

Evaluation of pollen secondary metabolites in exacerbation of non-allergic asthma, and development of computer-assisted LC-MS batch processing, clustering, and visualization

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Alen Bozicevic

aus

Basel, Basel-Stadt, Schweiz

Basel, 2017

Originaldokument gespeichert auf dem Dokumentenserver der Universität Basel

edoc.unibas.ch



Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Matthias Hamburger
Prof. Dr. Jean-Luc Wolfender

Basel, den 20. Juni 2017

Prof. Dr. Martin Spiess
Dekan

*Moju doktorsku disertaciju posvjećujem mojoj ženi i mojoj obitelji
što su mi uvijek bili blizu*

*I dedicate my doctoral thesis to my wife and my family
for always being on my side*

*Dedico la mia tesi di dottorato a mia moglie e alla mia famiglia
per essere stati sempre al mio fianco*

*Dedico minha tese de doutorado à minha esposa e minha família
por estarem sempre ao meu lado*

“... Human beings are like that. We want to live by each other’s happiness - not by each other’s misery. We don’t want to hate and despise one another. In this world there is room for everyone. And the good earth is rich and can provide for everyone. The way of life can be free and beautiful, but we have lost the way.

Greed has poisoned men’s souls, has barricaded the world with hate and has goose-stepped us into misery and bloodshed. We have developed speed, but we have shut ourselves in. Machinery that gives abundance has left us in want. Our knowledge has made us cynical, our cleverness hard and unkind. We think too much and feel too little. More than machinery we need humanity. More than cleverness we need kindness and gentleness. Without these qualities, life will be violent and all will be lost....

... Let us fight for a world of reason, a world where science and progress will lead to all men’s happiness....”

-The Great Dictator – Charlie Chaplin, 1940

Table of contents

SUMMARY	1
ZUSAMMENFASSUNG	4
1. AIM OF THE WORK	8
2. INTRODUCTION	11
2.1 Origin and uses of natural products	12
2.2 Strategies in analysis and discovery of natural products	15
<i>Progress from targeted to comprehensive profiling approach</i>	15
<i>HPLC-based activity profiling</i>	16
<i>Mass spectrometry (MS)</i>	16
<i>Nuclear magnetic resonance (NMR)</i>	19
2.3 Profiling, comprehensive processing, and visualization of GC/LC-MS data	24
2.4 Saponins are natural products with deterative properties	27
<i>Biosynthesis of saponins</i>	28
<i>Bioactivity of saponins</i>	30
2.5 Asthma	34
<i>Historical aspects of asthma</i>	34
<i>Pathophysiology of asthma</i>	35
<i>Causes of increased prevalence of asthma and allergies in Western countries</i>	38
2.6 Neuro-muscular pathways regulating airways mechanics	41
<i>Transient Receptor Potential Cation Channel member A1 (TRPA1)</i>	43
<i>Electrophiles and non-electrophiles with TRPA1 modulation properties</i>	45
<i>Type 1 Phosphatidylinositol 4-Phosphate 5-Kinase gamma (PIP5Kγ)</i>	45
<i>In vitro bioassay for assessing intracellular Ca²⁺ influx</i>	46

<i>Isolation and reparation of murine afferent neurons for Ca²⁺ imaging</i>	48
<i>In vitro bioassay for assessing tracheal tone modulation</i>	49
3. RESULTS AND DISCUSSION	52
3.1 Dammarane-type saponins from leaves of <i>Ziziphus spina-christi</i>	53
3.2 Automated comparative metabolite profiling of large LC-ESIMS datasets in an ACD/Labs add-in, and data clustering on a new open-source web platform FreeClust	138
3.3 Secondary metabolites in allergic plant pollen modulate murine afferent neurons and tracheal rings	168
4. CONCLUSIONS AND PERSPECTIVES	211

ACKNOWLEDGEMENTS

The period of my post graduate studies is going to be remembered as four very exciting, challenging, and educative years of my lifetime. For this extremely valuable experience, I would like to gratefully acknowledge many people. First, my big thanks go to my supervisor Prof. Dr. Matthias Hamburger for his valuable support, guidance, and enthusiastic encouragement during my projects and for giving me the freedom to explore my own ideas throughout my work.

Very big thanks go to my colleague Dr. Maria De Mieri for her excellent contribution and strong dedication in the NMR data evaluation and who helped me to deepen my knowledge in NMR. Thank you Maria, very much appreciated.

I would like to thank my colleague Dr. Eliane Garo for the excellent support, valuable advices, and generous help she gave me during the development of the analytical tool with ACD/Labs.

I would like to thank Dr. Christina Nassenstein for giving me the opportunity to visit her lab at the Anatomy and Cell Biology Institute in Giessen (Germany) where I was introduced into the asthma mouse model and where I had the possibility to have hands-on experience with different bioassays. Thank you for your great hospitality and your engagement in evaluating *in vitro* data!

I would like to thank Ms. Silke Wiegand for her time and passion in introducing me the *in vitro* bioassays and for the time invested in testing and evaluating all my samples.

A general big acknowledgement goes to the entire Anatomy and Cell Biology Institute at the University of Giessen, for generously offering their mice for our experiments.

I would like to acknowledge Dr. Hans De Bie, Dr. Gerd Rheinwald, Dr. Shahriar Jahanbakht, and Mr. Tim Salbert from Advanced Chemistry Development, Inc. (ACD/Labs) for their great collaboration in the design and development of the LC-ESIMS chromatograms automated processor and for sharing of outstanding and valuable knowledge in the field of analytical chemistry and data processing.

Big thanks go to Dr. Maciej Dobrzynski from the Institute of Cell Biology, University of Bern (Switzerland) for his important contribution in the design and development of the online platform for statistical analysis and clustering of data originating from LC-ESIMS.

I would also like to thank the co-authors for their help, discussions, contributions, and beneficial advices during the writing phase of the articles, which were part of my project and presented in this dissertation.

I would like to thank Dr. Samad Ebrahimi, our former lab member and great friend, for his dedication in introducing me the instruments when I started my PhD and for his continuous sharing of experience and ideas.

I want to acknowledge my former Master student Angela Di Benedetto for her great commitment in conducting her work and to always being on duty.

I wish to thank Orlando Fertig for his valuable technical support and his amazing patience and engagement in resolving daily technical issues.

Furthermore, I would like to thank all my former colleagues for the great working atmosphere and memorable time passed together in and outside of the lab.

Finally, on a more personal level, I would like to thank my wife, my parents and my sister, and friends for their moral support during my entire doctoral studies, being characterized by some harder moments but mostly remembered as unforgettable four years of my life. Hvala! Obrigado! Grazie! Danke! Thank you!

SUMMARY

Pollen project

Asthma accounts for one of the many forms of chronic pulmonary diseases. Rather than a single disease, asthma is a term that can include a multitude of clinical symptoms related to gas exchange limitation, airway hyper-responsiveness resulting in wheezing and coughing, and airway inflammation. This variety of symptoms depicts very well the complexity of asthma and suggests the involvement of a multitude of triggers, biological pathways, and molecular mechanisms, which can be related to an allergic or non-allergic asthma outcome.

On one side, the allergic asthma is characterized by the inflammation of the airway tissue in response to a load of allergic triggers. The inflammatory process is induced by the over-reaction of the immune system, and causes airways swelling and mucus over-production. Besides inflammation, pro-inflammatory cytokines like interleukins (IL-1 β), tumor necrosis factor (TNF), and interferon-gamma (IFN- γ) are involved in the smooth muscle contraction. On the other side, the non-allergic asthma is mediated by non-inflammatory processes. Non-allergic triggers are organic and inorganic airway irritants, which modulate neuro-muscular macromolecular targets (receptors, channels) generating airways hyper-responsiveness and smooth muscle contraction.

The Transient Receptor Potential cation channel A1 (TRPA1) is a ubiquitous Ca²⁺-selective protein highly expressed on somatosensory neurons and is activated by environmental irritants inducing pain, cold and itchiness. Its expression on afferent neurons in the airways makes this channel a target for small chemical irritants, resulting in airway hyper-responsiveness to contractile stimuli. Differently, the type I Phosphatidylinositol 4-Phosphate 5-Kinase (PIP5K) gamma isoform is a lipid kinase involved in the regulation of intracellular Ca²⁺ in smooth muscle cells. PIP5K γ produces phosphatidylinositol 4,5-bisphosphate (PIP2), a substrate necessary for phospholipase C β (PLC β)-mediated inositol triphosphate (IP3) production. Finally, IP3 induces the release of Ca²⁺ from the sarcoplasmic reticulum with consequent myosin phosphorylation and smooth muscle contraction. The polyamines spermidine and spermine are crucial cofactors for PIP5K γ activation.

The aim of this work was to investigate the potential role of plant pollen secondary metabolites in the non-allergic exacerbation of asthma symptoms. We performed the phytochemical study of pollen from 30 plant species, including the highly allergenic *Ambrosia artemisiifolia* and

Ambrosia psilostachya from the family Asteraceae, and the highly allergic *Betula pendula* and *Corylus avellana* from the non-Asteraceae.

The spectrophotometric quantification of electrophiles in extracts revealed a surprisingly low concentration in *Ambrosia artemisiifolia*, a high concentration in *Ambrosia psilostachya*, and an electrophiles concentration in non-Asteraceae that was comparable to some species from *Ambrosia*, *Artemisia*, and other genera from the Asteraceae family. Furthermore, we observed the presence of polyamines in their conjugated form, which were exclusively present in the Asteraceae pollen. Finally, extracts and pure compounds were tested in murine afferent neurons for their effect on Ca^{2+} influx measured via calcium imaging, and for constriction of murine tracheal rings in organ bath experiments.

A. psilostachya THF extract and the STLs mixture coronopilin (**13**) and parthenin (**14**) induced similar Ca^{2+} influx results inducing Ca^{2+} influx in 32% ($n = 207$) and 37% ($n = 372$) ($p < 0.05$) of cinnamaldehyde positive (CA+) neurons, respectively. *A. artemisiifolia* induced Ca^{2+} influx in only 15% ($n = 144$) of CA+ neurons (statistically not significant). Tracheal tone in naïve and metacholine (MCh) pre-constricted tracheal rings was recorded in response to vehicle (0.5% DMSO), extract, and compounds. In naïve tracheal rings *A. psilostachya* extract and conjugated polyamines had no significant effect on tracheal tone. In contrast, **13/14** induced significant tracheal relaxation ($p = 0.017$). After washing, the tracheal rings were treated with MCh, a muscarinic receptor agonist known to induce a strong tracheal constriction. Interestingly, the magnitude of constriction in the **13/14** pre-treated tracheal rings was significantly lower compared to control ($p < 0.05$). MCh-constricted tracheal rings were treated with either *A. psilostachya* extract, **13/14**, or conjugated polyamines. Conjugated polyamines had no effect on MCh-constricted tracheal rings, while *A. psilostachya* extract and **13/14** both caused significant relaxation ($p < 0.05$).

Ziziphus project

Glycosylated triterpenes are organic molecules belonging to the secondary metabolism of plants and other organisms such as marine invertebrates. The genus *Ziziphus* (Rhamnaceae) comprises a variety of spiny shrubs and small trees, which leaves are commonly used in the oriental and Middle-East folk medicine for the treatment of various diseases but also to produce natural detergents and shampoos. Phytochemical characterization of glycosylated triterpenes in the hydroalcoholic extract from *Ziziphus spina-christi* leaves resulted in the identification of ten

saponins, of which three were found to be novel dammarane-type aglycones. Their structures were elucidated by a combination of powerful analytical methods such as HPLC-PDA-TOF-MS and highly sensitive microprobe NMR.

ACD/Labs project

The technological development of LC-MS instrumentation has led to significant improvements in automation, performance, and sensitivity. Complex samples, such as plant extracts can now be analyzed routinely and in high-throughput mode. Software tools allow efficient deconvolution of LC-MS chromatograms to obtain comprehensive information on single constituents. However, the systematic and unbiased comparison of large numbers of complex LC-MS chromatograms remains a challenge. Existing software tools for comparative processing are widely used in scientific fields like systems biology, including disciplines like genomics, proteomic, and metabolomics. Nevertheless, still little application is observed in phytochemistry and drug discovery research. The reasons for this are diverse and include insufficient user friendliness and the method-oriented (MS or NMR) configuration of the large majority of the available metabolomics platforms. In our opinion, a valid alternative to the existing computational tools could be found in the implementation of automated data processing and statistical analysis in software packages designed for modern drug discovery. ACD/Labs (Advanced Chemistry Development, Inc., Toronto, Canada) is a commercial analytical platform integrated with a wide range of analytical and predictive functionalities such as LC-MS and NMR data processing and analysis (ACD/Spectrus[®]), metabolite identification (ACD/MetaSense[®]) and *in silico* drug likeliness predictions (ACD/Percepta[®]).

In this work we focused on creating a two-step protocol comprising a parallel comparative processing integrated in ACD/Labs, and a web platform developed in R language designed for clustering and visualization of chromatographic data. Initially, all relevant chromatographic and spectroscopic data (retention time, molecular ions with the respective ion abundance, and sample names) are automatically extracted and assembled in an Excel spreadsheet. Afterwards, the file is loaded into an online web application equipped with various statistical algorithms where the user can compare and visualize the results in intuitive 2D heat maps.

ZUSAMMENFASSUNG

Asthma Projekt

Das Asthma gehört zu einer Form der chronischen Lungenkrankheiten. Lieber als eine Krankheit, Asthma wird als ein Begriff bezeichnet welcher eine Vielzahl von klinischen Symptomen in Bezug zu Gasaustausch Beschränkungen, Atemwege Überempfindlichkeit, und Atemwege Entzündung einbezieht. Diese Vielseitigkeit von Symptomen wiedergibt sehr wohl die Asthma Komplexität und hinweist auf die Beteiligung von mehreren Auslösern, biologischen Signalwege, und molekulare Mechanismen, welche mit einem allergischen oder nicht-allergischen Asthma zusammenhängen.

Zu einer Seite, das allergische Asthma ist charakterisiert bei einer Atemwegeentzündung als Rückantwort zur Last von allergischen Auslösern. Der inflammatorische Prozess ist verursacht durch eine Überreaktion des Immunsystem, und führt zu Atemwege Anschwellung begleitet durch Schleim Überproduktion. Ausser Entzündung, die pro-inflammatorische Cytokinen wie Interleukine (IL-1 β), Tumornekrosefaktor (TNF), und Interferon-gamma (IFN- γ) sind auch beteiligt in der Konstriktion der glatten Muskulatur. Auf der anderen Seite, das nicht-allergische Asthma wird durch nicht-inflammatorischen Prozessen vermittelt. Zu den nicht-allergischen Auslösern gehören organische und anorganische Reizstoffe der Atemwege, welche die neuromuskulären makromolekularen Targets (Rezeptoren, Kanälen) modulieren und somit die Atemwege Überempfindlichkeit und Konstriktion verursachen.

Der TRPA1 (auf Englisch *Transient Receptor Potential cation channel A1*) ist ein ubiquitäres Ca²⁺-selektives Protein das auf somatosensorischen Neuronen überexprimiert ist, welches durch umweltbedingte Reizstoffe aktiviert wird. Ihre Expression auf den somatosensorischen Neuronen in den Atemwegen macht den TRPA1 Kanal ein beliebtes Target für chemische Reizstoffe, was zu Kontraktion-Überempfindlichkeit der Atemwege führt. Die PIP5K-gamma Isoform Lipid Kinase (auf Englisch *type I Phosphatidylinositol 4-Phosphate 5-Kinase (PIP5K) gamma isoform*) ist verbunden mit der Regulation der intrazelluläre Ca²⁺ Konzentration in die Zellen der glatten Muskulatur. PIP5K γ erzeugt phosphatidylinositol 4,5-biphosphat (PIP2), ein Substrat für die Phospholipase C (PLC)-vermittelte Produktion von Inositoltriphosphat (IP3). IP3 bewirkt über die Bindung an seinen spezifischen Rezeptor, den IP3-Rezeptor, die Freisetzung von Calcium-Ionen aus dem sarkoplasmatischen Retikulum. Dies hat zur Folge, dass Myosin durch die Miosyn-leichte-Ketten-Kinase phosphoryliert wird und zur Kontraktion der glatten Muskulatur

führt. Die Polyamine Spermidin und Spermin sind entscheidende Kofaktoren für die Aktivierung der PIP5K γ .

Der Zweck dieser Arbeit war die Ermittlung der potentiellen Rolle von Pollen Sekundärmetaboliten in der Exacerbation von nicht-allergischen Asthma Symptome. Wir führten die phytochemische Studie von Pollen aus 30 Pflanzenspezies vor, inklusive die hoch-allergische *Ambrosia artemisiifolia* und *Ambrosia psilostachya* aus der Familie Asteraceae, als auch die hoch-allergische *Betula pendula* und *Corylus avellana* aus den nich-Asteraceae. Die spektrophotometrische Quantifizierung von elektrophilen Stoffen in Extrakten zeigte eine überraschend tiefe Konzentration in *Ambrosia artemisiifolia*, eine hohe Konzentration in *Ambrosia psilostachya*, und eine Konzentration in nicht-Asteraceae die vergleichbar mit einigen Spezies aus *Ambrosia*, *Artemisia*, und anderen Genera aus der Familie Asteraceae war. Weiterhin, könnten wir die Anwesenheit von Polyaminen in deren konjugierte Form beobachten, und ihre Verteilung nur in Asteraceae Pollen beschliessen. Schlussendlich, wurden ausgewählte Extrakten und Reinstoffe in afferenten Mausneuronen für ihren Effekt auf der intrazelluläre Ca²⁺ Erhöhung mittels Calcium Imaging und auf Einschränkung von trachealen Mausringen in Organ Bad Experimenten getestet.

A. psilostachya THF Extrakt und die Mischung an STL Coronopilin (**13**) und Parthenin (**14**) zeigten ähnliche Ergebnissen und induzierten Ca²⁺ Influx in beziehungsweise 32% ($n = 207$) und 37% ($n = 372$) ($p < 0.05$) von Cinnamaldehyd-positive (CA+) Neuronen. *A. artemisiifolia* induzierte Ca²⁺ Influx in nur 15% ($n = 144$) von CA+ Neuronen (statistisch insignifikantes Ergebnis). Die Erregungsantwort des trachealen Tonus in naive und Metacholin (MCh) voreingeschränkte trachealen Ringen wurde nach Zugabe von Kontrolle (0.5% DMSO), Extrakt, und **13/14** gemessen. In den naiven trachealen Ringen zeigten *A. psilostachya* Extrakt und die konjugierten Polyamine kein signifikanter Effekt. Im Gegensatz dazu, **13/14** induzierte eine signifikante tracheale Relaxation ($p = 0.017$). Nach dem Abspülen, wurden die trachealen Ringen mit MCh (ein muskarinisches Rezeptor Agonist) eingeschränkt. Interessanterweise, die maximale MCh-vermittelte Einschränkung war deutlich tiefer in **13/14** vorbehandelten trachealen Ringen, im Vergleich zur Kontrolle ($p < 0.05$). MCh voreingeschränkte trachealen Ringen wurden mit entweder *A. psilostachya* Extrakt, konjugierten Polyaminen oder **13/14** behandelt. Die konjugierten Polyamine zeigten keinen Effekt, während *A. psilostachya* Extrakt und **13/14** bewirkten eine signifikante Relaxation ($p < 0.05$).

Ziziphus Projekt

Glykosylierte Triterpene sind organische Moleküle, die zum Sekundärstoffwechsel von Pflanzen sowie anderen Organismen wie marinen Invertebraten gehören. Zur *Ziziphus* Gattung (Rhamnaceae) gehören verschiedene Arten von dornigen Sträuchern und Büschen, deren Blätter in der Volksmedizin sowie zur Herstellung von natürlichen Reinigungsmitteln und Shampoo verwendet werden. Die phytochemische Charakterisierung von glykosylierten Triterpenen im hydroalkoholischen Extrakt aus *Ziziphus spina-christi* Blättern führte zur Identifizierung von zehn Saponinen, von denen drei ein neuartiges Dammaran-Typ Aglykon besitzen. Ihre Strukturaufklärung wurde mittels Kombination von leistungsstarken analytischen Verfahren wie HPLC-PDA-TOF-MS und hochsensibler Mikro NMR durchgeführt.

ACD Projekt

Die technologische Entwicklung von LC-MS Instrumenten führte zu signifikanten Verbesserungen in Automation, Leistung, und Empfindlichkeit. Komplexe Proben, wie Pflanzenextrakten können jetzt routinemässig und in Hochdurchsatz analysiert werden. Die Softwares erlauben effiziente Dekonvolution von hunderte Metaboliten in einem LC-MS Chromatogramm, und helfen somit die Informationen über den einzelnen Konstituenten zu enthalten. Jedoch, das systematische und unverfälschte Vergleich von grossen Zahlen von komplexen LC-MS Chromatogrammen bleibt eine Herausforderung. Die bestehenden Computerprogramme für komparatives Processing sind weitgehend in wissenschaftlichen Gebieten wie Systembiologie. Trotzdem, ein immer noch zu geringer Einsatz ist zu beobachten in Phytochemie und Drug Discovery. Die Gründe dafür sind unterschiedliche und könnten an eine ungenügende Benutzerfreundlichkeit oder eine methodenorientierte (MS oder NMR) Konfiguration der Mehrheit der verfügbaren metabolomischen Plattformen ansprechen werden. Unserer Meinung nach, eine zulässige Alternative zu den existierenden Computerprogrammen könnte man in der Implementierung von automatisierter Datenprozessierung und statistischen Analysen in Programmpaketen für moderne Drug Discovery suchen. ACD/Labs (Advanced Chemistry Development, Inc., Toronto, Canada) ist eine kommerzielle analytische Plattform die mit einer breiten Auswahl an Spektralanalysen und prädiktiven Funktionen, wie LC-MS Data-Prozessierung (ACD/Spectrum[®]), Metaboliten Identifizierung (ACD/MetaSense[®]), und *in-silico* Voraussage der pharmakologischen Eigenschaften von Stoffen (ACD/Percepta[®]) integriert ist. In

Kollaboration mit ACD/Labs haben wir ein automatisiertes LC-ESIMS Prozessierung Tool entwickelt, und somit die schon stark polyedrische analytische Ausstattung von ACD/Labs weiter verbessert. Ausserdem, haben wir eine externe, kostenfreie, on-line Clustering-Plattform in R entwickelt, um bei der Darstellung von großen und komplexen Datensätze zu helfen, aber auch, um die Cluster-Identifizierung und die Analyse der Metaboliten Verteilung zwischen den Proben zu erleichtern.

1. AIM OF THE WORK

Asthma project

Asthma exacerbations are acute events characterized by severe respiratory symptoms. Even though, diagnosis and treatment of such acute attacks are well managed, their causes are still poorly understood. Existing recognized causalities are mostly focused on inflammatory processes, though often it is still difficult to correlate inflammation with exacerbation events. The physiological regulation of airway constriction and relaxation is driven by the autonomous nervous system, somatosensory, and smooth muscular activities and some of the proteins involved are the TRPA1 channel and the lipid kinase PIP5K γ . TRPA1 is located on somatosensory neurons and responsible for pain, cold, and itching sensations. Its expression on afferent neurons in airways makes TRPA1 an extremely important target for small chemical irritants with or without electrophilic character, responsible for airway hyper reactivity to contractile stimuli. On the other hand, PIP5K γ is present in the smooth muscle cells where it produces PIP2, a substrate for PLC β -mediated IP3 production, which induces the release of Ca²⁺ from the sarcoplasmic reticulum, and consequently the smooth muscle contraction. Polyamines spermidine and spermine are crucial cofactors for the PIP5K γ activation, regulated by the co-localization of the polyamine-catabolizing enzyme spermidine/spermine N1-acetyltransferase. Our hypothesis is that plant pollen secondary metabolites could be involved in the non-allergic asthma exacerbation process by triggering specific targets in airways, and leading to airways constriction. To explore the biological effect of pollen secondary metabolites on airways we conducted a comprehensive phytochemical study on pollen originating from thirty different plant species, and tested selected pollen extracts and pure compounds on murine dorsal root ganglia neurons and on the murine tracheal preparations.

***Ziziphus* Project**

Saponins are naturally occurring chemical compounds originating from various plant species. The chemical peculiarity of this class of plant metabolites resides in the molecular structure characterized by an aliphatic (lipophilic) and a polar (hydrophilic) moiety. The resulting amphiphilic character of saponins makes them valuable, naturally occurring surfactants and this physico-chemical property is employed in pharmaceutical technology, for example to enhance the cell membrane penetration of macromolecules, or in cosmetics due to their deterative properties. *Ziziphus spina-christi* (Rhamnaceae) is an evergreen tree, which leaves extract is used

in North Africa, the Middle East, and Asia for the preparation of body-care preparations like shampoos. Considering the surprisingly little scientific literature about the saponin content in the leaves of *Z. spina-christi*, and its socio-cultural importance in Middle East regions, we performed an in-depth phytochemical profiling of saponins by means of HPLC coupled to PDA-ESIMS-ELSD and microprobe NMR on the hydro-alcoholic extract from *Z. spina-christi* leaves.

ACD/Labs Project

Our lab is focused on the discovery of new leads from natural products. This early stage of drug discovery comprises extraction and fractionation of crude extracts, and purification of active metabolites, which then undergo biological and pharmacological studies for bioactivity characterization and evaluation of drug-like properties. This implies the acquisition of analytical (e.g. LC-PDA-MS, NMR) and biological (e.g. bioactivity, ADME-Tox) data, for rationalization of lead structure with its bioactivity. This multi-step interdisciplinary process is furthermore streamlined through consolidated use of computational platforms, like ACD/Labs, integrated with a wide range of analytical and predictive functionalities for LC-MS and NMR data, metabolite identification, and *in silico* drug likeliness prediction tools. ACD/Spectrus Processor is a multi-technique data processing and chemical characterization tool, which besides other functionalities allows efficient deconvolution of LC-MS chromatograms to obtain comprehensive information on single constituents. Nevertheless, in the current configuration the user is still faced with the time-consuming processing and comparison of multiple chromatograms, and representation of results in complex tables with a difficult read-out.

In this project, we developed a two-step protocol consisting of an automated LC-ESIMS batch processing tool integrated in ACD/MS Workbook Suite, and a web platform designed for clustering and visualization of chromatographic results. The two-step protocol was meant for enabling user to perform automated processing and comparison of multiple LC-MS chromatograms in a systematic, comprehensive, and unbiased way. This solution is meant to represent a valid alternative to existing metabolic platforms like XCMS and MZmine.

2. INTRODUCTION

2.1 Origin and uses of natural products

Natural products are very diverse in chemical structure and biological origin. Their biological source is wide and belongs to organisms living in marine, freshwater, and terrestrial ecosystems. Besides plants, it includes other eukaryotic organisms such as fungi and animals, as well as prokaryotic bacteria and archaea. The nature is an important source of low molecular weight organic molecules originating from metabolic processes involved in the growth, adaptation, and defense of the living organisms. Their synthesis is orchestrated by a plethora of metabolic enzymes, regulators of complex biosynthetic pathways, and is able to produce compounds with great structural diversity and complexity [1].

Since ancient times, plants represent the most important source of natural products to humans in form of nutrition but are also an integral part of human culture and health. During the millennia of the evolutionary process and still now-a-days, humans have been using natural products from eukaryotic fungi, animal, and plant sources but also from prokaryotic bacteria. The 19th Century was characterized by great historical changes worldwide at the political, economic, and scientific levels. This was also the time when we started to discovering and understanding the rationale behind the natural products chemistry and biological activity. The increasing knowledge of chemistry and related sciences, led to the duplication, modification, inspiration, and *de novo* synthesis of organic molecules, which were then used and studied for therapeutic uses. Examples are penicillin, tetracycline, or streptomycin antibiotics [2], statins like mevastatin from *Penicillium citrinum*, and immunosuppressant drugs from fungi [3], anti-hypertensive and anti-platelet drugs from animal tissues or venoms [4], and many others. On the other hand, the nutritive and health beneficial properties of traditional edible plants are catching an increasing interest gained by the scientific community, and are retrieving the people's perception regarding their benefits, especially when originating from wild sources [5]. From ancient times until today, besides providing pharmacological and nutritive benefits to humans, plants have played an important role in cosmetic preparations varying from skin care to perfumery. Perfumes originating from essential oils, skin and anti-aging creams obtained from vegetal oils (e.g. olive oil), and shampoos prepared with naturally occurring surfactants are only some examples. Nevertheless, the increasing demand for safety and quality of cosmetic products from natural sources is pushing the research and development to be constantly up to date with scientific knowledge and technologies. Cosmetic products based on natural sources are mostly prepared

with extracts, that might be obtained from different biologic origins. The application of extracts can result beneficial for a synergistic effect of contained natural products on biologic systems but, at the same time, the quality and safety of the preparation cannot be addressed to a specific active ingredient. For this reason, appropriate analytical tools are needed [6] in order to determine the quality and safety of applied extracts by discriminating the beneficial from deleterious metabolites.

References

1. Hong, J.; *et al.* Natural product diversity and its role in chemical biology and drug discovery. *Curr. Opin. Chem. Bio.* **2011**, 15, 350-354.
2. Clardy, J.; *et al.* The natural history of antibiotics. *Curr. Biol.* **2009**, 19, 437-441.
3. Butler, M. S. The role of natural product chemistry in drug discovery. *J. Nat. Prod.* **2004**, 67, 2141-2153.
4. Koh, D. C. I.; *et al.* Snake venom components and their applications in biomedicine. *Cell. Mol. Life Sci.* **2006**, 63, 3030-3041.
5. Simopoulos, A. P. Omega-3 fatty acids and antioxidants in edible wild plants. *Biol. Res.* **2004**, 37, 263-277.
6. Liang, Y.-Z.; *et al.* Quality control of herbal medicines. *J. Chromatogr. B* **2004**, 812, 53-70.

2.2 Strategies in analysis and discovery of natural products

The main goal in the natural products research is the identification of molecular entities originating from an organism. The purpose of studying the natural products can be very diverse, and comprises the discovery of new leads; study of environmental effects on the organism metabolome; identification of beneficial or toxic agents in food products; characterization of cosmetic products based on natural sources; ect. Depending on the goal of the study, different approaches can be undertaken.

Progress from targeted to comprehensive profiling approach

Targeted profiling of crude extracts is a holistic approach consisting in a concept of separating and identifying specific classes of low molecular weight entities. In the past, this approach was based on the extraction, fractionation, and purification of molecules with the final structure elucidation performed by nuclear magnetic resonance (NMR) spectroscopy. The monitoring and partial identification of metabolites during the fractionation and purification process was performed by thin layer chromatography (TLC), where the detection was based on UV light and visible light after treatment with visualization reagents, respectively [1-3]. This classical approach was appropriate for identifying major phytochemicals belonging to a specific class of metabolites. The relatively high limit of detection and the high material loss rate between each purification step was demanding relatively large amounts of starting material and was of limited applicability for the identification of minor compounds. In fact, the relatively poor separation capabilities of gravitational or low pressure chromatography columns, the low sample recovery, as well as low detection sensitivity were limiting factors to perform an exhaustive phytochemical study. Now-a-days, the cost-extensive and time-consuming targeted approach got relieved in the process by modern and powerful analytical technologies. These include chromatographic techniques like high performance liquid chromatography (HPLC) or ultra-high performance liquid chromatography (UHPLC), which are coupled to photodiode array (PDA), evaporative light scattering (ELSD), mass spectrometry (MS), high resolution-mass spectrometry (HR-MS), and tandem mass spectrometry (MSⁿ) detectors. HPLC/UPLC-PDA-ESI-ELSD is used as state of the art methodology for metabolic studies in extracts originating from edible [4] or medicinal [5] plants. The application of sensitive and selective detectors like mass spectrometers brings up the detection of a broad variety of analytes present in traces. Besides that, the MS detection enables

the discrimination of co-eluting compounds via deconvolution, a computational process, which creates pure spectra for each component. This enables scientist to detect and differentiate hundreds of peaks in one chromatogram. The high separation performance and the chromatographic reproducibility of liquid chromatographers coupled with different detectors significantly accelerated the compound identification process. This led to a new approach in the natural products discovery, named as comprehensive metabolic profiling. It enabled the scientists the detection and identification of molecular entities characterized by a broad structural variety, and the differentiation of metabolites with subtle structural variations. The comprehensive metabolite profiling with the aid of computational support went one step further and led to the creation of a new concept in the natural product science: the metabolome fingerprinting. This approach rather than being used for the identification of metabolites was implemented for the comparison of metabolic profiles within or among biologic systems, and with the final goal of identifying metabolome patterns. Few more important aspects regarding the metabolic fingerprinting approach are described in chapter 2.3.

HPLC-based activity profiling

Alternatively to the comprehensive phytochemical study, the HPLC-based activity profiling approach is a strategy used in the identification and structural characterization of specific peaks in plant extracts or fractions. This is a miniaturized approach to accelerate dereplication and thus the discovery of new bioactive compounds in crude extracts. HPLC hyphenated to bioassays results in the most widespread method for separation and tracking of natural products bioactivity. However, HPLC separation conditions or the bioassay complexity are limiting factors for such fully on-line strategies. This can result in a non-applicable method or in a slowed-down process caused by the extensive bioassay measurements or bioassays non-applicable at microscales (e.g. calcium imaging). Recently, alternative MS-based methods based on detection of ligands binding to a specific immobilized target, NMR-based methods revealing ligand interaction with soluble proteins, and investigations on a microchip format such as *in vitro* or *in vivo* gene expression profiling on microarrays have been developed in order to track activity [6].

Mass spectrometry (MS)

For an accurate elucidation of molecular structures the molecular weight, molecular formula, and structure fragmentation patterns are some of the most important information that can be obtained

from the mass spectrometry (MS). The basic principle of the MS method is to provide analytes with a charge, make them move through an electrostatic field, and finally measure their molecular weight. Nevertheless, depending on the instrument used, this general working principle can strongly vary at the technical level of the instrument. Different ionization techniques and mass analyzers can provide more or less accurate mass values of the analyte. Some examples of ionization techniques are: electron ionization (EI) - the analytes ionization can be obtained through an electron beam, which causes the loss of an electron to the analyte producing a cation ($M + e^- = M^{+\bullet} + 2e^-$); chemical ionization (CI) - the ionization is obtained by adding a charged element obtained from a reagent gas, in this case methane, into the analyte instead of knocking out an electron from it ($M + CH_5^+ = [M + H]^+ + CH_4$) [7]; electrospray ionization (ESI) - the ionization takes place in a nebulizing voltage capillary, which produces a fine spray of (multiple) charged solvent droplets containing the analytes [8] (Figure 1A). The charged droplets are driven by an electrostatic field towards the cone where the high temperature and an inert nebulizing gas continuously reduce the droplets size by solvent evaporation (desolvation). The final outcome is the transfer of the charge from the droplet to the analyte, which results in creation of ions in a gaseous phase with consequent acceleration towards the mass analyzer [8, 9]. The solvent and the neutral molecules are removed from the system by vacuum. Differently from the ESI ionization, the atmospheric pressure chemical ionization (APCI) doesn't produce charged droplets (Figure 1B). The nebulization originates from a non-voltage heating capillary at atmospheric pressure. The heat induces desolvation of droplets containing neutral analytes until transformation of solvent molecules and analytes in the gas-phase. At this point, uncharged analyte and solvent molecules enter the region of corona discharge. The much more abundant solvent molecules get ionized and some of them transfer the charge to the analyte.

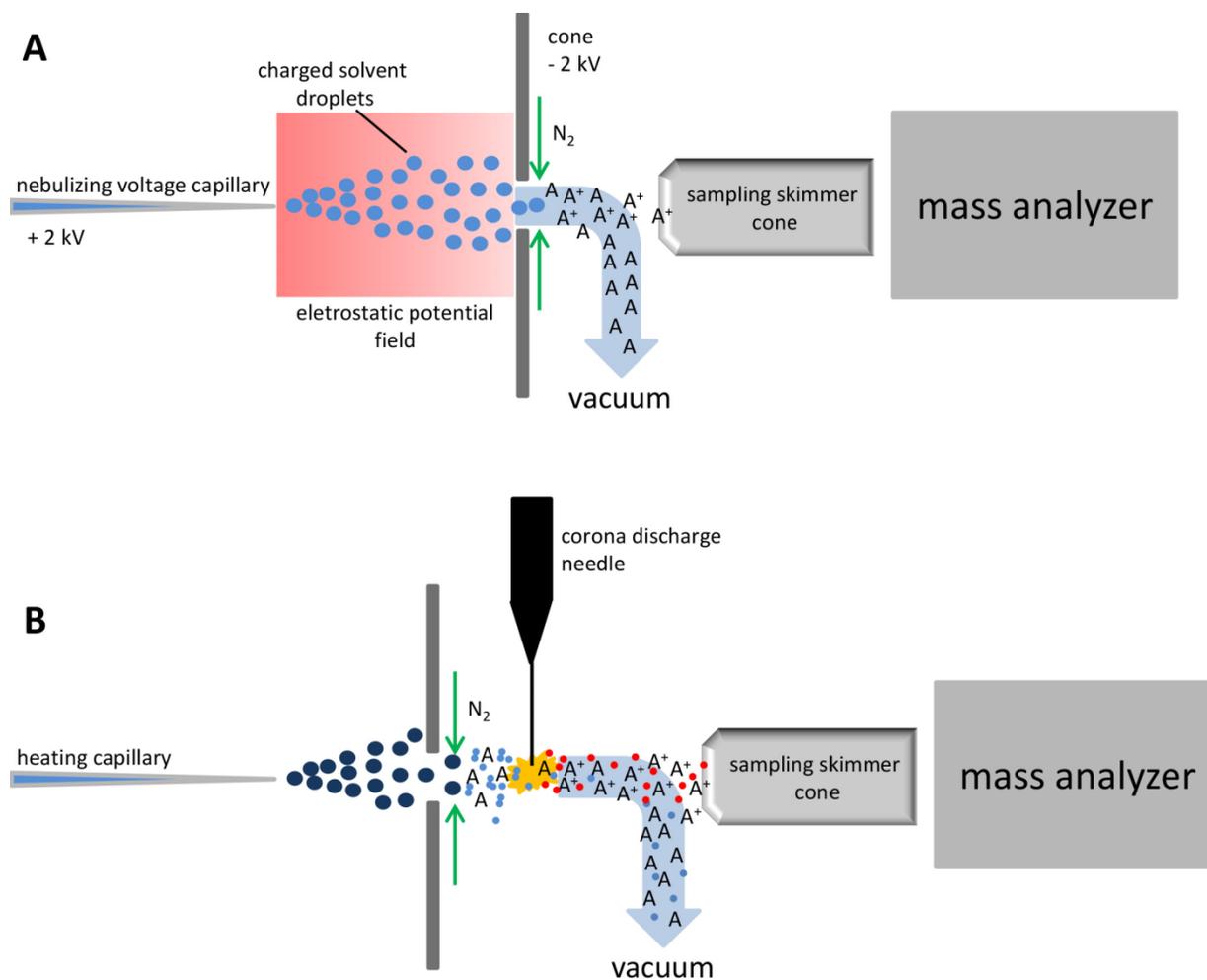


Figure 1: Graphical representation of ESI (A) and APCI (B) ionization sources. The red dots represent the charged solvent molecules, which after transferring the charge to the analyte become neutral (light blue) (B).

Depending on the information needed about an analyte, different mass analyzers can be used. Triple quadrupole (TQ) systems are suitable for classical metabolites profiling (scan mode) as well as analyte quantification and structural studies (selected/multiple reaction monitoring – SRM/MRM). In the scan mode all ions are directed to the Q3 where the detector is located and the result is shown in a mass chromatogram (Figure 2A). In the SRM mode, the focus is given on detection and characterization of one specific analyte. After ionization, the ion of interest gets filtered from other ions in the first quadrupole unit (Q1). In the collision cell the selected ion gets partially or totally fragmented depending on the strength of the applied collision energy. The so obtained fragments can be separated in the Q3 based on their m/z , and finally detected and quantified (Figure 2B).

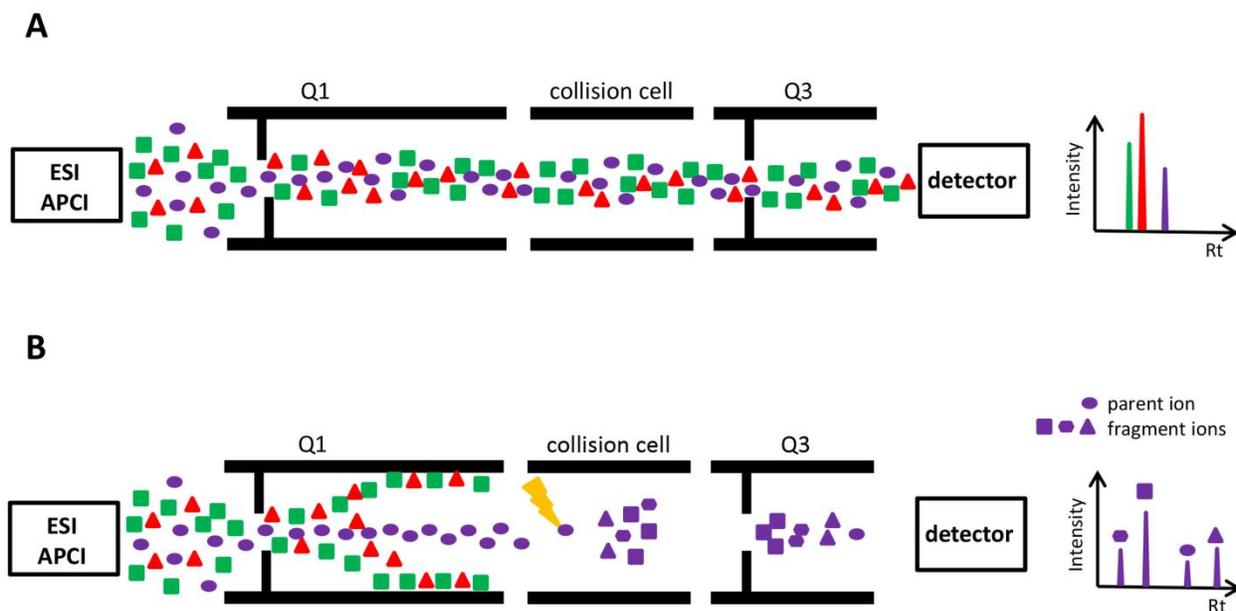


Figure 2: Schematic representation of a triple quadrupole (TQ). Full scan analysis allows the analyst to detect all ionized molecules carrying a m/z situated within the selected m/z range. This method is useful to detect a wide range of analytes with different masses. A possible limiting factor of the full scan detection is the lower detection sensitivity, which is indirectly proportional to the m/z scan range if assuming a constant event time (A). SRM analysis consists in the detection of a specific m/z , thus enabling a highly specific and sensitive analysis (B).

Differently from the triple quadrupole, ion trap systems are suitable for MS^n (tandem mass spectrometry) experiments, which are able to produce fragments of fragments and therefore provide information on the intramolecular atom connectivity [10]. Time-of-flight (TOF) and quantitative time-of-flight (Q-TOF) mass analyzers are instruments designed to deliver accuracy in milli mass units range providing accurate masses to calculate the molecular formula of a compound. UV spectroscopic data, the accurate mass, and the molecular formula of the analyte are fundamental information for a correct interpretation of nuclear magnetic resonance (NMR) spectra and for the correct structure assignment of organic compounds.

Nuclear magnetic resonance (NMR)

The NMR technique is based on the evaluation of the magnetic properties of atomic nuclei. Since organic molecules are principally composed of hydrogen and carbon atoms, only the proton NMR (1H NMR) and the carbon NMR (^{13}C NMR) are going to be discussed here. Both 1H NMR and the ^{13}C NMR work on same physical principles. Elemental isotopes 1H and ^{13}C are characterized by a spin (I) of 1/2. Assuming that the atomic nuclei are spherical, we can imagine them spinning around their axis and thus generate a magnetic dipole. By applying an external

magnetic field (B_0) two spin states are established. The $+1/2$ represents the magnetic moment of the lower energy state and is aligned with the external field, whereas the $-1/2$ (higher energy spin) is opposed to the external field. The energy difference (ΔE) between two spin states is proportional to the applied B_0 (Figure 3) [11]. Unfortunately, the nuclei are characterized by a very low ΔE between the upper (β) and the lower (α) spin state, due to an almost homogenous population of spin states (ΔE_1 , Figure 3A). To detect the net magnetization, namely the minuscule difference in population number between the upper and the lower spin states, it is necessary to apply stronger B_0 (Figure 4B). Thus, the application of higher external magnetic fields results in higher signal-to-noise ratios in the spectrum and consequently in higher sensitivity.

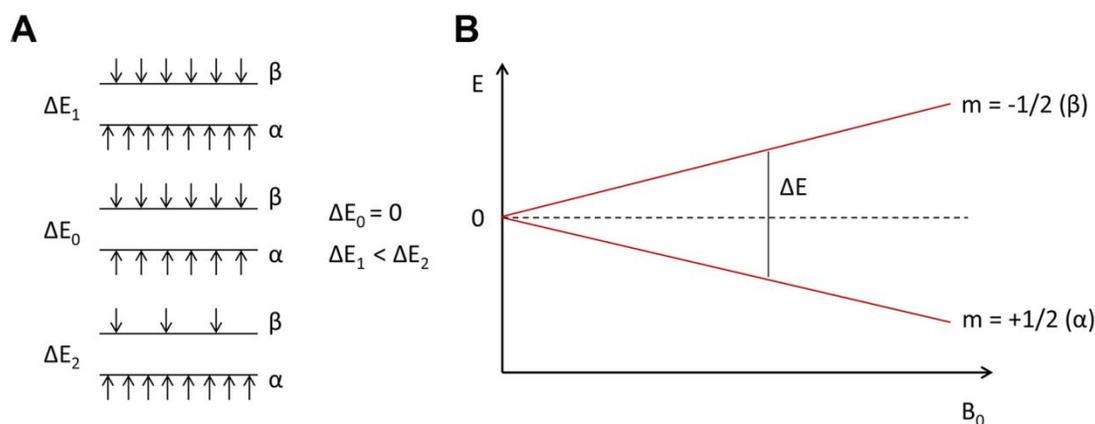


Figure 3: Distribution of spin state populations between upper (β) and lower (α) energy spin. ΔE_0 is observed when hypothetically the number of high energy spin states equals the low energy spin states. The increasing difference in population number between the upper and the lower spin states results in a higher energy difference ($\Delta E_1 < \Delta E_2$), assuming a constant external magnetic field (A). Energy difference between two neighboring spin states increases in a linear manner with an increasing B_0 (A) (adapted from [11]).

Since the discovery of one dimensional NMR spectroscopy by Bloch and Purcell in 1946 (Nobel Prize in Physics, 1952) [11], great technical advances have been made in the optimization and the increase of sensitivity of this technique. The application of higher magnetic fields together with the implementation of Pulse Fourier Transform (PFT) spectroscopy and the optimization of the probe performance, made it possible to increase the sensitivity of NMR instruments in order to detect naturally less occurring nuclei like ^{13}C and ^{15}N . These technical advances allowed a higher sample throughput and smaller quantities of samples needed.

For the structural elucidation of the molecule, a wide option of NMR experiments is available, depending on the nature of the analyte. The NMR experiments are classified into one-dimensional ^1H NMR and ^{13}C NMR experiments, two-dimensional mononuclear ^1H , ^1H -correlation spectroscopy (COSY) (Figure 4A), and heteronuclear ^1H , ^{13}C correlation experiments like heteronuclear multiple-bond correlation spectroscopy HMBC and heteronuclear single-quantum correlation spectroscopy (HSQC) (Figure 4B).

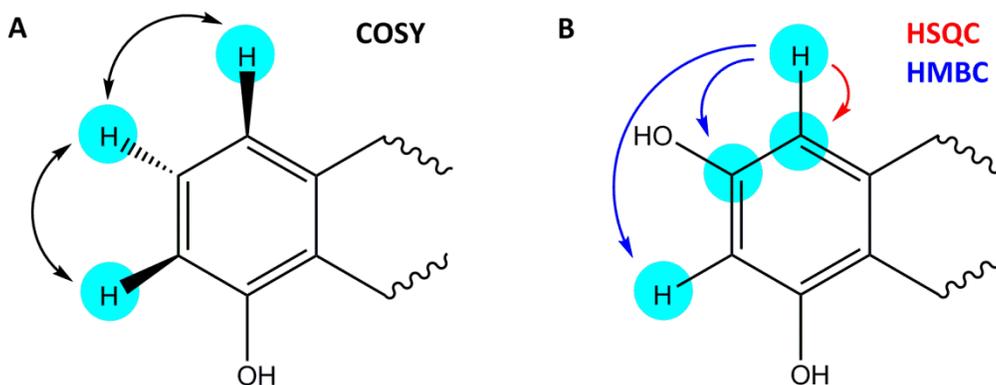


Figure 4: Examples of detected correlations for COSY (A), and HSQC/HMBC (B).

The correlation spectroscopy (COSY) is a two-dimensional experiment for detection of homonuclear correlations in coupled spins (Figure 4A). HMBC and HSQC are standard experiments in determining the covalent structure of the compound. HMBC detects heteronuclear correlations over longer ranges of about 2 to 4 bonds, whereas HSQC detects correlations between nuclei of two different types separated by one bond (Figure 4B). In case of more complex structures (e.g. glycosides, aliphatic chains, aliphatic rings), the presence of coupling systems results in overlapping signals. To accredit the right scalar coupling between nuclei, additional experiments like (^1H , ^1H)-total correlation spectroscopy TOCSY or (^1H , ^{13}C)-HSQC-TOCSY are necessary. At the end, the stereo chemical information of the analyte is provided by through-space correlation methods. The nuclear Overhauser effect (NOE) is a transfer of nuclear spin polarization through space between two atoms that are in a close spatial proximity, independently from the bond connectivity.

