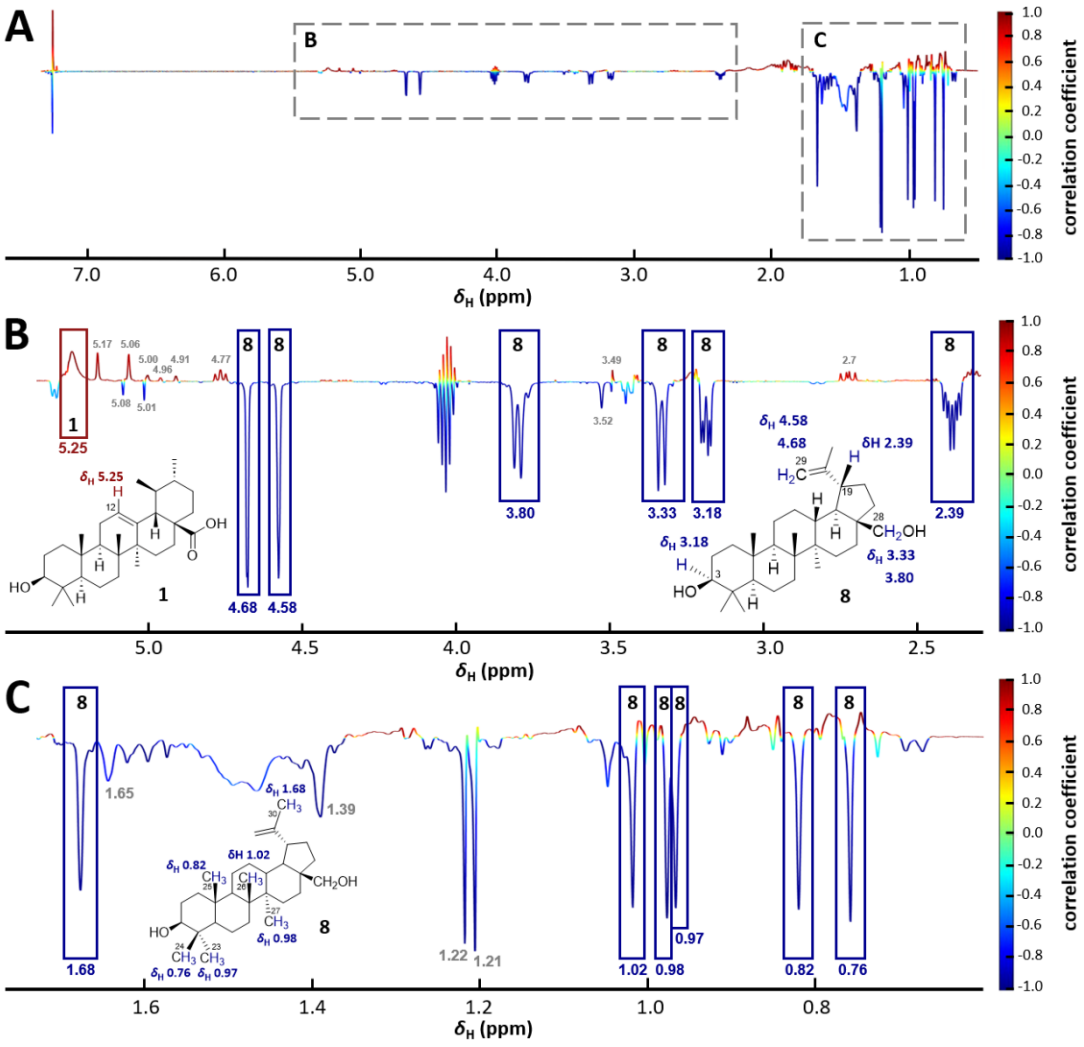


Figure 2. (A) 1D HetCA pseudospectrum of AMF1–AMF4 depicting TGR5 bioactivity with δ_H resonances on the x-axis and the correlation coefficient on the y-axis. (B) Zoom into the region δ_H 2.30–5.50 and assignment to the involved TTs **1** and **8**. (C) In the aliphatic region, only negatively correlated signals related to **8** could be determined.

Based on the 1D HetCA results, compound **1** was statistically determined to be active, and **8** to be inactive. This is consistent with the experimental TGR5 bioactivity data. Compounds **7** and **12** could not be assigned to any noticeable signals in the 1D pseudospectrum and were therefore not covered by 1D HetCA. Quantitatively, **8** was the main component in the AMF1–AMF4 set, based on the high intensity of its proton resonances. Due to overlapping signals, the 1D HetCA analysis provided no further insight.

To overcome the limitations of the 1D NMR-based approach, a 2D HetCA plot for the same sample set AMF1–AMF4 was generated. Lessons learned from the 1D HetCA results were considered when preparing the data for the 2D analysis. For instance, HSQC signals of artefacts were not included in the total peak list after manual peak picking. In contrast, essential marker signals in the 1D pseudospectrum, such as δ_H 5.25 for compound **1** and δ_H 3.18 for compound **8**, were specifically targeted for 2D HetCA. All cross-peaks in the 2D plot were unified to the same size to avoid overlooking less intense signals. The 2D HetCA of AMF1–AMF4 resulted in a 2D pseudospectrum with δ_H values on the x-axis, δ_C values on the left y-axis and the color-coded cc on the right y-axis (Figure 3).

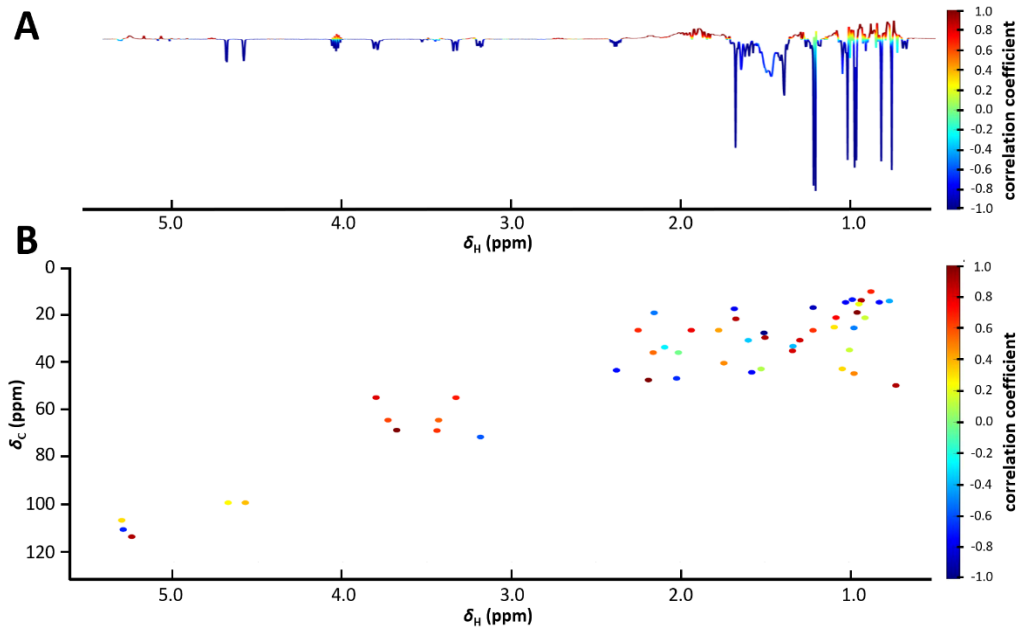


Figure 3. ^{1}H-based 1D HetCA pseudospectrum (A) and HSQC-based 2D HetCA pseudospectrum (B) for TGR5 bioactivity of AMF1–AMF4. The x-axis and left y-axis of the 2D plot shows the δ_H and δ_C values in ppm, the right y-axis represents the calculated correlation coefficient , color-coded from active (dark red), to inactive (dark blue) signals. Listed δ_H,δ_C value pairs see Table S4.

2D HetCA distributes the calculated signals in a 2D plot based on δ_H and δ_C values and allows better interpretability due to a distinction between similar δ_H values that overlap in the 1D HetCA plot. Evaluation of the δ_H 0.0–2.5 region of the 2D pseudospectrum allowed differentiation of active and inactive signals compared to the 1D plot in the same ppm range. Compound **1** could be assigned as positively correlated based on red cross-peaks at δ_H,δ_C 0.72,55.0; 0.93,15.2; 0.95,20.9; 1.08, 23.3; 1.33,38.8; 1.50,32.7; 1.67,23.9; and 2.19,52.5. The inactive compound **8** was confirmed by blue cross-peaks at δ_H,δ_C 0.76,15.1; 0.82,16.1; 0.97,28.1; 0.98,14.9; 1.02,16.3; 1.60,48.9; and 2.38,47.9, but two false calculations occurred (1.21,29.3 (dark orange) and 1.93,29.2 (medium red)). The moderately active status of compound **12** was demonstrated by signals of different colors at δ_H,δ_C 0.87,11.1 (light red), 0.90,23.4 (green), 0.94,16.9 (green), 0.97,49.5 (orange), 1.29,33.9 (medium red), 1.74,44.6 (orange), 1.77,29.2 (orange), 2.01,39.7 (green), and 2.16,39.7 (orange). Compound **7** could be correctly assigned as inactive based on signals at δ_H,δ_C 1.04,47.3 (yellow), 1.33,36.7 (light blue), and 2.10,37.2 (turquoise).

In the δ_H 3.00–5.50 region of the 1D HetCA plot, the signals at δ_H 4.58, 4.68, 3.80, and 3.33 belonging to compound **8**, were correctly calculated as negatively correlated. In the 2D plot, the cross-peaks calculated for these resonances are displayed as signals with a weak or moderate positive correlation with activity (yellow, orange). Additionally, two 2D HetCA calculations in the same range were incorrect. The δ_H,δ_C value pair 5.30,122.0 belonging to the moderately active compound **12** was predicted to be inactive (blue), and the δ_H,δ_C value pair 3.44,76.1 belonging to the inactive compound **7** was calculated to be moderately active (orange). In the same ppm area, the 2D approach correctly identified **12** as moderately active (based on δ_H,δ_C 3.73,71.2; 3.68,75.9; and 3.43,71.2), and **1** as active (based on δ_H,δ_C 5.25,126.0).

However, of a total of 51 calculated signals in the 2D plot, 34 signals were correctly assigned to the categories active, moderately active, and inactive, five were ambiguous, and twelve were incorrect. Dark red cross-peaks were obtained only for signals related to active compound **1** and moderately active compound **12**. Nine out of 14 dark blue cross-peaks were correctly assigned to the inactive compound **8**. While 1D HetCA only enabled the detection and correct TGR5 bioactivity classification of compounds **1** and **8**, the 2D HetCA results provided the mapping and classification of compounds of **1**, **7**, **8**, and **12**. More importantly, a potential subsequent targeted isolation would focus on the correct NMR signals of the 2D HetCA. Furthermore, not only was the dispersion and visibility of the signals in the 2D HetCA improved, but also a greater variation of the color code gradations between positive and negative correlations was observed.

Evaluation of 1D and 2D HetCA using a botanical extract. To test the applicability of 2D HetCA for complex extracts, in the second part of this study, both 1D and 2D HetCA were applied and compared for the analysis of bioactive TTs in *E. japonica*. [33] The plant is known to contain a wide variety of at least 47 pentacyclic TTs[33] in different quantitative ratios. [34] Many of the TTs differ only in one or a few functional groups[33], which complicates the search for bioactives based on conventional methods such as bioassay-guided fractionation. [3] Therefore, the aim was to determine

whether 2D HetCA allows to fish out compounds with RORγ bioactivity in a targeted way, and to compare the results of the 2D approach with those of the 1D plot.

Dried leaves of *E. japonica* were macerated with dichloromethane at room temperature to obtain an extract (EJD). EJD was fractionated by normal-phase flash chromatography into 45 microfractions (MFs), designated MF1–MF45, and tested for their inverse agonist activity on RORγ in luciferase reporter gene assays at a concentration of 3 μg/mL.

In total, 10 fraction packages were generated for consecutive MFs based on their ascending or descending activities on RORγ (Figure S7). All packages underwent UHPSFC-MS analysis and preliminary compound annotation by comparison with literature data. [33, 35, 36] The most promising packages with partially unknown constituents were prioritized for the 1D and 2D NMR HetCA approach. Although the fraction package MF33–35 contains known bioactive TTs such as **1** and **2** (Figure S8), it was selected due to additional peaks that could not be assigned to reference compounds (Figure S9). Therefore, ^{1}H NMR and HSQC spectra of MF33–35 were recorded (Figures S10, S11) and processed according to the protocol described for the proof-of-concept study.

1D and 2D NMR-based HetCA analyses of MF33–35. The 1D NMR-based HetCA analysis revealed positively correlated (red, up) and negatively correlated (blue, down) ^{1}H NMR resonances associated with activity on RORγ; the 2D NMR-based HetCA calculations yielded 207 cross-peaks with color coding ranging from dark red to dark blue (Figure 4).

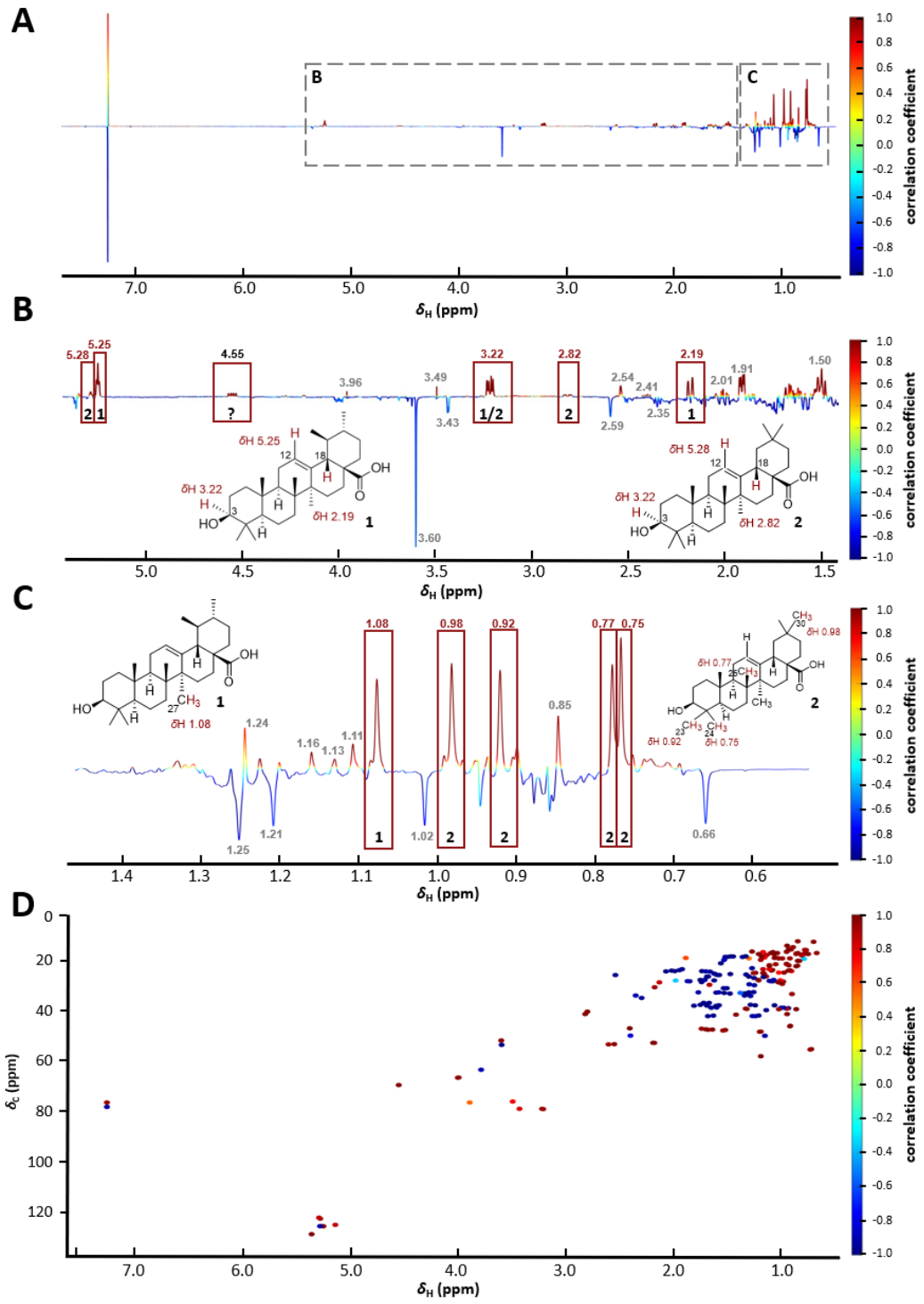


Figure 4. 1D HetCA pseudospectrum of MF33–35 based on RORγ activity (A), with enlarged regions δ_H 1.50–5.05 (B) and δ_H 0.50–1.50 (C). (D) 2D HetCA pseudospectrum, based on HSQC data, with δ_H values on the x-axis, δ_C values on the left y-axis, correlation coefficient on the right y-axis. Positively correlated signals shown in dark red, negatively correlated signals in dark blue.

Although many of the positively correlated features in the 1D plot matched the known bioactive compounds **1** and **2**, additional signals from the mixture could not be assigned. A positively correlated doublet of doublets at δ_H 4.55 did not match any reference TT and was therefore used as a marker feature in subsequent 2D NMR-based HetCA evaluations and for a targeted isolation.

In the 2D NMR-based HetCA, a variety of positively and negatively correlated δ_H,δ_C value pairs (Table S5) were observed, with most values located at the extremes of the scale (cc 1.0, dark red and cc -1.0, dark blue). A comparison with TT reference compounds **1** and **2** showed that approximately three different groups of signals could be distinguished. Seven signals occurred in the region of alkenes, between δ_H,δ_C 5.14–5.36,125.2–128.9. The second group, present between δ_H,δ_C 2.18–4.55,41.0–78.9, was assumed to contain protons located at secondary alcohol C-atoms or tertiary protons at C-18 of the triterpene scaffold near a carboxy function. The third group, which includes all other cross-peaks, presumably contains protons from methylene groups next to hydroxyl groups, next to C-12 with its double bond, or next to a carboxy function, as well as tertiary protons (e.g., at C-9 or C-5 of the TT scaffold), and various methylene and methyl groups.

To filter out the cross-peaks with the highest positve correlation to activity in the 2D pseudospectrum, signals with a cc >0.990 were evaluated. Among this group of 14 δ_H,δ_C pairs found, the marker feature at δ_H 4.55, with a corresponding δ_C value of 69.3, appeared in this list of most promising, positively correlated signals. The cross-peak values of this selection group spanned the entire 2D pseudo-spectrum (Table S6). Hence, it was expected that these δ_H,δ_C pairs would originate from protons/carbons at different positions of the scaffold, next to different functional groups. None of the 14 features could be assigned to reference compounds.

All signals previously assigned in the 1D HetCA process of the same sample set were also found to be positively correlated with bioactivity in the 2D HetCA plot, but with graded cc values between 0.934–0.985. Again, the marker feature at δ_H,δ_C 4.55,69.3 had the highest value among them with 0.9916. The evaluation of the region covering the methyl group proton cross-peaks, contained several positively correlated signals with high cc >0.97, such as δ_H,δ_C 0.86,21.9; 0.85,21.7;

0.85,19.7; 0.84,19.7; 0.79,17.0; 0.75,17.0; and 0.70,12.1. In contrast to the 1D HetCA results, the 2D plot allowed differentiation between signals with the same δ_H value, mostly present in the aliphatic region, together with a gradual assessment of the calculated bioactivity. All 14 cross-peak values from the selected group of active 2D HetCA candidates, including the main marker signal at δ_H,δ_C 4.55,69.3, showed higher signal intensity in MF33 compared to MF34 and MF35. Therefore, MF33 was selected for a targeted isolation of the underlying putative RORγ-modulating compound.

Phytochemical processing of MF33. MF33 was fractionated by flash chromatography into fractions A1–A8, which were subjected to ^{1}H NMR experiments to follow the main marker signal at δ_H 4.55. At this stage, **1** and **2** could already be separated (Figure S12). The NMR marker feature was detected in fractions A3–A5 (Figure S13). Since A3 contained the highest amount of the target compound, it was further fractionated with semi-preparative SFC into B1–B3 (Figure S14). Only in the ^{1}H NMR spectrum of B1, the signal at δ_H 4.55 could be detected (Figure S15).

1D and 2D NMR experiments and comparison with the literature[36, 37] revealed the target compound to be 2α,19α-dihydroxy-3-oxo-12-ursen-28-oic acid (**16**) ($C_{30}H_{46}O_5$, molecular weight 486.69 g/mol) with a purity of 95.8% based on ELSD. Methyl groups between δ_H 0.80–1.27 are clearly distinguishable, and the marker signal at δ_H 4.55 is assigned to C-2 (δ_C 69.3), which is bonded to a hydroxyl group (Table S7). This rare pentacyclic TT has already been described as constituent of *E. japonica*[33, 36] with anti-tumor and anti-HIV effects. [33] There have been no reports of its activity on RORγ. In addition, B2 and B3 were confirmed as 3-epimaslinic acid (**17**) and 3-epicorosolic acid (**18**) by comparison with previously reported spectral data[38, 39], with a purity of 88.9% and 94.1% based on ELSD results.

Bioactivity testing of compounds 16, 17, and 18 on RORγ. To validate the findings of the HetCA approach, compounds **16**, **17**, and **18** were tested in RORγ-Gal4 luciferase reporter gene assays. Additionally, a concentration-response curve and the IC_{50} of 2.15 µM was determined for compound **16** (Figure S16).

The results of the bioactivity studies confirmed the calculations of the 1D and 2D HetCA approach. The follow-up of the signal at δ_H,δ_C 4.55,69.3 led to the isolation of the bioactive compound **16**. Compounds **17** and **18** which do not share this marker feature showed less activity. Furthermore, the determined IC_{50} value of compound **16** in this study was lower than the RORγ activity previously reported by Zhou et al. for three main bioactive pentacyclic TTs known from *E. japonica*, namely methyl corosolate (6.51 μM), uvaol (**5**) (4.25 μM), and oleanolic acid (**2**) (8.59 μM). [35] Of the 14 most active cross-peaks derived from the 2D HetCA pseudospectrum, six belonged to compound **16** (5.36,128.9 at C-12; 4.55,69.3 at C-2; 1.70,47.1 at C-9; 1.24,24.5 at C-27; 1.19,57.8 at C-5; 1.03,28.2 at C-15) and two signals belonged to compound **18** (5.27,125.7 at C-12; 1.02,28.4 at C-23). This highlights the potential of the 2D HetCA approach for the targeted and rapid identification of bioactives in complex mixtures based on biochemometric data mining.

CONCLUSIONS

This study shows that 2D HetCA is a useful complement to the 1D HetCA approach. Although the evaluation of NMR data in 2D requires longer measurement times and more data interpretation, 2D HetCA provides reduced spectral overlap due to the second dimension (δ_C) and access to information that cannot be deduced from 1D HetCA pseudospectra. Further advantages are the finely subdivided gradation of the color coding, the possibility of removing undesirable signals from the data set in advance via manual peak picking, and the size standardization of the cross-peaks in the pseudospectrum, which gives even less intense signals a higher level of visibility.

In this study, a 2D HetCA approach was developed using an artificially mixed set of TTs. In a follow-up application example, 2D HetCA was able to pinpoint to the bioactive compound **16** in a dichloromethane *E. japonica* leaf extract (content of compound **16** in EJD 1.98%, Figure S17). This highlights the capability of the 2D HetCA approach to identify compounds that might otherwise be overlooked, even when they exhibit similar bioactivity to those present at higher concentrations. In summary, 2D HetCA provides a valuable tool in the search for bioactive compounds in a targeted, resource-efficient and unambiguous manner, to reduce the number of necessary bioactivity tests and

to fish out compounds from complex mixtures, without restrictions of the bioactivity data to specific pharmacological targets.

ASSOCIATED CONTENT

Supporting Information

Chemical information and sources of 15 triterpenes; Inverse agonist activity of selected 15 triterpenes on RORγ in RORγ-Gal4 luciferase assays; Reported activity on RORγ and TGR5 of selected 15 triterpenes; Compilation of the artificial microfractions (AMFs); ^{1}H NMR stack plot of compounds **1**, **7**, **8**, **12**, and **13**; HSQC spectra of of compounds **1**, **7**, **8**, **12**, and **13**; NMR data of AMF1–AMF4; Cutout of the superimposed ^{1}H NMR plot of compounds **1**, **7**, **8**, **12**, and **13**; Superimposed HSQC plot of compounds **1**, **7**, **8**, **12**, and **13**; AMF1–AMF4 2D HetCA pseudospectrum values; Inverse agonist activity of MF1–MF45 in RORγ-Gal4 luciferase assays; UHPSFC-ELSD results of MF33–35 and references **1** and **2**; UHPSFC-MS results of MF33–MF35 and references **1** and **2**; Stacked ^{1}H NMR plot of MF33–MF35; HSQC spectra of MF33–MF35; 2D HetCA results of MF33–MF35; 14 most active features from the MF33–MF35 2D HetCA; UHPSFC-ELSD chromatograms of MF33, A3, and compounds **1** and **2**; Overlayed ^{1}H NMR of A3–5 at δ_H 4.55; UHPSFC-ELSD results for A3, and B1–B3; Stacked plot of the ^{1}H NMR spectra of fractions B1–B3; ^{1}H NMR data of 2α,19α-dihydroxy-3-oxo-12-ursen-28-oic acid (**16**); Bioactivity testing of compounds **16**, **17**, and **18** on RORγ; UHPSFC-ELSD results for EJD integrated area under the curves; Supporting Information References.

AUTHOR INFORMATION

Corresponding Author

Ulrike Grienke – Division of Pharmacognosy, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria; orcid.org/0000-0003-0305-9270; E-mail: ulrike.grienke@univie.ac.at

Authors

Sigrid Adelsberger – Division of Pharmacognosy, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria; orcid.org/0000-0001-7310-3199

Alexander F. Perhal – Division of Pharmacognosy, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria; orcid.org/0000-0003-3922-6333

Patrik F. Schwarz – Division of Pharmacognosy, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria; orcid.org/0000-0002-3093-4176

Lorenza Bertaina – Division of Pharmacognosy, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria; orcid.org/0009-0000-7712-2768

Verena M. Dirsch – Division of Pharmacognosy, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria; orcid.org/0000-0002-9261-5293

Judith M. Rollinger – Division of Pharmacognosy, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria; orcid.org/0000-0001-6581-0774

Funding

This work was funded by the Austrian Science Fund (FWF); Grant-DOI: 10.55776/P35115.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Tobias Madl (Otto Loewi Research Center, Medicinal Chemistry, Research Unit Integrative Structural Biology, Medical University of Graz, Austria) for his help in 2D NMR data processing and Pia Fischer-Windsteig, Marcel Gühner, Zana Hoti, Martin M. Kraus, Leopold Neuhaus, and Zuzana Vaneková (Division of Pharmacognosy, Department of Pharmaceutical Sciences, University of Vienna, Austria) for technical assistance, and Nicolas Coudevylle and Jiri Mares (Division of Pharmaceutical Chemistry, Department of Pharmaceutical Sciences, University of Vienna, Austria) for assistance with NMR measurements.

REFERENCES

(1) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J Nat Prod* **2020**, *83* (3), 770-803. DOI: 10.1021/acs.jnatprod.9b01285 From NLM Medline.

(2) Wolfender, J. L.; Litaudon, M.; Touboul, D.; Queiroz, E. F. Innovative omics-based approaches for prioritisation and targeted isolation of natural products - new strategies for drug discovery. *Nat Prod Rep* **2019**, *36* (6), 855-868. DOI: 10.1039/c9np00004f From NLM Medline.

(3) Kellogg, J. J.; Todd, D. A.; Egan, J. M.; Raja, H. A.; Oberlies, N. H.; Kvalheim, O. M.; Cech, N. B. Biochemometrics for Natural Products Research: Comparison of Data Analysis Approaches and Application to Identification of Bioactive Compounds. *J Nat Prod* **2016**, *79* (2), 376-386. DOI: 10.1021/acs.jnatprod.5b01014 From NLM Medline.

(4) Britton, E. R.; Kellogg, J. J.; Kvalheim, O. M.; Cech, N. B. Biochemometrics to Identify Synergists and Additives from Botanical Medicines: A Case Study with *Hydrastis canadensis* (Goldenseal). *J Nat Prod* **2018**, *81* (3), 484-493. DOI: 10.1021/acs.jnatprod.7b00654 From NLM Medline.

(5) Langeder, J.; Döring, K.; Schmietendorf, H.; Grienke, U.; Schmidtke, M.; Rollinger, J. M. ^{1}H NMR-based biochemometric analysis of *Morus alba* extracts toward a multipotent herbal anti-infective. *J Nat Prod* **2023**, *86* (1), 8-17. DOI: 10.1021/acs.jnatprod.2c00481.

(6) Wasilewicz, A.; Areesanan, A.; Kirchweger, B.; Nicolay, S.; Waltenberger, E.; Beniddir, M. A.; Grundemann, C.; Rollinger, J. M.; Grienke, U. Combining the Strengths of MS and NMR in Biochemometrics: A Case Study on *Buddleja officinalis*. *J Nat Prod* **2024**, *in print*. DOI: 10.1021/acs.jnatprod.4c00847 From NLM Publisher.

(7) Amountzias, V.; Cheilari, A.; Vontzalidou, A.; Benaki, D.; Gikas, E.; Aligiannis, N. Can NMR-HetCA be a Reliable Prediction Tool for the Direct Identification of Bioactive Substances in Complex Mixtures? *Anal Chem* **2024**, *96* (50), 20090-20097. DOI: 10.1021/acs.analchem.4c05080 From NLM Medline.

(8) Grienke, U.; Foster, P. A.; Zwirchmayr, J.; Tahir, A.; Rollinger, J. M.; Mikros, E. 1H NMR-MS-based heterocovariance as a drug discovery tool for fishing bioactive compounds out of a complex mixture of structural analogues. *Scientific Reports* **2019**, *9* (1), 11113. DOI: 10.1038/s41598-019-47434-8.

(9) Aligiannis, N.; Halabalaki, M.; Chaita, E.; Kouloura, E.; Argyropoulou, A.; Benaki, D.; Kalpoutzakis, E.; Angelis, A.; Stathopoulou, K.; Antoniou, S.; et al. Heterocovariance Based Metabolomics as a Powerful Tool Accelerating Bioactive Natural Product Identification. *ChemistrySelect* **2016**, *1* (10), 2531-2535. DOI: 10.1002/slct.201600744.

(10) Zwirchmayr, J.; Grienke, U.; Hummelbrunner, S.; Seigner, J.; de Martin, R.; Dirsch, V. M.; Rollinger, J. M. A Biochemometric Approach for the Identification of In Vitro Anti-Inflammatory Constituents in Masterwort. *Biomolecules* **2020**, *10* (5), 679. DOI: 10.3390/biom10050679 From NLM Medline.

(11) Tahir, A.; Kampleitner, C.; Wirglauer, T.; Grienke, U.; Hoffmann, O.; Rollinger, J. M. Biochemometry-Based Discovery of Phenylpropanoids from *Azadirachta indica* Fruits as Inhibitors of In Vitro Osteoclast Formation. *Molecules* **2022**, *27* (11), 3611. DOI: 10.3390/molecules27113611 From NLM Medline.

(12) Zwirchmayr, J.; Cruz, C. D.; Grienke, U.; Tammela, P.; Rollinger, J. M. Biochemometry identifies ostruthin as pluripotent antimicrobial and anthelmintic agent from masterwort. *iScience* **2023**, *26* (9), 107523. DOI: 10.1016/j.isci.2023.107523 From NLM PubMed-not-MEDLINE.

(13) Zwirchmayr, J.; Schachner, D.; Grienke, U.; Rudzionyte, I.; de Martin, R.; Dirsch, V. M.; Rollinger, J. M. Biochemometry identifies suppressors of pro-inflammatory gene expression in *Pterocarpus santalinus* heartwood. *Phytochemistry* **2023**, *212*, 113709. DOI: 10.1016/j.phytochem.2023.113709 From NLM Medline.

(14) Dudley, J. A.; Park, S.; MacDonald, M. E.; Fetene, E.; Smith, C. A. Resolving overlapped signals with automated FitNMR analytical peak modeling. *J Magn Reson* **2020**, *318*, 106773. DOI: 10.1016/j.jmr.2020.106773 From NLM PubMed-not-MEDLINE.

(15) Korzhneva, D. M.; Ibraghimov, I. V.; Billeter, M.; Orekhov, V. Y. MUNIN: application of three-way decomposition to the analysis of heteronuclear NMR relaxation data. *J Biomol NMR* **2001**, *21* (3), 263-268. DOI: 10.1023/a:1012982830367 From NLM Medline.

(16) Egan, J. M.; van Santen, J. A.; Liu, D. Y.; Linington, R. G. Development of an NMR-Based Platform for the Direct Structural Annotation of Complex Natural Products Mixtures. *J Nat Prod* **2021**, *84* (4), 1044-1055. DOI: 10.1021/acs.jnatprod.0c01076 From NLM Medline.

(17) Cardenas, P. D.; Almeida, A.; Bak, S. Evolution of Structural Diversity of Triterpenoids. *Front Plant Sci* **2019**, *10*, 1523. DOI: 10.3389/fpls.2019.01523 From NLM PubMed-not-MEDLINE.

(18) Martelanc, M.; Vovk, I.; Simonovska, B. Separation and identification of some common isomeric plant triterpenoids by thin-layer chromatography and high-performance liquid chromatography. *J Chromatogr A* **2009**, *1216* (38), 6662-6670. DOI: 10.1016/j.chroma.2009.07.038 From NLM Medline.

(19) Vilkickyte, G.; Raudone, L. Optimization, Validation and Application of HPLC-PDA Methods for Quantification of Triterpenoids in *Vaccinium vitis-idaea* L. *Molecules* **2021**, *26* (6), 1645. DOI: 10.3390/molecules26061645 From NLM Medline.

(20) Rhourri-Frih, B.; Chaimbault, P.; Claude, B.; Lamy, C.; Andre, P.; Lafosse, M. Analysis of pentacyclic triterpenes by LC-MS. A comparative study between APCI and APPI. *J Mass Spectrom* **2009**, *44* (1), 71-80. DOI: 10.1002/jms.1472 From NLM Medline.

(21) Ladurner, A.; Schwarz, P. F.; Dirsch, V. M. Natural products as modulators of retinoic acid receptor-related orphan receptors (RORs). *Nat Prod Rep* **2021**, *38* (4), 757-781. DOI: 10.1039/d0np00047g From NLM Medline.

(22) Papadopoulos, M. G. E.; Perhal, A. F.; Medel-Lacruz, B.; Ladurner, A.; Selent, J.; Dirsch, V. M.; Kolb, P. Discovery and characterization of small-molecule TGR5 ligands with agonistic activity. *Eur J Med Chem* **2024**, *276*, 116616. DOI: 10.1016/j.ejmech.2024.116616 From NLM Medline.

(23) Zeng, J.; Li, M.; Zhao, Q.; Chen, M.; Zhao, L.; Wei, S.; Yang, H.; Zhao, Y.; Wang, A.; Shen, J.; et al. Small molecule inhibitors of RORgammat for Th17 regulation in inflammatory and autoimmune diseases. *J Pharm Anal* **2023**, *13* (6), 545-562. DOI: 10.1016/j.jpha.2023.05.009 From NLM PubMed-not-MEDLINE.

(24) Pols, T. W.; Noriega, L. G.; Nomura, M.; Auwerx, J.; Schoonjans, K. The bile acid membrane receptor TGR5 as an emerging target in metabolism and inflammation. *J Hepatol* **2011**, *54* (6), 1263-1272. DOI: 10.1016/j.jhep.2010.12.004 From NLM Medline.

(25) Duboc, H.; Tache, Y.; Hofmann, A. F. The bile acid TGR5 membrane receptor: from basic research to clinical application. *Dig Liver Dis* **2014**, *46* (4), 302-312. DOI: 10.1016/j.dld.2013.10.021 From NLM Medline.

(26) Guo, C.; Chen, W. D.; Wang, Y. D. TGR5, Not Only a Metabolic Regulator. *Front Physiol* **2016**, *7*, 646. DOI: 10.3389/fphys.2016.00646 From NLM PubMed-not-MEDLINE.

(27) Yang, H.; Luo, F.; Wei, Y.; Jiao, Y.; Qian, J.; Chen, S.; Gong, Y.; Tang, L. TGR5 protects against cholestatic liver disease via suppressing the NF-kappaB pathway and activating the Nrf2/HO-1 pathway. *Ann Transl Med* **2021**, *9* (14), 1158. DOI: 10.21037/atm-21-2631 From NLM PubMed-not-MEDLINE.

(28) Qiu, F.; McAlpine, J. B.; Lankin, D. C.; Burton, I.; Karakach, T.; Chen, S. N.; Pauli, G. F. 2D NMR barcoding and differential analysis of complex mixtures for chemical identification: the *Actaea* triterpenes. *Anal. Chem.* **2014**, *86* (8), 3964-3972. DOI: 10.1021/ac500188j.

(29) Schwarz, P. F.; Perhal, A. F.; Schoberl, L. N.; Kraus, M. M.; Kirchmair, J.; Dirsch, V. M. Identification of the Natural Steroid Sapogenin Diosgenin as a Direct Dual-Specific RORalpha/gamma Inverse Agonist. *Biomedicines* **2022**, *10* (9). DOI: 10.3390/biomedicines10092076 From NLM PubMed-not-MEDLINE.

(30) Perhal, A. F.; Schwarz, P. F.; Linder, T.; Mihovilovic, M. D.; Schnurch, M.; Dirsch, V. M. Identification and Characterization of a Leoligin-Inspired Synthetic Lignan as a TGR5 Agonist. *J Nat Prod* **2025**, *in print*. DOI: 10.1021/acs.jnatprod.5c00059 From NLM Publisher.

(31) Genet, C.; Strehle, A.; Schmidt, C.; Boudjelal, G.; Lobstein, A.; Schoonjans, K.; Souchet, M.; Auwerx, J.; Saladin, R.; Wagner, A. Structure-activity relationship study of betulinic acid, a novel and selective TGR5 agonist, and its synthetic derivatives: potential impact in diabetes. *J Med Chem* **2010**, *53* (1), 178-190. DOI: 10.1021/jm900872z From NLM Medline.

(32) Schwarz, P. F.; Perhal, A. F.; Preglej, T.; Breit, L.; Schmetterer, K. G.; Grienke, U.; Janneschütz, J.; Chen, Y.; Rochel, N.; Kirchmair, J.; et al. Primulagenin A is a potent inverse agonist of the nuclear receptor RAR-related orphan receptor gamma (RORγ). *bioRxiv preprint* **2025**. DOI: 10.1101/2025.04.01.646598.

(33) Zhu, X.; Wang, L.; Zhao, T.; Jiang, Q. Traditional uses, phytochemistry, pharmacology, and toxicity of *Eriobotrya japonica* leaves: A summary. *J Ethnopharmacol* **2022**, *298*, 115566. DOI: 10.1016/j.jep.2022.115566 From NLM Medline.

(34) Li, E. N.; Luo, J. G.; Kong, L. Y. Qualitative and quantitative determination of seven triterpene acids in *Eriobotrya japonica* Lindl. by high-performance liquid chromatography with photodiode array detection and mass spectrometry. *Phytochem Anal* **2009**, *20* (4), 338-343. DOI: 10.1002/pca.1134 From NLM Medline.

(35) Zhou, X.; Chen, H.; Wei, F.; Zhao, Q.; Su, Q.; Lei, Y.; Yin, M.; Tian, X.; Liu, Z.; Yu, B.; et al. The Inhibitory Effects of Pentacyclic Triterpenes from Loquat Leaf against Th17 Differentiation. *Immunol Invest* **2020**, *49* (6), 632-647. DOI: 10.1080/08820139.2019.1698599 From NLM Medline.

(36) Taniguchi, S.; Imayoshi, Y.; Kobayashi, E.; Takamatsu, Y.; Ito, H.; Hatano, T.; Sakagami, H.; Tokuda, H.; Nishino, H.; Sugita, D.; et al. Production of bioactive triterpenes by *Eriobotrya japonica* calli. *Phytochemistry* **2002**, *59* (3), 315-323. DOI: 10.1016/s0031-9422(01)00455-1 From NLM Medline.

(37) Xu, H. X.; Zeng, F. Q.; Wan, M.; Sim, K. Y. Anti-HIV triterpene acids from *Geum japonicum*. *J Nat Prod* **1996**, *59* (7), 643-645. DOI: 10.1021/np960165e From NLM Medline.

(38) Kumar, N. S.; Muthukuda, P. M.; Balasubramaniam, S. Triterpenes from *Euonymus Revolutus*. *Phytochemistry* **1985**, *24* (10), 2454-2455. DOI: 10.1016/S0031-9422(00)83068-X.

(39) Zhang, Q.; Lu, Z.; Li, X.; Zheng, Y.; Yao, D.; Gu, Y.; Huo, C.; Cong, B. Triterpenoids and Steroids from the Leaves of *Forsythia suspensa*. *Chemistry of Natural Compounds* **2015**, *51* (1), 178-180. DOI: 10.1007/s10600-015-1236-5.

SUPPORTING INFORMATION

Biochemometric 2D NMR-based heterocovariance analysis:
A targeted approach for identifying bioactive compounds in complex mixtures

Sigrid Adelsberger,[†,§] Alexander F. Perhal,[†] Lorenza Bertaina,[†,§] Patrik F. Schwarz,[†,§] Verena M. Dirsch,[†] Judith M. Rollinger,[†] Ulrike Grienke[,†]*

[†] Division of Pharmacognosy, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria
[§] Vienna Doctoral School of Pharmaceutical, Nutritional and Sport Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria

* Corresponding Author: Ulrike Grienke – Division of Pharmacognosy, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria; orcid.org/0000-0003-0305-9270; E-mail: ulrike.grienke@univie.ac.at

TABLE OF CONTENTS

Table S1. Chemical information and sources of 15 triterpenes.

triterpene	sum formula	molecular weight	provider	batch number	puritiy [%]
ursolic acid (**1**)	$C_{30}H_{48}O_3$	456.70	PhytoLab	11781	99.99
oleanolic acid (**2**)	$C_{30}H_{48}O_3$	456.70	Sigma Aldrich	SLBD5617V	99.60
echinocystic acid (**3**)	$C_{30}H_{48}O_4$	472.70	PhytoLab	11854	99.72
erythrodiol (**4**)	$C_{30}H_{50}O_2$	442.72	PhytoLab	10531	99.50
uvaol (**5**)	$C_{30}H_{50}O_2$	442.72	isolated in house	/	98.05
α-amyrin (**6**)	$C_{30}H_{50}O$	426.72	PhytoLab	10429	99.54
faradiol (**7**)	$C_{30}H_{50}O_2$	442.72	PhytoLab	18231	99.36
betulin (**8**)	$C_{30}H_{50}O_2$	442.72	PhytoLab	19269	99.54
lupeol (**9**)	$C_{30}H_{50}O$	426.72	Carl Roth	3991247	99.99
maslinic acid (**10**)	$C_{30}H_{48}O_4$	472.70	PhytoLab	6695	99.91
corosolic acid (**11**)	$C_{30}H_{48}O_4$	472.70	PhytoLab	7053	99.71
hederagenin (**12**)	$C_{30}H_{48}O_4$	472.70	PhytoLab	12808	99.82
bayogenin (**13**)	$C_{30}H_{48}O_5$	488.70	PhytoLab	10392	99.95
asiatic acid (**14**)	$C_{30}H_{48}O_5$	488.70	PhytoLab	6265	99.14
ganoderic acid A (**15**)	$C_{30}H_{44}O_7$	516.67	PhytoLab	9512	98.00

The purity of uvaol (**5**) was evaluated via Ultra-high performance supercritical fluid chromatography (UHPSFC) using an evaporative light scattering detector (ELSD) on a Waters Acquity UPC² device with a 1-aminoanthracene column (3.0 x 100 mm, 1.7 µm) as stationary phase at 45°C. The mobile phase consisted of supercritical carbon dioxide and methanol as a co-solvent, the flowrate was 1 mL/min. 1–2 mg of each TT were dissolved in n-hexane:isopropanol (both VWR Chemicals, Radnor, USA) 7:3. Areas under the curve were summed up to 100% and the purity was calculated as the percentage of the respective TT peak. A Waters Acquity Quadrupole Dalton (QDa) detector was used to evaluate the mass of the TTs. Methanol (VWR Chemicals, Radnor, USA) and double distilled water (Milli-Q-Plus ultrapure water device by Millipore Corporation, Bedford, MA, USA) 95:5 with 10 mM ammonium formate improved the ionization process. Waters software Empower 3 was used for data processing.

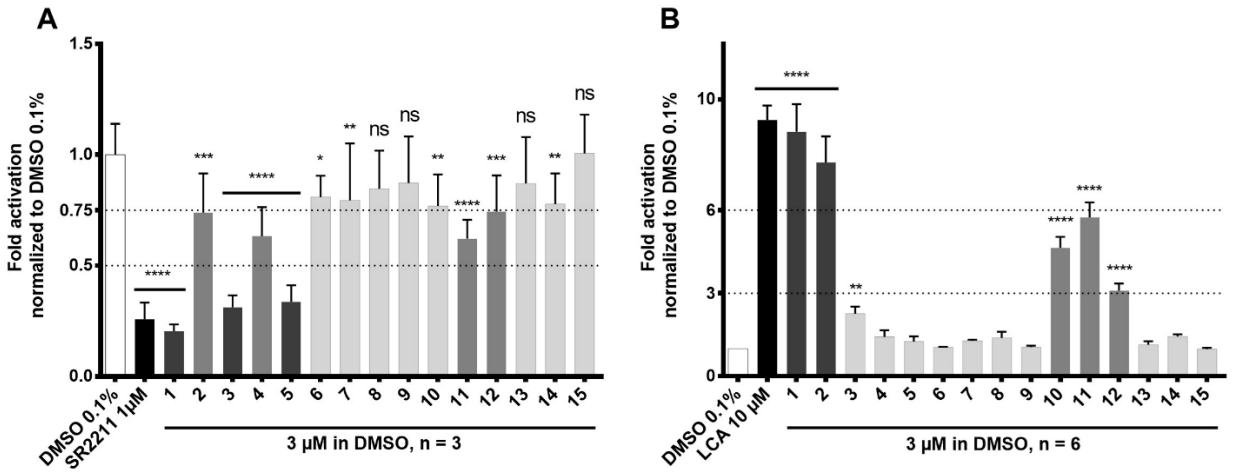


Figure S1. Inverse agonist activity of selected 15 triterpenes on ROR*γ* in ROR*γ*-Gal4 luciferase assays. The inverse agonist activity on ROR*γ* was compared to DMSO 0.1% as vehicle control, with SR2211 as positive control (A) and agonistic activity on TGR5, compared to DMSO 0.1% as vehicle control, with 10 µM lithocholic acid (LCA) as positive control in CRE-luciferase assays (B) of 15 selected TTs (3 µM). Results are expressed as fold activations, bars represent transactivation activities expressed as means ± standard deviation of n biological replicates measured in technical quadruplicates. One-way ANOVA followed by Dunnett's post hoc test were used for statistical analysis. **** $p \leq 0.0001$, *** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$ compared to vehicle control. For ROR*γ* inverse agonists, <0.5-fold activation was considered as active, 0.5–0.75-fold activation was classified as moderately active, and from >0.75-fold activation, compounds were considered inactive. For TGR5 agonists, >6-fold activation was classified as active, 3–6-fold as moderately active, and <3-fold as inactive. Bars of active compounds are marked in dark grey, of moderately active compounds in medium grey, and of inactive compounds marked in light grey.

Table S2. Reported activity on RORγ and TGR5 of selected 15 triterpenes. N/A means that no information is available. TTs such as **1**, [1] **2**, [1] **5**, [1] **11**, [2] **12**, [3] and to some extent also **14**, [2] are known as inverse agonists on ROR*γ* and therefore decrease the transcriptional activity below the basal level, [1] while **10**[2] shows no activity on the same target. Also, there are studies that confirm TTs, such as **1**, [4, 5] **2**, [4, 6] **10**, [7] and **11**, [7] to be active on TGR 5, while compounds **6**, [4] **8**, [4] **12**, [8] **13**, [8] and **14**, [4] are reported to be inactive or show weak activity.

Triterpene	Target(s)	Comment(s)
ursolic acid (**1**)	RORγt antagonist[9] / inverse agonist[1, 10]	IC_{50} = 0.68±0.1 mM co-activator binding to RORγt ligand binding domain (LBD) in TR-FRET assay[1]
	TGR5 agonist[4, 5]	EC_{50} = 1.43 µM, TGR5 luciferase assay[4]
oleanolic acid (**2**)	RORγ(t) inverse agonist[1]	EC_{50} = 8.589 µM RORγ-LBD:Gal4-DNA binding domain (DBD) luciferase reporter assay in Jurkat cells[1]
	TGR5 agonist[4, 6]	EC_{50} = 2.25 µM, TGR5 luciferase assay[4] EC_{50} = 1.42 µM, human TGR5 luciferase assay[6]
echinocystic acid (**3**)	RORγ(t) inverse agonist[11]	Best-fit IC_{50} = 2.127 µM RORγ-Gal4 luciferase assay[11]
erythrodiol (**4**)	RORγ(t) inverse agonist[11]	Best-fit IC_{50} = 0.456 µM RORγ-Gal4 luciferase assay[11]
uvaol (**5**)	RORγ(t) inverse agonist[1]	EC_{50} = 4.254 µM RORγ-LBD:Gal4-DBD luciferase reporter assay in Jurkat cells[1, 12]
α-amyrin (**6**)	TGR5 agonist[4]	EC_{50} > 10 µM TGR5 luciferase assay[4]
faradiol (**7**)	N/A	N/A
betulin (**8**)	TGR5 agonist[4]	EC_{50} > 10 µM, TGR5 luciferase assay[4]
lupeol (**9**)	N/A	N/A
maslinic acid (**10**)	no RORγ(t) activity[10]	GAL-RORγ-LBD reporter assay
	TGR5 agonist[7]	3.7 ± 0.7 µM, TGR5 luciferase reporter gene assay[7]
corosolic acid (**11**)	RORγ(t) inverse agonist[10]	EC = 2.52 µM, GAL-RORγ-LBD reporter assay[10]
	TGR5 agonist[7]	0.5 ± 1.0 µM, TGR5 luciferase reporter gene assay[7]
hederagenin (**12**)	reduced RORγ(t) activity[13]	Autoimmune encephalomyelitis (EAE) mouse model
	no TGR5 activity[8]	reporter gene-based luciferase assay in HEK 293T cells
bayogenin (**13**)	no TGR5 activity[8]	reporter gene-based luciferase assay in HEK 293T cells
asiatic acid (**14**)	RORγ(t) inverse agonist[10]	EC > 15 µM GAL-RORγ-LBD reporter assay[10]
	TGR5 agonist[4]	EC_{50} > 10 µM TGR5 luciferase assay[4]
ganoderic acid A (**15**)	N/A	N/A

Table S3. Compilation of the artificial microfractions (AMFs).

triterpene	ratios							
	AMF1	AMF2	AMF3	AMF4	AMF5	AMF6	AMF7	AMF8
ursolic acid (**1**)	3.50	1.00	0.50	0	0	0	0	0
hederagenin (**12**)	0	0	0.25	0.25	1.00	3.00	1.00	0.75
faradiol (**7**)	1.00	1.50	2.00	1.00	0.75	0	0	0
betulin (**8**)	1.50	3.50	5.00	6.00	5.00	3.50	2.00	1.50
bayogenin (**13**)	0	0	0	0	0.25	0.75	3.00	3.50

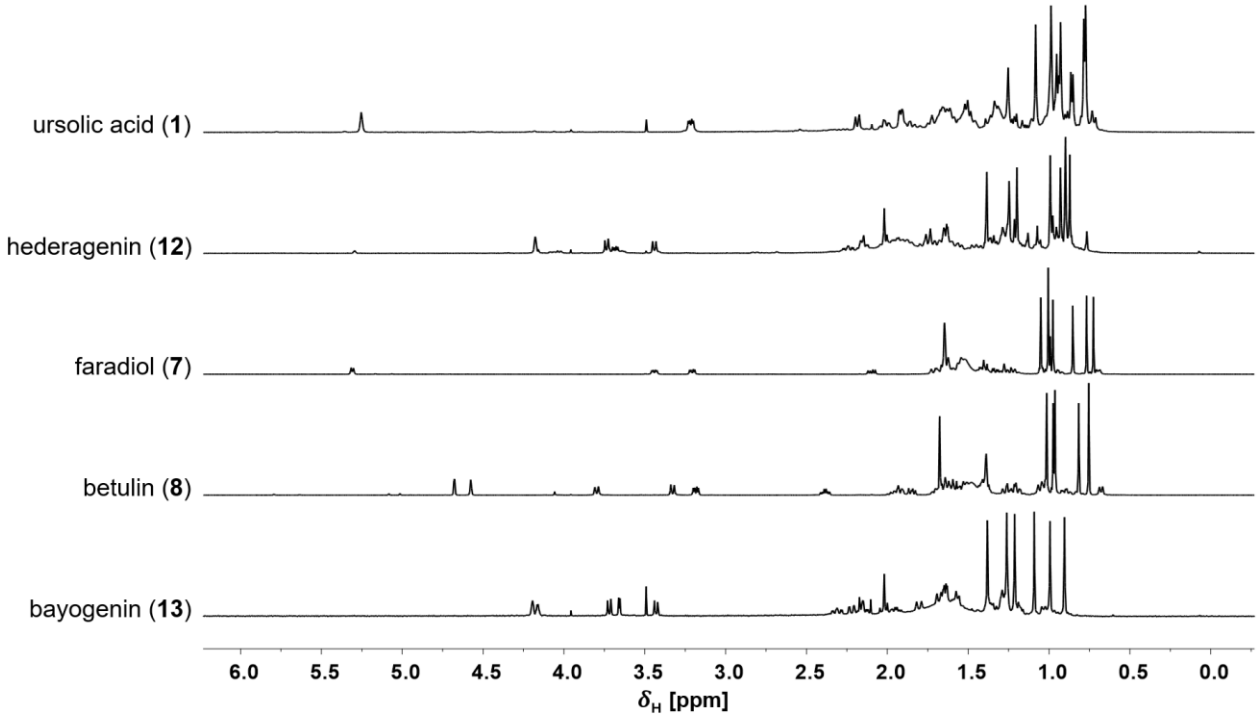


Figure S2. ^{1}H NMR stack plot of pure compounds ursolic acid (**1**), hederagenin (**12**), faradiol (**7**), betulin (**8**), and bayogenin (**13**). The measurement was performed at 500.19 MHz in deuterated chloroform-D1.

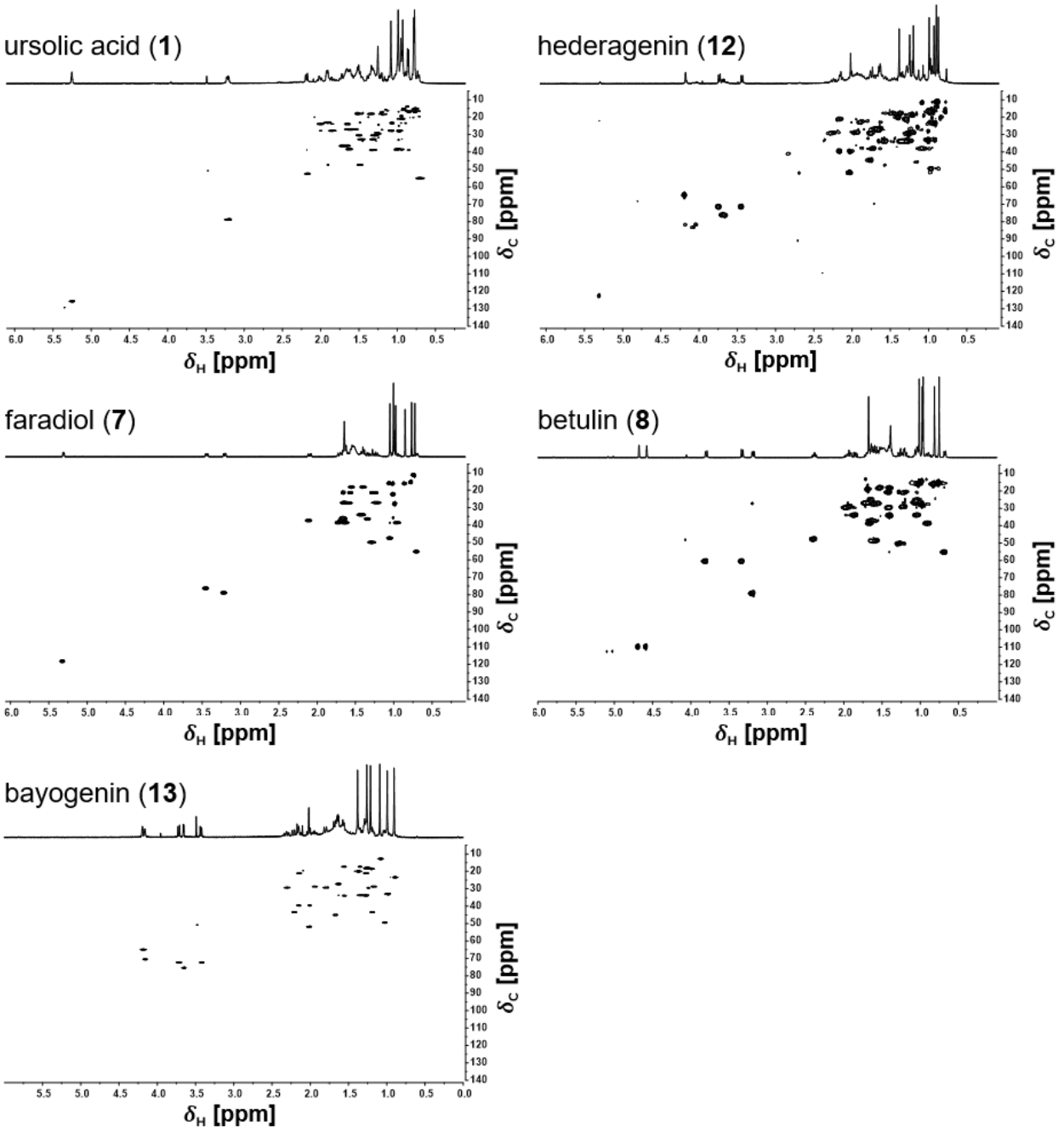


Figure S3. HSQC spectra of compounds **1**, **7**, **8**, **12**, and **13**.

HSQC measurements of the pure compounds ursolic acid (**1**), hederagenin (**12**), faradiol (**7**), betulin (**8**), and bayogenin (**13**) in deuterated chloroform-D1.

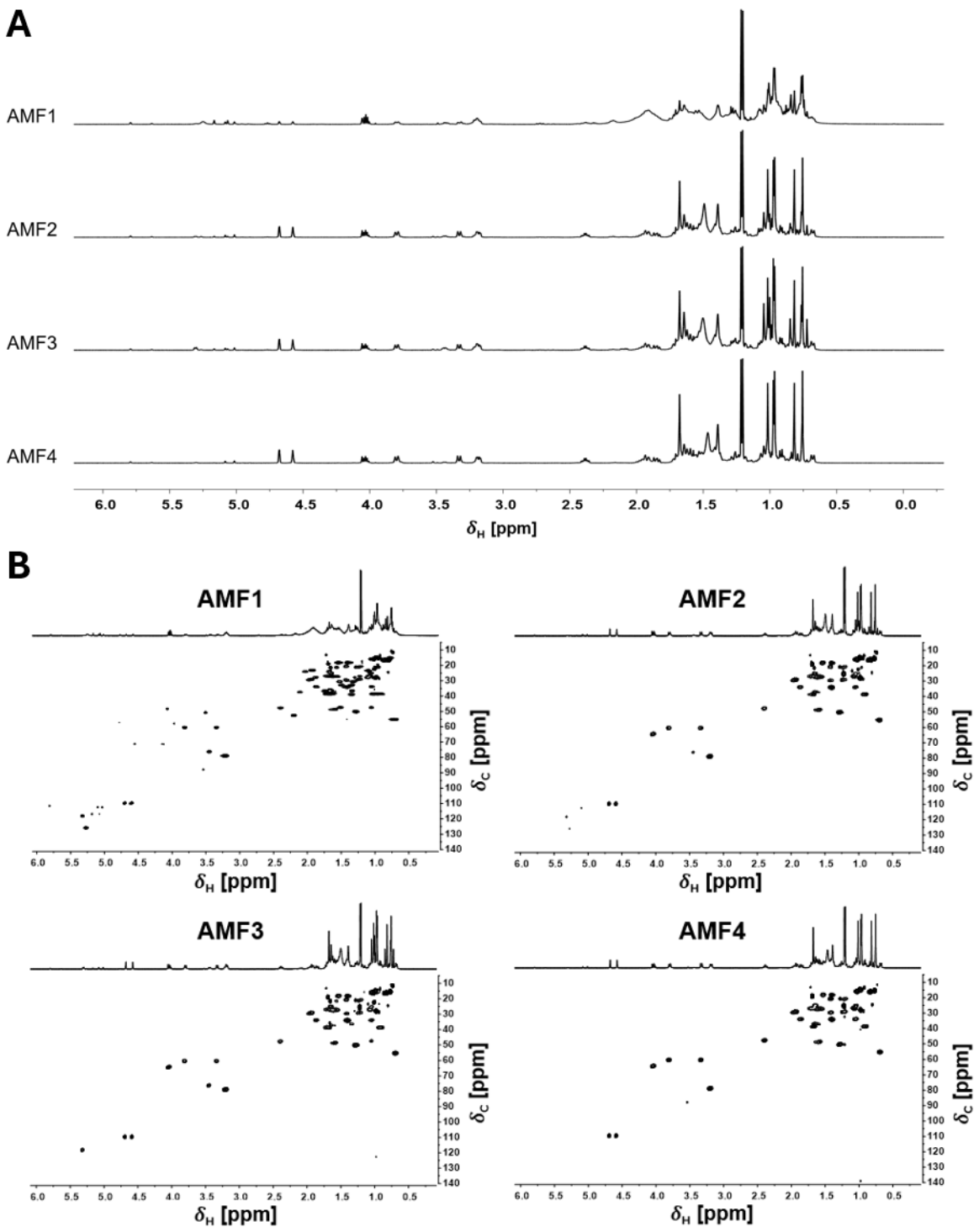


Figure S4. NMR data of AMF1–AMF4.

All samples contain varying concentrations of the pure compounds ursolic acid (**1**), hederagenin (**12**), faradiol (**7**), and betulin (**8**). (A) ^{1}H NMR stack plot of AMF1–AMF4. (B) The HSQC spectra for AMF1, AMF2, AMF3, and AMF4.

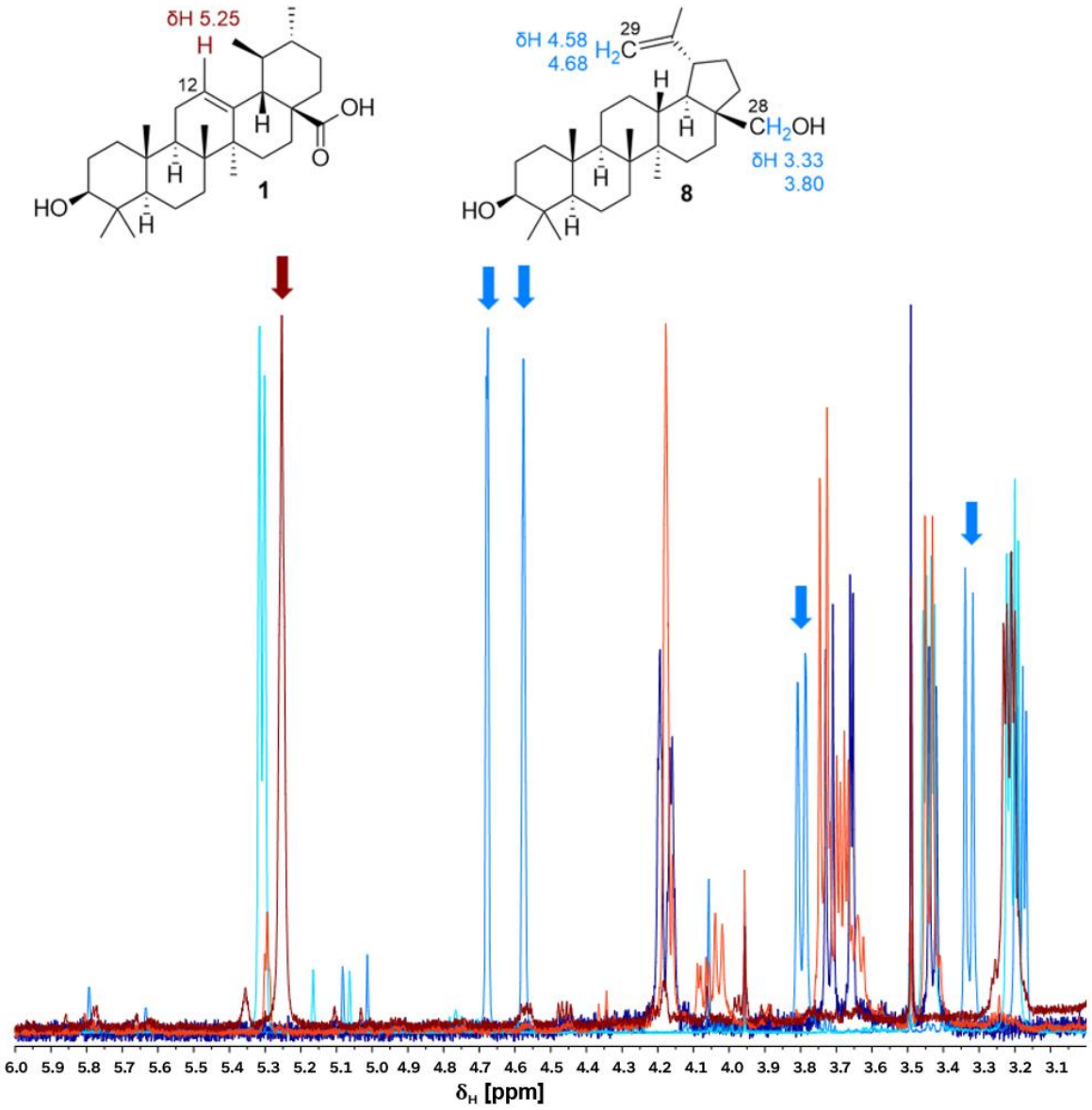


Figure S5. Cutout of the superimposed ^{1}H NMR plot of compounds **1**, **7**, **8**, **12**, and **13**.
The measurement was performed at 500.19 MHz in deuterated chloroform-D1 for ursolic acid (**1**) (dark red), faradiol (**7**) (turquoise), betulin (**8**) (light blue), hederagenin (**12**) (light red), and bayogenin (**13**) (dark blue), here shown in the δ_H region 3.0–6.0. Most of the signals are overlapped and cannot be clearly assigned to one of the pure compounds. However, one signal at δ_H 5.25 (marked with red arrow) stands out for (**1**), belonging to the proton at the double bond possessing position C-12. Additionally, four resonances at δ_H 3.33, 3.780, 4.58, and 4.68 (marked with light blue arrow) can be assigned to protons at the positions C-28 and C-29 of (**8**). For the pure compounds (**7**), (**12**), and (**13**) no significant signals could be found in this setting for clear assignment.

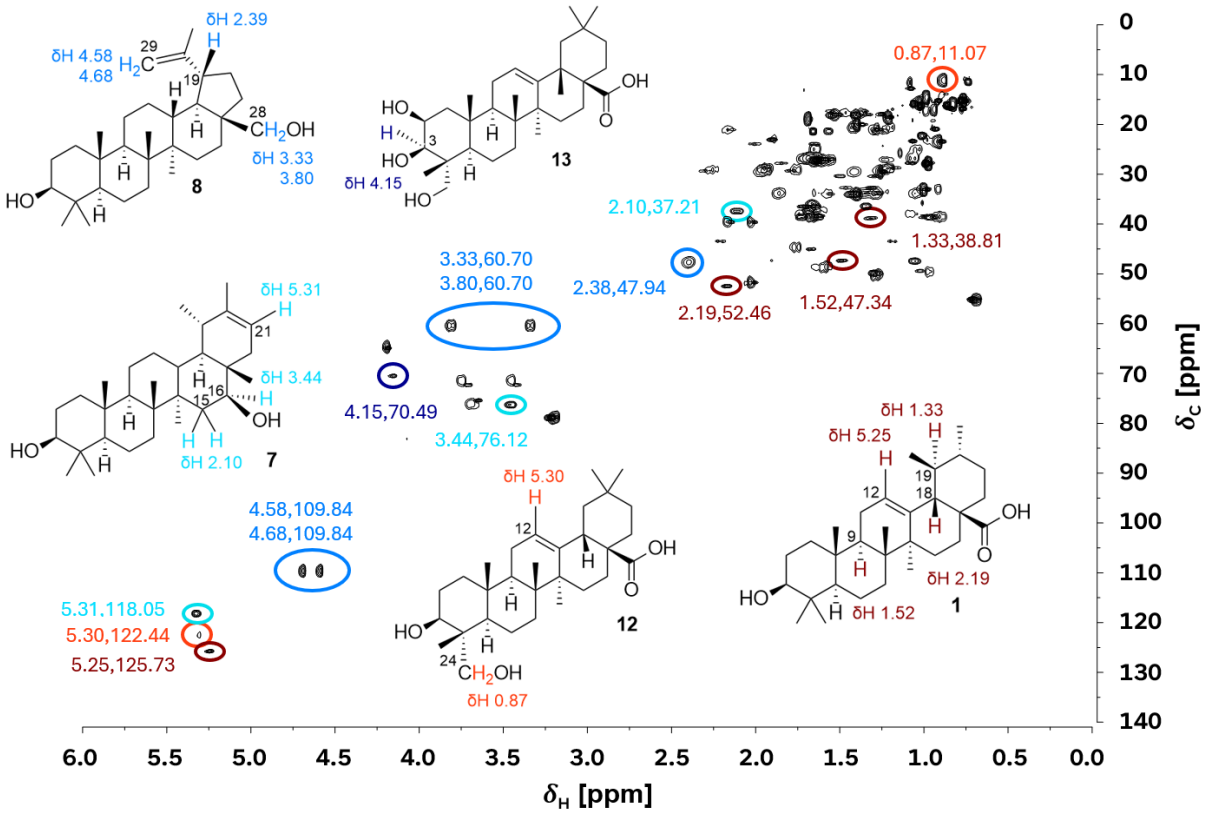
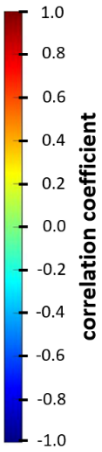


Figure S6. Superimposed HSQC plot of compounds **1**, **7**, **8**, **12**, and **13**.

HSQC measurements were performed in deuterated chloroform-D1, for ursolic acid (**1**), faradiol (**7**), betulin (**8**), hederagenin (**12**), and bayogenin (**13**). While based on the 1D NMR data, only (**1**) and (**8**) could be assigned, 2D NMR results allowed an unambiguous assignment for discriminant δ_H,δ_C pairs for all five pure compounds. Assignments are visually marked for (**1**) in dark red, for (**7**) in turquoise, for (**8**) in light blue, for (**12**) in light red, and for (**13**) in dark blue. Assignments could be made for (**1**) for protons in positions C-9, C-12, C-18, and C-19. For (**7**) for protons in positions C-15, C-16, and C-21. For (**8**) for protons in positions C-19, C-29, and the methyl protons in C-28. For (**12**) for protons in position C-12 and C-24. For (**13**) for proton in position C-3.

Table S4. AMF1–AMF4 2D HetCA pseudospectrum values. Value pairs of δ_H,δ_C for each cross-peak with the same color-code as the correlation coefficient, with assignments to compounds **1**, **7**, **8**, and **12**.

δ_H (ppm)	δ_C (ppm)	δ_H (ppm)	δ_C (ppm)	δ_H (ppm)	δ_C (ppm)	δ_H (ppm)	δ_C (ppm)
5.31	118.1 (7)	2.25	29.2 (12)	1.56	33.9 (12)	1.00	38.5 (1)
5.30	122.4 (12)	2.19	52.5 (1)	1.52	47.3 (1)	0.98	14.9 (8)
5.25	125.7 (1)	2.16	39.7 (12)	1.51	30.5 (1)	0.97	49.5 (12)
4.68	109.8 (8)	2.15	21.1 (12)	1.50	32.7 (1)	0.97	28.1 (8)
4.58	109.8 (8)	2.10	37.2 (7)	1.33	38.8 (1)	0.95	20.9 (1)
3.80	60.7 (8)	2.02	51.8 (12)	1.33	36.7 (7)	0.94	16.9 (12)
3.73	71.2 (12)	2.01	39.7 (12)	1.29	33.9 (12)	0.93	15.2 (1)
3.68	75.9 (12)	1.93	29.2 (8)	1.21	29.3 (8)	0.90	23.4 (12)
3.44	76.1 (7)	1.77	29.2 (12)	1.21	18.7 (12)	0.87	11.1 (12)
3.43	71.2 (12)	1.74	44.6 (12)	1.09	27.8 (1)	0.82	16.1 (8)
3.33	60.7 (8)	1.68	19.2 (8)	1.08	23.3 (1)	0.76	15.1 (8)
3.18	79.1 (8)	1.67	23.9 (1)	1.04	47.3 (7)	0.72	55.0 (1)
2.38	47.9 (8)	1.60	48.9 (8)	1.02	16.3 (8)		



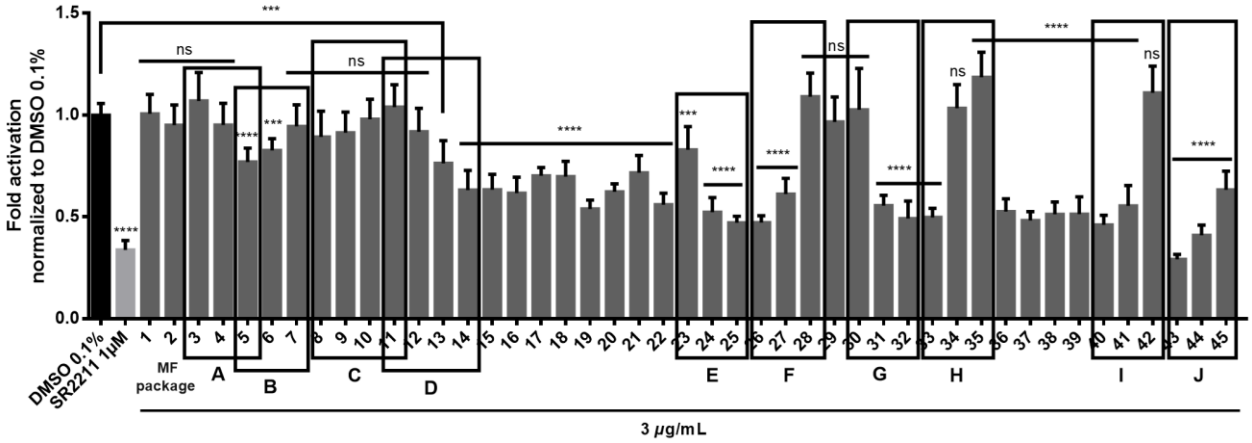


Figure S7. Inverse agonist activity of MF1–MF45 in RORγ-Gal4 luciferase assays. The sample concentrations were 3 µg/mL. Results are expressed as fold activations, bars represent transactivation activities expressed as means ± standard deviation of three biological replicates (n = 3) measured in technical quadruplicates. One-way ANOVA followed by Dunnett's post hoc test were used for statistical analysis. **** $p \leq 0.0001$, *** $p \leq 0.001$, ns $p > 0.05$ compared to vehicle control. Packages A–J were selected based on their ascending or descending activity over a minimum of three consecutive MFs.

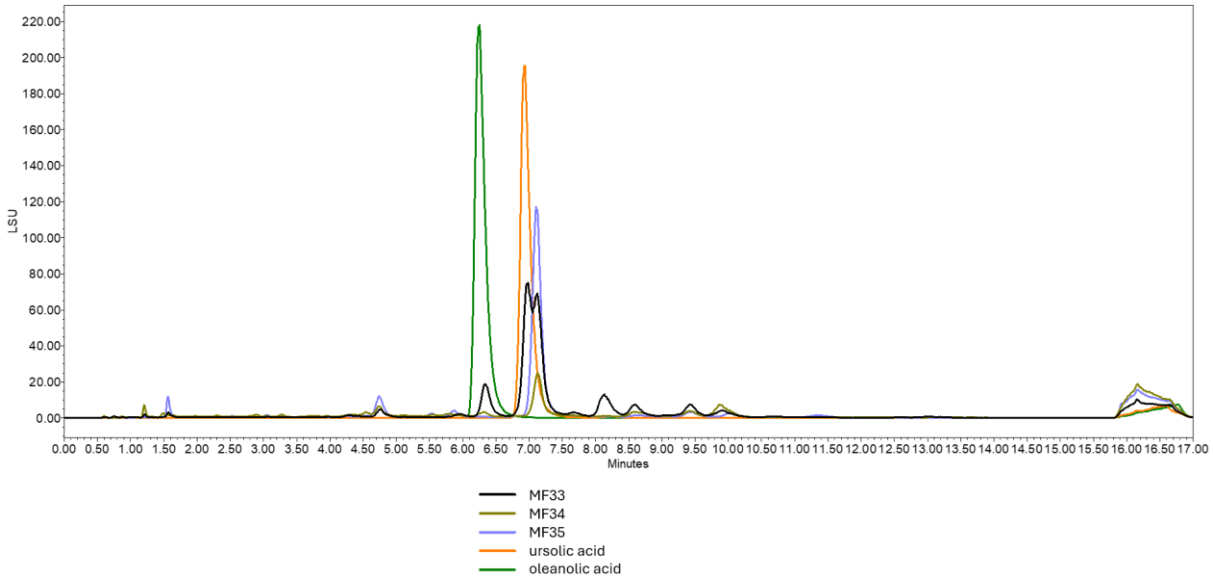


Figure S8. UHPSFC-ELSD results of MF33–35 and references **1** and **2**.

Ultra-high performance supercritical fluid chromatography (UHPSFC) results obtained via evaporative light scattering detector (ELSD) for microfractions MF33 (black), MF34 (olive), and MF35 (mauve) and reference compounds ursolic acid (**1**) (orange), oleanolic acid (**2**) (dark green).

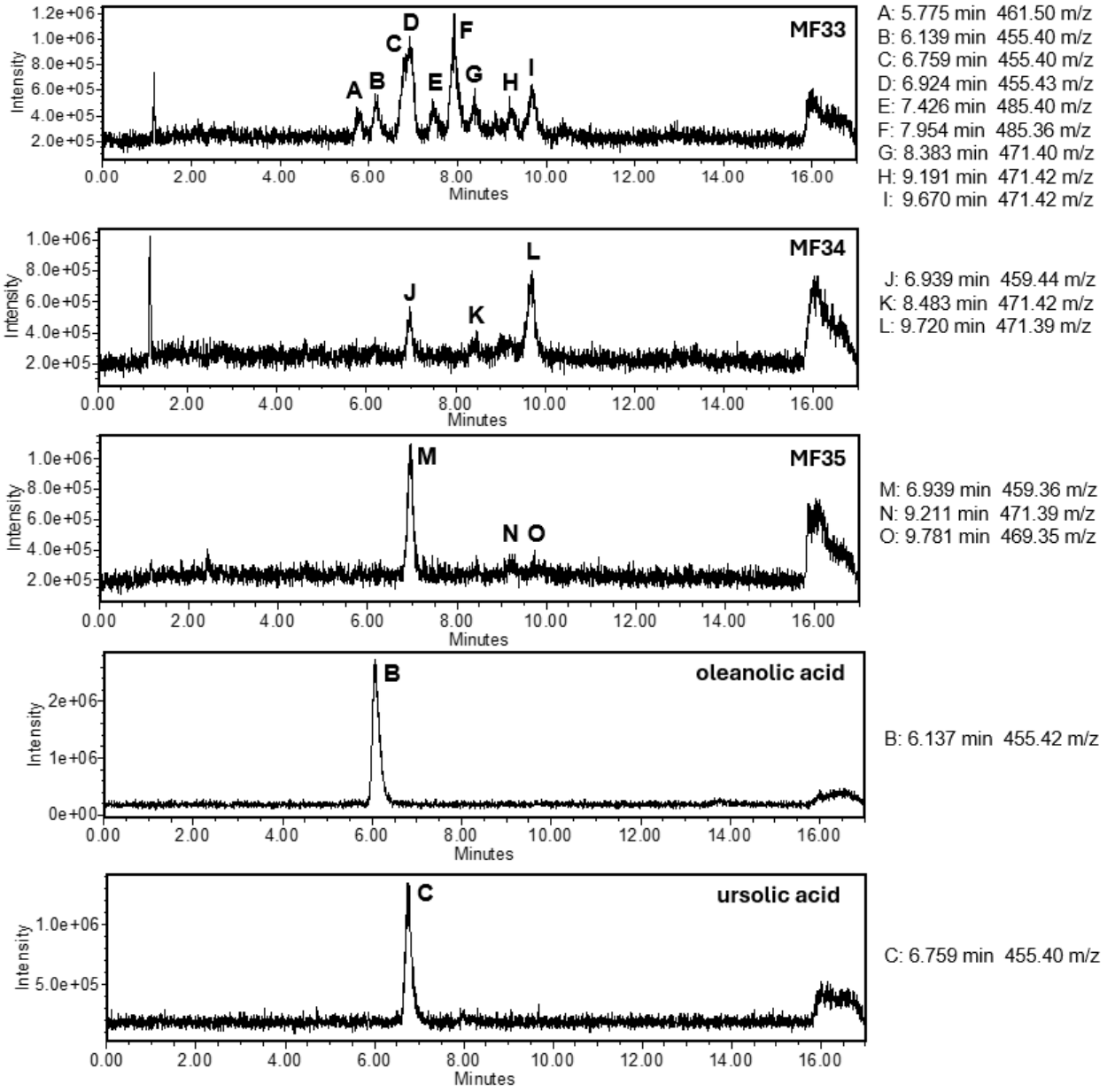


Figure S9. UHPSFC-MS results of MF33–35 and references **1** and **2**.

Ultra-high performance supercritical fluid chromatography (UHPSFC) results obtained via single quadrupole mass detector (Scan 100.00–1000.00 Da, negative ionization mode, cone voltage CV=30) for microfractions MF33, MF34, and MF35 and reference compounds ursolic acid (**1**), and oleanolic acid (**2**). Values for m/z are given for negative ionization mode.

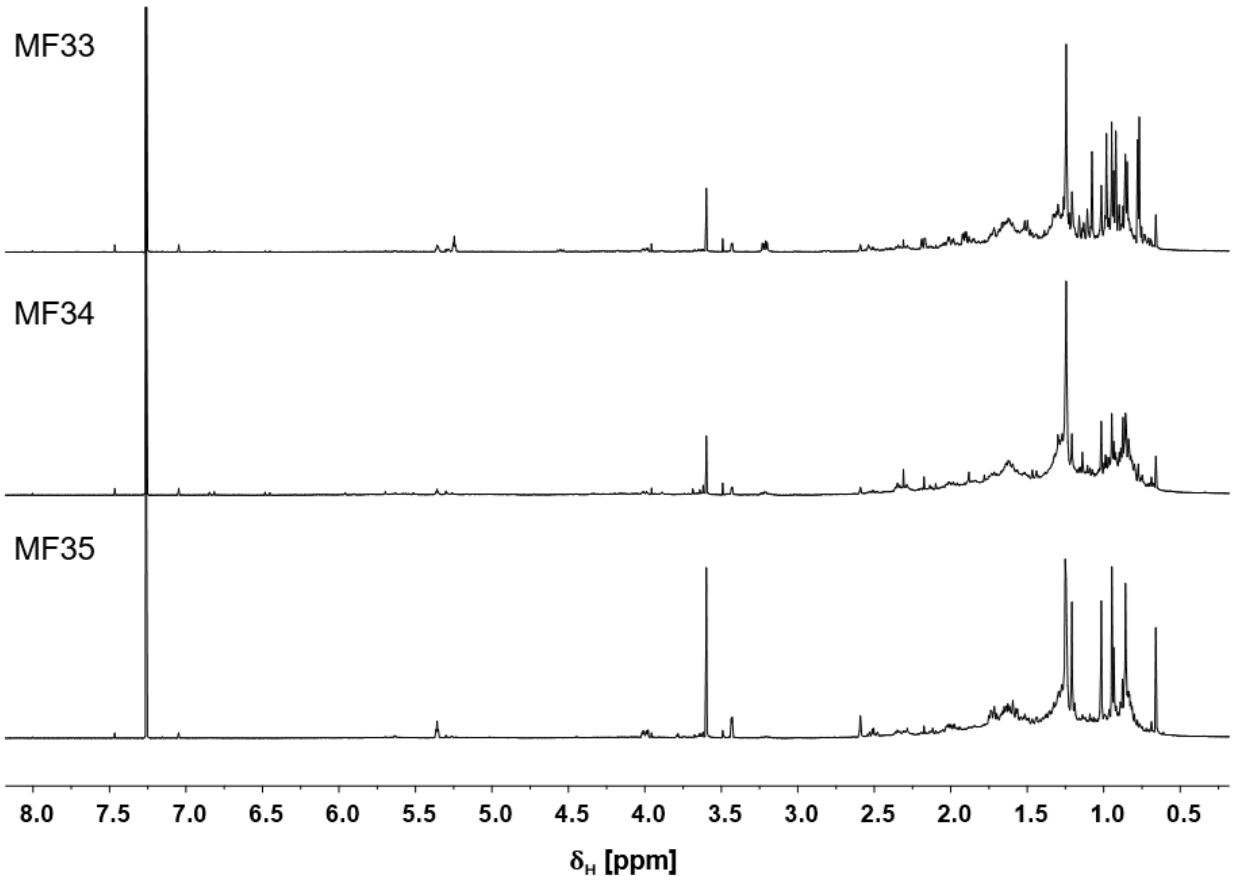


Figure S10. Stacked ^{1}H NMR plot of MF33–35.

Stacked ^{1}H NMR plot of microfractions (MFs) MF33–35 in deuterated chloroform-D1 at 500.19 MHz.

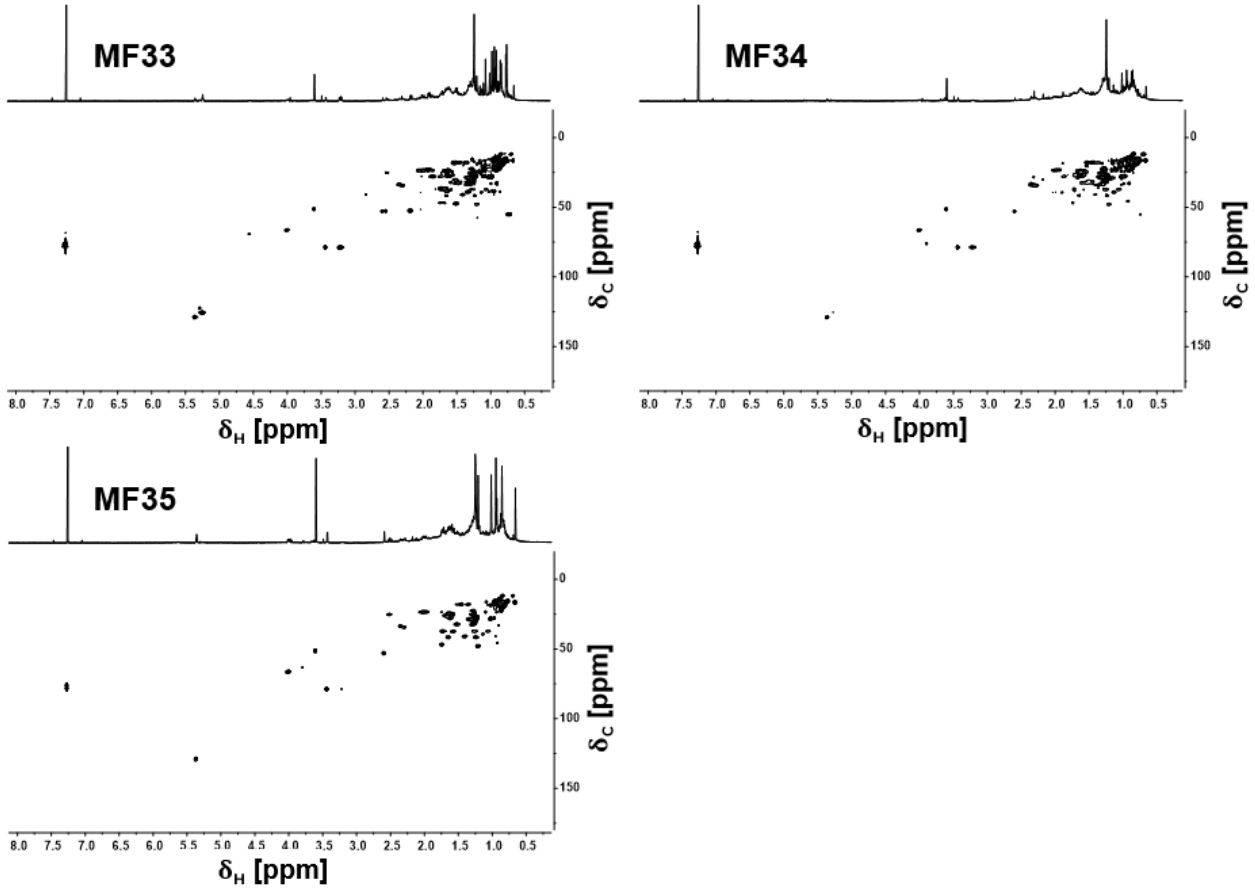


Figure S11. HSQC spectra of MF33–35.

The measurement was performed in deuterated chloroform-D1.

Table S5. 2D HetCA results of MF33–35.

The results for the microfractions (MFs) MF33–35 are given as pairs of δ_H,δ_C, together with the respective correlation coefficient (cc). The color-code of the cc, from dark blue with -1.00 to dark red with 1.00, is used for the font color of the δ_H,δ_C values.