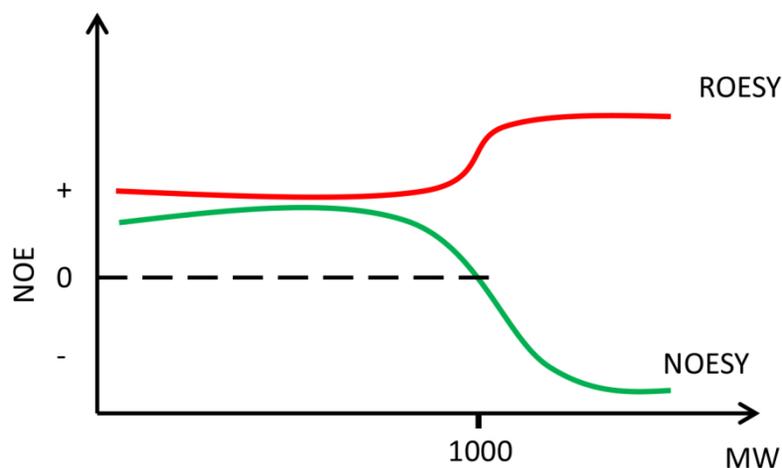


Figure 5: Diagram showing correlation between the molecular size and the relative increase/decrease of NOE. Figure adapted from [12].

The nuclear Overhauser effect spectroscopy (NOESY) provides a similar spectrum to COSY spectrum. However, rather than showing nuclei that are coupled through-bond, NOESY correlates nuclei that are spatially close to each other. For molecules with a molecular weight of around 1000 Da, the NOE is proximal to zero and therefore non detectable. In this case, measurement of rotating-frame nuclear Overhauser effect correlation spectroscopy (ROESY) experiment is a valid alternative (Figure 5).

References

1. Laskar, S.; *et al.* A modified spray reagent for the detection of amino acids on thin layer chromatography plates. *Amino Acids* **2001**, 21, 201-204.
2. Lemieux, R. U.; *et al.* Spray reagent for detection of carbohydrates. *Anal. Chem.* **1954**, 26, 920-921.
3. Ghosh, P.; *et al.* Spray reagent for the detection of coumarins and flavonoids on thin-layer plates. *J. Chromatogr.* **1987**, 403, 285-287.
4. Abbet, C.; *et al.* Comprehensive analysis of *Cirsium spinosissimum* Scop., a wild alpine food plant. *Food Chem.* **2014**, 160, 165-170.
5. Mohn, T.; *et al.* A comprehensive metabolite profiling of *Isatis tinctoria* leaf extracts. *Phytochemistry* **2009**, 70, 924-934.
6. Potterat, O.; M. Hamburger. Natural products in drug discovery - concepts and approaches for tracking bioactivity. *Curr. Org. Chem.* **2006**, 10, 899-920.
7. Munson, M. S. B.; Field, F. H. Chemical ionization mass spectrometry. I. General introduction. *JACS* **1966**, 88, 2621-2630.
8. Ho, C. S.; *et al.* Electrospray ionisation mass spectrometry: principles and clinical applications. *Clin. Biochem. Rev.* **2003**, 24, 3-12.
9. Pitt, J. J. Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clin. Biochem. Rev.* **2009**, 30, 19-34.
10. Glish, G. L.; Vachet, R. W. The basics of mass spectrometry in the twenty-first century. *Nat. Rev. Drug. Discov.* **2003**, 2, 140-150.
11. Friebolin, H. Ein- und zweidimensionale NMR-Spektroskopie: eine Einführung. **2013**, Wiley, 5th Edition.
12. Neuhaus, D. and Williamson, M. P. The nuclear overhauser effect in structural and conformational analysis. **2009**, Wiley, 2nd Edition.

2.3 Profiling, comprehensive data analysis, and visualization of mass spectrometry data

The term metabolomics describes a systematic study of exclusive metabolite fingerprints in biological systems, and is supported by analytical methods like gas and liquid chromatography coupled to mass spectrometry (GC/LC-MS), and/or nuclear magnetic resonance spectroscopy [1, 2]. The American biochemist Roger John Williams (1893 - 1988) formulated for the first time the term “metabolic patterns” in the book “*Individual metabolic patterns and human disease: an exploratory study utilizing predominantly paper chromatographic methods*” (1951) [4]. Williams and co-authors discussed the correlation of metabolic patterns in saliva and urine samples using paper chromatography, and represented the differences in metabolic composition and intensities between different samples using vector diagrams (Figure 6).

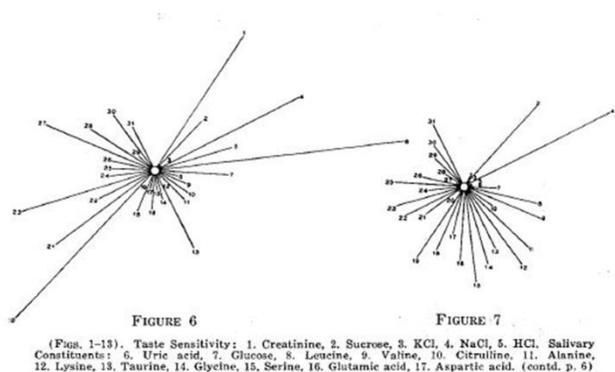


Figure 1: Correlations between salivary constituents and taste sensitivity to sodium chloride (adapted from [4]).

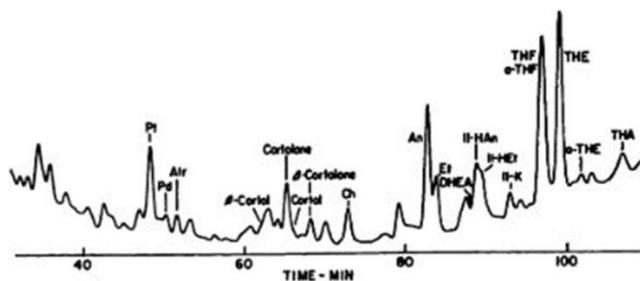


Figure 2: GC-MS profile of urinary samples performed by Horning and Horning (adapted from [5]).

The second half of the 20th century was a period where new technologies in the analytical chemistry were developed. The following commercialization of new analytical techniques such

as gas chromatography coupled to MS in the 1960's [6] enabled analysis of organic samples with higher selectivity, sensitivity, and peak resolution. The following development and modernization of GC-MS and other analytical techniques such as LC-MS and NMR expanded their use in the analysis of complex organic samples. The evolution of computational tools for analysis of chromatographic and spectroscopic data went in parallel with the improvement of analytical hardwares. One of the most important achievements in GC/LC-MS data processing was the creation in 1999 of AMDIS (Automated Mass Spectrometry Deconvolution and Identification System) algorithm, which enabled the automated identification of co-eluting components [7, 8], and a consequent discrimination of hundreds of peaks in a single chromatogram. The growing need for metabolic studies in complex biological systems demanded for further development of computer-assisted strategies in processing and comparison of multiple samples capable of accelerating the identification of metabolic differences between samples [9, 10]. This process finally evolved in modern free-access and commercial metabolomics platforms like XCMS, MetaboAnalyst, and MZmine integrated with multivariate statistics and designed for GC/LC-MS data processing, comparison, and visualization. The advanced computational tools gave the researchers the possibility to develop modern metabolomics studies in systems biology, like those pioneered by Prof. Jeremy Nicholson [11-13].

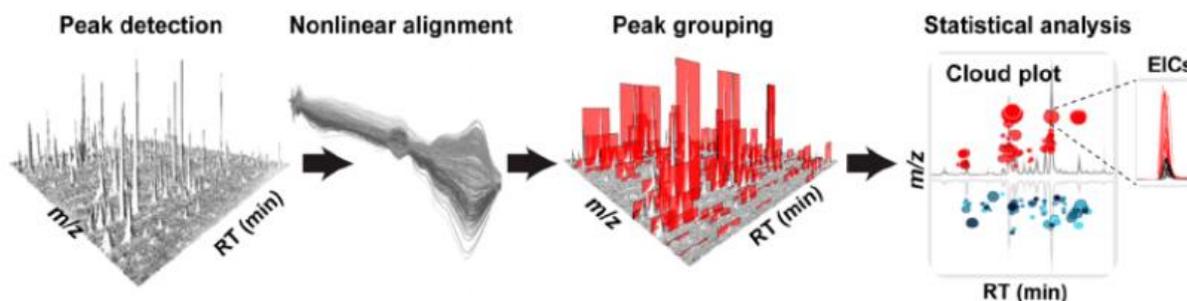


Figure 3. Example of computational functions in XCMS Metabolomics platform. It includes automatic peak detection, nonlinear alignment of chromatograms, peak grouping, and statistical analysis [14].

The most advanced achievements in the metabolomics have been published in 2015 by Sauer and Zamboni, and consist in a simultaneous and real-time analysis of hundreds of compounds in cells or organisms, skipping the chromatographic separation process [3].

References

1. Karpievitch, Y. V.; *et al.* Liquid chromatography mass spectrometry-based proteomics: biological and technological aspects. *Ann. Appl. Stat.* **2010**, *4*, 1797-1823.
2. Zhou, B.; *et al.* LC-MS-based metabolomics. *Mol. Biosyst.* **2012**, *8*, 470-481.
3. Link, H.; *et al.* Real-time metabolome profiling of the metabolic switch between starvation and growth. *Nat. Methods* **2015**, *12*, 1091-1097.
4. Individual metabolic patterns and human disease : an exploratory study utilizing predominantly paper chromatographic methods. Austin: The University of Texas Publication, **1951**. University of Texas Library. Web. 27 Apr. 2017. <https://repositories.lib.utexas.edu/handle/2152/7023>.
5. Horning, E. C.; Horning, M. G. Metabolic profiles: gas-phase methods for analysis of metabolites. *Clin. Chem.* **1971**, *17*, 802-809.
6. Hites, R. A. Development of gas chromatographic mass spectrometry. *Anal. Chem.* **2016**, *88*, 6955-6961.
7. Colby, B. N. Spectral deconvolution for overlapping GC/MS components. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 558-562.
8. Stein, S. E. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 770-781.
9. Jonsson, P.; *et al.* A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. *Anal. Chem.* **2004**, *76*, 1738-1745.
10. Wolfender, J. L.; *et al.* Plant metabolomics: from holistic data to relevant biomarkers. *Curr. Med. Chem.* **2013**, *20*, 1056-1090.
11. Nicholson, J. K.; Lindon, J.C. Systems biology: metabonomics. *Nature* **2008**, *455*, 1054-1056.
12. Holmes, E.; *et al.* Metabolic phenotyping in health and disease. *Cell* **2008**, *134*, 714-717.
13. Nicholson, J. K.; *et al.* Metabonomics: a platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discov.* **2002**, *1*, 153-161.
14. Johnson, C. H.; *et al.* Bioinformatics: the next frontier of metabolomics. *Anal. Chem.* **2014**, *87*.

2.4 Saponins as natural products with biological activity and deterative properties

Saponins (latin *sapo* meaning soap) are organic molecules deriving from plants and marine invertebrate secondary metabolism [1]. Their molecular structure consists of an apolar aglycon of triterpenic (C30), steroidal (C27), or alkaloid nature (e.g. solanine) [2, 3], substituted by monoglycosides (monodesmosidic) or glycosidic chains (Figure 1). The name saponin derives from the characteristic soap-foaming effect of these organic molecules when strongly shaken in water, a phenomenon related to their amphiphilic character.

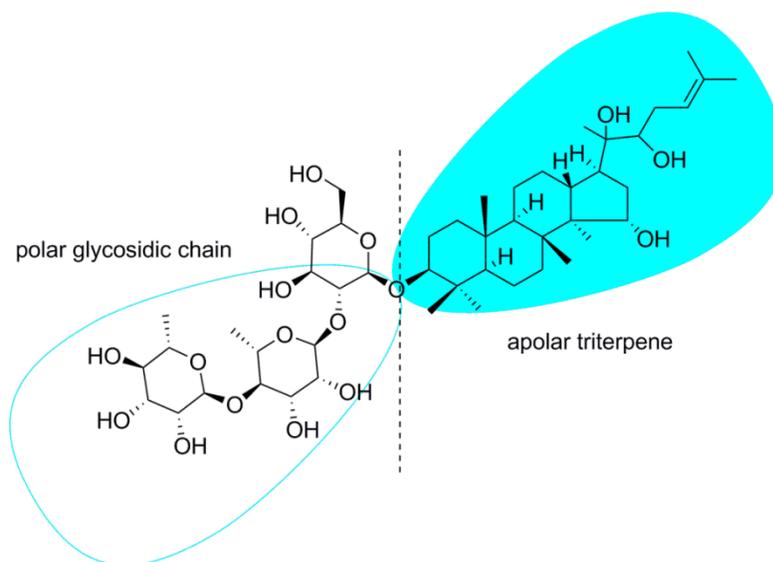


Figure 1. Saponin structure and its amphiphilic character. Compound adapted from [4].

Saponins are very abundant in plant families like Hippocastanaceae (*Aesculus hyppocastanum*), Araliaceae (*Hedera helix*, *Panax ginseng*), Asparagaceae (*Ruscus aculeatus*), Quillajaceae (*Quillaja saponaria*), and Caryophyllaceae (*Saponaria officinalis*) [5]. In the past, soap plants like *Saponaria officinalis* (soapwort), were used as a source of natural soaps or detergents for laundry and personal hygiene. The use of soap plants for washing purposes goes back to the ancient times, and an example worth of consideration is the treatment of the Holy Shroud of Tourin with extracts derived from the *Saponaria officinalis* (Figure 2). The reasons for its

extraordinary conservation are still debated, and the delicateness and antifungal properties of the characteristic triterpenoid saponins are considered as valid arguments [6].

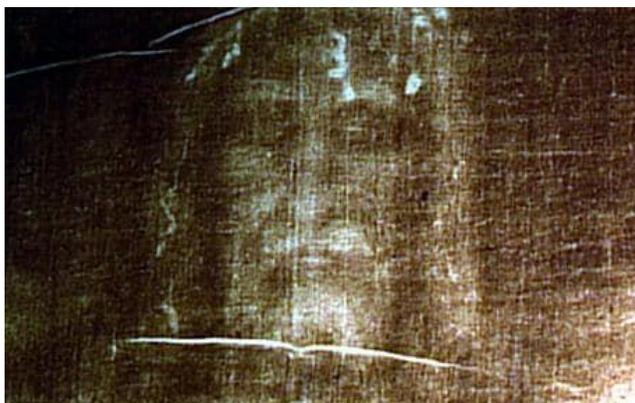


Figure 2. Shroud of Turin, a piece of cloth bearing the image of the Jesus of Nazareth [7].

Biosynthesis of saponins

The isoprene isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl pyrophosphate (DMAPP) are precursors for the biosynthesis of terpenoids, which constitute one of the largest families of natural products derived from plants [8]. Saponins are metabolites biogenetically originating from terpenoids, and their structure is characterized by an apolar triterpene derivative skeleton conjugated to a polar glycosidic moiety. The IPP/DMAPP biosynthesis follows two distinct pathways, namely the mevalonate pathway and the non-mevalonate pathway (Figure 3). In higher plants, both biogenetic pathways use the glycolysis-derived pyruvate as precursor for the IPP/DMAPP synthesis, although being active in two different cellular compartments. The mevalonate pathway operates in the cellular cytosol, whereas the non-mevalonate pathway takes part in the chloroplasts [9]. In the mevalonate pathway, the pyruvate is transformed into acetyl-CoA, which is subject to a series of condensation and reduction steps until forming the mevalonic acid. Consequent phosphorylation and decarboxylation steps transform the mevalonate (phosphorylated mevalonic acid) into IPP/DMAPP [10]. Differently, the non-mevalonate pathway, also called 1-deoxy-D-xylulose-5-phosphate (DOXP) pathway, starts with the condensation of pyruvate and glyceraldehyde 3-phosphate resulting in the DOXP-synthase (DXS)-mediated formation of DOXP. Additional enzymatically-orchestrated reactions lead to the formation of the isoprene IPP/DMAPP [11] (Figure 4), which follow a reductive condensation and cyclisation to finally form cyclic triterpenes and steroids. The so formed cyclic hydrocarbons

are then functionalized by Cytochrome P450-mediated oxygenation of the scaffold, which finally results in a target for type 1 glycosyltransferases responsible for their glycosylation [12].

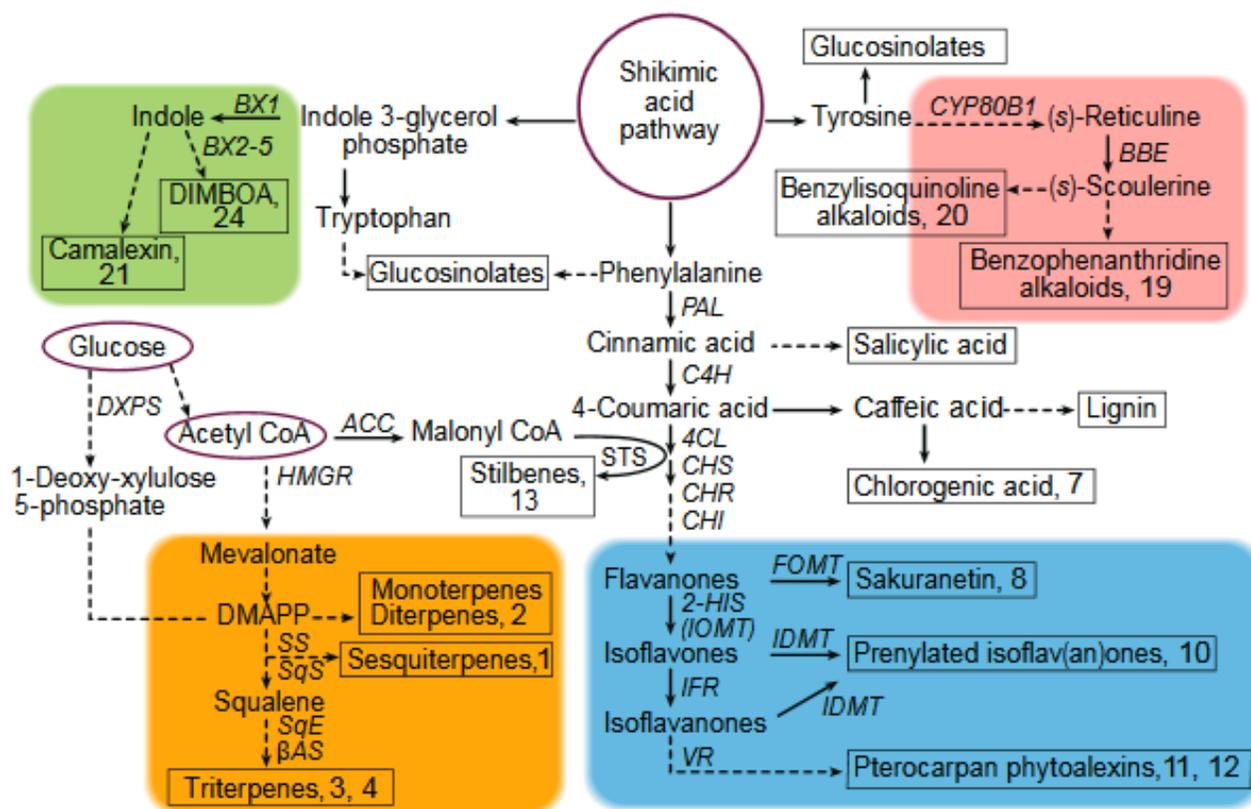


Figure 3. Biosynthesis of plant natural products with anti-microbial activity [13]. The mevalonate and the non-mevalonate (DOXP) biosynthetic pathways are described as biogenetic sources of isoprenoids IPP/DMAPP, building blocks for the synthesis of terpenes.

Besides being present in higher plants, the DOXP pathway is also operating in other organisms like eubacteria, algae, and cyanobacteria. On one hand, the absence of DOPX in mammals made the enzymes involved in this pathway interesting pharmacological targets for treatment of diseases such as malaria or cryptosporidiosis [14]. On the other hand, other organisms like cyanobacteria and *E. coli* have been characterized as highly performing hosts for the heterologous overexpression of specific enzymes [15], or for the engineered biosynthesis of natural products, like lycopene and terpenoids [16, 17].

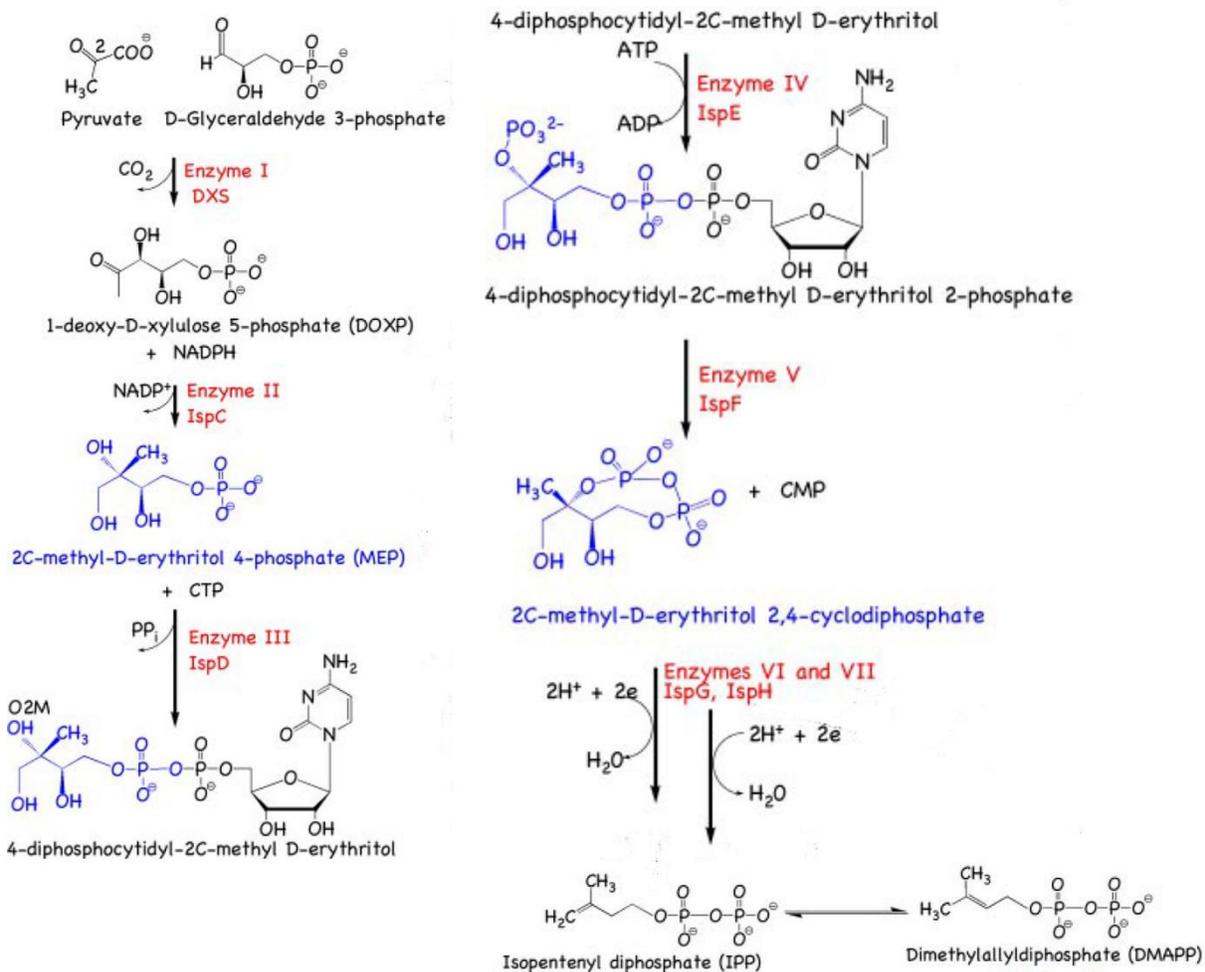


Figure 4. Non-mevalonate biosynthetic pathway of isoprene IPP and DMAPP, precursors in the terpene biosynthesis (adapted from [11]).

Bioactivity of saponins

Saponins are characteristic natural products in foods like beans, potatoes, quinoa, asparagus, and blackberries, and comprise a broad range of structurally diverse amphiphilic molecules. Besides their soap-like properties they were studied for their biological and pharmacological activities like anti-inflammatory [18] and venotonic properties [19]. Anti-microbial and anti-fungal properties were also reported, along with anti-tumor [20], and *in vitro* immunomodulation [21]. Another important feature of saponins is attributed to their effects on cell membranes where they can strongly affect the membrane integrity by different mechanisms depending on their chemical structure. This ability of saponins to modify membrane permeability can be used to facilitate the

passage of drug molecules or other natural products through the cell membrane [22], but can also result in undesired toxic hemolytic activity [23, 24].

Although saponins are known to be involved in vital morphological and biochemical processes in plants, still very little is known regarding the biological mechanisms controlled by these secondary metabolites [25]. On the other side, saponins have gained a lot of attention for their therapeutic potential, but also for valuable properties as cosmetic ingredients by increasing the function of microvasculature (e.g. *Aesculus hippocastanum*), and for their deterative properties [26]. This wide range of potential applications makes saponins highly interesting secondary metabolites, on which still a lot of research is needed for better understanding their biological contribution.

References

1. Ivanchina, N. V.; *et al.* Steroid glycosides from marine organisms. *Steroids* **2011**, 76, 425-454.
2. Moses, T.; *et al.* Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. *Crit. Rev. Biochem. Mol. Biol.* **2014**, 49, 439-462.
3. Funayama, S.; Cordell, G. A. Alkaloids – a treasury of poison and medicines. **2015**, Academic Press, 1-20.
4. Bozicevic, A.; *et al.*, Dammarane-type saponins from leaves of *Ziziphus spina-christi*. *Phytochemistry* **2017**, 138, 134-144.
5. Facino, R. M.; *et al.* Anti-elastase and anti-hyaluronidase activities of saponins and sapogenins from *Hedera helix*, *Aesculus hippocastanum*, and *Ruscus aculeatus*: factors contributing to their efficacy in the treatment of venous insufficiency. *Arch. Pharm.* **1995**, 328, 720-724.
6. Heller, J. H.; Adler, A. D. A chemical investigation of the Shroud of Turin. *Can. Soc. Forensic Sci. J.* **1981**, 14, 81-103.
7. The Telegraph. <http://www.telegraph.co.uk/news/religion/5137163/Turin-Shroud-could-be-genuine-as-carbon-dating-was-flawed.html>. (accessed 17 Feb, 2017).
8. Dubey, V. S.; *et al.* An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants. *J. Biosci.* **2003**, 28, 637.
9. Kuzuyama, T. Mevalonate and non-mevalonate pathways for the biosynthesis of isoprene units. *Biosci. Biotechnol. Biochem.* **2002**, 66, 1619-1627.
10. Mizioro, H. M. Enzymes of the mevalonate pathway of isoprenoid biosynthesis. *Arch. Biochem. Biophys.* **2011**, 505, 131-143.
11. Hunter, W. N. The non-mevalonate pathway of isoprenoid precursor biosynthesis. *J. Biol. Chem.* **2007**, 282, 21573-21577.
12. Thimmappa, R.; *et al.* Triterpene biosynthesis in plants. *Annu. Rev. Plant Biol.* **2014**, 65, 225-57.
13. Dixon, R. A. Natural products and plant disease resistance. *Nature* **2001**, 411, 843-847.
14. Hunter, W. N. Isoprenoid precursor biosynthesis offers potential targets for drug discovery against diseases caused by apicomplexan parasites. *Curr. Top. Med. Chem.* **2011**, 11, 2048-2059.
15. Pattanaik, B.; Lindberg, P. Terpenoids and their biosynthesis in cyanobacteria. *Life* **2015**, 5, 269-293.
16. Kim, S. W.; Keasling, J. D. Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in *Escherichia coli* enhances lycopene production. *Biotechnol. Bioeng.* **2001**, 72, 408-415.
17. Martin, V. J.; *et al.* Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* **2003**, 21, 796-802.
18. Just, M. J.; *et al.* Anti-inflammatory activity of unusual lupane saponins from *Bupleurum fruticosens*. *Planta Med.* **1998**, 64, 404-407.
19. Sirtori, C. R. Aescin: pharmacology, pharmacokinetics and therapeutic profile. *Pharmacol. Res.* **2001**, 44, 183-193.
20. Man, S.; *et al.* Chemical study and medical application of saponins as anti-cancer agents. *Fitoterapia* **2010**, 81, 703-714.
21. Bhardwaj, J.; *et al.* Immunomodulatory effect of tea saponin in immune T-cells and T-lymphoma cells via regulation of Th1, Th2 immune response and MAPK/ERK2 signaling pathway. *Immunopharmacol. Immunotoxicol.* **2014**, 36, 202-210.
22. Francis, G.; *et al.* The biological action of saponins in animal systems: a review. *Br. J. Nutr.* **2002**, 88, 587-605.
23. Baumann, E.; *et al.* Hemolysis of human erythrocytes with saponin affects the membrane structure. *Acta Histochem.* **2000**, 102, 21-35.

24. Oda, K.; *et al.* Adjuvant and haemolytic activities of 47 saponins derived from medicinal and food plants. *Biol. Chem.* **2000**, 381, 67-74.
25. Moses, T; *et al.* Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. *Crit. Rev. Biochem. Mol. Biol.* **2014**, 49, 439–462.
26. Chen, Y.-F.; *et al.* Foam properties and detergent abilities of the saponins from *Camellia oleifera*. *Int. J. Mol. Sci.* **2010**, 11, 4417–4425.

2.5 Asthma

Historical aspects of asthma

The term asthma is known since remote times in ancient Egyptians, Chinese, Japanese, and Greek civilities. Asthma as a medical term was already described in the classical period by the Greek physician and philosopher Hippocrates (460 – 370 BC) in the *Corpus Hyppocraticum*, an important collection of medical works. The first clinical descriptions of asthma were provided by the ancient Greek master clinician Aretaeus of Cappadocia [1, 2]. In the 12th century, *Treatise of Asthma* written by Maimonides (1135-1204 AD) were first documents describing environmental factors such as low temperature and air moisture, causing negative effects on asthmatics [3]. Further findings about asthma were published in the 17th century by the Belgium physician and pneumatic chemist Jean Baptiste Van Helmont (1579-1644 AD). In his posthumous work *Ortus Medicinae*, he described the pipes of the lungs as the anatomical origin of asthma [4]. These findings were followed by Bernardino Ramazzini (1633-1714 AD) in *De Morbis Artificum Diatriba*, a work focused on the occupational medicine, where he described the negative influence of organic dust on asthma [5]. All these findings were of great importance in order to understand the anatomical localization of the disease and its triggers. In parallel, other works were conducted on medical observations that were originally not linked to the asthma, such as the immune system and its functionality. For the future understanding of asthma pathophysiology, important contribution was given by the Viennese pediatrician Clemens von Pirquet Freiherr (1874–1929), who coined the term allergy as a protective and/or harmful immune system reaction [6] and by Prausnitz and Küstner in early 1920's by discovering the protein *reagin* [7], an antibody found in the blood of individuals that have a genetic predisposition to allergies. These were key findings that permitted Johansson and Ishizaka to discover the antibody IgE, and finally observe the causality of the serum IgE and asthma. Nowadays, asthma is recognized as a syndrome driven by “asthma susceptibility genes” expressed in the epithelium and innate immune pathways [8], and characterized by airway inflammation and airway hyper-responsiveness, where coughing, wheezing, and chest tightness are recognized as common symptoms [9].

Pathophysiology of asthma

The airway in humans is divided in upper and lower respiratory tract and is covered with a continuous sheet of epithelial cells. The upper respiratory tract is defined between the larynx and the nasal cavity, whereas the lower respiratory tract starts below the larynx and comprises trachea, bronchi, and bronchioles. The mucociliary apparatus consists of ciliated and secretory cells localized in the lower respiratory tract, and is part of the mechanical defense of the airways against inhaled pathogens and allergens. Beside the mucus, the airway epithelium itself represents a physical barrier protecting the tissue from the external environment. In this context, the airway epithelial tight junctions were revealed to play an integral role with the mucociliary apparatus in blocking the passage of allergens and thus mechanically blocking the passage of asthma-inducing factors [10]. Nevertheless, the function of the epithelial barrier can be compromised by the proteolytic character of many allergens [11] as well as chemicals present in the tobacco smoke [10], thus increasing the susceptibility of individuals to asthma.

Asthma is considered as a heterogeneous disease characterized by different symptoms [12], often correlated with the genetic predisposition of individuals and their exposure to environmental conditions as well as therapeutic adherence [9]. It results in airway morphological changes as a response to tissue injury caused by pathogens, allergens, but also inflammatory mediators generated by the immune response. These changes are known as “airway remodeling” and include a wide variety of changes at the tissue morphology and cellular function levels including (i) epithelial disruption caused by infiltration of inflammatory cells (lymphocytes, eosinophils, and mast cells) [13]; (ii) smooth muscle hypertrophy with consequent airway lumen reduction; (iii) tissue with excess interstitial fluid (edematous); (iv) blood vessel dilatation, and angiogenesis with endothelial swelling; (v) mucus gland hypertrophy with consequent mucus overproduction [14]. The past few decades have been characterized by a general decrease in the number of hospitalizations of patients suffering from asthma due to advanced sensitization and therapeutic programs [15]. Nevertheless, the hospitalization frequency due to an acute exacerbation of asthma remains still high [16]. The etiology of acute asthma exacerbations is attributed to infections of viral or bacterial origin, allergens, pollutants, and drugs. All this leads to typical symptoms such as airway hyper responsiveness and breathing difficulties [17]. Generally, asthma is characterized by an immunologic reaction to allergens of different nature.

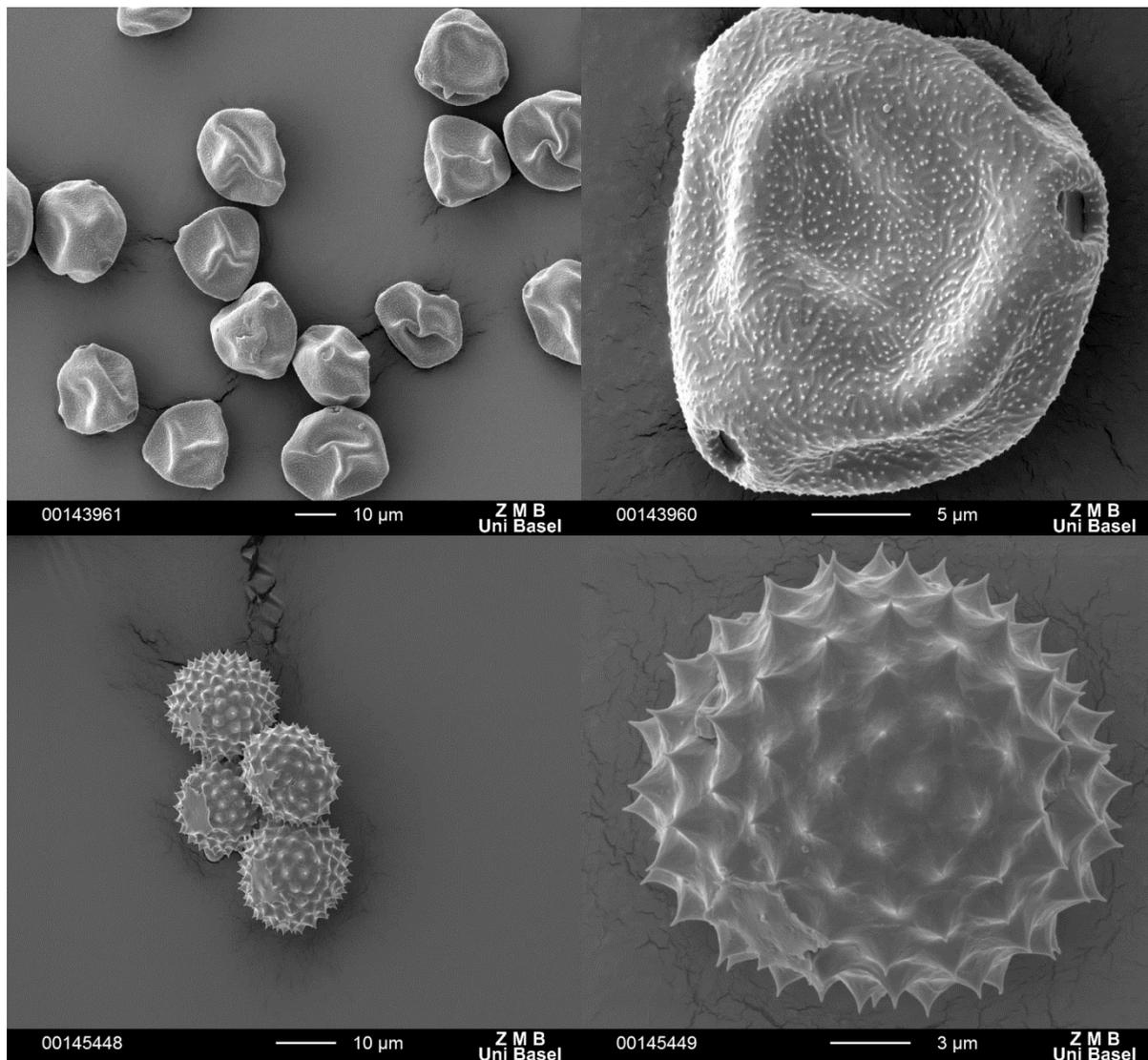


Figure 1. Images of pollen grains (*Corylus betulus*, up and *Ambrosia artemisiifolia*, down) recorded by electron microscopy at the Zentrum für Mikroskopie (ZMB), University of Basel.

Plant pollen (Figure 1) are known to induce respiratory allergy and consequent asthma via proteic allergens present in the grain. Nicotinamide adenine dinucleotide phosphate (reduced) oxidases, bioactive lipid mediators, and other irritants are defined as pollen allergens and are contained in pollen fragments able to penetrate lower airways inducing an immune-modulated asthmatic response [18]. The pollen characterized with the highest number of allergens belong to families of Poaceae, Cupressaceae, Asteraceae, Oleaceae, and Betulaceae, and are restricted to 29 protein families (out of 157 characterized pollen proteins, 18%) [19]. The key factor to initiate an allergenic response is constituted by the highly affine interaction between the allergen and the

host antibody. The presentation of an allergen by antigen-presenting cells induces the release of the cytokine IL-4 from the T_H2 cell. The interaction between the T_H2 and the B cells induces the transformation of B cells into plasma cells, which under stimulation of IL-4 secretes the antibody IgE. IgE consequently binds to FcεRI receptors present on the surface of basophils and mast cells, which results in the cells sensitization to the specific allergen (Figure 2). A later exposure to the same allergen followed by its cross-binding to the IgE, will induce mast cell degranulation and release of inflammatory mediators such as histamine, cytokines, and prostaglandins. Specific structural regions in the protein defined as epitopes are responsible for the recognition of an allergen by the antibody. In case of IgE antibodies, responsible for the type 1 hypersensitivity [20], high affinity is shown towards conformational epitopes [21]. Besides the allergen-antibody affinity, the immunologic reaction can be modulated by other mechanisms. Regulatory T-cells (Tregs) are important regulators of the immune response. Namely, Tregs downregulate the proliferation of effector T-cells inducing an immunosuppressive reaction [22], and therefore maintain a healthy immune response.

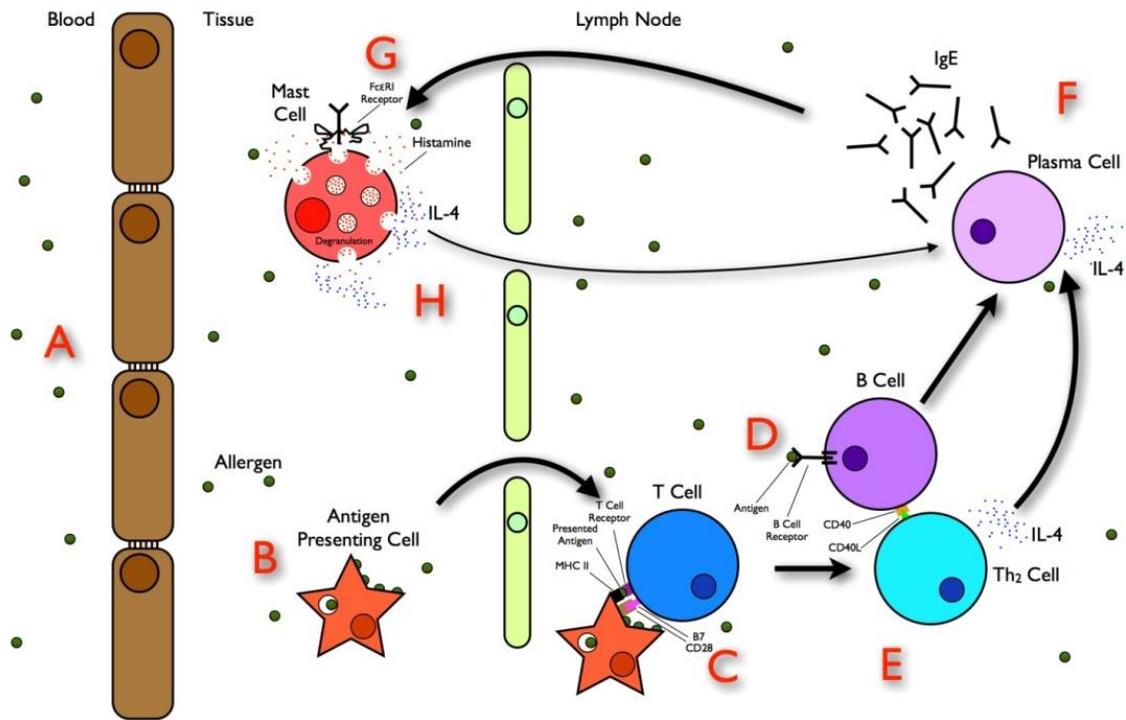


Figure 2. Allergenic reaction pathway. Early stage type 1 reaction and sensitization process [23].

The great multitude of cells and pathways involved in the immunologic and inflammatory reaction emphasizes the complexity of an allergy response. Furthermore, the individual genetic predisposition for the production of specific antibodies and receptors involved in the immunologic cascade makes it hard to predict the final clinical outcome.

Causes of increased prevalence of asthma and allergy in Western countries

Based on data presented in the World Health Organization asthma fact sheet (N° 307), last updated in 2013, asthma is a non-communicable disease which affects more than 200 million people worldwide. These numbers gain more importance if correlated with the worldwide uneven distribution of the asthma in single countries. The Global Initiative for Asthma (GINA) program performed a worldwide study on the burden of asthma based on literature published in the International Study of Asthma and Allergies in Childhood (ISAAC) and the European Community Respiratory Health Survey (ECHRS) [24]. These studies revealed a highest asthma prevalence in industrialized countries, and were correlated with the so called “westernized lifestyle” defined in four points: (1) “paradox of modern housing”. It ascribes the modern housing environment (warm and humid) favorable for the proliferation of strong allergens such as house dust mite [25]; (2) the “hygiene hypothesis” is being largely discussed and it supports the direct correlation of the lower microbial pressure, especially in the early childhood, with a higher prevalence of asthma [26, 27]; (3) lot of studies have been performed to understand how the sedentary lifestyle and the consequent obesity can be correlated to asthma [28]. Recently, Baffi and coworkers [29, 30] tried to give a rationale between asthma and obesity. On the mechanistic aspect, they pointed out the increasing accumulation of fat tissue around the chest wall and abdomen as a direct cause for losing lung volume capacity. Furthermore, they highlighted the bi-directional relationship between asthma and obesity (low physical activity leads to obesity, and obesity leads to low physical activity); (4) the westernized lifestyle is characterized by an increasing consumption of processed food. The “dietary hypothesis” proposes the changes of specific nutrients content like antioxidants, lipids, and oligo elements – caused by extensive food processing and storage – as contribution to the increase in prevalence of asthma and allergy [31].

References

1. Tekiner, H. Aretaeus of Cappadocia and his treatises on diseases. *Turk. Neurosurg.* **2015**, 25, 508-512.
2. Karamanou, M.; Androutsos, G. Aretaeus of Cappadocia and the first clinical description of asthma. *Am. J. Respir. Crit. Care Med.* **2011**, 184, 1420-1421.
3. Rosner, F. Moses Maimonides' treatise on asthma. *Thorax* **1981**, 36, 245-251.
4. Parascandola, J.; Ihde, A. J. History of the pneumatic trough. *Isis* **1969**, 60, 351-361.
5. Ramazzini, B. De morbis artificum diatriba [diseases of workers]. 1713. *Am. J. Public Health* **2001**, 91, 1380-1382.
6. Igea, J. M. The history of the idea of allergy. *Allergy* **2013**, 68, 966-973.
7. Holgate, S. T. A brief history of asthma and its mechanisms to modern concepts of disease pathogenesis. *Allergy Asthma Immunol. Res.* **2010**, 2, 165-171.
8. Moffatt, M. F.; *et al.* A large-scale, consortium-based genomewide association study of asthma. *N. Engl. J. Med.* **2010**, 363, 1211-1221.
9. Koczulla, A. R.; *et al.* New concepts in asthma: clinical phenotypes and pathophysiological mechanisms. *Drug Discov. Today* **2017**, 22, 388-396.
10. Xiao, C.; *et al.* Defective epithelial barrier function in asthma. *J. Allergy Clin. Immunol.* **2011**, 128, 549-556.
11. Jacquet, A. Interactions of airway epithelium with protease allergens in the allergic response. *Clin. Exp. Allergy* **2011**, 41, 305-311.
12. Wenzel, S. E. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat. Med.* **2012**, 18, 716-725.
13. O'Donnell, R.; *et al.* Inflammatory cells in the airways in COPD. *Thorax* **2006**, 61, 448-454.
14. Graham, S. D. Epidemiology, pathology, and pathophysiology, **2012**, Oxford University Press, 2nd Edition.
15. Goodman, D. C.; *et al.* Trends in pediatric asthma hospitalization rates: regional and socioeconomic differences. *Pediatrics* **1998**, 101, 208-213.
16. Dougherty, R. H.; Fahy, J. V. Acute exacerbations of asthma: epidemiology, biology and the exacerbation-prone phenotype. *Clin. Exp. Allergy* **2009**, 39, 193-202.
17. Backer, J. F. W.; Currie, G. P. Asthma, **2012**, Oxford University Press, 2nd Edition.
18. Taylor, P. E.; *et al.* Links between pollen, atopy and the asthma epidemic. *Int. Arch. Allergy Immunol.* **2007**, 144, 162-170.
19. Radauer, C.; Breiteneder, H. Pollen allergens are restricted to few protein families and show distinct patterns of species distribution. *J. Allergy Clin. Immunol.* **2006**, 117, 141-147.
20. Gould, H. J.; *et al.* The biology of IgE and the basis of allergic disease. *Annu. Rev. Immunol.* **2003**, 21, 579-628.
21. Pomes, A. Relevant B cell epitopes in allergic disease. *Int. Arch. Allergy Immunol.* **2010**, 152, 1-11.
22. Bettelli, E.; *et al.* Reciprocal developmental pathways for the generation of pathogenic effector Th17 and regulatory T cells. *Nature* **2006**, 441, 235-238.
23. https://en.wikipedia.org/wiki/Allergy#/media/File:The_Allergy_Pathway.jpg (accessed on 19.08.2016).
24. Masoli, M.; *et al.* The global burden of asthma: executive summary of the GINA Dissemination Committee report. *Allergy* **2004**, 59, 469-478.
25. Cinteza, M.; Daian, C. House dust mite – the paradox. *Mædica* **2014**, 9, 313-315.
26. Brooks, C.; *et al.* The hygiene hypothesis in allergy and asthma: an update. *Curr. Opin. Allergy Clin. Immunol.* **2013**, 13, 70-77.

27. Strachan, D. P. Hay fever, hygiene, and household size. *BMJ* **1989**, 299, 1259-1260.
28. Delgado, J.; et al. Obesity and asthma. *J. Investig. Allergol. Clin. Immunol.* **2008**, 18, 420-425.
29. Baffi, C. W.; et al. Asthma and obesity: mechanisms and clinical implications. *Asthma Res. Pract.* **2015**, 1, 1-7.
30. Baruwa, P.; Sarmah, K. R. Obesity and asthma. *Lung India* **2013**, 30, 38-46.
31. Devereux, G. The increase in the prevalence of asthma and allergy: food for thought. *Nat. Rev. Immunol.* **2006**, 6, 869-874.

2.6 Neuro-muscular pathways regulating airway mechanics

Airway smooth muscle contraction is a result of cyclic cross-bridging of actin and smooth muscle myosin, which is regulated by the myosin phosphorylation state. The reversible phosphorylation of the myosin is simultaneously coordinated on one side by the myosin light chain kinase (MLCK), and on the other side by the myosin light chain phosphatase (MLCP). Increase in cytosolic calcium (Ca^{2+}), by modulation of regulatory proteins (e.g. lipid kinase PIP5K γ) [1], induces the messenger protein calmodulin (CaM)-mediated activation of the myosin light chain kinase (MLCK), responsible for phosphorylation of the myosin filament. In parallel, RhoA directly activates Rho-associated coiled-coil protein kinase (ROCK), which phosphorylates and thereby inactivates the myosin light chain phosphatase (MLCP) responsible for myosin dephosphorylation, thus dilatation [2]. Besides a direct control of airway smooth muscle contraction by activation of regulatory mechanisms in the myocytes, an indirect modulation through the sympathetic and para-sympathetic nervous systems also plays a central role. Muscarinic and adrenergic receptor subunits are coupled to different G protein subunits, and lead to constriction or dilatation of airway. For example, the sympathetic α_1 adrenoceptor coupled to the G protein subunit $G_{aq/11}$ in efferent neurons leads to a phospholipase C (PLC)-mediated increase of cytosolic Ca^{2+} in airway smooth muscle (ASM). Differently, the adrenergic β_2 receptor activates the subunit G_s resulting in cyclic adenosine monophosphate (cAMP)-mediated decrease of Ca^{2+} , and thus inducing dilatation [3]. The muscarinic receptors (subtypes M_1 - M_5) are also involved in the contraction of the ASM, but regulating different pathways. Namely, the stimulation of M_1 , M_3 , and M_5 subtypes promotes the PLC-driven production of inositol triphosphate, which has as consequence the increase of cytosolic Ca^{2+} , released from the sarcoplasmic reticulum [4]. On the other side, the muscarinic subtypes M_2 and M_4 are indirectly involved in the ASM contraction through the inhibition of the adenylate cyclase, responsible to produce cAMP [1]. Tachykinin peptides are neuropeptides involved in many biological processes including neurogenic inflammation, mediated by the substance P, and the bronchoconstriction induced by neurokinin A [5].

Furthermore, the ASM contractility was also shown to be remote controlled at the endogenous level by many mediators released from organs (Figure 1) [6]. For example, epinephrine (or adrenaline) released from the adrenal glands induces a β_2 -agonist mediated ASM dilatation, whereas glucocorticoids showed effects on gene expression of ASM associated with an

improvement in airway hyper responsiveness [7]. Leptin, a hormone released from adipose tissue, debilitates the parasympathetic constricting effect on ASM through M₃ muscarinic receptor [8]. On the other hand, the blood factor fibrin has been correlated with an increased Airway Hyper Responsiveness in asthma (AHR), caused by a decrease in alveolar surface tension [6]. Guanylin, and in a smaller extent uroguanylin, were characterized by relaxing effects on ASM due to increased cGMP production [9], whereas insulin, the pancreatic hormone responsible for the regulation of blood glucose levels, has been characterized as a mediator in the ASM regulation by coordinating the availability of glucose, as energy source for ASM contraction [6]. Besides the adrenergic and muscarinic receptors, the voltage-gated ion channels and Transient Receptor Potential (TRP) ion channels were characterized as important proteins in the regulation of the airway smooth muscle tone.

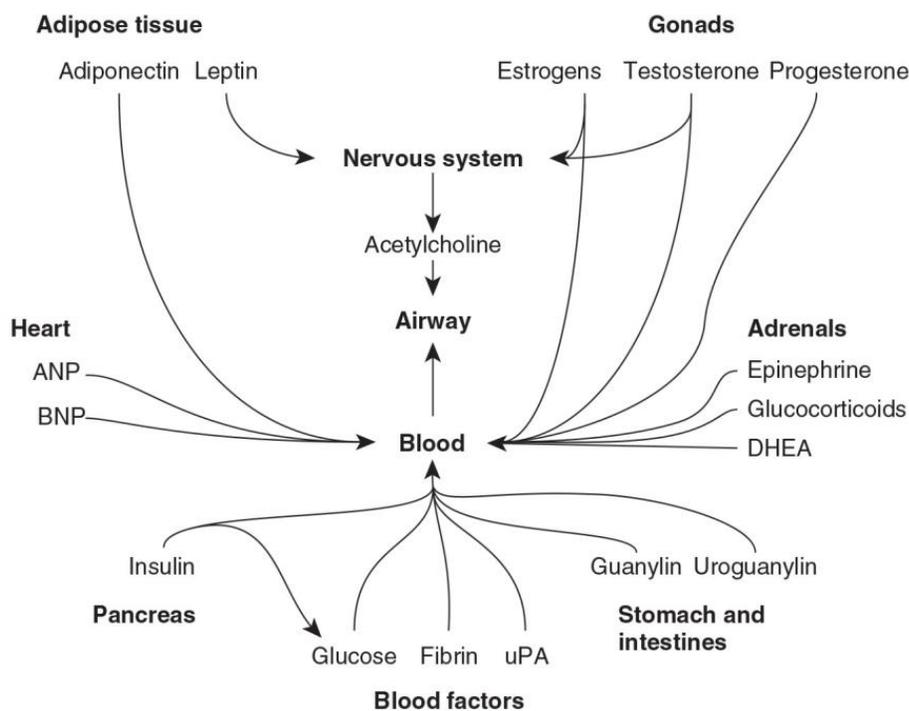


Figure 1. Endogenous mediators contribute to the direct or indirect regulation of airway smooth muscle relaxation or contraction. It can result in a direct regulation by contracting or relaxing the airway smooth muscle (ASM), or an indirect regulation by modulating the ASM responsiveness to spasmogens and bronchodilators. Mediators like estrogen, testosterone, adiponectin, and leptin indirectly control the ASM responsiveness via the parasympathetic nervous system [6].

Transient receptor potential cation channel, member A1 (TRPA1)

In 1995 Wes and colleagues characterized for the first time the transient receptor potential channel-related protein 1 (TRPC1), a human homolog of *trp* and with 40% similarity to *Drosophila* TRP [10]. Today, the transient receptor potential (TRP) family of ion channels represents an ensemble of 30+ cation channels divided in seven main families, based on sequence homology: the TRPC (Canonical), the TRPV (Vanilloid), the TRPM (Melastatin), the TRPP (Polycystin), the TRPML (Mucolipin), the TRPA (Ankyrin), and the TRPN (NOMPC - no mechanoreceptor potential C) family [11]. The human TRPA1 is expressed in hair cells [12], dorsal root ganglia and trigeminal ganglia neurons [13], and was first described as ANKTM1 by Story and colleagues in 2003 [14]. TRPA1 is a transmembrane 120-130 kDa protein encoded by the *trpa1* gene and functions as a homotetramer (four units of TRPA1 protein) (Figure 2) [15]. It is a Ca²⁺ permeable non-selective cation channel and belongs to the large TRP family of ion channels, characterized as sensory proteins, and involved in important homeostatic functions. TRPA1 is expressed on somatosensory neurons but recently was also found on non-neuronal cells like vascular smooth muscle, keratinocytes, and endothelium cells [16]. Both covalent binding of electrophiles to nucleophilic cysteine residues, and lock-and-key interaction through non-covalent intermolecular forces belong to TRPA1 activation mechanisms [17] (Chart 1, Chart 2). Each TRPA1 unit of the channel contains 31 cysteines, most of them present on the intracellular fragment and all potentially involved in the covalent binding. Nevertheless, only Cys415, Cys422, and Cys622 have been shown to be involved in the channel activation in humans (Figure 3). On the other hand, the murine TRPA1 activation was shown to be modulated by cysteines located in regions of the Ankyrin repeat domain different from those in humans [18], sometimes showing species-specific biological effects [19].

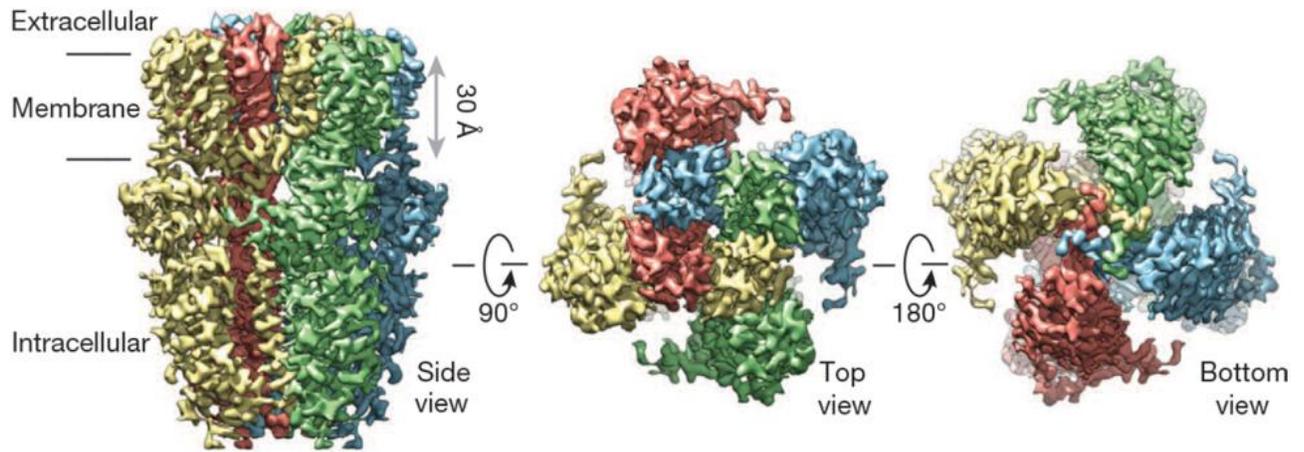


Figure 2. Human TRPA1 3D density map with side, top, and bottom view [15].

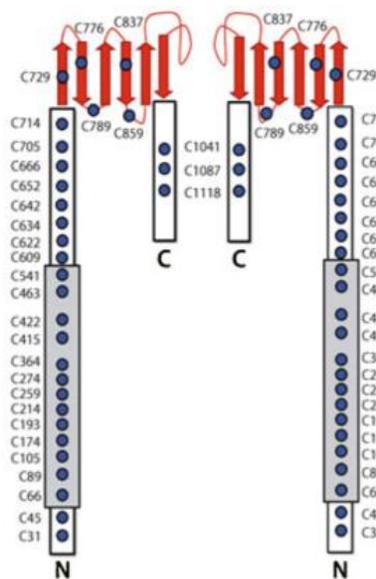


Figure 3. Schematic representation of TRPA1 dimer with all 31 cysteines present in the protein structure (blue circles) [18].

Electrophiles and non-electrophiles with TRPA1 modulation properties

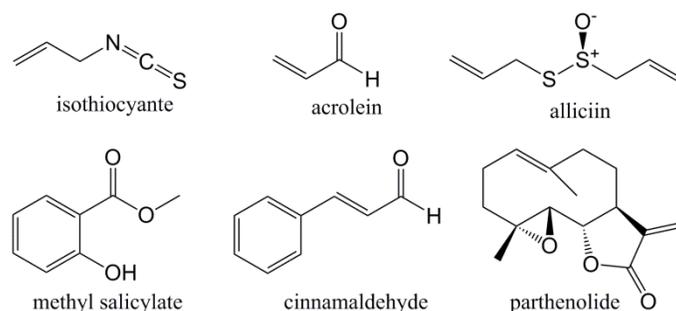


Chart 1. Electrophiles from various origins as TRPA1 activators via covalent binding to cysteine and/or lysine residues [6].

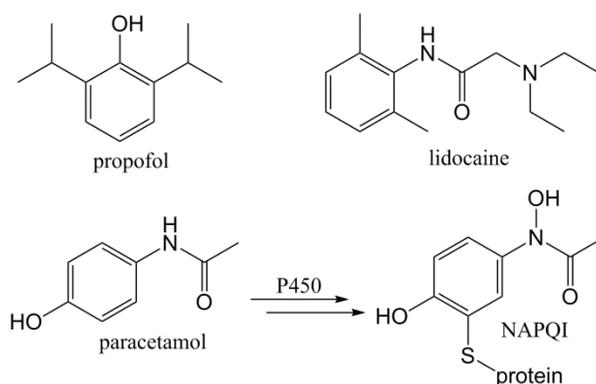


Chart 2. Non-electrophiles as TRPA1 activators via lock-and-key mechanism [6]. Paracetamol is a non-electrophilic compound, which in case of overdose induces depletion of glutathione and undergoes metabolic transformation to the electrophilic *N*-acetyl-*p*-benzo-quinone imine (NAPQI).

Type I Phosphatidylinositol 4-Phosphate 5-Kinase gamma (PIP5K γ)

Phosphatidylinositols (PtdIns) are important secondary messengers in eukaryotic cells involved in the regulation of a plethora of cellular functions like cellular metabolism, protein synthesis, and vesicle trafficking. The signaling specificity of PtdIns is attributed to the different positions and extent of phosphorylation of the inositol ring, which is regulated by phosphatases and phosphoinositide lipid kinases (PIKs). PIKs are diversified in three major families: PtdIns 3-kinases (PI3Ks), PtdIns 4-kinases (PI4Ks), and PtdIns-P (PIP) kinases (PIP5Ks and PIP4Ks). Type I PIP5K is one of the three PIP kinases and PIP5K γ (PIP5K1C) is one of them [20]. The lipid kinase PIP5K γ is the most important regulator of intracellular Ca²⁺ in smooth muscle cells.

PIP5K γ produces phosphatidylinositol 4,5-bisphosphate (PIP₂), substrate for PLC beta1 mediated IP₃ production, which induces the release of Ca²⁺ from the sarcoplasmic reticulum with consequent induction of smooth muscle contraction. Spermidine and spermine are crucial cofactors for PIP5K γ and its activation is regulated by co-localization of the polyamine-catabolizing enzyme spermidine/spermine N1-acetyltransferase (SSAT) (Figure 4) [1].

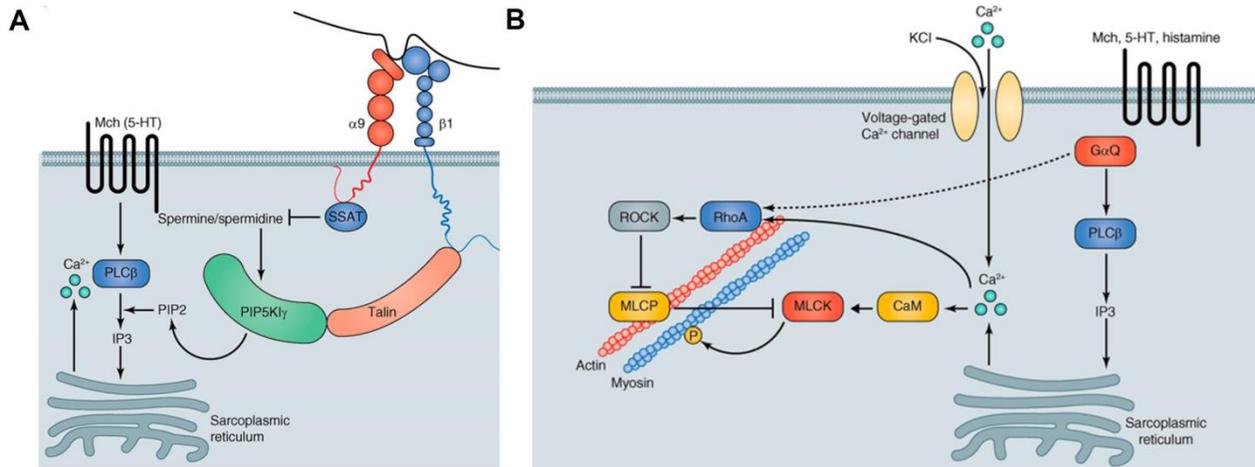


Figure 4. Schematic representation of lipid kinase PIP5K γ -mediated Ca²⁺ release from sarcoplasmic reticulum into cytosol. Polyamines spermidine and spermine are co-factors regulating the activity of the PIP5K γ (A). Ca²⁺-mediated smooth muscle contraction (B) [1].

In vitro bioassay for assessing intracellular Ca²⁺ influx

Calcium is an important biologic messenger regulating many cellular processes with complex spatial (cellular compartments) and temporal forms and coordinates the function of multicellular systems [21]. The Ca²⁺ homeostasis is regulated by calcium-regulating hormones [22]. The perturbation of the Ca²⁺ homeostatic state has been in focus of lot of studies and it was suggested to be the major cause in neurodegenerative diseases like Alzheimer [23-26] and Parkinson [27], and was correlated with the modulation of blood pressure and obesity [28]. The extraordinary importance of Ca²⁺ in many physiological processes led to the development of sophisticated strategies to monitor the Ca²⁺ physiologic dynamics at all levels, from cellular to intact organisms [29]. Calcium imaging method implies the cell loading with chemical or genetically encoded indicators [30, 31]. By the selection of the appropriate indicator, different aspects must be taken into consideration. For example, the binding affinity of the Ca²⁺ to the indicator is a very

important parameter. In fact, depending on the compartmental Ca^{2+} concentrations an indicator with a convenient dissociation constant must be chosen. The different spectral properties of different indicators can also significantly influence the final read-out. Single wavelength indicators are characterized by changing fluorescent intensities at a stable excitation and emission wavelengths, whereas ratiometric indicators have the advantage of shifting their excitation or emission curve upon binding. Ca^{2+} indicators are present in different forms. Salts and dextran conjugates both require invasive introduction in the cell, but the dextran conjugates were shown to not compartmentalize and thus permitting the cytosolic Ca^{2+} measurement over a long period of time. Acetoxymethyl esters (AM) are a form of indicators that permit an easy cellular loading due to their increased lipophilicity [32]. Fura-2AM is a membrane-permeable AM ester derivative of aminopolycarboxylic acid (Fura-2). In the cytosol, Fura-2AM gets metabolized by esterases to Fura-2, an active calcium-binding ratiometric dye, which is characterized by an emission shift after binding to calcium (Figure 5). The intensity ratio is calculated at $340_{\text{free}}/380_{\text{bound}}$ nm, corresponding to wavelengths where difference of fluorescence between bound and free indicator is maximum. Recording of intensities and the calculation of ratios is performed by sophisticated digital microscopy coupled to advanced computing technologies for final data analysis.

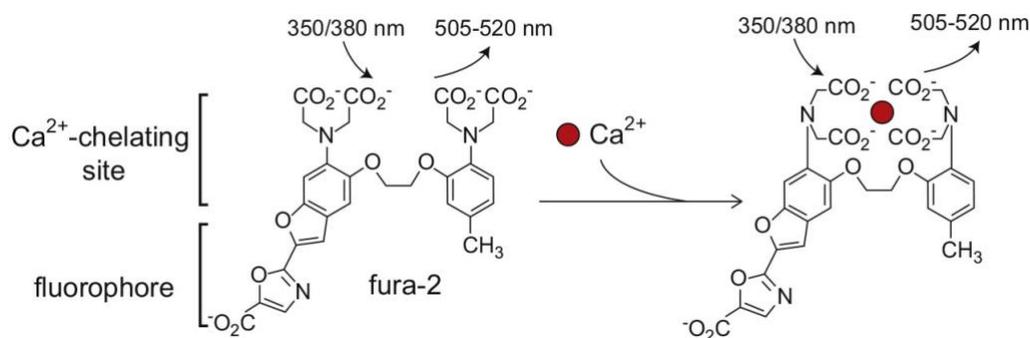


Figure 5. Molecular structure and chelation mechanism of Fura-2 [32].

For the *in vitro* bioassays we used the laboratory mouse strain C57BL/6. This is one of the most used mouse strains, from which good models of human diseases can be developed.

Isolation and preparation of murine afferent neurons for Ca²⁺ imaging

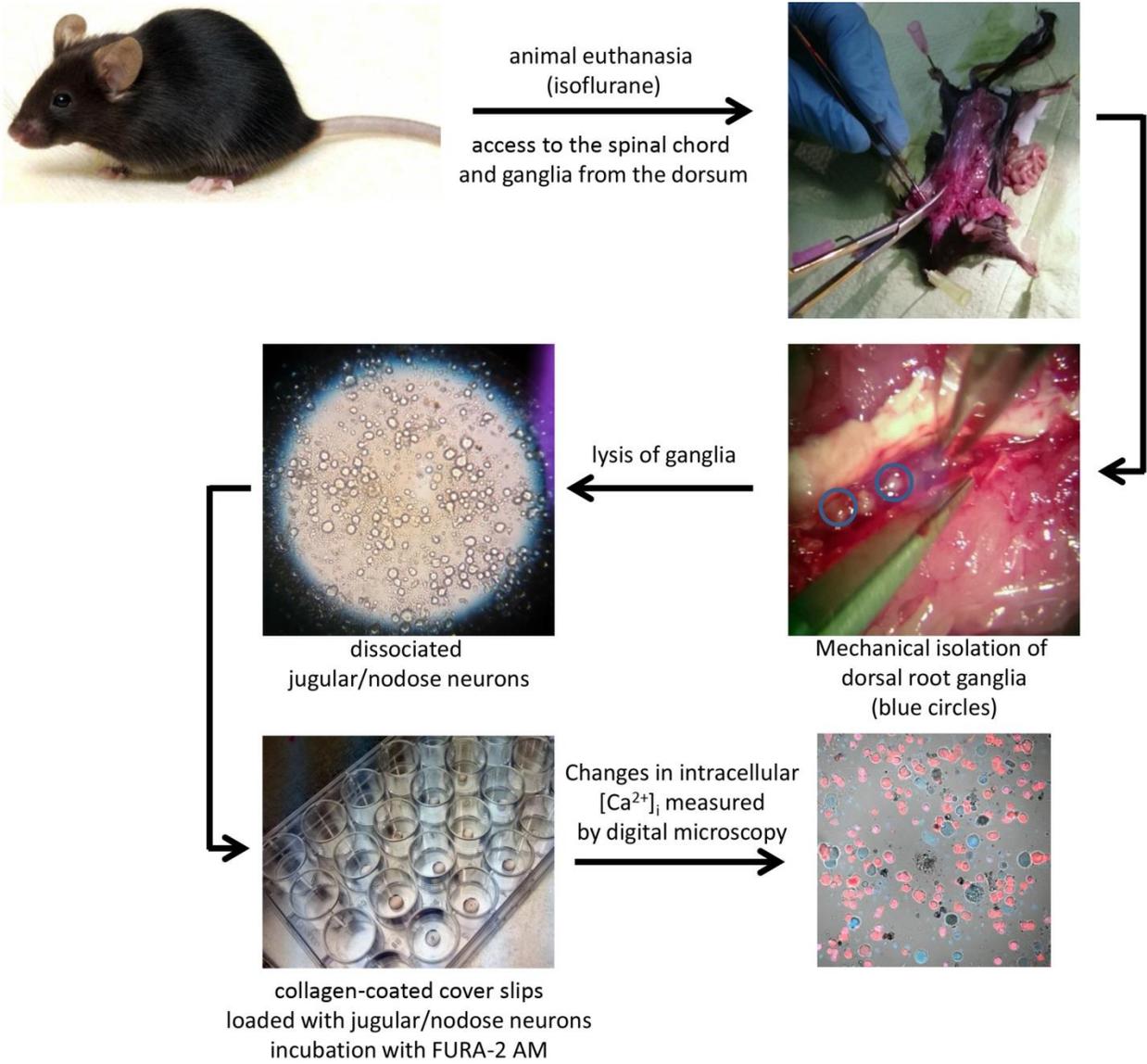


Figure 6. Schematic workflow for isolation of jugular/nodose neurons from male and female C57BL/6 mice (age range: 10 to 16 weeks) After mechanical isolation of dorsal root ganglia and their enzymatic lysis, dissociated jugular/nodose neurons are fixed on collagen-coated coverslips and incubated with Fura-2 AM for 24 h. For more details, see Bozicevic et al., JNP.

In vitro bioassay for assessing tracheal tone modulation

The murine tracheal rings preparations are being widely used for exploring the physiological and pathophysiological mechanisms in airways. *In vitro* organ bath experiments with wild-type and knock-out mice models largely contributed to the functional characterization of Transient Receptor Potential (TRP) ion channels [33], muscarinic receptors [34], and for the evaluation of different asthmatic phenotypes [35]. The organ bath experiments are performed with specific equipment consisting of a tissue bath, force transducer, and an analog to digital converter (A/D). The tissue bath is a 10-15 mL flask provided with water jacket for circulating a warming solution (37°C), a fritted glass inlet to bubble oxygen (95%/5% O₂/CO₂ mixture), and inlet and outlet ports for changing solutions. The isometric tension of the tracheal ring preparations is measured with two stainless steel rods (one fixed and one connected to the force transducer) with an L-shaped end threaded in the inner part of the ring. Contraction of the trachea creates tension on the force transducer, which is converted to a voltage signal at the preamplifier (Figure 7). The relaxation can be also measured in this configuration because the tracheal rings are maintained at a constant tension of 0.5 grams. Finally, the A/D converter transforms the signal from analog to digital, which is then transferred to acquisition software.

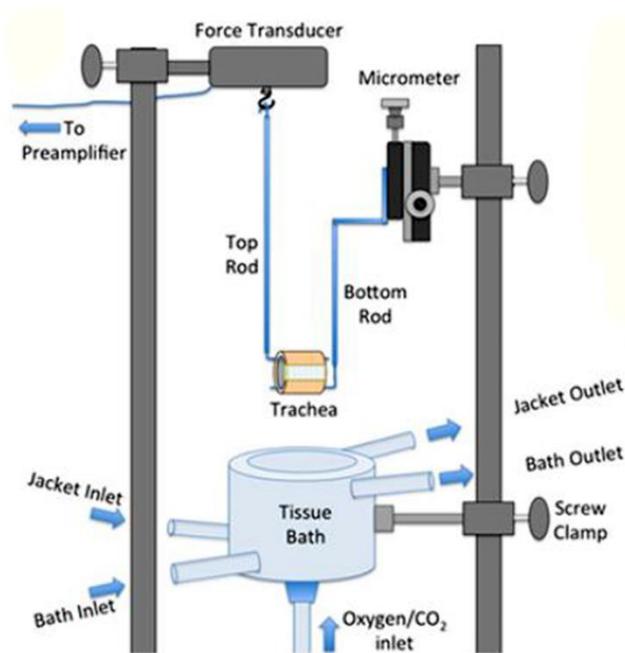


Figure 7. adapted from [36].

References

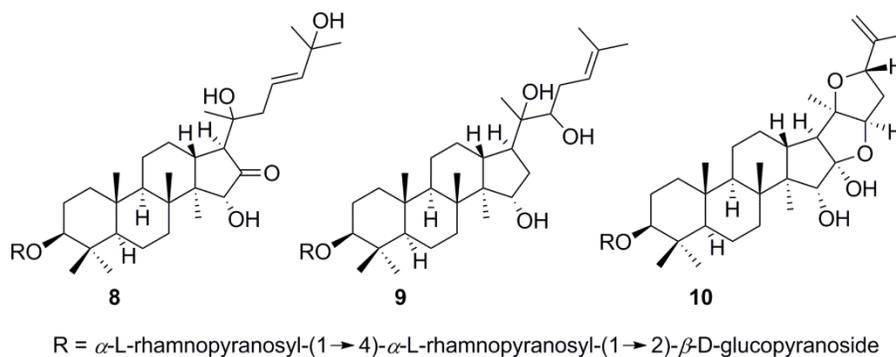
1. Erle, D. J.; Sheppard, D. Cell biology of disease: the cell biology of asthma. *J. Cell Biol.* **2014**, 205, 621-631.
2. Webb, R. C. Smooth muscle contraction and relaxation. *Adv. Physiol. Educ.* **2003**, 27, 201-206.
3. Billington, C. K. cAMP regulation of airway smooth muscle function. *Pulm. Pharmacol. Ther.* **2016**, 26, 112-120.
4. Canning, B. J.; Fischer, A. Neural regulation of airway smooth muscle tone. *Respir. Physiol.* **2001**, 125, 113-127.
5. Boot, J. D.; *et al.* Effect of an NK1/NK2 receptor antagonist on airway responses and inflammation to allergen in asthma. *Am. J. Respir. Crit. Care Med.* **2007**, 175, 450-457.
6. Bossé, Y. Endocrine regulation of airway contractility is overlooked. *J. Endocrinol.* **2014**, 222, 61-73.
7. Yick, C. Y.; *et al.* Glucocorticoid-induced changes in gene expression of airway smooth muscle in patients with asthma. *Am. J. Respir. Crit. Care Med.* **2013**, 187, 1076-1084.
8. Arteaga-Solis, E.; *et al.* Inhibition of leptin regulation of parasympathetic signaling as a cause of extreme body weight-associated asthma. *Cell Metab.* **2013**, 17, 35-48.
9. Ohbayashi, H.; *et al.* Effects of uroguanylin and guanylin against antigen-induced bronchoconstriction and airway microvascular leakage in sensitized guinea-pigs. *Life Sci.* **1998**, 62, 1833-1844.
10. Wes, P. D.; *et al.* TRPC1, a human homolog of a *Drosophila* store-operated channel. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, 92, 9652-9656.
11. Pedersen, S. F.; *et al.* TRP channels: an overview. *Cell Calcium* **2005**, 38, 233-252.
12. Corey, D. P.; *et al.* TRPA1 is a candidate for the mechanosensitive transduction channel of vertebrate hair cells. *Nature* **2004**, 432, 723-730.
13. Nassenstein, C.; *et al.* Expression and function of the ion channel TRPA1 in vagal afferent nerves innervating mouse lungs. *J. Physiol.* **2008**, 586, 1595-1604.
14. Story, G. M.; *et al.* ANKTM1, a TRP-like channel expressed in nociceptive neurons, is activated by cold temperatures. *Cell* **2003**, 112, 819-829.
15. Paulsen, C. E.; *et al.* Structure of the TRPA1 ion channel suggests regulatory mechanisms. *Nature* **2015**, 520, 511-517.
16. Fernandes, E. S., *et al.* The functions of TRPA1 and TRPV1: moving away from sensory nerves. *Br. J. Pharmacol.* **2012**, 166, 510-521.
17. Nilius, B.; *et al.* The transient receptor potential channel TRPA1: from gene to pathophysiology. *Pflugers Arch., EJP* **2012**, 464, 425-458.
18. Macpherson, L. J.; *et al.* Noxious compounds activate TRPA1 ion channels through covalent modification of cysteines. *Nature* **2007**, 445, 541-545.
19. Leffler, A.; *et al.* Activation of TRPA1 by membrane permeable local anesthetics. *Mol. Pain* **2011**, 7, 62.
20. Brown, J. R.; Auger, K. R. Phylogenomics of phosphoinositide lipid kinases: perspectives on the evolution of second messenger signaling and drug discovery. *BMC Evol. Biol.* **2011**, 11, 4.
21. Thomas, A. P.; *et al.* Spatial and temporal aspects of cellular calcium signaling. *FASEB J.* **1996**, 10, 1505-1517.
22. Peacock, M. Calcium metabolism in health and disease. *Clin. J. Am. Soc. Nephrol.* **2010**, 5, 23-30.
23. Hölscher, C. Possible causes of alzheimer's disease: amyloid fragments, free radicals, and calcium homeostasis. *Neurobiol. Dis.* **1998**, 5, 129-141.
24. Kawahara, M. Disrupted calcium homeostasis implicated in Alzheimer disease pathogenesis. *Curr. Alzheimer Res.* **2004**, 1, 87-95.

25. Dreses-Werringloer, U.; *et al.* A polymorphism in CALHM1 influences Ca²⁺ homeostasis, Abeta levels, and Alzheimer's disease risk. *Cell* **2008**, 133, 1149-1161.
26. Cheung, K. H.; *et al.* Mechanism of Ca²⁺ disruption in Alzheimer's disease by presenilin regulation of InsP3 receptor channel gating. *Neuron* **2008**, 58, 871-883.
27. Chan, C. S. *et al.* Calcium homeostasis, selective vulnerability and Parkinson's disease. *Trends Neurosci.* **2009**, 32, 249-256.
28. Zemel, M. B. Calcium modulation of hypertension and obesity: mechanisms and implications. *J. Am. Coll. Nutr.* **2001**, 20, 428-435.
29. Russell, J. T. Imaging calcium signals in vivo: a powerful tool in physiology and pharmacology. *Br. J. Pharmacol.* **2011**, 163, 1605-1625.
30. Hoi, H.; *et al.* Highlightable Ca²⁺ indicators for live cell imaging. *J. Am. Chem. Soc.* **2013**, 135, 46-49.
31. Paredes, R. M.; *et al.* Chemical calcium indicators. *Methods* **2008**, 46, 143-151.
32. Grienberger, C.; Konnerth, A. Imaging calcium in neurons. *Neuron* **2012**, 73, 862-885.
33. Zhou, Y.; *et al.* Sensitivity of bronchopulmonary receptors to cold and heat mediated by transient receptor potential cation channel subtypes in an *ex vivo* rat lung preparation. *Respir. Physiol. Neurobiol.* **2011**, 177, 327-332.
34. Garssen, J.; *et al.* Functional characterization of muscarinic receptors in murine airways. *Br. J. Pharmacol.* **1993**, 109, 53-60.
35. Herlihy, J. T.; *et al.* Assessment of airway hyperresponsiveness in murine tracheal rings. Humana Press: Totowa, NJ, **2013**. Springer Link. Web. 14 Feb. 2017. http://link.springer.com/protocol/10.1007%2F978-1-62703-496-8_20.
36. Semenov, I.; *et al.* *In vitro* measurements of tracheal constriction using mice. *J. Vis. Exp.* **2012**, (64), 3703.

3. RESULTS AND DISCUSSION

3.1 Dammarane-type saponins from leaves of *Ziziphus spina-christi*

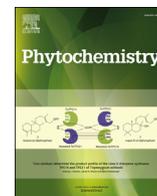
Alen Bozicevic, Maria De Mieri, Angela Di Benedetto, Frank Gafner, Matthias Hamburger
Phytochemistry **2017**, 138, 134-144. doi: 10.1016/j.phytochem.2017.02.028.



Phytochemical profiling of *Ziziphus spina-christi* leaves led to the characterization of 10 dammarane-type saponins and 12 phenolic compounds. Isolation was achieved by gel chromatography on Sephadex LH20, open column chromatography on silica gel, and semi-preparative HPLC with PDA and ELSD detectors. Structural characterization was performed by extensive 1D and 2D NMR, mass spectrometry, and by GC-MS of sugar derivatives. A biosynthetic pathway leading to three previously undescribed sapogenins is proposed. The saponin profiles in *Z. spina-christi* leaves of four different origins were compared by means of HPLC-ESIMS.

Extraction of plant material, fractionation (Sephadex, open column silica gel, and HPLC), writing of the manuscript, and preparation of the figures were my contribution. Interpretation of analytical data for structural elucidation (mass spectrometry, NMR) was a contribution of both Maria De Mieri and me. The isolation and analysis of flavonoids was a contribution of Angela Di Benedetto with my support.

Alen Bozicevic



Dammarane-type saponins from leaves of *Ziziphus spina-christi*



Alen Bozicevic^{a,1}, Maria De Mieri^{a,1}, Angela Di Benedetto^a, Frank Gafner^b,
Matthias Hamburger^{a,*}

^a Institute of Pharmaceutical Biology, University of Basel, Klingelbergstrasse 50, CH-4056, Basel, Switzerland

^b Mibelle Biochemistry, Mibelle AG, 5033, Buchs, Switzerland

ARTICLE INFO

Article history:

Received 28 October 2016

Received in revised form

21 February 2017

Accepted 27 February 2017

Available online 6 March 2017

Keywords:

Ziziphus spina-christi

Rhamnaceae

Dammarane saponins

Lotogenin

LC-MS

ABSTRACT

Phytochemical profiling of *Ziziphus spina-christi* leaves led to the characterization of 10 dammarane-type saponins and 12 phenolic compounds. Isolation was achieved by gel chromatography on Sephadex LH20, open column chromatography on silica gel, and semi-preparative HPLC with PDA and ELSD detectors. Structural characterization was performed by extensive 1D and 2D NMR, mass spectrometry, and by GC-MS of sugar derivatives. A biosynthetic pathway leading to three previously undescribed saponin profiles is proposed. The saponin profiles in *Z. spina-christi* leaves of four different origins were compared by means of HPLC-ESIMS.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Ziziphus species (Rhamnaceae) occur in warm-temperate and sub-tropical regions around the world, and are commonly used in folk medicine for the treatment of various diseases, such as digestive disorders, weakness, liver complaints, obesity, urinary problems, diabetes, skin infections, fever, diarrhea, and insomnia (Abdel-Zaher et al., 2005). *Ziziphus spina-christi* (L.) Desf. (Christ's Thorn Jujube, Sedr in Farsi, and Konar in Arabic) is a wild evergreen tree characterized by spiny branches and small orange fruits that is native of subtropical eastern Africa, the Middle East, and the Indian subcontinent, but also widespread in northwestern Africa. In some countries of the Middle East, leaves of *Z. spina-christi* are harvested from the wild or from cultures to be mainly sold on local markets. The Iranian province of Khuzestan is known for the extensive cultivation of *Z. spina-christi*, and leaves are used in Iran as a natural detergent and shampoo due to the saponin content. Surprisingly, the phytochemistry of *Z. spina-christi* leaves is poorly studied, and only a limited number of polyphenols (Nawwar et al., 1984) and jujubogenin glycosides have been reported up to now (Mahran et al., 1996). More information is available regarding the

phytochemical composition of *Z. spina-christi* and *Z. jujuba* fruits. Pawlowska et al. (2009) identified numerous flavonoids in both species, and various phenolics, triterpenoids, sterols, cyclopeptide alkaloids, and amino acids have been reported from *Z. jujuba* (Gao et al., 2013). In view of a possible use of *Z. spina-christi* leaf extract as a cosmetic ingredient due to its surfactant properties, a phytochemical analysis of the leaves was performed with an emphasis on saponins. Finally, the saponin profiles in *Z. spina-christi* leaf samples of four different origins were compared by means of HPLC-ESIMS, in order to assess the variability of saponin patterns.

2. Results and discussion

Leaves of *Z. spina-christi* were submitted to sequential extraction with EtOAc and 35% (v/v) ethanol. After evaporation of the hydroalcoholic extract, the residue was suspended in water and partitioned with *n*-BuOH to remove sugars and other highly polar compounds. The *n*-BuOH layer was separated by gel filtration on Sephadex LH20, open column chromatography on silica gel, and semi-preparative HPLC to afford ten triterpene saponins **1–10**. Of these, compounds **8–10** contained new dammarane type aglycons. In addition, eleven phenolic compounds (**11–21**) were identified, including the new quercetin glycoside **11**. The HPLC profile of the *n*-BuOH layer, and peaks corresponding to compounds **1–21** are shown in Fig. 1.

Structure elucidation was achieved by means of extensive 1D and 2D NMR analysis and HRESIMS. In addition, sugar moieties in

* Corresponding author.

E-mail addresses: alen.bozicevic@unibas.ch (A. Bozicevic), maria.demieri@unibas.ch (M. De Mieri), angi.dibenedetto@stud.unibas.ch (A. Di Benedetto), matthias.hamburger@unibas.ch (M. Hamburger).

¹ These authors contributed equally.

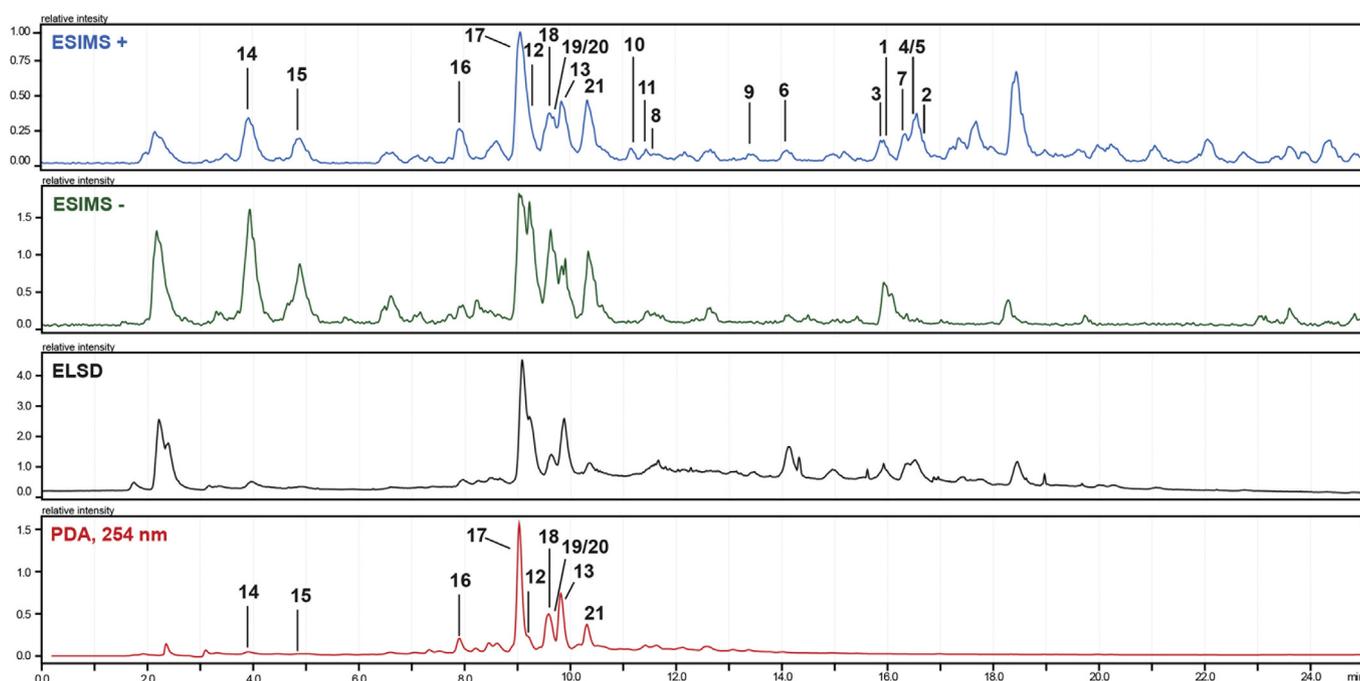


Fig. 1. HPLC profile of the *n*-BuOH layer of *Ziziphus spina-christi* leaf extract recorded with PDA (254 nm), ELSD, and ESIMS (positive and negative ion mode). Peak numbers designate saponins **1–10** and polyphenols **11–21**.

saponins **1**, **6**, and **9**, and in flavonoid **11** were identified by GC-MS after acid hydrolysis followed by derivatization with L-cystein methyl ester and subsequent silylation, and comparison with derivatized reference sugars. In case of minor saponins **2–5**, **7**, **8** and **10**, we did not perform a hydrolysis due to the limited amount of material. We here assume the same absolute configuration of sugars as found in saponins **1**, **6**, and **9**, given the same nature of individual sugars, and the relatedness of sugars chains to those in **1**, **6** and **9**.

Saponins **1** and **2** were identified as jujuboside B1 (Matsuda et al., 1999) and christinin A (Mahran et al., 1996) which had been previously reported from leaves of *Z. jujuba* and *Z. spina-christi*, respectively (Fig. 2).

The negative ion HRESIMS spectrum of **3** showed a signal at m/z 969.5150 $[M-H]^-$ corresponding to a molecular formula of $C_{49}H_{78}O_{19}$ (calcd for $C_{49}H_{77}O_{19}$: 969.5137). 1D and 2D NMR spectra of **3** were virtually superimposable to those of **2**. The only differences were in the presence of an oxygenated methine (δ_H 4.51, br s) in **3** vs a methylene in **2**, and of an acetyl moiety (δ_H 2.05, s; δ_C 169.5) in **3**. The location of the oxygenated methine at C-22 was inferred from its COSY correlation with H-23, and corroborated through HMBC correlations with C-17, C-20, and C-21 (Fig. 3). An HMBC correlation of H-22 with the carboxyl group at δ 169.5 established the location of the acetyl moiety. The relative configuration of the triterpene skeleton corresponded to that reported for christinin A (Fig. 2). The ROESY contact of H-23/H-15a (Fig. 3) indicated a β -orientation of H-23, while the α -orientation of the 22-acetoxy moiety was inferred from the small J_{HH} coupling of H-22/H-23 (br s), and corroborated by the absence of ROESY contacts of H-22/H-17. Thus, the structure of **3** was established as 22 α -acetoxy christinin A (Fig. 2).

Saponins **4** and **5** were obtained as an inseparable mixture. Molecular formula of $C_{49}H_{78}O_{18}$ and $C_{50}H_{78}O_{20}$ were deduced from ions at m/z 953.5161 $[M-H]^-$, (calcd for $C_{49}H_{77}O_{18}$: 953.5188) and 997.5072 $[M-H]^-$, (calcd for $C_{50}H_{77}O_{20}$: 997.5086) in the HRESIMS spectra of **4** and **5**, respectively. Inspection of the NMR data revealed the presence of jujubogenin as aglycon in both saponins, and of a trisaccharidic moiety that closely resembled that of **2**.

Additional signals were detected in the mixture of **4** and **5**, namely of a methylene group (δ_H 3.50, δ_C 70.7), a methyl group (δ_H 1.99, δ_C 21.6) (Fig. S8), and signals attributable to carboxylic groups in the DEPTq spectrum (δ 167.8 and δ 168.7) (Fig. S9). Hence, saponins were esterified with an acetyl and malonyl moiety, respectively. The signals corresponding to the CH_2 -6 of the glucose moiety (δ_H 4.31 and 4.09/ δ_C 64.7) appeared downfield compared to those in **2** (δ_H 3.65 and 3.45/ δ_C 61.0), and thereby indicated attachment of the acyl moieties at this position. The saponins were thus identified as jujubogenin 3- β -O-(6-O-acetyl)- β -D-glucopyranosyl-(1 \rightarrow 3)-[α -D-fucopyranosyl-(1 \rightarrow 2)]- α -L-arabinopyranoside (**4**) and jujubogenin 3- β -O-(6-O-malonyl)- β -D-glucopyranosyl-(1 \rightarrow 3)-[α -D-fucopyranosyl-(1 \rightarrow 2)]- α -L-arabinopyranoside (**5**), and were named as christinin A1 and A2, respectively (Fig. 2).

Compound **6** had a molecular formula of $C_{48}H_{80}O_{18}$, as deduced from the signal at m/z 943.5363 (calcd for $C_{48}H_{79}O_{18}$: 943.5345) in the negative ion HRESIMS spectrum. 1H and ^{13}C NMR spectra showed the presence of six methyl singlets in the high field region (δ 0.74, 0.80, 0.83, 0.93, 1.02, and 1.15), of a 3-methyl-2-butenyl group (δ 1.58 and 1.65, H₃-26 and 27; δ 5.15, H-24; δ 2.13, H₂-23), and of two oxygenated quaternary carbons (δ 109.5, C-16; and δ 76.5, C-20). The NMR data of the aglycon were in good accord with those of lotogenin (Renault et al., 1997). In addition, three anomeric protons were detected in the 1H NMR spectrum at δ 4.24 (d, $J = 7.4$ Hz), δ 5.34 (br s), and δ 5.03 (br s), with ^{13}C resonances of their corresponding carbon atoms at δ 104.0, δ 98.9, and δ 101.0. By 1H - 1H -COSY, 1D selective TOCSY, and 2D HSQC-TOCSY experiments the three sugars spin systems were assigned to one β -glucose and two units of α -rhamnose, respectively (Table 3). β -Glucose was identified by the J values of H-1' to H-5' that showed all-*trans* diaxial interactions. The α -rhamnose units were identified by the small J values of H-2''/H-2''' and their vicinal protons (dd, $^3J_{H-2/H-3} = 3.6$ Hz, and $^3J_{H-2/H-1} =$ br s) (same for compound **11**), while the ^{13}C -NMR shifts of C-3''/C-3''' (δ 71.2 and 70.9) and C-5''/C-5''' (δ 65.9 and 68.8) confirmed the α configuration at the anomeric carbons. Long range $^3J_{HC}$ HMBC correlations of H-1''' (δ 5.03) of Rha II and C-4'' (δ 78.4) of Rha I, and H-1'' (δ 5.34) of Rha I with C-2' (δ 75.0) of

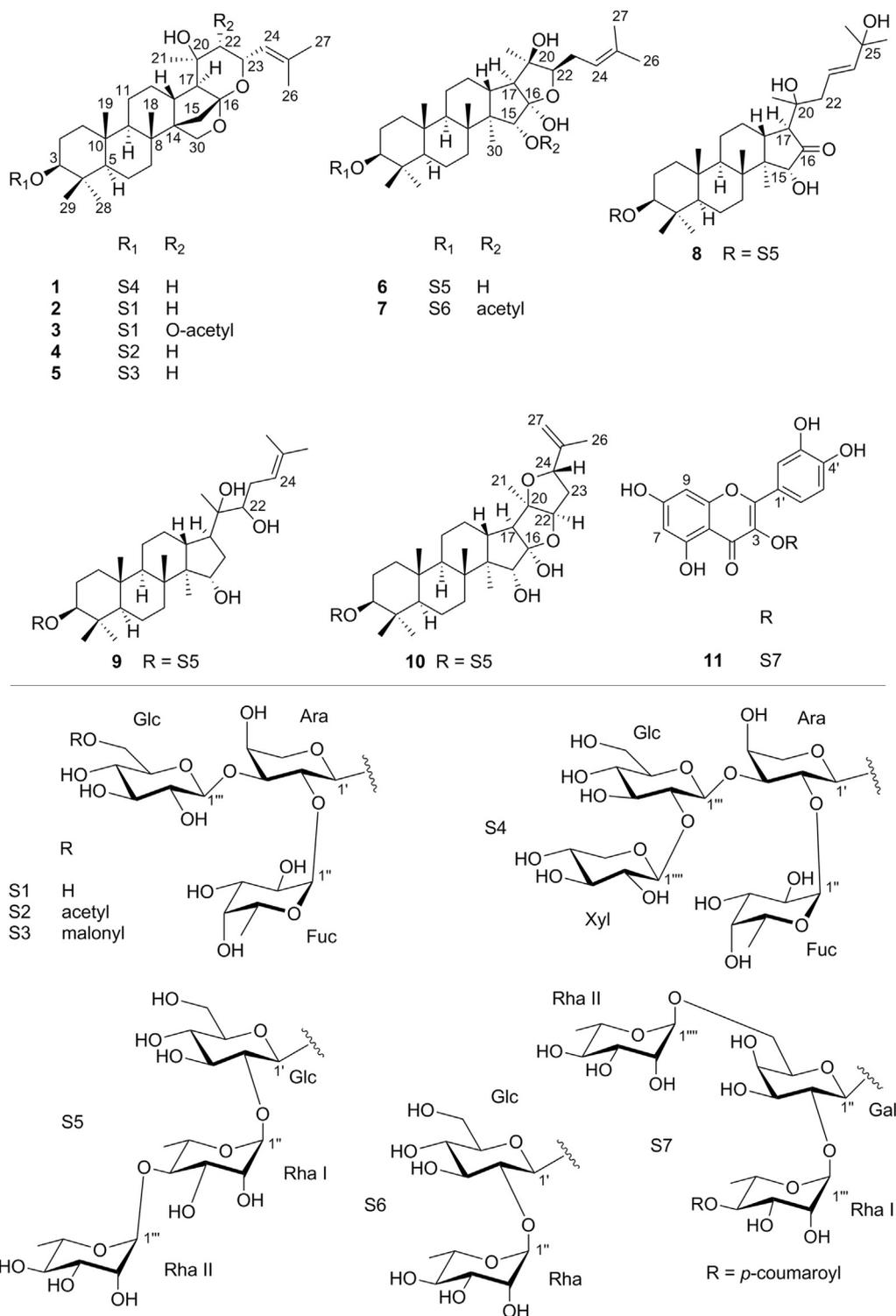


Fig. 2. Structures of saponins 1–10 and flavonoid 11.