δ_H (ppm)	δ_C (ppm)	cc	δ_H (ppm)	δ_C (ppm)	cc	δ_H (ppm)	δ_C (ppm)	cc	δ_H (ppm)	δ_C (ppm)	cc
5.36	128.9	0.995	1.70	36.9	-0.985	1.29	33.8	-0.922	1.02	18.8	0.986
5.29	122.3	0.869	1.70	23.9	-0.905	1.27	41.8	-0.922	1.01	27.5	0.737
5.28	125.7	-0.916	1.69	37.6	-0.993	1.27	22.7	-0.949	1.00	38.5	-0.879
5.28	122.7	0.957	1.69	26.0	-1.000	1.27	16.0	0.984	0.99	28.9	0.954
5.27	125.7	0.998	1.68	47.2	0.955	1.26	32.2	-0.957	0.99	27.9	0.938
5.25	125.7	0.978	1.68	41.8	-0.992	1.26	24.6	0.942	0.99	18.8	0.937
5.14	125.2	0.882	1.68	26.0	-0.979	1.25	29.5	-0.999	0.98	38.4	-0.947
4.55	69.3	0.992	1.67	36.5	-0.973	1.25	26.6	-0.931	0.98	28.0	0.957
4.00	66.4	0.955	1.67	29.3	0.906	1.24	41.8	-0.999	0.98	24.4	0.967
4.00	66.3	0.979	1.66	24.0	-0.984	1.24	37.5	0.998	0.95	41.0	0.983
3.90	76.3	0.553	1.65	38.4	-0.975	1.24	24.5	-0.967	0.95	38.8	0.923
3.79	63.2	-0.891	1.64	47.2	0.970	1.21	41.6	0.976	0.95	21.0	0.956
3.60	53.3	-0.977	1.63	41.6	-0.962	1.21	27.5	-0.978	0.95	20.9	0.955
3.60	51.5	0.973	1.62	27.5	-0.966	1.21	27.3	0.973	0.95	16.1	0.980
3.49	75.9	0.770	1.62	24.8	-0.938	1.21	17.0	0.990	0.95	13.7	0.989
3.43	78.8	0.829	1.60	38.4	-0.946	1.20	48.0	0.969	0.93	38.6	-0.985
3.22	78.9	0.934	1.60	37.7	-0.997	1.19	57.8	0.925	0.93	23.6	0.953
3.21	78.9	0.957	1.60	32.4	-0.950	1.19	48.0	0.927	0.93	21.5	0.941
2.82	41.0	0.985	1.61	25.2	-0.982	1.19	19.3	1.000	0.93	15.2	0.985
2.80	40.0	0.983	1.59	25.4	-0.995	1.18	18.0	0.995	0.92	45.9	0.986
2.60	53.1	0.925	1.58	37.5	-0.913	1.17	26.2	0.970	0.92	45.7	0.990
2.55	53.0	0.970	1.58	32.3	-0.961	1.17	24.6	0.770	0.92	23.5	0.958
2.54	25.4	-0.977	1.58	25.4	-1.000	1.17	23.0	0.874	0.92	18.7	0.932
2.41	46.7	1.000	1.56	20.7	-0.996	1.17	16.2	-0.997	0.91	15.2	0.908
2.40	49.6	-0.838	1.54	47.6	0.940	1.16	24.8	0.984	0.91	23.8	0.790
2.35	33.6	-0.909	1.53	30.6	-0.972	1.15	49.7	0.765	0.90	33.0	0.953
2.30	34.5	-0.940	1.53	32.5	-1.000	1.15	18.4	-0.938	0.89	14.2	0.965
2.19	52.5	0.985	1.53	19.3	-0.905	1.14	39.4	-0.997	0.86	39.0	1.000
2.18	52.5	0.962	1.52	47.4	0.977	1.14	26.0	0.935	0.86	21.9	0.989
2.18	30.3	0.934	1.52	18.2	-0.963	1.14	17.4	0.895	0.86	17.0	0.967
2.13	28.3	0.856	1.51	30.6	-0.967	1.11	27.7	0.975	0.86	16.9	0.967
2.07	23.9	-0.997	1.50	18.2	-0.986	1.11	23.0	-0.945	0.86	16.7	0.967
2.02	23.6	-0.993	1.48	32.8	-0.980	1.11	21.7	0.979	0.85	21.7	0.977
1.99	23.8	-0.997	1.47	25.0	-0.993	1.11	21.5	0.803	0.85	19.7	0.988
1.98	27.6	-0.358	1.47	18.0	-1.000	1.10	27.8	-0.982	0.85	16.6	0.961
1.95	23.4	-0.982	1.45	18.0	-1.000	1.10	16.6	0.999	0.85	11.9	0.963
1.92	23.0	-0.967	1.42	41.2	0.974	1.09	23.6	0.978	0.84	19.7	0.988
1.89	18.6	0.610	1.38	32.5	-0.562	1.08	27.9	0.969	0.83	19.7	0.958
1.87	27.8	-0.920	1.36	18.2	-1.000	1.08	23.3	-0.966	0.81	17.5	0.914
1.85	27.9	-0.965	1.34	33.6	-0.963	1.08	16.2	0.989	0.80	17.1	0.958
1.81	37.6	-0.962	1.34	32.7	-0.989	1.07	27.3	-0.890	0.79	18.9	-0.354
1.77	32.3	-0.935	1.33	38.8	0.970	1.07	17.0	0.992	0.79	17.0	0.977
1.75	37.3	-1.000	1.33	25.2	-0.955	1.06	37.5	-0.999	0.78	15.3	0.918
1.75	27.3	-0.990	1.33	17.9	-0.999	1.06	18.6	0.983	0.77	15.6	0.967
1.74	46.8	0.938	1.32	38.9	0.974	1.05	21.1	0.947	0.77	15.4	0.967
1.74	28.2	-0.973	1.32	33.0	-0.986	1.04	18.5	0.933	0.75	17.0	0.972
1.74	23.8	-0.994	1.30	32.3	-0.996	1.04	16.6	0.970	0.73	55.2	0.961
1.72	36.6	-0.999	1.30	29.1	-0.935	1.03	28.2	0.998	0.72	55.0	0.977
1.72	36.5	-0.995	1.30	26.0	-0.999	1.02	39.0	0.958	0.70	12.1	0.974
1.71	27.3	-0.939	1.30	18.8	0.526	1.02	28.4	0.990	0.67	16.6	0.946
1.70	47.1	1.000	1.29	37.3	-1.000	1.02	24.4	0.716			

Table S6. 14 most active features from the MF33–MF35 2D HetCA. The HSQC-based data sets were calculated for RORγ activity of MF33–35, listed as pairs of δ_H,δ_C, with the corresponding correlation coefficient (cc).

δ_H (ppm)	δ_C (ppm)	cc
2.41	46.7	1.0000
1.70	47.1	0.9999
0.86	39.0	0.9998
1.19	57.8	0.9996
1.10	16.6	0.9989
1.03	28.2	0.9984
1.24	24.5	0.9983
5.27	125.7	0.9981
1.18	18.0	0.9951
5.36	128.9	0.9950
1.07	17.0	0.9919
4.55	69.3	0.9916
1.21	17.0	0.9904
1.02	28.4	0.9903

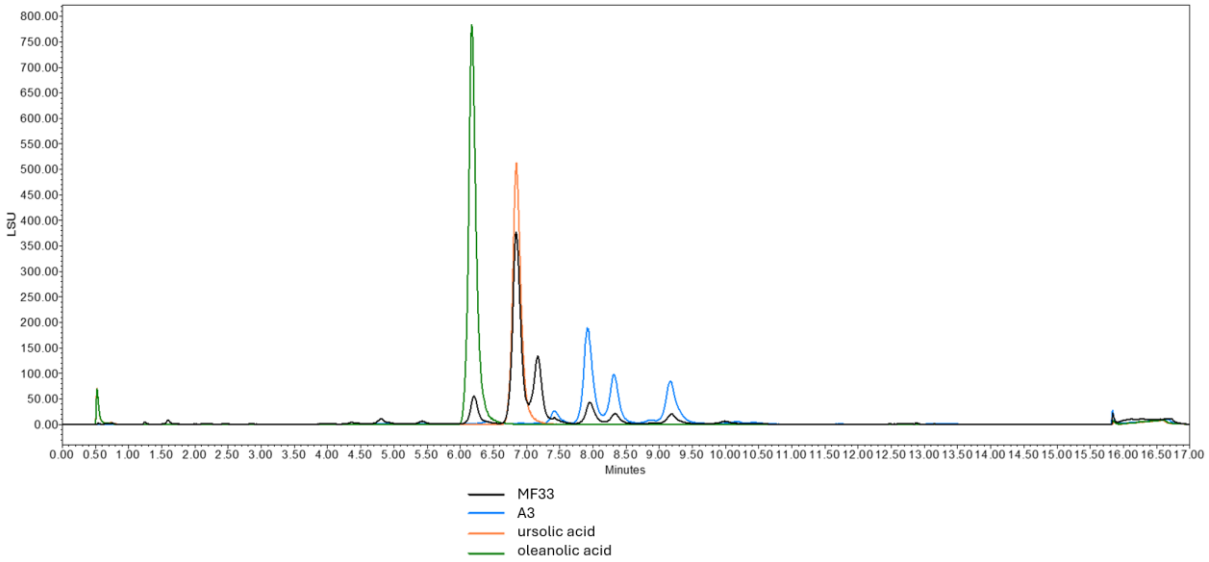


Figure S12. UHPSFC-ELSD chromatograms of MF33, A3, and compounds **1** and **2**.

Overlayed chromatograms of microfraction MF33, fraction A3, ursolic acid (**1**), and oleanolic acid (**2**) obtained via ultra-high performance supercritical fluid chromatography (UHPSFC), equipped with an evaporative light scattering detector (ELSD). While **1** (orange) and **2** (green) are present in MF33, after separation, fraction A3 does not contain both reference compounds.

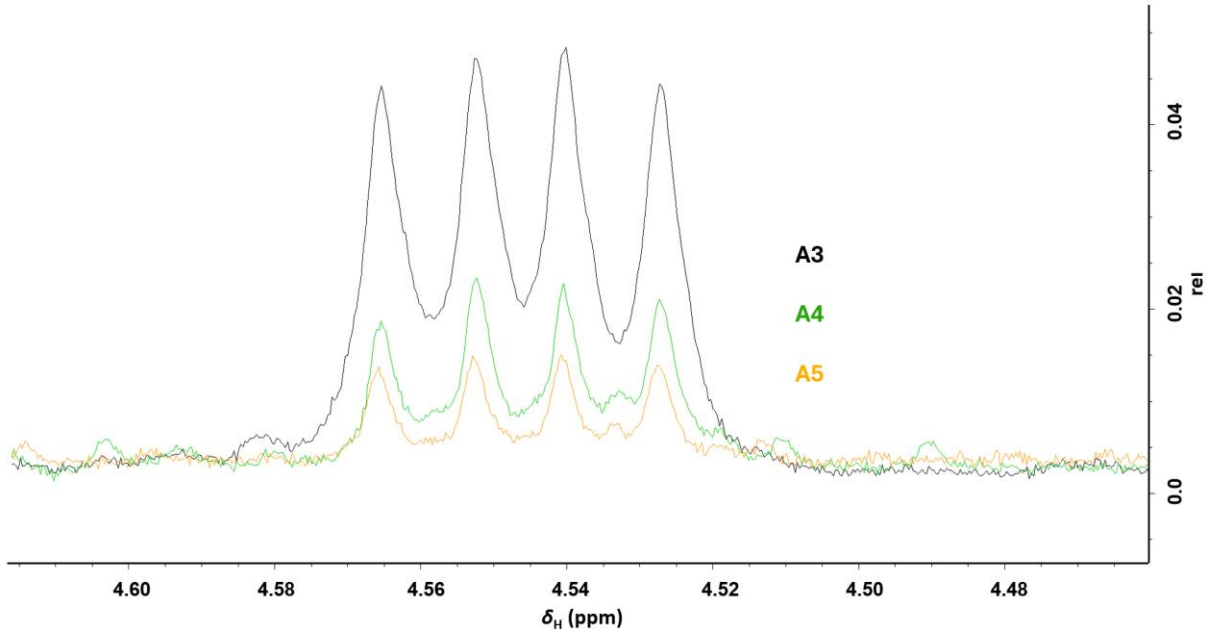


Figure S13. Overlayed ^{1}H NMR of A3–A5 at δ_H 4.55.

Proton NMR signal of the marker feature at δ_H 4.55, measured in deuterated chloroform-D1 at 500.19 MHz, with decreasing intensity from fraction A3 (black), to fraction A4 (green), to fraction A5 (orange).

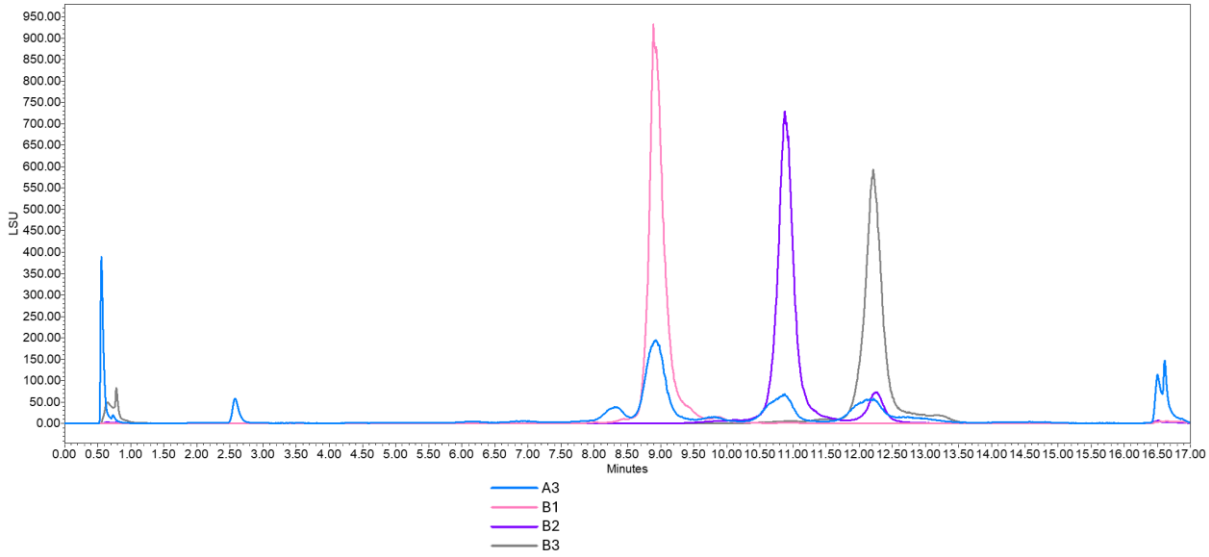


Figure S14. UHPSFC-ELSD results for A3, and B1–B3.

Ultra-high performance supercritical fluid chromatography (UHPSFC) results of an evaporative light scattering detector (ELSD) for fraction A3 (light blue) and its sub-fractions B1 (old rose), B2 (violet), and B3 (gray), obtained via semi-preparative SFC.

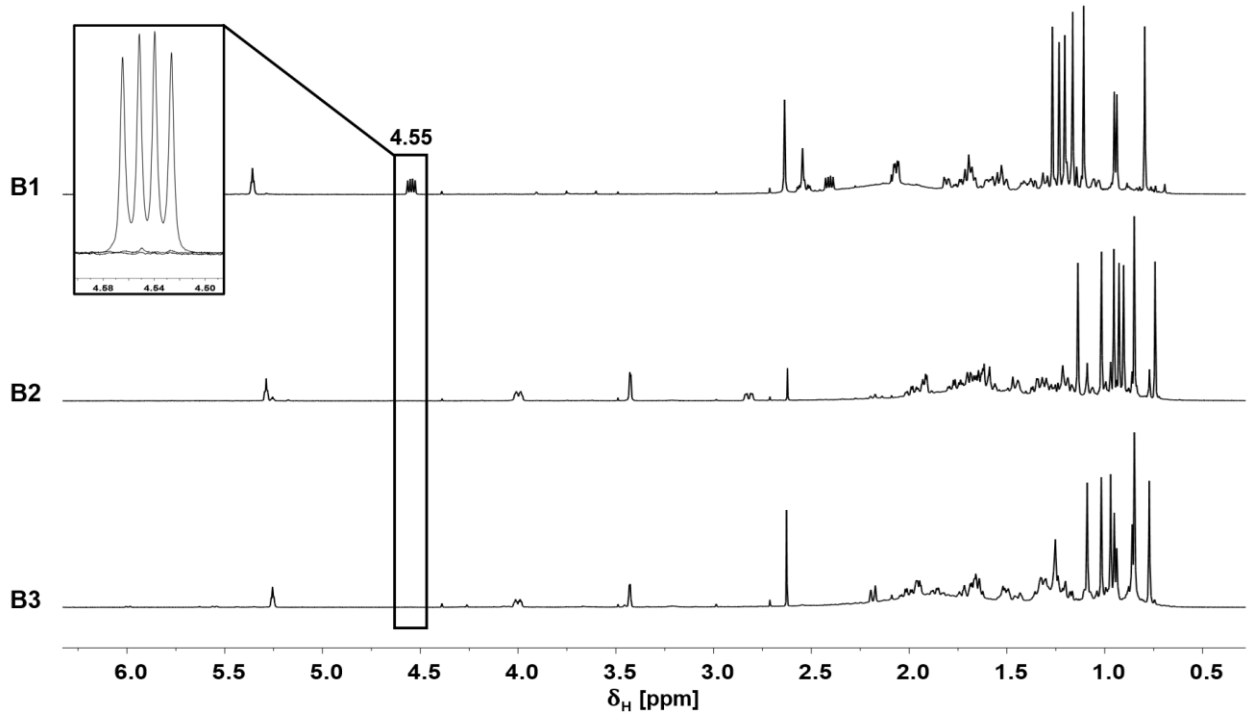


Figure S15. Stacked plot of the ^{1}H NMR spectra of fractions B1–B3.

The area of the doublet of doublets marker signal in the three spectra is marked with a black rectangle, with a zoom into the superimposed spectra at the respective δ_H region at δ_H 4.55.

Table S7. ^{1}H NMR data of 2*α*,19*α*-dihydroxy-3-oxo-12-ursen-28-oic acid (**16**).

The measurements were performed in deuterated chloroform-D1, at 500.19 MHz for ^{1}H NMR and 125.77 MHz for ^{13}C NMR.

pos	∫ H	multiplicity	δH	δC	C	δC HMBC							δH COSY
1	1.74	dd, J (6.58/12.53)	1.15, 2.40	49.6	CH2	16.0	37.8	57.8	69.3	216.8			
2	0.98	dd, J (6.57/12.60)	4.55	69.3	CH	49.6	216.8						1.15, 2.40
3				216.8	C								
4				47.8	C								
5		overlapped	1.19	57.8	CH								
6		overlapped	1.19, 1.53	19.3	CH2								
7		overlapped	1.38, 1.53	32.5	CH2								
8				41.4	C								
9		overlapped	1.70	47.1	CH								
10				37.8	C								
11	2.98	dd, J (3.41/8.97)	1.70, 2.07	23.9	CH2	40.2	47.1	128.9	138.3				
12	1.00	t, J (3.47)	5.36	128.9	CH	23.9	41.4	47.1	53.0	73.2			1.70, 2.07
13				138.3	C								
14				40.2	C								
15		overlapped	1.03, 1.74	28.2	CH2								
16		overlapped	1.59, 2.54	25.4	CH2								
17				47.9	C								
18		overlapped	2.55	53.0	CH	25.4	41.2	47.9	73.2	128.9	138.3	182.4	1.05, 1.74, 1.59
19				73.2	C								
20		overlapped	1.42	41.2	CH								
21		overlapped	1.30, 1.69	26.0	CH2								
22		overlapped	1.69, 1.81	37.6	CH2								
23	2.98	s	1.16	24.8	CH3	21.7	47.8	57.8	216.8				
24	2.95	s	1.11	21.7	CH3	24.8	47.8	57.8	216.8				
25	2.91	s	1.27	16.0	CH3	37.8	47.1	49.6	57.8				
26	2.76	s	0.80	17.1	CH3	32.5	40.2	41.4	47.1				
27	3.02	s	1.24	24.5	CH3	28.2	40.2	41.4	138.3				
28				182.4	C								
29	2.79	s	1.21	27.5	CH3	41.2	53.0	73.2					
30	3.38	d, J (6.70)	0.95	16.1	CH3	26.0	41.2	73.2					

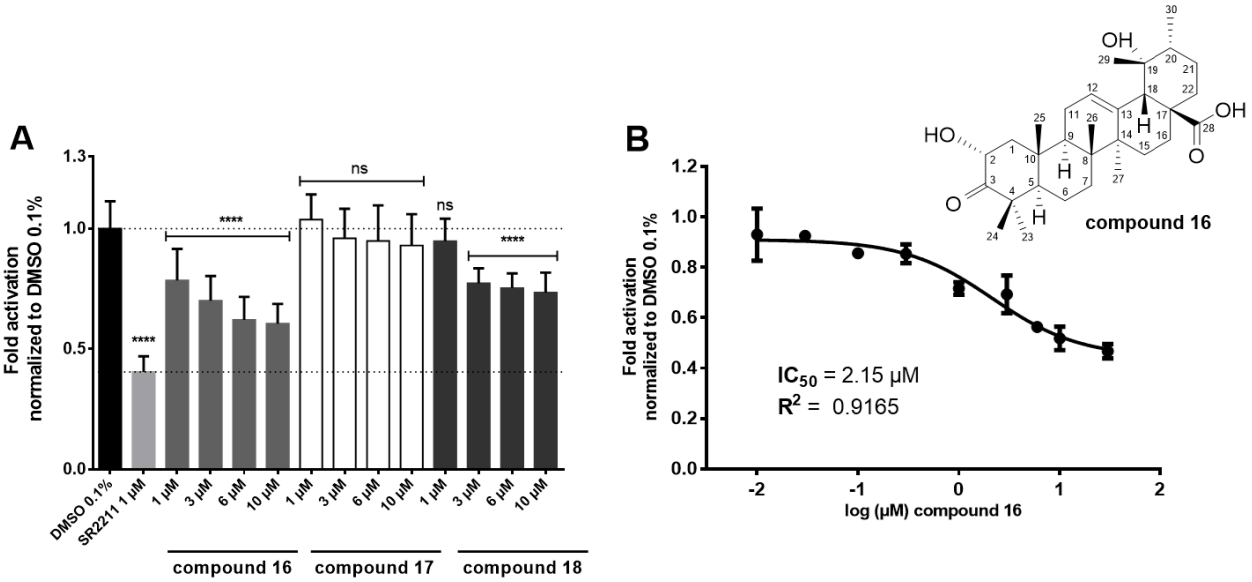
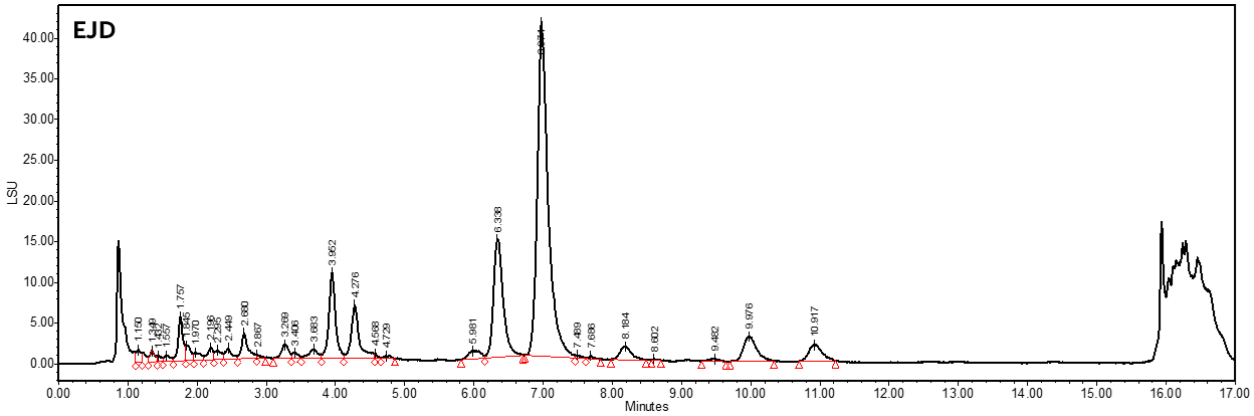


Figure S16. Bioactivity testing of compounds **16**, **17**, and **18** on RORγ. (A) RORγ inverse agonistic activities of 2α,19α-dihydroxy-3-oxo-12-ursen-28-oic acid (**16**), 3-epimaslinic acid (**17**), and 3-epicorosolic acid (**18**) in RORγ-Gal4 luciferase reporter gene assays. Bars (A) and data points (B) represent the mean ± standard deviation of three biological replicates (n = 3), measured in technical quadruplicates. One-way ANOVA followed by Dunnett's post hoc test were used for statistical analysis. **** $p \leq 0.0001$, *** $p \leq 0.001$, ns $p > 0.05$ compared to vehicle control. (B) Concentration-response curve for the RORγ inverse agonistic activity of compound **16**. The IC_{50} value of **16** was determined using nonlinear regression.

S-24



retention time [min]	area [µV*sec]	area [%]		retention time [min]	area [µV*sec]	area [%]
1.150	7703	0.80		3.952	70303	7.33
1.349	7093	0.74		4.276	53151	5.54
1.432	2763	0.29		4.568	1665	0.17
1.557	4628	0.48		4.729	2234	0.23
1.757	26841	2.80		5.981	12646	1.32
1.845	9365	0.98		6.338	141280	14.73
1.970	6544	0.68		6.974	443721	46.28
2.196	8209	0.86		7.489	2175	0.23
2.295	7734	0.81		7.686	1670	0.17
2.449	8905	0.93		8.184	18970	1.98
2.680	20177	2.10		8.602	467	0.05
2.867	1548	0.16		9.482	3768	0.39
3.269	12613	1.32		9.976	40839	4.26
3.406	4153	0.43		10.917	26630	2.78
3.683	11061	1.15				

Figure S17. UHPSFC-ELSD results for EJD with integrated area under the curves.

Ultra-high performance supercritical fluid chromatography (UHPSFC) results of an evaporative light scattering detector (ELSD) for *Eriobotrya japonica* dichloromethane crude extract (EJD), with integrated area under the curves. Proportions of the signals are given in the table included. EJD contained 1.98 % of compound **16** (at min 8.184), 0.05 % of compound **17** (at min 8.602), and 0.39 % of compound **18** (at min 9.482).

Supporting Information References

(1) Ladurner, A.; Schwarz, P. F.; Dirsch, V. M. Natural products as modulators of retinoic acid receptor-related orphan receptors (RORs). *Nat Prod Rep* **2021**, *38* (4), 757-781. DOI: 10.1039/d0np00047g From NLM Medline.

(2) Pastwinska, J.; Karas, K.; Salkowska, A.; Karwaciak, I.; Chalaskiewicz, K.; Wojtczak, B. A.; Bachorz, R. A.; Ratajewski, M. Identification of Corosolic and Oleanolic Acids as Molecules Antagonizing the Human RORgammaT Nuclear Receptor Using the Calculated Fingerprints of the Molecular Similarity. *Int J Mol Sci* **2022**, *23* (3), 1906. DOI: 10.3390/ijms23031906 From NLM Medline.

(3) Guan, D.; Li, Y.; Zhao, X.; Wang, K.; Guo, Y.; Dong, N.; Cui, Y.; Gao, Y.; Wang, M.; Wang, J.; et al. Hederagenol improves multiple sclerosis by modulating Th17 cell differentiation. *IUBMB Life* **2024**, *76* (10), 845-857. DOI: 10.1002/iub.2863 From NLM Medline.

(4) Genet, C.; Strehle, A.; Schmidt, C.; Boudjelal, G.; Lobstein, A.; Schoonjans, K.; Souchet, M.; Auwerx, J.; Saladin, R.; Wagner, A. Structure-activity relationship study of betulinic acid, a novel and selective TGR5 agonist, and its synthetic derivatives: potential impact in diabetes. *J Med Chem* **2010**, *53* (1), 178-190. DOI: 10.1021/jm900872z From NLM Medline.

(5) Lo, S. H.; Li, Y.; Cheng, K. C.; Niu, C. S.; Cheng, J. T.; Niu, H. S. Ursolic acid activates the TGR5 receptor to enhance GLP-1 secretion in type 1-like diabetic rats. *Naunyn Schmiedebergs Arch Pharmacol* **2017**, *390* (11), 1097-1104. DOI: 10.1007/s00210-017-1409-9 From NLM Medline.

(6) Sato, H.; Genet, C.; Strehle, A.; Thomas, C.; Lobstein, A.; Wagner, A.; Mioskowski, C.; Auwerx, J.; Saladin, R. Anti-hyperglycemic activity of a TGR5 agonist isolated from Olea europaea. *Biochem Biophys Res Commun* **2007**, *362* (4), S-26 793-798. DOI: 10.1016/j.bbrc.2007.06.130 From NLM Medline.

(7) Ladurner, A.; Zehl, M.; Grienke, U.; Hofstadler, C.; Faur, N.; Pereira, F. C.; Berry, D.; Dirsch, V. M.; Rollinger, J. M. Allspice and Clove As Source of Triterpene Acids Activating the G Protein-

Coupled Bile Acid Receptor TGR5. *Front Pharmacol* **2017**, *8*, 468. DOI: 10.3389/fphar.2017.00468 From NLM PubMed-not-MEDLINE.

(8) Kirchweger, B.; Kratz, J. M.; Ladurner, A.; Grienke, U.; Langer, T.; Dirsch, V. M.; Rollinger, J. M. In Silico Workflow for the Discovery of Natural Products Activating the G Protein-Coupled Bile Acid Receptor 1. *Front Chem* **2018**, *6*, 242. DOI: 10.3389/fchem.2018.00242 From NLM PubMed-not-MEDLINE.

(9) Xu, T.; Wang, X.; Zhong, B.; Nurieva, R. I.; Ding, S.; Dong, C. Ursolic acid suppresses interleukin-17 (IL-17) production by selectively antagonizing the function of RORgamma t protein. *J Biol Chem* **2011**, *286* (26), 22707-22710. DOI: 10.1074/jbc.C111.250407 From NLM Medline.

(10) Pastwinska, J.; Karas, K.; Salkowska, A.; Karwaciak, I.; Chalaskiewicz, K.; Wojtczak, B. A.; Bachorz, R. A.; Ratajewski, M. Identification of Corosolic and Oleanolic Acids as Molecules Antagonizing the Human RORgammaT Nuclear Receptor Using the Calculated Fingerprints of the Molecular Similarity. *Int J Mol Sci* **2022**, *23* (3). DOI: 10.3390/ijms23031906 From NLM Medline.

(11) Schwarz, P. F.; Perhal, A. F.; Preglej, T.; Breit, L.; Schmetterer, K. G.; Grienke, U.; Janneschütz, J.; Chen, Y.; Rochel, N.; Kirchmair, J.; et al. Primulagenin A is a potent inverse agonist of the nuclear receptor RAR-related orphan receptor gamma (RORγ). *bioRxiv preprint* **2025**. DOI: 10.1101/2025.04.01.646598.

(12) Zhou, X.; Chen, H.; Wei, F.; Zhao, Q.; Su, Q.; Lei, Y.; Yin, M.; Tian, X.; Liu, Z.; Yu, B.; et al. The Inhibitory Effects of Pentacyclic Triterpenes from Loquat Leaf against Th17 Differentiation. *Immunol Invest* **2020**, *49* (6), 632-647. S-27 DOI: 10.1080/08820139.2019.1698599 From NLM Medline.

(13) Guan, D.; Li, Y.; Zhao, X.; Wang, K.; Guo, Y.; Dong, N.; Cui, Y.; Gao, Y.; Wang, M.; Wang, J.; et al. Hederagenol improves multiple sclerosis by modulating Th17 cell differentiation. *IUBMB Life* **2024**. DOI: 10.1002/iub.2863 From NLM Publisher.

7.2.1.2 Review of Manuscript I

As demonstrated in **Manuscript I**, 2D HetCA based on 2D NMR data provides access to pseudospectral information that cannot be obtained using the 1D NMR-based method with ^{1}H spectra, particularly in regions with a high density of overlapping signals. Since the chemical shift range for ^{13}C in a ^{1}H-^{13}C HSQC spectrum is relatively broad, spectral overlap is reduced, and the additional information from the second dimension (δ_C) enhances the significance of the sum of individual signals as a biochemical fingerprint.[187] The HSQC-based 2D HetCA approach not only leverages the advantages of 2D NMR experiments, but also statistically integrates bioactivity data from the corresponding sample sets.

As also demonstrated in **Manuscript I**, a key advantage in the search for bioactive entities is that the HetCA approach relies neither on the identification of all compounds in the tested sample nor on the type of pharmacological target used in the bioactivity assays. The biochemometric method can correlate activity levels with individual chemical features, regardless of their assignment to specific compounds, thereby highlighting the most promising candidate resonances in an unbiased manner and without prior identification of the respective compound, even across different types of target molecules.

A general drawback of NMR experiments is their low sensitivity, with LOD typically in the low µM range.[188] This parameter depends on instrumentation factors such as magnetic field strength, solvent suppression techniques, the use of cryoprobes, and available measurement time, which is associated with the number of scans and the signal-to-noise ratio.[187-189] Since the sensitivity of HetCA is inherently linked to that of the NMR experiment, some signals may not be statistically correlated with bioactivity, even if they originate from bioactive compounds, because their intensities fall below the LOD and are therefore excluded from the HetCA calculations.[152]

A recently published study by Amountzias et al. evaluated various factors that contribute to false-positive and false-negative identification of bioactive compounds in 1D HetCA.[152] In addition to challenges common to all phytochemical approaches (e.g., limited solubility, chemical shift variations in NMR spectra caused by acidic or basic properties in the absence of buffering conditions, precipitation during sample preparation, and synergistic activity effects), the researchers described specific hurdles related to HetCA.[152]

For instance, if the concentration variance of an inactive compound coincidentally mirrors that of an active one, false-positive results may occur.[152] Moreover, the broad distribution of a compound across several consecutive MFs can result in concentrations below the threshold required to elicit a measurable effect in bioactivity assays.[152] In such cases, bioactive compounds might go undetected.[152] Nevertheless, Amountzias et al. also highlighted the possibility of false-positive results for compounds consistently present across multiple MFs.[152] All these issues should be considered when applying 2D HetCA approaches.

The signal intensity in an NMR spectrum is directly proportional to the number of nuclei responsible for a given resonance[188] and ^{1}H NMR is widely applied in quantitative approaches within the field of NP chemistry.[190] For 2D NMR methods such as HSQC, however, application in quantitative studies is hindered, not only due to long experiment durations.[191] The multipulse nature of this technique leads to a peak-dependent nuclei response and impacts the quantitative information of HSQC, since the proportionality coefficient between peak volume and concentration differs for each signal.[191-193] Nevertheless, several approaches have been developed and established to overcome this hurdle and to measure accurate analyte concentrations using quantitative HSQC (Q-HSQC).[191-198] As an outlook, the quantitative interpretability of 2D HetCA could be enhanced by using Q-HSQC-derived datasets for heterocovariance-based calculations.

Furthermore, integrating MS-based datasets could represent a promising extension of the new 2D HetCA approach, as already demonstrated by Wasilewicz et al. in the context of ^{1}H NMR-based HetCA.[130]

Additionally, the 2D HetCA method could be expanded by linking it to natural compound databases, which may be particularly advantageous for structural classes with numerous similar congeners. Automatic linking with databases such as Complex Mixture Analysis by NMR (COLMAR)[173] or Small Molecule Accurate Recognition Technology (SMART)[199], which is an NMR-based ML tool, could offer significant advantages in identifying the compounds highlighted in the 2D HetCA analysis. This could be especially helpful for the initial signal assignment and subsequent plausibility check of the data.

In summary, 2D HetCA should not be used as the sole tool for identifying bioactive constituents in complex mixtures derived from natural sources. Rather, it serves as a valuable complement to 1D HetCA and other techniques, such as classical phytochemical approaches, and represents a suitable method for collecting additional data that can significantly support such investigations. Overall, 2D HetCA offers promising opportunities to support the rapid and resource-efficient identification of bioactive compounds in multicomponent NP mixtures.

8 Biopharmaceutical profiling (BPP)

The identification of bioactive compounds that exhibit strong activity (low IC_{50} values) in combination with minimal cytotoxicity (high CC_{50} values), and thereby possess a favorable selectivity index (SI; CC_{50}/IC_{50}), constitutes a crucial step in drug development.[200-203] The SI offers preliminary insights into whether a compound demonstrates selectivity for a specific target, and various selectivity metrics have been developed to evaluate the safety profiles of compounds.[200,203] As a rough estimate, compounds with SI values ≥ 10 are generally considered selective.[202,204]

Sufficient bioactivity is the essential prerequisite for advancing a compound to further stages of investigation. However, it merely represents an initial starting point on the long path from a bioactive candidate to a fully developed pharmaceutical agent.[205] Early evaluation of a compound's overall suitability for drug development is essential, given the extensive time, resources, and financial investment required. The process from target identification to final drug approval typically spans approximately 14 years and may incur costs amounting to hundreds of millions of US dollars.[206]

A lack of adequate selectivity may be one of the factors contributing to the failure of drug candidates to demonstrate efficacy in clinical trials.[203] In addition, inadequate pharmacokinetic (PK) and ADME properties, as well as safety concerns, are further reasons why a bioactive and selective drug candidate may fail to successfully complete the drug development process.[207-210]

One commonly used tool to assess the 'drug-likeness' of a compound is Lipinski's Rule of Five (Ro5).[211] It evaluates physicochemical properties such as molecular weight (< 500 Da), the number of hydrogen bond donors (< 5), hydrogen bond acceptors (< 10), and the logP (the logarithmic partition coefficient in a biphasic water–octanol system) value (< 5), to predict the potential oral bioavailability of a molecule.[211,212] Notably, NPs were explicitly excluded from this system at the time of its publication[16,74], and they are often found in the chemical space beyond Ro5 (bRo5).[213]

Biopharmaceutical profiling (BPP) serves as a more refined and crucial method for assessing the drug-likeness of candidates, understood in terms of PK and safety at the *in vivo* level, and prioritizing compounds that show the most promise for further development.[207,214,215] BPP enables the systematic analysis and evaluation of a drug candidate to facilitate early assessment of whether its physicochemical, PK, and biopharmaceutical properties justify progression within the drug development pipeline toward a therapeutic product.[210,215] It should be noted that the delineation between the respective areas is inconsistently applied across the literature.

However, physicochemical and biopharmaceutical properties encompass parameters such as stability, pH-dependent solubility, pKa ($-\log_{10}$ of the acid dissociation equilibrium constant), lipophilicity (represented by logP), permeability, and ADME-related descriptors.[210,216-218] These properties, in conjunction with biological factors including membrane structures and metabolic

processes, determine the compound's interaction with physiological barriers encountered along its path to the therapeutic site of action.[210,215,219]

Compounds exhibiting favourable biopharmaceutical properties are more likely to achieve safe and efficacious concentrations at the target of interest.[210] In contrast, poor solubility, limited permeability, safety concerns, or unfavourable ADME profiles may severely compromise a drug candidate's suitability for clinical application, ultimately leading to development failure.[210] Kerns and Di provided an overview of the applications of BPP-related information for ADME prediction and the selection of candidates for costly *in vivo* assays.[215]

BPP is commonly integrated into lead candidate optimization studies for synthetic small molecules to gain insights into prospective bioavailability within the physiological milieu.[210,215] In contrast, such systematic evaluation is not routinely applied to drug candidates derived from NPs.[220] Nevertheless, it has been demonstrated that the rate and extent of bioavailability of preparations derived from NP extracts are highly dependent on their biopharmaceutical properties.[220] However, the unique characteristics of NPs require new methods to evaluate their therapeutic potential compared with traditional small molecule drug candidates.[4]

8.1 Biopharmaceutics Classification System (BCS)

Some BPP methods are based on, or have been influenced by, the Biopharmaceutics Classification System (BCS), a framework developed to provide insights into potential oral bioavailability.[207] The BCS was introduced in 1995 by Amidon et al. with the aim of classifying compounds based on their aqueous solubility and intestinal permeability, in combination with the dissolution behaviour of the pharmaceutical formulation.[221,222] This classification is founded on the assumption that solubility and permeability are key factors determining the rate and extent of oral drug absorption, which is in return relevant to the clinical efficacy of a medicinal product.[220,222,223] Therefore, the BCS system proposes a classification into four distinct categories based on the combinations of high or low solubility with high or low permeability.[221]

Reddy and Karunakar defined the four classes, based on the work of Amidon et al., as follows[221,222]:

(i) BCS Class I: High solubility and high permeability
Drugs in this class are well absorbed, with dissolution (or gastric emptying in case of rapid dissolution) being the limiting factor. Example: Metoprolol.

(ii) BCS Class II: Low solubility and high permeability
These drugs exhibit high absorption but low dissolution rates, making dissolution the rate-limiting step. Example: Glibenclamide.

(iii) BCS Class III: High solubility and low permeability
While these compounds dissolve rapidly, their limited permeability restricts absorption. Drugs in this class often show considerable variability in absorption. Example: Cimetidine.

(iv) BCS Class IV: Low solubility and low permeability
Poor solubility combined with limited permeability results in low bioavailability. Compounds in this class are rarely developed or marketed, although some examples exist. Example: Hydrochlorothiazide.

In addition, Reddy and Karunakar provided information on the threshold parameters for the BCS classification.[222] According to their definition, solubility is considered high when the highest dose strength of a drug dissolves in 250 mL or less water across a pH range of 1 to 7.5 at 37 °C.[222] High permeability was defined as an extent of absorption exceeding 90% of the administered dose, relative to intravenous administration.[222] Rapid dissolution is defined as the dissolution of at least 85% of the tested drug within 30 minutes in 900 mL or less of buffer solution under specified apparatus conditions.[222]

The BCS classification enables the estimation of *in vitro–in vivo* correlations (IVIVC) for immediate-release products.[221,222] For Classes I and II, Amidon et al. state that IVIVC generally exists, provided that the dissolution rate for Class I drugs is slower than the gastric emptying rate and that the administered dose for Class II drugs is not excessively high.[221] For Classes III and IV, the authors note that IVIVC is limited or does not exist.[221]

BCS studies are not routinely or systematically conducted in the field of NP drug development.[220,224] Reasons for this may include the complexity of NP chemistry with bRo5 properties[213,224,225], and high variability in the content of marker compounds.[226] Nevertheless, some examples of BCS-related bioavailability investigations of herbal multicomponent products have been reported.[220] Blume and Schug described such investigations for silymarin tablet formulations and for products derived from *Hypericum perforatum* L.[220] In 2016, Li et al. investigated the characteristics of 109 NP-derived drugs with regard to their BCS classification.[224] They applied an extended concept of the BCS, called the Biopharmaceutics Drug Disposition Classification System (BDDCS), which is based on the observation of a strong correlation between intestinal permeability and the extent of metabolism of compounds.[224,227] It has been found that drugs exhibit high permeability if elimination is metabolism-driven, and low permeability if renal and biliary excretion of the unchanged drug determine the route of elimination.[224,227] Consequently, instead of categorizing drugs based on solubility and permeability, the BDDCS classifies them

according to solubility and metabolism.[224] Additionally, the BDDCS also incorporates mechanisms of drug clearance and the effects of uptake and efflux transporters on ADME.[223]

The results of these studies by Li and co-workers suggest that the majority of plant-derived NPs exhibit high solubility and extensive metabolism.[224] 50% were classified as belonging to Class I (high solubility, extensive metabolism), 22% to Class II (low solubility, extensive metabolism), 28% to Class III (high solubility, poor metabolism), and none to Class IV (low solubility, poor metabolism).[224] The authors conclude that many NPs, i.e., those in Class I, are expected to readily cross membranes by passive diffusion and are not expected to interact with transporters.[224] Additionally, Varma et al. pointed out that this correlation between intestinal permeability and metabolism could be used to predict the elimination route based on knowledge of the corresponding permeability values.[223]

In general, the main goal of the various classification systems in this field is to assess potential bottlenecks in drug development and to identify suitable formulation strategies for the active ingredient at an early stage.[222-224,227]

8.2 Pharmacokinetic (PK) synergy

The use of preparations derived from natural compounds, containing precisely defined combinations of various plant extracts as found in traditional medicine systems, suggests that the physicochemical and biopharmaceutical properties of NPs may depend on the composition of the formulation.[226] PK synergy refers to various effects that can occur when active pharmaceutical ingredients (APIs) are administered as part of a multicomponent mixture, as opposed to their isolated form.[228-230]

Panossian clearly summarized these effects, which include additive (1 + 1 = 2), potentiating (0 + 1 > 1), amplifying (1 + 1 > 2), synergistic (0 + 0 > 0), as well as antagonistic (1 + 1 < 2) and attenuating (1 + 1 < 2) interactions.[226] PK synergy is particularly relevant for herbal preparations, which are typically composed of complex multicomponent mixtures.[231] Synergies in PK properties among compounds within herbal extracts have been described across all stages of ADME.[230]

There is a growing body of evidence suggesting that herbal extracts may exhibit greater therapeutic efficacy than their isolated APIs.[231-237] For example, the peak plasma concentrations and areas under the curve (AUCs) of tanshinone IIA and cryptotanshinone, two diterpenoids derived from the root of *Salvia miltiorrhiza* Bge., were significantly increased when administered as part of extracts rather than as isolated compounds in rats.[232] Additionally, significantly higher levels of ginsenoside Re were detected in mouse serum samples when it was administered within a plant extract compared to the pure compound.[233] Moreover, paeoniflorin, a monoterpenoid glycoside found in *Paeonia lactiflora* Pall.,[238] exhibited significantly higher AUC values when orally administered to

rats as part of a plant decoction compared to pure paeoniflorin.[234] Zhang et al. reported that the bioavailability of osthole, a coumarin derivative, was significantly increased when it was orally administered to rats as part of a traditional multicomponent prescription compared to when the isolated compound was given.[235] A publication from 2015 also demonstrated that the PK behavior of hypaconitine, one of the main aconitum alkaloids in *Aconitum carmichaelii* Debx., was influenced by the presence of other compounds in extracts and decoctions, as reflected in parameters such as AUC, maximum concentration (c_{max}), and time to maximum concentration (t_{max}).[236] Finally, Gilbert and Alves listed several examples of synergy effects reported in the literature, including investigations of *Zingiber officinale* Rosc., *Ginkgo biloba* L., *Glycyrrhiza glabra* L., *Artemisia annua* L., *Camelia sinensis* (L.) Kuntze, among others.[237]

These findings suggest that understanding PK synergies can provide valuable insights into whether a pure compound or a multicomponent mixture would exhibit favourable properties in an NP-based therapeutic product.