Glc established the trisaccharide chain. ROESY contacts of H-1''' (δ 5.03) and H-4'' (δ 3.33), H-1'' (δ 5.34) and H-2' (δ 3.28) (Fig. 3) supported this assignment. Attachment of the trisaccharide moiety at C-3 of the aglycon via a β -oriented glycosidic linkage was inferred from the $^3J_{\text{HC}}$ HMBC correlation between H-1' (δ 4.24) and C-3 (δ 88.3) of the aglycon, and by the J_{HH} coupling of H-3 (δ 3.03, dd, J = 11.4, and 4.0 Hz) which was indicative of its axial disposition in the chair-like conformation. The relative configuration of the

aglycon was identical to that reported for lotogenin, as indicated by key ROESY correlations (Fig. 3). The absolute configurations of glucose and rhamnose were determined as described for compound 1, 6, and 9. Thus, compound 6 was established as lotogenin 3- β -O- α -L-rhamnopyranosyl-(1 \rightarrow 4)- α -L-rhamnopyranosyl-(1 \rightarrow 2)- β -D-glucopyranoside, and was named as lotside III (Fig. 2).

A molecular formula of C₄₄H₇₂O₁₅ was assigned to compound 7

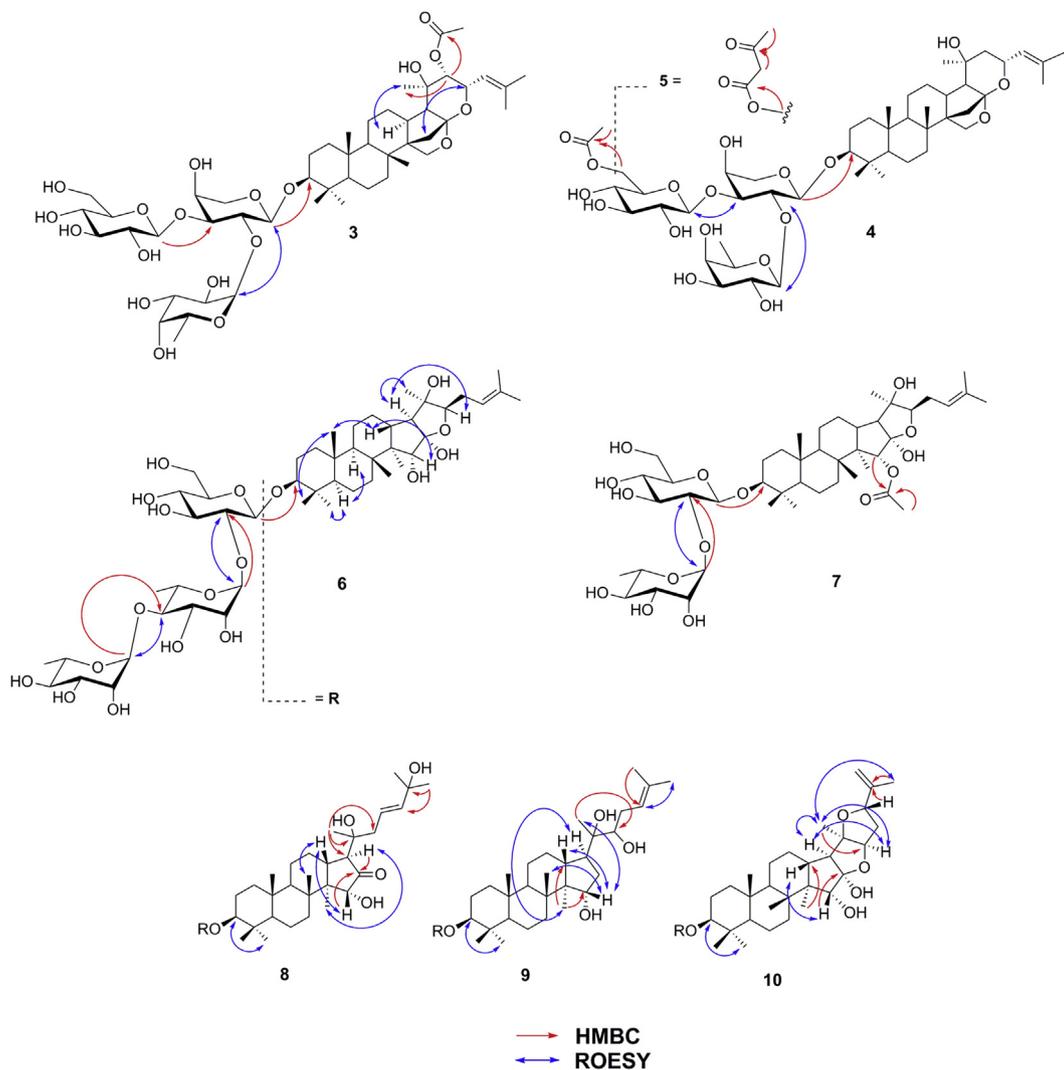


Fig. 3. Key ^1H - ^1H ROESY and HMBC correlations in compounds 3–10.

on the basis of a molecular ion at m/z 839.4849 in the negative ion HRESIMS spectrum (calcd for $\text{C}_{44}\text{H}_{71}\text{O}_{15}$: 839.4871). Inspection of the NMR data indicated lotogenin as aglycon, and the presence of an acetyl moiety (δ_{H} , 1.98, s; δ_{C} 21.6, C-32; δ_{C} 169.9, C-31). Attachment of the acetyl group at C-15 was established by an HMBC correlation of H-15 (δ 5.06) to C-31 (δ 169.9), and confirmed by the downfield shift of H-15 compared to H-15 in **6** (δ 3.73). The two sugars moieties in **7** were identified by COSY and 1D selective TOCSY experiments as β -glucopyranoside and α -rhamnopyranoside (Figs. S18 and S22). Key HMBC correlations of H-1'' (δ 5.24) and C-2' (δ 76.6), H-1' (δ 4.23) and C-3 (δ 87.8) established the interglycosidic linkage and the attachment of the disaccharide to C-3 of lotogenin (Renault et al., 1997) (Fig. 3). Hence, the structure of **7** was established as 15-acetoxy-lotogenin 3- β -O- α -L-rhamnopyranosyl-(1 \rightarrow 2)- β -D-glucopyranoside, and the compound was named as 15-acetoxy lotoside IV (Fig. 2).

Saponin **8** showed a molecular ion at m/z 943.5339 in the negative ion HRESIMS spectrum, corresponding to a molecular formula of $\text{C}_{48}\text{H}_{80}\text{O}_{18}$ (calcd for $\text{C}_{48}\text{H}_{79}\text{O}_{18}$: 943.5345). Chemical shifts of three anomeric protons at δ 4.25 (d, $J = 7.4$ Hz), δ 5.34 (br s), and δ 5.06 (br s) with corresponding carbons at δ 103.5, 98.7, and 100.5 were identical to those of the sugar moiety in **6**. Analysis of key HMBC and ROESY correlations confirmed the trisaccharide moiety as α -rhamnopyranosyl-(1 \rightarrow 4)- α -rhamnopyranosyl-(1 \rightarrow 2)- β -

glucopyranoside (Fig. 3). Additional resonances detected in the HSQC and HMBC spectra were eight methyls, seven methylenes, eight methines, and seven quaternary carbons including a carbonyl (δ 216.1) indicating a lotogenin-like aglycon (Renault et al., 1997). Compared to lotogenin, compound **8** lacked of the quaternary carbon attached to a hemiacetal group (δ 109.5). This was replaced by a carbonyl (δ 216.1), as confirmed by HMBC correlations of H-17/C-16 and H-15/C-16 (Fig. 3). Consequently, C-20 and C-22 were located on a side chain, and the spin system was identified as shown in Fig. 2. The relative configuration of the triterpene skeleton was corroborated by key ROESY correlations of H-17/H₃-30 and H-13/H-15 (Fig. 3), which established the relative configuration at C-17 and C-15. However, the relative configuration at C-20 could not be established due to the free rotation of the side chain, and limited amounts of material precluded derivatization. The aglycon was identified as 3 β ,15 α ,20,25-tetraol-dammar-23(24)-ene-16-one, and this new dammarane-type triterpenoid was named sidrigenin. Thus, compound **8** was sidrigenin 3-O- α -L-rhamnopyranosyl-(1 \rightarrow 4)- α -L-rhamnopyranosyl-(1 \rightarrow 2)- β -D-glucopyranoside (Fig. 2).

Compound **9** showed a molecular ion of m/z 929.5563 in the negative HRESIMS spectrum corresponding to a molecular formula of $\text{C}_{48}\text{H}_{82}\text{O}_{17}$ (calcd for $\text{C}_{48}\text{H}_{81}\text{O}_{17}$: 929.5552). The sugar moiety was identical to that of **8**, but the aglycon portion was slightly different. By means of COSY, HSQC, and HMBC experiments C-16 was

identified as a methylene (δ_{H} 1.82 and 1.24), while the methylene C-22 in compound **8** (δ_{C} 42.3) was replaced by a hydroxylated methine (δ_{C} 75.3). The signal of the only olefinic proton in **9** (δ 5.21, br t, $J = 7.0$ Hz; H-24) was coupled with a methylene group (δ 2.08, ddd, $J = 13.0, 7.0, 3.0$ Hz; H-23a; δ 1.80, m; H-23b) due to the presence of a double bond at positions 24 and 25 (Table 2). The relative configuration of the tetracyclic portion of the aglycon was established by analysis of J couplings, and confirmed by ROESY data (Fig. 3). However, the relative configuration at the stereogenic carbons C-21 and C-22 could not be established by NMR due to the flexibility of the side chain. Formation of an acetonide (2,2-dimethoxypropane and catalytic amount of *p*-toluenesulfonic acid in dry acetone at r.t. for 2 h) (Ihre et al., 1998) was attempted with the aglycon after acid hydrolysis, but failed due to decomposition of the aglycon during the reaction. The aglycon was thus established as 3 β ,15 α ,20,22-tetraol-dammar-24(25)-ene and was named as konarigenin. Hence, **9** was konarigenin 3-*O*- α -L-rhamnopyranosyl-

(1 \rightarrow 4)- α -L-rhamnopyranosyl-(1 \rightarrow 2)- β -D-glucopyranoside (Fig. 2).

Compound **10** had a molecular formula of $\text{C}_{48}\text{H}_{78}\text{O}_{18}$ (m/z 941.5167 in the negative HRESIMS; calcd for $\text{C}_{48}\text{H}_{77}\text{O}_{18}$: 941.5188). ^1H and ^{13}C -NMR chemical shifts of the trisaccharide moiety were superimposable to those of **6**, **8**, and **9** (Table 3, Table 4). The aglycon of **10** showed close similarity to lotogenin, but with differences in the fragment of C23–C27. The olefinic methine at C-24 of lotogenin was replaced in **10** by an oxygenated sp^3 methine (δ 3.95, br s), and the 2-methylpropenyl in lotogenin was replaced in **10** by an isopropylene residue (δ 4.89, br s, and 4.71, br s; H-27a and H-27b; and δ 1.66, s, H₃-26). Furthermore, an epoxy bridge between C-20 and C-24 was in accord with the molecular formula and required degrees of hydrogen deficiency. The relative configuration was determined by ROESY experiments (Fig. 3). Key ROESY contacts of H₃-21 with H-17 and H-22 established the cofacial orientation of these groups. Due to the close chemical shifts of H-22 and H-24

Table 1
 ^1H and ^{13}C spectroscopic data for the aglycon moieties of **3–6** (DMSO- d_6) (δ in ppm, J in Hz).

No.	3		4		5		6	
	$\delta_{\text{H}}^{\text{a}}$	$\delta_{\text{C}}^{\text{b}}$, type	$\delta_{\text{H}}^{\text{a}}$	$\delta_{\text{C}}^{\text{b}}$, type	$\delta_{\text{H}}^{\text{a}}$	$\delta_{\text{C}}^{\text{b}}$, type	$\delta_{\text{H}}^{\text{a}}$	δ_{C} , type
1	1.58 0.88, m	38.1, CH ₂	1.58 0.88, m	38.1, CH ₂	1.58 0.88, m	38.1, CH ₂	1.58 0.88, m	39.0, CH ₂
2	1.71, m 1.54	25.5, CH ₂	1.71, m 1.54	25.5, CH ₂	1.71, m 1.54	25.5, CH ₂	1.90 1.50	26.1, CH ₂
3	3.03, dd (12.0, 4.0)	86.9, CH	3.03, dd (12.0, 4.0)	86.9, CH	3.03, dd (12.0, 4.0)	86.9, CH	3.03, dd (11.4, 4.0)	88.3, CH
4	–	39.3, C	–	39.3, C	–	39.3, C	–	38.8, C
5	0.71	55.1, CH	0.71	55.1, CH	0.71	55.1, CH	0.68, br d (12.0)	56.1, CH
6	1.46 1.39	17.2, CH ₂	1.46 1.39	17.2, CH ₂	1.46 1.39	17.2, CH ₂	1.41 1.34	17.6, CH ₂
7	1.47 1.37	35.1, CH ₂	1.47 1.37	35.1, CH ₂	1.47 1.37	35.1, CH ₂	1.47 1.35	35.5, CH ₂
8	–	36.5, C	–	36.5, C	–	36.5, C	–	40.6, C
9	0.83, br d (12.0)	51.7, CH	0.83, br d (12.0)	51.7, CH	0.83, br d (12.0)	51.7, CH	1.22, m	50.5, CH
10	–	36.4, C	–	36.4, C	–	36.4, C	–	36.5, C
11	1.51 1.34	20.7, CH ₂	1.51 1.34	20.7, CH ₂	1.51 1.34	20.7, CH ₂	1.40 1.14	20.6, CH ₂
12	1.69 1.59	27.3, CH ₂	1.69 1.59	27.3, CH ₂	1.69 1.59	27.3, CH ₂	1.68 1.14	25.6, CH ₂
13	2.44, m	35.2, CH	2.44, m	35.2, CH	2.44, m	35.2, CH	1.93, m	34.9, CH
14	–	52.9, C	–	52.9, C	–	52.9, C	–	52.2, C
15a	1.94, d (8.2)	35.4, CH	1.94, d (8.2)	35.4, CH ₂	1.94, d (8.2)	35.4, CH ₂	3.73	75.6, CH
15b	1.05	–	1.05	–	1.05	–	–	–
16	–	108.5, C	–	109.4, C	–	109.4, C	–	109.5, C
17	1.10, br d (6.2)	48.3, CH	0.84	52.8, CH	0.84	52.8, CH	1.69, d (11.4)	61.9, CH
18	1.03, s	18.0, CH ₃	1.03, s	18.0, CH ₃	1.03, s	18.0, CH ₃	1.02, s	16.0, CH ₃
19	0.79, s	15.7, CH ₃	0.79, s	15.7, CH ₃	0.79, s	15.7, CH ₃	0.80, s	16.6, CH ₃
20	–	68.8, C	–	68.8, C	–	68.8, C	–	76.5, C
21	0.96, s	24.5, CH ₃	1.03, s	29.3, CH ₃	1.03, s	29.3, CH ₃	1.15, s	26.9, CH ₃
22	4.51, br s	73.3, CH	1.37 1.24	44.2, CH ₂	1.37 1.24	44.2, CH ₂	3.66	87.5, CH
23	4.73, br d (8.0)	67.8, CH	4.59	67.2, CH	4.59	67.2, CH	2.13, m	27.2, CH ₂
24	4.99, d (8.0)	121.0, CH	5.08, br d (8.0)	125.9, CH	5.08, br d (8.0)	125.9, CH	5.15, br t (7.0)	122.5, CH
25	–	135.2, C	–	133.5, C	–	133.5, C	–	131.1, C
26	1.61, s	17.9, CH ₃	1.61, s	17.9, CH ₃	1.61, s	17.9, CH ₃	1.58, br s	17.8, CH ₃
27	1.65, s	25.0, CH ₃	1.65, s	25.0, CH ₃	1.65, s	25.0, CH ₃	1.65, br s	25.7, CH ₃
28	0.91, s	26.7, CH ₃	0.91, s	27.6, CH ₃	0.91, s	27.6, CH ₃	0.93, s	27.4, CH ₃
29	0.70, s	15.5, CH ₃	0.70, s	16.0, CH ₃	0.70, s	16.0, CH ₃	0.74, s	15.8, CH ₃
30a	3.88	64.2, CH ₂	3.88 d (7.2)	64.2, CH ₂	3.88 d (7.2)	64.2, CH ₂	0.83, s	9.5, CH ₃
30b	3.78, d (7.2)	–	3.78, d (7.2)	–	3.78, d (7.2)	–	–	–
31	–	169.5, C	–	–	–	–	–	–
32	2.05, s	20.3, CH ₃	–	–	–	–	–	–

^a Overlapped ^1H -NMR signals are reported without multiplicities.

^b ^{13}C shifts were extracted from HSQC and HMBC data.

Table 2
¹H and ¹³C spectroscopic data for the aglycon moieties of **7–10** (DMSO-*d*₆) (δ in ppm, *J* in Hz).

No.	7		8		9		10	
	δ _H ^a	δ _C , type	δ _H ^a	δ _C ^b , type	δ _H ^a	δ _C ^b , type	δ _H ^a	δ _C ^b , type
1	1.57 0.89	38.8, CH ₂	1.58 0.94	38.5, CH ₂	1.57 0.92	38.9, CH ₂	1.58 0.94	39.0, CH ₂
2	1.89 1.50	26.0, CH ₂	1.93, m 1.53	25.5, CH ₂	1.91 1.52	25.6, CH ₂	1.91 1.52	25.6, CH ₂
3	3.01, dd (11.4, 4.0)	87.8, CH	3.03, dd (11.5, 4.0)	87.8, CH	3.02, dd (11.5, 4.0)	87.9, CH	3.05, dd (11.4, 4.0)	87.8, CH
4	—	40.6, C	—	38.5, C	—	38.0, C	—	39.3, C
5	0.68, br d (12.0)	55.6, CH	0.73	55.7, CH	0.70, m	55.5, CH	0.71, br d (11.5)	56.1, CH
6	1.37 1.32	17.6, CH ₂	1.44 1.39	17.0, CH ₂	1.42 1.33	17.2, CH ₂	1.45 1.38	17.1, CH ₂
7	1.42 0.98	34.6, CH ₂	1.55 1.35	35.0, CH ₂	1.59 1.40	35.0, CH ₂	1.44 1.25	35.4, CH ₂
8	—	40.6, C	—	39.6, C	—	39.3, C	—	40.2, C
9	1.22, m	50.3, CH	1.37, m	50.6, CH	1.25, m	50.6, CH	1.28, m	50.2, CH
10	—	36.5, C	—	36.7, C	—	36.4, C	—	37.5, C
11	1.40 1.13	20.7, CH ₂	1.51 1.28	20.6, CH ₂	1.41 1.10	21.0, CH ₂	1.44 1.16	20.5, CH ₂
12	1.67 1.18	25.2, CH ₂	1.81, m 1.30	26.3, CH ₂	1.74 1.21	27.0, CH ₂	1.75 1.26	24.8, CH ₂
13	2.02, m	34.9, CH	2.24, ddd (11.0, 10.3, 3.5)	34.1, CH	1.69, m	39.2, CH	2.07, m	37.5, CH
14	—	52.4, C	—	46.6, C	—	49.4, C	—	53.1, C
15	5.06, br s	76.7, CH	3.95	80.0, CH	3.86, dd (8.5, 8.5)	71.2, CH	4.16, br s	82.0, CH
16	—	109.9, C	—	216.1, C	1.82 1.24	32.9, CH ₂	—	118.7, C
17	1.75, d (11.4)	62.5, CH	1.70, d (10.3)	53.5, CH	1.67	42.9, CH	1.94, d (12.0)	61.0, CH
18	1.02, s	15.9, CH ₃	1.11, s	15.2, CH ₃	0.96, s	14.9, CH ₃	1.00, s	15.9, CH ₃
19	0.82, s	16.4, CH ₃	0.85, s	15.9, CH ₃	0.80, s	16.0, CH ₃	0.81, s	16.3, CH ₃
20	—	76.3, C	—	72.3, C	—	75.3, C	—	76.1, C
21	1.15, s	27.2, CH ₃	1.06, s	15.8, CH ₃	0.94, s	20.1, CH ₃	1.18, s	26.9, CH ₃
22a	3.70	87.6, CH	2.32, dd (13.3, 8.0)	42.3, CH ₂	3.20	75.3, CH	3.99	86.1, CH
22b	—	—	2.15, dd (13.3, 6.0)	—	—	—	—	—
23a	2.16, m	27.1, CH ₂	5.49, ddd (15.7, 8.0, 6.0)	121.1, CH	2.08, ddd (13.0, 7.0, 3.0)	30.0, CH ₂	1.60, m	34.6, CH ₂
23b	—	—	—	—	1.80, m	—	1.50	—
24	5.15, br t (7.0)	122.3, CH	5.56, d (15.7)	142.0, CH	5.21, br t (7.0)	122.6, CH	3.95, br s	70.6, CH
25	—	131.4, C	—	68.2, C	—	129.8, C	—	148.9, C
26	1.58, br s	17.8, CH ₃	1.14, s	29.8, CH ₃	1.54, s	17.4, CH ₃	1.66, s	17.7, CH ₃
27a	1.65, br s	25.6, CH ₃	1.14, s	29.8, CH ₃	1.65, s	25.1, CH ₃	4.89, br s	109.0, CH ₂
27b	—	—	—	—	—	—	4.71, br s	—
28	0.93, s	27.2, CH ₃	0.94, s	26.8, CH ₃	0.93, s	27.0, CH ₃	0.94, s	27.1, CH ₃
29	0.75, s	16.0, CH ₃	0.75, s	15.1, CH ₃	0.73, s	15.3, CH ₃	0.74, s	15.4, CH ₃
30	1.00, s	10.1, CH ₃	0.69, s	9.2, CH ₃	0.82, s	9.4, CH ₃	0.81, s	8.7, CH ₃
31	—	169.9, C	—	—	—	—	—	—
32	1.98, s	21.6, CH ₃	—	—	—	—	—	—

^a Overlapped ¹H-NMR signals are reported without multiplicities.

^b ¹³C shifts were extracted from HSQC and HMBC data.

resonances, the relative configuration at C-24 could not be determined by a ROESY spectrum. Spatial proximity of H₃-21 and H₃-26 was inferred from a selective ROESY experiment (Fig. S40) that indicated an α -orientation of the isopropylene group. The new aglycon of **10** was thus established as (3 β ,15 α ,16 α ,22*R*,24*S*)-3,15,16-triol-16,22:20,24-diepoxy-dammar-25(27)-ene, and named siconigenin. Hence, saponin **10** was identified as siconigenin 3-*O*- α -L-rhamnopyranosyl-(1 \rightarrow 4)- α -L-rhamnopyranosyl-(1 \rightarrow 2)- β -D-glucopyranoside (Fig. 2).

From a biogenetic perspective, the aglycon of **10** can be rationalized as being formed from aglycons of **8** and **9** via 20,22-dihydroxy,16-keto-dammar-24,25-ene (**a**) as a putative intermediate (Fig. 4). This intermediate would then undergo cyclization to a hemiketal, and subsequent nucleophilic attack of the C-22 hydroxy function onto the 24,25 olefin, followed by oxidation/reduction steps would lead to the aglycon of **10**.

Saponin profiles of *Z. spina-christi* leaves of different geographical origin were compared in order to evaluate possible qualitative and semi-quantitative differences that might occur. HPLC-ESIMS chromatograms of the *n*-BuOH portion of hydroalcoholic extracts of four samples are shown in Fig. S67. Besides the sample from Dezful (Iran) that was used in the phytochemical characterization,

samples from Caesarea (Israel), Nurabad (Iran), and a commercial sample purchased in a bazar in Tehran (Iran) were analyzed (Fig. 5). Jujubogenin glycosides **1–4** were the major saponins in samples from Caesarea and Nurabad, as compared to glycosides of lotogenin, sidragenin, konarigenin, and siconigenin (compounds **6–10**). The sample purchased at the bazar contained the lowest concentration as well as diversity of triterpene glycosides. Jujubogenin glycoside **5** was a major saponin in the sample from Dezful, but was not found in the other three samples. Lotogenin glycosides **6** and **7** were absent in the sample from the bazar. Sidragenin and siconigenin glycosides **8** and **10** were present in all samples in comparable concentrations, whereas konarigenin glycoside **9** was detected only in the samples from Dezful and Nurabad (Fig. 5).

The structure of the previously undescribed flavonol glycoside **11** (Fig. 2) was established as follows. A molecular ion was observed at *m/z* 901.2469 [M-H]⁻ in the HRESIMS spectrum corresponding to a molecular formula of C₄₂H₄₆O₂₂ (calcd for C₄₂H₄₅O₂₂: 901.2481). The UV spectrum was typical of a flavonol, with UV absorption maxima at 273 and 312 nm. The ¹H-NMR and ¹³C-NMR data (Table 5) showed characteristic signals of quercetin, with three proton resonances at δ 7.62 (dd, *J* = 8.5 and 2.0 Hz), 7.50 (d, *J* = 2.0 Hz), and 6.80 (d, *J* = 8.5 Hz), and two *meta*-coupled

Table 3
¹H and ¹³C spectroscopic data for the sugar moieties of **3–6** (DMSO-*d*₆) (δ in ppm, J in Hz).

Sugar	No	3		4		5		6	
		δ _H ^e	δ _C ^f , type	δ _H ^e	δ _C ^f , type	δ _H ^e	δ _C ^f , type	δ _H ^e	δ _C ^f , type
Glc ^a	1'	4.33, d (7.7)	102.8, CH	4.33, d (7.7)	102.8, CH	4.33, d (7.7)	102.8, CH	4.24, d (7.4)	104.0, CH
	2'	3.06, dd (8.5, 7.7)	73.2, CH	3.06, dd (8.5, 7.7)	73.2, CH	3.06, dd (8.5, 7.7)	73.2, CH	3.28, dd (9.0, 7.4)	75.0, CH
	3'	3.17, dd (9.0, 8.5)	76.6, CH	3.17, dd (9.0, 8.5)	76.6, CH	3.17, dd (9.0, 8.5)	76.6, CH	3.34, dd (9.0, 9.0)	78.3, CH
	4'	3.11, dd (9.0, 9.0)	69.3, CH	3.11, dd (9.0, 9.0)	69.3, CH	3.11, dd (9.0, 9.0)	69.3, CH	3.09	70.4, CH
	5'	3.14, ddd (9.0, 5.0, 2.0)	76.6, CH	3.14, ddd (9.0, 5.0, 2.0)	76.6, CH	3.14, ddd (9.0, 5.0, 2.0)	76.6, CH	3.09	76.6, CH
	6a'	3.65, br d (11.2, 2.0)	60.8, CH ₂	4.31 (11.2, 2.0)	64.7, CH ₂	4.31 (11.2, 2.0)	64.7, CH ₂	3.65, br d (11.0, 2.0)	61.1, CH ₂
	6b'	3.44, dd (11.2, 5.0)	–	4.09 (11.2, 5.0)	–	4.09 (11.2, 5.0)	–	3.45, dd (11.0, 5.0)	–
	31	–	–	–	167.8, ^g C	–	168.7, ^g C	–	–
	32	–	–	1.99, s	21.6, CH ₃	3.50	70.7, CH ₂	–	–
	33	–	–	–	–	–	168.7, ^g C	–	–
	Rha I ^b	1''	–	–	–	–	–	5.34, br s	98.9, CH
2''		–	–	–	–	–	3.63, br d (3.6)	70.7, CH	
3''		–	–	–	–	–	3.68, dd (9.0, 3.6)	71.2, CH	
4''		–	–	–	–	–	3.33, dd (9.0, 9.0)	78.4, CH	
5''		–	–	–	–	–	3.97, dq (9.0, 6.0)	65.9, CH	
6''		–	–	–	–	–	1.11, d (6.0)	18.3, CH ₃	
Rha II ^b	1'''	–	–	–	–	–	5.03, br s	101.0, CH	
	2'''	–	–	–	–	–	3.71, br d (3.6)	70.7, CH	
	3'''	–	–	–	–	–	3.37	70.9, CH	
	4'''	–	–	–	–	–	3.19, dd (9.0, 9.0)	72.0, CH	
	5'''	–	–	–	–	–	3.49, dd (9.0, 6.0)	68.8, CH	
	6'''	–	–	–	–	–	1.10, d (6.0)	17.8, CH ₃	
Ara ^c	1''	4.29, d (6.2)	103.6, CH	4.29, d (6.2)	103.6, CH	4.29, d (6.2)	103.6, CH	–	
	2''	3.73, dd (8.5, 6.2)	71.9, CH	3.73, dd (8.5, 6.2)	71.9, CH	3.73, dd (8.5, 6.2)	71.9, CH	–	
	3''	3.70, dd (8.5, 3.0)	81.6, CH	3.70, dd (8.5, 3.0)	81.6, CH	3.70, dd (8.5, 3.0)	81.6, CH	–	
	4''	3.87, m	66.6, CH	3.87, m	66.4, CH	3.87, m	66.4, CH	–	
	5a''	3.68	64.3, CH ₂	3.68	64.3, CH ₂	3.68	64.3, CH ₂	–	
	5b''	3.39	–	3.39	–	3.39	–	–	
Fuc ^d	1'''	5.31, br s	100.0, CH	5.31, br s	100.0, CH	5.31, br s	100.0, CH	–	
	2'''	3.63, br d (6.0)	70.1, CH	3.63, br d (6.0)	70.1, CH	3.63, br d (6.0)	70.1, CH	–	
	3'''	3.60, m	65.0, CH	3.60, m	65.0, CH	3.60, m	65.0, CH	–	
	4'''	3.56, m	72.1, CH	3.56, m	72.1, CH	3.56, m	72.1, CH	–	
	5'''	4.11, br q (6.2)	66.0, CH	4.11, br q (6.2)	66.0, CH	4.11, br q (6.2)	66.0, CH	–	
	6'''	1.07, d (6.2)	16.5, CH ₃	1.07, d (6.2)	16.0, CH ₃	1.07, d (6.2)	16.0, CH ₃	–	

^a Glc = β-D-glucopyranosyl.

^b Rha = α-L-rhamnopyranosyl.

^c Ara = α-L-arabinopyranosyl.

^d Fuc = β-D-fucopyranosyl.

^e Overlapped ¹H-NMR signals are reported without multiplicities.

^f ¹³C shifts were extracted from HSQC and HMBC data.

^g Interchangeable spin systems of sugars were resolved by 1D selective TOCSY and 2D HSQC-TOCSY.

resonances at δ 6.37 (br s) and 6.19 (br s) attributable to rings B and A, respectively. Three anomeric protons at δ 5.53 (d, *J* = 7.8 Hz), 5.12 (br s), and 4.40 (br s) indicated the presence of 3 sugars. By ¹H-¹H-COSY, 1D selective TOCSY, and 2D HSQC-TOCSY experiments the three sugars spin systems were assigned to one β-galactose and two α-rhamnose units (Table 5). The β-galactosyl residue was characterized by *J* values of H-1'' (δ 5.53, d, *J*_{H-1/H-2} = 7.8 Hz) and H-3'' (δ 3.62, dd, *J*_{H-3/H-2} = 9.0 Hz, *J*_{H-3/H-4} = 3.5 Hz). The two rhamnosyl residues were identified from the vicinal coupling constants of H-2'''/H-2'''' (dd, ³*J*_{H-2/H-3} = 3.2 Hz, ³*J*_{H-2/H-1} = br s), and the α configuration was established on the basis of the ¹³C-NMR shifts of C-3'''/C-3'''' (δ 68.0 and 70.5) and C-5'''/C-5'''' (δ 65.5 and 67.8) (Agrawal et al., 1985). The linkage position of sugar residues was established by HMBC and 2D ROESY data (Figs. S41 and S45). HMBC correlations between H-1'' (δ 5.53) and C-3 (δ 135.0) indicated that the galactose unit was attached to the hydroxyl group at C-3 of the flavonoid. Likewise, the HMBC correlation between H-1''' (δ 5.12) and C-2'' (δ 75.1) established the fragment of α-rhamnopyranosyl-(1 → 2)-β-galactopyranoside. The downfield shift of H-4''' (δ 4.83) and its HMBC correlation with a carboxyl carbon (δ 165.7) revealed that the Rha I unit was esterified at C-4''' by a *trans-p*-coumaroyl moiety. The attachment of the second α-rhamnose was established by a NOESY contact between the methylenic proton of CH₂-6b'' of Gal (δ 3.25, m) and H-1'''' (δ 4.40, br s) of Rha II. Hence, compound

11 was identified as quercetin 3-*O*-(4-*O-trans-p*-coumaroyl)-α-L-rhamnopyranosyl-(1 → 2)-[α-L-rhamnopyranosyl-(1 → 6)]-β-D-galactopyranoside.

Phenolic compounds **12–21** were identified on the basis of UV, ESIMS, and NMR data comparison as: 3',5'-di-*C*-β-glucosylphloracetone (**12**) (Ogawa et al., 2001), quercetin 3-*O*-β-xylopyranosyl-(1 → 2)-α-rhamnopyranoside 4'-*O*-α-rhamnopyranoside (**13**) (Nawwar et al., 1984), epigallocatechin (**14**) (Liao et al., 2014), gallicocatechin (**15**) (Nomizu et al., 2008), quercetin 3-*O*-α-rhamnopyranosyl-(1 → 6)-α-rhamnopyranosyl-(1 → 2)-β-galactopyranoside (**16**) (Yasukawa et al., 1989), prodelpinidin (**17**) (Fujii et al., 2013), quercetin (**18**) (Shul'ts et al., 2012), kaempferol 3-*O*-robinobioside (**19**) (Pawlowska et al., 2009), kaempferol 3-*O*-rutinoside (**20**) (Kazuma et al., 2003), quercetin 3-*O*-α-arabinosyl-(1 → 2)-α-rhamnoside (**21**) (Nielsen et al., 2005).

3. Conclusions

A phytochemical profiling of *Z. spina-christi* leaves led to the characterization of 10 dammarane-type saponins and 12 known polyphenols. Eight saponins (**3–10**) are reported here for the first time. For the three previously undescribed aglycons in saponins **8–10**, a possible biosynthetic pathway leading from **8** and **9** to the unprecedented *bis*-tetrahydrofuran moiety in **10** is proposed.

Table 4
¹H and ¹³C spectroscopic data for the sugar moieties of **7–10** (DMSO-*d*₆) (δ in ppm, *J* in Hz).

Sugar	No	7		8		9		10	
		δ _H ^c	δ _C , type	δ _H ^c	δ _C ^d , type	δ _H ^c	δ _C ^d , type	δ _H ^c	δ _C ^d , type
Glc ^a	1'	4.23, d (7.4)	103.9, CH	4.25, d (7.4)	103.5, CH	4.25, d (7.7)	103.3, CH	4.25, d (7.4)	103.5, CH
	2'	3.24, dd (9.0, 7.4)	76.6, CH	3.28, dd (9.0, 7.4)	75.0, CH	3.27, dd (9.0, 7.7)	74.9, CH	3.28, dd (9.0, 7.4)	75.0, CH
	3'	3.31, dd (9.0, 9.0)	78.2, CH	3.34, dd (9.5, 9.0)	77.9, CH	3.33, dd (9.5, 9.0)	78.0, CH	3.34, dd (9.0, 9.0)	78.3, CH
	4'	3.09	70.5, CH	3.09	70.2, CH	3.09	70.1, CH	3.07	70.2, CH
	5'	3.09	76.1, CH	3.09	76.2, CH	3.09	76.0, CH	3.09	76.2, CH
	6a'	3.64, br d (11.0)	61.1, CH ₂	3.65, br d (11.0)	60.8, CH ₂	3.65, br d (11.0)	60.5, CH ₂	3.66, br d (11.0)	61.1, CH ₂
Rhal ^b	6b'	3.47, dd (11.0, 5.0)		3.44, dd (11.0, 5.0)		3.45, dd (11.0, 5.0)		3.45, dd (11.0, 5.0)	
	1''	5.24, br s	99.8, CH	5.34, br s	98.7, CH	5.33, brs	98.4, CH	5.33, br s	98.7, CH
	2''	3.73, br d (3.6)	70.7, CH	3.62, br d (3.6)	70.5, CH	3.64, dd (3.6)	70.3, CH	3.63, br d (3.6)	70.7, CH
	3''	3.49, dd (9.0, 3.6)	71.2, CH	3.68, dd (9.0, 3.6)	70.7, CH	3.68, dd (9.0, 3.6)	70.6, CH	3.68, dd (9.0, 3.6)	71.1, CH
	4''	3.19, dd (9.0, 9.0)	78.4, CH	3.34, dd (9.0, 9.0)	78.0, CH	3.33, dd (9.0, 9.0)	78.0, CH	3.34, dd (9.0, 9.0)	78.2, CH
	5''	3.81, dq (9.0, 6.0)	65.9, CH	3.97, dq (9.0, 6.0)	65.6, CH	3.96, dq (9.0, 6.0)	65.5, CH	3.98, dq (9.0, 6.0)	65.9, CH
Rhall ^b	6''	1.07, d (6.0)	17.9, CH ₃	1.11, d (6.0)	17.7, CH ₃	1.10, d (6.0)	17.8, CH ₃	1.10, d (6.0)	17.6, CH ₃
	1'''	–		5.06, br s	100.5, CH	5.06, br s	100.4, CH	5.06, br s	100.8, CH
	2'''	–		3.69, br d (3.5)	70.4, CH	3.71, br d (3.5)	70.3, CH	3.71, br d (3.5)	70.6, CH
	3'''	–		3.39, dd (9.0, 3.5)	70.4, CH	3.38, dd (9.0, 3.5)	70.4, CH	3.38, dd (9.0, 3.5)	70.5, CH
	4'''	–		3.19, dd (9.0, 9.0)	71.9, CH	3.19, dd (9.0, 9.0)	71.6, CH	3.19, dd (9.0, 9.0)	72.0, CH
	5'''	–		3.49, dq (9.0, 6.0)	68.3, CH	3.48, dq (9.0, 6.0)	68.2, CH	3.49, dq (9.0, 6.0)	68.8, CH
6'''	–		1.11, d (6.0)	17.3, CH ₃	1.10, d (6.0)	17.4, CH ₃	1.10, d (6.0)	17.6, CH ₃	

Spin systems of sugars were resolved by 1D selective TOCSY and 2D HSQC-TOCSY.

^a Glc = β-D-glucopyranosyl.

^b Rha = α-L-rhamnopyranosyl.

^c Overlapped ¹H-NMR signals are reported without multiplicities.

^d ¹³C shifts were extracted from HSQC and HMBC data.

Semiquantitative analysis of saponin profiles in four geographically distinct leaf samples showed that lotogenin glycosides were the major saponins and occurred in all samples analyzed, while glycosides of the other aglycons were more restricted in occurrence.

4. Experimental section

4.1. Solvents and chemicals

For extraction and column chromatography technical grade solvents (Scharlau, Spain) were used after distillation, and HPLC grade solvents (Macron, Norway) were used for HPLC. HPLC grade water was obtained from an EASY-pure II water purification system. DMSO was from Scharlau (Spain). Deuterated solvents were purchased from Armar Chemicals (Switzerland). Sephadex LH-20 was purchased from GE Healthcare. Silica gel 60 (40–63 μm, 70–230 mesh ASTM) was from Merck (Germany).

4.2. Plant material

Z. spina-christi leaves were originating from the province of Khuzestan, Dezful County, Iran. Leaves were collected in February 2014 by Mr. Hamid Ranjbar, and identified by Dr. Ali Sonboli, Department of Biology, Medicinal Plants and Drugs Research Institute, Shahid Beheshti University, Tehran, Iran. A voucher specimen (MPH-2409) has been deposited at the Herbarium of the Medicinal Plants and Drugs Research Institute, Shahid Beheshti University, Tehran. The samples originating from Caesarea (Israel), Nurabad (Iran), and a sample purchased from a bazar in Tehran (Iran) were provided by Mibelle AG, Buchs, Switzerland.

4.3. NMR spectroscopy

NMR spectra were recorded on a 500 MHz Avance III™ spectrometer (Bruker BioSpin) equipped with a 1 mm TXI microprobe (¹H-NMR) or a 5 mm BBO probe (¹³C-NMR) at 18.0 and 23.0 °C, respectively. Standard pulse sequences of the software package Topspin 3.0 were used.

4.4. HPLC analysis

HPLC profiling of extracts was carried out on a HPLC system (Shimadzu, USA) (system 1) consisting of a degasser, binary high pressure mixing pump, column thermostat, PDA detector (SDP-M20A) with thermostatted UV cell, and a triple quadrupole MS with ESI interface (ESIMS) (LCMS-8030) coupled via T split. Both positive and negative ion HPLC-ESIMS spectra were simultaneously recorded in the range of *m/z* 160 to 1500; capillary voltage: 4.5 V; scan speed: 6000 u/sec; event time: 0.150 s. The second line of the T-splitter was connected to an Alltech 3300 evaporative light scattering detector. ELSD conditions were as follows: nitrogen flow: 2.5 l/min, temperature: 55 °C, detector gain: 8. Data acquisition and analysis were performed with LabSolutions software (Shimadzu, USA).

High-resolution spectra were recorded with a microTOF (Bruker Daltonic, Germany) mass spectrometer with ESI interface (Bruker) connected to an 1100 series HPLC (Agilent, USA) (system 2). Spectra were recorded in both positive and negative mode in the range of *m/z* 160 to 1500; capillary voltage: 4.0 V in positive and 3.0 V in negative mode. For calibration, a reference solution of sodium formate 0.1% (m/v) in *iso*-PrOH/H₂O (1:1) containing 5 mM NaOH was used. Data acquisition and processing were performed using HyStar 3.0 software (Bruker).

The *n*-BuOH fraction was dissolved in MeOH at a concentration of 10 mg/ml and filtered. Separation conditions for the *n*-BuOH fraction were: SunFire C₁₈ column (3.5 μm, 150 × 3.0 mm I.D.) (Waters, Ireland) equipped with a guard column (20.0 × 3.0 mm I.D.); mobile phase A: H₂O with 0.1% (v/v) formic acid, mobile phase B: MeCN with 0.1% (v/v) formic acid; gradient profile: 15% B isocratic for 2 min, 15–70% B in 18 min, 100% B in 5 min; flow rate: 0.4 ml/min; column temperature: 25.0 °C; sample injection volume: 6 μl.

4.5. Extraction and isolation

Dried leaves of *Z. spina-christi* (900 g) were ground with a RLB mill (Rotor AG, Switzerland) and extracted by maceration with

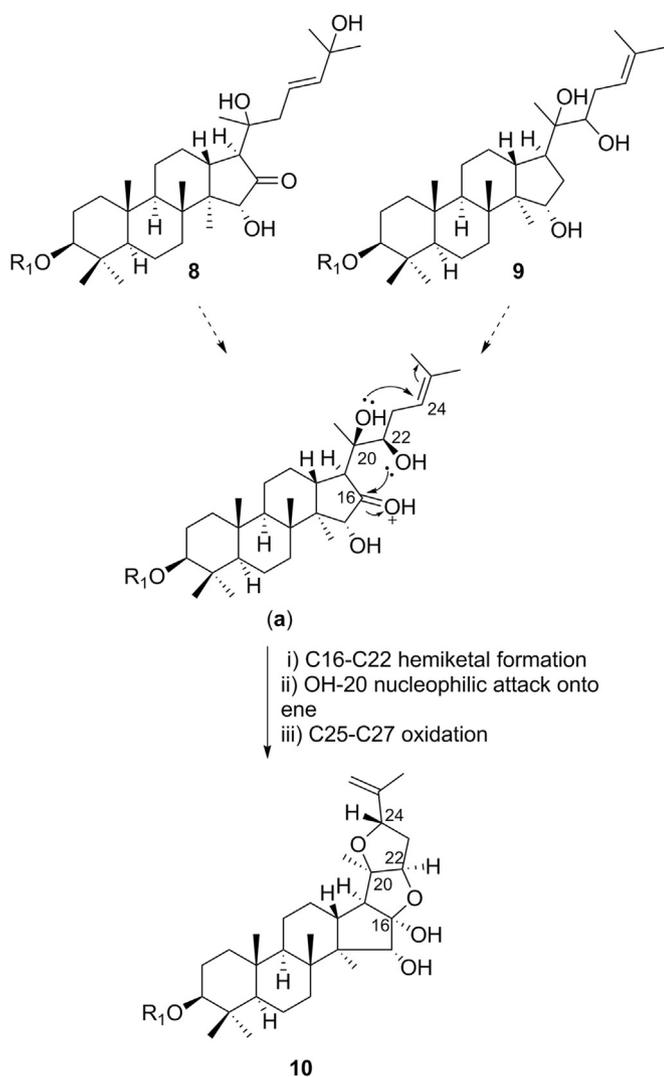


Fig. 4. Proposed biosynthesis of aglycon in saponin **10**.

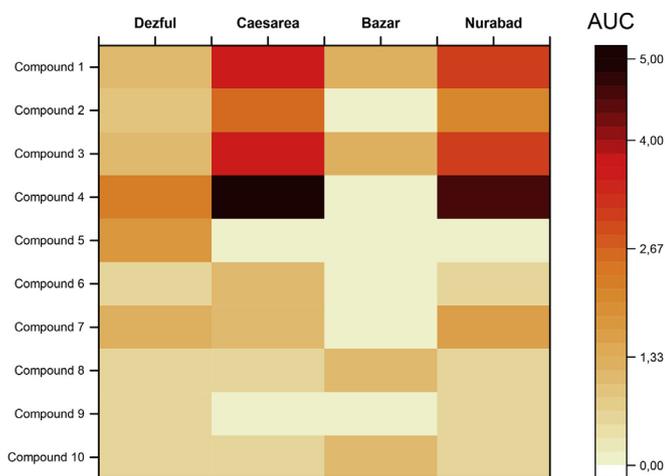


Fig. 5. 2D heat map representing semi-quantitative distribution of saponins **1–10** in leaf samples of four different origins.

EtOAc (2 × 2 l), followed by 35% EtOH v/v (3 × 4 l). After evaporation 30 g of EtOAc and 119 g of 35% ethanolic extracts were obtained. For the ethanolic extract, fifteen aliquots of 7 g each were

suspended in distilled water (600 ml), ultrasonicated for 15 min, and partitioned with *n*-BuOH (3 × 200 ml). After evaporation a total of 64 g of aqueous and 24 g of *n*-BuOH fraction were obtained (Fig. S68).

A portion of the *n*-BuOH fraction (12 g) was separated on a Sephadex LH-20 column (100 × 7 cm, I.D.) eluted with MeOH at a flow rate of 1.5 ml/min using the B-688 pump (Büchi, Switzerland). Fractions were collected with a SuperFrac fraction collector (Pharmacia Biotech, Sweden). A total of 11 fractions (A-K) were combined on the basis of TLC patterns. The saponin containing fraction A (2.9 g) was further separated on an open column (75 × 4 cm, I.D.) packed with silica gel 60 (40–63 μm, 300 g). Separation was performed with a step gradient of CHCl₃-MeOH-H₂O 80:20:2 (1 l), 75:30:3 (1 l), and 65:35:3.5 (1 l) at a flow rate of 5 ml/min. The eluate was combined in fractions A1–A13 based on TLC pattern. Fractions A2 (15.7 mg), A3 (44.1 mg), A5 (38.4 mg), A6 (23.3 mg), A7 (98.2 mg), A9 (72.8 mg), A11 (59.7 mg), A12 (45.5 mg) were selected for final purification of compounds. This step was achieved by HPLC utilizing an Alliance 2690 instrument (Waters, USA) coupled via a T-splitter to an Alltech 2000ES (USA) evaporative light scattering detector (system 3). Separations were carried out at 25 °C on a SunFire C₁₈ column (5 μm, 150 × 10 mm, I.D.) (Waters, Ireland) equipped with a pre-column (10 × 10 mm, I.D.). Gradients of H₂O/MeCN were used. The flow-rate was 4 ml/min. Finally, we obtained compound **1** (3.2 mg) and compound **2** (5.1 mg) from A3; compound **3** (2.8 mg) from A2; compounds **4–5** (2.1 mg) from A6; compound **6** (2.2 mg) from A7; compound **7** (2.8 mg) from A5; compound **8** (1.7 mg) from A9; compound **9** (1.7 mg) from A11–A12; compound **10** (1.3 mg) from A11. Sephadex fractions B–K contained polyphenolic compounds. Compound **11** (4.7 mg) from fraction D (300.1 mg), **12** (4.3 mg) and **13** (6.2 mg) from fraction E (273.4 mg), and **14** (14.8 mg) and **15** (12.7 mg) from fraction H (283.6 mg) were isolated by means of semi-preparative HPLC. Compounds **16–22** were identified in fractions B–C, F–G, I–K comparing their UV spectra and MS spectroscopic data with those available in literature.

4.5.1. 22 α -Acetoxy-christinin A (**3**)

Amorphous white solid; $[\alpha]_{\text{D}}^{25}$ -27.9 (c 0.05, MeOH); for ¹³C NMR and ¹H NMR spectroscopic data, see Table 1 and Table 3; (-)-ESIMS *m/z* 969; (-)-HRESIMS *m/z* 969.5150 [M-H]⁻ (calcd for C₄₉H₇₇O₁₉: 969.5137).

4.5.2. Christinin A1 (**4**); christinin A2 (**5**)

Amorphous white solid; for ¹³C NMR and ¹H NMR spectroscopic data, see Tables 1 and 3; (-)-ESIMS *m/z* 953; (-)-HRESIMS *m/z* 953.5161 [M-H]⁻ (calcd for C₄₉H₇₇O₁₈: 953.5188) (**4**). ESIMS *m/z* 997; (-)-HRESIMS *m/z* 997.5072 [M-H]⁻ (calcd for C₅₀H₇₇O₂₀: 997.5086) (**5**).

4.5.3. Lotoside III (**6**)

Amorphous white solid; $[\alpha]_{\text{D}}^{25}$ -50.4 (c 0.12, MeOH); for ¹³C NMR and ¹H NMR spectroscopic data, see Tables 1 and 3; (-)-ESIMS *m/z* 943; (-)-HRESIMS *m/z* 943.5363 [M-H]⁻ (calcd for C₄₈H₇₉O₁₈: 943.5345).

4.5.4. 15-Acetoxy-lotoside IV (**7**)

Amorphous white solid; $[\alpha]_{\text{D}}^{25}$ -36.2 (c 0.31, MeOH); for ¹³C NMR and ¹H NMR spectroscopic data, see Tables 2 and 4; (-)-ESIMS *m/z* 839; (-)-HRESIMS *m/z* 839.4849 [M-H]⁻ (calcd for C₄₄H₇₁O₁₅: 839.4871).