8.3 BPP and PK synergies of anti-infective MDAAs

MDAAs, prenylated flavonoids found in the root bark of *M. alba* but not in other parts of the plant such as the leaves, fruits, or twigs, exhibit noteworthy anti-infective properties.[122,123] In a 2016 study, Grienke et al. demonstrated that selected MDAAs inhibited both the viral neuraminidase (NA) of Influenza A virus (IAV) and the bacterial NA of *S. pneumoniae*.[123]

NA is a surface glycoprotein of the IAV that exhibits sialidase activity.[239-241] Along with hemagglutinin (HA), it is one of the two major antigens of the viral envelope and can cleave the α-glycosidic bond between sialic acid (N-acetylneuraminic acid) and the adjacent sugar at terminal, non-reducing ends.[239-241] NA thus enables the release of newly formed daughter viruses, which (without its sialidase activity) would remain attached to the infected cell via their own HA binding to sialic acid residues.[239-241] The viral NA of IAV is targeted by NA inhibitors such as oseltamivir (Tamiflu®, orally administered), zanamivir (Relenza®, inhaled), peramivir (Rapivab®, administered intravenously), and laninamivir (Inavir®, inhaled, single-dose administration).[240,242,243]

However, the occurrence of NA is not limited to viruses;[244] it has also been reported for bacteria such as *S. pneumoniae*, *Haemophilus influenzae*, and *Pseudomonas aeruginosa*.[244] Humans also possess four types of NA, called NEU1–NEU4, which, however, differ significantly in structure, as they are presumed to be monomers rather than tetramers like those of IVA.[239,245,246] Human NA performs functions in cellular metabolism.[245]

It has been reported that secondary infections with bacterial pathogens can occur NA-mediated during ongoing IAV infections, most commonly involving *S. pneumoniae*.[123,247-249] The resulting lethal synergism is based on several factors.[123,247,248,250] Among them, infection of epithelial cells

in the upper respiratory tract by IAV disrupts epithelial barrier functions and causes tissue damage, thereby increasing host susceptibility to bacterial infection.[123,247,248,250] Furthermore, the cleavage of sialic acids provides a nutrient source for *S. pneumoniae*.[123,251] The NA of *S. pneumoniae* primarily enhance bacterial virulence by facilitating access to receptors on lung epithelial cells, but can also promote viral pathogenicity by supporting viral release.[123,247,252-254] In addition, the bacterium can form a biofilm composed of pathogens and extracellular matrix components.[123,255]

Secondary infections contribute significantly to morbidity and mortality associated with IAV infections and are more frequently associated with complications such as pneumonia.[248,254] Overall, the availability of dual NA inhibitors for therapeutic use could therefore represent a substantial advantage.[124]

As reported by Grienke et al. in 2016, MDAAs, in contrast to the approved NA inhibitor oseltamivir, not only exhibited dual antiviral and antibacterial NA activity but also inhibited pneumococcal biofilm formation and planktonic growth without harming lung epithelial cells.[123] They are therefore considered promising dual-acting anti-infectives capable of preventing the lethal synergism between the NAs of IAV and *S. pneumoniae*.[123] MDAAs demonstrated significant activity in inhibiting pneumococcal growth (SGC minimal inhibitory concentration (MIC) = 2.73 µM; SGD MIC = 8.72 µM) and biofilm formation (SGC minimal biofilm inhibitory concentration (MBIC) = 1.08 µM; SGD MBIC = 5.89 µM).[123] Furthermore, an extract enriched in prenylated constituents also showed activity against pneumococcal NA.[123]

Additionally, Langeder et al., using a ^{1}H NMR-based HetCA approach, demonstrated that MDAAs were solely responsible for the *in vitro* anti-influenza activity of MDAA-containing extracts.[122] SGC was identified as the main contributor to this antiviral effect.[122] Specifically, MDAA-enriched extracts and pure compounds SGD, SGG, and SGC were well tolerated by the lung epithelial cell lines A549 and Calu-3.[122] The extract, SGG, and SGC showed broad anti-influenza activity, including activity against IAV and Influenza B virus (IBV) strains resistant to the NA inhibitor oseltamivir.[122] Notably, sanggenons inhibited IAV replication in Madin-Darby Canine Kidney (MDCK) cells in a dose-dependent manner, but showed no effect in bronchial epithelial cells.[122]

Anti-inflammatory activities of extracts and compounds derived from the root bark of *M. alba* have also been reported in the literature.[122,256-258] Traditional applications, such as those in Traditional Chinese Medicine (TCM), where the root bark is referred to as ***sāng bái pí***, indicate its use in inflammatory lung conditions such as bronchitis.[123,256]

As summarized by Lim et al., methanolic and ethanolic extracts of the root bark of *M. alba*, as well as isolated prenylated flavonoids, exhibited anti-inflammatory effects in various assays.[256] Moreover, Dat et al. reported anti-inflammatory activity of SGC and SGO through inhibition of NF-κB activation.[258] Several prenylated flavonoids from *M. alba* were also found to inhibit interleukin 6

(IL-6) production and were described as potentially anti-inflammatory.[257] Furthermore, Langeder et al. investigated the anti-inflammatory potential of MDAA-enriched extracts and observed reduced expression of CXC Motif Chemokine Ligand 10 (IP-10), IL-6, and interferon ß (IFN-β) in IAV-infected bronchial cells.[122] The obtained MIC values for the compounds SGD, SGG, and SGC ranged from 1.7 to 3.16 µM against *S. pneumoniae* strains and from 3.16 to 10 µM against *S. aureus* strains, with SGC being the most active compound.[122]

As an additional effect, as described in **Publication I**, compounds such as SGC (IC_{50} = 4.6 µM), SGG (IC_{50} = 7.6 µM), and SGO (IC_{50} = 8.0 µM) exhibited anti-SARS-CoV-2 activity in a spike protein immunostaining assay in Calu-3 cells, while showing only moderate cytotoxicity (SGC: CC_{50} = 23.6 µM; SGG: CC_{50} = 49.1 µM; SGO: CC_{50} = 24.9 µM).[119] An extract enriched with 29% MDAAs also inhibited SARS-CoV-2 replication (IC_{50} = 23.0 µg/mL).[119]

In summary, MDAAs demonstrate promising potential for drug development against IAV and SARS-CoV-2, while simultaneously exhibiting antibacterial and anti-inflammatory properties.[119,122,123,256-258]

Therefore, Langeder et al. conducted studies to evaluate the oral bioavailability of MDAAs and MDAA-containing extracts.[125] Selected pure anti-infective MDAAs, namely SGC and SGD, along with an extract containing 29% MDAAs, were orally administered to mice for seven days.[125] The aim was to assess toxicity, as well as anti-influenza and anti-inflammatory effects.[125] Serum and tissue samples (lung, spleen, liver) were collected from different groups of mice (placebo-/mock-treated, combined with IAV-infected/mock-infected) for analysis.[125] The maximum tolerated dose (MTD) was determined to be 100 mg/kg, with only moderate efficacy observed.[125] Additionally, the authors found pronounced differences in extraction recovery rates for SGC and SGD depending on the tissue environment, and they suspect that differences in stereochemical configuration may be responsible.[125] Repeated administration did not result in the accumulation of MDAAs.[125] In general, the measured concentrations of SGC and SGD in the various tissues were extremely low.[125] The authors reported no statistically significant treatment effects for either MDAA and therefore recommend intravenous or inhalation administration over oral delivery in future studies.[125]

Based on these results (anti-inflammatory, antiviral, and antibacterial effects, along with low plasma/tissue concentrations after oral application), alternative routes of administration should be explored.[119,122,123,125,256-258] In the literature on the traditional use of *M. alba* root bark preparations, respiratory infections are repeatedly mentioned as a primary indication.[123,256,257] Therefore, inhalational administration with local drug delivery directly to the lungs appears to be the most promising approach.

BPP AND PK SYNERGIES OF ANTI-INFECTIVE MDAAs

To assess the suitability of MDAA-enriched extracts and pure MDAAs for inhalation administration, studies on their BPP and PK synergy effects were therefore required. Notably, no studies applying established protocols for NPs in this context have been reported in the literature.

In **Publication II**, strategies tailored to NPs were developed for BPP and PK synergy, with a particular focus on the inhalational administration of both multicomponent mixtures (two MDAA-containing extracts, namely MA21 and MA60) and isolated natural compounds (SGC and SGD). Based on the structural characteristics of MDAAs, it was hypothesized that sanggenons may exhibit low permeability across the air-blood epithelial barrier, thereby enabling efficient and sustained concentrations in the epithelial lining fluid (ELF) of the lungs.

8.3.1 Publication II

Biopharmaceutical profiling of anti-infective sanggenons from *Morus alba* root bark for inhalation administration

Jacqueline Schwarzinger*, **Sigrid Adelsberger***,
Karin Ortmayr, Sarah Luise Stellnberger, Ammar Tahir,
Gabriela Hädrich, Verena Pichler, Judith M. Rollinger,
Ulrike Grienke and Lea Ann Dailey

* equal contribution

International Journal of Pharmaceutics: X 2024, 8, 100272.
DOI: 10.1016/j.ijpx.2024.100272

Impact Factor (2023): 5.179; Pharmacology & Pharmacy: Q1

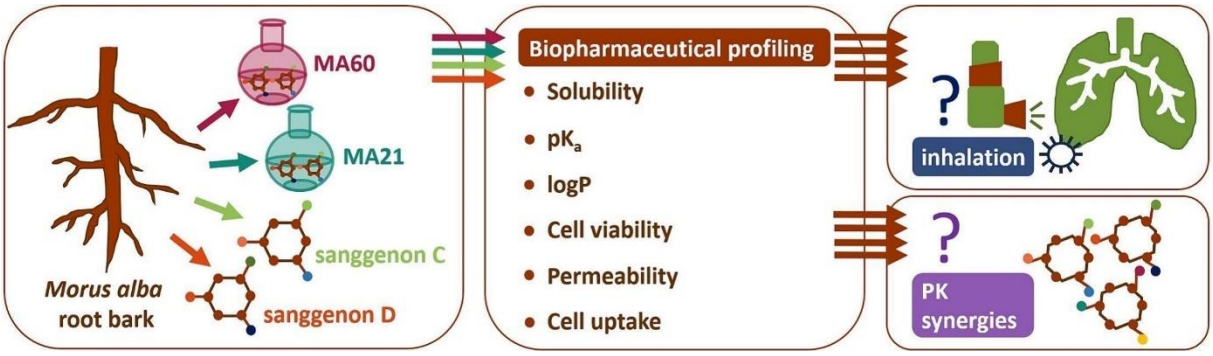


Figure D: Graphical abstract of **Publication II**

In this study, I developed appropriate sample preparation procedures and extraction protocols for the analytical part of the pH-dependent solubility and permeability studies, and I performed the corresponding quantitative measurements. Furthermore, Jacqueline Schwarzinger and I were responsible for evaluating the datasets obtained from all experiments, focusing on potential differences between pure compounds and those in extracts. We were jointly responsible for drafting and writing the manuscript, and we created all figures, tables, and the graphical abstract.

International Journal of Pharmaceutics: X 8 (2024) 100272

Contents lists available at ScienceDirect

International Journal of Pharmaceutics: X

journal homepage: www.sciencedirect.com/journal/international-journal-of-pharmaceutics-x

Biopharmaceutical profiling of anti-infective sanggenons from *Morus alba* root bark for inhalation administration

Jacqueline Schwarzinger [a,b,1], Sigrid Adelsberger [b,c,1], Karin Ortmayr [c], Sarah Luise Stellnberger [b,d], Ammar Tahir [c], Gabriela Hädrich [a], Verena Pichler [d], Judith M. Rollinger [c], Ulrike Grienke [c,*], Lea Ann Dailey [a,*]

[a] Division of Pharmaceutical Technology and Biopharmaceutics, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria
[b] Vienna Doctoral School of Pharmaceutical, Nutritional and Sport Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria
[c] Division of Pharmacognosy, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria
[d] Division of Pharmaceutical Chemistry, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria

ARTICLE INFO

Keywords:
Morus alba
Moraceae
Inhalation
Biopharmaceutical profiling
Pharmacokinetic synergy
Permeability

ABSTRACT

Mulberry Diels-Alder-type adducts (MDAAs), isolated from *Morus alba* root bark, exhibit dual activity against viral and bacterial pathogens but show sobering efficacy following oral administration. Inhalation administration may overcome issues with oral bioavailability and improve efficacy for the treatment of respiratory infections. To assess the suitability of MDAAs for inhalation administration, physicochemical (e.g. pH, pK_a, logP, pH-dependent solubility) and biopharmaceutical (epithelial cytotoxicity, permeability, and uptake) properties of two bioactive MDAA stereoisomers sanggenon C (SGC) and sanggenon D (SGD) were evaluated as isolated natural compounds and within parent extracts (MA21, MA60). Despite their structural similarity, SGD exhibited a 10-fold higher solubility than SGC across pH 1.2–7.4, with slight increases at neutral pH. Both compounds were more soluble in isolated form than in the parent extracts. The more lipophilic SGC was found to be more cytotoxic when compared to SGD, indicating a better cellular penetration, which was confirmed by uptake studies. Nonetheless, SGC and SGD exhibited no measurable permeability across intact Calu-3 monolayers, highlighting their potential for increased lung retention and improved local anti-infective activity following inhalation administration. Results suggest that SGC and SGD in isolated form, rather than as extracts, are promising candidates for pulmonary drug delivery to treat lung infections.

1. Introduction

Poor or undesired biopharmaceutical properties are, next to adverse effects and lack of efficacy, one of the main reasons for lead candidate failure during non-clinical and clinical drug development (Prentis et al., 1988; Waring et al., 2015). Currently, biopharmaceutical profiling studies are often performed in the lead candidate optimization phase for synthetic small molecule drug candidates (Kerns and Di, 2003; Venkatesh and Lipper, 2000) but not routinely investigated for promising compounds isolated from natural sources, such as herbal extracts (Blume and Schug, 2000). Nonetheless, there is a growing body of literature documenting the biopharmaceutical behaviour of natural product-derived compounds in isolated versus extract form. The term "pharmacokinetic synergy" (PK synergy) has recently been used to describe the observation that important physicochemical and biopharmaceutical properties of natural product-derived active pharmaceutical ingredients (APIs), such as solubility and cell permeability, can differ substantially when the key constituents are in isolated form or in their corresponding herbal extract (Li et al., 2011; Machado et al., 2020; Zhao et al., 2020). Many studies found an increased solubility of the analyte in the presence of coexisting compounds (Jürgenliemk and Nahrstedt, 2003; Keung et al., 1996; Kornievskaya et al., 2007; Ma et al., 2016; Xu et al., 2007). Interestingly, PK synergies between constituents of natural product extracts have been reported to increase or decrease the permeability of

* Corresponding authors.
E-mail addresses: ulrike.grienke@univie.ac.at (U. Grienke), leaann.dailey@univie.ac.at (L.A. Dailey).
[1] J. Schwarzinger and S. Adelsberger contributed equally to this work.

https://doi.org/10.1016/j.ijpx.2024.100272
Received 14 June 2024; Received in revised form 31 July 2024; Accepted 1 August 2024
Available online 5 August 2024

plant constituents (Ta et al., 2023; Zheng et al., 2012; Zheng et al., 2015; Zhu et al., 2013). In some cases, PK synergies of the extract have been shown to improve therapeutic efficacy compared to the purified bioactive compound (Hengjumrut et al., 2018; Joo et al., 2010; Song et al., 2007; Weathers et al., 2011; Wu et al., 2009; Zhang et al., 2014; Zhang et al., 2015). Thus, understanding PK synergies is an important step towards designing a suitable target product profile (TPP) for therapeutic agents derived from natural products. This type of investigation addresses the question of whether the extract or isolated bioactive agent would exhibit favourable properties in the drug product.

The white mulberry tree, i.e. *Morus alba* L. (Moraceae), has a long history of traditional medicinal use in Asian countries (Singh et al., 2013). Especially root bark preparations have been reported as oral treatment options for various lung inflammatory disorders such as acute and chronic bronchitis due to their antimicrobial and anti-inflammatory activity (Batiha et al., 2023; Chan et al., 2016; Lim et al., 2013). Mulberry Diels-Alder-type adducts (MDAAs) such as sanggenon C (SGC), sanggenon D (SGD), and sanggenon G (SGG), are prenylated flavonoids formed via [4 + 2]-cycloaddition between chalcones and dehydroprenylphenols (Luo et al., 2021, 2022) (Fig. 1) with promising pharmacodynamic activity as anti-infectives. Recent in vitro studies show that MDAAs combat the lethal synergism of viruses (influenza A virus, SARS-CoV-2) and bacteria (pneumococci) (Grienke et al., 2016; Langeder et al., 2023a; Wasilewicz et al., 2023).

However, a major drawback is their low in vivo bioavailability and low plasma/tissue concentration after oral application (Langeder et al., 2023b). Lipinski's rule of five characterizes the properties of a substance, such as molecular weight (< 500 Da), hydrogen bond donors (< 5) and acceptors (< 10), as well as the logP value (< 5), to determine its "drug-likeness" and predict its potential for good oral bioavailability (Lipinski et al., 1997). The poor oral bioavailability of SGC and SGD could be due to the complex molecular structure of the sanggenons, which does not adhere to Lipinski's rule of five because of their high molecular weight (708.7 g/mol) and the presence of eight hydrogen bond donors and 12 hydrogen bond acceptors. The lack of oral bioavailability may be overcome by local drug delivery to the lung. Since respiratory infections are one of the primary target disease areas for MDAAs, the option of developing an inhalation product instead of an oral formulation is an attractive solution to circumvent this dilemma and at the same time benefit from a targeted application.

Marchand and Couet have extensively investigated the aerosol administration of antibiotics for pulmonary infections, to understand molecular properties favourable for lung retention and therefore improved local efficacy. They observed that molecules characterized by a large, hydrophilic structure and low permeability across lung epithelial monolayers exhibit favourable pulmonary pharmacokinetics, with high lung concentrations and low systemic exposure. The investigated antibiotics showed the following P_{app} values: tobramycin ($P_{app} < 0.05 \times 10^{-6}$ cm/s), aztreonam (P_{app} $0.07 \pm 0.02 \times 10^{-6}$ cm/s), and colistin (P_{app} $0.04 \pm 0.02 \times 10^{-6}$ cm/s), which are defined as low P_{app} values (Galindo Bedor et al., 2016; Gontijo et al., 2014a; Gontijo et al., 2014b; Marchand et al., 2010, 2015, 2016, 2018). In comparison, metoprolol, which represents the border between high and low permeability, has a P_{app} of $10.3 \pm 0.0 \times 10^{-6}$ cm/s in Calu-3 cells (Bosquillon et al., 2017).

Taking these observations into consideration, we hypothesize that sanggenons may also exhibit prolonged retention in the lungs. However, no physicochemical and biopharmaceutical data on sanggenons as a class of molecules have been published. To address this lack of information, two MDAA-containing extracts (MA21 and MA60) and their isolated major MDAA stereoisomers (SGC and SGD) were selected as representative compounds for the evaluation of (i) pH-dependent solubility, (ii) pK_a, (iii) lipophilicity (logP), (iv) cell viability, (v) epithelial permeability and (vi) cell uptake. Due to their structural characteristics, we hypothesize that sanggenons might exhibit low permeability characteristics and therefore show a potential for prolonged lung retention. The study provides relevant data for the possible development of an inhaled drug product based on MDAA-containing extracts or purified bioactive compounds (i.e. sanggenons) and acts as a general template for biopharmaceutical profiling of natural product-derived APIs. Such molecules can often show a surprising degree of chemical complexity compared to standard synthetic small-molecule APIs and can therefore present a substantial challenge for subsequent formulation.

2. Materials and methods

2.1. Materials

Fetal bovine serum (FBS), phosphate-buffered saline (PBS), Hanks' balanced salt solution (HBSS), fluorescein sodium salt, metoprolol tartrate and dimethyl sulfoxide (DMSO, purity 99.8%, CAS: 67–68-05) were obtained from Sigma-Aldrich (St. Louis, USA). Dulbecco's modified Eagle medium/nutrient mixture F-12 (DMEM/F-12), GlutaMAX™ supplement, penicillin-streptomycin (10,000 U/mL) solution, trypsin-EDTA (0.25%) and thiazolyl blue tetrazolium bromide (MTT) were purchased from Thermo Fisher Scientific (Waltham, USA). Ammonia solution Rotipuran® ($NH_3 + H_2O$), was purchased from Carl Roth GmbH + Co. KG (Karlsruhe, Germany). LC-MS grade methanol (MeOH), was purchased from VWR Chemicals (Radnor, USA). Double-distilled deionized water (dd H_2O) for preparation of solubility and permeability samples was produced by a Milli-Q-Plus ultrapure water device (Millipore Corporation, Bedford, MA, USA) and by an Arium® Pro Ultrapure Lab Water System from Sartorius (Göttingen, Germany). Reagent grade tert-butyl methyl ether (MTBE), was purchased from Sigma-Aldrich (St. Louis, USA). Formic acid (FA) p.a., used as an additive for mobile phases, was purchased from Chem-lab nv (Zedelgem, Belgium).

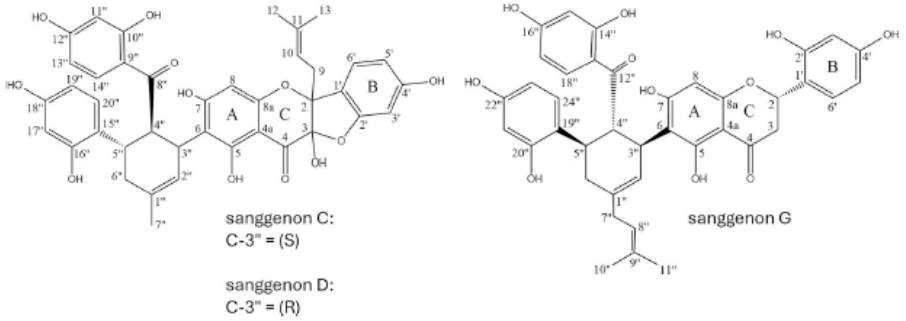


Fig. 1. Structures of stereoisomers sanggenon C (SGC) and sanggenon D (SGD), both prenylated in position C-2, and sanggenon G (SGG) with an isoprenyl unit in position C-1".

2.2. Extracts and pure compounds

One batch of two respective *M. alba* root bark extracts, MA21 and MA60, (Langeder et al., 2023a) were used for this study. MA21 is a hydroethanolic extract (extracted at room temperature with 60 vol% ethanol) containing 1.0% of SGC and 1.1% of SGD, with a total sanggenon content of 2.6%. MA60 was produced by pressurised liquid extraction using *n*-hexane for defatting and isopropanol-petroleum ether (2:1) for extraction to obtain a high content of sanggenons in the final extract. MA60 contains 6.9% of SGC and 10.7% of SGD with an overall total sanggenon content of 29.0%. Both MDAAs, SGC and SGD (purity >98%), were previously isolated from MA60 as described in Langeder et al., 2023a. For quantitative analysis of the permeability assay, SGG was used as an internal standard. This compound, with a purity of 98% according to UPLC-ELSD detection, was also previously isolated and analysed (Langeder et al., 2023a).

2.3. pH-dependent solubility profile

The thermodynamic solubility was evaluated with an adapted miniaturized shake-flask method in a pH-dependent manner (Glomme et al., 2005; Veseli et al., 2020). Weighed masses (SGC, SGD: 0.15 mg; MA21: 14.2 mg; MA60: 1.4 mg) were selected to exceed the estimated solubility values determined in preliminary studies by a factor of five, to ensure that saturation conditions were reached. The masses of the extracts were then chosen to contain the same quantity of sanggenons as the purified compounds to improve comparability. MDAA samples were first prepared as stock solutions in methanol. The desired concentrations were then pipetted into new vials, and methanol was evaporated to obtain the required quantities. A volume of 100 µL of the following buffers was added to the sample: HCl buffer (pH 1.2), acetate buffer (pH 4.5), and phosphate buffer (pH 6.8 and 7.4). After five minutes of sonication, the suspensions were incubated under shaking conditions (250 rpm at 24 ± 0.24 °C) for 24 h (Incubator hood TH15 + Compact Shaker KS 15 A control from Edmund Bühler GmbH, Bodelshausen, Germany). Separation of dissolved and non-dissolved compounds was achieved via centrifugation at 6708*g* for 20 min (MIKRO 200/200R, Andreas Hettich GmbH & Co. KG, Tuttlingen, Germany). Supernatants were removed and stored at −20 °C until measurement. Shortly before the measurement, 80 µL of the thawed supernatant samples were centrifuged and neutralised with 0.1% aqueous ammonia, if necessary, to adjust the pH. A 3-cycle MTBE-based extraction protocol using 3 × 750 µL MTBE was then implemented. Quality control (QC) evaluations included low, middle, and high-concentration samples from the calibration curve range. Blanks were prepared by applying the same extraction protocol to the respective buffer systems. Quantitative analysis was performed using an Acquity UPLC H-class system by Waters equipped with an Acquity BEH Phenyl column (2.1 × 100 mm, 1.7 µm) and a PDA detector. Detailed information about the preparation of samples and blanks, QC samples and QC blanks, as well as information about method validation and quantitation using UPLC-PDA can be found in the electronic supplementary information (ESI).

2.4. pK_a of sanggenons C and D

Experimental pK_a ($-\log_{10}$ of the acid dissociation equilibrium constant) values of SGC and SGD were evaluated using the UV-metric pK_a measurement method of the SiriusT3 automated titrator instrument and the corresponding software SiriusT3Control and SiriusT3Refine (Pion Inc., Billerica, USA). The SiriusT3 apparatus uses a deuterium lamp to determine the UV/Vis absorbance spectrum of the test solutions at each pH point. The standard aqueous UV-metric pK_a measurement protocol was conducted according to manufacturer instructions. Briefly, a 10 mM stock solution of SGC and SGD was prepared in DMSO ($n = 5$ replicate samples) and 5 µL was added to a 4 mL Sirius T3 flat bottom tube. This was mixed with 25 µL phosphate buffer (14.7 mM K_2HPO_4 and 0.15 M KCl in MilliQ water) and diluted with 1.5 mL ionic strength-adjusted water (ISA water ≡ 0.15 M KCl in in MilliQ water). The subsequent acid/base titration and measurement of UV absorbance was automatically performed by the SiriusT3 at 25.0 ± 0.5 °C. The pH titration was performed in triplicate titrations between pH 1 and 13 by adding acid (0.5 M HCl) and base (0.5 M KOH) under an argon atmosphere. Turbidity was monitored by a turbidity sensor to ensure the absence of precipitation of the test compounds during the experiment, since accurate pK_a measurements are limited to fully dissolved substances. In case turbidity occurred, individual titrations were not used for further analysis, resulting in a total of 11 analysable titrations across the 5 replicate samples for both SGC and SGD.

2.5. logP of sanggenons C and D

LogP (partition coefficient) evaluation was performed via SiriusT3. The potentiometric logP measurement relies on assessing the partition profile of the test substance within a two-phase water-octanol solvent system. Changes in the relative volume ratio of octanol to water lead to shifts in the apparent pK_a, resulting from the partitioning between non-ionised and ionised forms. LogP values were then estimated based on these assessments. SGC and SGD (SGC: 0.91 mg, SGD: 1.00 mg) were weighed and transferred to the SiriusT3 instrument, which automatically added 1.5 mL ISA water and partition solvent (ISA water-saturated octanol). The pH titration ($n = 3$ technical replicates) was performed automatically with acid (0.5 M KCl) and base (0.5 M KOH) solutions.

2.6. Calu-3 cell viability

The cytotoxicity of SGC, SGD, MA21 and MA60 in the Calu-3 cell line was evaluated via MTT (3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide) assay (Berridge and Tan, 1993). Calu-3 cells, a human bronchial cell line, obtained from LGC Standards (Cat No. ATCC-HTB-55) were used from passage number 21–31 for MTT experiments. The cells were seeded in a density of 3×10^4 cells/well in 96 well plates from VWR Chemicals (cat. no. 734–2327, Radnor, USA). After 24 h, the medium was removed and the Calu-3 cells were exposed to SGC (10–80 µg/mL), SGD (40–180 µg/mL), MA21 (200–1600 µg/mL) or MA60 (50–350 µg/mL) for 24 h. This concentration range enabled the determination of CC_{50} values. The pure compounds and the extracts were diluted in medium (10% FBS, 1% P/S) before addition to the cells. Triton-X100 (0.25%) served as a positive control, and cells treated only with medium (10% FBS, 1%P/S) served as a negative control. After 24 h incubation with the test solutions, the supernatants were removed, cell monolayers were washed with 100 µL PBS buffer, and 200 µL of medium and 20 µL of MTT in PBS (5 mg/mL) were applied to each well. After incubation for 3.5 h, the MTT solution was removed and 100 µL DMSO were added to dissolve the formazan crystals. To ensure complete solubilization of the formazan crystals, plates were incubated on a shaking platform (200 rpm) under the exclusion of light for 20 min, and the absorbance was recorded at 570 nm with an Epoch 2 spectrophotometer (BioTek Instruments, Inc., Vermont, USA).

The cell viability was calculated according to Eq. (1).

$$\textit{Cell viability}\ (\%) = \frac{OD_T - OD_P}{OD_N - OD_P}\ x\ 100 \qquad (1)$$

where OD_T is the optical density of the test well at 570 nm, OD_P is the optical density of the positive control (no viability) and OD_N is the optical density of the negative control (100% viability).

2.7. Permeability

Calu-3 cells were seeded on PET membranes of PS inserts (1.1 cm^2, 0.4 µm) from Sarstedt (cat. no. 83.3931.040, Nümbrecht, Germany) with a seeding density of 1×10^5 cells/insert, placed in a 12 well plate

(cat. no. 734–2324, VWR Chemicals, Radnor, USA). Calu-3 cells were used from passages 36–39 for permeability assays. After approximately two weeks of culturing in a liquid-liquid interface, an intact monolayer with tight junctions formed, which is the requirement for the subsequently performed permeability assays. Monolayer integrity was assessed by two methods: 1) the transepithelial electrical resistance or TEER value and 2) the addition of fluorescein (40 µg/mL). Cell layers with TEER values exceeding 1000 Ω cm^2 (reached at days 12–14) on the day of the experiment were considered suitable for permeability assays, as was a lack of measurable fluorescein transport. Immediately before permeability studies, the medium was replaced with HBSS and following a 30 min incubation time, TEER values were measured. Permeability studies were performed in a bidirectional manner for SGC, SGD, MA21, and MA60. Metoprolol served as a reference for comparison with available data from the literature. For permeability assessment in the apical to the basolateral direction (A-B), 700 µL of test solutions (SGC/SGD: 20 µg/mL, MA21: 600 µg/mL, MA60: 100 µg/mL in HBSS) were added to the apical compartment and 1700 µL of buffer to the basolateral chamber. For studies from the basolateral to the apical chamber (B-A), 700 µL of buffer was added to the apical compartment and 1700 µL of the same test solutions to the basolateral compartment. 200 µL were withdrawn from donor and acceptor compartments at two time points: directly after the addition of both solutions ($t = 0$ h) and after 120 min ($t = 2$ h). For the duration of the permeability studies, plates were placed on a shaking platform at 37 ± 0.37 °C at 200 rpm. After incubation, cells were carefully washed and the solutions were replaced with HBSS. After a further 30-min incubation, TEER values were again evaluated to assess cell integrity.

The quantitative analysis of permeability samples was performed on a UHPLC-ESI-MS system (Thermo Fisher Scientific, CA) comprising a Dionex UltiMate 3000 coupled to an LTQ XL linear ion trap mass spectrometer, as described in detail in the Supporting information. SGG was used as an internal standard for quantitation (ratio analyte/internal standard). The apparent permeability coefficient (P_{app}) was calculated using Eq. (2):

$$P_{app} = \frac{dQ}{dt} + \frac{1}{A * C_0} \quad (2)$$

where dQ/dt was the amount of analyte present in the acceptor compartment as a function of time, A was the surface area of the cell monolayer (1.1 cm^2) and C_0 was the initial concentration of the test compound in the donor compartment. The acceptor compartment in the direction A-B is the basal chamber, and in the direction B-A the apical chamber. P_{app} is calculated in the dimension x 10^{-6} cm/s.

2.8. Calu-3 cell uptake

To determine the uptake and retention of SGC and SGD in Calu-3 cells, cells were seeded in DMEM-F12 with 10% FBS at high cell density (5 × 10^5 cells/cm^2) in 24-well plates (cat. no. 142475, Thermo Fisher Scientific). The 2D cell monolayer was allowed to equilibrate for 72 h before the fresh medium was supplied, and SGC and SGD were added to culture supernatants at a final concentration of 50 µg/mL. At 4, 12, and 24 h after addition, cell extracts were generated in triplicates. To that end, the medium was aspirated from the respective wells, and cells were washed with pre-warmed ammonium carbonate solution (75 mM, pH 7.4, 37 °C). Immediately after removal of the wash solvent, 300 µL cold extraction solvent (40:40:20 acetonitrile:methanol:water, −20 °C) was added per well to lyse cells in situ and stop the enzymatic activity. The bottom of each well was subsequently scraped and the cell extracts were transferred to 96-well storage plates and stored at −80 °C until analysis. On the day of MS analysis, the cell extract samples were centrifuged at 6 °C for 10 min to remove cell debris, and the cell-free extract was transferred to a fresh 96-well plate for measurement. Quantification of SGC and SGD was performed using flow-injection time-of-flight mass spectrometry (FIA-TOFMS) in negative ion mode as described previously (Fuhrer et al., 2011). SGC and SGD signals were identified by matching accurate masses (for both [M-H]- at 707.2127 m/z). To account for possible influences of the complex matrix in the cell extract samples, a matrix-matched series of calibration samples were prepared, in which increasing concentrations of SGC were spiked into extracts of cells that had not previously been treated with either SGC or SGD. The calibration samples were injected alongside the cell extract samples and used to establish a calibration function using linear regression. The ion intensities measured were converted into SGC and SGD concentrations and normalized by the cell number in each sample, which was determined at each time point in separate cell cultures using trypan blue staining and cell counting in a Thermo Fisher Countess II Automated cell counter after trypsin-mediated cell detachment. This procedure eventually yielded estimates of SGC and SGD contents in pg/cell.

2.9. Statistical analysis

Data analysis was performed with Microsoft Excel version 2404 and GraphPad Prism version 10.2.3 (403). Results are expressed as mean ± standard deviation (SD). To test for normal distribution the skewness of the data set and the Kolmogorov-Smirnov method were used. A two-sided Student's t-test was performed for parametric data sets. For non-parametric data the Mann-Whitney U test was used instead. For statistical significance $p \leq 0.05$ was set as the cut off. Quantitative and statistical analysis of the cellular uptake of SGC and SGD was performed in Matlab 2020b with custom analysis scripts using the Matlab functions *regress* and *t-test2*.

3. Results and discussion

3.1. pH-dependent solubility profile, pK_a and logP

In pharmaceutical research early solubility characterisation helps to overcome potential bioavailability difficulties and to choose the best administration route as well as to plan the appropriate formulation strategy (Arnott and Planey, 2012). Thermodynamic solubility is widely regarded as the gold standard for accurately representing true solubility. It defines the maximum amount of a solid compound that can dissolve in a solvent at equilibrium. This method is based on saturating the solvent with excess compound until the dissolution equilibrium is attained. Factors such as buffer composition, time, temperature, mixing conditions, and polymorphism can influence the results (Alsenz and Kansy, 2007; Veseli et al., 2019). The European Medicines Agency (EMA) recommends obtaining biopharmaceutically relevant drug solubility data within a physiologically acceptable pH range (International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH Harmonised Guideline: Biopharmaceutics Classification System-Based Biowaiver (M9), 2019). pH-dependent solubility profiles refer to solubility at a given pH, measured in a defined pH-buffered system. The solubility of a substance may change in response to variations in pH caused by factors such as the degree of ionisation. This phenomenon can lead to significant differences in the solubility of ionised and non-ionised molecules at a defined pH value (Bergström et al., 2004; Veseli et al., 2019).

The assessment of pH-dependent solubility of the isolated stereoisomers, SGC and SGD, showed a distinct increase with ascending pH values for both compounds, indicating progressive ionisation of the hydroxyl groups over the range of acidic to neutral pH (Fig. 2A). Furthermore, SGD exhibited an approximately 10-fold higher solubility over SGC at all pH values tested, indicating a profound effect of the steric orientation on the solubility. This was unexpected in light of the high structural similarity between the two stereoisomers. However, differences in solubility for similar flavonoid stereoisomers have been mentioned in the literature (Zheng et al., 2020). Also, the influence on

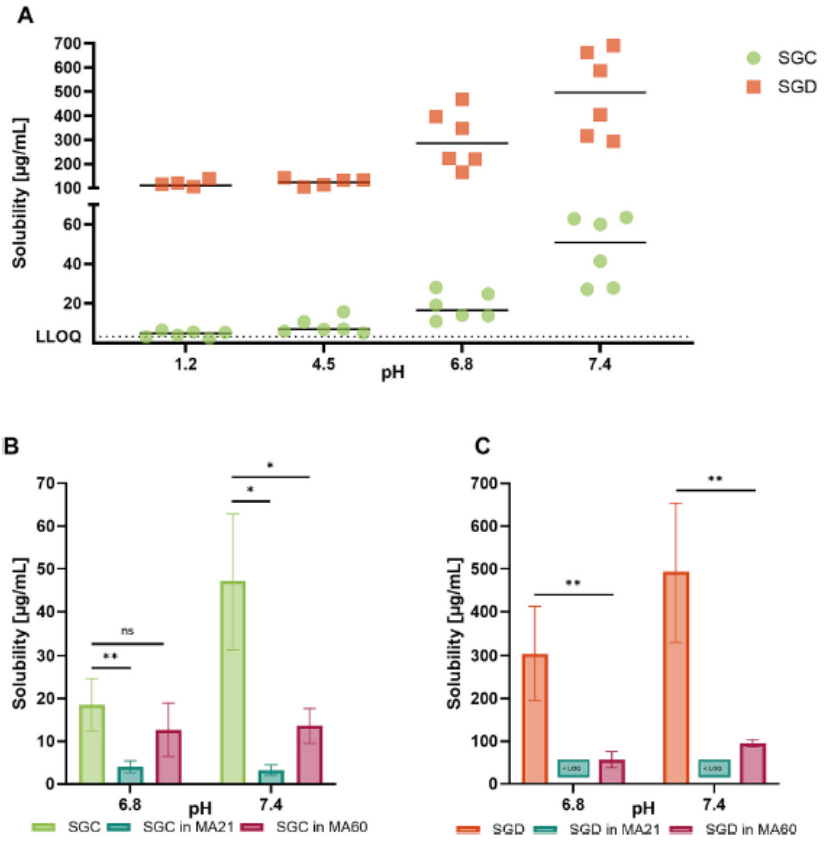


Fig. 2. A) Solubility of sanggenon C (SGC, in green) and sanggenon D (SGD, in orange) in four different buffer solutions with pH values of 1.2, 4.5, 6.8 and 7.4 ($n = 6$; black lines indicate the mean values). Solubility of SGC in B) and solubility of SGD in C), both at pH 6.8 and pH 7.4 compared to their solubility in extracts MA21 and MA60 at the same conditions expressed as the mean ± SD ($n = 3$). Values for SGD in MA21 at respective pH values were lower than the LOQ. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the solubility based on the torsion angle of ring C and B in the flavonoid structure has been described before (Chebil et al., 2007).

The solubility of SGC and SGD within the two multi-component extract mixtures, i.e. MA21 and MA60, at the two neutral pH values, 6.8 and 7.4 (Fig. 2B-C) showed in some cases a significant decrease in solubility compared to the isolated compounds. This is an unusual observation in the field of natural product research, where the opposite behaviour, i.e., a solubility enhancement of APIs due to endogenous solubilization enhancers, is more often reported (Zhao et al., 2020). An increase in water solubility in the presence of a Puerariae radix extract was found for daidzin by Keung and colleagues (Keung et al., 1996). Similarly, the dissolution rate of hypericin was increased by coexisting flavonoids in a *Hypericum perforatum* crude extract (Jürgenliemk and Nahrstedt, 2003). In other examples, an increased solubility of berberine was observed in the presence of Coptidis rhizoma extract (Ma et al., 2016) and the solubility of schizandrin was increased when tested in a complex decoction preparation derived from Schisandrae fructus (Xu et al., 2007).

The lower solubilities of SGC and SGD at acidic pH values might be an explanation for the low serum and tissue concentrations reported by Langeder et al. after oral administration to mice, even at high dose concentrations of up to 100 mg/kg (Langeder et al., 2023b). Even increases in stomach pH caused by the intake of food may not necessarily improve the pH-dependent solubility of SGC and SGD in the stomach following oral administration (Dressman et al., 1990; Langeder et al., 2023b; Simonian et al., 2005). In contrast, the pH of the respiratory tract has been reported to be approximately pH 6.8 (Effros and Chinard, 1969) which provides more favourable dissolution conditions for MDAAs administered via the pulmonary route, at least in terms of pH-effect alone.

To date, no studies have reported experimentally determined values for the ionisation constant (K_a) of MDAAs, including SGC and SGD. Since the pH-dependent protonation state not only affects physicochemical properties such as solubility, dissolution rate and lipophilicity, but also ADME properties, such as permeability (Manallack et al., 2013), this data can be essential and useful for the prediction of API behaviour in the physiological milieu. There are currently different pK_a prediction tools available, such as pKalc, ADMET Predictor, ACD/pK_a or Marvin, to gain first insights into the ionisation behaviour of test compounds (Manallack et al., 2013; Prankerd, 2007). When dealing with the comparably large and complex chemical structures of natural product-derived compounds, like SGC and SGD which both have eight hydroxyl groups, in silico pK_a prediction can be a challenge because each functional group of these multi-protic molecules has a specific pK_a value (Navo and Jiménez-Osés, 2021). In both sanggenons, the respective hydroxyl groups are present at the positions C-3, C-5, and C-7 of the benzopyrone structure as well as at positions C-10″, C-12″, C-16″, and C-18″ connected to the flavonoid core via the tetrasubstituted cyclohexene ring and at position C-4′ in ring B (Fig. 1). Calculated pK_a values, generated via MarvinSketch, range from 7.26 to 11.04 (Table 1), with no possibility to differentiate between the two stereoisomers SGC and SGD.

A total of eight pK_a values were measured for SGC and SGD with the UV-metric standard titration method (SiriusT3 apparatus) and compared to in silico determined values (MarvinSketch, ChemAxon;

Table 1
Experimentally determined (SiriusT3) and calculated (*MarvinSketch, ChemAxon) pK_a ($n = 5$) and logP ($n = 3$) values for sanggenon C (SGC) and sanggenon D (SGD) presented as the mean ± S.D.