4.5.5. Sidrigenin 3-O- α -L-rhamnopyranosyl-(1 \rightarrow 4)- α -L-rhamnopyranosyl-(1 \rightarrow 2)- β -D-glucopyranoside (**8**)

Amorphous white solid; $[\alpha]_{\text{D}}^{25}$ -76.8 (c 0.14, MeOH); for ¹³C NMR and ¹H NMR spectroscopic data, see Tables 2 and 4; (-)-ESIMS

Table 5
¹H and ¹³C spectroscopic data of **11** (DMSO-*d*₆) (δ in ppm, *J* in Hz).

No.	aglycon				sugars	
	δ _H	δ _C ^c , type			δ _H ^e	δ _C ^c , type
2	—	156.1, C	Gal ^a	1''	5.53, d (7.8)	99.1, CH
3	—	135.0, C		2''	3.86, dd (9.0, 7.5)	75.1, CH
4	—	d		3''	3.62, dd (9.0, 3.5)	73.5, CH
5	—	d		4''	3.62, d (3.7)	68.3, CH
6	—	d		5''	3.57	73.1, CH
7	6.19, br s	98.7, CH		6a''	3.59	65.0, CH ₂
8	—	165.0, C		6b''	3.25	
9	6.37, br s	93.4, CH				
10	—	100.5, C	Rha I ^b	1'''	5.12, br s	100.0, CH
1'	—	d		2'''	3.85, dd (3.2, br s)	70.7, CH
2'	7.50, d (2.0)	115.5, CH		3'''	3.87, dd (9.0, 3.2)	68.0, CH
3'	—	145.4, C		4'''	4.83, dd (9.0, 9.0)	73.6, CH
4'	—	147.8, C		5'''	4.14, dq (9.0, 6.0)	65.5, CH
5'	6.80, d (8.5)	115.0, CH		6'''	0.73, d (6.0)	16.5, CH ₃
6'	7.62, dd (8.5, 2.0)	121.2, CH				
1''''	—	125.5, C	Rha II ^b	1''''	4.40, br s	99.8, CH
2''''	7.36, d (8.0)	130.0, CH		2''''	3.42, dd (3.2, br s)	70.1, CH
3''''	6.79, d (8.0)	115.3, CH		3''''	3.32, dd (9.0, 3.2)	70.5, CH
4''''	—	159.5, C		4''''	3.10, dd (9.0, 9.0)	71.5, CH
5''''	6.79, d (8.0)	115.3, CH		5''''	3.37, dq (9.0, 6.0)	67.8, CH
6''''	7.36, d (8.0)	130.0, CH		6''''	1.06, d (6.0)	17.5, CH ₃
7''''	7.35, d (16.0)	143.8, CH				
8''''	6.23, d (16.0)	114.0, CH				
9''''	—	165.7, C				

Spin systems of sugars were resolved by 1D selective TOCSY and 2D HSQC-TOCSY.

^a Gal = β-D-galactopyranosyl.^b Rha = α-L-rhamnopyranosyl.^c ¹³C shifts were extracted from HSQC and HMBC data.^d Not detected.^e Overlapped ¹H-NMR signals are reported without multiplicities.

m/z 943; (-)-HRESIMS *m/z* 943.5339 [M-H]⁻ (calcd for C₄₈H₇₉O₁₈: 943.5345).

4.5.6. Konarigenin 3-O-α-L-rhamnopyranosyl-(1 → 4)-α-L-rhamnopyranosyl-(1 → 2)-β-D-glucopyranoside (**9**)

Amorphous white solid; [α]²⁵_{Na} -30.4 (c 0.16, MeOH); for ¹³C NMR and ¹H NMR spectroscopic data, see Tables 2 and 4; (-)-ESIMS *m/z* 929; (-)-HRESIMS *m/z* 929.5563 [M-H]⁻ (calcd for C₄₈H₈₁O₁₇: 929.5552).

4.5.7. Siconigenin 3-O-α-L-rhamnopyranosyl-(1 → 4)-α-L-rhamnopyranosyl-(1 → 2)-β-D-glucopyranoside (**10**)

Amorphous white solid; [α]²⁵_{Na} -30.7 (c 0.26, MeOH); for ¹³C NMR and ¹H NMR spectroscopic data, see Tables 2 and 4; (-)-ESIMS *m/z* 941; (-)-HRESIMS *m/z* 941.5167 [M-H]⁻ (calcd for C₄₈H₇₇O₁₈: 941.5188).

4.5.8. Quercetin 3-O-(4-O-trans-p-coumaroyl)-α-L-rhamnopyranosyl-(1 → 2)-[α-L-rhamnopyranosyl-(1 → 6)]-β-D-galactopyranoside (**11**)

Amorphous yellow solid; [α]²⁵_{Na} -57.2 (c 0.12, MeOH); UV (MeOH) λ_{max} (log ε) 273 (2.95), 312 (4.01) nm; ¹H NMR and ¹³C NMR data, see Table 5; (-)-ESIMS *m/z* 901; (-)-HRESIMS *m/z* 901.2469 [M-H]⁻ (calcd for C₄₂H₄₅O₂₂: 901.2481).

4.6. Acid hydrolysis and sugar analysis

Hydrolysis of sugars and GC-MS analysis of derivatives was performed according to Abbet et al. (2011). Compounds **1** (0.8 mg), **6** (1.5 mg), **9** (1.7 mg), and **11** (1.2 mg) were hydrolyzed separately. After heating at 100 °C for 1 h in 2 M TFA (1 ml), the mixture was extracted with CH₂Cl₂ (3 × 1.0 ml). The aq. phase was freeze-dried, and re-dissolved in dry pyridine (200 μl) containing 5 mg/ml L-

cysteine methyl ester hydrochloride. The reaction mixture was heated at 60 °C for 1 h, followed by silylation with hexamethyldisilazane and chlorotrimethylsilane (Fluka) in pyridine (3:1:10, 300 μl) at 60 °C for 30 min (Chai et al., 2007). After silylation, pyridine was evaporated, and the solid residue extracted with *n*-hexane. GC-MS analysis was performed on a 5890 Series II gas chromatograph coupled to a HP 5971A mass detector (Hewlett Packard, USA). The separation was carried out on a DB-225 MS column (30 m × 0.25 mm, I.D., Waters, USA); column temp. 150 °C for 2 min, then gradient of 58 °C/min to 210 °C, then 10 °C/min to 240 °C. Comparison of the retention times of derivatized reference sugars with those obtained from samples resulted in L-arabinose (Rt 24.48 min), D-glucose (Rt 28.64 min), D-xylose (Rt 24.47 min), and D-fucose (Rt 25.96 min) in compound **1**, and D-glucose (Rt 28.64 min) and L-rhamnose (Rt 25.66 min) in compound **6**. Sugars in **9** and **11** were analyzed at a different time and were identified as D-glucose and L-rhamnose (**9**), and L-rhamnose (24.49 min) and D-galactose (26.04 min) (**11**). Derivatives of L-arabinose (Rt 24.48 min) and D-xylose (Rt 24.47 min) had very similar retention times resulting in peak overlapping. Confirmation of L-arabinose and D-xylose in **1** was obtained in a subsequent spiking experiment with derivatized reference sugars leading to unambiguous separation (D-arabinose, Rt 24.92 min); L-xylose, Rt 24.84 min) (Figs. S47–S66).

Acknowledgements

The authors are grateful to Orlando Fertig, Division of Pharmaceutical Biology, University of Basel, for performing the sugar analysis, to Dr. Samad Ebrahimi, for provision of the *Z. spina-christi* leaf sample from Dezfoul, and to Dr. Frank Gafner, Mibelle Biochemistry, for provision of leaf samples from Caesarea, Nurabad, and from the bazar in Tehran.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.phytochem.2017.02.028>.

References

- Abbet, C., Neuburger, M., Wagner, T., Quitschau, M., Hamburger, M., Potterat, O., 2011. Phyteumosides A and B: new saponins with unique triterpenoid aglycons from *Phyteuma orbiculare* L. *Org. Lett.* 13, 1354–1357.
- Abdel-Zaher, A.O., Salim, S.Y., Assaf, M.H., Abdel-Hady, R.H., 2005. Antidiabetic activity and toxicity of *Zizyphus spina-christi* leaves. *J. Ethnopharmacol.* 101, 129–138.
- Agrawal, P.K., Jain, D.C., Gupta, R.K., Thakur, R.S., 1985. Carbon-13 NMR spectroscopy of steroidal saponinins and steroidal saponins. *Phytochemistry* 24, 2479–2496.
- Chai, X.Y., Xu, Z.R., Ren, H.Y., Shi, H.M., Lu, Y.M., Li, F.F., Tu, P.F., 2007. Itosides A – I, new phenolic glycosides from *Itoa orientalis*. *Helv. Chim. Acta* 90, 2176–2185.
- Fujii, W., Toda, K., Matsumoto, K., Kawaguchi, K., Kawahara, S., Hattori, Y., Fujii, H., Makabe, H., 2013. Syntheses of prodelphinidin B1, B2, and B4 and their anti-tumor activities against human PC-3 prostate cancer cell lines. *Tetrahedron Lett.* 54, 7188–7192.
- Gao, Q.H., Wu, C.S., Wang, M., 2013. The jujube (*Zizyphus jujuba* Mill.) fruit: a review of current knowledge of fruit composition and health benefits. *J. Agric. Food Chem.* 61, 3351–3363.
- Ihre, H., Hult, A., Fréchet, J.M.J., Gitsov, I., 1998. Double-stage convergent approach for the synthesis of functionalized dendritic aliphatic polyesters based on 2,2-bis(hydroxymethyl)propionic acid. *Macromolecules* 31, 4061–4068.
- Kazuma, K., Noda, N., Suzuki, M., 2003. Malonylated flavonol glycosides from the petals of *Clitoria ternatea*. *Phytochemistry* 62, 229–237.
- Liao, C.R., Kuo, Y.H., Ho, Y.L., Wang, C.Y., Yang, C., Lin, C.W., Chang, Y.S., 2014. Studies on cytotoxic constituents from the leaves of *Elaeagnus oldhamii* Maxim. in non-small cell lung cancer A549 cells. *Molecules* 19, 9515.
- Mahran, G., Glombitza, K.W., Mirhom, Y.W., Hartmann, R., Michel, C.G., 1996. Novel saponins from *Zizyphus spina-christi* growing in Egypt. *Planta Med.* 62, 163–165.
- Matsuda, H., Murakami, T., Ikebata, A., Yamahara, J., Yoshikawa, M., 1999. Structure elucidation and immunological adjuvant activity of novel protojubilogenin type triterpene bisdesmosides, protojubilosides A, B, and B₁, from the seeds of *Zizyphus jujuba* var. *spinosa* (Zizyphi Spinosi Semen). *Chem. Pharm. Bull.* 47, 1744–1748.
- Nawwar, M.A.M., Ishak, M.S., Michael, H.N., Buddrust, J., 1984. Leaf flavonoids of *Zizyphus spina-christi*. *Phytochemistry* 23, 2110–2111.
- Nielsen, A.H., Olsen, C.E., Møller, B.L., 2005. Flavonoids in flowers of 16 *Kalanchoë blossfeldiana* varieties. *Phytochemistry* 66, 2829–2835.
- Nomizu, K., Hashida, K., Makino, R., Ohara, S., 2008. Antioxidants from steamed used tea leaves and their reaction behavior. *Biosci. Biotech. Biochem.* 72, 1682–1689.
- Ogawa, K., Kawasaki, A., Omura, M., Yoshida, T., Ikoma, Y., Yano, M., 2001. 3:5-Di-C-β-glucopyranosylphloretin, a flavonoid characteristic of the genus *Fortunella*. *Phytochemistry* 57, 737–742.
- Pawłowska, A.M., Camangi, F., Bader, A., Braca, A., 2009. Flavonoids of *Zizyphus jujuba* L. and *Zizyphus spina-christi* L. Willd. (Rhamnaceae) fruits. *Food Chem.* 112, 858–862.
- Renault, J.H., Ghedira, K., Thepenier, P., Lavaud, C., Zeches-Hanrot, M., Le Men-Olivier, L., 1997. Dammarane saponins from *Zizyphus lotus*. *Phytochemistry* 44, 1321–1327.
- Shul'ts, E.E., Ganbaatar, Z., Petrova, T.N., Shakirov, M.M., Bagryanskaya, I.Y., Taraskin, V.V., Radnaeva, L.D., Otgonsuren, D., Pokrovskii, A.G., Tolstikov, G.A., 2012. Plant coumarins. IX. * Phenolic compounds of *Ferulopsis hystrix* growing in Mongolia. Cytotoxic activity of 8,9-dihydrofurocoumarins. *Chem. Nat. Compd.* 48, 211–217.
- Yasukawa, K., Sekine, H., Takido, M., 1989. Two flavonol glycosides from *Lysimachia fortunei*. *Phytochemistry* 28, 2215–2216.

Supporting information

Dammarane-type saponins from leaves of *Ziziphus spina-christi*

Alen Bozicevic, Maria De Mieri, Angela Di Benedetto, Matthias Hamburger

Institute of Pharmaceutical Biology, University of Basel, Klingelbergstrasse 50, CH-4056 Basel, Switzerland

E-mail addresses:

Corresponding author: matthias.hamburger@unibas.ch

Co-authors: alen.bozicevic@unibas.ch, maria.demieri@unibas.ch,
angi.dibenedetto@stud.unibas.ch

Corresponding author:

Prof. Matthias Hamburger, Institute of Pharmaceutical Biology, University of Basel,
Klingelbergstrasse 50, CH-4056 Basel, Switzerland

Contents

Figure S1: ^1H -NMR spectrum of compound 3 (500 MHz, DMSO-d_6).....	5
Figure S2: ^1H - ^1H COSY spectrum of compound 3 (500 MHz, DMSO-d_6).....	5
Figure S3: Overlay of HSQC and HMBC spectra of compound 3 (500 and 125 MHz, DMSO-d_6).....	6
Figure S4: Overlay of HSQC and TOCSY spectra of compound 3 (500 and 125 MHz, DMSO-d_6).....	6
Figure S5: 2D ^1H - ^1H ROESY spectrum of compound 3 (500 and 125 MHz, DMSO-d_6).....	7
Figure S6: Selective 1D TOCSY spectrum of compound 3 (500 MHz, DMSO-d_6).....	7
Figure S7: ^1H -NMR spectrum of compounds 4-5 (500 MHz, DMSO-d_6).....	8
Figure S8: Overlay of HSQC and HMBC spectra of compound 4-5 (500 and 125 MHz, DMSO-d_6).....	8
Figure S9: DEPTq spectrum of compounds 4-5 (500 MHz, DMSO-d_6).....	9
Figure S10: ^1H -NMR spectrum of compound 6 (500 MHz, DMSO-d_6).....	9
Figure S11: ^1H - ^1H COSY spectrum of compound 6 (500 MHz, DMSO-d_6).....	10
Figure S12: Overlay of HSQC and HMBC spectra of compound 6 (500 and 125 MHz, DMSO-d_6).....	10
Figure S13: 2D ^1H - ^1H ROESY spectrum of compound 6 (500 MHz, DMSO-d_6).....	11
Figure S14: Overlay of HSQC and TOCSY spectra of compound 6 (500 and 125 MHz, DMSO-d_6).....	11
Figure S15: Selective 1D TOCSY spectrum of compound 6 (500 MHz, DMSO-d_6).....	12
Figure S16: DEPTq spectrum of compound 6 (125 MHz, DMSO-d_6).....	12
Figure S17: ^1H -NMR spectrum of compound 7 (500 MHz, DMSO-d_6).....	13
Figure S18: ^1H - ^1H COSY spectrum of compound 7 (500 MHz, DMSO-d_6).....	13
Figure S19: Overlay of HSQC and HMBC spectra of compound 7 (500 and 125 MHz, DMSO-d_6).....	14
Figure S20: 2D ^1H - ^1H ROESY spectrum of compound 7 (500 MHz, DMSO-d_6).....	14
Figure S21: Overlay of HSQC and TOCSY spectra of compound 7 (500 and 125 MHz, DMSO-d_6).....	15
Figure S22: Selective 1D TOCSY spectrum of compound 7 (500 MHz, DMSO-d_6).....	15
Figure S23: DEPTq spectrum of compound 7 (500 MHz, DMSO-d_6).....	16
Figure S24: ^1H -NMR spectrum of compound 8 (500 MHz, DMSO-d_6).....	16
Figure S25: ^1H - ^1H COSY spectrum of compound 8 (500 MHz, DMSO-d_6).....	17
Figure S26: Overlay of HSQC and HMBC spectra of compound 8 (500 and 125 MHz, DMSO-d_6).....	17
Figure S27: 2D ^1H - ^1H ROESY spectrum of compound 8 (500 MHz, DMSO-d_6).....	18
Figure S28: Overlay of HSQC and TOCSY spectra of compound 8 (500 and 125 MHz, DMSO-d_6).....	18
Figure S29: ^1H -NMR spectrum of compound 9 (500 MHz, DMSO-d_6).....	19
Figure S30: ^1H - ^1H COSY spectrum of compound 9 (500 MHz, DMSO-d_6).....	19
Figure S31: Overlay of HSQC and HMBC spectra of compound 9 (500 and 125 MHz, DMSO-d_6).....	20
Figure S32: 2D ^1H - ^1H ROESY spectrum of compound 9 (500 MHz, DMSO-d_6).....	20

Figure S33: Overlay of HSQC and TOCSY spectra of compound 9 (500 and 125 MHz, DMSO-d ₆).....	21
Figure S34: Selective 1D TOCSY spectrum of compound 9 (500 MHz, DMSO-d ₆).....	21
Figure S35: ¹ H-NMR spectrum of compound 10 (500 MHz, DMSO-d ₆).....	22
Figure S36: ¹ H- ¹ H COSY spectrum of compound 10 (500 MHz, DMSO-d ₆).....	22
Figure S37: Overlay of HSQC and HMBC spectra of compound 10 (500 and 125 MHz, DMSO-d ₆).....	23
Figure S38: 2D ¹ H- ¹ H ROESY spectrum of compound 10 (500 MHz, DMSO-d ₆).....	23
Figure S39: Overlay of HSQC and TOCSY spectra of compound 10 (500 and 125 MHz, DMSO-d ₆).....	24
Figure S40: 1D selective ROESY compound 10 (500 MHz, DMSO-d ₆).....	24
Figure S41: Key ¹ H- ¹ H ROESY and HMBC correlations of compound 11	25
Figure S42: ¹ H-NMR spectrum of compound 11 (500 MHz, DMSO-d ₆).....	26
Figure S43: ¹ H- ¹ H COSY spectrum of compound 11 (500 MHz, DMSO-d ₆).....	26
Figure S44: Overlay of HSQC and HMBC spectra of compound 11 (500 and 125 MHz, DMSO-d ₆).....	27
Figure S45: 2D ¹ H- ¹ H ROESY spectrum of compound 11 (500 MHz, DMSO-d ₆).....	27
Figure S46: Selective 1D TOCSY spectrum of compound 11 (500 MHz, DMSO-d ₆).....	28
Figure S47: GC-MS chromatogram of hydrolyzed sugars from compound 1	28
Figure S48: GC-MS chromatogram of hydrolyzed sugars from compound 6	29
Figure S49: GC-MS chromatogram of derivatized _D -arabinose.....	29
Figure S50: GC-MS chromatogram of derivatized _D -fucose.....	30
Figure S51: GC-MS chromatogram of derivatized _D -glucose.....	30
Figure S52: GC-MS chromatogram of derivatized _D -rhamnose.....	31
Figure S53: GC-MS chromatogram of derivatized _D -xylose.....	31
Figure S54: GC-MS chromatogram of derivatized _L -arabinose.	32
Figure S55: GC-MS chromatogram of derivatized _L -fucose.....	32
Figure S56: GC-MS chromatogram of derivatized _L -glucose.	33
Figure S57: GC-MS chromatogram of derivatized _L -rhamnose.	33
Figure S58: GC-MS chromatogram of derivatized _L -xylose.....	34
Figure S59: GC-MS chromatogram of hydrolyzed sugars from compound 9	34
Figure S60: GC-MS chromatogram of hydrolyzed sugars from compound 11	35
Figure S61: GC-MS chromatogram of derivatized _D -glucose.....	35
Figure S62: GC-MS chromatogram of derivatized _D -galactose.	36
Figure S63: GC-MS chromatogram of derivatized _D -rhamnose.....	36
Figure S64: GC-MS chromatogram of derivatized _L -glucose.	37
Figure S65: GC-MS chromatogram of derivatized _L -galactose.....	37

Figure S66: GC-MS chromatogram of derivatized L-rhamnose..... 38

Figure S67: HPLC-ESIMS traces recorded in positive and negative ion mode of *Ziziphus spina-christi* of four different origins. Peak numbering designates compounds **1-21**. 38

Figure S68: ¹H-NMR spectrum of water and *n*-BuOH phases (500 MHz, DMSO-d₆)..... 39

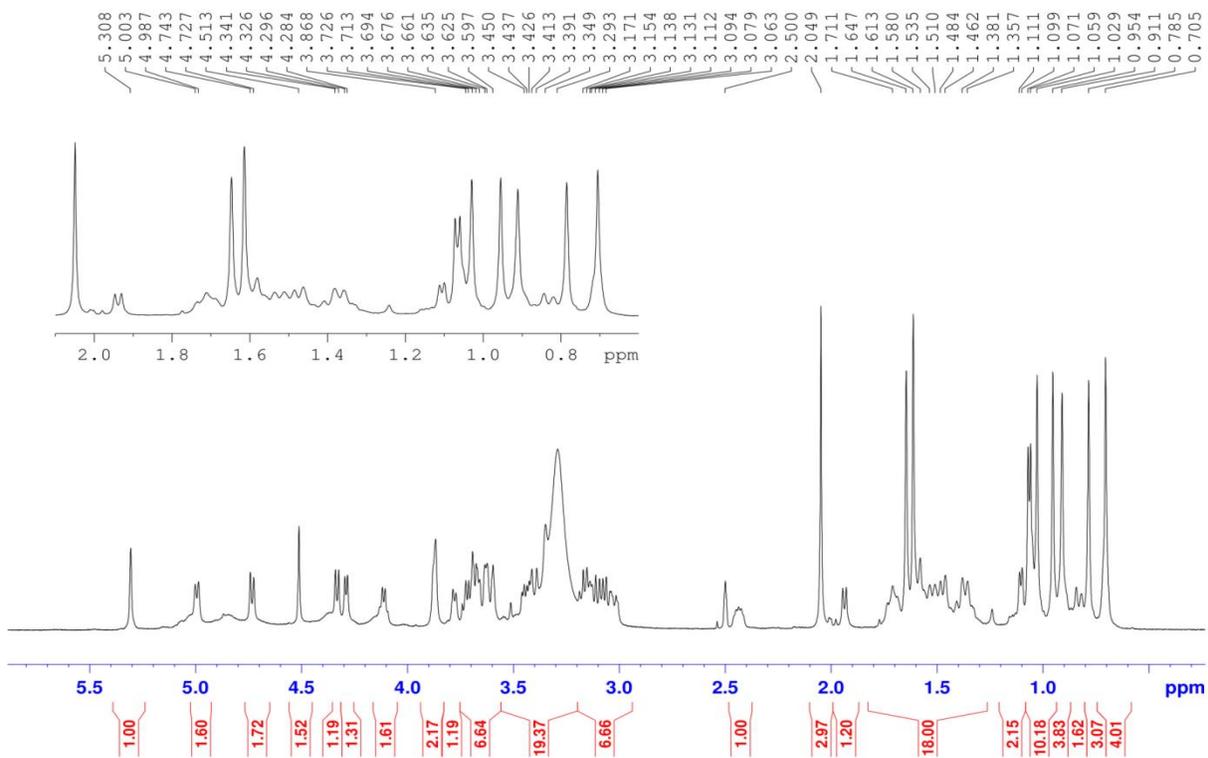


Figure S1: ^1H -NMR spectrum of compound **3** (500 MHz, DMSO-d_6).

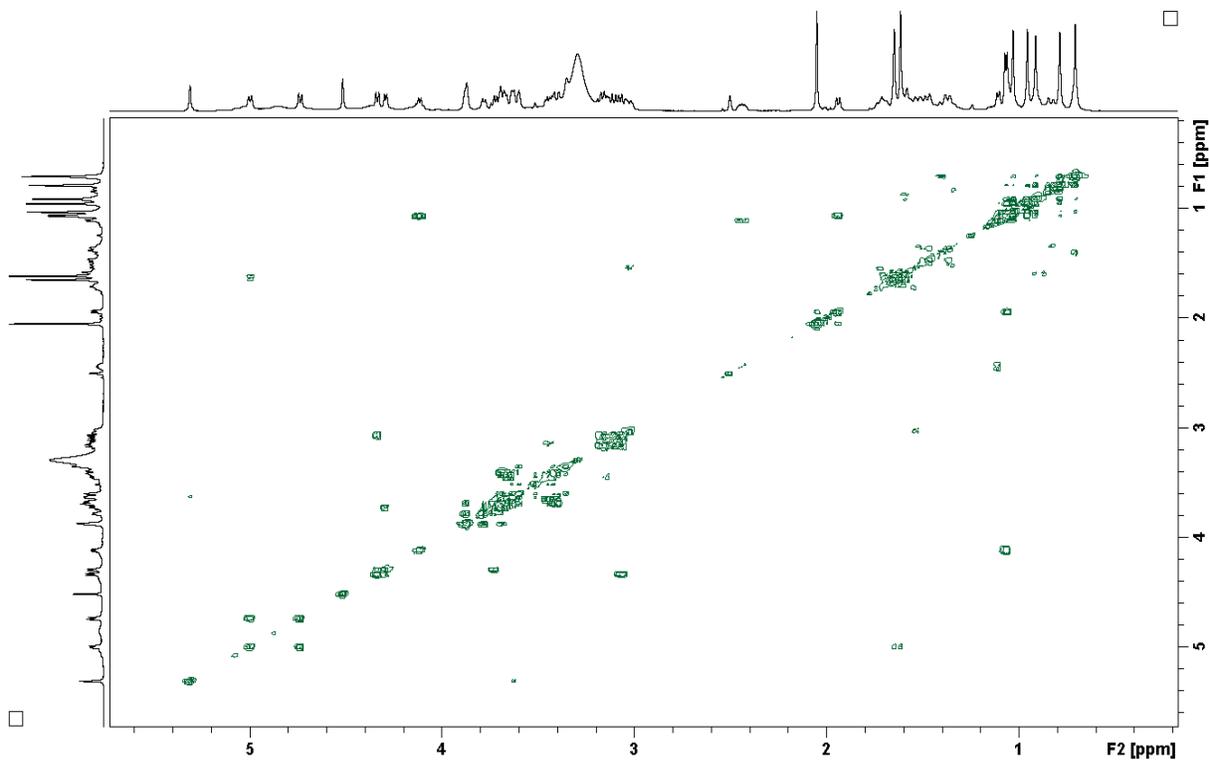


Figure S2: ^1H - ^1H COSY spectrum of compound **3** (500 MHz, DMSO-d_6).

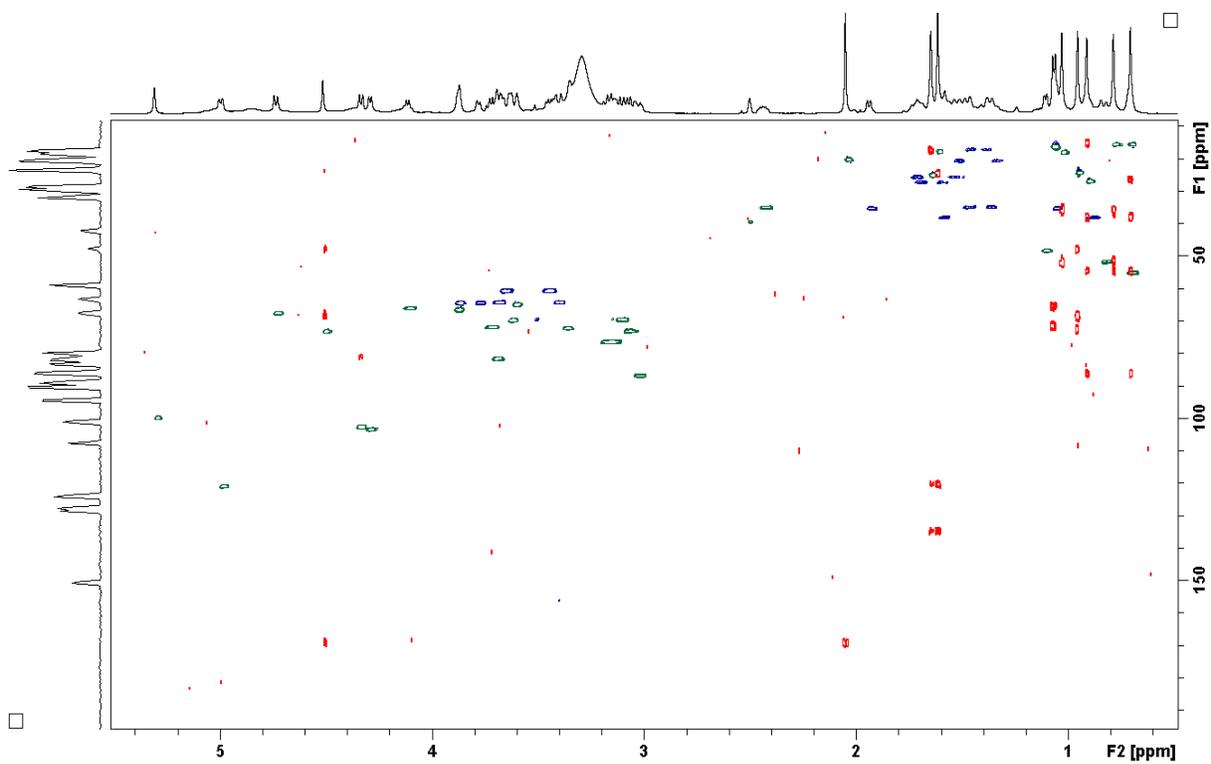


Figure S3: Overlay of HSQC (green-blue) and HMBC (red) spectra of compound **3** (500 and 125 MHz, DMSO-d₆).

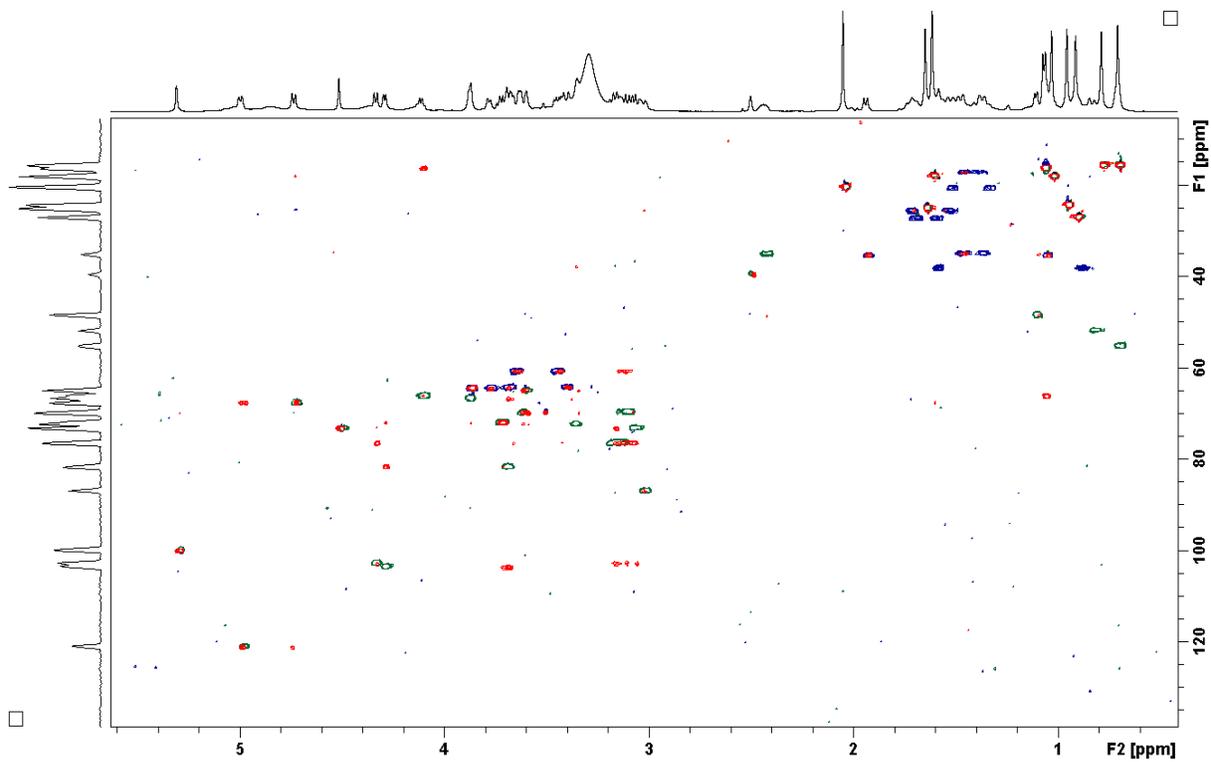


Figure S4: Overlay of HSQC and TOCSY spectra of compound **3** (500 and 125 MHz, DMSO-d₆).

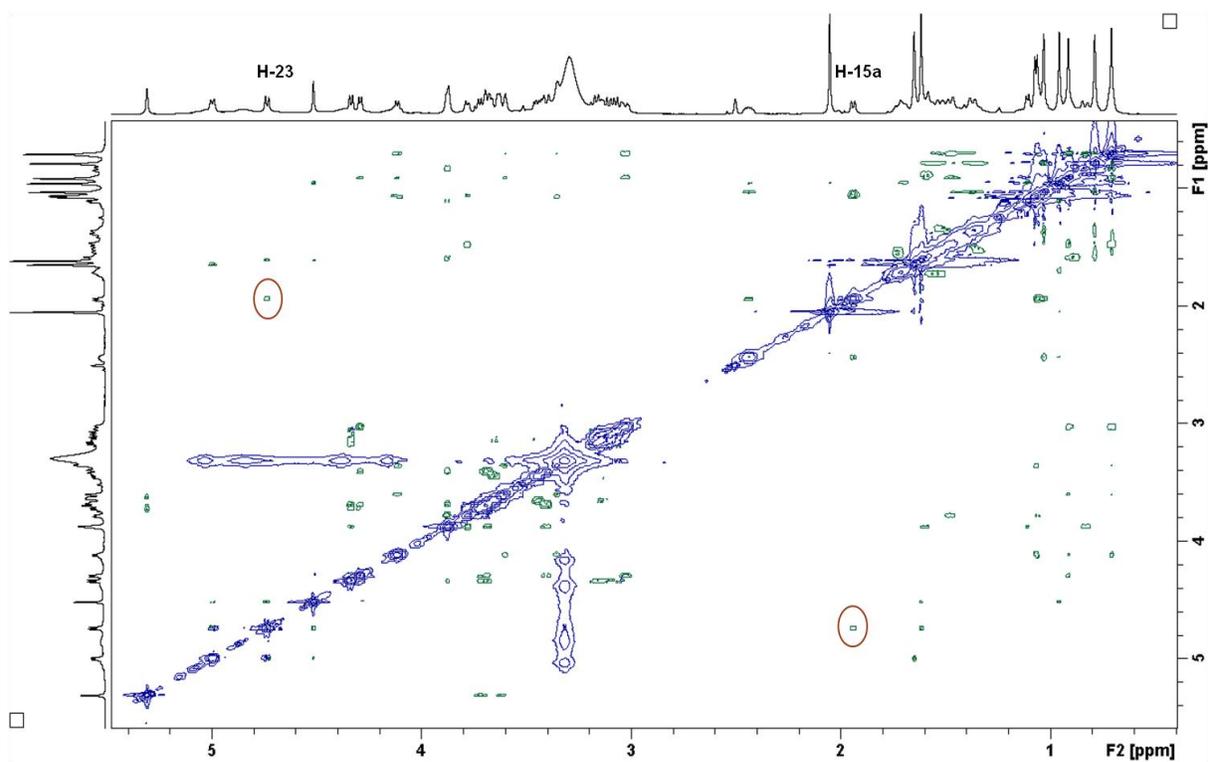


Figure S5: 2D ^1H - ^1H ROESY spectrum of compound **3** (500 and 125 MHz, DMSO-d_6).

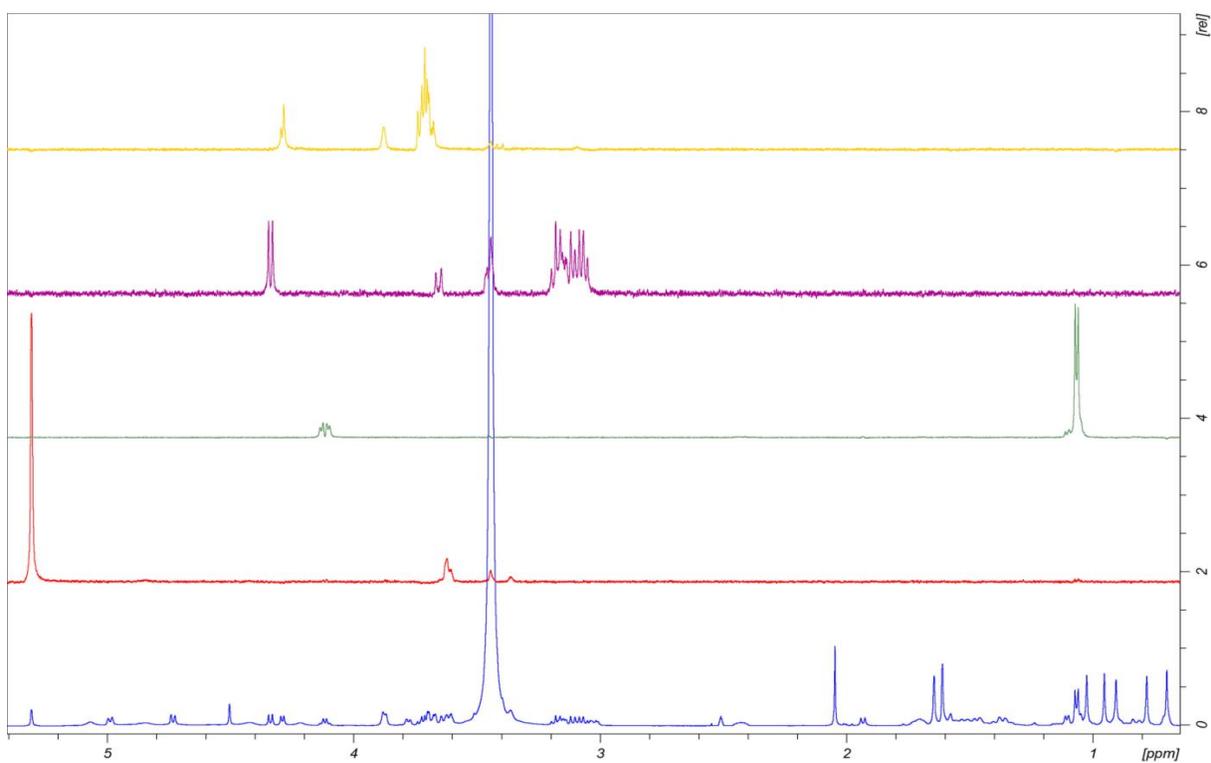


Figure S6: Selective 1D TOCSY spectrum of compound **3** (500 MHz, DMSO-d_6).

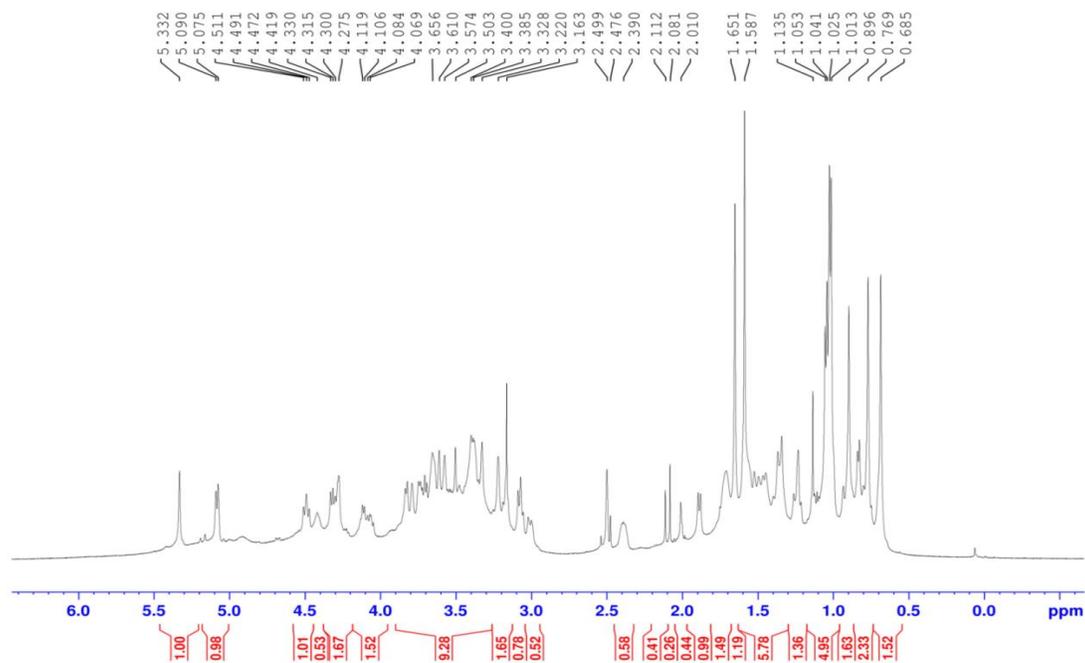


Figure S7: $^1\text{H-NMR}$ spectrum of compounds **4-5** (500 MHz, DMSO-d_6).

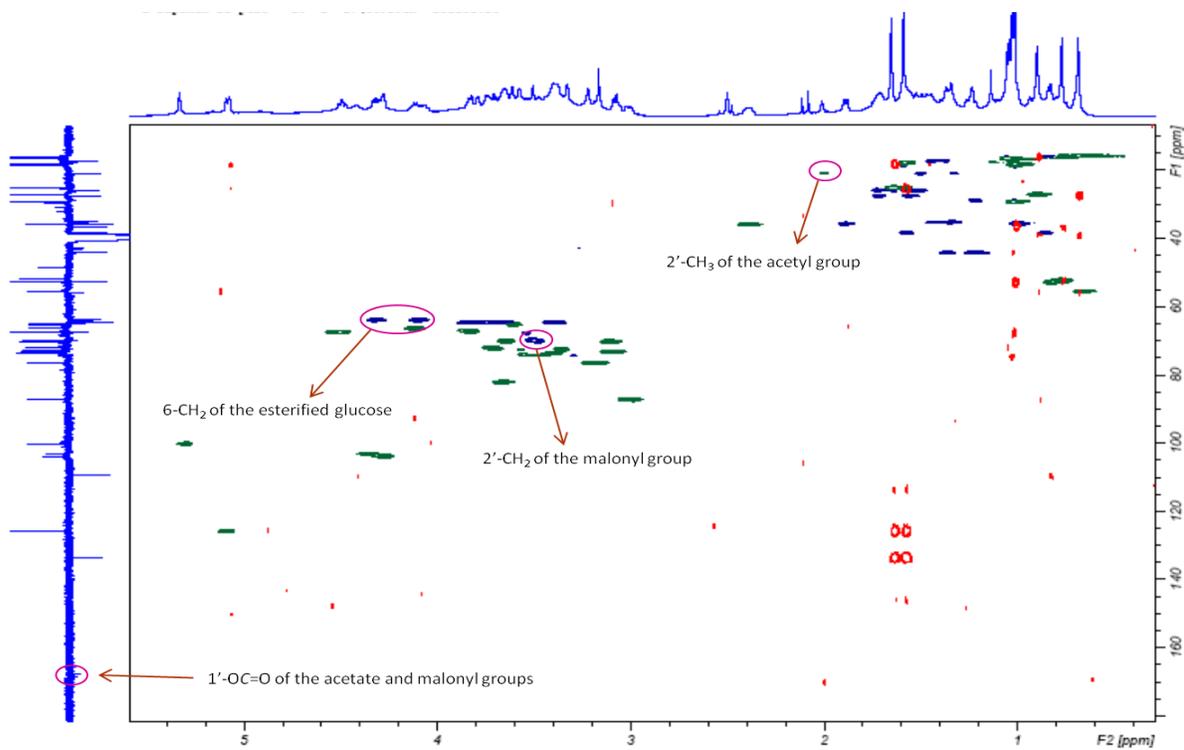


Figure S8: Overlay of HSQC (green-blue) and HMBC (red) spectra of compound **4-5** (500 and 125 MHz, DMSO-d_6).

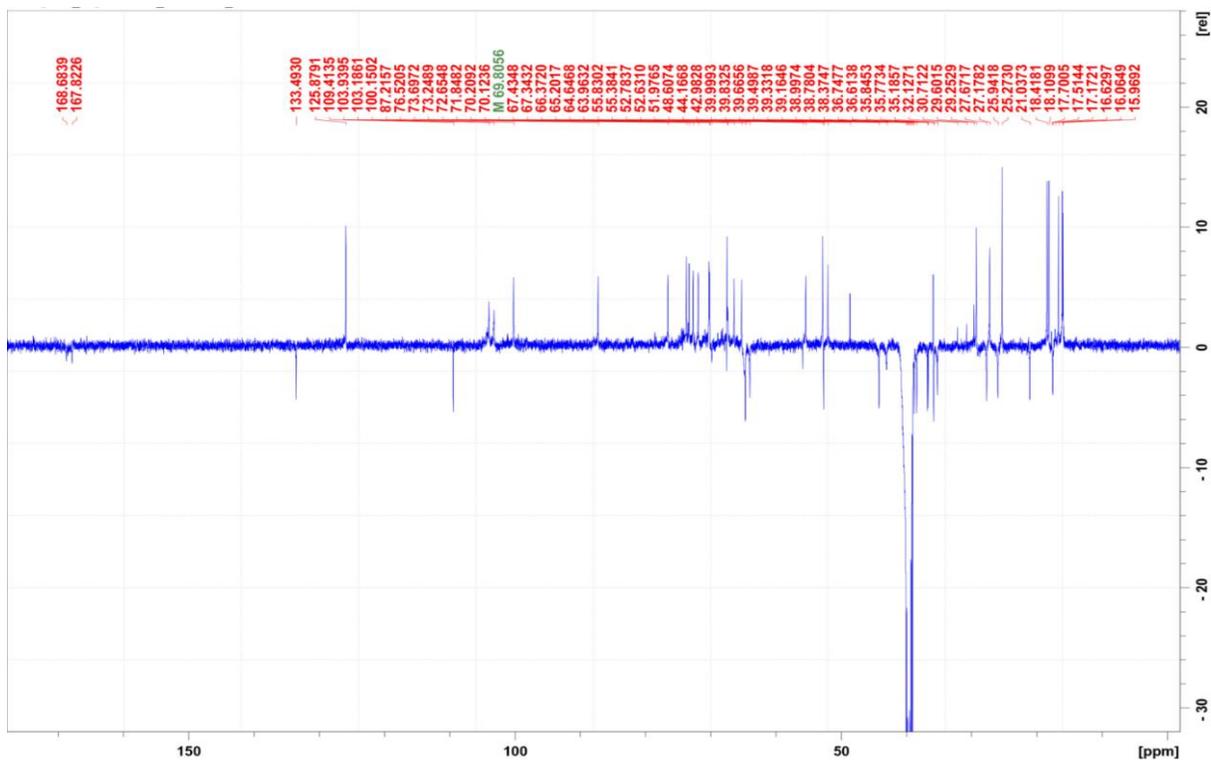


Figure S9: DEPTq spectrum of compounds 4-5 (500 MHz, DMSO-d₆).

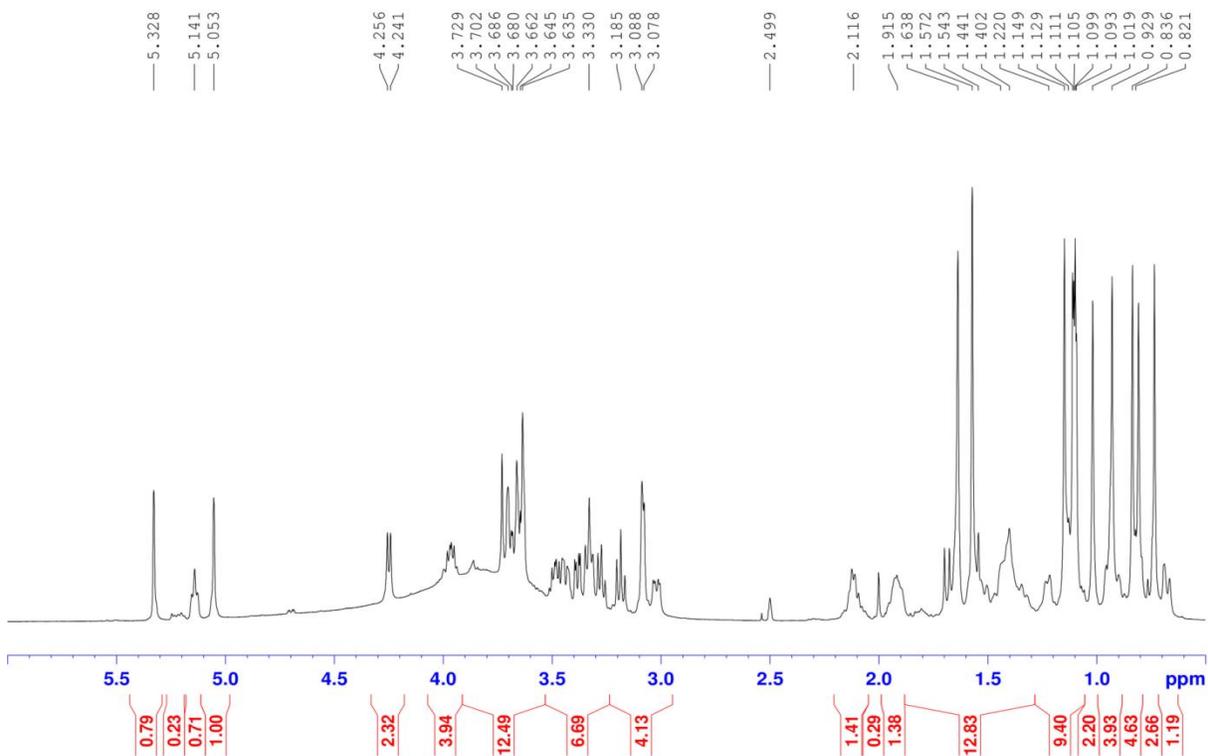


Figure S10: ¹H-NMR spectrum of compound 6 (500 MHz, DMSO-d₆).

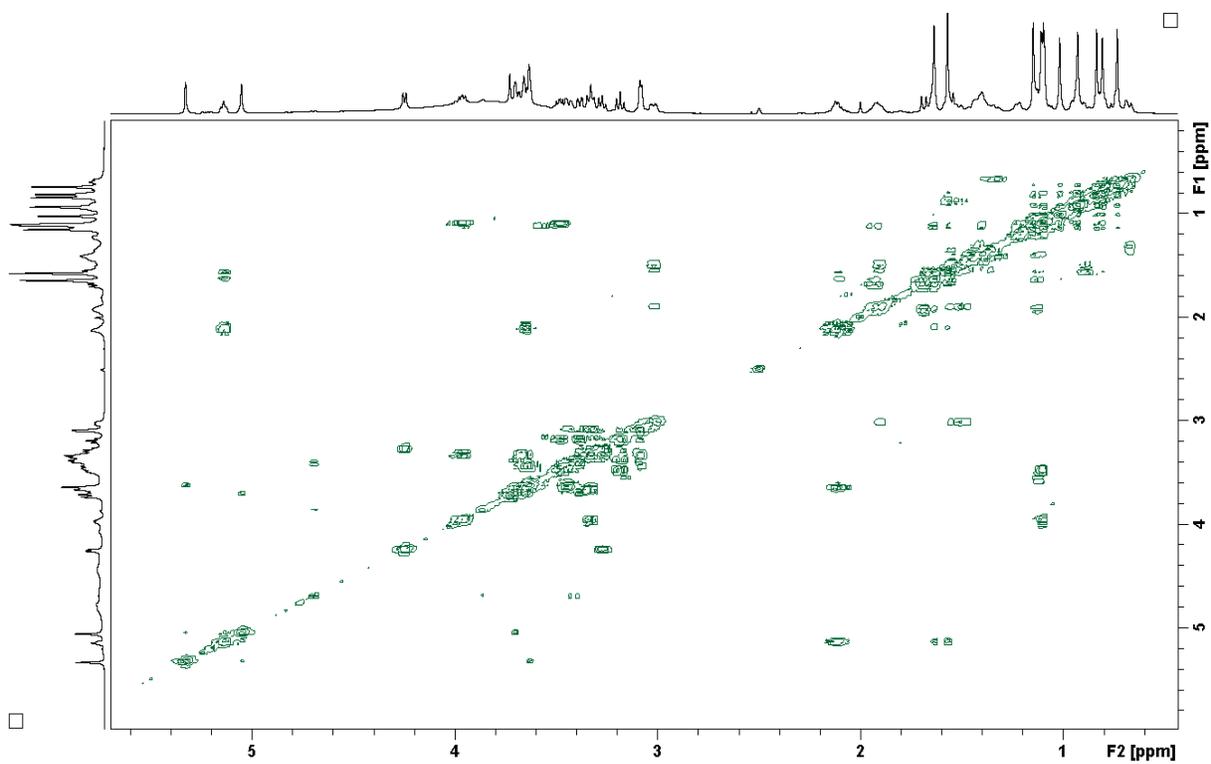


Figure S11: ¹H-¹H COSY spectrum of compound **6** (500 MHz, DMSO-d₆).

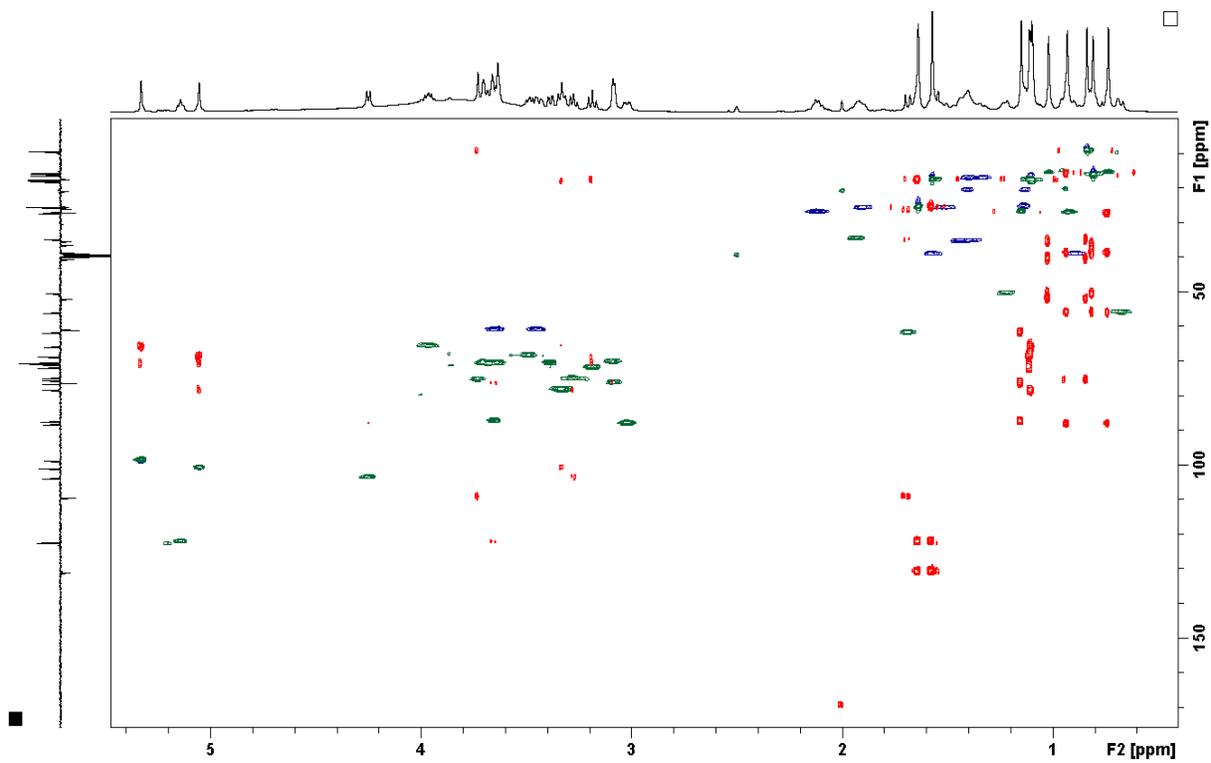


Figure S12: Overlay of HSQC (green-blue) and HMBC (red) spectra of compound **6** (500 and 125 MHz, DMSO-d₆).

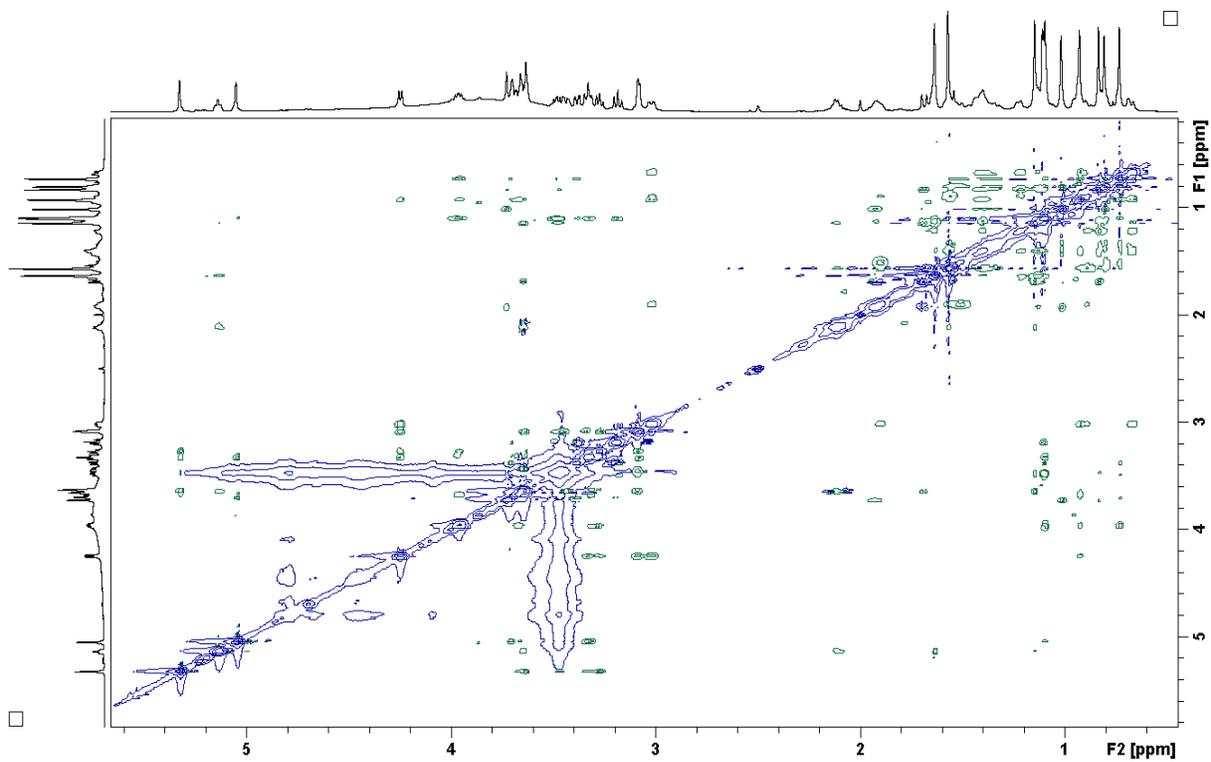


Figure S13: 2D ¹H-¹H ROESY spectrum of compound **6** (500 MHz, DMSO-d₆).

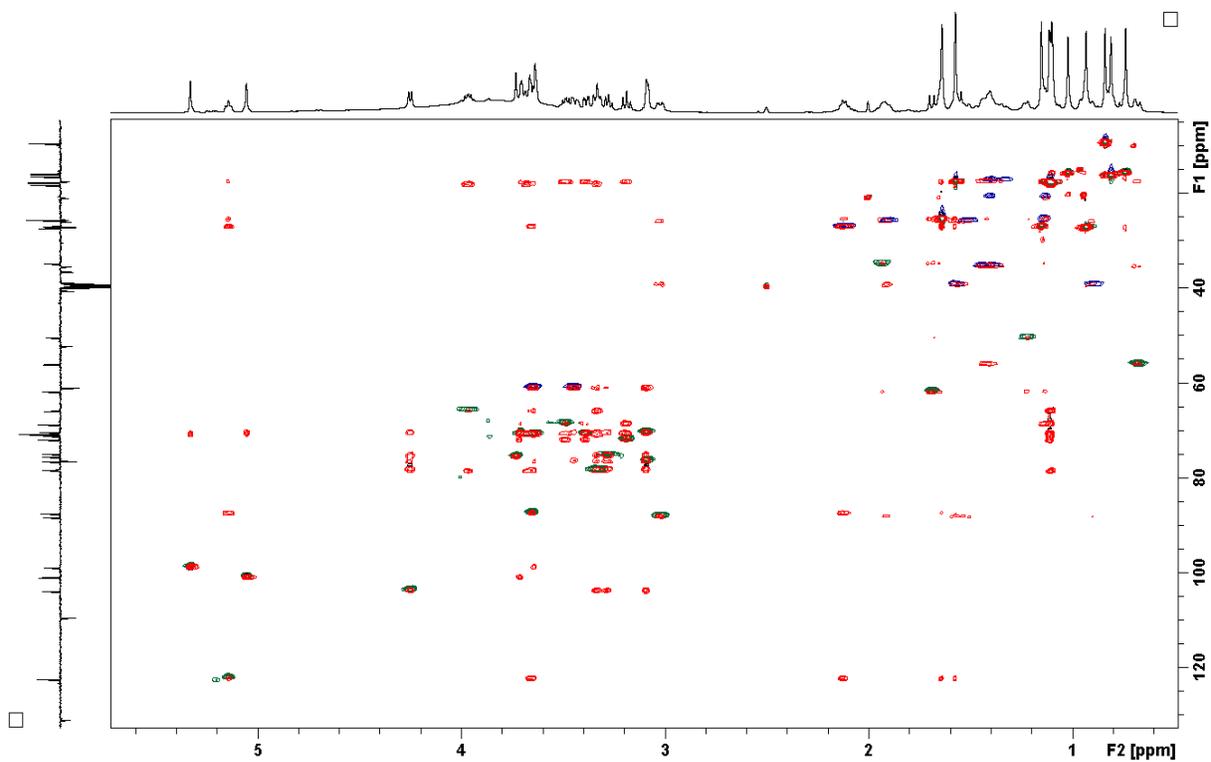


Figure S14: Overlay of HSQC and TOCSY spectra of compound **6** (500 and 125 MHz, DMSO-d₆).

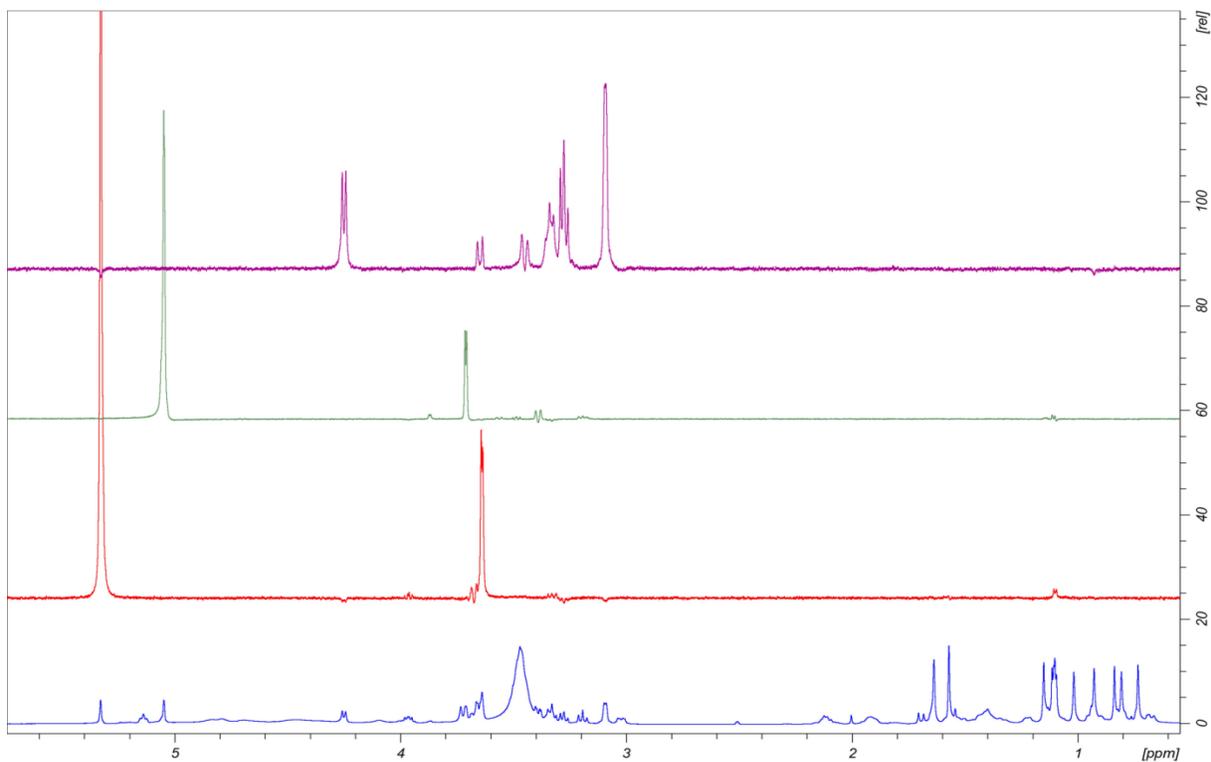


Figure S15: Selective 1D TOCSY spectrum of compound **6** (500 MHz, DMSO-d₆).

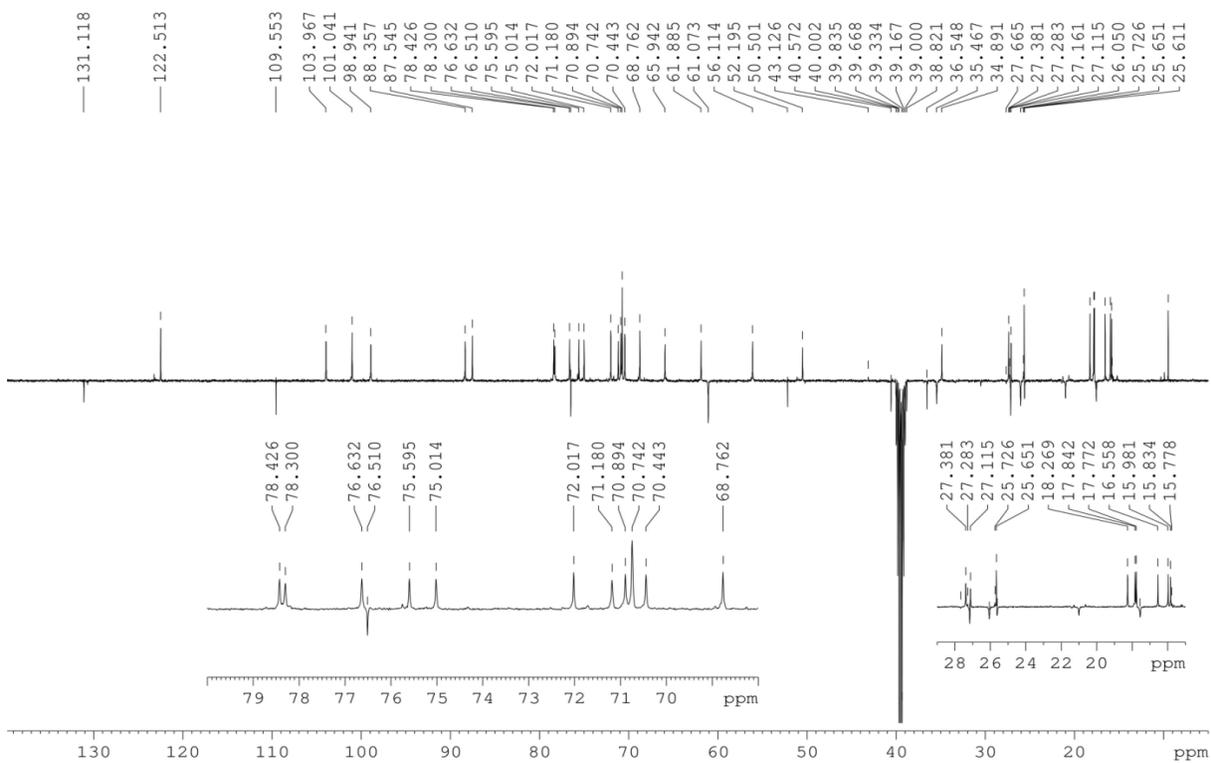


Figure S16: DEPTq spectrum of compound **6** (125 MHz, DMSO-d₆).

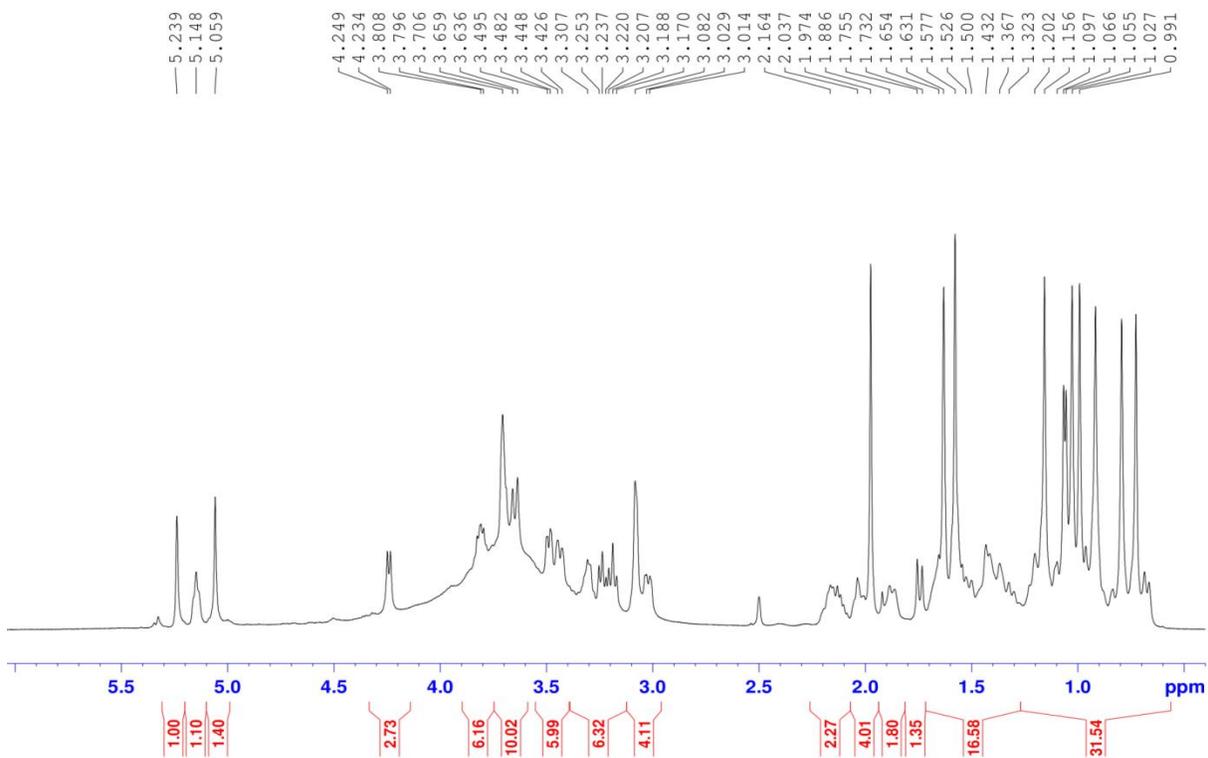


Figure S17: ^1H -NMR spectrum of compound **7** (500 MHz, DMSO-d_6).

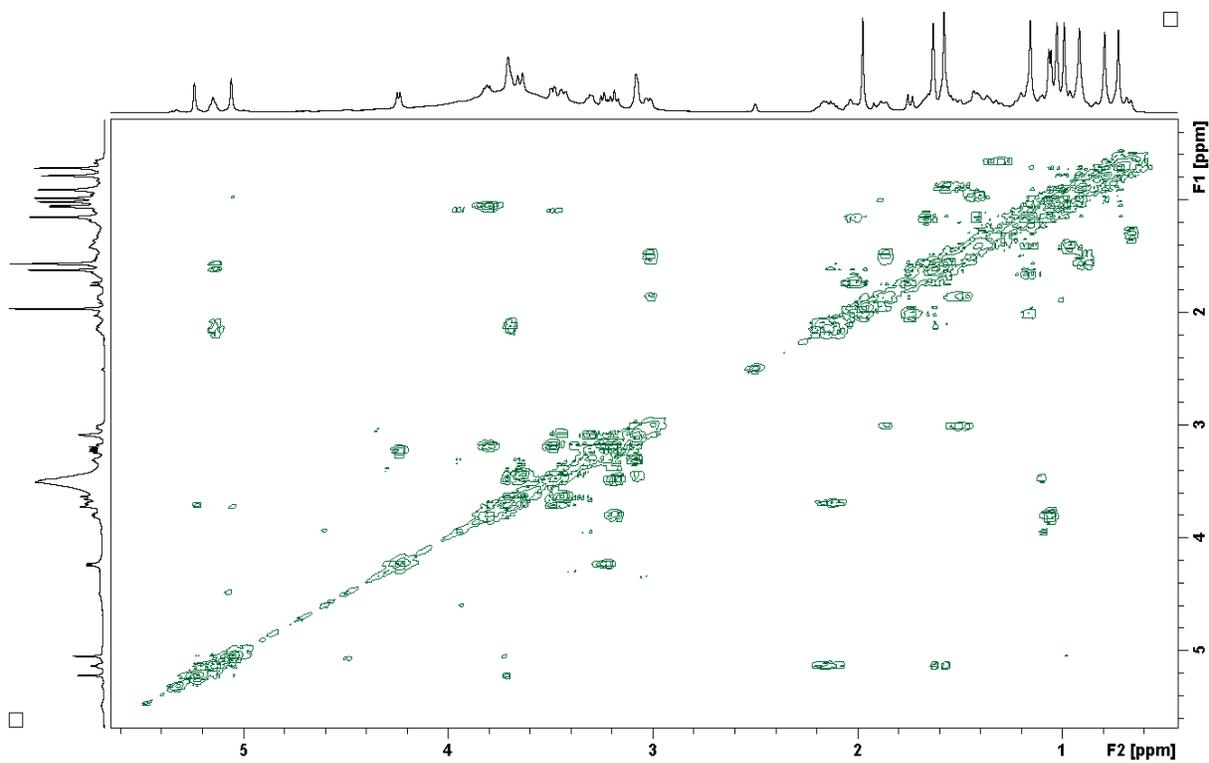


Figure S18: ^1H - ^1H COSY spectrum of compound **7** (500 MHz, DMSO-d_6).

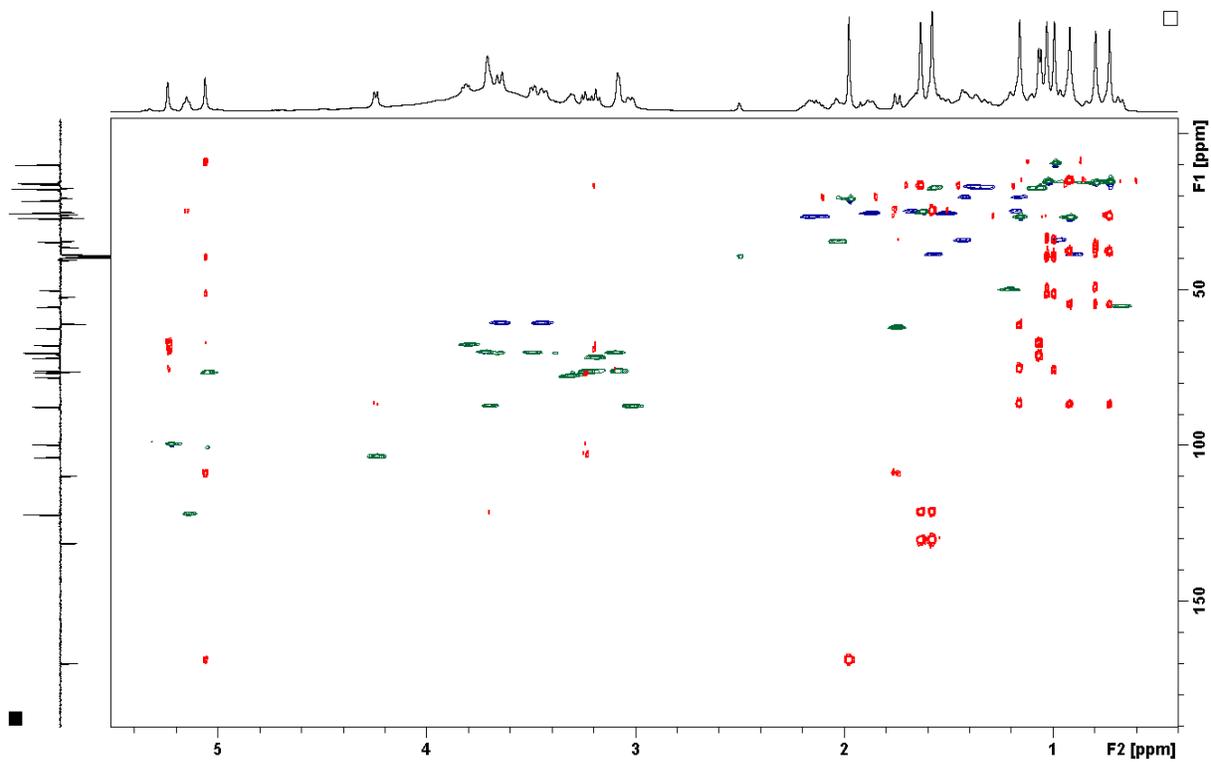


Figure S19: Overlay of HSQC (green-blue) and HMBC (red) spectra of compound **7** (500 and 125 MHz, DMSO- d_6).

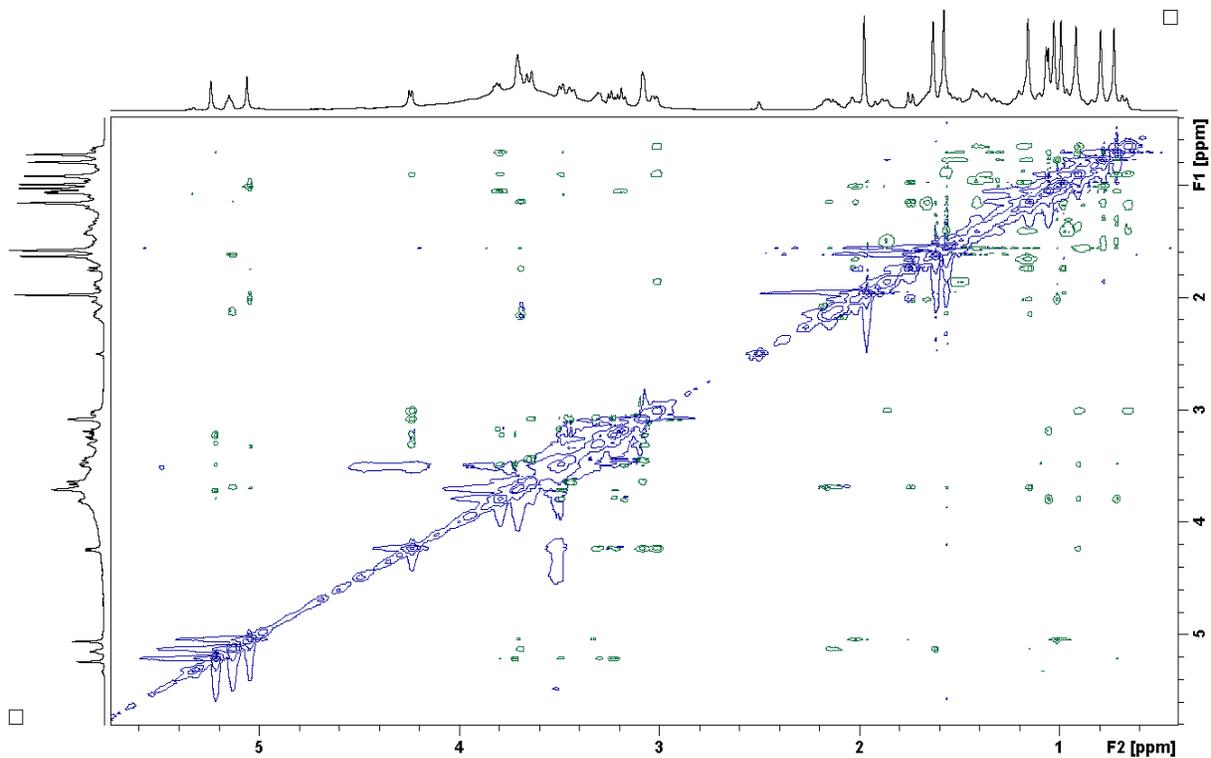


Figure S20: 2D ^1H - ^1H ROESY spectrum of compound **7** (500 MHz, DMSO- d_6).

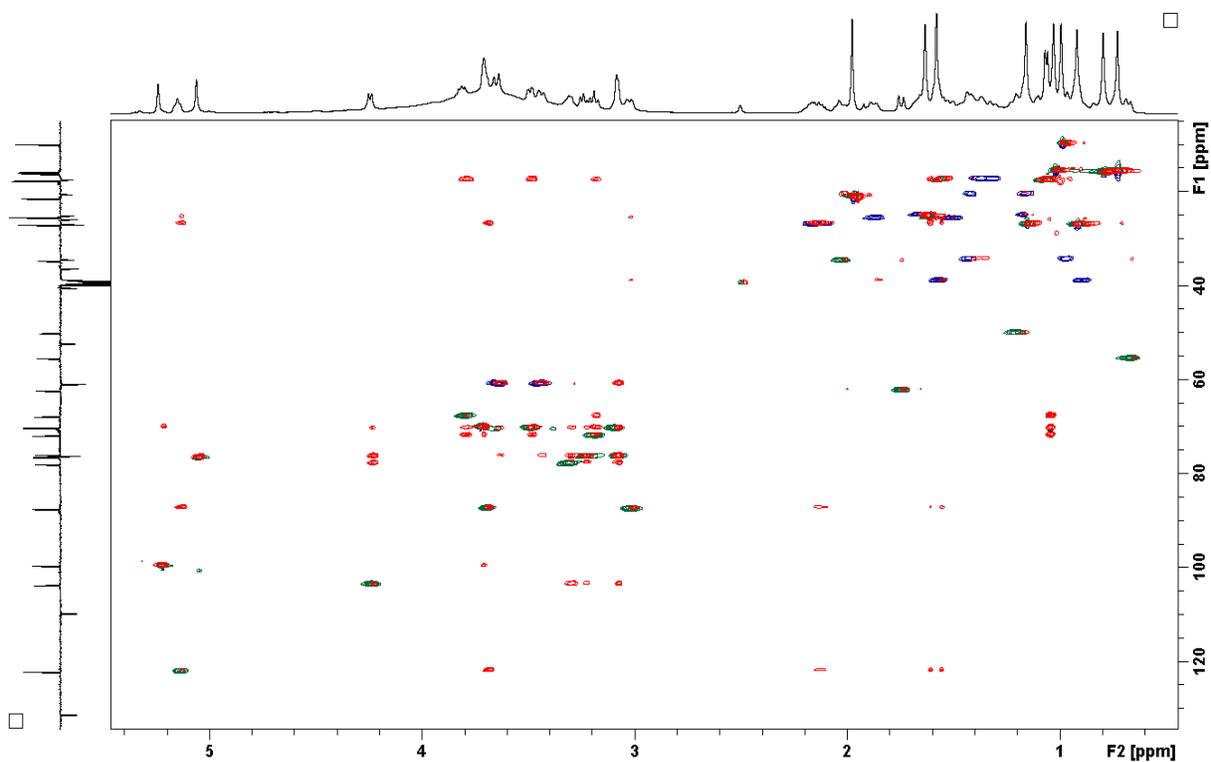


Figure S21: Overlay of HSQC and TOCSY spectra of compound 7 (500 and 125 MHz, DMSO-d₆).

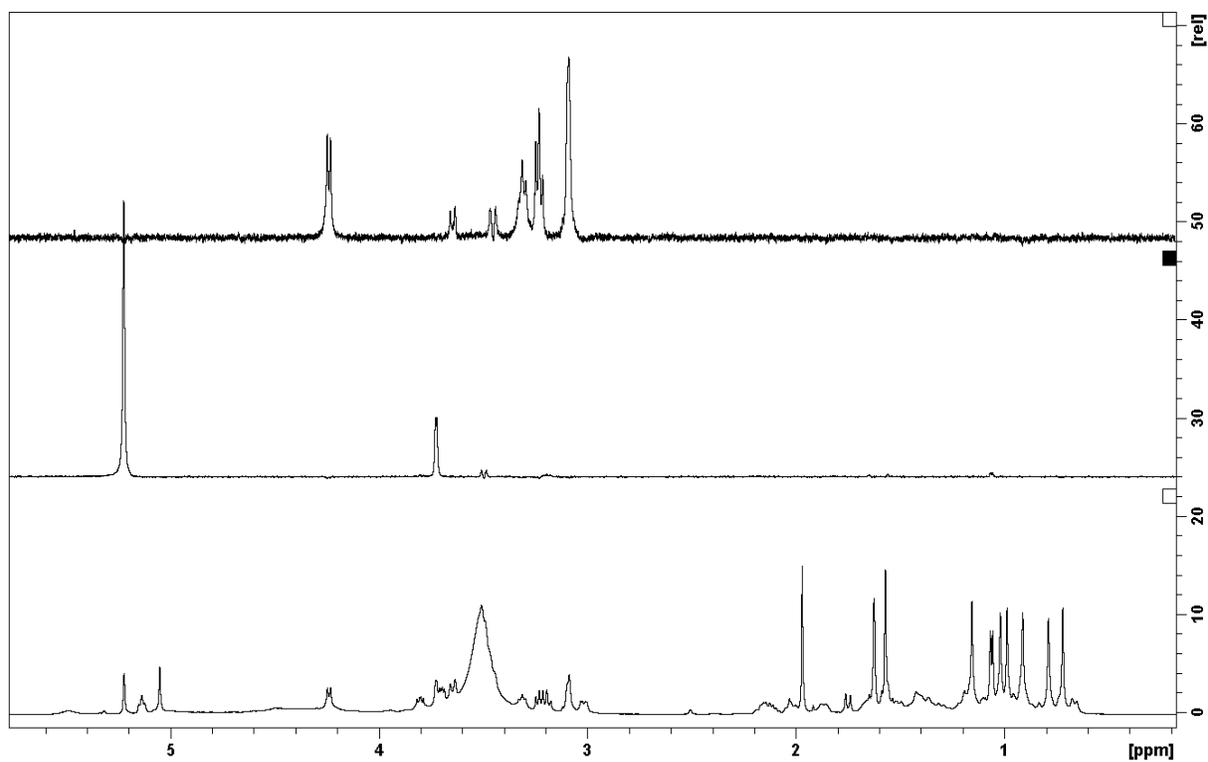


Figure S22: Selective 1D TOCSY spectrum of compound 7 (500 MHz, DMSO-d₆).

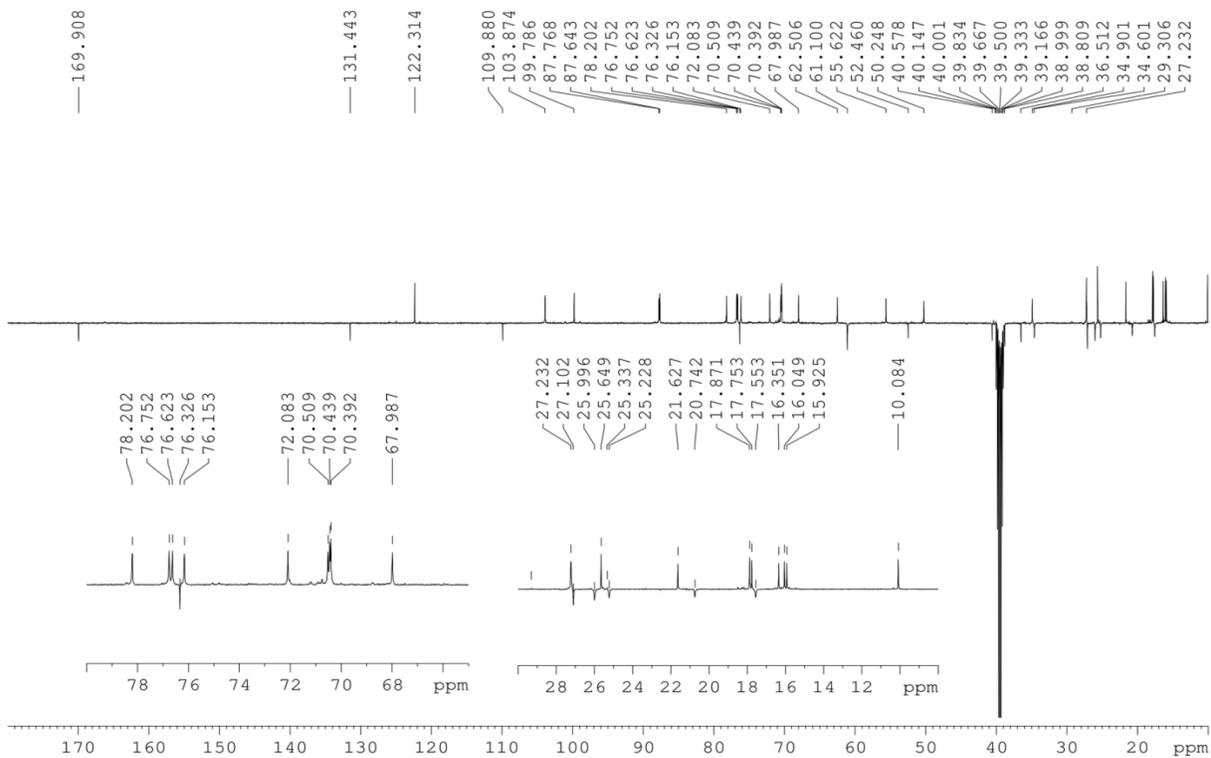


Figure S23: DEPTq spectrum of compound **7** (500 MHz, DMSO-d₆).

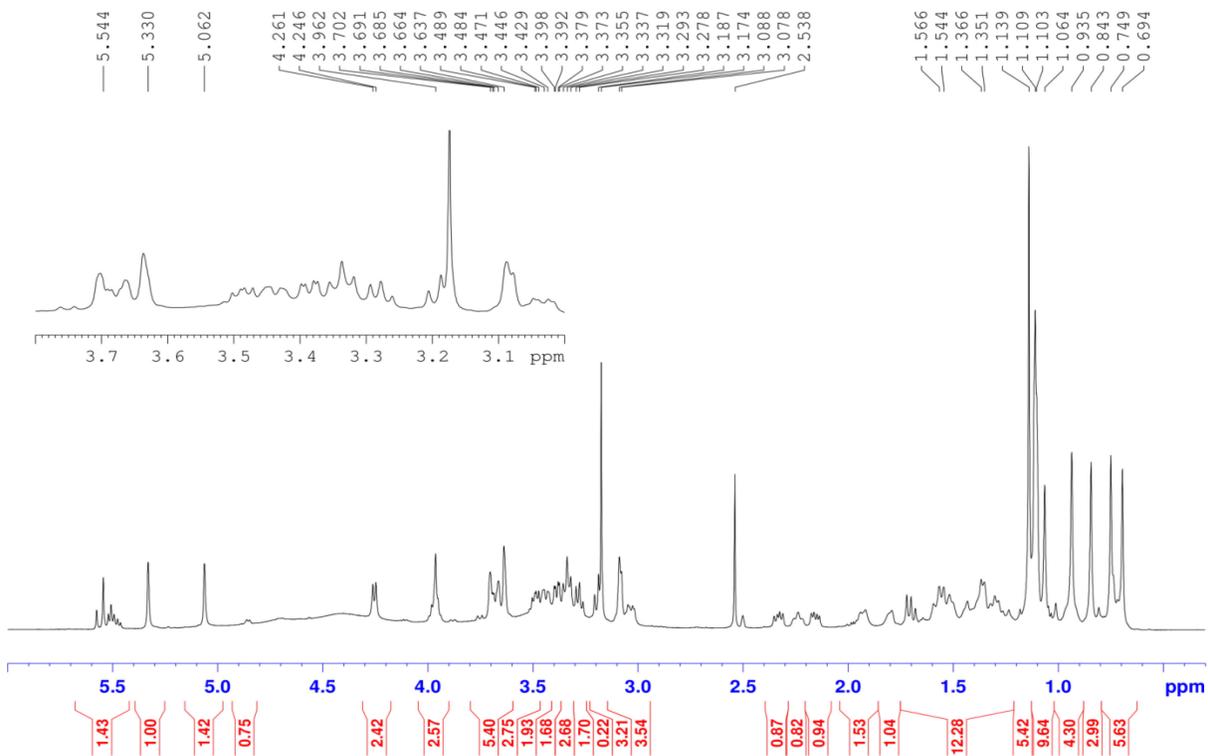


Figure S24: ¹H-NMR spectrum of compound **8** (500 MHz, DMSO-d₆).

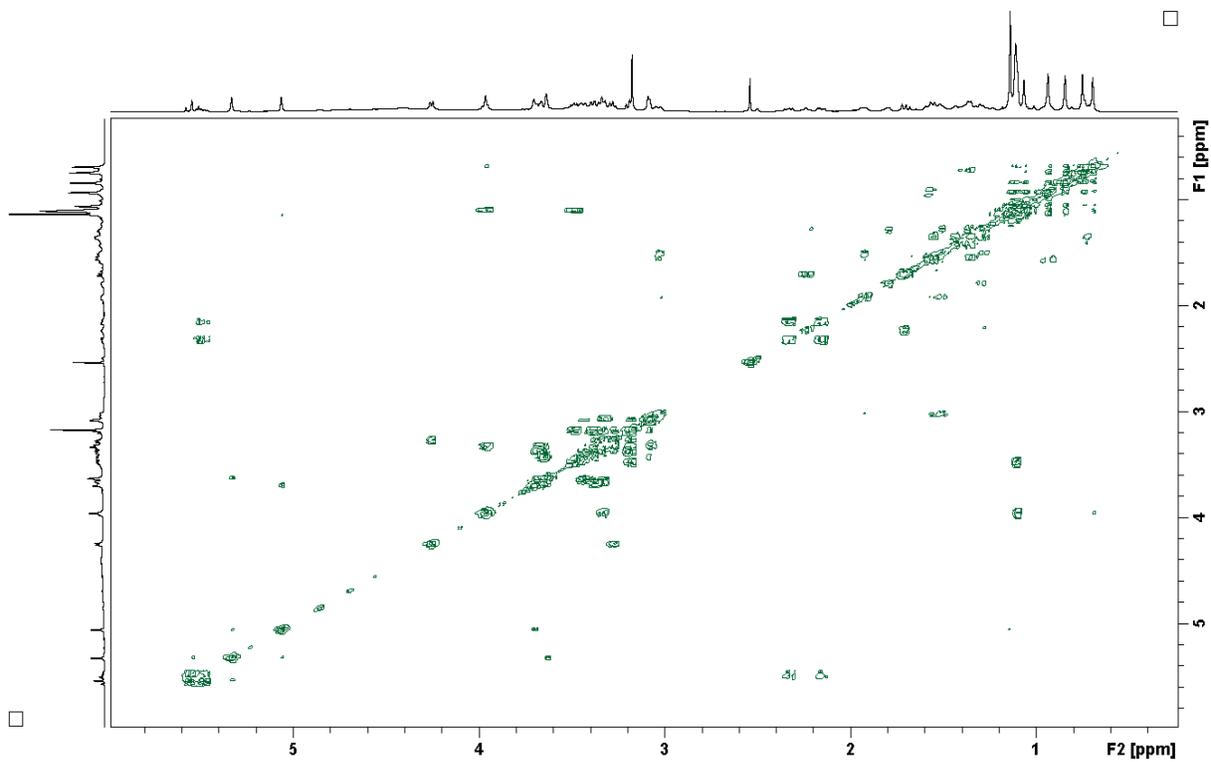


Figure S25: ¹H-¹H COSY spectrum of compound **8** (500 MHz, DMSO-d₆).

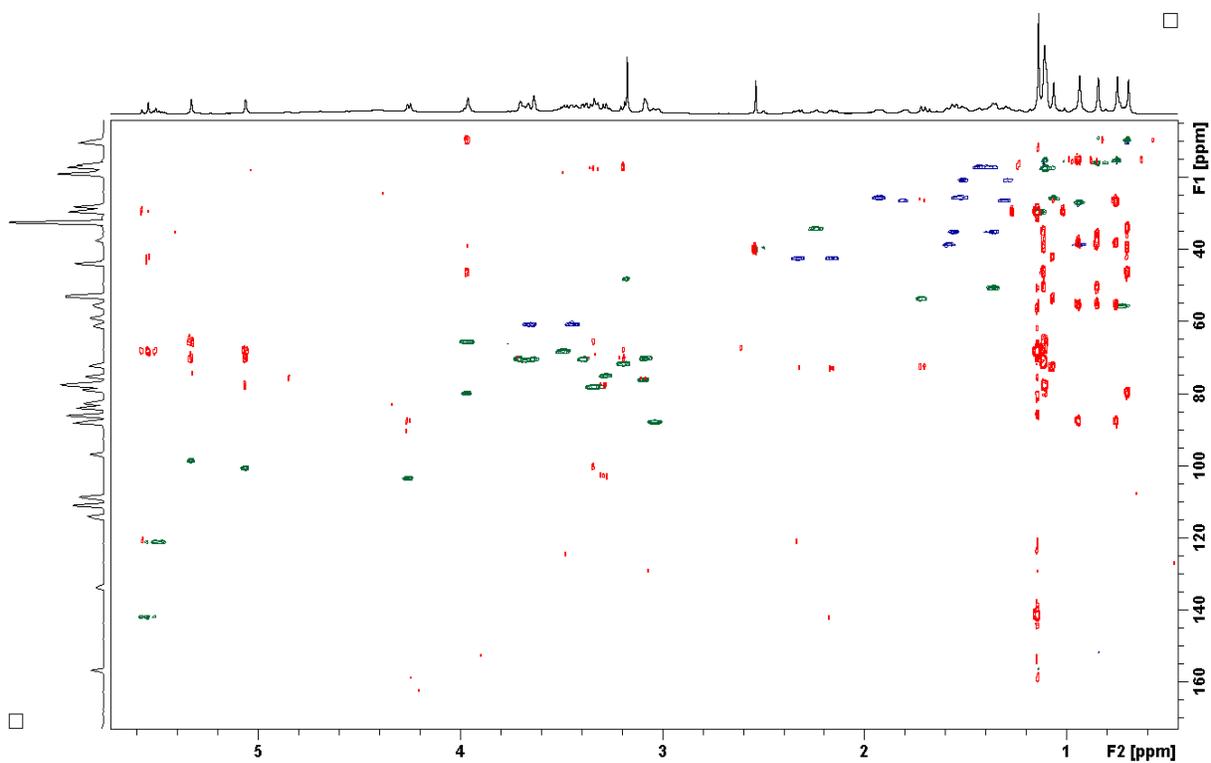


Figure S26: Overlay of HSQC (green-blue) and HMBC (red) spectra of compound **8** (500 and 125 MHz, DMSO-d₆).