Property	SGC (experimental)	SGD (experimental)	SGC and SGD (calculated)
pK_a 1	2.64 ± 0.69	3.27 ± 0.37	7.26
pK_a 2	5.20 ± 0.80	6.37 ± 0.13	7.90
pK_a 3	6.74 ± 0.14	7.25 ± 0.32	8.55
pK_a 4	7.97 ± 0.06	7.99 ± 0.26	8.96
pK_a 5	9.00 ± 0.04	8.99 ± 0.08	9.32
pK_a 6	9.84 ± 0.05	9.94 ± 0.04	9.74
pK_a 7	10.70 ± 0.02	10.78 ± 0.01	10.46
pK_a 8	12.05 ± 0.07	12.04 ± 0.07	11.04
logP	3.48 ± 0.05	2.92 ± 0.07	–

Table 1). Due to the complex structures of SGC and SGD, it was not possible to distinguish and assign the eight pK_a values to the eight proton positions. However, it was observed that four out of eight experimentally determined pK_a values were substantially lower than the in silico results. Further studies, including a set of derivatives with the removal of the respective phenolic hydroxy groups, would be needed to fully assign the pK_a to the respective functional group (Fuguet et al., 2023). However, the mean absolute error (MAE) for the pK_a of SGC and SGC are 1.47 and 1.21, respectively, when comparing the pK_a values closest to each other. This discrepancy may be caused by the fact that the MarvinSketch algorithm only considers molecular structures and their functional groups when predicting pK_a values, without considering the potential influence of adjacent groups. Difficulties in predicting the pK_a values of drug-like molecules can be challenging due to tautomerization, heterocycles, conformational flexibility of large molecules, the ability to form intramolecular hydrogen bonds and the presence of various titratable sites since the protonation state of one functional group can affect the proton dissociation tendency of adjacent groups (Işik et al., 2018). The experimental results did roughly correspond with the solubility data where an increased solubility was observed at pH 6.8 and pH 7.4, suggesting that higher deprotonation ratios of both sanggenons at higher pH levels lead to higher solubility. However, differences in pK_a values and therefore ionisation states of SGC and SGD cannot be completely responsible for the 10-fold higher solubility of SGD compared to SGC, since the first three pK_a values of SGC were significantly lower than the corresponding values of SGD, which would theoretically suggest that under acidic pH conditions, a higher fraction of the SGC molecule should carry anionic charges and therefore be more soluble. Since we observe the opposite, we infer other mechanisms (e.g. tautomerization, conformational state, intramolecular hydrogen bond formation) to be more dominant. These results underscore the importance of experimentally assessing pK_a values to elucidate their impact on physico-chemical features, as prediction models still fall short, especially for natural products.

Lipophilicity is also a key factor influencing the pharmacokinetic behaviour of drug substances, including the transfer across cell membranes, distribution kinetics into tissues, absorption, toxicity and protein binding characteristics of a drug, as well as solubility characteristics in aqueous environments (Gleeson, 2008; Hughes et al., 2008; Testa et al., 2000). Thus, the evaluation of lipophilicity is a standard parameter in biopharmaceutical profiling (Arnott and Planey, 2012). Lipophilicity is typically characterized by the logarithmic partition coefficient (logP) between octanol and water (Leo et al., 1971), whereby high logP values (usually >3) represent a higher lipophilicity (Arnott and Planey, 2012). The experimentally determined logP values of SGC and SGD (Table 1) were in the moderately lipophilic range (~3–3.5), and SGC partitioned to a greater extent in octanol compared to SGD, which also aligns well with the solubility data.

3.2. Cytotoxicity and uptake in Calu-3 cells

Calu-3 cell viability was evaluated via an MTT assay to determine suitable sample concentrations for subsequent permeability investigations. Calu-3 cells, human bronchial epithelial cells, are commonly used for drug transport studies in the lung (Forbes and Ehrhardt, 2005; Mathias et al., 2002). In line with the results of the physicochemical characterisation studies, the stereoisomers SGC and SGD also showed different cytotoxicity profiles, with SGC exhibiting a higher cytotoxicity (Fig. 3A-B). The respective half-maximal cytotoxic concentrations (CC_{50}) were calculated from the dose-response curves giving a CC_{50} value of 47.8 µg/mL for SGC and 120.5 µg/mL for SGD (Fig. 3A-B). The observation that SGC is more cytotoxic than SGD is supported by the high logP of SGC indicating that it can permeate more easily across cell membranes compared to SGD.

Exposure of Calu-3 cells to the multi-component extract mixtures, which have a lower content of total MDAAs (MA21 contains 2.6% and MA60 contains 29.0% MDAAs in total) also provided interesting insights. Extract MA60, with a 10-fold higher total sanggenon content, exhibited greater cytotoxicity (CC_{50} = 238.1 µg/mL) compared to the hydroethanolic extract MA21 (CC_{50} = 822.8 µg/mL; Fig. 3A-B), indicating that increased levels of sanggenons correlate with higher cytotoxicity. When MA21 and MA60 cell viability data are plotted against the respective concentrations of SGC and SGD in each extract (Fig. 3B-C) both extracts demonstrated greater cytotoxicity compared to the individual sanggenons. This indicates the likely presence of other unique and currently unidentified compounds in each extract which may contribute to cytotoxicity. Theoretically, if either SGC or SGD were the main cytotoxic component in the mixture, all three curves should overlap. The mismatch of the curves therefore implies that further constituents in the multi-component mixture enhance cytotoxic effects.

To rule out that SGC and/or SGD are retained within the epithelial cells and thereby prevented to exert their full therapeutic properties in the extracellular space (e.g., neuraminidase inhibition, inhibition of pneumococcal biofilm formation), an assessment of sanggenon uptake in Calu-3 cell monolayers was performed. Calu-3 cells seeded at high confluence were exposed to either SGC or SGD after an initial equilibration period of 72 h. For both SGC and SGD, the concentration tested was 50 µg/mL, at which no cytotoxic effects were observed during a 4-h and 12-h incubation period evaluated via MTT assay (Fig. 4B). After 4, 12 and 24 h, the cell layer was washed with ammonium carbonate solution, before cells were extracted by cold solvent extraction for the analysis of intracellular SGC and SGD contents by mass spectrometry. The analysis of the cell extract samples revealed that SGC showed a distinctly higher accumulation than SGD within Calu-3 cells already at 4 h after addition (2.0 ± 0.1 vs. 0.3 ± 0.2 pg/cell, Fig. 4A), which continued to increase until the intracellular SGC content reached a concentration of 8.1 ± 0.3 pg/cell after 24 h (Fig. 4A). In contrast, SGD levels only marginally increased during the same 24 h period indicating that SGC shows greater penetration of Calu-3 cells, consistent with its higher lipophilicity compared to SGD (higher logP, Table 1) facilitating its entry through the cell membrane. Intriguingly, the low intracellular accumulation of SGD prompts the hypothesis that SGD would remain in the extracellular/apical space and neither enters nor permeates the epithelial barrier. Thus, despite their structural similarity, SGC and SGD exhibit different permeation behaviour, and may thus localize differently at the infection site.

3.3. Bidirectional permeability of SGC and SGD

Unlike orally administered drugs which generally aim for a high systemic bioavailability and therefore benefit from a high permeability, many inhaled drugs, including anti-infective agents for the treatment of lung infections, have local targets in the lung (Gontijo et al., 2014b; Marchand et al., 2015, 2016, 2018). Consequently, a low-to-moderate permeability is desirable for inhaled drugs, since such compounds are

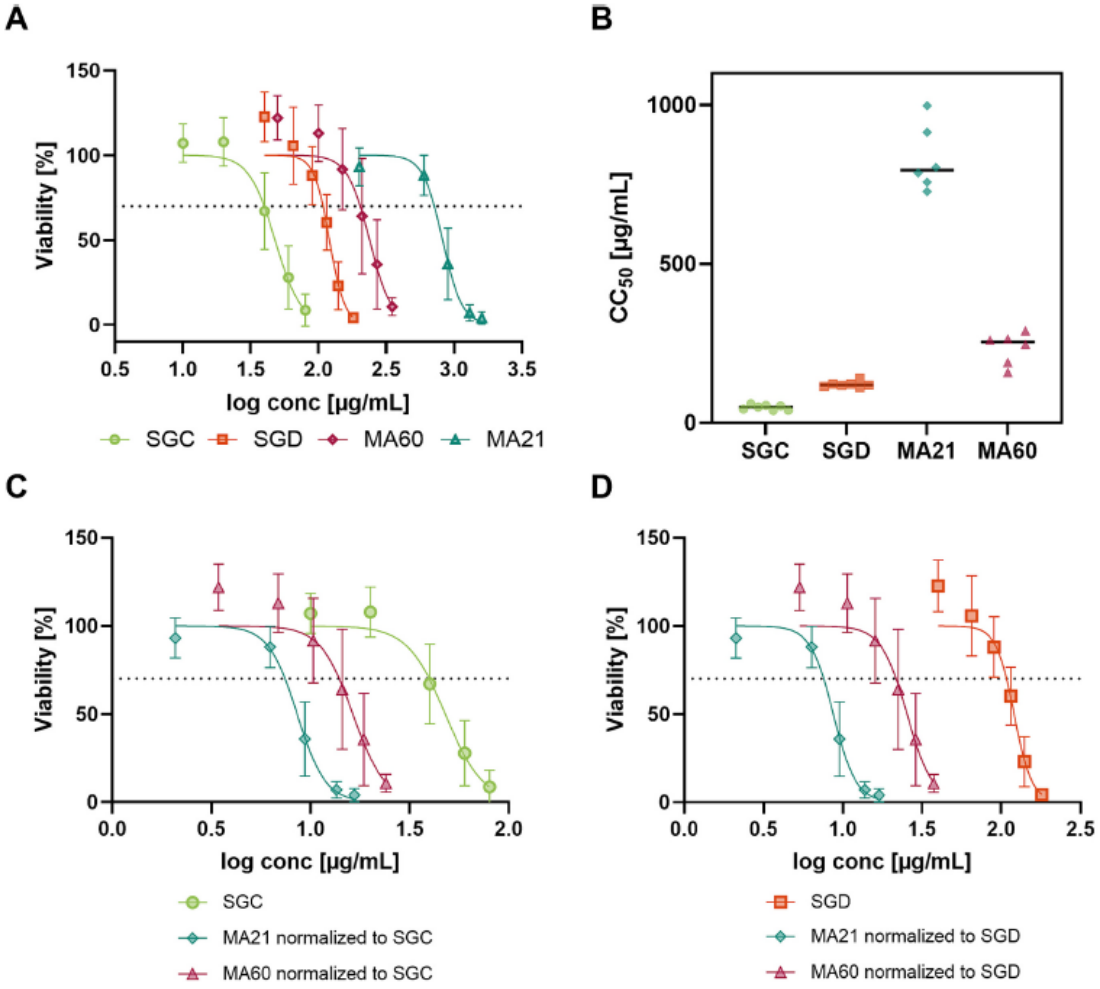


Fig. 3. A) Calu-3 cell viability after 24 h in the presence of sanggenon C (SGC), sanggenon D (SGD), MA21 and MA60. Reduction under 70% of viability was considered cytotoxic (dotted line). B) CC_{50} values of SGC, SGD, MA21 and MA60. MA21 and MA60 cell viability data was also plotted against the respective concentrations of (C) SGC and (D) SGD in each extract mixture to assess whether SGC or SGD were the main cytotoxic constituents. Data is expressed as the mean ± SD of $n = 6$ separate experiments with different passage numbers.

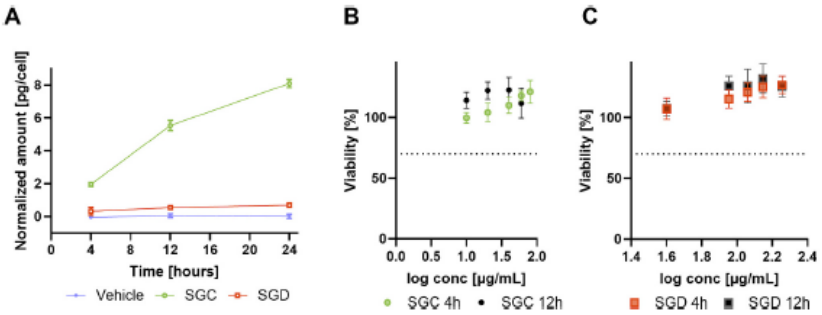


Fig. 4. A) Cellular uptake of SGC and SGD in Calu-3 cells. Cell monolayers were incubated with 50 µg/mL compound or vehicle (0.25% DMSO) and cell extracts were generated at 4, 12 and 24 h after washing cell layers with ammonium carbonate solution (75 mM, pH 7.4, 37 °C) to remove residual medium. SGC and SGD contents were additionally normalized to the number of cells, obtained by cell counting. Data points represent the mean ± standard deviation of three replicates. Calu-3 cell viability after 4 h and 12 h treatment with sanggenon C (SGC) (B) and sanggenon D (SGD) (C). Reduction under 70% of viability was considered cytotoxic (dotted line).

permeable enough to reach their targets, but at the same time do not diffuse too quickly from the lung into the plasma, where they are less effective and can cause systemic side effects (Brillault and Tewes, 2020; Guo et al., 2021; Hastedt et al., 2022; Zhou et al., 2015). Fang and colleagues analysed the relationship between flavonoid structure and the ability to permeate through a Caco-2 cell layer and found that passive diffusion is the major transport mechanism for this compound class. Moreover, they described a broad range of apparent permeability coefficients ($0.29 \pm 0.01 \times 10^{-6}$ cm/s to $33.90 \pm 3.55 \times 10^{-6}$ cm/s in direction A-B, and $0.23 \pm 0.03 \times 10^{-6}$ cm/s to $42.19 \pm 3.11 \times 10^{-6}$ cm/s in direction B-A) for different tested flavonoids such as myricetin, daidzein, and kaempferide (Fang et al., 2017). In many cases, a low permeability of flavonoids has been reported (Zhao et al., 2019). Interestingly, PK synergies between coexisting constituents of natural product extracts have been reported to increase or decrease the permeability of plant constituents (Machado et al., 2020). For example, improved permeability behaviour was found for the flavonoid baicalin when present as constituent of the root extract of *Angelica dahurica* (Zhu et al., 2013). In contrast, a decreased permeability has been reported for alkaloids from Ephedrae herba in different plant extracts on the transport compared to pure alkaloid compounds (Zheng et al., 2015). In literature, there are also examples described with no changes in permeability due to matrix effects; e.g., no influence of the *Prunella vulgaris* extract matrix was observed for the permeability of rosmarinic acid in Caco-2 cells. Also, ursolic acid, a pentycyclic triterpene, showed no difference in permeability behaviour in the same cell line, when comparing the results obtained from the isolated compound and as constituent in an extract of *Salvia officinalis* (Qiang et al., 2011). The different examples demonstrate that it is not possible to extrapolate PK synergy in natural products, thus making experimental investigations indispensable.

The permeability evaluation of SGC, SGD, MA21, and MA60 was conducted bidirectionally across a Calu-3 monolayer. For this setup (Fig. 5A), the test sample was added to either the apical chamber (in the A-B direction) or the basolateral chamber (in the B-A direction) at subtoxic concentrations: SGC (20 µg/mL), SGD (20 µg/mL), MA21 (600 µg/mL), and MA60 (100 µg/mL). The concentration of permeated SGC or SGD was subsequently determined in the acceptor compartment after a 2-h incubation period. As a routine control, a small volume of donor solution was taken immediately after addition to the donor compartment ($t = 0$ h). We observed an unexpected reduction in the SGC and SGD concentrations in the donor compartment in our t = 0 h control samples (Fig. 5B). The reduction in concentration was more pronounced for SGC than SGD, possibly indicating the binding of the more hydrophobic compound either to residual proteins in the well from the serum-supplemented medium or to the plastic components of the wells themselves. Control experiments showed no indication that SGC or SGD bind to serum proteins. However, both SGC and SGD might interact with certain plastic materials (Fig. 5C-D). As neither the molecular structures nor the logP values of the sanggenons were indicative of a high adsorptive behaviour of these compounds to different surfaces, this observation is important, especially for the design of future studies with this class of compounds.

Despite the significant reduction of the free compounds in the donor chamber at t = 0 h, the remaining concentration of free drug was sufficiently high to maintain sink conditions and be detectable in small amounts above the limit of quantitation. Metoprolol (500 µM) served as a reference compound to distinguish high from low permeability and ensure consistency of results with the literature (International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH Harmonised Guideline: Biopharmaceutics Classification System-Based Biowaiver (M9), 2019; Shah and Amidon, 2014; Zur et al., 2014). Bosquillon et al. determined the P_{app} for metoprolol (10 µM in HBSS) in Calu-3 cells with $10.3 \pm 0.0 \times 10^{-6}$ cm/s, in 16HBE14o- cells with $24.5 \pm 2.5 \times 10^{-6}$ cm/s and in NHBE cells with $11.3 \pm 0.7 \times 10^{-6}$ cm/s (Bosquillon et al., 2017). The mean P_{app} values determined for metoprolol in this study were $25.4 \pm 6.9 \times 10^{-6}$ cm/s (A-B) and $23.3 \pm 4.9 \times 10^{-6}$ cm/s (B-A), indicating a good correspondence with the

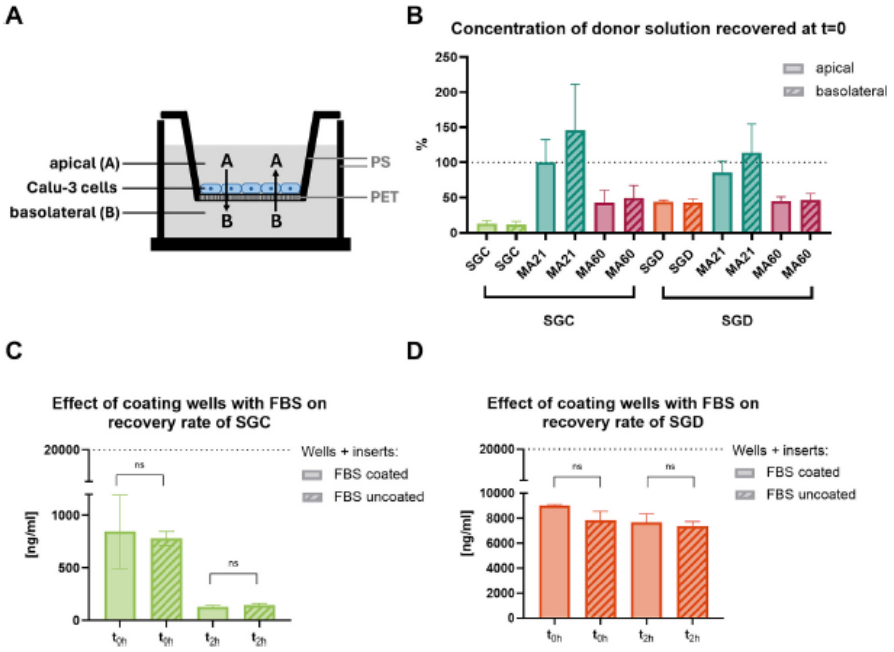


Fig. 5. A) Shows the permeability setup. Calu-3 cells are cultured on a membrane, thereby creating a barrier between the acceptor and donor chamber (well plate insert: membrane material: polyethylene therephthalate (PET), frame material: polystyrene (PS) and well plates: PS). B) Recovered SGC and SGD concentrations in the donor compartment in the t = 0 h control samples. Control experiment for serum-binding evaluated for sanggenon C (SGC; C) and sanggenon D (SGD; D). The dotted line represents the applied concentration. Values represent the mean ± standard deviation of $n = 3$ technical replicates.

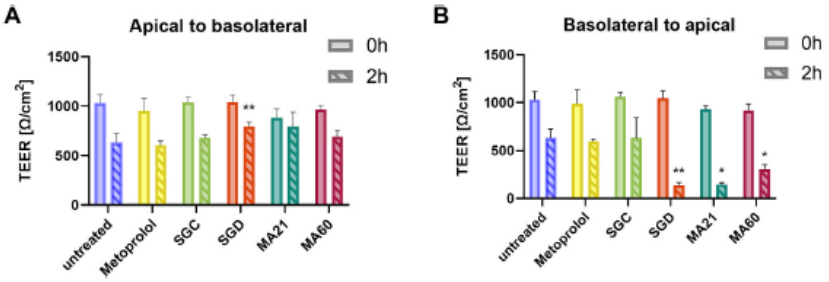


Fig. 6. Changes in TEER before and after permeability study in apical to the basolateral direction (A) and basolateral to the apical direction (B) expressed as the mean ± S.D. (n = 3). The statistical analysis was performed using Mann-Whitney *U* test. Comparisons were made between the untreated groups (2 h) and the treated groups (2 h).

literature. The slightly higher P_{app} values measured in our study are likely due to higher initial metoprolol concentrations used compared to studies from the literature.

Interestingly, SGC and SGD could not be detected in the acceptor chamber after 2 h of incubation (A-B direction). Neither by applying the isolated compounds nor the extracts MA60 and MA21, any evidence for the permeability of SGC or SGD was detectable. Small amounts of SGD (<0.6%) were detected in the acceptor chamber when the B-A permeability was investigated after applying either (i) isolated SGD = 35.4 ± 15.1 ng/mL, (ii) MA21 = 7.9 ± 4.2 ng/mL, or (iii) MA60 = 10.6 ± 5.7 ng/mL (n = 3 experiments with different passage numbers). However, this minor increase in B-A permeability also coincided with a loss in TEER only observed when SGD was added to the basolateral chamber (Fig. 6A-B). The observed decrease in TEER during experiments may result from a loss of epithelial integrity. Cellular damage can be ruled out, as a subsequent MTT assay revealed no cytotoxic effects. Additionally, TEER measurements taken after 24 h of incubating with medium at 37 °C, showed recovery to initial values, suggesting a temporary integrity loss. Control experiments demonstrated that the TEER decrease was specific to this particular batch of Calu-3 cells as it could not be repeated with an independent batch of Calu-3 cells (Fig. S1, Supporting Information). Despite the TEER decrease, permeation of the non-permeation marker, fluorescein, did not increase in these monolayers (Supporting Information). A decrease in TEER exclusively in the B–A direction is rare but has been reported in Calu-3 cells as a property of the mycotoxin, deoxynivalenol (Diesing et al., 2011; Manford et al., 2005) and in 16HBE14o- human bronchial epithelial cell line possibly due to the influence of DMSO (Manford et al., 2005).

3.4. Implications of this study for pulmonary delivery of sanggenons and beyond

The relevance of this interdisciplinary study lies in the recognition that biopharmaceutical profiling and in-depth in vitro investigations of PK synergy can provide important insights into (i) whether the purified bioactive compound or the parent extract will perform more favourably in vivo and (ii) which administration route is likely to provide optimal therapeutic activity. This information will help to inform the planning of in vivo proof-of-concept efficacy studies and formulation strategies. In the case of the two isolated MDAA stereoisomers studied here, SGC and SGD, the rare observation was made that the isolated bioactive compounds exhibited more favourable properties compared to the parent extracts (MA21 and MA60), for instance in terms of improved solubility. It is important to note that despite the high structural similarity of the herein investigated stereoisomers, SGD exhibited a 10-fold higher solubility in aqueous medium than SGC at all four tested pH values. This difference in solubility behaviour between stereoisomers is currently not easily predicted by in silico models and thus emphasises the need for experimental studies, such as those described here.

Therapeutic strategies for SGC and SGD targeting the respiratory tract may hold high potential for both preventive and acute antiviral and antibacterial treatments and associated superinfections. This could include antibacterial effects due to the inhibition of biofilm formation, circumvention of virus release from the cells after viral infection and prevention of secondary infections due to neuraminidase inhibition (Grienke et al., 2016; Wasilewicz et al., 2023). Based on the low uptake and permeability in Calu-3 cells, the tested MDAAs are expected to be able to exert their inhibitory effect on neuraminidases of *Streptococcus pneumoniae* and influenza strains, and potentially also in a dual manner, in the extracellular space (Grienke et al., 2016). Also, the inhibition of planktonic growth and biofilm formation of *S. pneumoniae* is estimated as possible due to low permeability and uptake. On the other hand, the previously determined anti-SARS-CoV-2 activities of SGC and SGD (IC_{50} values of 7.2 µM and 13.9 µM, respectively, (Wasilewicz et al., 2023)), which have been reported to be based on inhibition of the viral main protease, could be constrained since the enzyme is localized within the cell (Fan et al., 2004; Wasilewicz et al., 2023).

The results of this study are currently being utilised to plan future in vivo proof-of-concept studies in a murine infection model using intra-pulmonary delivery of SGC and SGD formulations. Since the efficacy of anti-infective agents is a complex interplay between PK behaviour at the infection site and efficacy parameters such as the minimal inhibitory concentration (MIC) or minimum biofilm inhibitory concentration (MBIC) it is hypothesized that the possibility of high lung retention of sanggenons, due to their low permeability properties, will be highly favourable for therapeutic efficacy. To ensure the application of the highest possible concentrations, optimized formulation of SGC and SGD will be necessary. Future formulation strategies for inhaled sanggenons should focus on both dry powders and liquid formulations for nebulization.

4. Conclusion

Our study showed that biopharmaceutical profiling is a highly useful, albeit currently under-utilised strategy to evaluate the suitability of isolated natural compounds and multi-component mixtures for their intended therapeutic applications. In this study, we were able to show that the two MDAAs, SGC and SGD, demonstrate more favourable PK properties, including e.g. higher solubility as isolated compounds rather than as constituents of multi-component extracts. Furthermore, the low permeability behaviour of SGC and SGD suggests a high potential for lung delivery via inhalation administration to treat acute respiratory tract infections. Ongoing research, including planned in vivo proof-of-concept studies using inhalation administration will show whether the PK advantage of lung delivery for poorly permeable anti-infectives proves therapeutically relevant.

CRediT authorship contribution statement

Jacqueline Schwarzinger: Investigation, Methodology, Visualization, Writing – original draft. **Sigrid Adelsberger:** Investigation, Methodology, Visualization, Writing – original draft. **Karin Ortmayr:** Investigation, Methodology, Visualization. **Sarah Luise Stellnberger:** Investigation, Methodology, Writing – review & editing. **Ammar Tahir:** Methodology. **Gabriela Hädrich:** Investigation, Methodology, Supervision, Writing – review & editing. **Verena Pichler:** Methodology, Supervision, Writing – review & editing. **Judith M. Rollinger:** Funding acquisition, Supervision, Writing – review & editing. **Ulrike Grienke:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing. **Lea Ann Dailey:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare no competing financial interest.

Data availability

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

Acknowledgments

The authors thank Prof. Nicola Zamboni and Prof. Uwe Sauer for providing access to the FIA-TOFMS platform at the Institute of Molecular Systems Biology, ETH Zürich, Switzerland. This research was funded in whole or in part by the Austrian Science Fund (FWF) [10.55776/P35115; 10.55776/P34028].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijpx.2024.100272.

References

Alsenz, J., Kansy, M., 2007. High throughput solubility measurement in drug discovery and development. Adv. Drug Deliv. Rev. 59 (7), 546–567. https://doi.org/10.1016/j.addr.2007.05.007.

Arnott, J.A., Planey, S.L., 2012. The influence of lipophilicity in drug discovery and design. Expert Opin. Drug Discov. 7 (10), 863–875. https://doi.org/10.1517/17460441.2012.714363.

Batiha, G.E.S., Al-Snafi, A.E., Thuwaini, M.M., Teibo, J.O., Shaheen, H.M., Akomolafe, A. P., Teibo, T.K.A., Al-kuraishy, H.M., Al-Garbeeb, A.I., Alexiou, A., Papadakis, M., 2023. Morus alba: a comprehensive phytochemical and pharmacological review. Naunyn-Schmiedeberg's Arch. Pharmacol. 396 (7), 1399–1413. https://doi.org/10.1007/s00210-023-02434-4.

Bergström, C.A.S., Luthman, K., Artursson, P., 2004. Accuracy of calculated pH-dependent aqueous drug solubility. Eur. J. Pharm. Sci. 22 (5), 387–398. https://doi.org/10.1016/j.ejps.2004.04.006.

Berridge, M.V., Tan, A.S., 1993. Characterization of the cellular reduction of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT): subcellular localization, substrate dependence, and involvement of mitochondrial electron transport in MTT reduction. Arch. Biochem. Biophys. 303 (2), 474–482. https://doi.org/10.1006/abbi.1993.1311.

Blume, H.H., Schug, B.S., 2000. Biopharmaceutical characterisation of herbal medicinal products: are in vivo studies necessary? Eur. J. Drug Metab. Pharmacokinet. 25, 41–48. https://doi.org/10.1007/BF03190057.

Bosquillon, C., Madlova, M., Patel, N., Clear, N., Forbes, B., 2017. A comparison of drug transport in pulmonary absorption models: isolated perfused rat lungs, respiratory epithelial cell lines and primary cell culture. Pharm. Res. 34 (12), 2532–2540. https://doi.org/10.1007/s11095-017-2251-y.

Brillault, J., Tewes, F., 2020. Control of the lung residence time of highly permeablemolecules after nebulization: example of the fluoroquinolones. Pharmaceutics 12 (4). https://doi.org/10.3390/pharmaceutics12040387.

Chan, E.W.C., Lye, P.Y., Wong, S.K., 2016. Phytochemistry, pharmacology, and clinical trials of Morus alba. Chin. J. Nat. Med. 14 (1), 17–30. https://doi.org/10.3724/SP.J.1009.2016.00017.

Chebil, L., Humeau, C., Anthoni, J., Dehez, F., Engasser, J.-M., Ghoul, M., 2007. Solubility of flavonoids in organic solvents. J. Chem. Eng. Data 52 (5), 1552–1556. https://doi.org/10.1021/je7001094.

Diesing, A.K., Nossol, C., Dänicke, S., Walk, N., Post, A., Kahlert, S., Rothkötter, H.J., Kluess, J., 2011. Vulnerability of polarised intestinal porcine epithelial cells to mycotoxin deoxynivalenol depends on the route of application. PLoS One 6 (2). https://doi.org/10.1371/journal.pone.0017472.

Dressman, J.B., Berardi, R.R., Dermentzoglou, L.C., Russell, T.L., Schmaltz, S.P., Barnett, J.L., Jarvenpaa, K.M., 1990. Upper gastrointestinal (GI) pH in young, healthy men and women. Pharm. Res. 7 (7), 756–761. https://doi.org/10.1023/A:1015827908309.

Effros, R.M., Chinard, F.P., 1969. The in vivo pH of the extravascular space of the lung. J. Clin. Invest. 48 (11), 1983–1996. https://doi.org/10.1172/JCI106164.

Fan, K., Wei, P., Feng, Q., Chen, S., Huang, C., Ma, L., Lai, B., Pei, J., Liu, Y., Chen, J., Lai, L., 2004. Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. J. Biol. Chem. 279 (3), 1637–1642. https://doi.org/10.1074/jbc.M310875200.

Fang, Y., Cao, W., Xia, M., Pan, S., Xu, X., 2017. Study of structure and permeability relationship of flavonoids in caco-2 cells. Nutrients 9 (12). https://doi.org/10.3390/nu9121301.

Forbes, B., Ehrhardt, C., 2005. Human respiratory epithelial cell culture for drug delivery applications. Eur. J. Pharm. Biopharm. 60 (2), 193–205. https://doi.org/10.1016/j.ejpb.2005.02.010.

Fuguet, E., Ràfols, C., Mañé, M., Ruiz, R., Bosch, E., 2023. Acidity constants of hydroxyl groups placed in several flavonoids: two flavanones, two flavones and five flavonols. Talanta 253, 124096. https://doi.org/10.1016/J.TALANTA.2022.124096.

Fuhrer, T., Heer, D., Begemann, B., Zamboni, N., 2011. High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection–time-of-flight mass spectrometry. Anal. Chem. 83 (18), 7074–7080. https://doi.org/10.1021/ac201267k.

Galindo Bedor, D.C., Marchand, S., Lamarche, I., Laroche, J., Pereira De Santana, D., Couet, W., 2016. Biopharmaceutical characterization of nebulized antimicrobial agents in rats: 5. Oseltamivir carboxylate. Antimicrob. Agents Chemother. 60 (8), 5085–5087. https://doi.org/10.1128/AAC.00909-16.

Gleeson, M.P., 2008. Generation of a set of simple, interpretable ADMET rules of thumb. J. Med. Chem. 51 (4), 817–834. https://doi.org/10.1021/jm701122q.

Glomme, A., März, J., Dressman, J.B., 2005. Comparison of a miniaturized shake-flask solubility method with automated potentiometric acid/base titrations and calculated solubilities. J. Pharm. Sci. 94 (1), 1–16. https://doi.org/10.1002/jps.20212.

Gontijo, A.V.L., Brillault, J., Grégoire, N., Lamarche, I., Gobin, P., Couet, W., Marchand, S., 2014a. Biopharmaceutical characterization of nebulized antimicrobial agents in rats: 1. Ciprofloxacin, moxifloxacin, and grepafloxacin. Antimicrob. Agents Chemother. 58 (7), 3942–3949. https://doi.org/10.1128/AAC.02818-14.

Gontijo, A.V.L., Grégoire, N., Lamarche, I., Gobin, P., Couet, W., Marchand, S., 2014b. Biopharmaceutical characterization of nebulized antimicrobial agents in rats: 2. Colistin. Antimicrob. Agents Chemother. 58 (7), 3950–3956. https://doi.org/10.1128/AAC.02819-14.

Grienke, U., Richter, M., Walther, E., Hoffmann, A., Kirchmair, J., Makarov, V., Nietzsche, S., Schmidtke, M., Rollinger, J.M., 2016. Discovery of prenylated flavonoids with dual activity against influenza virus and *Streptococcus pneumoniae*. Sci. Rep. 6 https://doi.org/10.1038/SREP27156.

Guo, Y., Bera, H., Shi, C., Zhang, L., Cun, D., Yang, M., 2021. Pharmaceutical strategies to extend pulmonary exposure of inhaled medicines. Acta Pharm. Sin. B 11 (8), 2565–2584. Chinese Academy of Medical Sciences. https://doi.org/10.1016/j.apsb.2021.05.015.

Hastedt, J.E., Bäckman, P., Cabal, A., Clark, A., Ehrhardt, C., Forbes, B., Hickey, A.J., Hochhaus, G., Jiang, W., Kassinos, S., Kuehl, P.J., Prime, D., Son, Y.J., Teague, S., Tehler, U., Wylie, J., 2022. IBCS: 1. Principles and Framework of an Inhalation-Based Biopharmaceutics Classification System. Mol. Pharm. 19 (7), 2032–2039. https://doi.org/10.1021/acs.molpharmaceut.2c00113.

Hengjumrut, P., Anukunwithaya, T., Tantisira, M.H., Tantisira, B., Khemawoot, P., 2018. Comparative pharmacokinetics between madecassoside and asiaticoside presented in a standardised extract of *Centella asiatica*, ECa 233 and their respective pure compound given separately in rats. Xenobiotica 48 (1), 18–27. https://doi.org/10.1080/00498254.2016.1273562.

Hughes, J.D., Blagg, J., Price, D.A., Bailey, S., DeCrescenzo, G.A., Devraj, R.V., Ellsworth, E., Fobian, Y.M., Gibbs, M.E., Gilles, R.W., Greene, N., Huang, E., Krieger-Burke, T., Loesel, J., Wager, T., Whiteley, L., Zhang, Y., 2008. Physiochemical drug properties associated with in vivo toxicological outcomes. Bioorg. Med. Chem. Lett. 18 (17), 4872–4875. https://doi.org/10.1016/j.bmcl.2008.07.071.

International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, 2019. ICH Harmonised Guideline: Biopharmaceutics Classification System-Based Biowaiver (M9). https://database.ich.org/sites/default/files/M9_Guideline_Step4_2019_1116.pdf (Accessed 29 July 2024).

Işik, M., Levorse, D., Rustenburg, A.S., Ndukwe, I.E., Wang, H., Wang, X., Reibarkh, M., Martin, G.E., Makarov, A.A., Mobley, D.L., Rhodes, T., Chodera, J.D., 2018. pKa measurements for the SAMPL6 prediction challenge for a set of kinase inhibitor-like fragments. J. Comput. Aided Mol. Des. 32 (10), 1117–1138. https://doi.org/10.1007/s10822-018-0168-0.

Joo, K.M., Lee, J.H., Jeon, H.Y., Park, C.W., Hong, D.K., Jeong, H.J., Lee, S.J., Lee, S.Y., Lim, K.M., 2010. Pharmacokinetic study of ginsenoside Re with pure ginsenoside Re and ginseng berry extracts in mouse using ultra performance liquid chromatography/mass spectrometric method. J. Pharm. Biomed. Anal. 51 (1), 278–283. https://doi.org/10.1016/j.jpba.2009.08.013.

Jürgenliemk, G, Nahrstedt, A., 2003. Dissolution, solubility and cooperativity of phenolic compounds from *Hypericum perforatum* L. in aqueous systems. Pharmazie 58 (3), 200–203. https://www.pnas.org/doi/10.1073/pnas.93.9.4284.

Kerns, E.H., Di, L., 2003. Pharmaceutical profiling in drug discovery. Drug Discov. Today 8 (7), 316–323. https://doi.org/10.1016/S1359-6446(03)02649-7.

Keung, W.-M., Lazo, O., Kunze, L., Vallee, B.L., 1996. Potentiation of the bioavailability ofdaidzin by an extract of Radix puerariae (antidipsotropic activity/dose-responsecurve). Biochemistry 93 (9), 4284–4288. https://www.pnas.org/doi/10.1073/pnas.93.9.4284.

Kornievskaya, V.S., Kruppa, A.I., Polyakov, N.E., Leshina, T.V., 2007. Effect of glycyrrhizic acid on lappaconitine phototransformation. J. Phys. Chem. B 111 (39), 11447–11452. https://doi.org/10.1021/jp0739770.

Langeder, J., Döring, K., Schmietendorf, H., Grienke, U., Schmidtke, M., Rollinger, J.M., 2023a. 1H NMR-based biochemometric analysis of *Morus alba* extracts toward a multipotent herbal anti-infective. J. Nat. Prod. 86 (1), 8–17. https://doi.org/10.1021/acs.jnatprod.2c00481.

Langeder, J., Koch, M., Schmietendorf, H., Tahir, A., Grienke, U., Rollinger, J.M., Schmidtke, M., 2023b. Correlation of bioactive marker compounds of an orally applied *Morus alba* root bark extract with toxicity and efficacy in BALB/c mice. Front. Pharmacol. 14 https://doi.org/10.3389/fphar.2023.1193118.

Leo, A., Hansch, C., Elkins, D., 1971. Partition coefficients and their uses. Chem. Rev. 71 (6), 525–616. https://doi.org/10.1021/cr60274a001.

Li, C., Lin, G., Zuo, Z., 2011. Pharmacological effects and pharmacokinetics properties of Radix Scutellariae and its bioactive flavones. Biopharm. Drug Dispos. 32 (8), 427–445. https://doi.org/10.1002/bdd.771.

Lim, H.J., Jin, H.G., Woo, E.R., Lee, S.K., Kim, H.P., 2013. The root barks of *Morus alba* and the flavonoid constituents inhibit airway inflammation. J. Ethnopharmacol. 149 (1), 169–175. https://doi.org/10.1016/j.jep.2013.06.017.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J., 1997. Experimental andcomputational approaches to estimate solubility and permeability in drug discoveryand development settings. Adv. Drug Deliv. Rev. 23 (1–3), 3–25. https://doi.org/10.1016/S0169-409X(96)00423-1.

Luo, S.-Y., Tang, Z.-Y., Li, Q., Weng, J., Yin, S., Tang, G.-H., 2021. Total synthesis of Mulberry Diels–Alder-type adducts Kuwanons G and H. J. Org. Chem. 86 (6), 4786–4793. https://doi.org/10.1021/acs.joc.1c00229.

Luo, S.Y., Zhu, J.Y., Zou, M.F., Yin, S., Tang, G.H., 2022. Mulberry Diels–Alder-type adducts: isolation, structure, bioactivity, and synthesis. Nat. Prod. Bioprospect. 12 (1) https://doi.org/10.1007/s13659-022-00355-y.

Ma, B.L., Yin, C., Zhang, B.K., Dai, Y., Jia, Y.Q., Yang, Y., Li, Q., Shi, R., Wang, T.M., Wu, J.S., Li, Y.Y., Lin, G., Ma, Y.M., 2016. Naturally occurring proteinaceous nanoparticles in Coptidis Rhizoma extract act as concentration-dependent carriers that facilitate berberine absorption. Sci. Rep. 6 https://doi.org/10.1038/srep20110.

Machado, R.D., de Morais, M.C., da Conceição, E.C., Vaz, B.G., Chaves, A.R., Rezende, K. R., 2020. Crude plant extract versus single compounds for vitiligo treatment: ex vivo intestinal permeability assessment on Brosimum gaudichaudii Trécul. J. Pharm. Biomed. Anal. 191 https://doi.org/10.1016/j.jpba.2020.113593.

Manallack, D.T., Prankerd, R.J., Yuriev, E., Oprea, T.I., Chalmers, D.K., 2013. The significance of acid/base properties in drug discovery. Chem. Soc. Rev. 42 (2), 485–496. https://doi.org/10.1039/c2cs35348b.

Manford, F., Tronde, A., Jeppsson, A.B., Patel, N., Johansson, F., Forbes, B., 2005. Drug permeability in 16HBE14o- airway cell layers correlates with absorption from the isolated perfused rat lung. Eur. J. Pharm. Sci. 26 (5), 414–420. https://doi.org/10.1016/j.ejps.2005.07.010.

Marchand, S., Boisson, M., Mehta, S., Adier, C., Mimoz, O., Grégoire, N., Couet, W., 2018. Biopharmaceutical characterization of nebulized antimicrobial agents in rats: 6. Aminoglycosides. Antimicrob. Agents Chemother. 62, e01261–18. https://doi.org/10.1128/AAC.01261-18.

Marchand, S., Gobin, P., Brillault, J., Baptista, S., Adier, C., Olivier, J.C., Mimoz, O., Couet, W., 2010. Aerosol therapy with colistin methanesulfonate: a biopharmaceutical issue illustrated in rats. Antimicrob. Agents Chemother. 54 (9), 3702–3707. https://doi.org/10.1128/AAC.00411-10.

Marchand, S., Grégoire, N., Brillault, J., Lamarche, I., Gobin, P., Couet, W., 2015. Biopharmaceutical characterization of nebulized antimicrobial agents in rats: 3. Tobramycin. Antimicrob. Agents Chemother. 59 (10), 6646–6647. https://doi.org/10.1128/AAC.01647-15.

Marchand, S., Grégoire, N., Brillault, J., Lamarche, I., Gobin, P., Couet, W., 2016. Biopharmaceutical characterization of nebulized antimicrobial agents in rats. 4. Aztreonam. Antimicrob. Agents Chemother. 60 (5), 3196–3198. https://doi.org/10.1128/AAC.00165-16.

Mathias, N.R., Timoszyk, J., Stetsko, P.I., Megill, J.R., Smith, R.L., Wall, D.A., 2002. Permeability characteristics of Calu-3 human bronchial epithelial cells: in vitro - in vivo correlation to predict lung absorption in rats. J. Drug Target. 10 (1), 31–40. https://doi.org/10.1080/10611860290007504.

Navo, C.D., Jiménez-Osés, G., 2021. Computer prediction of p K_a values in small molecules and proteins. ACS Med. Chem. Lett. 12 (11), 1624–1628. American Chemical Society. https://doi.org/10.1021/acsmedchemlett.1c00435.

Prankerd, R.J., 2007. Critical compilation of pKa values for pharmaceutical substances. In: Profiles of Drug Substances, Excipients and Related Methodology, 33, pp. 1–33. https://doi.org/10.1016/S0099-5428(07)33001-3.

Prentis, R., Lis, Y., Walker, S., 1988. Pharmaceutical innovation by the seven UK-owned pharmaceutical companies (1964-1985). Br. J. Clin. Pharmacol. 25 (3), 387–396. https://doi.org/10.1111/j.1365-2125.1988.tb03318.x.

Qiang, Z., Ye, Z., Hauck, C., Murphy, P.A., McCoy, J.A., Widrlechner, M.P., Reddy, M.B., Hendrich, S., 2011. Permeability of rosmarinic acid in *Prunella vulgaris* and ursolic acid in *Salvia officinalis* extracts across Caco-2 cell monolayers. J. Ethnopharmacol. 137 (3), 1107–1112. https://doi.org/10.1016/j.jep.2011.07.037.

Shah, V.P., Amidon, G.L., 2014. G.L. Amidon, H. Lennernas, V.P. Shah, and J.R. Crison. Atheoretical basis for a biopharmaceutic drug classification: the correlation of in vitrodrug product dissolution and in vivo bioavailability, Pharm Res 12, 413-420, 1995-Backstory of BCS. AAPS J. 16 (5), 894–898. https://doi.org/10.1208/s12248-014-9620-9.

Simonian, H.P., Vo, L., Doma, S., Fisher, R.S., Parkman, H.P., 2005. Regional postprandial differences in pH within the stomach and gastroesophageal junction. Dig. Dis. Sci. 50 (12), 2276–2285. https://doi.org/10.1007/s10620-005-3048-0.

Singh, R., Bagchi, A., Bagachi, A., Semwal, A., bharadwaj, A., 2013. Traditional uses, phytochemistry and pharmacology of *Morus alba* Linn.: a review. J. Med. Plant Res. 7 (9), 461–469. https://doi.org/10.5897/JMPR012.1079.

Song, M., Hang, T.J., Zhang, Z., Chen, H.Y., 2007. Effects of the coexisting diterpenoid tanshinones on the pharmacokinetics of cryptotanshinone and tanshinone IIA in rat. Eur. J. Pharm. Sci. 32 (4–5), 247–253. https://doi.org/10.1016/j.ejps.2007.07.007.

Ta, W., Yang, X., Wang, J., Han, C., Hua, R., Lu, W., 2023. Comparison of intestinal absorption characteristics between rhubarb traditional Chinese medicine preparation and activity ingredients using in situ and in vitro model. Chin. Herb. Med. 15 (1), 117–122. https://doi.org/10.1016/j.chmed.2022.09.004.

Testa, B., Crivori, P., Reist, M., Carrupt, P.-A., 2000. The influence of lipophilicity on thepharmacokinetic behavior of drugs: concepts and examples. Perspect. Drug Discov. Design 19, 179–211. https://doi.org/10.1023/A:1008741731244.

Venkatesh, S., Lipper, R.A., 2000. Role of the development scientist in compound lead selection and optimization. J. Pharm. Sci. 89 (2), 145–154. https://doi.org/10.1002/(SICI)1520-6017(200002)89:2<145::AID-JPS2>3.0.CO;2-6.

Veseli, A., Žakelj, S., Kristl, A., 2019. A review of methods for solubility determination in biopharmaceutical drug characterization. Drug Dev. Ind. Pharm. 45 (11), 1717–1724. Taylor and Francis Ltd. https://doi.org/10.1080/03639045.2019.1665062.

Veseli, A., Kristl, A., Žakelj, S., 2020. Proof-of-concept for a miniaturized shake-flask biopharmaceutical solubility determination by sonic mixing. Pharmazie 75 (12), 626–631. https://doi.org/10.1691/ph.2020.0063.

Waring, M.J., Arrowsmith, J., Leach, A.R., Leeson, P.D., Mandrell, S., Owen, R.M., Pairaudeau, G., Pennie, W.D., Pickett, S.D., Wang, J., Wallace, O., Weir, A., 2015. An analysis of the attrition of drug candidates from four major pharmaceutical companies. Nat. Rev. Drug Discov. 14 (7), 475–486. Nature Publishing Group. https://doi.org/10.1038/nrd4609.

Wasilewicz, A., Kirchweger, B., Bojkova, D., Abi Saad, M.J., Langeder, J., Bütikofer, M., Adelsberger, S., Grienke, U., Cinatl, J., Petermann, O., Scapozza, L., Orts, J., Kirchmair, J., Rabenau, H.F., Rollinger, J.M., 2023. Identification of natural products inhibiting SARS-CoV-2 by targeting viral proteases: a combined in silico and in vitro approach. J. Nat. Prod. 86 (2), 264–275. https://doi.org/10.1021/acs.jnatprod.2c00843.

Weathers, P.J., Arsenault, P.R., Covello, P.S., McMickle, A., Teoh, K.H., Reed, D.W., 2011. Artemisinin production in *Artemisia annua*: studies in planta and results of a novel delivery method for treating malaria and other neglected diseases. Phytochem. Rev. 10 (2), 173–183. https://doi.org/10.1007/s11101-010-9166-0.

Wu, H., Zhu, Z., Zhang, G., Zhao, L., Zhang, H., Zhu, D., Chai, Y., 2009. Comparative pharmacokinetic study of paeoniflorin after oral administration of pure paeoniflorin, extract of Cortex Moutan and Shuang-Dan prescription to rats. J. Ethnopharmacol. 125 (3), 444–449. https://doi.org/10.1016/j.jep.2009.07.019.

Xu, M., Wang, G., Xie, H., Huang, Q., Wang, W., Jia, Y., 2007. Pharmacokinetic comparisons of schizandrin after oral administration of schizandrin monomer, Fructus Schisandrae aqueous extract and Sheng-Mai-San to rats. J. Ethnopharmacol. 115 (3), 483–488. https://doi.org/10.1016/j.jep.2007.10.016.

Zhang, L., Yan, R., Su, R., Yang, C., Liu, S., Yu, X., Chang, X., Zhang, S., Liu, C., Xu, M., Zeng, W., Chen, Y., Wang, Q., 2014. Bioavailability enhancement of osthole after oral administration of Bushen Yizhi prescription extract to rats followed by *Cnidium monnieri* (L.) Cusson fruits extract in comparison to pure osthole at different doses. J. Ethnopharmacol. 152 (2), 266–271. https://doi.org/10.1016/j.jep.2013.09.045.

Zhang, W., Zhang, H., Sun, S., Sun, F.F., Chen, J., Zhao, L., Zhang, G.Q., 2015. Comparative pharmacokinetics of hypaconitine after oral administration of pure hypaconitine, *aconitum carmichaelii* extract and sini decoction to rats. Molecules 20 (1), 1560–1570. https://doi.org/10.3390/molecules20011560.

Zhao, J., Yang, J., Xie, Y., 2019. Improvement strategies for the oral bioavailability ofpoorly water-soluble flavonoids: an overview. Int. J. Pharm. 570 https://doi.org/10.1016/j.ijpharm.2019.118642.

Zhao, Q., Luan, X., Zheng, M., Tian, X.H., Zhao, J., Zhang, W.D., Ma, B.L., 2020. Synergistic mechanisms of constituents in herbal extracts during intestinal absorption: focus on natural occurring nanoparticles. Pharmaceutics 12 (2). https://doi.org/10.3390/pharmaceutics12020128.

Zheng, K.Y.Z., Choi, R.C.Y., Guo, A.J.Y., Bi, C.W.C., Zhu, K.Y., Du, C.Y.Q., Zhang, Z.X., Lau, D.T.W., Dong, T.T.X., Tsim, K.W.K., 2012. The membrane permeability of Astragali Radix-derived formononetin and calycosin is increased by Angelicae

Sinensis Radix in Caco-2 cells: a synergistic action of an ancient herbal decoction Danggui Buxue Tang. J. Pharm. Biomed. Anal. 70, 671–679. https://doi.org/10.1016/j.jpba.2012.05.018.

Zheng, M., Zhou, H., Wan, H., Chen, Y.L., He, Y., 2015. Effects of herbal drugs in Mahuang decoction and their main components on intestinal transport characteristics of Ephedra alkaloids evaluated by a Caco-2 cell monolayer model. J. Ethnopharmacol. 164, 22–29. https://doi.org/10.1016/j.jep.2015.01.043.

Zheng, D., Ruan, Y.T., Yin, Z.P., Zhang, Q.F., 2020. A comparison of solubility, stability, and bioavailability between astilbin and neoastilbin isolated from smilax glabra rhizoma. Molecules 25 (20). https://doi.org/10.3390/molecules25204728.

Zhou, Q.T., Leung, S.S.Y., Tang, P., Parumasivam, T., Loh, Z.H., Chan, H.K., 2015. Inhaledformulations and pulmonary drug delivery systems for respiratory infections. Adv. Drug Deliv. Rev. 85, 83–99. https://doi.org/10.1016/j.addr.2014.10.022.

Zhu, M.L., Liang, X.L., Zhao, L.J., Liao, Z.G., Zhao, G.W., Cao, Y.C., Zhang, J., Luo, Y., 2013. Elucidation of the transport mechanism of baicalin and the influence of a Radix Angelicae Dahuricae extract on the absorption of baicalin in a Caco-2 cell monolayer model. J. Ethnopharmacol. 150 (2), 553–559. https://doi.org/10.1016/j.jep.2013.09.011.

Zur, M., Gasparini, M., Wolk, O., Amidon, G.L., Dahan, A., 2014. The Low/high bcs permeability class boundary: physicochemical comparison of metoprolol and labetalol. Mol. Pharm. 11 (5), 1707–1714. https://doi.org/10.1021/mp500152y.

Supporting information

Biopharmaceutical profiling of anti-infective sanggenons from *Morus alba* root bark for inhalation administration

Authors:

Jacqueline Schwarzinger[a,b,#], Sigrid Adelsberger[b,c,#], Karin Ortmayr[c], Sarah Stellnberger[b,d], Ammar Tahir[c], Gabriela Hädrich[a], Verena Pichler[d], Judith M. Rollinger[c], Ulrike Grienke[c,], Lea Ann Dailey[a,*]*

[#] J. Schwarzinger and S. Adelsberger contributed equally to this work

Correspondence:

[*] U. Grienke (ulrike.grienke@univie.ac.at) and L.A. Dailey (leaann.dailey@univie.ac.at)

Affiliations:

[a] *Division of Pharmaceutical Technology and Biopharmaceutics, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria*

[b] *Vienna Doctoral School of Pharmaceutical, Nutritional and Sport Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria*

[c] *Division of Pharmacognosy, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria*

[d] *Division of Pharmaceutical Chemistry, Department of Pharmaceutical Sciences, Faculty of Life Sciences, University of Vienna, Josef-Holaubek-Platz 2, 1090 Vienna, Austria*

Table of contents

1. Sample preparation for quantitation of MDAAs

In the following chapters, preparation steps for the quantitation of samples from (a) pH-dependent solubility profile assays and (b) bidirectional permeability in Calu-3 cells testing series are listed. For method validation of (a), quality control (QC) samples have been used, which represent samples with known concentrations of sanggenon C (SGC) and sanggenon D (SGD) that were processed in the same way as test samples. For the evaluation of (b), sanggenon G (SGG) was used as internal standard (IS) to minimize errors such as changes in detector response and to assess the quality of the measurement process.

1.1. pH-dependent solubility profile

Different protocols for sample preparation of the pH-dependent solubility profile assays are listed, beginning with descriptions for the production of stock and working solutions up to method evaluation.

1.1.1. Preparation of stock and working solutions, calibration standards and QC working samples

Table S1 shows stock and working solutions for calibration curve and QC working samples. Twelve QC working samples containing both SGC and SGD were prepared for low (6.25 µg/mL), middle (25 µg/mL) and high (100 µg/mL) concentration range. For QC blanks, amounts of respective buffer were used. All samples were prepared in the solvent methanol (MeOH).

Table S1. Overview of the preparation of stock and working solutions, calibration standards and QC working samples for the pH-dependent solubility assay.

	sanggenon C (SGC) [µg/mL]	**sanggenon D (SGD) [µg/mL]**
calibration curve standards:		
2 stock solutions	1000	-
	-	1000
1 working solution	400	400
9 standards	0.78	0.78
	1.56	1.56
	3.13	3.13
	6.25	6.25
	12.5	12.5
	25.0	25.0
	50.0	50.0
	100	100
	200	200
QC samples:		
2 stock solutions	1000	-
	-	1000
1 working solution	400	400
12 working samples	6.25	6.25
	25.0	25.0
	100	100
	6.25	6.25
	25.0	25.0
	100	100
	6.25	6.25
	25.0	25.0
	100	100
	6.25	6.25
	25.0	25.0
	100	100

1.1.2. Prearrangement of QC samples

For the prearrangement of QC samples, 80 µL of the respective QC working samples were used, following the protocol described in Table S2.

Table S2. Prearrangement of QC samples

• 80 µL of the QC working sample were pipetted into a 1.5 mL Eppendorf tube
• Evaporation of MeOH via GeneVac
• 80 µL of respective buffer solution were added
• Shaking on a vortex shaker for 60 min
• Ultrasonication for 10 min including cooling of the bath with ice pads
• Centrifugation in a mini centrifuge for 30 s with 5400 rpm
• Neutralization with 0.1% aqueous ammonia solution depending on the respective buffer - 90 µL for buffer pH 1.2 - 80 µL for buffer pH 4.5 - 30 µL for buffer pH 6.8 - sample with buffer pH 7.4 were not manipulated
• Vortexing for 10 s
• Centrifugation for 30 s with 5400 rpm

1.1.3. Preparation of solubility samples

Table S3 describes the preparation protocol for solubility samples (volume 80 µL). For setting up of blanks, 80 µL of respective buffer solution in a 1.5 mL Eppendorf tube were processed in the same way, beginning with neutralization. After these steps, the application of the MTBE (methyl tertiary butyl ether)-based extraction protocol was possible.

Table S3. Preparation of solubility samples and blanks.

• Centrifugation for 15 s with 5400 rpm
• Neutralization with 0.1% aqueous ammonia solution depending on the respective buffer - 90 µL for buffer pH 1.2 - 80 µL for buffer pH 4.5 - 30 µL for buffer pH 6.8 - sample with buffer pH 7.4 were not manipulated
• Vortexing for 10 s
• Evaporation of MeOH

1.1.4. MTBE-based extraction protocol and completion of sample preparation

The MTBE-based extraction protocol was implemented as a three-cycle procedure with 3 x 750 µL as extraction solvent and can be seen in Table S4. This procedure was followed for all samples, blanks, QC samples and QC blanks.

Table S4. Sequence of the MTBE-based extraction protocol.

• Addition of 750 µL MTBE to the 1.5 mL Eppendorf tube containing the sample
• Shaking on a vortex shaker for 60 min
• Ultrasonication for 10 min including cooling of the bath with ice pads
• Centrifugation for 30 s with 5400 rpm
• Removal (after pipetting up and down 5 times) and transfer of 650 µL supernatant into a second 1.5 mL Eppendorf collector tube
• Removal of the solvent in both Eppendorf tubes via sample concentrator
• Start of next cycle with addition of 750 µL MTBE or storage of the sample at -20 °C after cycle 3

Shortly before the measurement, the completion of sample preparation was done in accordance with the protocol in Table S5.

Table S5. Completion of sample preparation before the measurement.

• Thawing of the sample at room temperature
• Addition of 80 µL MeOH
• Vortexing for 10 s
• Shaking for 10 min
• Ultrasonication for 10 min
• Centrifugation for 30 s with 5400 rpm
• Pipetting of the sample into a 300 µL HPLC vial insert

1.1.5. Evaluation of the MTBE-based extraction protocol

The MTBE-based extraction protocol was the basis for the preparation of all samples of the pH-dependent solubility assay. Therefore, the method was validated, including 12 different QC combinations of pH values (1.2, 4.5, 6.8, 7.4) and concentrations (6.25, 25 and 100 µg/mL), available in 3, 4 or 6 replicates. For pH 1.2 in low concentration range only duplicates were definable and drawn upon calculation.

To show the extraction efficiency for this extraction protocol, recovery was calculated as percentage of the known analyte amount for both SGC and SGD. Figure S1 shows the recovery results for both SGC and SGD.

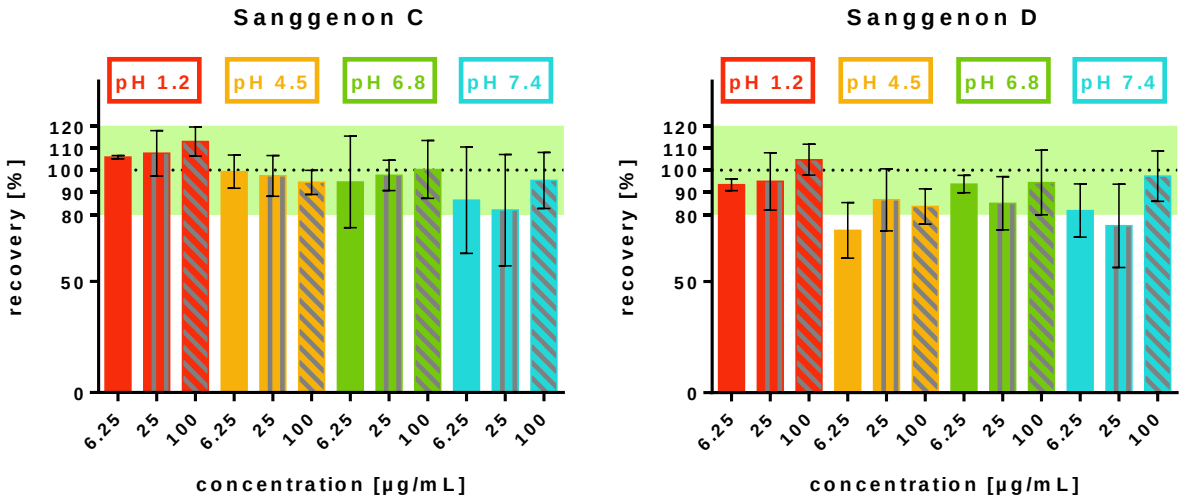


Figure S1. Recovery results for the evaluation of the efficiency of the MTBE-based extraction protocol as part of the pH-dependent solubility assay. Left for sanggenon C (SGC), right for sanggenon D (SGD).

Over all parameters, the results for SGC were higher than for SGD, except for 100 µg/mL at pH 7.4. For compound SGC, the recovery value means were in the acceptable range of ± 15%, with exception of 25 µg/mL at pH 7.4 (18% deviation). For compound SGD the data availability was more complex with good recovery of ± 15% for pH 1.2 and pH 6.8. However, recovery results for pH 4.5 and pH 7.4 for SGD were partly too low. The lowest result was found for pH 4.5 and 6.25 µg/mL (-28% from expected value).

1.1.6. Measurement of solubility samples via UPLC-PDA method

Quantitative analysis measurements were performed on a Waters Acquity UPLC H-class system using a PDA detector at 205 nm wavelength. Parameters of the measurement are listed in Table S6.

Table S6. List of parameters for UPLC-PDA measurement of pH-dependent solubility assay samples.

device	Waters Acquity UPLC H-Class system
detector	PDA, wavelength 205 nm
autosampler temperature [°C]	8
injection volume [µL]	5
mobile phase	double-distilled water with 0.1% FA (solvent A) acetonitrile with 0.1% FA (solvent B)
column	Acquity BEH Phenyl column 2.1 x 100 mm, 1.7 µm
column temperature [°C]	40
flow rate [mL/min]	0.3
gradient program [time (min)/% B]	0/5, 1/50, 6.9/50, 7/98, 8.9/98, 9/5, 10/5
total run time [min]	10 min
equilibration time before next injection [min]	2.0 min
data acquisition and processing	Waters Empower 3

1.1.7. Evaluation of UPLC-PDA quantitation method

Linearity for solubility quantitation method was calculated via regression equation for nine calibration levels (0.78, 1.56, 3.13, 6.25, 12.5, 25, 50, 100 and 200 µg/mL in methanol) for both SGC and SGD. A visually evaluated signal-to-noise ratio higher than 3 times allowed the evaluation of limit of detection (LOD), higher than 10 times the limit of quantitation (LOQ). Intra-day precision was calculated via QC samples, if possible, with n = 6 replicates. The inter-day precision was calculated via repeated measurement of QC samples on three different days. The number of available replicates for the solubility samples ranged from 3–6. Table S7 gives an overview of obtained validation parameters.

Table S7. Validation parameters of the UPLC-PDA quantitation method of the pH-dependent solubility assay.

	sanggenon C (SCG)	**sanggenon D (SGD)**
levels used	9	9
linear range [µg/mL]	0.78–200	0.78–200
R^2	0.9998	0.9983
LOD [µg/mL]	1.56	1.56
LOQ [µg/mL]	3.13	3.13
Precision as standard deviation		
intra-day measurement [%]	0.4–4.1	0.7–3.6
intra-day QC samples [%]	0.1–21.3	0.0–17.1
inter-day QC samples [%]	0.8–25.0	2.8–25.0

1.2. Bidirectional permeability assay in Calu-3 cells

This chapter describes different protocols for sample preparation of the bidirectional permeability assay in Calu-3 cells including an overview of stock and working solutions, sample preparation protocols, measurement parameters and method validation.

1.2.1. Preparation of stock and working solutions and calibration standards

For the quantitation of samples from the bidirectional permeability assay in Calu-3 cells, the preparation of stock solutions containing SGC, SGD and SGG, all in methanol, was necessary. The first two stock solutions have been used for calibration curve dilution series, the latter was used for internal standard (IS) purposes. Table S8 allows an overview of stock and working solutions.

Table S8. Overview of stock and working solutions for the bidirectional permeability assay in Calu-3 cells.

	sanggenon C (SGC) [µg/mL]	**sanggenon D (SGD) [µg/mL]**	**sanggenon G (SGG) [µg/mL]**
calibration curve standards:			
2 stock solutions	1000	-	-
	-	1000	-
2 working solutions	100	100	-
	10	10	-
internal standard (IS):			
1 stock solution	-	-	1000
3 working solutions	-	-	100
	-	-	10
	-	-	0.5

Based on knowledge from pre-trials, concentration range of SGC and SGD in permeability samples was expected to be extensive, from low ng range to high µg range. Therefore, standards for two different calibration curve rows (high concentration, low concentration) were prepared. Both series were spiked with IS (100 ng/mL for low concentration range and 1000 ng/mL for high concentration range).

Table S9. Low and high concentration range calibration standard series of the bidirectional permeability assay in Calu-3 cells.

	sanggenon C (SGC) [ng/mL]	**sanggenon D (SGD) [ng/mL]**	**sanggenon G (SGG) [ng/mL]**
Low concentration range calibration standards			
8 levels for SGC, 6 levels for SGD	3.13	3.13	100
	6.25	6.25	100
	12.5	12.5	100
	25.0	25.0	100
	50.0	50.0	100
	75.0	75.0	100
	150	-	100
	200	-	100
High concentration range calibration standards			
8 levels for SGC, 7 levels for SGD	19.5	19.5	1000
	39.1	39.1	1000
	78.1	78.1	1000
	156.3	156.3	1000
	312.5	312.5	1000
	625	625	1000
	1250	1250	1000
	2500	-	1000

1.2.2. Preparation of permeability samples

For the high concentration range, 10 µL of the homogenized permeability samples were spiked with 10 µL of 10 µg/mL IS in methanol and filled up with 80 µL methanol. Thus, dilution factor for each high range sample was 1:10. Blanks for this concentration range were prepared by use of 10 µL pure HBSS.

Table S10 describes the sample preparation for low concentration range samples and respective blanks. Hence, 200 µL of the sample have been dissolved in 50 µL solvent only, the concentration factor for each sample in the low concentration range was 4:1. To ensure the same matrix conditions as in the permeability samples, the low concentration range calibration curve standards were provided in vials containing residue of 200 µL evaporated HBSS buffer.

Table S10. Preparation protocol for low concentration range permeability samples.

• Addition of 10 µL of 0.5 µg/mL IS in methanol to 200 µL of the sample (or to 200 µL pure HBSS for blanks) available in a 1.5 mL Eppendorf tube
• Vortexing for 10 s
• Evaporation of the solvent via GeneVac device
• Addition of 50 µL of a methanol:double distilled water = 1:1 mixture
• Vortexing for 10 s
• Shaking for 15 min
• Centrifugation for 2 s with short spin function of a mini centrifuge
• Transfer of the sample into a 300 µL HPLC vial insert

1.2.3. Measurement of permeability samples via UHPLC-ESI-MS method

Quantitative analysis measurements were performed on a UHPLC-ESI-MS system (Thermo Fisher Scientific, CA) containing a Dionex UltiMate 3000 coupled to a LTQ XL linear ion trap mass spectrometer. Calibration curves for sanggenon C (SGC) included 8 levels for the low concentration range and 8 levels for the high concentration range; calibration curves for sanggenon D (SGD) were calculated based on 6 levels for the low and 7 levels for the high concentration range (see Table S9). Sanggenon G (SGG) was used as an internal standard for quantitation (ratio analyte/internal standard). Parameters of the measurement can be found in Table S11. The total run time for each sample was 7 min, with a gradient step of 10% solvent A for 1.0 min at the beginning to get rid of buffer constituents, followed by 6 min of actual separation. To avoid buffer constituents transport to the MS detector, the divert valve setting of the LTQ XL linear ion trap device included a transfer of the flow to the waste container in the first 2.41 min and from 5.06 min to 6.52 min. To enhance the sensitivity of the quantitative analysis, single ion monitoring recording was chosen as scanning mode for the MS detector for [M-H]- with two narrow m/z windows: 692.20–694.20 for internal standard SGG and 706.23–708.23 for compounds SGC and SGD.

Table S11. Parameters for UHPLC-ESI-MS measurement of permeability samples.

chromatographic separation and detection	Dionex UltiMate 3000 coupled to a LTQ XL linear ion trap mass spectrometer
autosampler temperature [°C]	12
injection volume [µL]	5
mobile phase	double-distilled water with 0.1% FA (solvent A) acetonitrile with 0.1% FA (solvent B)
flow rate [mL/min]	0.35
gradient program [time (min)/% B]	0/10, 0.99/10, 1/50, 3/75, 3.5/98, 4.5/98, 4.6/50, 7/50
column	Acquity BEH C_{18} column, 2.1 x 50 mm, 1.7 µm
column temperature [°C]	50
total run time [min]	7 min
equilibration time before next injection [min]	1.6 min
ionisation	negative mode
HESI source heater temperature [°C]	400
sheath gas/aux gas/sweep gas flow rates [arbitrary units]	35/20/0.02
source voltage [kV]	2.5
capillary temperature [°C]	400
data acquisition and processing	XCalibur software (version 4.27.0.19)

1.2.4. Measurement of metoprolol containing permeability samples via UHPLC-ESI-MS method

Due to different molecular weight of metoprolol compared with sanggenons, metoprolol containing permeability samples were analyzed in a separate UHPLC-ESI-MS measurement. The same device combination and parameters for permeability samples were used as described in chapter 1.2.3, but with some adaptions that are listed in Table S12.

Table S12. List of method adaptions for metoprolol containing permeability samples compared to sanggenon containing permeability samples.

gradient program [time (min)/% B]	0/10, 1.5/98, 2.3/98, 2.4/10, 4.5/10
column	Kinetex C_{18} column, 150 x 2.1 mm, 1.7 µm, 100 Å
Injection volume [µL]	0.1
total run time [min]	4.5 min
ionisation	positive mode
sheath gas/aux gas/sweep gas flow rates [arbitrary units]	35/15/2
source voltage [kV]	3.5
capillary temperature [°C]	400
data acquisition and processing	XCalibur software (version 4.27.0.19)

Single ion monitoring recording was set to m/z values 266.30–270.20 for [M+H]+. To avoid buffer constituent transport to the MS detector, the divert valve setting was changed to source inlet status from only 2.80 min to 3.50 min.

1.2.5. Validation of UHPLC-ESI-MS method

Validation parameters of the UHPLC-ESI-MS method used for the measurement of MDAA containing samples are shown in Table S13.

Table S13. Validation parameters of the UHPLC-ESI-MS quantitation method of the bidirectional permeability in Calu-3 cell assay samples.

	sanggenon C (SC)	**sanggenon D (SD)**
Low concentration range calibration standards		
levels used	8	6
linear range [ng/mL]	3.13–200	3.13–75.0
R^2	0.9999	0.9989
LOD [ng/mL]	3.13	3.13
LOQ [ng/mL]	6.25	6.25
High concentration range calibration standards		
levels used	8	7
linear range [ng/mL]	19.5–2500	19.5–1250
R^2	0.9992	0.9980
LOD [ng/mL]	9.8	9.8
LOQ [ng/mL]	19.5	19.5

Each permeability sample was analysed in n = 6 replicates (except MA60 in apical compartment, direction A-B at t = 2h with n = 5 replicates) with 2 replicates on 1 well plate. Thus, the 6 replicates were derived from 3 different passage numbers of Calu-3 cells.

1.3. Results of the pH-dependent solubility profile

Results of the pH-dependent solubility profile of pure compounds SGC and SGD and both compounds within extracts MA 21 and MA60 are summarised in Table S14.

Table S14. Solubility of sanggenon C (SGC) and sanggenon D (SGD) as isolated stereoisomers and within the two multicomponent extract mixtures, MA21 and MA60, in four different buffer solutions with pH values of 1.2, 4.5, 6.8 and 7.4.

Matrix	Compound	pH 1.2 solubility (µg/mL)	pH 4.5 solubility (µg/mL)	pH 6.8 solubility (µg/mL)	pH 7.4 solubility (µg/mL)
isolated	SGC	4.47 ± 1.43	8.53 ± 3.66	18.5 ± 6.2	47.1 ± 15.7
MA21	SGC	< LOQ -	< LOQ -	4.05 ± 1.38	3.24 ± 1.21
MA60	SGC	6.88 ± 9.70	< LOQ -	12.7 ± 6.3	13.6 ± 4.1
isolated	SGD	111 ± 18	121 ± 17	304 ± 108	493 ± 161
MA21	SGD	< LOQ -	< LOQ -	< LOQ -	< LOQ -
MA60	SGD	20.6 ± 14.6	19.7 ± 1.3	56.8 ± 18.9	94.9 ± 9.5

2. Permeability control experiments

2.1. Materials and Method

To better understand the cause of the one-directional TEER decrease observed when 20 µg/mL SGD samples were tested for permeation from the basolateral to apical direction, further studies were conducted to investigate this effect and its impact on the results. The experimental setup was as follows: SGC and SGD were added to either the basolateral or apical chambers of the wells or inserts. This time, fluorescein at a concentration of 40 µg/mL was always added to the apical chamber. Fluorescein was included because it is a non-permeability marker, and any slight influence on monolayer integrity would be detected by its rapid permeation through non-intact monolayers to the acceptor chamber. Inserts with and without cells, with the addition of just fluorescein, served as controls. Furthermore, MTT assays were conducted directly on the cell monolayer following the permeability studies to evaluate if cytotoxic processes were present.

The experiment was conducted with two different batches of Calu-3 cells, each from a different LOT, to determine whether the observed phenomenon was general or specific to a particular batch or passage number range. Calu-3 cells from batch 1, also used in the MDAA permeability studies (2.7), were from passages 21-22, while cells from batch 2 had a passage number of P12.

The experiment was initiated as described in section 2.7. Calu-3 cells were seeded on PET inserts (1.1 cm^2, 0.4 µm; Sarstedt, Nümbrecht, Germany) at a density of 1 x 10^5 cells/insert, and placed in 12-well plates (VWR Chemicals, Radnor, USA). The cells were cultured for two weeks in a liquid-liquid interface to form an intact monolayer with tight junctions, which was assessed by measuring TEER values. Before testing the permeability of fluorescein in the presence of SGC/SGD, the medium was replaced with HBSS, and after a 30-minute incubation, TEER values were measured. Fluorescein permeability (donor concentration 40 µg/mL) was measured in the presence of 20 µg/mL of SGC or SGD in a bidirectional setup. For cell integrity assessment in the apical to basolateral direction (A-B), 520 µL of test solutions (SGC/SGD: 20 µg/mL; fluorescein 40 µg/mL) were added to the apical compartment, and 1700 µL of buffer was added to the basolateral chamber. For studies from the basolateral to the apical chamber (B-A), 520 µL of 40 µg/mL fluorescein in HBSS buffer was added to the apical compartment, and 1700 µL of the same SGC/SGD solutions were added to the basolateral compartment. Samples of 20 µL were withdrawn from the apical chamber at two time points: immediately after the addition of both solutions (t = 0 h) and after 120 minutes (t = 2 h). The

donor samples were diluted with 180 µL of HBSS. From the basolateral chambers, samples of 200 µL were taken at the same time points. The plates were then placed on a shaking platform at 37 ± 0.37°C and 200 rpm for 2 hours. After incubation, the cells were washed, and the solutions were replaced with HBSS. After a further 30-minute incubation, TEER values were measured again to assess cell integrity. Samples taken at t_{0h} and t_{2h} were pipetted into a 96-well plate, and absorbance was measured at 490 nm using an EPOCH2 spectrophotometer (BioTek Instruments, Inc., Vermont, USA). A calibration curve with fluorescein, ranging from 20 to 0.5 mg/mL, was created to calculate the fluorescein concentration in the samples.

To assess potential cytotoxic effects on the monolayer, cell culture medium (10% FBS, 1% P/S) containing 0.45 mg/mL MTT reagent was added to the inserts (300 µL) and wells (1500 µL). Cells treated with fluorescein (40 µg/mL) served as a control. Samples were then incubated for 3.5 hours at 37°C. Following incubation, the MTT solution was removed, and DMSO was added to dissolve the formazan crystals (300 µL for inserts; 1500 µL for wells). The plates were incubated on a shaking platform at 200 rpm in the absence of light for 30 minutes. Subsequently, 100 µL samples from each well were pipetted into a 96-well plate, and the absorbance was measured at 570 nm using an Epoch 2 spectrophotometer.

2.2. Results and Discussion

The comparison of TEER values revealed significant differences in behaviour between the batches (**Figure S1 A-B**). While batch 2 exhibited higher overall TEER values compared to batch 1, a greater decrease in TEER was observed for batch 2 during the experiment, ultimately resulting in approximately the same TEER value after the experiment. Interestingly, the decrease in TEER observed in batch 1 during the measurement of the B-A transport of SGD was not observed in batch 2, indicating that this effect may be specific to the cell batch. In the MDAA permeability studies (**Figure 5**), the applied batch 1 exhibited a significant decrease in TEER values when SGD was directed from the basolateral to the apical side (B-A). This pronounced decline in TEER values was observed both in the previous and current experiments with batch 1 and not batch 2. Additionally, a decrease in TEER values was also observed with SGC in the B-A direction, although it was not as substantial as the decrease caused by SGD B-A.

To determine whether the observed decline in TEER influenced the permeability results, we recreated the experiments and added fluorescein, a non-permeability marker, to

the cells. In an intact Calu-3 monolayer, no fluorescein permeation occurs to the other chamber. The experimental setup compared fluorescein permeation on inserts without a cell monolayer to that on a Calu-3 monolayer, as well as on a Calu-3 monolayer in the presence of SGC or SGD, in both transport directions. Particular focus was given to those conditions experiencing a reduction in TEER. However, no fluorescein permeation was detected in the transport studies involving a Calu-3 monolayer (LOQ 0.646 µg/mL; LOD 0.213 µg/mL; **Figure S1 C**). Fluorescein permeation only occurred through the membranes of inserts without a cell monolayer. None of the other tested substances affected the integrity of the Calu-3 monolayer to the extent that fluorescein permeability increased. This indicates that, despite the observed decline in TEER of the Calu-3 monolayer, there was no impact on the permeability assessments. Additionally, the remeasurement of TEER in the B-A directed SGD-test wells, after 24 hours of incubation with medium at 37°C, showed a recovery in TEER, suggesting that the TEER decrease is only a temporary effect.

Subsequently, a MTT assay was conducted to evaluate whether cytotoxic effects occurred on the Calu-3 monolayers, particularly to ascertain if the B-A directed TEER reduction caused by SGD was due to cell-damaging processes. Neither batch 1 nor batch 2 showed a decrease in cell viability after the permeability studies, implying that the TEER decline is not caused by damaged cells (**Figure S1 D**).

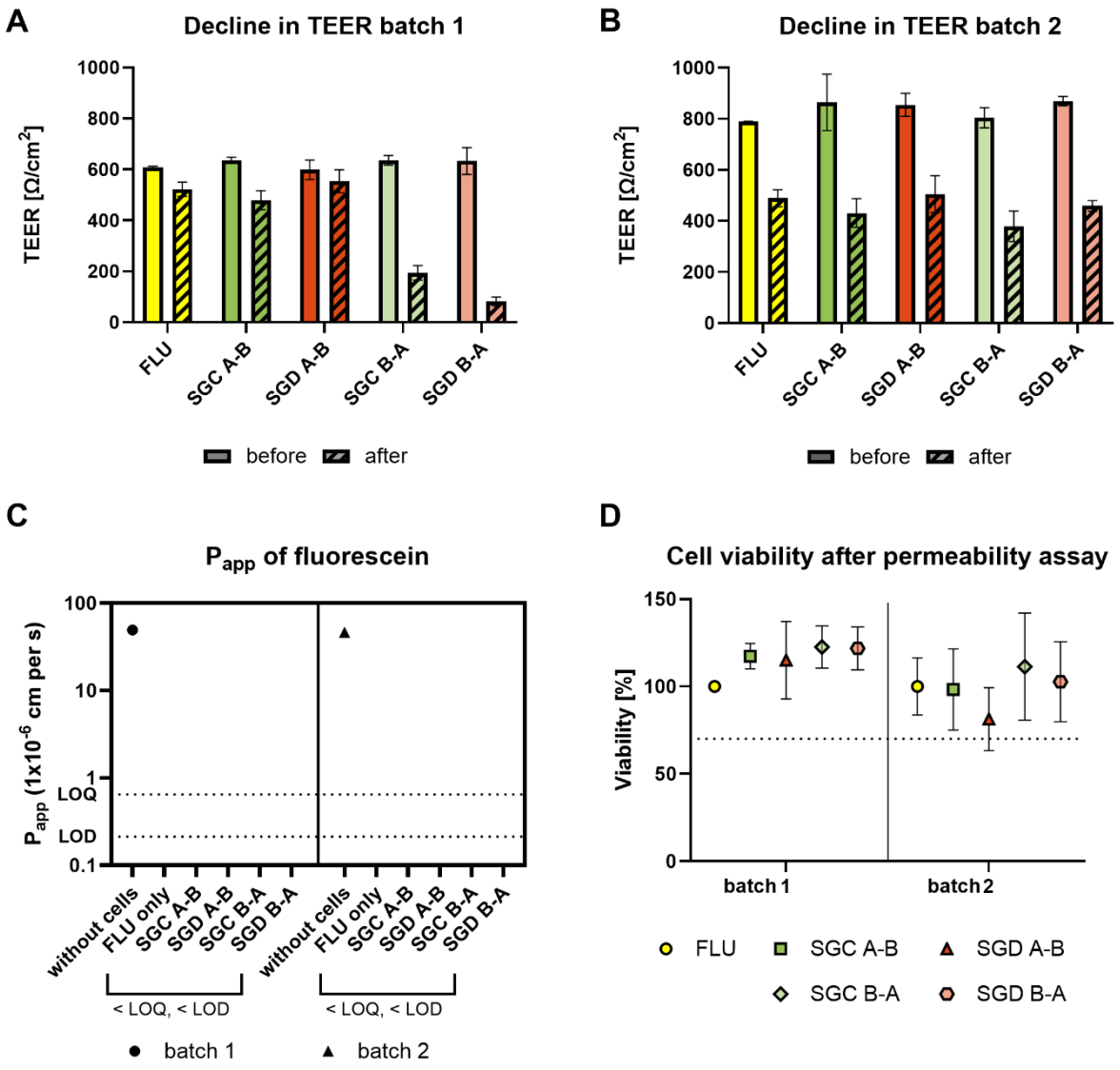


Figure S2. A) Decline in TEER values of Calu-3 cell batch 1 before and after the permeability experiment. B) Decline in TEER values of Calu-3 cell batch 2 before and after the permeability experiment. C) Comparison of apparent permeability (Papp) values of fluorescein on inserts without a cell monolayer versus on a Calu-3 monolayer, in the presence of fluorescein alone or in a bidirectional setup with SGC or SGD. D) Cell viability assessed for both batches of cells after the permeability studies with the Calu-3 monolayer. Reduction under 70 % of viability was considered as cytotoxic (dotted line).

In conclusion, the observed one-directional decrease in TEER caused by SGD appears to be a cell-batch specific phenomenon with an unknown molecular mechanism. Importantly, despite the decline in epithelial resistance, there was no significant increase in permeability, indicating that the robustness of the permeability results remains unaffected.

3. Results of uptake in Calu-3 cells

Results of the uptake experiment via flow-injection time-of-flight mass spectrometry (FIA-TOFMS) in negative ion mode for ion m/z 707.2127 with validation parameters of 3.14E+05 for slope, 6.58E+04 for intercept and 0.9840 for R^2 are listed in Table S15.

Table S15. Overview of the measurement data of the uptake in Calu-3 cells. Negative concentration values observed in vehicle-treated cell extracts are due to numerical errors close to the detection limit (e.g. ion intensities lower than the intercept of the regression model).

Treatment	Treatment concentration	Time [hours]	Ion intensity [counts]			Concentration [pg/µL]			Cell number [cells/µL]	Normalized concentration [pg/cell]			Mean [pg/cell]	SD [pg/cell]
			1	2	3	1	2	3		1	2	3		
Sanggenon C	50 µg/mL	4	1.41E+06	1.28E+06	1.29E+06	4284	3851	3902	2050	2.09	1.88	1.90	1.96	0.12
Sanggenon C	50 µg/mL	12	4.35E+06	4.28E+06	3.92E+06	13620	13397	12248	2362	5.77	5.67	5.19	5.54	0.31
Sanggenon C	50 µg/mL	24	5.70E+06	5.37E+06	5.54E+06	17926	16862	17411	2148	8.34	7.85	8.11	8.10	0.25
Sanggenon D	50 µg/mL	4	2.07E+05	1.54E+05	4.30E+05	449	281	1158	1946	0.23	0.14	0.60	0.32	0.24
Sanggenon D	50 µg/mL	12	4.23E+05	4.17E+05	3.85E+05	1137	1116	1014	2014	0.57	0.55	0.50	0.54	0.03
Sanggenon D	50 µg/mL	24	5.54E+05	7.59E+05	5.91E+05	1552	2206	1671	2638	0.59	0.84	0.63	0.69	0.13
Vehicle (DMSO)	0.25%	4	1.89E+04	2.87E+04	4.97E+04	-149	-118	-51	2689	-0.06	-0.04	-0.02	-0.04	0.02
Vehicle (DMSO)	0.25%	12	3.73E+04	2.03E+05	8.11E+04	-91	436	49	2316	-0.04	0.19	0.02	0.06	0.12
Vehicle (DMSO)	0.25%	24	1.45E+04	1.95E+05	1.92E+04	-163	411	-148	2513	-0.07	0.16	-0.06	0.01	0.13

8.3.2 Review of Publication II

A growing body of literature suggests that the inhalation of antimicrobial agents may offer advantages over other routes of administration in the treatment of lung infections, as it enables targeted drug delivery with higher local concentrations at the site of infection while reducing systemic exposure and associated toxicity.[259-261] In several studies, inhaled antimicrobial agents have been investigated with respect to their molecular properties to better understand how pulmonary retention can be improved, thereby enhancing local efficacy in the lungs.[259,260,262-266] The term 'pulmonary retention' refers to the slow absorption of a compound from lung tissue into the systemic circulation.[267] It has been shown that large, hydrophilic molecules with low permeability exhibit favorable PK properties across lung epithelial monolayers.[259,262-267]

In 2022, Hastedt and colleagues presented the so-called inhalation-based Biopharmaceutics Classification System (iBCS), a concept that had already been discussed during a 2014 workshop due to the absence of a BCS model for inhalable medicinal products.[268,269] The iBCS was developed to enable a qualitative assessment of technical and clinical risks based on the physicochemical properties of a drug candidate (solubility, represented by the dose number D_o, and permeability) together with key attributes of its drug product (dose and dissolution).[267,268] In addition, the iBCS aims to provide an analytical basis for IVIVC evaluations of inhaled drugs.[268]

When implementing the iBCS, Hastedt et al. used model compounds in combination with physiologically based PK modeling (PBPK) to classify drugs into four groups, owing to the limited number of approved orally inhalable drugs on the market.[268] Their aim was to establish a system grounded in the same principles as the BCS developed for orally administered drugs, namely focusing on the rate and extent of uptake, while taking into account the specific physiological factors associated with pulmonary drug delivery.[268]

An important distinction is that oral dosage forms and inhalable drugs differ fundamentally in their optimal product characteristics.[268] Oral drugs are typically intended for systemic activity and should therefore exhibit high permeability and systemic bioavailability.[268] In contrast, compounds with low permeability are advantageous for inhalation, as they allow for high concentrations in the ELF of the lung while maintaining low systemic exposure.[263,265,268]

Since most inhaled drugs are intended for local activity, low permeability is generally considered advantageous for inhalation therapy, as it promotes pulmonary retention.[259,267,268] In contrast, high permeability can result in faster clearance from the ELF and potentially increased systemic exposure, which may compromise the desired local efficacy profile.[268] Based on this knowledge, the iBCS can provide decision support in the early stages of drug development by identifying drug candidates with favorable properties for inhalation administration that warrant further investigation.[267]

The four classes of the iBCS and their implications for the discovery and development of inhalable drugs have been described by Forbes et al. as follows[267]:

(i) iBCS Class I: High solubility (D_o < 1) and high permeability
Ideal for systemically acting drugs with fast systemic absorption.

(ii) iBCS Class II: Low solubility (D_o > 1) and high permeability
The bioavailability might be limited due to low solubility and dissolution rate, requiring a suitable formulation to overcome this challenge.

(iii) iBCS Class III: High solubility (D_o < 1) and low permeability
Limited permeability restricts absorption and contributes to prolonged retention within the lungs.

(iv) iBCS Class IV: Low solubility (D_o > 1) and low permeability
Class IV is considered the most challenging for inhalation, since the accumulation of undissolved drug, together with a prolonged residence time in the lungs, can lead to toxic effects.