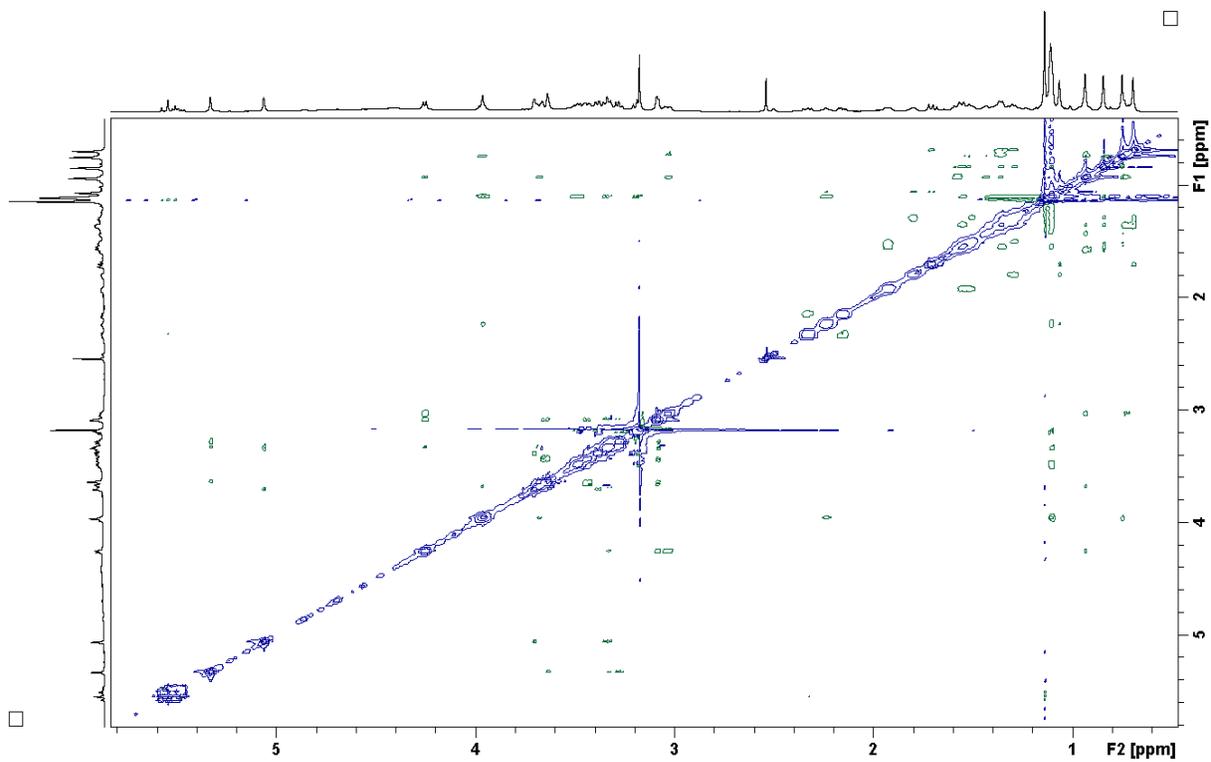


Figure S27: 2D ¹H-¹H ROESY spectrum of compound **8** (500 MHz, DMSO-d₆).

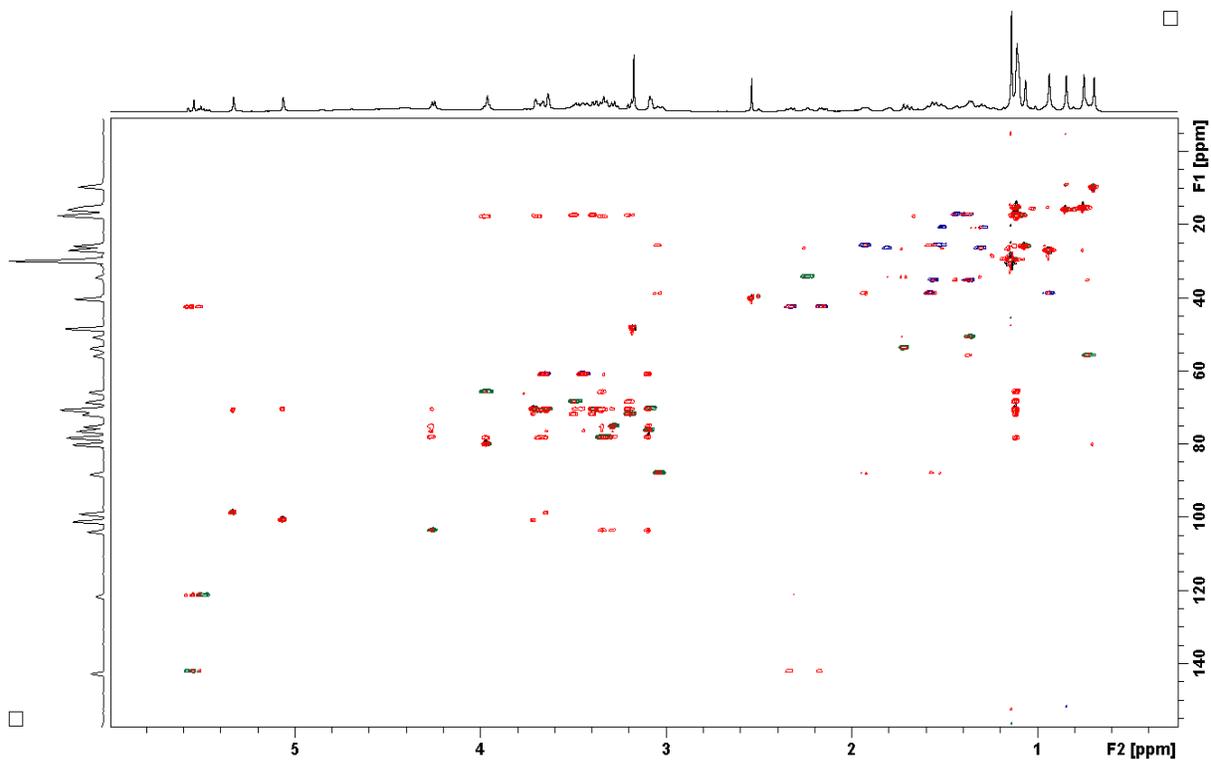


Figure S28: Overlay of HSQC and TOCSY spectra of compound **8** (500 and 125 MHz, DMSO-d₆).

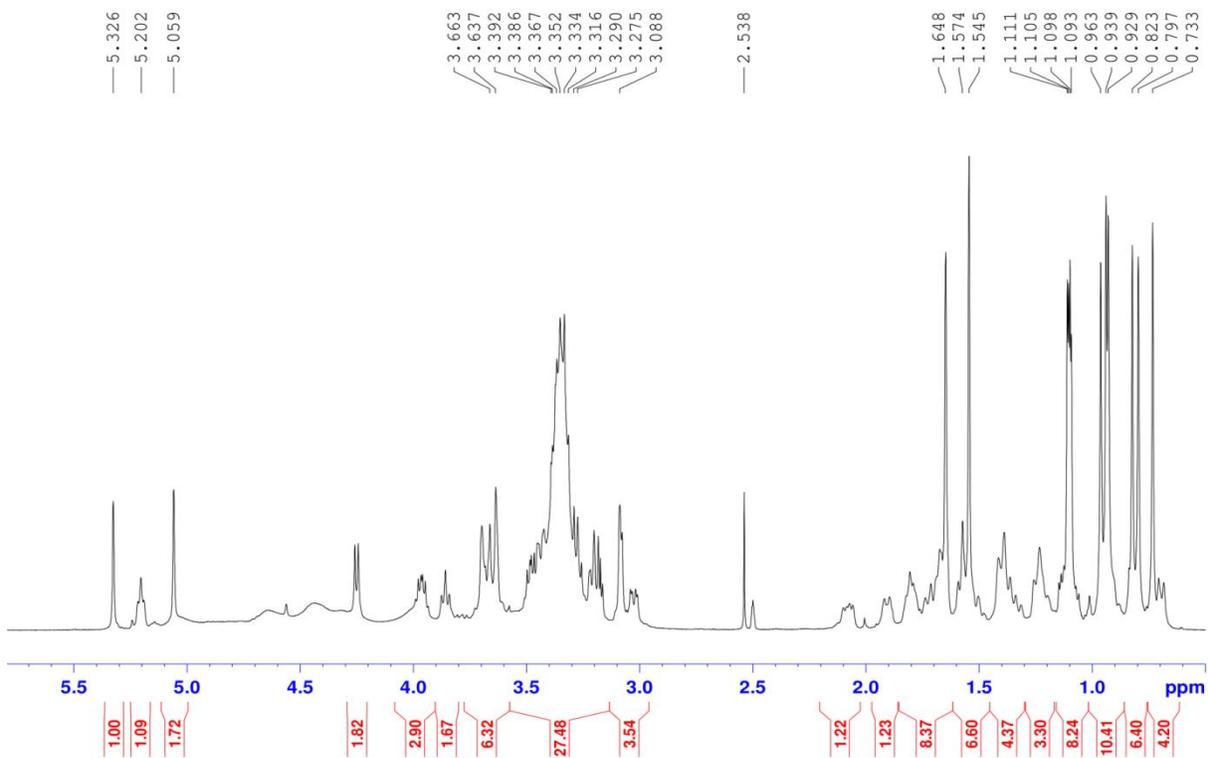


Figure S29: ^1H -NMR spectrum of compound **9** (500 MHz, DMSO-d_6).

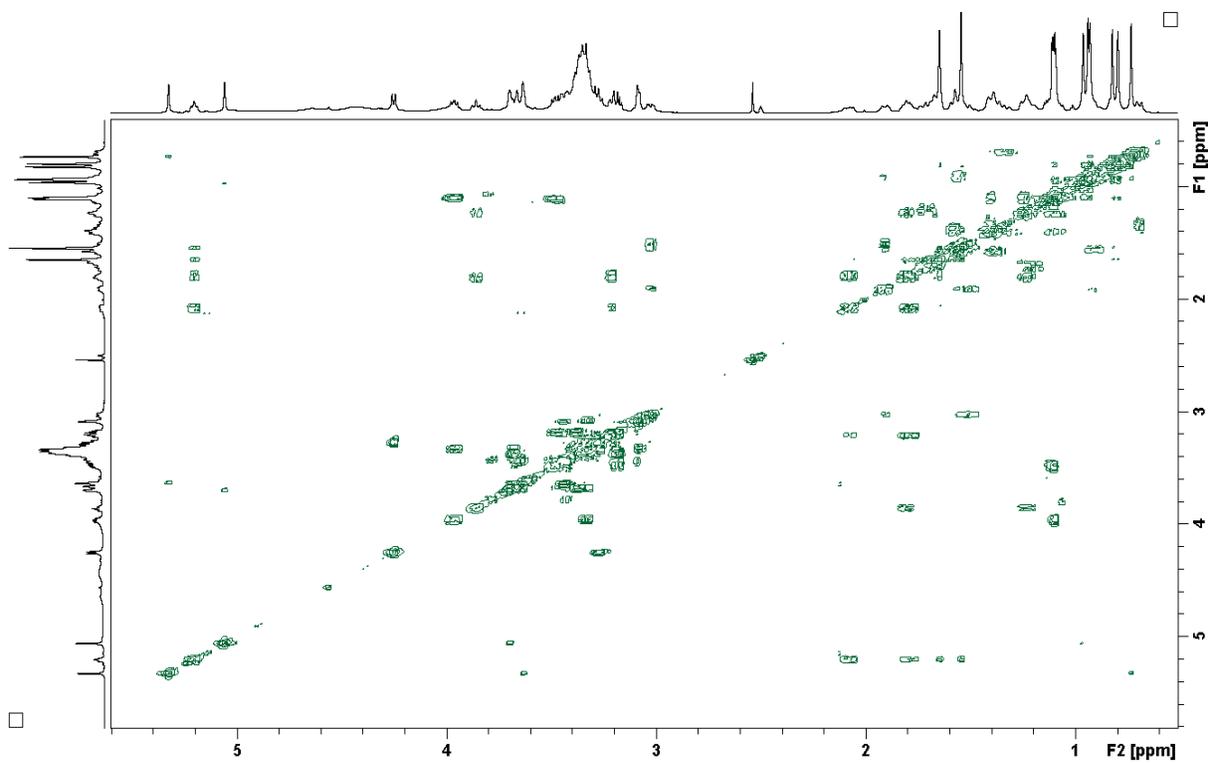


Figure S30: ^1H - ^1H COSY spectrum of compound **9** (500 MHz, DMSO-d_6).

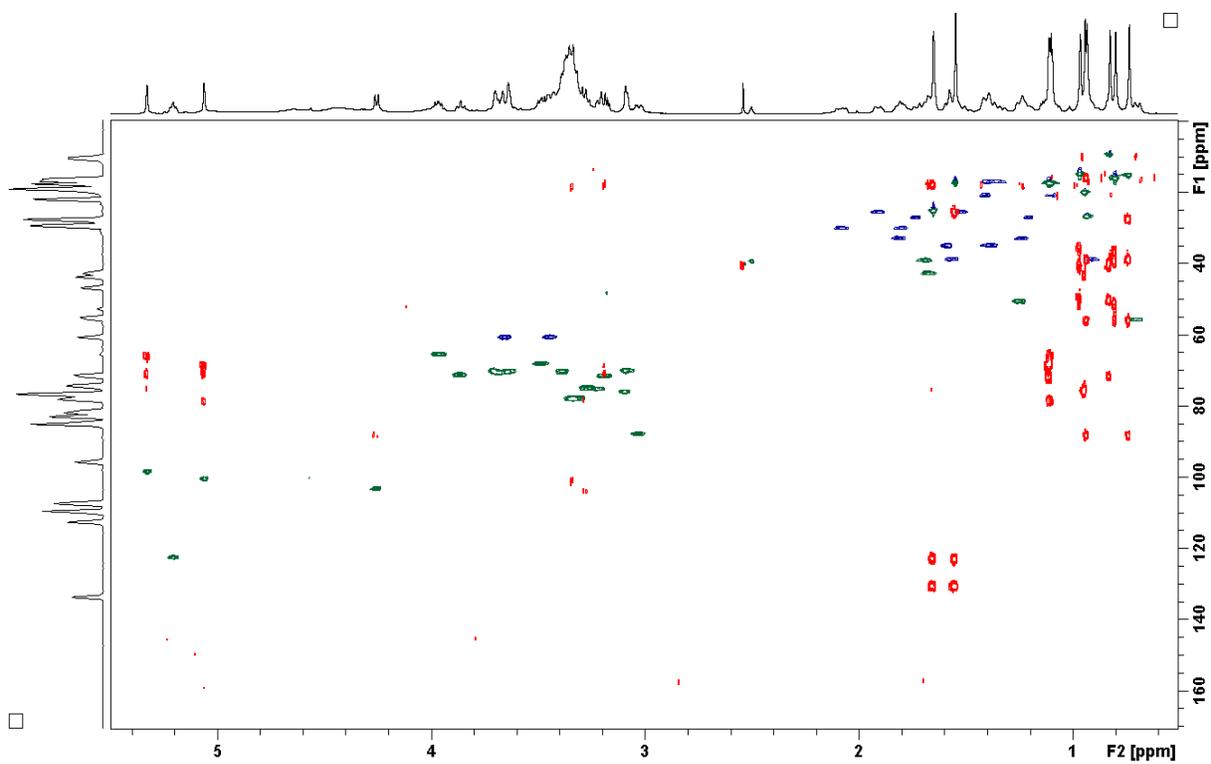


Figure S31: Overlay of HSQC (green-blue) and HMBC (red) spectra of compound **9** (500 and 125 MHz, DMSO- d_6).

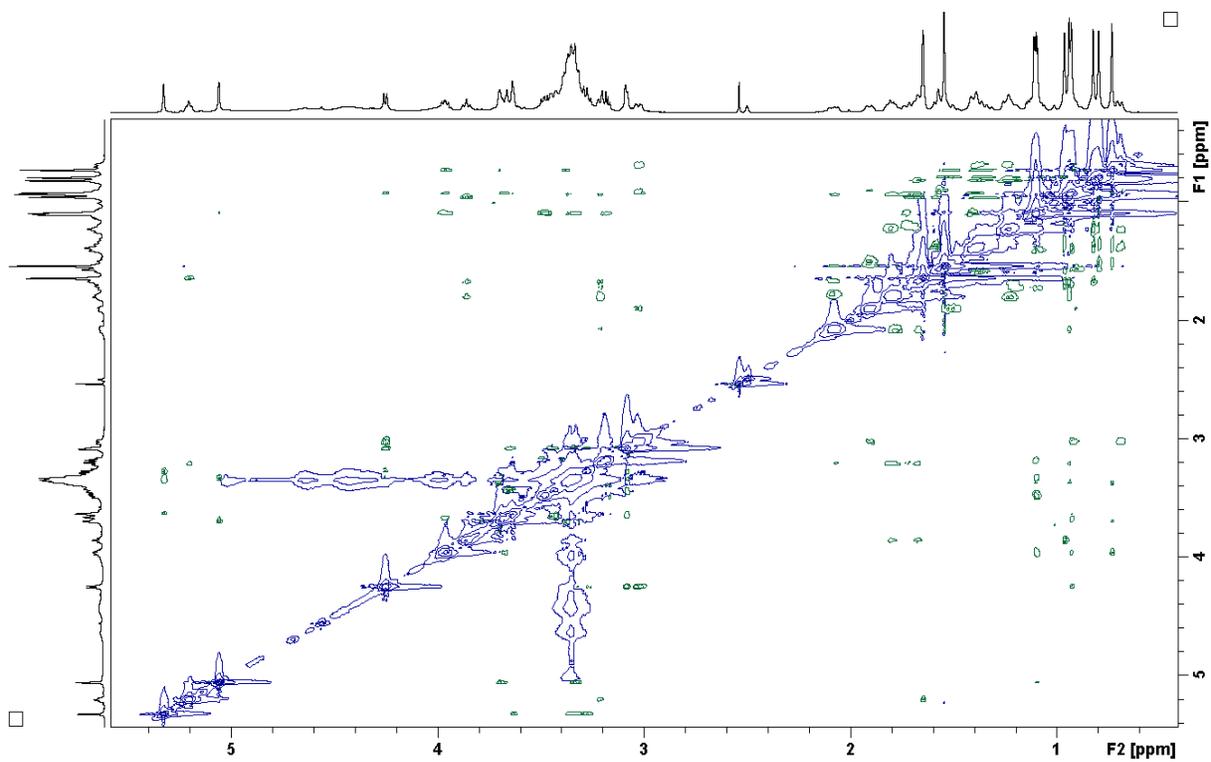


Figure S32: 2D ^1H - ^1H ROESY spectrum of compound **9** (500 MHz, DMSO- d_6).

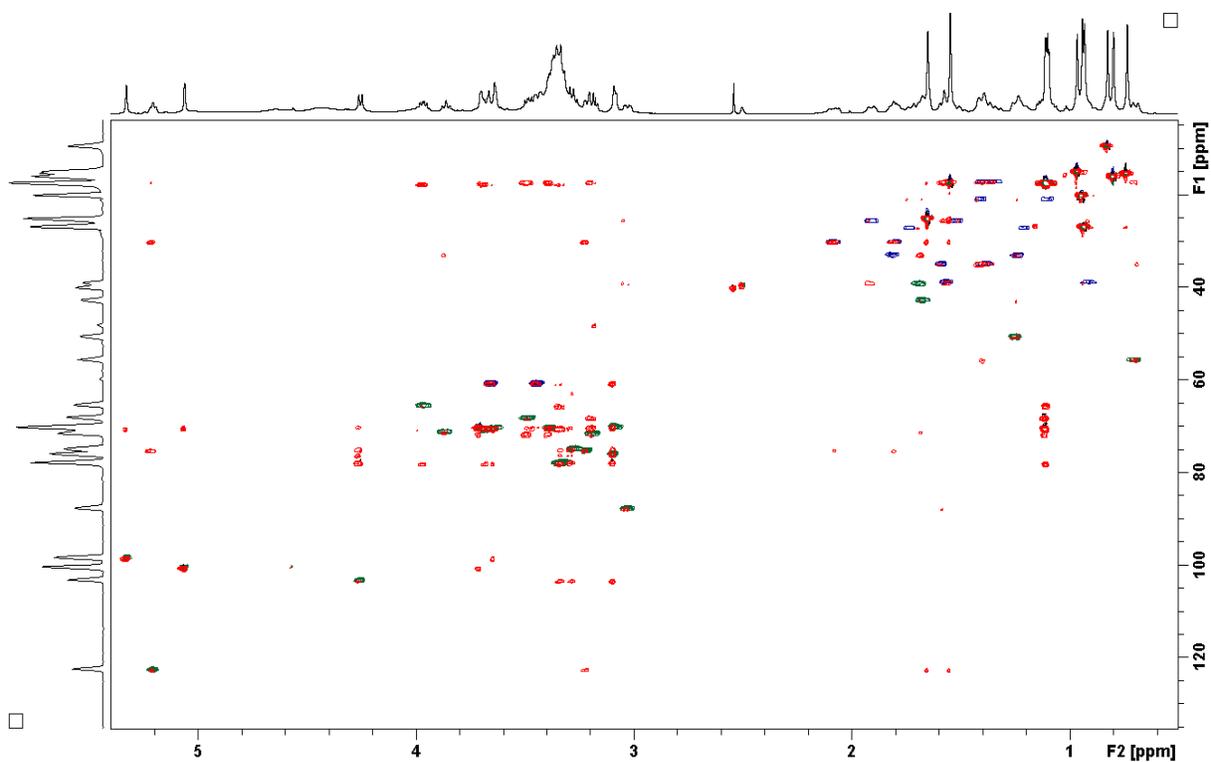


Figure S33: Overlay of HSQC and TOCSY spectra of compound **9** (500 and 125 MHz, DMSO- d_6).

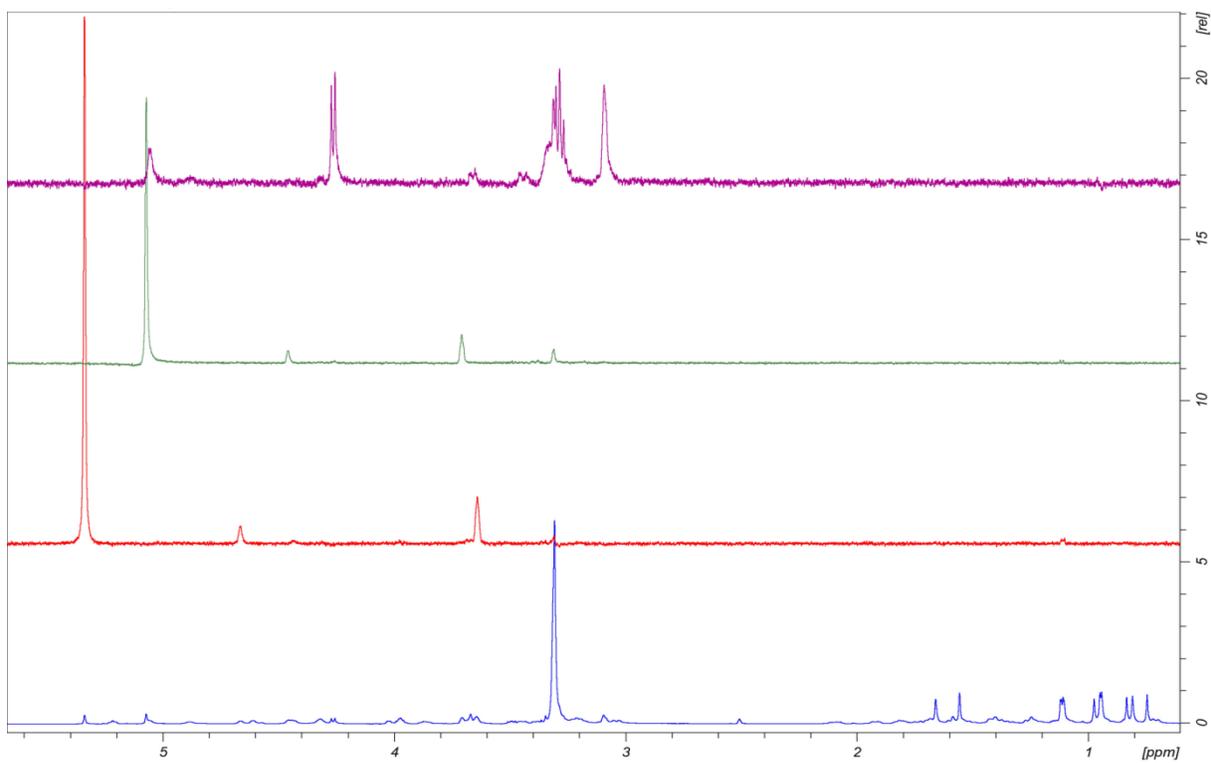


Figure S34: Selective 1D TOCSY spectrum of compound **9** (500 MHz, DMSO- d_6).

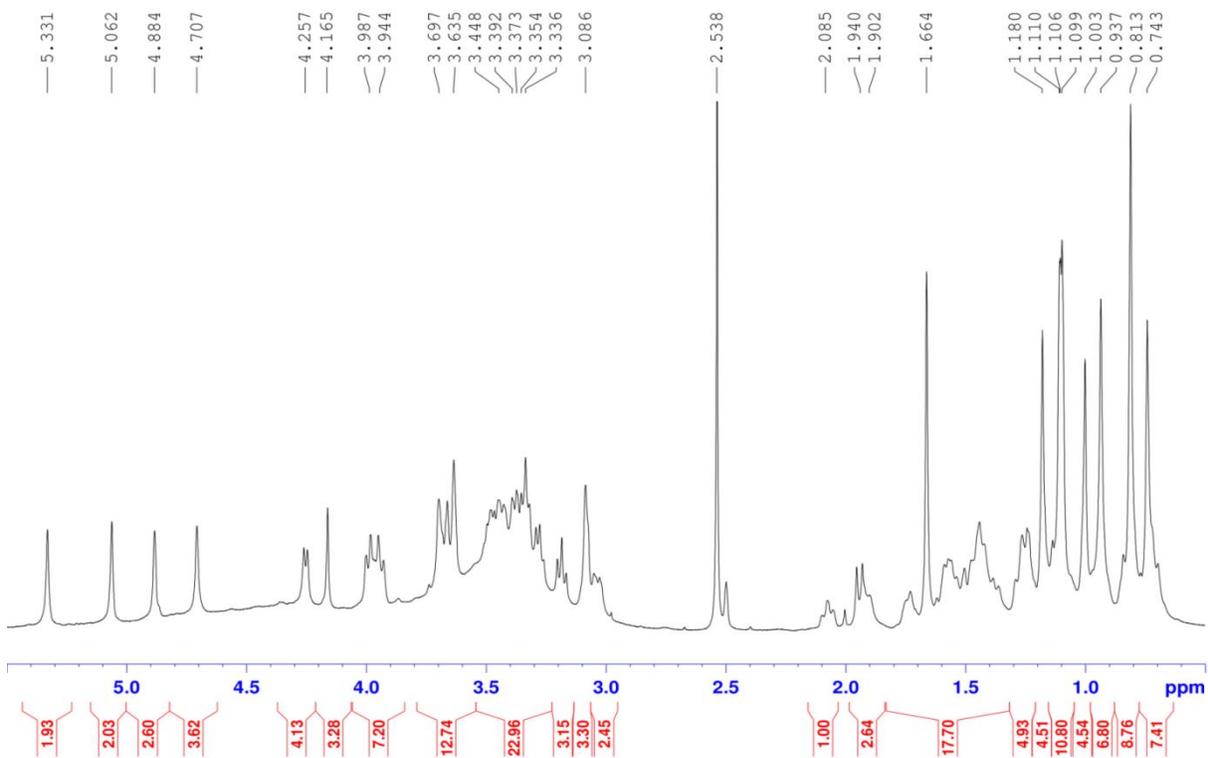


Figure S35: $^1\text{H-NMR}$ spectrum of compound **10** (500 MHz, DMSO-d_6).

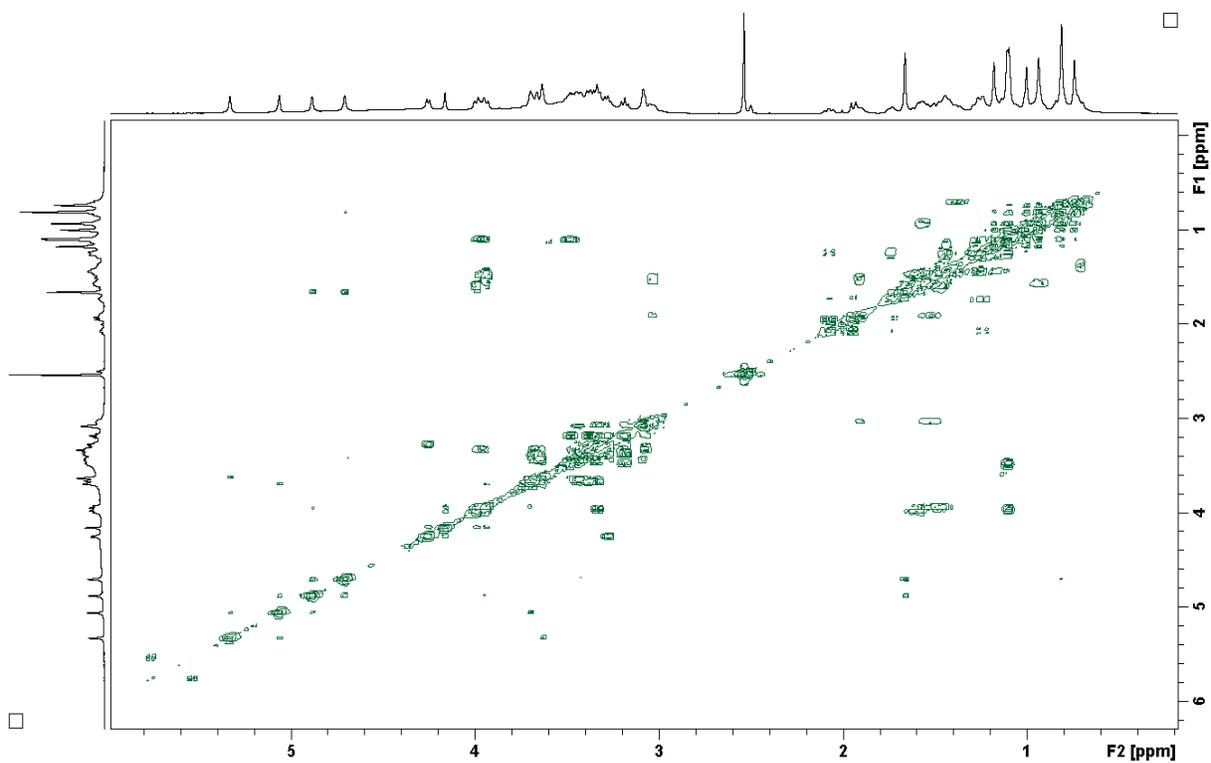


Figure S36: $^1\text{H-}^1\text{H}$ COSY spectrum of compound **10** (500 MHz, DMSO-d_6).

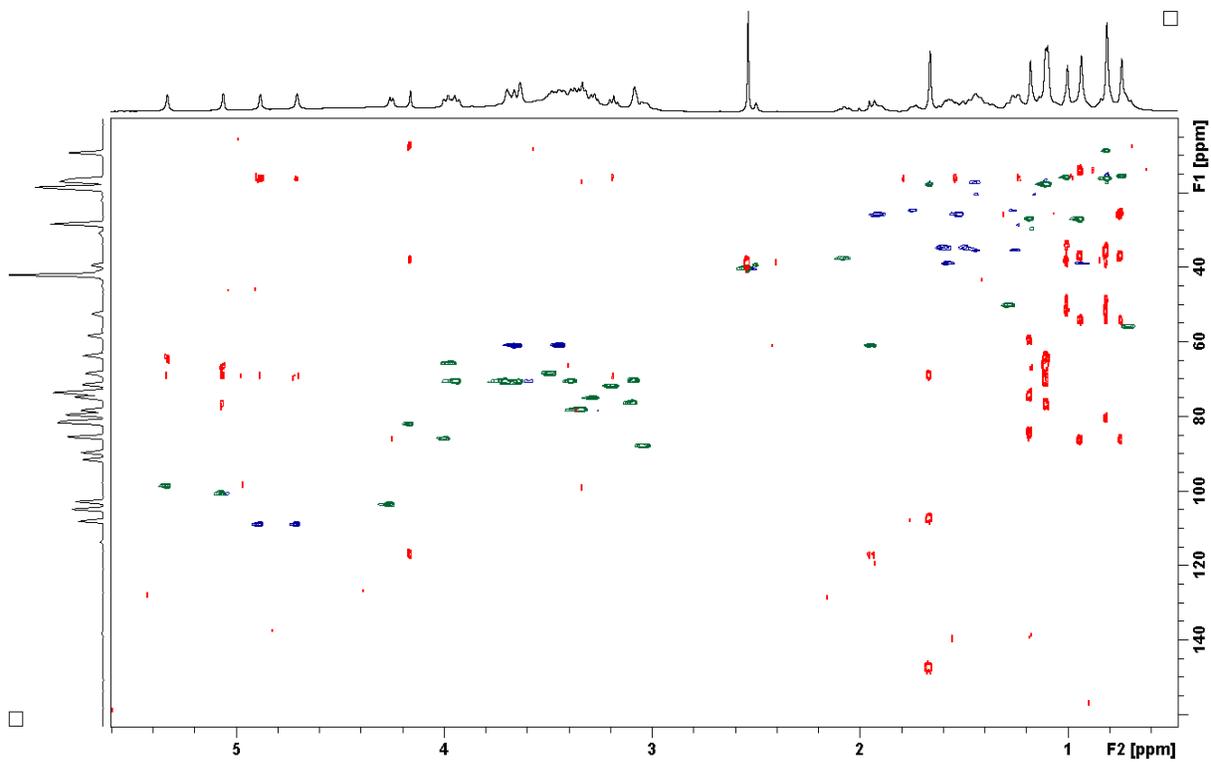


Figure S37: Overlay of HSQC (green-blue) and HMBC (red) spectra of compound **10** (500 and 125 MHz, DMSO-d₆).

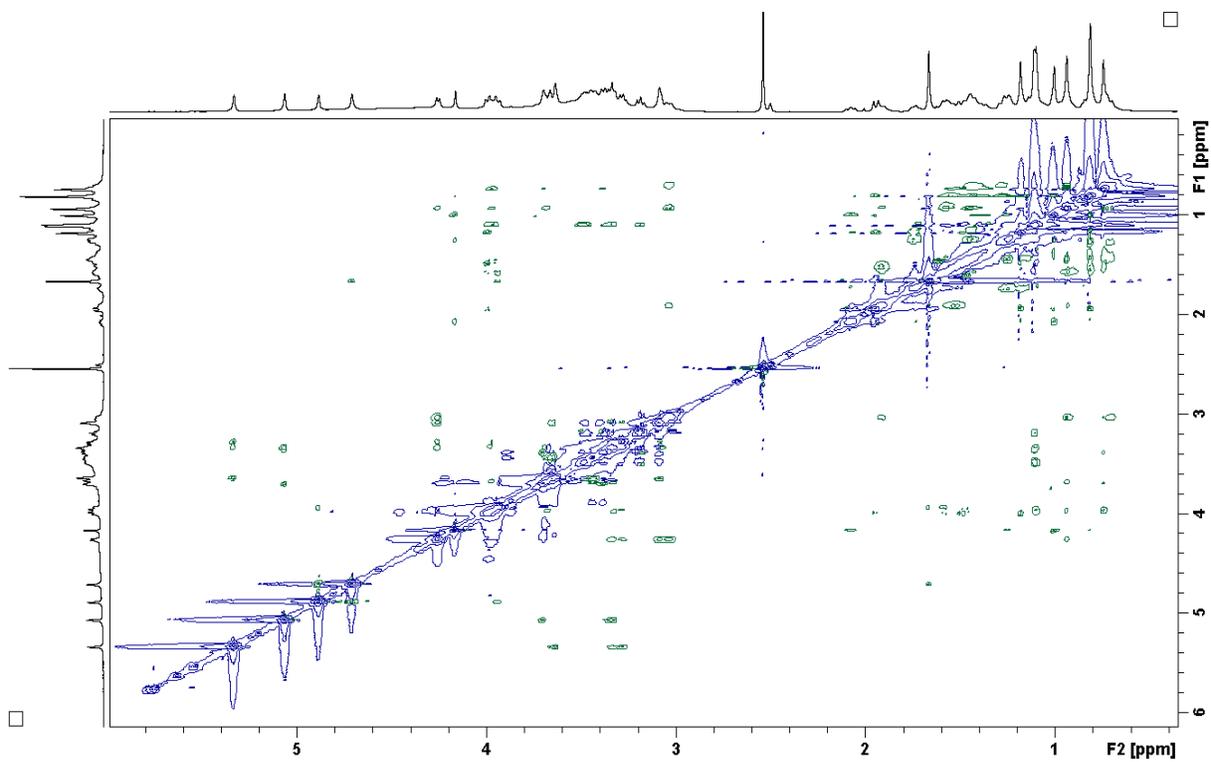


Figure S38: 2D ¹H-¹H ROESY spectrum of compound **10** (500 MHz, DMSO-d₆).

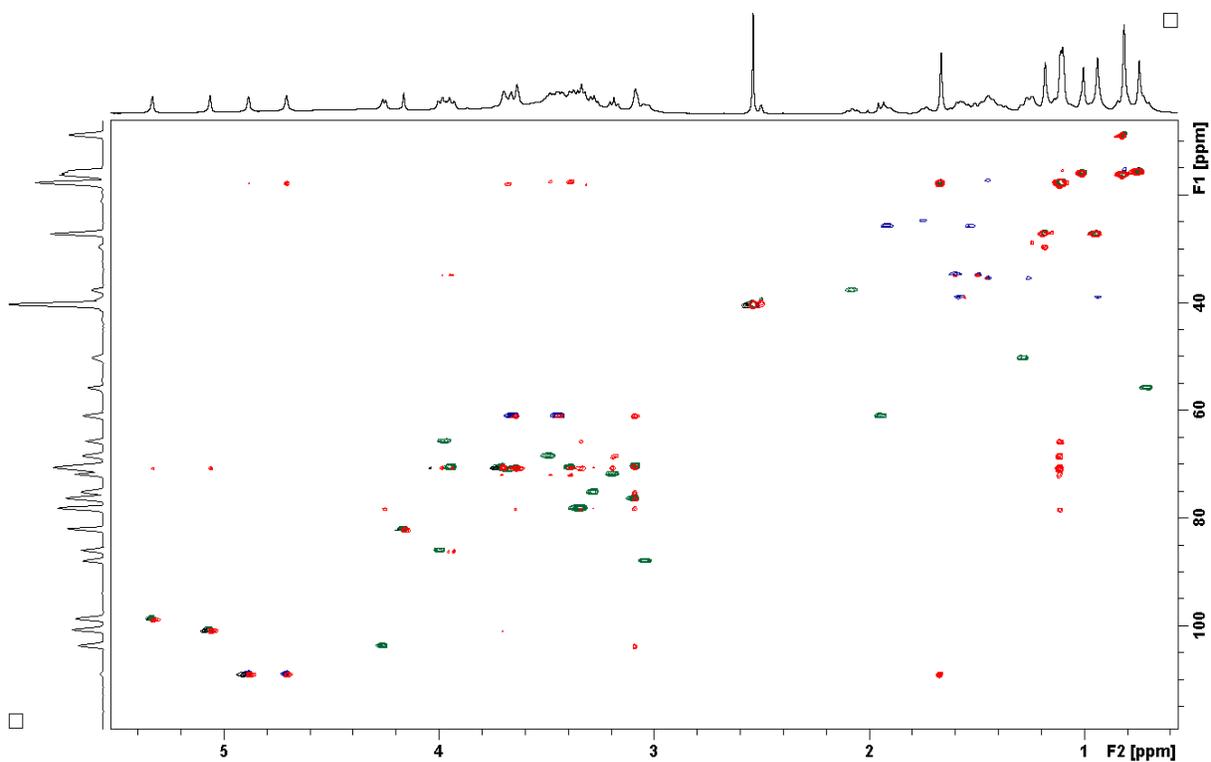


Figure S39: Overlay of HSQC and TOCSY spectra of compound **10** (500 and 125 MHz, DMSO- d_6).

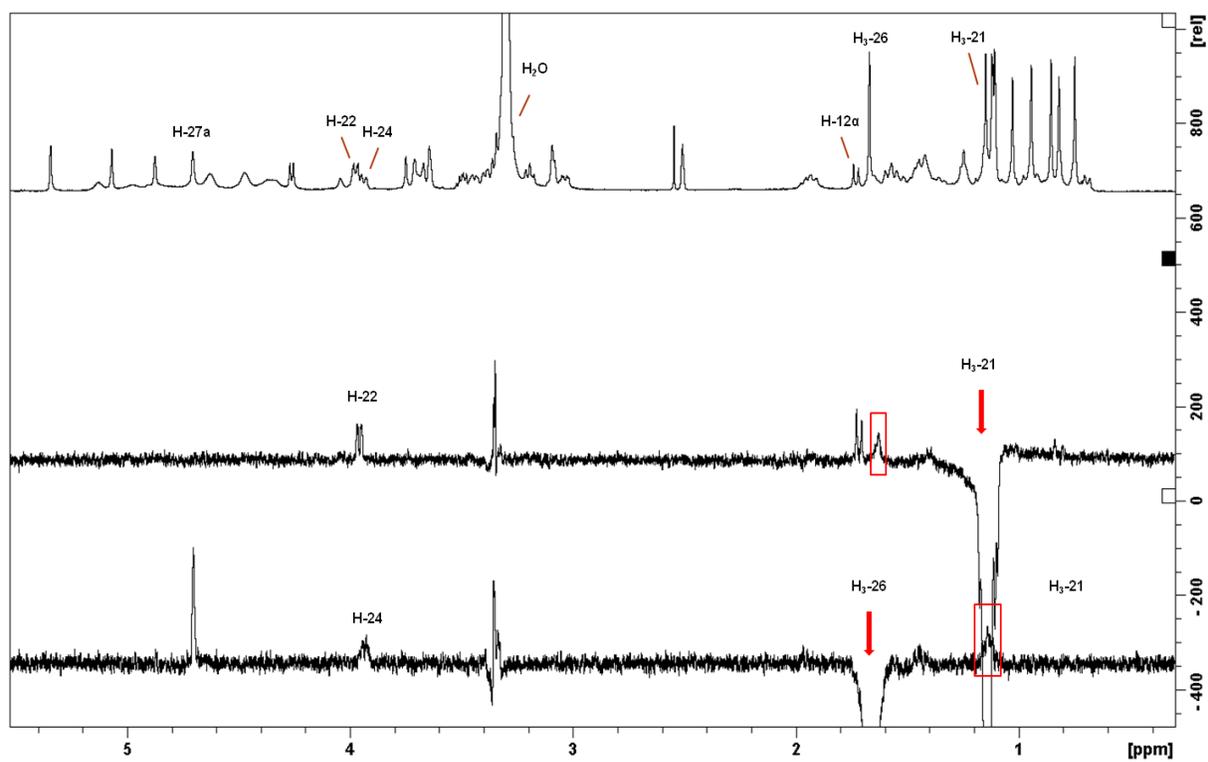


Figure S40: 1D selective ROESY compound **10** (500 MHz, DMSO- d_6).

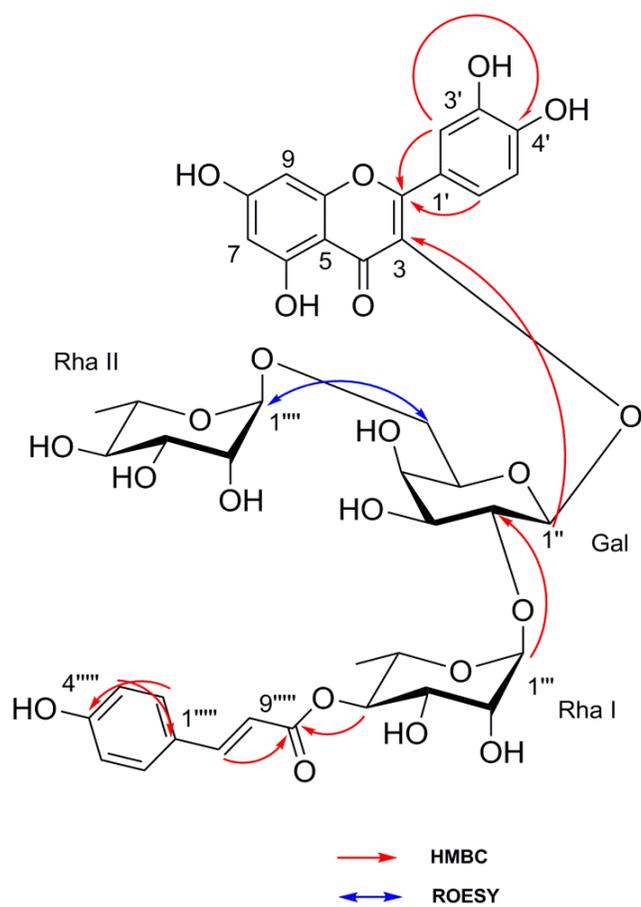


Figure S41: Key ^1H - ^1H ROESY and HMBC correlations of compound **11**.

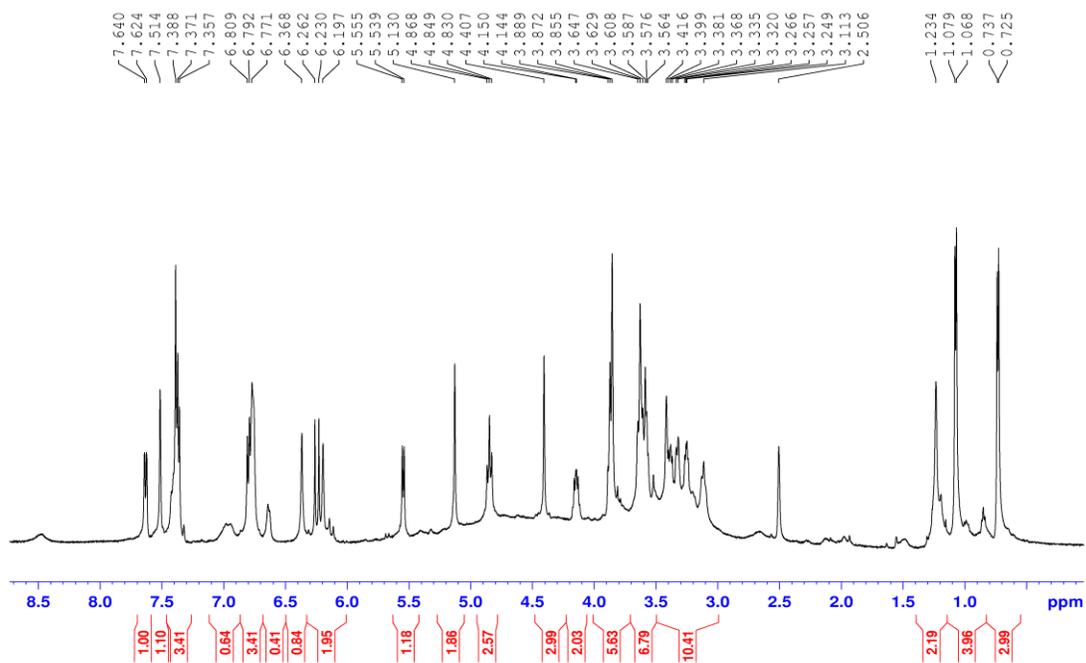


Figure S42: ^1H -NMR spectrum of compound **11** (500 MHz, DMSO-d_6).

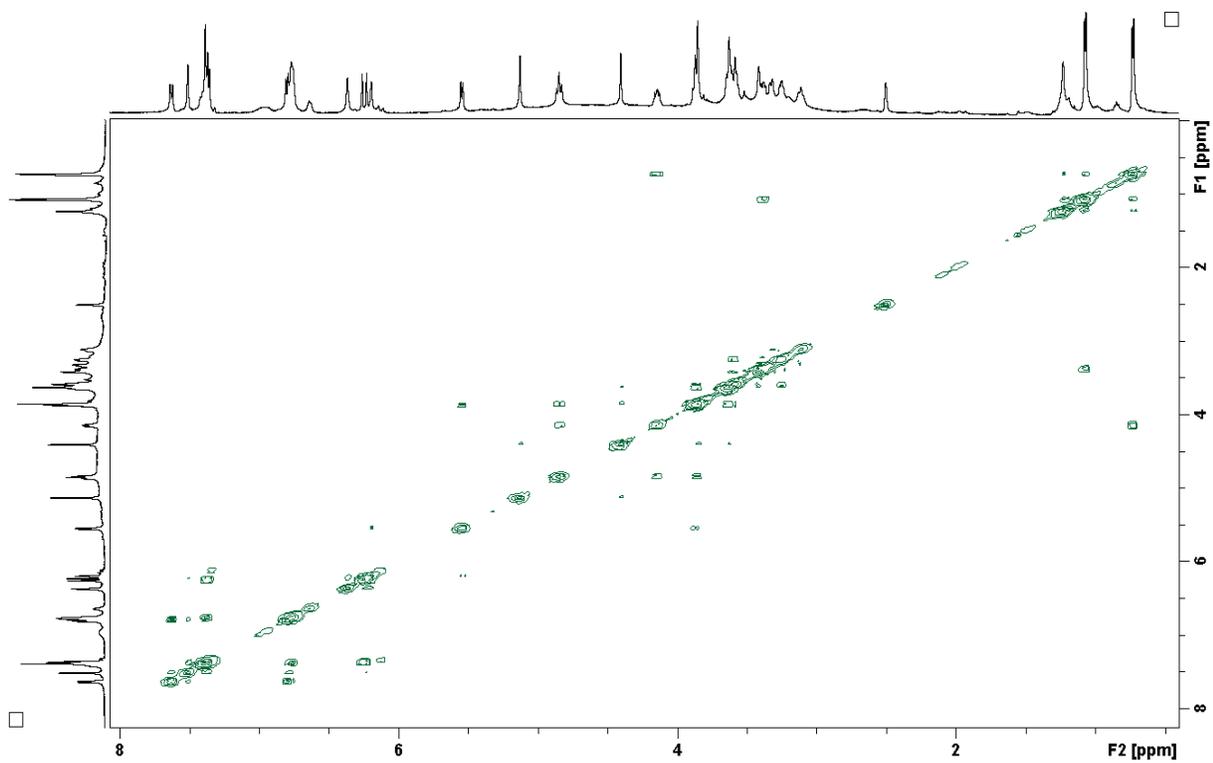


Figure S43: ^1H - ^1H COSY spectrum of compound **11** (500 MHz, DMSO-d_6).