The dose number for a specific lung region I (D_{o_i}) describes the fraction of the nominal dose (label claim) that is deposited within the available fluid volume in that region.[269] To calculate this value, the regional dose (M_i), the available volume of ELF for dissolution (V_i), and the regional solubility of the drug (C_{s_i}) must be known (eq 1).[269]

$$D_{o_i} = \frac{(M_i / V_i)}{C_{s_i}} \qquad \text{eq. 1}$$

The threshold between low and high solubility is defined by a D_{o_i} = 1.[269] When the drug is fully dissolved in V_i, the D_{o_i} is < 1 (high solubility); if only a fraction is dissolved, the D_{o_i} is > 1 (low solubility).[269]

Hastedt et al. pointed out that most marketed drugs intended for inhalation with local activity are classified as iBCS Class II or Class III products.[269] The Class III drug tobramycin, for instance, exhibits high solubility along with low permeability across biological membranes, making it particularly well-suited for locally targeted inhalation therapy.[224,263,264,270] The authors therefore conclude that prolonged pulmonary retention may enhance efficacy.[269]

As shown in **Publication II**, the two tested pure MDAAs, i.e., SGC and SGD, exhibited either no measurable permeability (SGC) or very low permeability (SGD, < 0.6% of the applied amount) in Calu-3 cells.[271] Moreover, both MDAAs displayed pH-dependent solubility as pure compounds.[271] In a buffer with a pH of 6.8, which simulates pulmonary conditions, a solubility of 18.5 µg/mL was measured for pure SGC, compared to 304 µg/mL for isolated SGD.[271] Particularly notable was the approximately tenfold higher solubility of SGD, a difference that was observed consistently across all tested pH values.[271]

With regard to the iBCS, a classification of SGC and SGD into one of the four classes I–IV is not possible, as these are pure compounds and classification requires the investigation of the API's behavior within a formulation, as described in equation 1.[267,269] In addition, the BDCCS evaluation of 109 NPs provided by Li et al. did not include flavonoids and therefore the results of **Publication II** could not be compared with related datasets.[224]

However, the studies presented in **Publication II** provide valuable insights into whether inhalation is a viable route of administration and whether further investigation focusing on formulation development is warranted.[271] This is indeed the case for SGC and SGD, as both compounds exhibit favorably low permeability.[271] The solubility data of both MDAAs suggest that improving solubility through appropriate formulation strategies could be beneficial.[271]

After 4 hours, only limited cell uptake was detected for both SGC (2.0 pg/cell) and SGD (0.3 pg/cell), with a maximum increase to 8.1 pg/cell for SGC after 24 hours.[271] SGD thus appears to remain in the extracellular space to a greater extent than SGC and may be less able to cross the epithelial barrier, which aligns with its lower lipophilicity values (SGC: logP = 3.48; SGD: logP = 2.92).[271] This is considered advantageous for therapeutic effects that potentially take place in the extracellular space, such as NA inhibition or disruption of *S. pneumoniae* biofilm formation.[271]

These findings, combined with the reported antiviral, antibacterial, and anti-inflammatory activities of the compounds and their traditional use in respiratory tract diseases such as bronchitis, support the conclusion that inhalation could be a highly beneficial route of administration and should be further investigated *in vivo*.[119,122,123,125,256-258] This could be especially beneficial for potential treatment options for secondary infections.[123,124,251]

In the context of PK synergy, **Publication II** showed that the two pure MDAAs appear more promising for inhalation than the multicomponent extracts MA21 (containing 2.6% MDAAs) and MA60 (containing 29.0% MDAAs), particularly in terms of solubility.[271] Based on the available literature, the solubility of APIs can be enhanced by co-existing compounds such as flavonoids, which act as endogenous solubility enhancers.[230,272,273] Since the flavonoids in the extracts obtained from *M. alba* contain prenylated structures, it is conceivable that the prenyl moiety

significantly alters this behavior. However, further studies are necessary to substantiate this preliminary assumption and to gain more insigths into potential PK synergy effects.

Increases, decreases, as well as no changes in the permeability of NPs due to the influence of co-present compounds in an extract have all been reported in the literature.[228,274-276] For isolated SGC, which exhibits no permeability under the test conditions, no improvement in permeability was observed in the extract either.[271] For SGD, a decrease in permeability in the basolateral-to-apical (B-A) direction was evident due to the influences of the extracts.[271] However, the results for SGD should be interpreted with caution, as a simultaneous decrease in transepithelial electrical resistance (TEER) was also observed.[271] This may indicate a temporary loss of epithelial integrity, as TEER is an indicator of the tight junction status and should therefore remain constant throughout the experiments.[271,277] However, this type of B-A TEER decrease has already been described in the literature, although the underlying causes could not always be clearly identified.[277,278] These findings underscore the need for further optimization and standardization of testing methods when evaluating the permeability of NPs.

From a technical perspective, improving the solubility of the MDAAs should be a key consideration in the development of an inhalable pharmaceutical product based on the pure compounds SGC and SGD. Simultaneously, the selected formulation vehicle should support efficient and stable nebulization, as this is considered the most suitable delivery method given the physicochemical properties of the compounds.[259,263,264,266] Furthermore, effective release of the API at the target site must be ensured while maintaining the favorable permeability characteristics, in order to support localized drug release, prolonged pulmonary retention, and high therapeutic efficacy.[267,269,279]

9 Conclusion

This doctoral thesis spans the entire spectrum from the discovery of bioactive natural products (NPs) to their potential application, with a focus on translational aspects. To this end, a biochemometric two-dimensional statistical heterocovariance analysis (2D HetCA) approach was developed to identify promising bioactive compounds, and a biopharmaceutical profiling (BPP) method was established to evaluate suitability of mulberry Diels–Alder-type adducts (MDAAs) for inhalation administration.

2D HetCA based on HSQC datasets appears to be a promising method for determining whether further processing is advisable at an early stage of the phytochemical workup, namely, after a single microfractionation step. The method is capable of specifically identifying those microfractions (MFs) that contain compounds most likely correlated with bioactivity. While a few basic prerequisites must be met for the method to be applicable (e.g., overlapping constituents across at least three consecutive MFs and a consistently increasing or decreasing bioactivity profile across the same MFs), 2D HetCA enables biochemometric analysis for any pharmacological target with a measurable activity.

In **Manuscript I**, the new 2D HetCA approach was first established in a proof-of-concept study using artificially mixed triterpenes (TTs). Subsequently, 2α,19α-dihydroxy-3-oxo-12-ursen-28-oic acid, a minor triterpenoid derived from a dichloromethane leaf extract of *Eriobotrya japonica* Lindl., was successfully identified and later confirmed as active on the retinoic acid receptor-related orphan receptor gamma (RORγ) with an IC_{50} = 2.15 µM.

The study demonstrated that the new approach enabled access to spectral regions that could not be analyzed using 1D HetCA based on ^{1}H spectra due to the high density of overlapping signals. This improvement resulted from enhanced dispersion and, consequently, better visibility of cross-peaks. Additionally, the greater variation in the color code allowed for more precise differentiation between HSQC value pairs that were positively or negatively correlated with bioactivity.

It is beneficial that the method is generally independent of the isolation and structural elucidation of the bioactive compounds. Another advantage may lie in the overall robustness of NMR measurements, although their inherently low sensitivity must be taken into account. However, 2D HetCA requires appropriate 2D NMR measurements, careful dataset preparation, and subsequent data analysis and interpretation.

As potential future developments, the linking of 2D HetCA with NMR databases, as well as the integration of MS-based datasets, should be emphasized. Overall, 2D HetCA is a valuable technique for the rapid and targeted search for bioactive constituents in complex mixtures derived from natural sources.

Another focus of this doctoral thesis was the establishment of a BPP method for NPs. Two mulberry Diels–Alder-type adducts (MDAAs) from the root bark of *Morus alba* L., namely sanggenon C (SGC) and sanggenon D (SGD), were identified in **Publication I** as inhibitors of the main protease (M^{pro}), exhibiting *in vitro* activity against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) through a time-efficient and cost-effective target-based virtual screening (TBVS) approach.

Subsequently, **Publication II** clearly demonstrated that both MDAAs, due to their low permeability in Calu-3 cells, exhibit favorable properties for inhalation administration. The isolated compounds SGC and SGD generally showed better suitability for pulmonary delivery compared to the corresponding compounds in MDAA-containing multicomponent mixtures MA21 (2.6% MDAAs) and MA60 (29.0% MDAAs).

One current challenge in drug development is the difficulty in assessing the suitability of drug candidates for inhalation at an early stage, for example, based on predicted permeability. This is partly due to the fact that formulation-related parameters are also required for classification within systems such as the inhalation Biopharmaceutics Classification System (iBCS) which allows for the estimation of *in vitro–in vivo* correlations (IVIVC). This poses a problem for decision-making during the evaluation process, as it is essential to determine as early as possible whether further testing of a drug candidate is likely to be successful, and to exclude candidates with a high risk of failure from the outset. This applies not only to NPs, but to pure compounds in general.

Additionally, predicting pharmacokinetic (PK) synergy effects in complex extracts containing NPs remains particularly challenging, and at the moment experimental studies in this area are indispensable. For example, a literature review on the influence of flavonoids in extracts on solubility and permeability yielded no references to publications describing comprehensive and systematic investigations in this area. Furthermore, there are no specific studies that have investigated whether, and to what extent, the degree of flavonoid prenylation affects solubility.

Developing new approaches that allow for more precise and robust predictions would therefore be highly desirable. To this end, it may be advantageous to establish methods initially based on testing well-defined artificial mixtures rather than complex extracts whose exact composition can likely only be approximated due to the large number of NPs they contain.

In regard to the isolated MDAA compounds of **Publication II**, as a future step, it seems reasonable to develop a formulation that takes advantage of the low permeability, enhances the solubility of SGC and SGD, and enables their targeted release at the site of infection in the lung.

This doctoral thesis successfully demonstrates how the broad spectrum, from the search for bioactive compounds to the initial evaluation of their potential use in a defined application area, can be addressed. 2D HetCA and BPP have proven to be effective and valuable tools, whose further development could significantly contribute to NP drug discovery and development.

10 List of abbreviations

ADMET	absorption, distribution, metabolism, excretion, and toxicity
AI	artificial intelligence
anti-TB	anti-tuberculosis
APIs	active pharmaceutical ingredients
AUC	area under the curve
B-A	basolateral-to-apical
BCS	the Biopharmaceutics Classification System
BDDCS	the Biopharmaceutics Drug Disposition Classification System
BPP	biopharmaceutical profiling
bRo5	beyond Rule of Five
CADD	computer-aided drug design
COLMAR	Complex Mixture Analysis by NMR
c_{max}	maximum concentration
cryo-EM	cryo-electron microscopy
CX3CL1	C-X3-C Motif Chemokine Ligand 1
DAFdiscovery	Data Fusion-based Discovery
DL	deep learning
D_o	dose number
DTT	dithiothreitol
ELF	epithelial lining fluid
ELINA	Eliciting Nature's Activities
ELSD	evaporative light scattering detector
GC	gas chromatography
GC-MS	gas chromatography-mass spectrometry
GNPS	the Global Natural Product Social Molecular Networking
GNPS-FBMN	GNPS-feature-based molecular networking
HA	hemagglutinin
HetCA	statistical heterocovariance analysis
HMBC	Heteronuclear Multiple Bond Correlation spectroscopy
^{1}H NMR	proton nucelar magnetic resonance
HPLC	high-performance liquid chromatography
HPTLC	high-performance thin-layer chromatography
HSCCC	high-speed countercurrent chromatography
HSQC	Heteronuclear Single Quantum Coherence spectroscopy
HTS	high-throughput screening
HTVS	high-throughput virtual screening
HUVECs	human umbilical vein endothelial cells
IAV	Influenza A virus

iBCS	inhalation-based Biopharmaceutics Classification System
IBV	Influenza B virus
IFN-β	interferon ß
IL	interleukin
IVIVC	*in vitro–in vivo* correlations
IP-10	CXC Motif Chemokine Ligand 10
LBDD	ligand-based drug design
LBVS	ligand-based virtual screening
LC-MS	liquid chromatography-mass spectrometry
LOD	limit of detection
logP	logarithmic partition coefficient in a biphasic water–octanol system
MADByTE	Metabolomics and Dereplication by Two-Dimensional Experiments
MA21	*Morus alba* L. extract containing 2.6% MDAAs
MA60	*Morus alba* L. extract containing 29.0% MDAAs
MBIC	minimal biofilm inhibitory concentration
MDAAs	mulberry Diels–Alder-type adducts
MDCK	Madin-Darby Canine Kidney
MFs	microfractions
M_i	regional dose
MIC	minimal inhibitory concentration
ML	machine learning
MN	molecular networking
M^{pro}	main protease, nsp5, 3C-like protease ($3CL^{pro}$)
MS	mass spectrometry
MS^2	tandem mass spectrometry, MS/MS
MTD	maximum tolerated dose
NA	neuraminidase
NF-κB	nuclear factor-κB
NMR	nuclear magnetic resonance
NPs	natural products
nsp	non-structural protein
ORFs	open reading frames
PAINS	pan-assay interference compound
PBPK	physiologically based pharmacokinetic modeling
PK	pharmacokinetic
PL^{pro}	papain-like protease, domain of non-structural protein 3 (nsp3)
PLS	partial least squares
qHNMR	quantitative ^{1}H NMR
Q-HSQC	quantitative Heteronuclear Single Quantum Coherence spectroscopy
QSAR	quantitative structure-activity relationship

RdRp	RNA-dependent RNA polymerase
RORγ	retinoic acid receptor-related orphan receptor gamma
Ro5	Rule of Five
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SFC	supercritical fluid chromatography
SGC	sanggenon C
SGD	sanggenon D
SGG	sanggenon G
SGO	sanggenon O
SHY	statistical heterospectroscopy
SI	selectivity index
SMART	Small Molecule Accurate Recognition Technology
STOCSY	statistical total correlation spectroscopy
TBDD	target-based drug design
TBVS	target-based virtual screening
TCM	Traditional Chinese Medicine
TEER	transepithelial electrical resistance
TLC	thin-layer chromatography
t_{max}	time to maximum concentration
TOCSY	total correlation spectroscopy
TTs	triterpenes
UHPSFC	ultra-high performance supercritical fluid chromatography
VHs	virtual hits
V_i	volume of epithelial lining fluid
VS	virtual screening
XR	X-ray crystallography
1D	one-dimensional
1D HetCA	^{1}H NMR-based statistical heterocovariance analysis
2D	two-dimensional
2D HetCA	HSQC-based statistical heterocovariance analysis
3D	three-dimensional

11 References

(1) Dias, D. A.; Urban, S.; Roessner, U. *Metabolites* **2012**, *2* (2), 303-336.
DOI: https://doi.org/10.3390/metabo2020303

(2) Atanasov, A. G.; Waltenberger, B.; Pferschy-Wenzig, E. M.; Linder, T.; Wawrosch, C.; Uhrin, P.; Temml, V.; Wang, L.; Schwaiger, S.; Heiss, E. H.; Rollinger, J. M.; Schuster, D.; Breuss, J. M.; Bochkov, V.; Mihovilovic, M. D.; Kopp, B.; Bauer, R.; Dirsch, V. M.; Stuppner, H. *Biotechnol Adv* **2015**, *33* (8), 1582-1614.
DOI: https://doi.org/10.1016/j.biotechadv.2015.08.001

(3) Bernardini, S.; Tiezzi, A.; Laghezza Masci, V.; Ovidi, E. *Nat Prod Res* **2018**, *32* (16), 1926-1950.
DOI: https://doi.org/10.1080/14786419.2017.1356838

(4) Romano, J. D.; Tatonetti, N. P. *Front Genet* **2019**, *10*, 368.
DOI: https://doi.org/10.3389/fgene.2019.00368

(5) Zaynab, M.; Fatima, M.; Sharif, Y.; Zafar, M. H.; Ali, H.; Khan, K. A. *Microb Pathog* **2019**, *137*, 103728.
DOI: https://doi.org/10.1016/j.micpath.2019.103728

(6) Zaynab, M.; Fatima, M.; Abbas, S.; Sharif, Y.; Umair, M.; Zafar, M. H.; Bahadar, K. *Microb Pathog* **2018**, *124*, 198-202.
DOI: https://doi.org/10.1016/j.micpath.2018.08.034

(7) Rollinger, J. M.; Langer, T.; Stuppner, H. *Curr Med Chem* **2006**, *13* (13), 1491-1507.
DOI: https://doi.org/10.2174/092986706777442075

(8) Mullowney, M. W.; Duncan, K. R.; Elsayed, S. S.; Garg, N.; van der Hooft, J. J. J.; Martin, N. I.; Meijer, D.; Terlouw, B. R.; Biermann, F.; Blin, K.; Durairaj, J.; Gorostiola Gonzalez, M.; Helfrich, E. J. N.; Huber, F.; Leopold-Messer, S.; Rajan, K.; de Rond, T.; van Santen, J. A.; Sorokina, M.; Balunas, M. J.; Beniddir, M. A.; van Bergeijk, D. A.; Carroll, L. M.; Clark, C. M.; Clevert, D. A.; Dejong, C. A.; Du, C.; Ferrinho, S.; Grisoni, F.; Hofstetter, A.; Jespers, W.; Kalinina, O. V.; Kautsar, S. A.; Kim, H.; Leao, T. F.; Masschelein, J.; Rees, E. R.; Reher, R.; Reker, D.; Schwaller, P.; Segler, M.; Skinnider, M. A.; Walker, A. S.; Willighagen, E. L.; Zdrazil, B.; Ziemert, N.; Goss, R. J. M.; Guyomard, P.; Volkamer, A.; Gerwick, W. H.; Kim, H. U.; Muller, R.; van Wezel, G. P.; van Westen, G. J. P.; Hirsch, A. K. H.; Linington, R. G.; Robinson, S. L.; Medema, M. H. *Nat Rev Drug Discov* **2023**, *22* (11), 895-916.
DOI: https://doi.org/10.1038/s41573-023-00774-7

(9) Dorogan, M.; Namballa, H. K.; Harding, W. W. *J Med Chem* **2024**, *67* (15), 12463-12484.
DOI: https://doi.org/10.1021/acs.jmedchem.4c00537

(10) Aggarwal, P. R.; Mehanathan, M.; Choudhary, P. *J Plant Physiol* **2024**, *303*, 154379.
DOI: https://doi.org/10.1016/j.jplph.2024.154379

(11) Chevrette, M. G.; Gutierrez-Garcia, K.; Selem-Mojica, N.; Aguilar-Martinez, C.; Yanez-Olvera, A.; Ramos-Aboites, H. E.; Hoskisson, P. A.; Barona-Gomez, F. *Nat Prod Rep* **2020**, *37* (4), 566-599.
DOI: https://doi.org/10.1039/c9np00048h

(12) Evans, D. A.; Green, C. L. *Chemical Society Reviews* **1973**, *2* (1).
DOI: https://doi.org/10.1039/cs9730200075

(13) Wei, J.; Wang, L.; Zhu, J.; Zhang, S.; Nandi, O. I.; Kang, L. *PLoS One* **2007**, *2* (9), e852.
DOI: https://doi.org/10.1371/journal.pone.0000852

(14) Li, S.; Jiang, S.; Jia, W.; Guo, T.; Wang, F.; Li, J.; Yao, Z. *Food Chem* **2024**, *432*, 137231. DOI: https://doi.org/10.1016/j.foodchem.2023.137231

(15) Rollinger, J. M. *Phytochemistry Letters* **2009**, *2*, 53-58. DOI: https://doi.org/10.1016/j.phytol.2008.12.002

(16) Atanasov, A. G.; Zotchev, S. B.; Dirsch, V. M.; International Natural Product Sciences Taskforce; Supuran, C. T. *Nat Rev Drug Discov* **2021**, *20* (3), 200-216. DOI: https://doi.org/10.1038/s41573-020-00114-z

(17) Ekiert, H. M.; Szopa, A. *Molecules* **2020**, *25* (23). DOI: https://doi.org/10.3390/molecules25235769

(18) Chopra, B.; Dhingra, A. K. *Phytother Res* **2021**, *35* (9), 4660-4702. DOI: https://doi.org/10.1002/ptr.7099

(19) Chen, Y.; Garcia de Lomana, M.; Friedrich, N. O.; Kirchmair, J. *J Chem Inf Model* **2018**, *58* (8), 1518-1532. DOI: https://doi.org/10.1021/acs.jcim.8b00302

(20) Newman, D. J.; Cragg, G. M. *J Nat Prod* **2020**, *83* (3), 770-803. DOI: https://doi.org/10.1021/acs.jnatprod.9b01285

(21) McChesney, J. D.; Venkataraman, S. K.; Henri, J. T. *Phytochemistry* **2007**, *68* (14), 2015-2022. DOI: https://doi.org/10.1016/j.phytochem.2007.04.032

(22) Kingston, D. G. *J Nat Prod* **2011**, *74* (3), 496-511. DOI: https://doi.org/10.1021/np100550t

(23) Chen, Y.; Kirchmair, J. *Mol Inform* **2020**, *39* (12), e2000171. DOI: https://doi.org/10.1002/minf.202000171

(24) Stone, S.; Newman, D. J.; Colletti, S. L.; Tan, D. S. *Nat Prod Rep* **2022**, *39* (1), 20-32. DOI: https://doi.org/10.1039/d1np00039j

(25) Feher, M.; Schmidt, J. M. *J Chem Inf Comput Sci* **2003**, *43* (1), 218-227. DOI: https://doi.org/10.1021/ci0200467

(26) Lachance, H.; Wetzel, S.; Kumar, K.; Waldmann, H. *J Med Chem* **2012**, *55* (13), 5989-6001. DOI: https://doi.org/10.1021/jm300288g

(27) Tay, D. W. P.; Yeo, N. Z. X.; Adaikkappan, K.; Lim, Y. H.; Ang, S. J. *Sci Data* **2023**, *10* (1), 296. DOI: https://doi.org/10.1038/s41597-023-02207-x

(28) Ogawa, K.; Sakamoto, D.; Hosoki, R. *Chem Pharm Bull (Tokyo)* **2023**, *71* (7), 486-494. DOI: https://doi.org/10.1248/cpb.c23-00039

(29) Lawson, A. D. G.; MacCoss, M.; Heer, J. P. *J Med Chem* **2018**, *61* (10), 4283-4289. DOI: https://doi.org/10.1021/acs.jmedchem.7b01120

(30) Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J. *Nat Rev Drug Discov* **2015**, *14* (2), 111-129. DOI: https://doi.org/10.1038/nrd4510

(31) Hert, J.; Irwin, J. J.; Laggner, C.; Keiser, M. J.; Shoichet, B. K. *Nat Chem Biol* **2009**, *5* (7), 479-483.
DOI: https://doi.org/10.1038/nchembio.180

(32) Lavecchia, A.; Di Giovanni, C. *Curr Med Chem* **2013**, *20* (23), 2839-2860.
DOI: https://doi.org/10.2174/09298673113209990001

(33) Scott, K. A.; Ropek, N.; Melillo, B.; Schreiber, S. L.; Cravatt, B. F.; Vinogradova, E. V. *Current Research in Chemical Biology* **2022**, *2*.
DOI: https://doi.org/10.1016/j.crchbi.2022.100028

(34) Vaskevicius, A.; Trumpickaite, G.; Parafjanovic, E.; Manakova, E.; Mickeviciute, A.; Gedgaudas, M.; Kojis, T.; Paketuryte-Latve, V.; Smirnov, A.; Baranauskiene, L.; Grazulis, S.; Zubriene, A.; Dudutiene, V.; Matulis, D. *Chemistry* **2025**, *31* (18), e202404409.
DOI: https://doi.org/10.1002/chem.202404409

(35) Domingo-Fernandez, D.; Gadiya, Y.; Mubeen, S.; Bollerman, T. J.; Healy, M. D.; Chanana, S.; Sadovsky, R. G.; Healey, D.; Colluru, V. *iScience* **2023**, *26* (9), 107729.
DOI: https://doi.org/10.1016/j.isci.2023.107729

(36) Pirintsos, S.; Panagiotopoulos, A.; Bariotakis, M.; Daskalakis, V.; Lionis, C.; Sourvinos, G.; Karakasiliotis, I.; Kampa, M.; Castanas, E. *Molecules* **2022**, *27* (13).
DOI: https://doi.org/10.3390/molecules27134060

(37) Gangwal, A.; Lavecchia, A. *J Med Chem* **2025**, *68* (4), 3948-3969.
DOI: https://doi.org/10.1021/acs.jmedchem.4c01257

(38) Sneader, W. *Drug Discovery: A History, Wiley* **2005**.
DOI: https://doi.org/10.1002/0470015535

(39) Fabricant, D. S.; Farnsworth, N. R. *Environ Health Perspect* **2001**, *109 Suppl 1* (Suppl 1), 69-75.
DOI: https://doi.org/10.1289/ehp.01109s169

(40) Buenz, E. J.; Verpoorte, R.; Bauer, B. A. *Annu Rev Pharmacol Toxicol* **2018**, *58*, 509-530.
DOI: https://doi.org/10.1146/annurev-pharmtox-010617-052703

(41) Caesar, L. K.; Cech, N. B. *Nat Prod Rep* **2019**, *36* (6), 869-888.
DOI: https://doi.org/10.1039/c9np00011a

(42) Devereaux, A. L.; Mercer, S. L.; Cunningham, C. W. *ACS Chem Neurosci* **2018**, *9* (10), 2395-2407.
DOI: https://doi.org/10.1021/acschemneuro.8b00150

(43) Hamilton, G. R.; Baskett, T. F. *Can J Anaesth* **2000**, *47* (4), 367-374.
DOI: https://doi.org/10.1007/BF03020955

(44) Hänsel, R.; Sticher, O. *Pharmakognosie – Phytopharmazie*; Wissenschaftliche Verlagsgesellschaft, **2010**.

(45) Weller, M. G. *Sensors (Basel)* **2012**, *12* (7), 9181-9209.
DOI: https://doi.org/10.3390/s120709181

(46) Kellogg, J. J.; Paine, M. F.; McCune, J. S.; Oberlies, N. H.; Cech, N. B. *Nat Prod Rep* **2019**, *36* (8), 1196-1221.
DOI: https://doi.org/10.1039/c8np00065d

(47) Khin, M.; Knowles, S. L.; Crandall, W. J.; Jones, D. D., Jr.; Oberlies, N. H.; Cech, N. B.; Houriet, J. *J Pharm Biomed Anal* **2021**, *197*, 113965.
DOI: https://doi.org/10.1016/j.jpba.2021.113965

(48) Zwirchmayr, J.; Cruz, C. D.; Grienke, U.; Tammela, P.; Rollinger, J. M. *iScience* **2023**, *26* (9), 107523.
DOI: https://doi.org/10.1016/j.isci.2023.107523

(49) Nothias, L. F.; Nothias-Esposito, M.; da Silva, R.; Wang, M.; Protsyuk, I.; Zhang, Z.; Sarvepalli, A.; Leyssen, P.; Touboul, D.; Costa, J.; Paolini, J.; Alexandrov, T.; Litaudon, M.; Dorrestein, P. C. *J Nat Prod* **2018**, *81* (4), 758-767.
DOI: https://doi.org/10.1021/acs.jnatprod.7b00737

(50) Wolfender, J. L.; Litaudon, M.; Touboul, D.; Queiroz, E. F. *Nat Prod Rep* **2019**, *36* (6), 855-868.
DOI: https://doi.org/10.1039/c9np00004f

(51) Bindseil, K. U.; Jakupovic, J.; Wolf, D.; Lavayre, J.; Leboul, J.; van der Pyl, D. *Drug Discov Today* **2001**, *6* (16), 840-847.
DOI: https://doi.org/10.1016/s1359-6446(01)01856-6

(52) Inglese, J.; Johnson, R. L.; Simeonov, A.; Xia, M.; Zheng, W.; Austin, C. P.; Auld, D. S. *Nat Chem Biol* **2007**, *3* (8), 466-479.
DOI: https://doi.org/10.1038/nchembio.2007.17

(53) Schneider, G. *Nat Rev Drug Discov* **2018**, *17* (2), 97-113.
DOI: https://doi.org/10.1038/nrd.2017.232

(54) Sukuru, S. C.; Jenkins, J. L.; Beckwith, R. E.; Scheiber, J.; Bender, A.; Mikhailov, D.; Davies, J. W.; Glick, M. *J Biomol Screen* **2009**, *14* (6), 690-699.
DOI: https://doi.org/10.1177/1087057109335678

(55) Macarron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Cirovic, D. A.; Garyantes, T.; Green, D. V.; Hertzberg, R. P.; Janzen, W. P.; Paslay, J. W.; Schopfer, U.; Sittampalam, G. S. *Nat Rev Drug Discov* **2011**, *10* (3), 188-195.
DOI: https://doi.org/10.1038/nrd3368

(56) Altenburger, R.; Scholze, M.; Busch, W.; Escher, B. I.; Jakobs, G.; Krauss, M.; Kruger, J.; Neale, P. A.; Ait-Aissa, S.; Almeida, A. C.; Seiler, T. B.; Brion, F.; Hilscherova, K.; Hollert, H.; Novak, J.; Schlichting, R.; Serra, H.; Shao, Y.; Tindall, A.; Tollefsen, K. E.; Umbuzeiro, G.; Williams, T. D.; Kortenkamp, A. *Environ Int* **2018**, *114*, 95-106.
DOI: https://doi.org/10.1016/j.envint.2018.02.013

(57) Vidar, W. S.; Baumeister, T. U. H.; Caesar, L. K.; Kellogg, J. J.; Todd, D. A.; Linington, R. G.; O, M. K.; Cech, N. B. *J Nat Prod* **2023**, *86* (4), 655-671.
DOI: https://doi.org/10.1021/acs.jnatprod.2c00518

(58) Thorne, N.; Auld, D. S.; Inglese, J. *Curr Opin Chem Biol* **2010**, *14* (3), 315-324.
DOI: https://doi.org/10.1016/j.cbpa.2010.03.020

(59) Bisson, J.; McAlpine, J. B.; Friesen, J. B.; Chen, S. N.; Graham, J.; Pauli, G. F. *J Med Chem* **2016**, *59* (5), 1671-1690.
DOI: https://doi.org/10.1021/acs.jmedchem.5b01009

(60) Li, J. W.; Vederas, J. C. *Science* **2009**, *325* (5937), 161-165.
DOI: https://doi.org/10.1126/science.1168243

(61) Henrich, C. J.; Beutler, J. A. *Nat Prod Rep* **2013**, *30* (10), 1284-1298.
DOI: https://doi.org/10.1039/c3np70052f

(62) Tu, Y.; Jeffries, C.; Ruan, H.; Nelson, C.; Smithson, D.; Shelat, A. A.; Brown, K. M.; Li, X. C.; Hester, J. P.; Smillie, T.; Khan, I. A.; Walker, L.; Guy, K.; Yan, B. *J Nat Prod* **2010**, *73* (4), 751-754.
DOI: https://doi.org/10.1021/np9007359

(63) Van Drie, J. H. *J Comput Aided Mol Des* **2007**, *21* (10-11), 591-601.
DOI: https://doi.org/10.1007/s10822-007-9142-y

(64) Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W., Jr. *Pharmacol Rev* **2014**, *66* (1), 334-395.
DOI: https://doi.org/10.1124/pr.112.007336

(65) Sadybekov, A. V.; Katritch, V. *Nature* **2023**, *616* (7958), 673-685.
DOI: https://doi.org/10.1038/s41586-023-05905-z

(66) Sadri, A. *J Med Chem* **2023**, *66* (18), 12651-12677.
DOI: https://doi.org/10.1021/acs.jmedchem.2c01737

(67) Clark, D. E. *Expert Opinion Drug Discovery* **2008**, *3*, 841-851.
DOI: https://doi.org/10.1517/17460440802281978

(68) Wermuth, C. G.; Villoutreix, B.; Grisono, S.; Olivier, A.; Rocher, J.-P. Chapter 4 - Strategies in the Search for New Lead Compounds or Original Working Hypotheses. In *The Practice of Medicinal Chemistry (Fourth Edition)*, **2015**; pp 73-79.

(69) Bender, A.; Glen, R. C. *Org Biomol Chem* **2004**, *2* (22), 3204-3218.
DOI: https://doi.org/10.1039/B409813G

(70) Rollinger, J. M.; Stuppner, H.; Langer, T. *Prog Drug Res* **2008**, *65*, 211, 213-249.
DOI: https://doi.org/10.1007/978-3-7643-8117-2_6

(71) Ma, D.-L.; Chan, D. S.-H.; Leung, C.-H. *Chemical Science* **2011**, *2*, 1656-1665.
DOI: https://doi.org/10.1039/c1sc00152c

(72) Brusotti, G.; Cesari, I.; Dentamaro, A.; Caccialanza, G.; Massolini, G. *J Pharm Biomed Anal* **2014**, *87*, 218-228.
DOI: https://doi.org/10.1016/j.jpba.2013.03.007

(73) Kirchweger, B.; Rollinger, J. M. *Prog Chem Org Nat Prod* **2019**, *110*, 239-271.
DOI: https://doi.org/10.1007/978-3-030-14632-0_7

(74) Beutler, J. A. *Curr Protoc Pharmacol* **2019**, *86* (1), e67.
DOI: https://doi.org/10.1002/cpph.67

(75) Wilson, B. A. P.; Thornburg, C. C.; Henrich, C. J.; Grkovic, T.; O'Keefe, B. R. *Nat Prod Rep* **2020**, *37* (7), 893-918.
DOI: https://doi.org/10.1039/c9np00068b

(76) Pauli, G. F.; Chen, S. N.; Friesen, J. B.; McAlpine, J. B.; Jaki, B. U. *J Nat Prod* **2012**, *75* (6), 1243-1255.
DOI: https://doi.org/10.1021/np300066q

(77) Olgac, A.; Orhan, I. E.; Banoglu, E. *Future Med Chem* **2017**, *9* (14), 1665-1686.
DOI: https://doi.org/10.4155/fmc-2017-0124

(78) Seidel, T.; Wieder, O.; Garon, A.; Langer, T. *Mol Inform* **2020**, *39* (11), e2000059.
DOI: https://doi.org/10.1002/minf.202000059

(79) Huang, H.; Zhang, G.; Zhou, Y.; Lin, C.; Chen, S.; Lin, Y.; Mai, S.; Huang, Z. *Front Chem* **2018**, *6*, 138.
DOI: https://doi.org/10.3389/fchem.2018.00138

(80) Bassani, D.; Moro, S. *Molecules* **2023**, *28* (9).
DOI: https://doi.org/10.3390/molecules28093906

(81) Meijer, D.; Beniddir, M. A.; Coley, C. W.; Mejri, Y. M.; Ozturk, M.; van der Hooft, J. J. J.; Medema, M. H.; Skiredj, A. *Nat Prod Rep* **2025**, *42* (4), 654-662.
DOI: https://doi.org/10.1039/d4np00008k

(82) Chi, J.; Shu, J.; Li, M.; Mudappathi, R.; Jin, Y.; Lewis, F.; Boon, A.; Qin, X.; Liu, L.; Gu, H. *Trends Analyt Chem* **2024**, *178*.
DOI: https://doi.org/10.1016/j.trac.2024.117852

(83) Gupta, R.; Srivastava, D.; Sahu, M.; Tiwari, S.; Ambasta, R. K.; Kumar, P. *Mol Divers* **2021**, *25* (3), 1315-1360.
DOI: https://doi.org/10.1007/s11030-021-10217-3

(84) Hubert, J.; Nuzillard, J.-M.; Renault, J.-H. *Phytochemistry Reviews* **2015**, *16* (1), 55-95.
DOI: https://doi.org/10.1007/s11101-015-9448-7

(85) Wolfender, J. L.; Nuzillard, J. M.; van der Hooft, J. J. J.; Renault, J. H.; Bertrand, S. *Anal Chem* **2019**, *91* (1), 704-742.
DOI: https://doi.org/10.1021/acs.analchem.8b05112

(86) Bender, A.; Cortes-Ciriano, I. *Drug Discov Today* **2021**, *26* (4), 1040-1052.
DOI: https://doi.org/10.1016/j.drudis.2020.11.037

(87) Rollinger, J. M.; Wolber, G. Computational Approaches for the Discovery of Natural Lead Structures. In *Bioactive compounds from natural sources, Natural products as lead compounds in drug discovery*, CRC Press, Boca Raton, **2011**; pp 167-186.

(88) Puzyn, T.; Rasulev, B.; Gajewicz, A.; Hu, X.; Dasari, T. P.; Michalkova, A.; Hwang, H. M.; Toropov, A.; Leszczynska, D.; Leszczynski, J. *Nat Nanotechnol* **2011**, *6* (3), 175-178.
DOI: https://doi.org/10.1038/nnano.2011.10

(89) Moro, S.; Bacilieri, M.; Deflorian, F. *Expert Opin Drug Discov* **2007**, *2* (1), 37-49.
DOI: https://doi.org/10.1517/17460441.2.1.37

(90) Puzyn, T.; Leszczynska, D.; Leszczynski, J. *Small* **2009**, *5* (22), 2494-2509.
DOI: https://doi.org/10.1002/smll.200900179

(91) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, II; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. *J Med Chem* **2014**, *57* (12), 4977-5010.
DOI: https://doi.org/10.1021/jm4004285

(92) Syahdi, R. R.; Jasial, S.; Maeda, I.; Miyao, T. *ACS Omega* **2024**, *9* (37), 38957-38969.
DOI: https://doi.org/10.1021/acsomega.4c05433

(93) Lima, A. G.; Penteado, A. B.; de Jesus, J. G.; de Paula, V. J. R.; Ferraz, W. R.; Trossini, G. H. G. *Journal of the Brazilian Chemical Society* **2024**, *35* (10), 1-29.
DOI: https://doi.org/10.21577/0103-5053.20240112

(94) Davis, R. L. *iScience* **2020**, *23* (9), 101487.
DOI: https://doi.org/10.1016/j.isci.2020.101487

(95) Zhou, G.; Rusnac, D. V.; Park, H.; Canzani, D.; Nguyen, H. M.; Stewart, L.; Bush, M. F.; Nguyen, P. T.; Wulff, H.; Yarov-Yarovoy, V.; Zheng, N.; DiMaio, F. *Nat Commun* **2024**, *15* (1), 7761.
DOI: https://doi.org/10.1038/s41467-024-52061-7

(96) Wei, H.; McCammon, J. A. *npj Drug Discovery* **2024**, *1* (1).
DOI: https://doi.org/10.1038/s44386-024-00001-2

(97) Sams-Dodd, F. *Drug Discov Today* **2005**, *10* (2), 139-147.
DOI: https://doi.org/10.1016/S1359-6446(04)03316-1

(98) Zheng, W.; Thorne, N.; McKew, J. C. *Drug Discov Today* **2013**, *18* (21-22), 1067-1073.
DOI: https://doi.org/10.1016/j.drudis.2013.07.001

(99) von Delft, A.; Hall, M. D.; Kwong, A. D.; Purcell, L. A.; Saikatendu, K. S.; Schmitz, U.; Tallarico, J. A.; Lee, A. A. *Nat Rev Drug Discov* **2023**, *22* (7), 585-603.
DOI: https://doi.org/10.1038/s41573-023-00692-8

(100) Ullrich, S.; Nitsche, C. *Bioorg Med Chem Lett* **2020**, *30* (17), 127377.
DOI: https://doi.org/10.1016/j.bmcl.2020.127377

(101) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. *Science* **2020**, *368* (6489), 409-412.
DOI: https://doi.org/10.1126/science.abb3405

(102) Wu, C.; Liu, Y.; Yang, Y.; Zhang, P.; Zhong, W.; Wang, Y.; Wang, Q.; Xu, Y.; Li, M.; Li, X.; Zheng, M.; Chen, L.; Li, H. *Acta Pharm Sin B* **2020**, *10* (5), 766-788.
DOI: https://doi.org/10.1016/j.apsb.2020.02.008

(103) Lee, J. T.; Yang, Q.; Gribenko, A.; Perrin, B. S., Jr.; Zhu, Y.; Cardin, R.; Liberator, P. A.; Anderson, A. S.; Hao, L. *mBio* **2022**, *13* (4), e0086922.
DOI: https://doi.org/10.1128/mbio.00869-22

(104) Hillen, H. S.; Kokic, G.; Farnung, L.; Dienemann, C.; Tegunov, D.; Cramer, P. *Nature* **2020**, *584* (7819), 154-156.
DOI: https://doi.org/10.1038/s41586-020-2368-8

(105) Klemm, T.; Ebert, G.; Calleja, D. J.; Allison, C. C.; Richardson, L. W.; Bernardini, J. P.; Lu, B. G.; Kuchel, N. W.; Grohmann, C.; Shibata, Y.; Gan, Z. Y.; Cooney, J. P.; Doerflinger, M.; Au, A. E.; Blackmore, T. R.; van der Heden van Noort, G. J.; Geurink, P. P.; Ovaa, H.; Newman, J.; Riboldi-Tunnicliffe, A.; Czabotar, P. E.; Mitchell, J. P.; Feltham, R.; Lechtenberg, B. C.; Lowes, K. N.; Dewson, G.; Pellegrini, M.; Lessene, G.; Komander, D. *EMBO J* **2020**, *39* (18), e106275.
DOI: https://doi.org/10.15252/embj.2020106275

(106) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. *Nature* **2020**, *582* (7811), 289-293.
DOI: https://doi.org/10.1038/s41586-020-2223-y

(107) Prada Gori, D. N.; Ruatta, S.; Fló, M.; Alberca, L. N.; Bellera, C. L.; Park, S.; Heo, J.; Lee, H.; Park, K.-H. P.; Pritsch, O.; Shum, D.; Comini, M. A.; Talevi, A. *Frontiers in Drug Discovery* **2023**, *2*.
DOI: https://doi.org/10.3389/fddsv.2022.1082065

(108) Owen, D. R.; Allerton, C. M. N.; Anderson, A. S.; Aschenbrenner, L.; Avery, M.; Berritt, S.; Boras, B.; Cardin, R. D.; Carlo, A.; Coffman, K. J.; Dantonio, A.; Di, L.; Eng, H.; Ferre, R.; Gajiwala, K. S.; Gibson, S. A.; Greasley, S. E.; Hurst, B. L.; Kadar, E. P.; Kalgutkar, A. S.; Lee, J. C.; Lee, J.; Liu, W.; Mason, S. W.; Noell, S.; Novak, J. J.; Obach, R. S.; Ogilvie, K.; Patel, N. C.; Pettersson, M.; Rai, D. K.; Reese, M. R.; Sammons, M. F.; Sathish, J. G.; Singh, R. S. P.; Steppan, C. M.; Stewart, A. E.; Tuttle, J. B.; Updyke, L.; Verhoest, P. R.; Wei, L.; Yang, Q.; Zhu, Y. *Science* **2021**, *374* (6575), 1586-1593.
DOI: https://doi.org/10.1126/science.abl4784

(109) Unoh, Y.; Uehara, S.; Nakahara, K.; Nobori, H.; Yamatsu, Y.; Yamamoto, S.; Maruyama, Y.; Taoda, Y.; Kasamatsu, K.; Suto, T.; Kouki, K.; Nakahashi, A.; Kawashima, S.; Sanaki, T.; Toba, S.; Uemura, K.; Mizutare, T.; Ando, S.; Sasaki, M.; Orba, Y.; Sawa, H.; Sato, A.; Sato, T.; Kato, T.; Tachibana, Y. *J Med Chem* **2022**, *65* (9), 6499-6512.
DOI: https://doi.org/10.1021/acs.jmedchem.2c00117

(110) Tuttle, J. B.; Allais, C.; Allerton, C. M. N.; Anderson, A. S.; Arcari, J. T.; Aschenbrenner, L. M.; Avery, M.; Bellenger, J.; Berritt, S.; Boras, B.; Boscoe, B. P.; Buzon, L. M.; Cardin, R. D.; Carlo, A. A.; Coffman, K. J.; Dantonio, A.; Di, L.; Eng, H.; Farley, K. A.; Ferre, R. A.; Gajiwala, K. S.; Gibson, S. A.; Greasley, S. E.; Hurst, B. L.; Kadar, E. P.; Kalgutkar, A. S.; Lachapelle, E. A.; Lanyon, L. F.; Lee, J.; Lee, J.; Lian, Y.; Liu, W.; Martinez-Alsina, L. A.; Mason, S. W.; Noell, S.; Novak, J.; Obach, R. S.; Ogilvie, K.; O'Neil, S. V.; Ostner, G.; Owen, D. R.; Patel, N. C.; Pettersson, M.; Singh, R. S.; Rai, D. K.; Reese, M. R.; Sakata, S.; Sammons, M. F.; Sathish, J. G.; Sharma, R.; Steppan, C. M.; Stewart, A.; Updyke, L.; Verhoest, P. R.; Wei, L.; Wright, S. W.; Yang, E.; Yang, Q.; Zhu, Y. *J Med Chem* **2025**, *68* (7), 7003-7030.
DOI: https://doi.org/10.1021/acs.jmedchem.4c02561

(111) Eltayb, W. A.; Abdalla, M.; Rabie, A. M. *ACS Omega* **2023**, *8* (6), 5234-5246.
DOI: https://doi.org/10.1021/acsomega.2c03881

(112) Sathish, J. G.; Bhatt, S.; DaSilva, J. K.; Flynn, D.; Jenkinson, S.; Kalgutkar, A. S.; Liu, M.; Manickam, B.; Pinkstaff, J.; Reagan, W. J.; Shirai, N.; Shoieb, A. M.; Sirivelu, M.; Vispute, S.; Vitsky, A.; Walters, K.; Wisialowski, T. A.; Updyke, L. W. *Int J Toxicol* **2022**, *41* (4), 276-290.
DOI: https://doi.org/10.1177/10915818221095489

(113) Syed, Y. Y. *Drugs* **2024**, *84* (6), 721-728.
DOI: https://doi.org/10.1007/s40265-024-02039-y

(114) Ip, J. D.; Wing-Ho Chu, A.; Chan, W. M.; Cheuk-Ying Leung, R.; Umer Abdullah, S. M.; Sun, Y.; Kai-Wang To, K. *EBioMedicine* **2023**, *91*, 104559.
DOI: https://doi.org/10.1016/j.ebiom.2023.104559

(115) Biernacki, K.; Ciupak, O.; Dasko, M.; Rachon, J.; Flis, D.; Budka, J.; Inkielewicz-Stepniak, I.; Czaja, A.; Rak, J.; Demkowicz, S. *J Enzyme Inhib Med Chem* **2024**, *39* (1), 2290910.
DOI: https://doi.org/10.1080/14756366.2023.2290910

(116) Cao, Q.; Ding, Y.; Xu, Y.; Li, M.; Zheng, R.; Cao, Z.; Wang, W.; Bi, Y.; Ning, G.; Xu, Y.; Zhao, R. *Front Med* **2023**, *17* (6), 1068-1079.
DOI: https://doi.org/10.1007/s11684-023-1037-3

(117) Jiang, R.; Han, B.; Xu, W.; Zhang, X.; Peng, C.; Dang, Q.; Sun, W.; Lin, L.; Lin, Y.; Fan, L.; Lv, D.; Shao, L.; Chen, Y.; Qiu, Y.; Han, L.; Kong, W.; Li, G.; Wang, K.; Peng, J.; Lin, B.; Tong, Z.; Lu, X.; Wang, L.; Gao, F.; Feng, J.; Li, Y.; Ma, X.; Wang, J.; Wang, S.; Shen, W.; Wang, C.; Yan, K.; Lin, Z.; Jin, C.; Mao, L.; Liu, J.; Kushnareva, Y.; Kotoi, O.; Zhu, Z.; Royal, M.; Brunswick, M.; Ji, H.; Xu, X.; Lu, H. *NEJM Evid* **2024**, *3* (6), EVIDoa2400026.
DOI: https://doi.org/10.1056/EVIDoa2400026

(118) Borroto-Esoda, K.; Wilfret, D.; Tong, X.; Plummer, A.; Kearney, B.; Kwong, A. D. *Microbiol Spectr* **2024**, *12* (2), e0298023.
DOI: https://doi.org/10.1128/spectrum.02980-23

(119) Wasilewicz, A. K., Benjamin; Bojkova, Denisa; Abi Saad, Marie; Langeder, Julia; Bütikofer, Matthias; Adelsberger, Sigrid; Grienke, Ulrike; Cinatl, Jindrich; Petermann, Olivier; Scapozza, Leonardo; Orts, Julien; Kirchmair, Johannes; Rabenau, H.; Rollinger, Judith. **2024**.
DOI: https://doi.org/10.1021/acs.jnatprod.2c00843

(120) Luo, S. Y.; Tang, Z. Y.; Li, Q.; Weng, J.; Yin, S.; Tang, G. H. *J Org Chem* **2021**, *86* (6), 4786-4793.
DOI: https://doi.org/10.1021/acs.joc.1c00229

(121) Luo, S. Y.; Zhu, J. Y.; Zou, M. F.; Yin, S.; Tang, G. H. *Nat Prod Bioprospect* **2022**, *12* (1), 31.
DOI: https://doi.org/10.1007/s13659-022-00355-y

(122) Langeder, J.; Doring, K.; Schmietendorf, H.; Grienke, U.; Schmidtke, M.; Rollinger, J. M. *J Nat Prod* **2023**, *86* (1), 8-17.
DOI: https://doi.org/10.1021/acs.jnatprod.2c00481

(123) Grienke, U.; Richter, M.; Walther, E.; Hoffmann, A.; Kirchmair, J.; Makarov, V.; Nietzsche, S.; Schmidtke, M.; Rollinger, J. M. *Sci Rep* **2016**, *6*, 27156.
DOI: https://doi.org/10.1038/srep27156

(124) Walther, E.; Xu, Z.; Richter, M.; Kirchmair, J.; Grienke, U.; Rollinger, J. M.; Krumbholz, A.; Saluz, H. P.; Pfister, W.; Sauerbrei, A.; Schmidtke, M. *Front Microbiol* **2016**, *7*, 357.
DOI: https://doi.org/10.3389/fmicb.2016.00357

(125) Langeder, J.; Koch, M.; Schmietendorf, H.; Tahir, A.; Grienke, U.; Rollinger, J. M.; Schmidtke, M. *Front Pharmacol* **2023**, *14*, 1193118.
DOI: https://doi.org/10.3389/fphar.2023.1193118

(126) Guerra, Y.; Celi, D.; Cueva, P.; Perez-Castillo, Y.; Giampieri, F.; Alvarez-Suarez, J. M.; Tejera, E. *ACS Omega* **2022**, *7* (49), 44542-44555.
DOI: https://doi.org/10.1021/acsomega.2c05766

(127) Yan, H.; Zhang, R.; Yan, G.; Liu, X.; Li, N.; Chen, Y. *Phytother Res* **2024**, *38* (3), 1161-1164.
DOI: https://doi.org/10.1002/ptr.7829

(128) Hu, H.; Yi, X.; Xue, L.; Baell, J. B. *JACS Au* **2024**, *4* (12), 4883-4891.
DOI: https://doi.org/10.1021/jacsau.4c00851

(129) Beniddir, M. A.; Kang, K. B.; Genta-Jouve, G.; Huber, F.; Rogers, S.; van der Hooft, J. J. J. *Nat Prod Rep* **2021**, *38* (11), 1967-1993.
DOI: https://doi.org/10.1039/d1np00023c

(130) Wasilewicz, A.; Areesanan, A.; Kirchweger, B.; Nicolay, S.; Waltenberger, E.; Beniddir, M. A.; Grundemann, C.; Rollinger, J. M.; Grienke, U. *J Nat Prod* **2024**, *in print*.
DOI: https://doi.org/10.1021/acs.jnatprod.4c00847

(131) Kellogg, J. J.; Todd, D. A.; Egan, J. M.; Raja, H. A.; Oberlies, N. H.; Kvalheim, O. M.; Cech, N. B. *J Nat Prod* **2016**, *79* (2), 376-386.
DOI: https://doi.org/10.1021/acs.jnatprod.5b01014

(132) Fox Ramos, A. E.; Evanno, L.; Poupon, E.; Champy, P.; Beniddir, M. A. *Nat Prod Rep* **2019**, *36* (7), 960-980.
DOI: https://doi.org/10.1039/c9np00006b

(133) Martens, H.; Bruun, S. W.; Adt, I.; Sockalingum, G. D.; Kohler, A. *Journal of Chemometrics* **2007**, *20* (8-10), 402-417.
DOI: https://doi.org/10.1002/cem.1015

(134) Kvalheim, O. M.; Chan, H.-y.; Benzie, I. F. F.; Szeto, Y.-t.; Tzang, A. H.-c.; Mok, D. K.-w.; Chau, F.-t. *Chemometrics and Intelligent Laboratory Systems* **2011**, *107* (1), 98-105.
DOI: https://doi.org/10.1016/j.chemolab.2011.02.002

(135) Britton, E. R.; Kellogg, J. J.; Kvalheim, O. M.; Cech, N. B. *J Nat Prod* **2018**, *81* (3), 484-493.
DOI: https://doi.org/10.1021/acs.jnatprod.7b00654

(136) Inui, T.; Wang, Y.; Pro, S. M.; Franzblau, S. G.; Pauli, G. F. *Fitoterapia* **2012**, *83* (7), 1218-1225.
DOI: https://doi.org/10.1016/j.fitote.2012.06.012

(137) Qiu, F.; Cai, G.; Jaki, B. U.; Lankin, D. C.; Franzblau, S. G.; Pauli, G. F. *J Nat Prod* **2013**, *76* (3), 413-419.
DOI: https://doi.org/10.1021/np3007809

(138) Ramos Alvarenga, R. F.; Wan, B.; Inui, T.; Franzblau, S. G.; Pauli, G. F.; Jaki, B. U. *J Nat Prod* **2014**, *77* (3), 603-610.
DOI: https://doi.org/10.1021/np400872m

(139) Caesar, L. K.; Kellogg, J. J.; Kvalheim, O. M.; Cech, R. A.; Cech, N. B. *Planta Med* **2018**, *84* (9-10), 721-728.
DOI: https://doi.org/10.1055/a-0590-5223

(140) Salem, M. A.; Perez de Souza, L.; Serag, A.; Fernie, A. R.; Farag, M. A.; Ezzat, S. M.; Alseekh, S. *Metabolites* **2020**, *10* (1).
DOI: https://doi.org/10.3390/metabo10010037

(141) Weckwerth, W. *Bioanalysis* **2010**, *2* (4), 829-836.
DOI: https://doi.org/10.4155/bio.09.192

(142) Hou, Y.; Braun, D. R.; Michel, C. R.; Klassen, J. L.; Adnani, N.; Wyche, T. P.; Bugni, T. S. *Anal Chem* **2012**, *84* (10), 4277-4283.
DOI: https://doi.org/10.1021/ac202623g

(143) Aligiannis, N.; Halabalaki, M.; Chaita, E.; Kouloura, E.; Argyropoulou, A.; Benaki, D.; Kalpoutzakis, E.; Angelis, A.; Stathopoulou, K.; Antoniou, S.; Sani, M.; Krauth, V.; Werz, O.; Schütz, B.; Schäfer, H.; Spraul, M.; Mikros, E.; Skaltsounis, L. A. *ChemistrySelect* **2016**, *1* (10), 2531-2535.
DOI: https://doi.org/10.1002/slct.201600744

(144) Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M.; Moore, B. S.; Laskin, J.; Bandeira, N.; Dorrestein, P. C. *Proceedings of the National Academy of Sciences* **2012**, *109* (26), E1743-E1752.
DOI: https://doi.org/10.1073/pnas.1203689109

(145) Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; de Felicio, R.; Fenner, A.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonsi, H. M.; Gerwick, W. H.; Dorrestein, P. C. *J Nat Prod* **2013**, *76* (9), 1686-1699.
DOI: https://doi.org/10.1021/np400413s

(146) Frank, A. M.; Bandeira, N.; Shen, Z.; Tanner, S.; Briggs, S. P.; Smith, R. D.; Pevzner, P. A. *Journal of Proteome Research* **2008**, *7* (1), 113-122.
DOI: https://doi.org/10.1021/pr070361e

(147) Wan, K. X.; Vidavsky, I.; Gross, M. L. *Journal of the American Society for Mass Spectrometry* **2002**, *13* (1), 85-88.
DOI: https://doi.org/10.1016/S1044-0305(01)00327-0

(148) Nothias, L. F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; Aicheler, F.; Aksenov, A.; Alka, O.; Allard, P.-M.; Barsch, A.; Cachet, X.; Caraballo, M.; Da Silva, R. R.; Dang, T.; Garg, N.; Gauglitz, J. M.; Gurevich, A.; Isaac, G.; Jarmusch, A. K.; Kameník, Z.; Kang, K. B.; Kessler, N.; Koester, I.; Korf, A.; Gouellec, A. L.; Ludwig, M.; Christian, M. H.; McCall, L.-I.; McSayles, J.; Meyer, S. W.; Mohimani, H.; Morsy, M.; Moyne, O.; Neumann, S.; Neuweger, H.; Nguyen, N. H.; Nothias-Esposito, M.; Paolini, J.; Phelan, V. V.; Pluskal, T.; Quinn, R. A.; Rogers, S.; Shrestha, B.; Tripathi, A.; van der Hooft, J. J. J.; Vargas, F.; Weldon, K. C.; Witting, M.; Yang, H.; Zhang, Z.; Zubeil, F.; Kohlbacher, O.; Böcker, S.; Alexandrov, T.; Bandeira, N.; Wang, M.; Dorrestein, P. C. *bioRxiv* **2019**, 812404.
DOI: https://doi.org/10.1101/812404

(149) Wasilewicz, A.; Bojkova, D.; Beniddir, M. A.; Cinatl, J., Jr.; Rabenau, H. F.; Grienke, U.; Rollinger, J. M.; Kirchweger, B. *J Ethnopharmacol* **2024**, *319* (Pt 2), 117206.
DOI: https://doi.org/10.1016/j.jep.2023.117206

(150) Wasilewicz, A.; Zwirchmayr, J.; Kirchweger, B.; Bojkova, D.; Cinatl, J., Jr.; Rabenau, H. F.; Rollinger, J. M.; Beniddir, M. A.; Grienke, U. *Front Mol Biosci* **2023**, *10*, 1202394.
DOI: https://doi.org/10.3389/fmolb.2023.1202394

(151) Perez De Souza, L.; Alseekh, S.; Brotman, Y.; Fernie, A. R. *Expert Rev Proteomics* **2020**, *17* (4), 243-255.
DOI: https://doi.org/10.1080/14789450.2020.1766975

(152) Amountzias, V.; Cheilari, A.; Vontzalidou, A.; Benaki, D.; Gikas, E.; Aligiannis, N. *Anal Chem* **2024**, *96* (50), 20090-20097.
DOI: https://doi.org/10.1021/acs.analchem.4c05080

(153) Crockford, D. J.; Holmes, E.; Lindon, J. C.; Plumb, R. S.; Zirah, S.; Bruce, S. J.; Rainville, P.; Stumpf, C. L.; Nicholson, J. K. *Anal Chem* **2006**, *78* (2), 363-371.
DOI: https://doi.org/10.1021/ac051444m

(154) Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. *Anal Chem* **2005**, *77* (5), 1282-1289.
DOI: https://doi.org/10.1021/ac048630x

(155) Chaita, E.; Gikas, E.; Aligiannis, N. *Phytochem Anal* **2017**, *28* (2), 125-131.
DOI: https://doi.org/10.1002/pca.2670

(156) Boka, V.-I.; Stathopoulou, K.; Benaki, D.; Gikas, E.; Aligiannis, N.; Mikros, E.; Skaltsounis, A.-L. *Phytochemistry Letters* **2017**, *20*, 379-385.
DOI: https://doi.org/10.1016/j.phytol.2017.03.001

(157) Michalea, R.; Stathopoulou, K.; Polychronopoulos, P.; Benaki, D.; Mikros, E.; Aligiannis, N. *J Ethnopharmacol* **2020**, *257*, 111547.
DOI: https://doi.org/10.1016/j.jep.2018.10.008

(158) Grienke, U.; Foster, P. A.; Zwirchmayr, J.; Tahir, A.; Rollinger, J. M.; Mikros, E. *Scientific Reports* **2019**, *9* (1), 11113.
DOI: https://doi.org/10.1038/s41598-019-47434-8

(159) Zwirchmayr, J.; Grienke, U.; Hummelbrunner, S.; Seigner, J.; de Martin, R.; Dirsch, V. M.; Rollinger, J. M. *Biomolecules* **2020**, *10* (5), 679.
DOI: https://doi.org/10.3390/biom10050679

(160) Tahir, A.; Kampleitner, C.; Wirglauer, T.; Grienke, U.; Hoffmann, O.; Rollinger, J. M. *Molecules* **2022**, *27* (11), 3611.
DOI: https://doi.org/10.3390/molecules27113611

(161) Zwirchmayr, J.; Schachner, D.; Grienke, U.; Rudzionyte, I.; de Martin, R.; Dirsch, V. M.; Rollinger, J. M. *Phytochemistry* **2023**, *212*, 113709.
DOI: https://doi.org/10.1016/j.phytochem.2023.113709

(162) Dudley, J. A.; Park, S.; MacDonald, M. E.; Fetene, E.; Smith, C. A. *J Magn Reson* **2020**, *318*, 106773.
DOI: https://doi.org/10.1016/j.jmr.2020.106773

(163) Korzhneva, D. M.; Ibraghimov, I. V.; Billeter, M.; Orekhov, V. Y. *J Biomol NMR* **2001**, *21* (3), 263-268.
DOI: https://doi.org/10.1023/a:1012982830367

(164) Egan, J. M.; van Santen, J. A.; Liu, D. Y.; Linington, R. G. *J Nat Prod* **2021**, *84* (4), 1044-1055.
DOI: https://doi.org/10.1021/acs.jnatprod.0c01076

(165) Cardenas, P. D.; Almeida, A.; Bak, S. *Front Plant Sci* **2019**, *10*, 1523.
DOI: https://doi.org/10.3389/fpls.2019.01523

(166) Acebey-Castellon, I. L.; Voutquenne-Nazabadioko, L.; Doan Thi Mai, H.; Roseau, N.; Bouthagane, N.; Muhammad, D.; Le Magrex Debar, E.; Gangloff, S. C.; Litaudon, M.; Sevenet, T.; Hung, N. V.; Lavaud, C. *J Nat Prod* **2011**, *74* (2), 163-168.
DOI: https://doi.org/10.1021/np100502y

(167) Martelanc, M.; Vovk, I.; Simonovska, B. *J Chromatogr A* **2009**, *1216* (38), 6662-6670.
DOI: https://doi.org/10.1016/j.chroma.2009.07.038

(168) Vilkickyte, G.; Raudone, L. *Molecules* **2021**, *26* (6), 1645.
DOI: https://doi.org/10.3390/molecules26061645

(169) Rhourri-Frih, B.; Chaimbault, P.; Claude, B.; Lamy, C.; Andre, P.; Lafosse, M. *J Mass Spectrom* **2009**, *44* (1), 71-80.
DOI: https://doi.org/10.1002/jms.1472

(170) Xia, Q.; Zhang, H.; Sun, X.; Zhao, H.; Wu, L.; Zhu, D.; Yang, G.; Shao, Y.; Zhang, X.; Mao, X.; Zhang, L.; She, G. *Molecules* **2014**, *19* (11), 17478-17535.
DOI: https://doi.org/10.3390/molecules191117478

(171) Qiu, F.; McAlpine, J. B.; Lankin, D. C.; Burton, I.; Karakach, T.; Chen, S. N.; Pauli, G. F. *Anal. Chem.* **2014**, *86* (8), 3964-3972.
DOI: https://doi.org/10.1021/ac500188j

(172) Bakiri, A.; Hubert, J.; Reynaud, R.; Lambert, C.; Martinez, A.; Renault, J. H.; Nuzillard, J. M. *J Chem Inf Model* **2018**, *58* (2), 262-270.
DOI: https://doi.org/10.1021/acs.jcim.7b00653

(173) Borges, R. M.; Ferreira, G. A.; Campos, M. M.; Teixeira, A. M.; Costa, F. D. N.; das Chagas, F. O.; Colonna, M. *Phytochem Anal* **2023**, *34* (4), 385-392.
DOI: https://doi.org/10.1002/pca.3229

(174) Braunschweiler, L.; Ernst, R. R. *J. Magn. Reson.* **1983**, *53*, 521-528.
DOI: https://doi.org/10.1016/0022-2364(83)90226-3

(175) Borges, R. M.; das Neves Costa, F.; Chagas, F. O.; Teixeira, A. M.; Yoon, J.; Weiss, M. B.; Crnkovic, C. M.; Pilon, A. C.; Garrido, B. C.; Betancur, L. A.; Forero, A. M.; Castellanos, L.; Ramos, F. A.; Pupo, M. T.; Kuhn, S. *Phytochem Anal* **2023**, *34* (1), 48-55.
DOI: https://doi.org/10.1002/pca.3178

(176) Margueritte, L.; Duciel, L.; Bourjot, M.; Vonthron-Senecheau, C.; Delsuc, M. A. *Faraday Discuss* **2019**, *218*, 441-458.
DOI: https://doi.org/10.1039/c8fd00242h

(177) Zhu, X.; Wang, L.; Zhao, T.; Jiang, Q. *J Ethnopharmacol* **2022**, *298*, 115566.
DOI: https://doi.org/10.1016/j.jep.2022.115566
(178) Kim, T. M.; Paudel, K. R.; Kim, D. W. *J Ethnopharmacol* **2020**, *253*, 112082.
DOI: https://doi.org/10.1016/j.jep.2019.112082

(179) Ito, H.; Kobayashi, E.; Li, S. H.; Hatano, T.; Sugita, D.; Kubo, N.; Shimura, S.; Itoh, Y.; Tokuda, H.; Nishino, H.; Yoshida, T. *J Agric Food Chem* **2002**, *50* (8), 2400-2403.
DOI: https://doi.org/10.1021/jf011083l

(180) Huang, Y.; Li, J.; Wang, R.; Wu, Q.; Li, Y. H.; Yu, S. C.; Cheng, W. M.; Wang, Y. Y. *Inflamm Res* **2007**, *56* (2), 76-82.
DOI: https://doi.org/10.1007/s00011-006-5185-0

(181) Uto, T.; Sakamoto, A.; Tung, N. H.; Fujiki, T.; Kishihara, K.; Oiso, S.; Kariyazono, H.; Morinaga, O.; Shoyama, Y. *Int J Mol Sci* **2013**, *14* (2), 4106-4120.
DOI: https://doi.org/10.3390/ijms14024106

(182) Zhang, J.; Li, Y.; Chen, S. S.; Zhang, L.; Wang, J.; Yang, Y.; Zhang, S.; Pan, Y.; Wang, Y.; Yang, L. *Int J Mol Sci* **2015**, *16* (2), 2913-2941.
DOI: https://doi.org/10.3390/ijms16022913

(183) Liu, Y.; Zhang, W.; Xu, C.; Li, X. *Int J Mol Sci* **2016**, *17* (12), 1983.
DOI: https://doi.org/10.3390/ijms17121983

(184) Wu, Y.-x.; Jian, T.-y.; Lv, H.; Ding, X.-q.; Zuo, Y.-y.; Ren, B.-r.; Chen, J.; Li, W.-l. *Brazilian Journal of Pharmacognosy* **2018**, *28* (2), 239-242.
DOI: https://doi.org/10.1016/j.bjp.2018.02.006

(185) Mogole, L.; Omwoyo, W.; Mtunzi, F. *Heliyon* **2020**, *6* (8), e04736.
DOI: https://doi.org/10.1016/j.heliyon.2020.e04736

(186) Kuraoka-Oliveira, A. M.; Radai, J. A. S.; Leitao, M. M.; Lima Cardoso, C. A.; Silva-Filho, S. E.; Leite Kassuya, C. A. *J Ethnopharmacol* **2020**, *249*, 112418.
DOI: https://doi.org/10.1016/j.jep.2019.112418

(187) Oman, T.; Tessem, M. B.; Bathen, T. F.; Bertilsson, H.; Angelsen, A.; Hedenstrom, M.; Andreassen, T. *BMC Bioinformatics* **2014**, *15* (1), 413.
DOI: https://doi.org/10.1186/s12859-014-0413-z

(188) Simmler, C.; Napolitano, J. G.; McAlpine, J. B.; Chen, S. N.; Pauli, G. F. *Curr Opin Biotechnol* **2014**, *25*, 51-59.
DOI: https://doi.org/10.1016/j.copbio.2013.08.004

(189) Rubim de Santana, P. I.; Diz de Almeida, J. S. F.; Franca, T. C. C.; Junker, J. *J Anal Methods Chem* **2022**, *2022*, 7490691.
DOI: https://doi.org/10.1155/2022/7490691

(190) Jaki, B. U.; Bzhelyansky, A.; Pauli, G. F. *Magn Reson Chem* **2021**, *59* (1), 7-15.
DOI: https://doi.org/10.1002/mrc.5099

(191) Martineau, E.; Dumez, J. N.; Giraudeau, P. *Magn Reson Chem* **2020**, *58* (5), 390-403.
DOI: https://doi.org/10.1002/mrc.4899

(192) Farjon, J. Intrinsically quantitative 2D HSQC NMR: A tool for deciphering complex mixtures. Annual Reports on NMR Spectroscopy, Vol. 109; 2023; pp 1-27.

(193) Giraudeau, P. *Magn Reson Chem* **2014**, *52* (6), 259-272.
DOI: https://doi.org/10.1002/mrc.4068

(194) Heikkinen, S.; Toikka, M. M.; Karhunen, P. T.; Kilpelainen, I. A. *J Am Chem Soc* **2003**, *125* (14), 4362-4367.
DOI: https://doi.org/10.1021/ja029035k

(195) Rai, R. K.; Tripathi, P.; Sinha, N. *Anal Chem* **2009**, *81* (24), 10232-10238.
DOI: https://doi.org/10.1021/ac902405z

(196) Pathan, M.; Akoka, S.; Tea, I.; Charrier, B.; Giraudeau, P. *Analyst* **2011**, *136* (15), 3157-3163.
DOI: https://doi.org/10.1039/c1an15278e

(197) Rai, R. K.; Sinha, N. *Anal Chem* **2012**, *84* (22), 10005-10011.
DOI: https://doi.org/10.1021/ac302457s

(198) Wu, Y. S.; Li, B. X.; Long, Y. Y. *RSC Adv* **2022**, *12* (9), 5349-5356.
DOI: https://doi.org/10.1039/d1ra08423b

(199) Reher, R.; Kim, H. W.; Zhang, C.; Mao, H. H.; Wang, M.; Nothias, L. F.; Caraballo-Rodriguez, A. M.; Glukhov, E.; Teke, B.; Leao, T.; Alexander, K. L.; Duggan, B. M.; Van Everbroeck, E. L.; Dorrestein, P. C.; Cottrell, G. W.; Gerwick, W. H. *J Am Chem Soc* **2020**, *142* (9), 4114-4120.
DOI: https://doi.org/10.1021/jacs.9b13786

(200) Wang, T.; Pulkkinen, O. I.; Aittokallio, T. *Front Pharmacol* **2022**, *13*, 1003480.
DOI: https://doi.org/10.3389/fphar.2022.1003480

(201) Artemenko, A. G.; Muratov, E. N.; Kuz'min, V. E.; Kovdienko, N. A.; Hromov, A. I.; Makarov, V. A.; Riabova, O. B.; Wutzler, P.; Schmidtke, M. *J Antimicrob Chemother* **2007**, *60* (1), 68-77.
DOI: https://doi.org/10.1093/jac/dkm172

(202) Indrayanto, G.; Putra, G. S.; Suhud, F. *Profiles Drug Subst Excip Relat Methodol* **2021**, *46*, 273-307.
DOI: https://doi.org/10.1016/bs.podrm.2020.07.005

(203) Bosc, N.; Meyer, C.; Bonnet, P. *BMC Bioinformatics* **2017**, *18* (1), 17.
DOI: https://doi.org/10.1186/s12859-016-1413-y

(204) Pena-Moran, O. A.; Villarreal, M. L.; Alvarez-Berber, L.; Meneses-Acosta, A.; Rodriguez-Lopez, V. *Molecules* **2016**, *21* (8).
DOI: https://doi.org/10.3390/molecules21081013

(205) Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L. *Br J Pharmacol* **2011**, *162* (6), 1239-1249.
DOI: https://doi.org/10.1111/j.1476-5381.2010.01127.x

(206) Ambreen, S.; Umar, M.; Noor, A.; Jain, H.; Ali, R. *Eur J Med Chem* **2025**, *284*, 117164.
DOI: https://doi.org/10.1016/j.ejmech.2024.117164

(207) Anuta, V.; Nitulescu, G. M.; Dinu-Pirvu, C. E.; Olaru, O. T. *Molecules* **2014**, *19* (10), 16381-16401.
DOI: https://doi.org/10.3390/molecules191016381

(208) Hann, M. M.; Keseru, G. M. *Nat Rev Drug Discov* **2012**, *11* (5), 355-365.
DOI: https://doi.org/10.1038/nrd3701

(209) Lohani, S.; Cooper, H.; Jin, X.; Nissley, B. P.; Manser, K.; Rakes, L. H.; Cummings, J. J.; Fauty, S. E.; Bak, A. *J Pharm Sci* **2014**, *103* (10), 3007-3021.
DOI: https://doi.org/10.1002/jps.24088

(210) Venkatesh, S.; Lipper, R. A. *J Pharm Sci* **2000**, *89* (2), 145-154.
DOI: https://doi.org/10.1002/(SICI)1520-6017(200002)89:2<145::AID-JPS2>3.0.CO;2-6

(211) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Advanced drug delivery reviews* **1997**, *23*, 3-25.
DOI: https://doi.org/10.1016/S0169-409X(96)00423-1

(212) Leo, A.; Hansch, C.; Elkins, D. *Chemical Reviews* **1971**, *71* (6), 525-616.
DOI: https://doi.org/10.1021/cr60274a001

(213) Doak, B. C.; Over, B.; Giordanetto, F.; Kihlberg, J. *Chem Biol* **2014**, *21* (9), 1115-1142.
DOI: https://doi.org/10.1016/j.chembiol.2014.08.013

(214) Bergstrom, C. A.; Holm, R.; Jorgensen, S. A.; Andersson, S. B.; Artursson, P.; Beato, S.; Borde, A.; Box, K.; Brewster, M.; Dressman, J.; Feng, K. I.; Halbert, G.; Kostewicz, E.; McAllister, M.; Muenster, U.; Thinnes, J.; Taylor, R.; Mullertz, A. *Eur J Pharm Sci* **2014**, *57*, 173-199.
DOI: https://doi.org/10.1016/j.ejps.2013.10.015

(215) Kerns, E. H.; Di, L. *Drug Discov Today* **2003**, *8* (7), 316-323.
DOI: https://doi.org/10.1016/s1359-6446(03)02649-7

(216) Avdeef, A. *Curr Top Med Chem* **2001**, *1* (4), 277-351.
DOI: https://doi.org/10.2174/1568026013395100

(217) Yazdanian, M. *Curr Protoc Pharmacol* **2013**, *Chapter 9*, Unit9 17.
DOI: https://doi.org/10.1002/0471141755.ph0917s60

(218) Panchagnula, R.; Thomas, N. S. *Int J Pharm* **2000**, *201* (2), 131-150.
DOI: https://doi.org/10.1016/s0378-5173(00)00344-6

(219) Morak-Mlodawska, B.; Jelen, M.; Martula, E.; Korlacki, R. *Int J Mol Sci* **2023**, *24* (8).
DOI: https://doi.org/10.3390/ijms24086970

(220) Blume, H. H.; Schug, B. S. *Eur J Drug Metab Pharmacokinet* **2000**, *25* (1), 41-48.
DOI: https://doi.org/10.1007/BF03190057

(221) Amidon, G. L.; Lennernas, H.; Shah, V. P.; Crison, J. R. *Pharm Res* **1995**, *12* (3), 413-420.
DOI: https://doi.org/10.1023/a:1016212804288

(222) Reddy, B. B. K.; Karunakar, A. *Dissolution Technologies* **2011**, *18* (1), 31-37.
DOI: https://doi.org/10.14227/dt180111p31

(223) Varma, M. V.; Gardner, I.; Steyn, S. J.; Nkansah, P.; Rotter, C. J.; Whitney-Pickett, C.; Zhang, H.; Di, L.; Cram, M.; Fenner, K. S.; El-Kattan, A. F. *Mol Pharm* **2012**, *9* (5), 1199-1212.
DOI: https://doi.org/10.1021/mp2004912
(224) Li, J.; Larregieu, C. A.; Benet, L. Z. *Chin J Nat Med* **2016**, *14* (12), 888-897.
DOI: https://doi.org/10.1016/S1875-5364(17)30013-4

(225) Harvey, A. L. *Drug Discov Today* **2008**, *13* (19-20), 894-901.
DOI: https://doi.org/10.1016/j.drudis.2008.07.004

(226) Panossian, A. *Front Pharmacol* **2023**, *14*, 1199516.
DOI: https://doi.org/10.3389/fphar.2023.1199516

(227) Wu, C. Y.; Benet, L. Z. *Pharm Res* **2005**, *22* (1), 11-23.
DOI: https://doi.org/10.1007/s11095-004-9004-4

(228) Machado, R. D.; de Morais, M. C.; da Conceicao, E. C.; Vaz, B. G.; Chaves, A. R.; Rezende, K. R. *J Pharm Biomed Anal* **2020**, *191*, 113593.
DOI: https://doi.org/10.1016/j.jpba.2020.113593

(229) Li, C.; Lin, G.; Zuo, Z. *Biopharm Drug Dispos* **2011**, *32* (8), 427-445.
DOI: https://doi.org/10.1002/bdd.771

(230) Zhao, Q.; Luan, X.; Zheng, M.; Tian, X. H.; Zhao, J.; Zhang, W. D.; Ma, B. L. *Pharmaceutics* **2020**, *12* (2).
DOI: https://doi.org/10.3390/pharmaceutics12020128

(231) Yang, Y.; Zhang, Z.; Li, S.; Ye, X.; Li, X.; He, K. *Fitoterapia* **2014**, *92*, 133-147.
DOI: https://doi.org/10.1016/j.fitote.2013.10.010

(232) Song, M.; Hang, T. J.; Zhang, Z.; Chen, H. Y. *Eur J Pharm Sci* **2007**, *32* (4-5), 247-253.
DOI: https://doi.org/10.1016/j.ejps.2007.07.007

(233) Joo, K. M.; Lee, J. H.; Jeon, H. Y.; Park, C. W.; Hong, D. K.; Jeong, H. J.; Lee, S. J.; Lee, S. Y.; Lim, K. M. *J Pharm Biomed Anal* **2010**, *51* (1), 278-283.
DOI: https://doi.org/10.1016/j.jpba.2009.08.013

(234) Wu, H.; Zhu, Z.; Zhang, G.; Zhao, L.; Zhang, H.; Zhu, D.; Chai, Y. *J Ethnopharmacol* **2009**, *125* (3), 444-449.
DOI: https://doi.org/10.1016/j.jep.2009.07.019

(235) Zhang, L.; Yan, R.; Su, R.; Yang, C.; Liu, S.; Yu, X.; Chang, X.; Zhang, S.; Liu, C.; Xu, M.; Zeng, W.; Chen, Y.; Wang, Q. *J Ethnopharmacol* **2014**, *152* (2), 266-271.
DOI: https://doi.org/10.1016/j.jep.2013.09.045

(236) Zhang, W.; Zhang, H.; Sun, S.; Sun, F. F.; Chen, J.; Zhao, L.; Zhang, G. Q. *Molecules* **2015**, *20* (1), 1560-1570.
DOI: https://doi.org/10.3390/molecules20011560

(237) Gilbert, B.; Alves, L. F. *Curr Med Chem* **2003**, *10* (1), 13-20.
DOI: https://doi.org/10.2174/0929867033368583

(238) Ou, X.; Yu, Z.; Pan, C.; Zheng, X.; Li, D.; Qiao, Z.; Zheng, X. *Front Pharmacol* **2025**, *16*, 1551368.
DOI: https://doi.org/10.3389/fphar.2025.1551368

(239) McAuley, J. L.; Gilbertson, B. P.; Trifkovic, S.; Brown, L. E.; McKimm-Breschkin, J. L. *Front Microbiol* **2019**, *10*, 39.
DOI: https://doi.org/10.3389/fmicb.2019.00039
(240) Laborda, P.; Wang, S. Y.; Voglmeir, J. *Molecules* **2016**, *21* (11).
DOI: https://doi.org/10.3390/molecules21111513

(241) Grienke, U.; Schmidtke, M.; von Grafenstein, S.; Kirchmair, J.; Liedl, K. R.; Rollinger, J. M. *Nat Prod Rep* **2012**, *29* (1), 11-36.
DOI: https://doi.org/10.1039/c1np00053e

(242) Sunagawa, S.; Higa, F.; Cash, H. L.; Tateyama, M.; Uno, T.; Fujita, J. *Influenza Other Respir Viruses* **2013**, *7* (1), 1-3.
DOI: https://doi.org/10.1111/j.1750-2659.2012.00351.x

(243) Abed, Y.; Saim-Mamoun, A.; Boivin, G. *Rev Med Virol* **2021**, *31* (3), e2175.
DOI: https://doi.org/10.1002/rmv.2175

(244) Soong, G.; Muir, A.; Gomez, M. I.; Waks, J.; Reddy, B.; Planet, P.; Singh, P. K.; Kaneko, Y.; Wolfgang, M. C.; Hsiao, Y. S.; Tong, L.; Prince, A. *J Clin Invest* **2006**, *116* (8), 2297-2305.
DOI: https://doi.org/10.1172/JCI27920

(245) Magesh, S.; Suzuki, T.; Miyagi, T.; Ishida, H.; Kiso, M. *J Mol Graph Model* **2006**, *25* (2), 196-207.
DOI: https://doi.org/10.1016/j.jmgm.2005.12.006

(246) Chavas, L. M.; Tringali, C.; Fusi, P.; Venerando, B.; Tettamanti, G.; Kato, R.; Monti, E.; Wakatsuki, S. *J Biol Chem* **2005**, *280* (1), 469-475.
DOI: https://doi.org/10.1074/jbc.M411506200

(247) McCullers, J. A.; Tuomanen, E. I. *Front Biosci* **2001**, *6*, D877-889.
DOI: https://doi.org/10.2741/mccullers

(248) Nakamura, S.; Davis, K. M.; Weiser, J. N. *J Clin Invest* **2011**, *121* (9), 3657-3665.
DOI: https://doi.org/10.1172/JCI57762

(249) Plotkowski, M. C.; Puchelle, E.; Beck, G.; Jacquot, J.; Hannoun, C. *Am Rev Respir Dis* **1986**, *134* (5), 1040-1044.
DOI: https://doi.org/10.1164/arrd.1986.134.5.1040

(250) Keil, J. M.; Rafn, G. R.; Turan, I. M.; Aljohani, M. A.; Sahebjam-Atabaki, R.; Sun, X. L. *J Med Chem* **2022**, *65* (20), 13574-13593.
DOI: https://doi.org/10.1021/acs.jmedchem.2c01258

(251) Siegel, S. J.; Roche, A. M.; Weiser, J. N. *Cell Host Microbe* **2014**, *16* (1), 55-67.
DOI: https://doi.org/10.1016/j.chom.2014.06.005

(252) Tong, H. H.; McIver, M. A.; Fisher, L. M.; DeMaria, T. F. *Microb Pathog* **1999**, *26* (2), 111-119.
DOI: https://doi.org/10.1006/mpat.1998.0257

(253) McCullers, J. A.; Bartmess, K. C. *J Infect Dis* **2003**, *187* (6), 1000-1009.
DOI: https://doi.org/10.1086/368163

(254) McCullers, J. A. *Nat Rev Microbiol* **2014**, *12* (4), 252-262.
DOI: https://doi.org/10.1038/nrmicro3231

(255) Buommino, E.; Scognamiglio, M.; Donnarumma, G.; Fiorentino, A.; D'Abrosca, B. *Mini Rev Med Chem* **2014**, *14* (14), 1169-1182.
DOI: https://doi.org/10.2174/1389557515666150101095853

(256) Lim, H. J.; Jin, H. G.; Woo, E. R.; Lee, S. K.; Kim, H. P. *J Ethnopharmacol* **2013**, *149* (1), 169-175.
DOI: https://doi.org/10.1016/j.jep.2013.06.017

(257) Chang, Y.-S.; Jin, H.-G.; Lee, H.; Lee, D.-S.; Woo, E.-R. *Natural Product Sciences* **2019**, *25* (3).
DOI: https://doi.org/10.20307/nps.2019.25.3.268

(258) Dat, N. T.; Binh, P. T.; Quynh le, T. P.; Huong, H. T.; Minh, C. V. *Immunopharmacol Immunotoxicol* **2012**, *34* (1), 84-88.
DOI: https://doi.org/10.3109/08923973.2011.580755

(259) Galindo Bedor, D. C.; Marchand, S.; Lamarche, I.; Laroche, J.; Pereira de Santana, D.; Couet, W. *Antimicrob Agents Chemother* **2016**, *60* (8), 5085-5087.
DOI: https://doi.org/10.1128/AAC.00909-16

(260) Gontijo, A. V. L.; Brillault, J.; Grégoire, N.; Lamarche, I.; Gobin, P.; Couet, W.; Marchand, S. *Antimicrobial Agents and Chemotherapy* **2014**, *58* (7), 3942-3949.
DOI: https://doi.org/doi:10.1128/AAC.02818-14

(261) Hagerman, J. K.; Hancock, K. E.; Klepser, M. E. *Expert Opin Drug Deliv* **2006**, *3* (1), 71-86.
DOI: https://doi.org/10.1517/17425247.3.1.71

(262) Marchand, S.; Gobin, P.; Brillault, J.; Baptista, S.; Adier, C.; Olivier, J. C.; Mimoz, O.; Couet, W. *Antimicrob Agents Chemother* **2010**, *54* (9), 3702-3707.
DOI: https://doi.org/10.1128/AAC.00411-10

(263) Marchand, S.; Grégoire, N.; Brillault, J.; Lamarche, I.; Gobin, P.; Couet, W. *Antimicrob Agents Chemother* **2015**, *59* (10), 6646-6647.
DOI: https://doi.org/10.1128/aac.01647-15

(264) Marchand, S.; Boisson, M.; Mehta, S.; Adier, C.; Mimoz, O.; Grégoire, N.; Couet, W. *Antimicrobial Agents and Chemotherapy* **2018**, *62* (10), e01261-01218.
DOI: https://doi.org/doi:10.1128/AAC.01261-18

(265) Marchand, S.; Grégoire, N.; Brillault, J.; Lamarche, I.; Gobin, P.; Couet, W. *Antimicrobial Agents and Chemotherapy* **2016**, *60* (5), 3196-3198.
DOI: https://doi.org/doi:10.1128/AAC.00165-16

(266) Gontijo, A. V.; Gregoire, N.; Lamarche, I.; Gobin, P.; Couet, W.; Marchand, S. *Antimicrob Agents Chemother* **2014**, *58* (7), 3950-3956.
DOI: https://doi.org/10.1128/AAC.02819-14

(267) Forbes, B.; Backman, P.; Cabal, A.; Clark, A.; Ehrhardt, C.; Hastedt, J. E.; Hickey, A. J.; Hochhaus, G.; Jiang, W.; Kassinos, S.; Kuehl, P. J.; Olsson, B.; Prime, D.; Son, Y. J.; Teague, S.; Tehler, U.; Wylie, J. *Mol Pharm* **2025**, *22* (4), 1740-1751.
DOI: https://doi.org/10.1021/acs.molpharmaceut.4c01534

(268) Hastedt, J. E.; Backman, P.; Cabal, A.; Clark, A.; Ehrhardt, C.; Forbes, B.; Hickey, A. J.; Hochhaus, G.; Jiang, W.; Kassinos, S.; Kuehl, P. J.; Prime, D.; Son, Y. J.; Teague, S.; Tehler, U.; Wylie, J. *Mol Pharm* **2022**, *19* (7), 2032-2039.
DOI: https://doi.org/10.1021/acs.molpharmaceut.2c00113

(269) Hastedt, J. E.; Backman, P.; Cabal, A.; Clark, A.; Ehrhardt, C.; Forbes, B.; Hickey, A. J.; Hochhaus, G.; Jiang, W.; Kassinos, S.; Kuehl, P. J.; Prime, D.; Son, Y. J.; Teague, S.; Tehler, U.; Wylie, J. *Mol Pharm* **2024**, *21* (1), 164-172.
DOI: https://doi.org/10.1021/acs.molpharmaceut.3c00685

(270) Floroiu, A.; Loretz, B.; Kramer, J.; Lehr, C. M. *Eur J Pharm Biopharm* **2024**, *197*, 114206.
DOI: https://doi.org/10.1016/j.ejpb.2024.114206

(271) Schwarzinger, J.; Adelsberger, S.; Ortmayr, K.; Stellnberger, S. L.; Tahir, A.; Hadrich, G.; Pichler, V.; Rollinger, J. M.; Grienke, U.; Dailey, L. A. *Int J Pharm X* **2024**, *8*, 100272.
DOI: https://doi.org/10.1016/j.ijpx.2024.100272

(272) Choi, Y. H.; van Spronsen, J.; Dai, Y.; Verberne, M.; Hollmann, F.; Arends, I. W.; Witkamp, G. J.; Verpoorte, R. *Plant Physiol* **2011**, *156* (4), 1701-1705.
DOI: https://doi.org/10.1104/pp.111.178426

(273) Wagner, H.; Ulrich-Merzenich, G. *Phytomedicine* **2009**, *16* (2-3), 97-110.
DOI: https://doi.org/10.1016/j.phymed.2008.12.018

(274) Zhu, M. L.; Liang, X. L.; Zhao, L. J.; Liao, Z. G.; Zhao, G. W.; Cao, Y. C.; Zhang, J.; Luo, Y. *J Ethnopharmacol* **2013**, *150* (2), 553-559.
DOI: https://doi.org/10.1016/j.jep.2013.09.011

(275) Zheng, D.; Ruan, Y. T.; Yin, Z. P.; Zhang, Q. F. *Molecules* **2020**, *25* (20).
DOI: https://doi.org/10.3390/molecules25204728

(276) Qiang, Z.; Ye, Z.; Hauck, C.; Murphy, P. A.; McCoy, J. A.; Widrlechner, M. P.; Reddy, M. B.; Hendrich, S. *J Ethnopharmacol* **2011**, *137* (3), 1107-1112.
DOI: https://doi.org/10.1016/j.jep.2011.07.037

(277) Diesing, A. K.; Nossol, C.; Danicke, S.; Walk, N.; Post, A.; Kahlert, S.; Rothkotter, H. J.; Kluess, J. *PLoS One* **2011**, *6* (2), e17472.
DOI: https://doi.org/10.1371/journal.pone.0017472

(278) Manford, F.; Tronde, A.; Jeppsson, A. B.; Patel, N.; Johansson, F.; Forbes, B. *Eur J Pharm Sci* **2005**, *26* (5), 414-420.
DOI: https://doi.org/10.1016/j.ejps.2005.07.010

(279) Backman, P.; Cabal, A.; Clark, A.; Ehrhardt, C.; Forbes, B.; Hastedt, J.; Hickey, A.; Hochhaus, G.; Jiang, W.; Kassinos, S.; Kuehl, P. J.; Prime, D.; Son, Y. J.; Teague, S. P.; Tehler, U.; Wylie, J. *Mol Pharm* **2022**, *19* (7), 2040-2047.
DOI: https://doi.org/10.1021/acs.molpharmaceut.2c00112