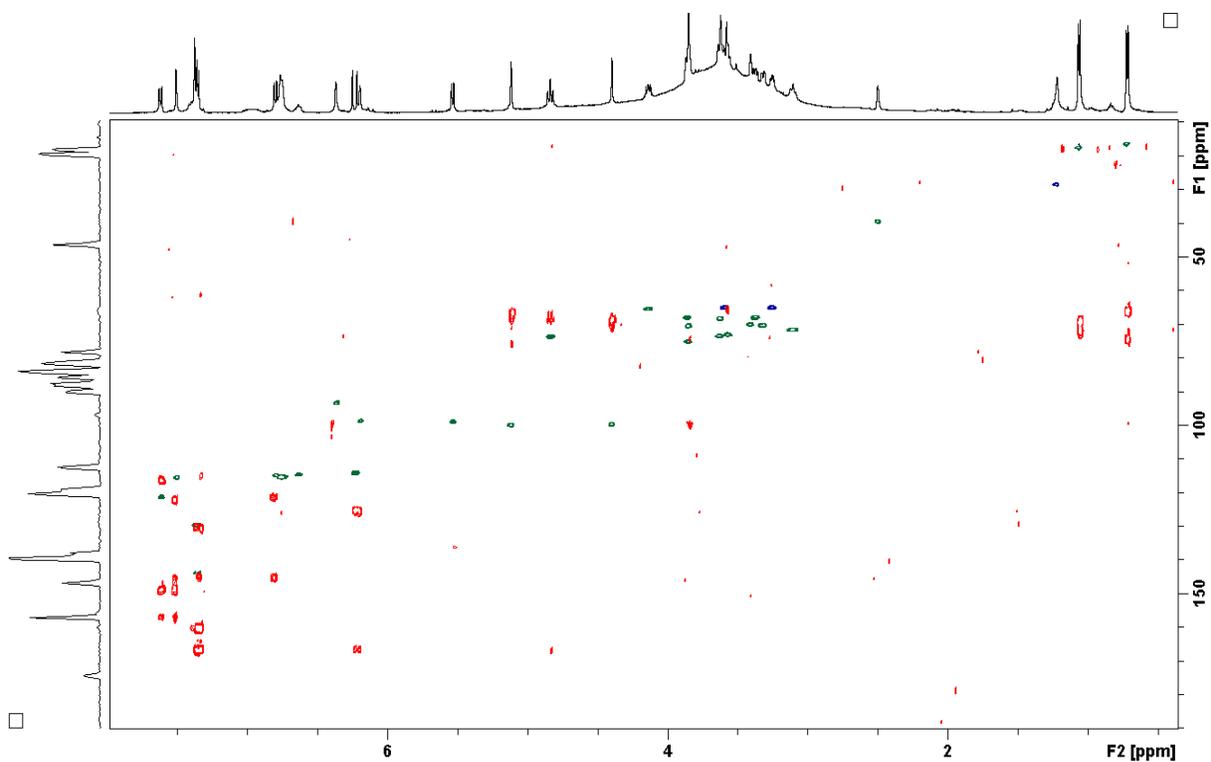


Figure S44: Overlay of HSQC (green-blue) and HMBC (red) spectra of compound **11** (500 and 125 MHz, DMSO- d_6).

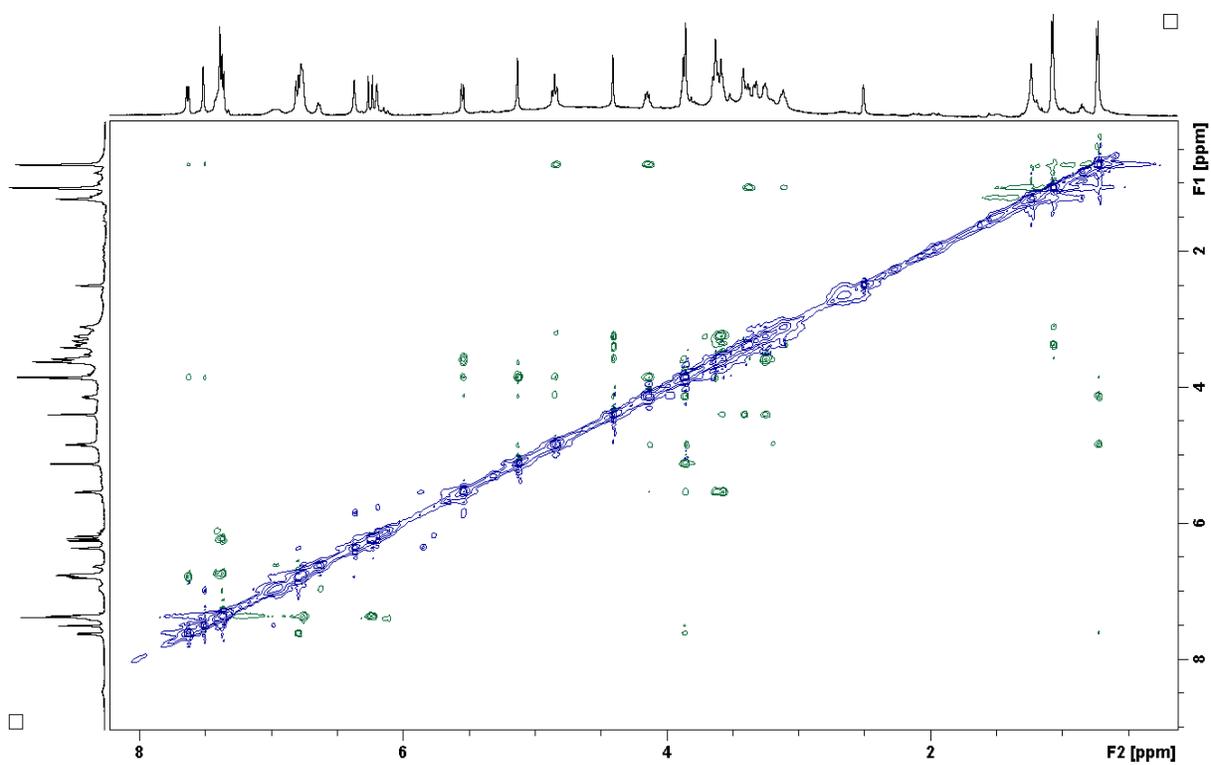


Figure S45: 2D ^1H - ^1H ROESY spectrum of compound **11** (500 MHz, DMSO- d_6).

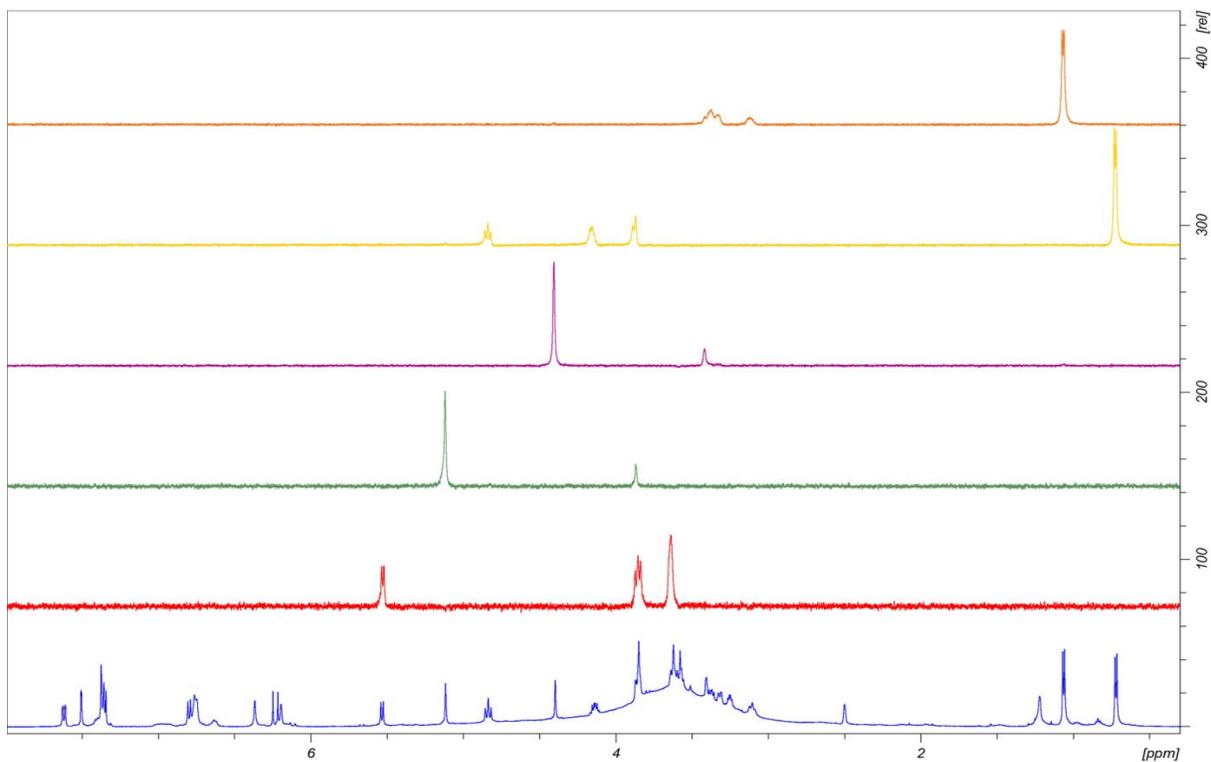


Figure S46: Selective 1D TOCSY spectrum of compound **11** (500 MHz, DMSO-d₆).

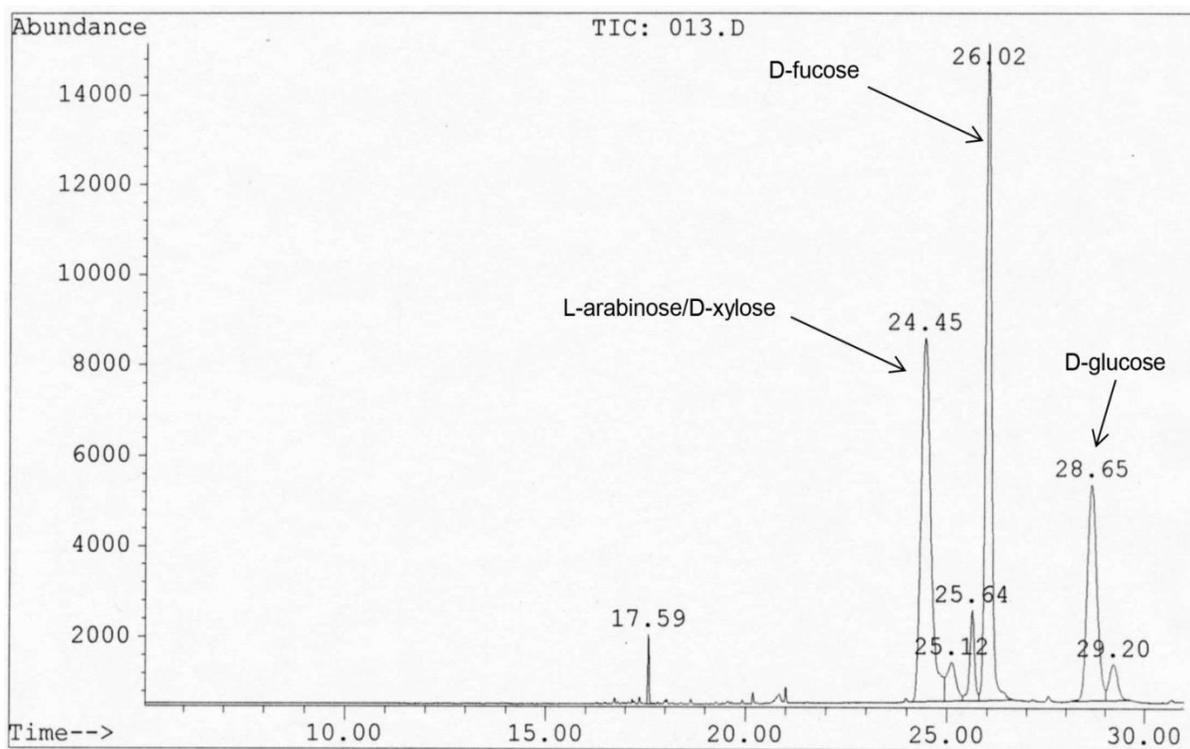


Figure S47: GC-MS chromatogram of hydrolyzed sugars from compound **1**.

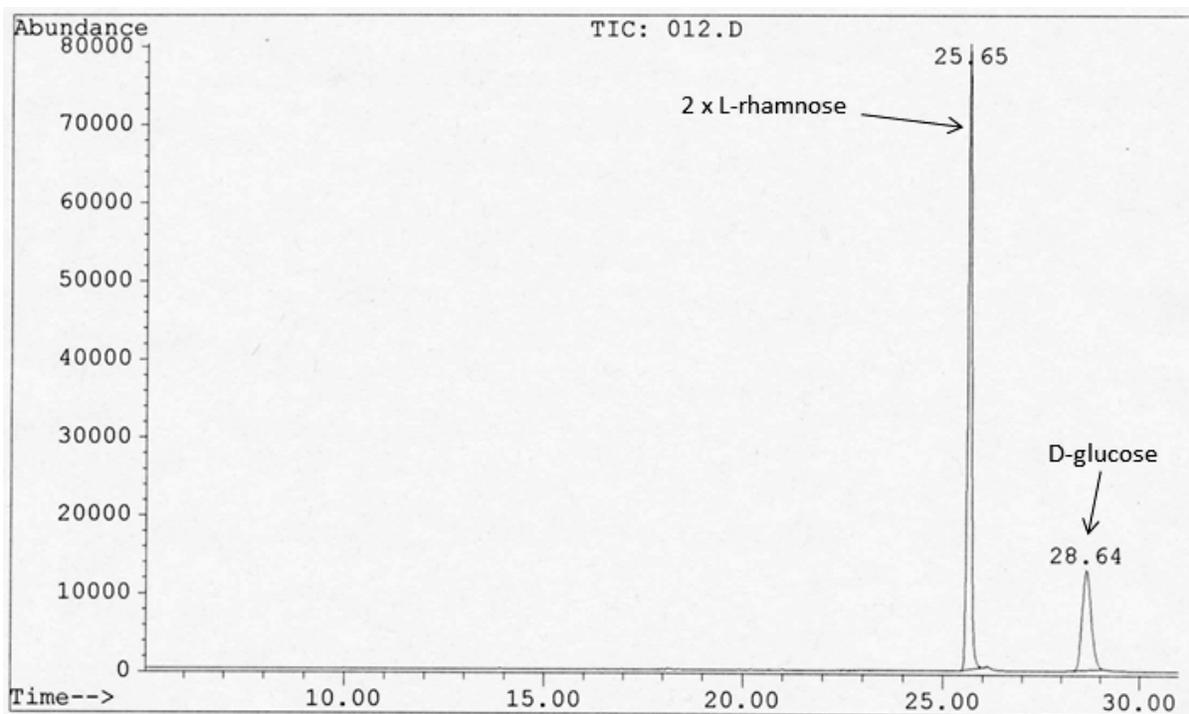


Figure S48: GC-MS chromatogram of hydrolyzed sugars from compound 6.

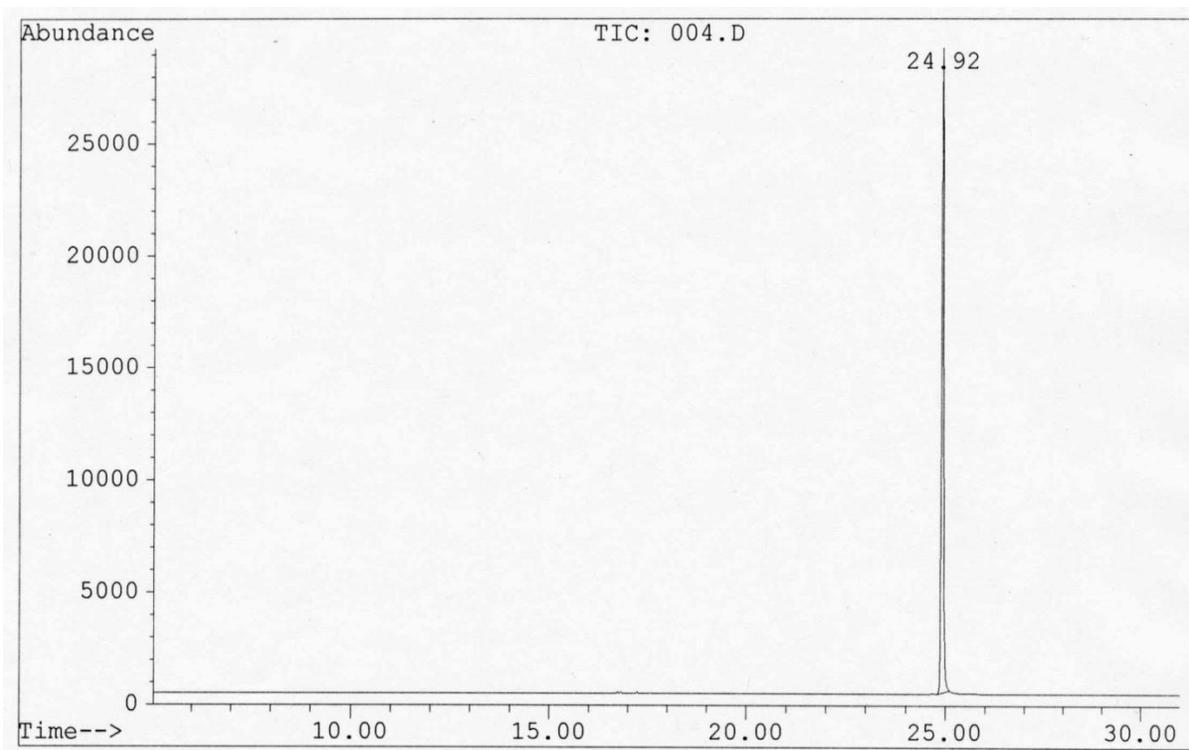


Figure S49: GC-MS chromatogram of derivatized D-arabinose.

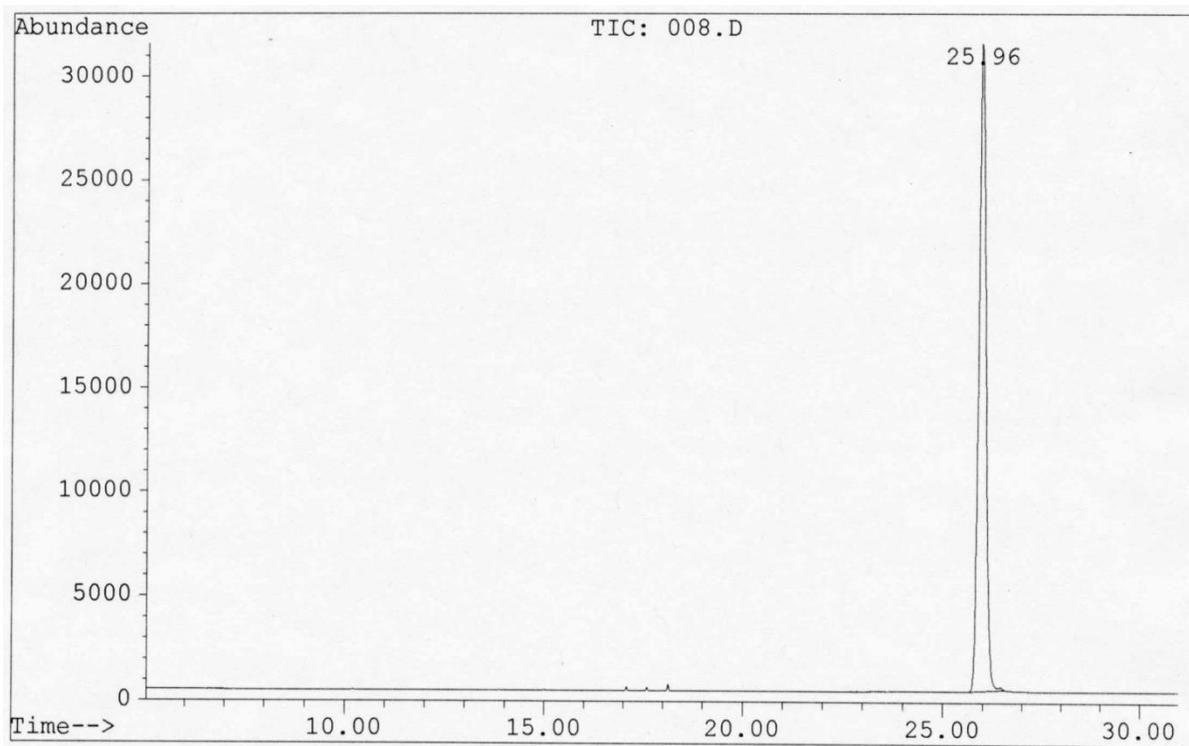


Figure S50: GC-MS chromatogram of derivatized D-fucose.

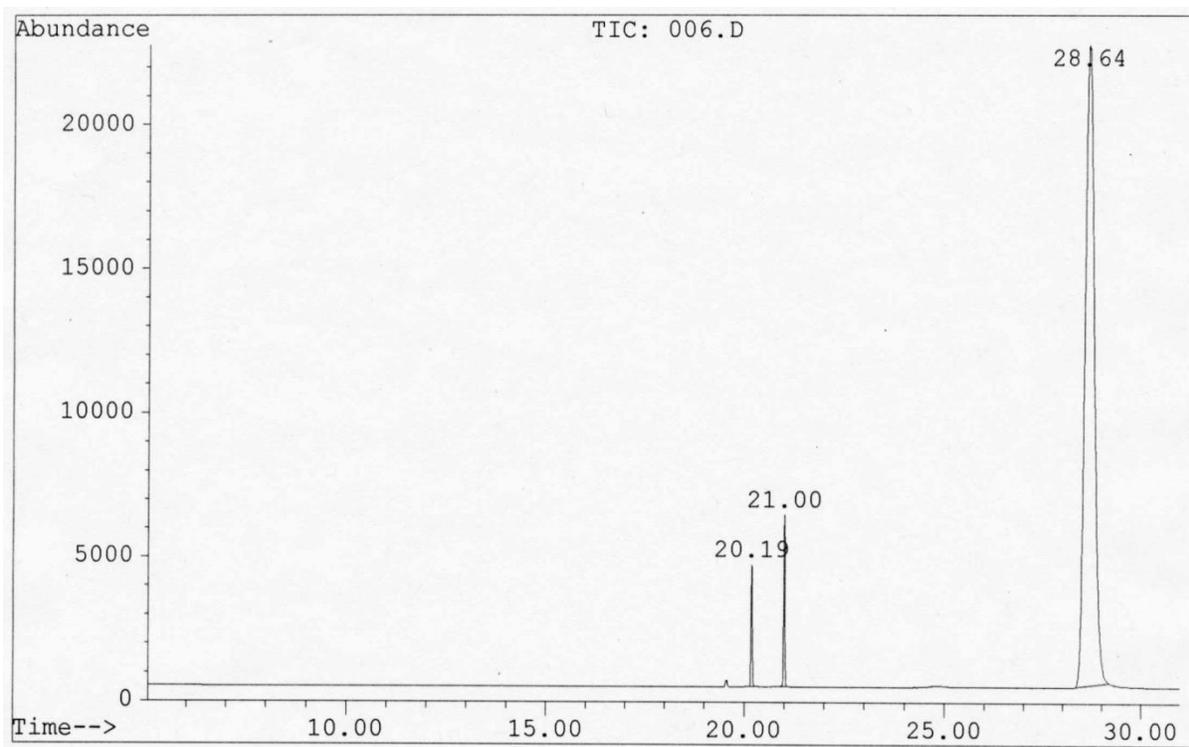


Figure S51: GC-MS chromatogram of derivatized D-glucose.

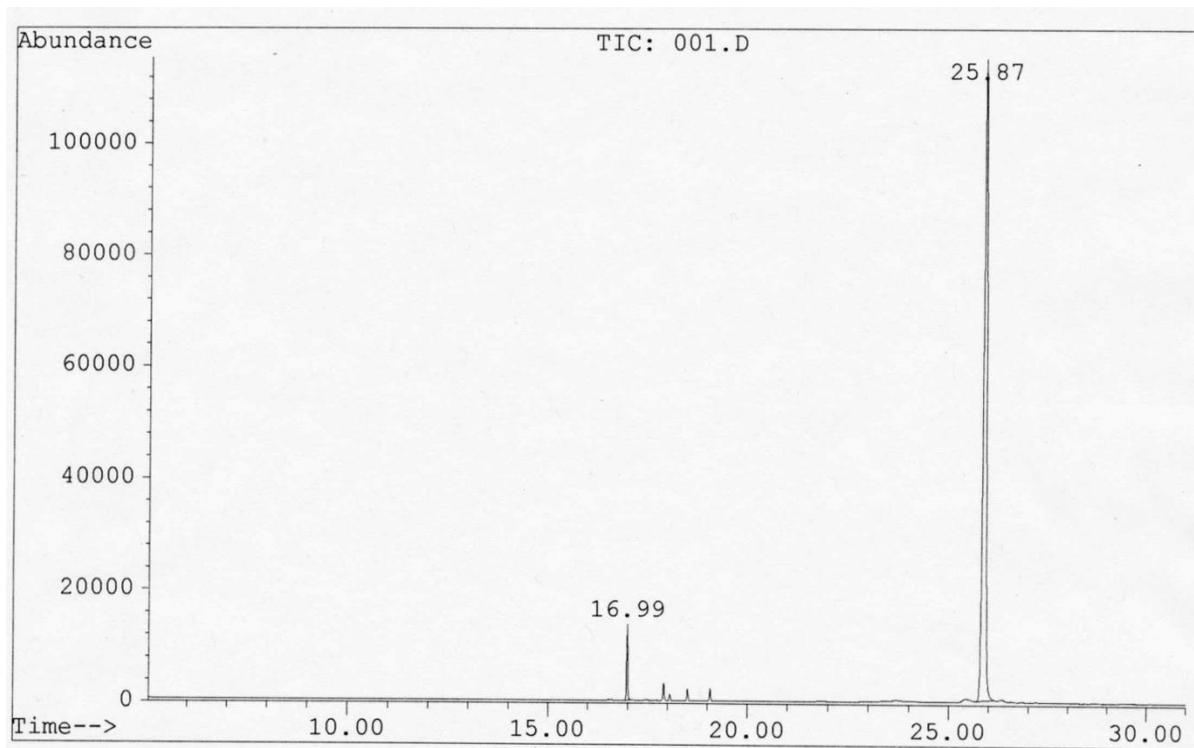


Figure S52: GC-MS chromatogram of derivatized D-rhamnose.

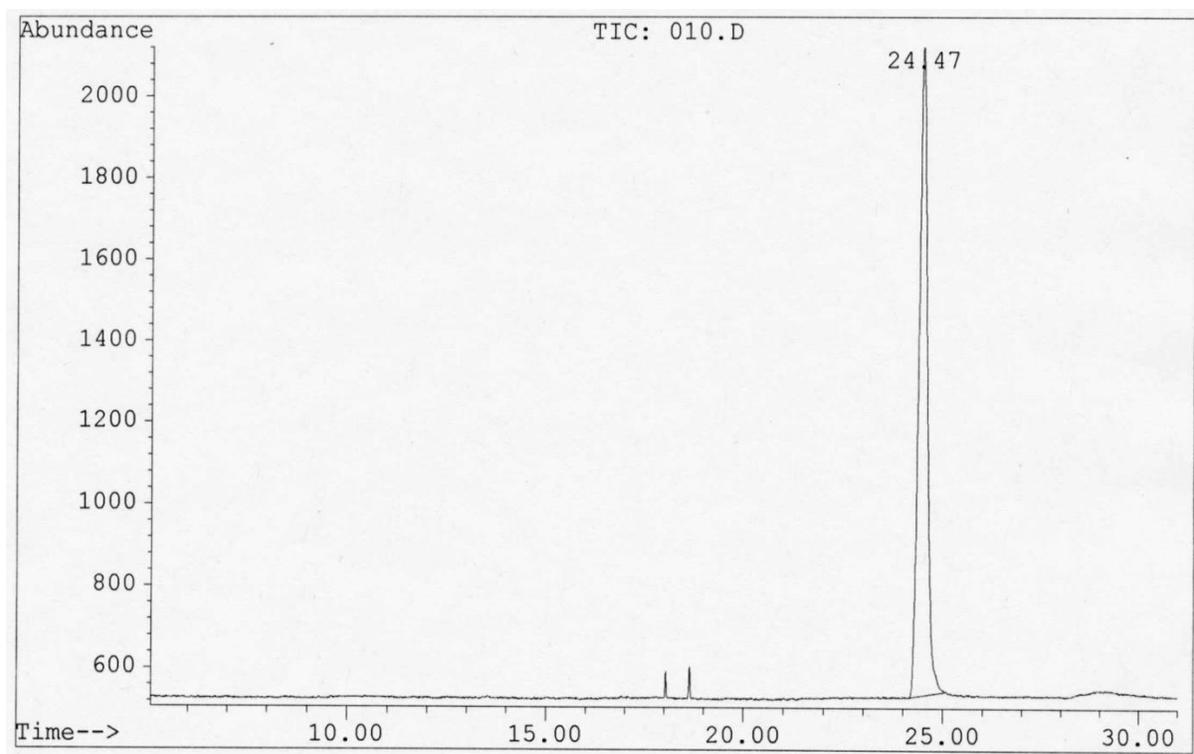


Figure S53: GC-MS chromatogram of derivatized D-xylose.

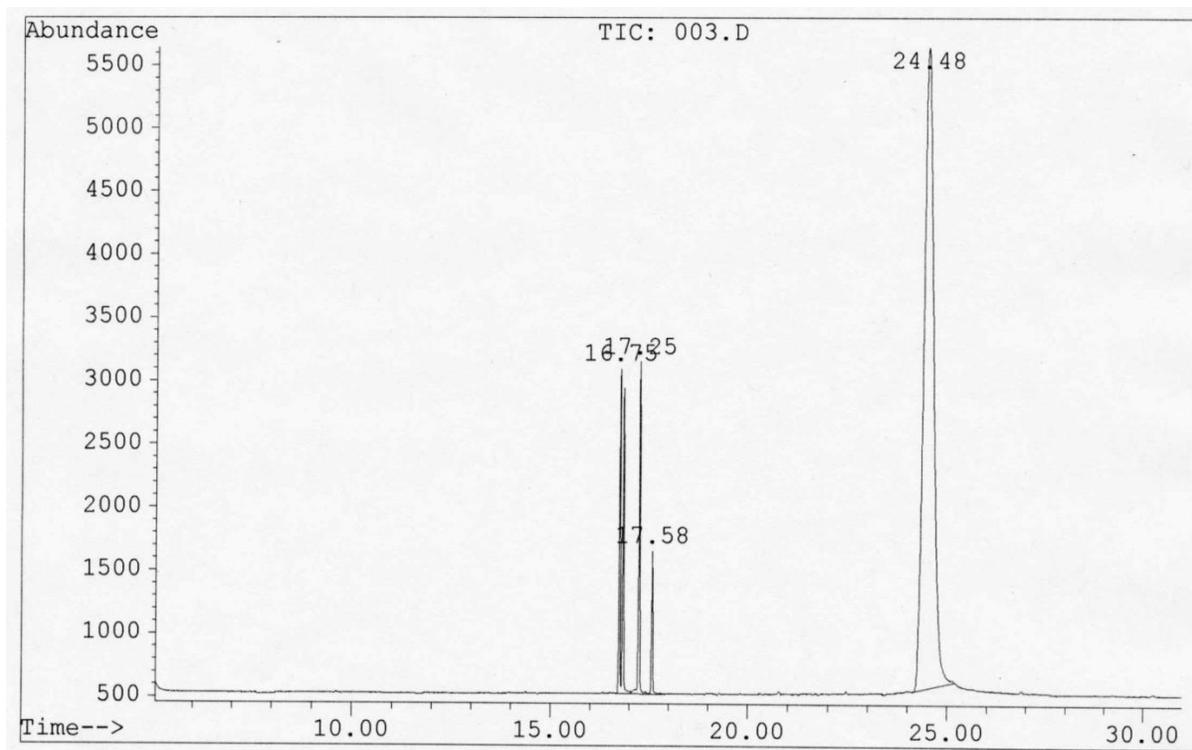


Figure S54: GC-MS chromatogram of derivatized L-arabinose.

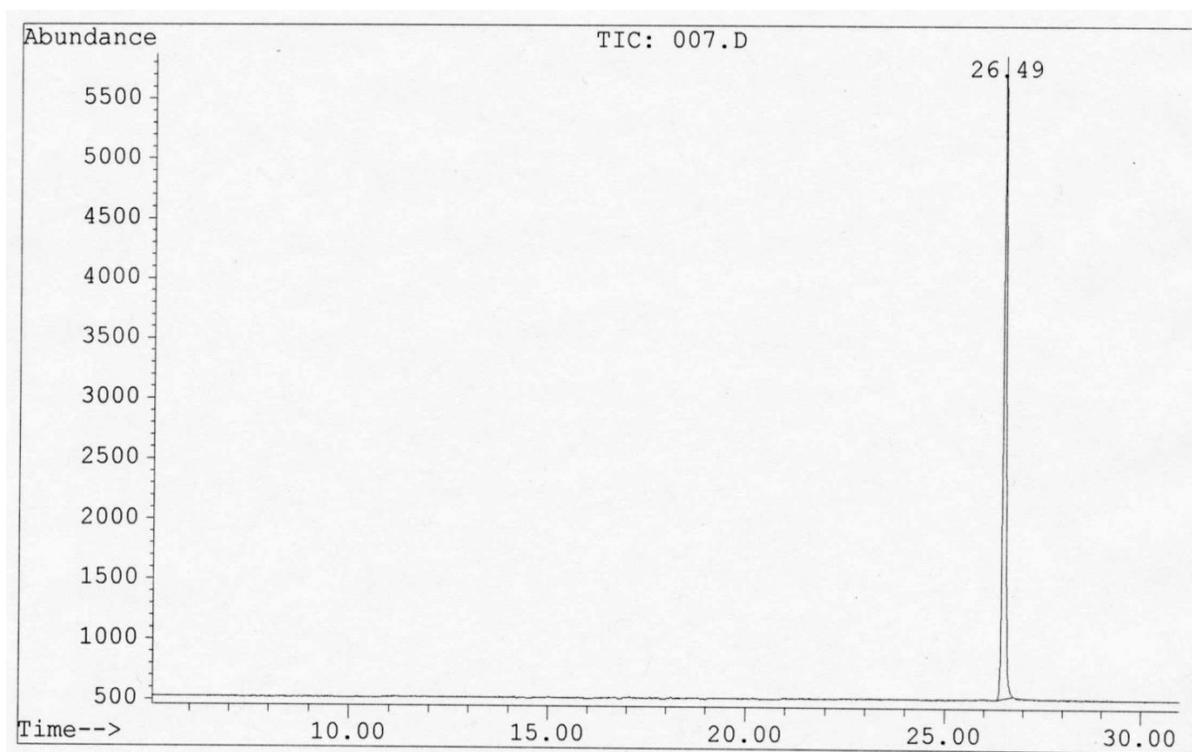


Figure S55: GC-MS chromatogram of derivatized L-fucose.

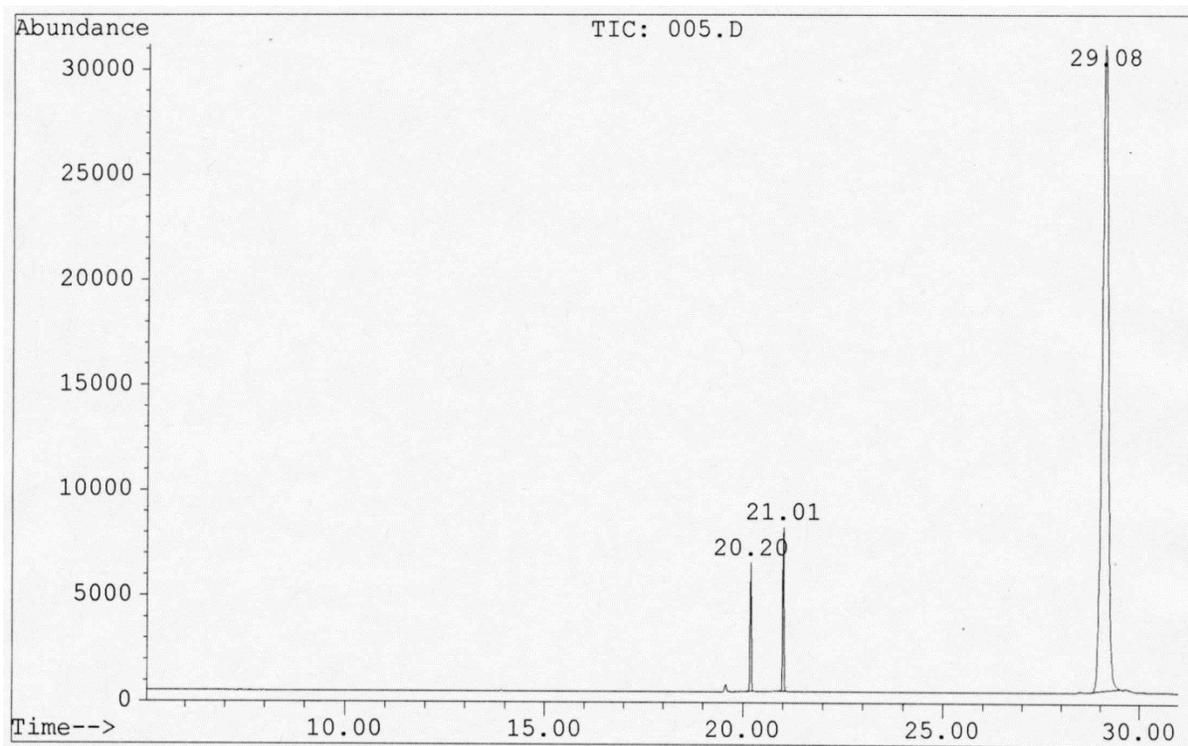


Figure S56: GC-MS chromatogram of derivatized L-glucose.

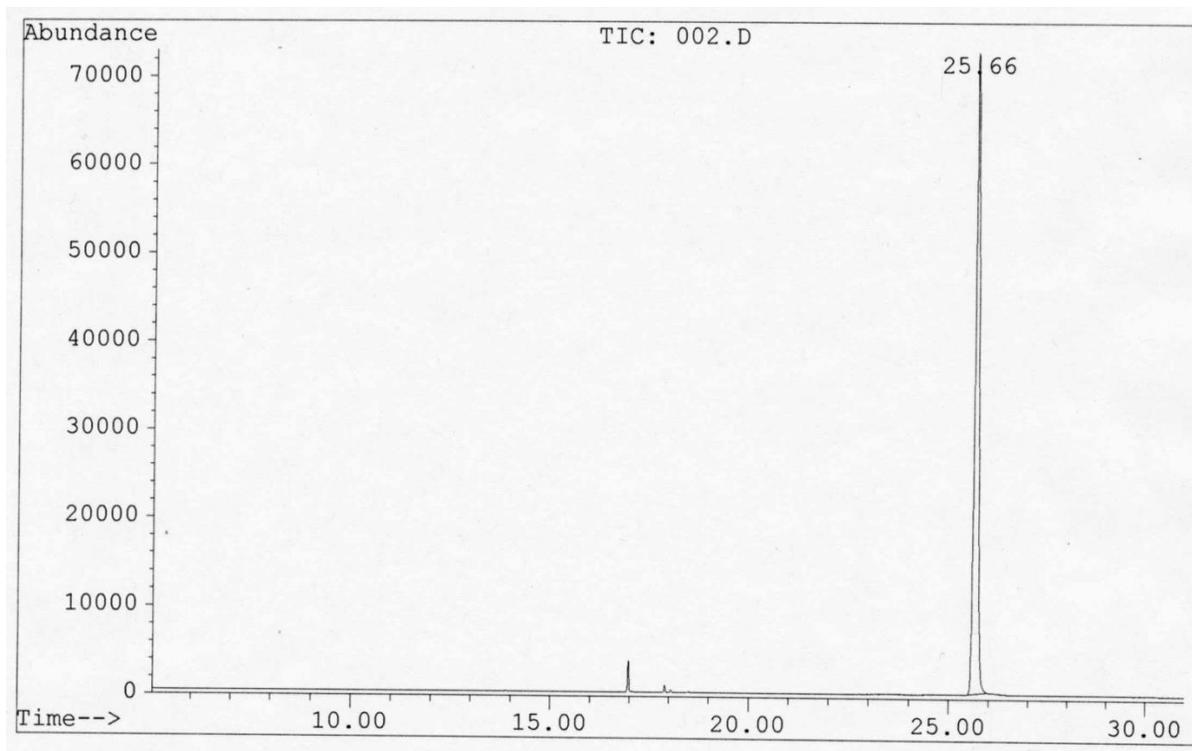


Figure S57: GC-MS chromatogram of derivatized L-rhamnose.

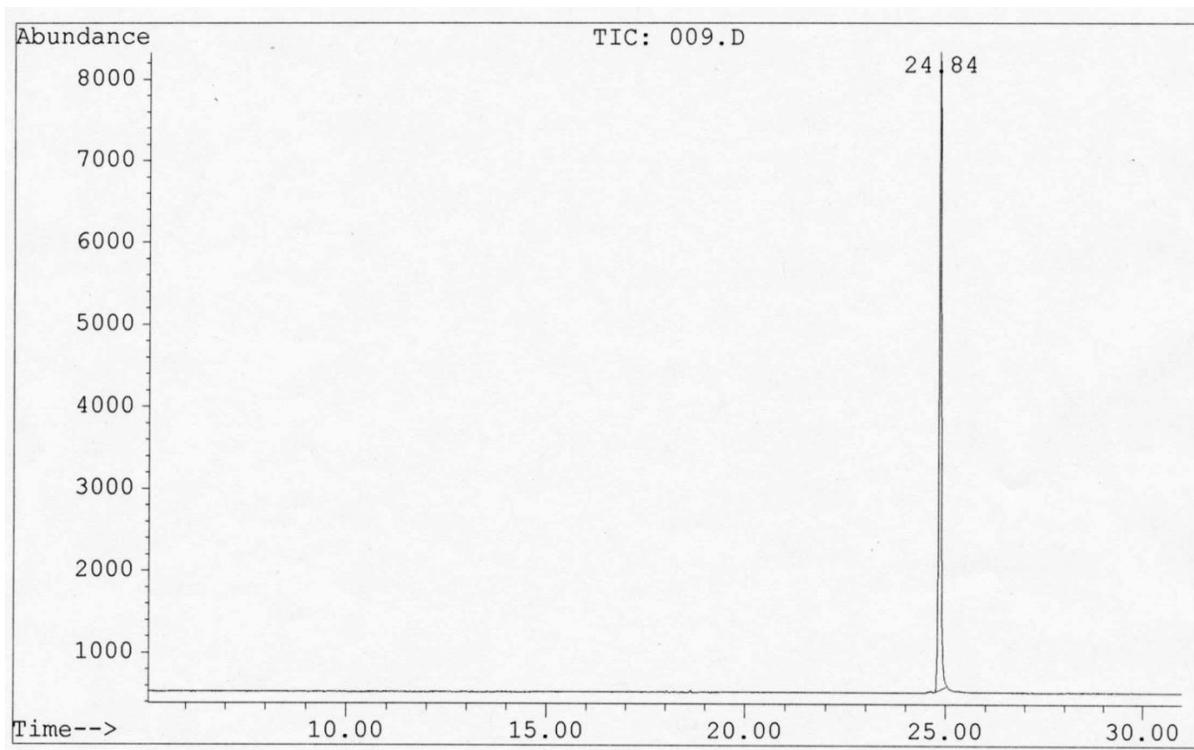


Figure S58: GC-MS chromatogram of derivatized L-xylose.

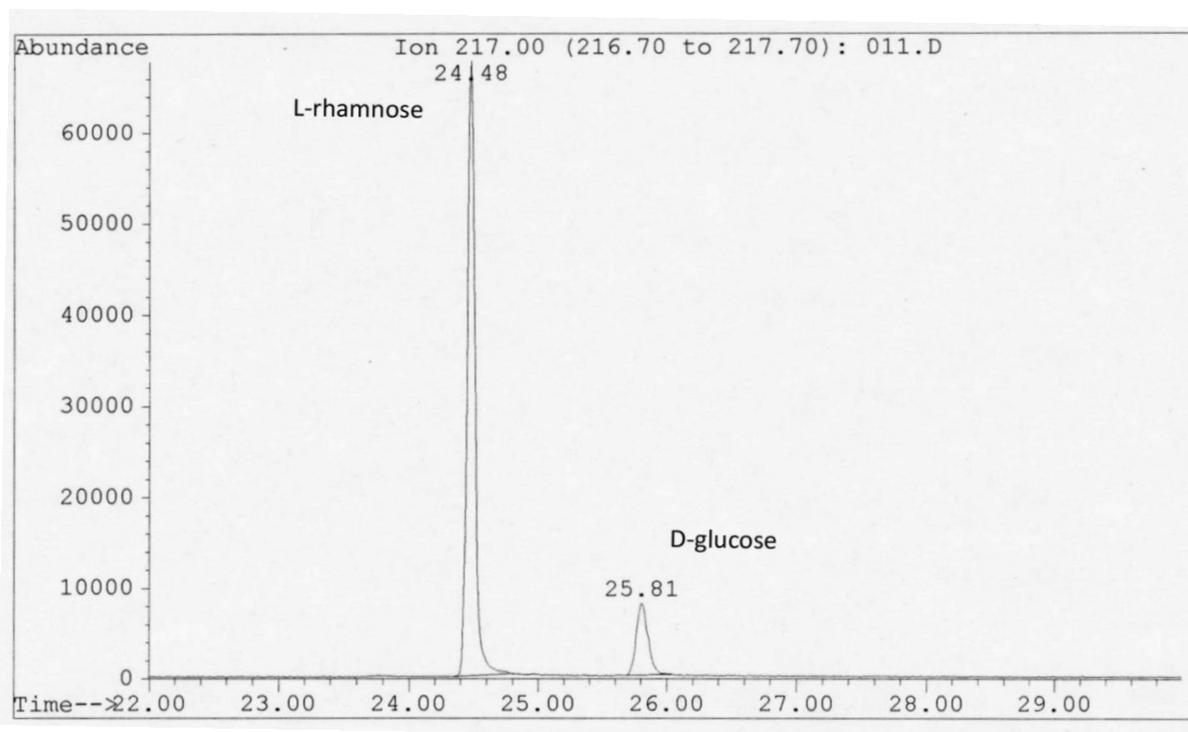


Figure S59: GC-MS chromatogram of hydrolyzed sugars from compound **9**. Sugar retention times in compounds **3** and **6** differ from those in **9** and **11** because have been measured on columns with different length.

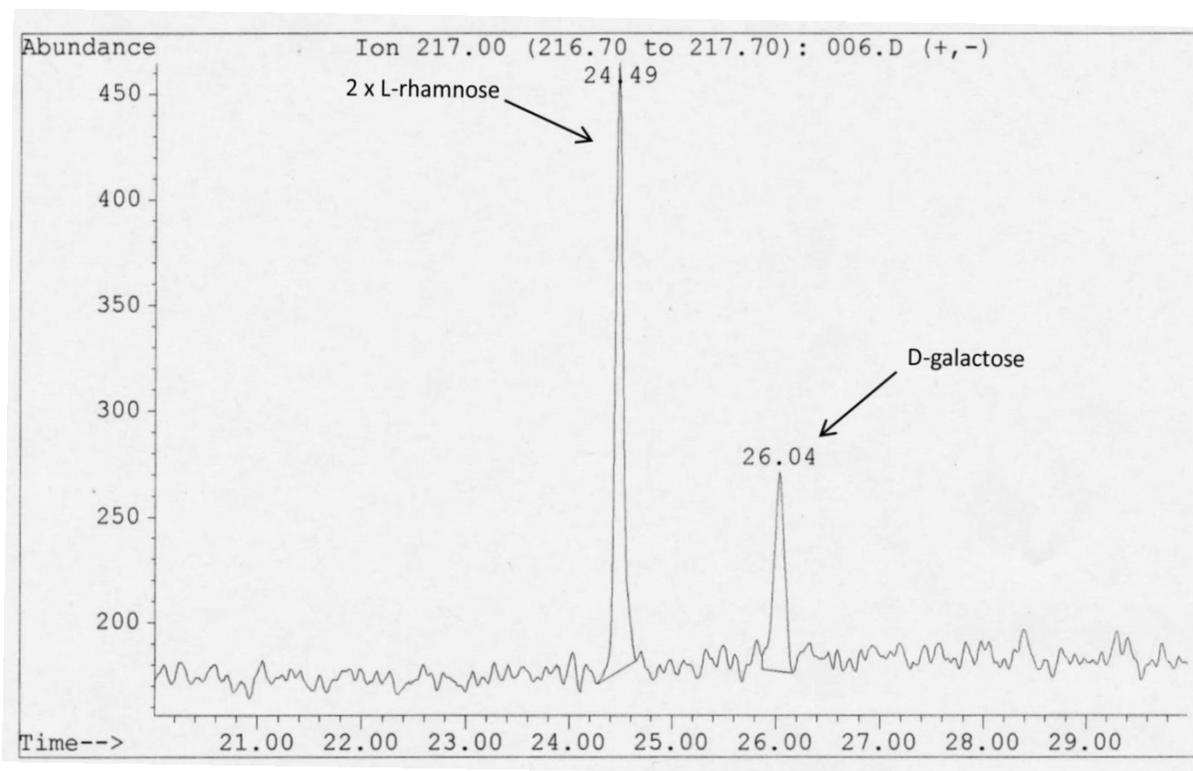


Figure S60: GC-MS chromatogram of hydrolyzed sugars from compound **11**.

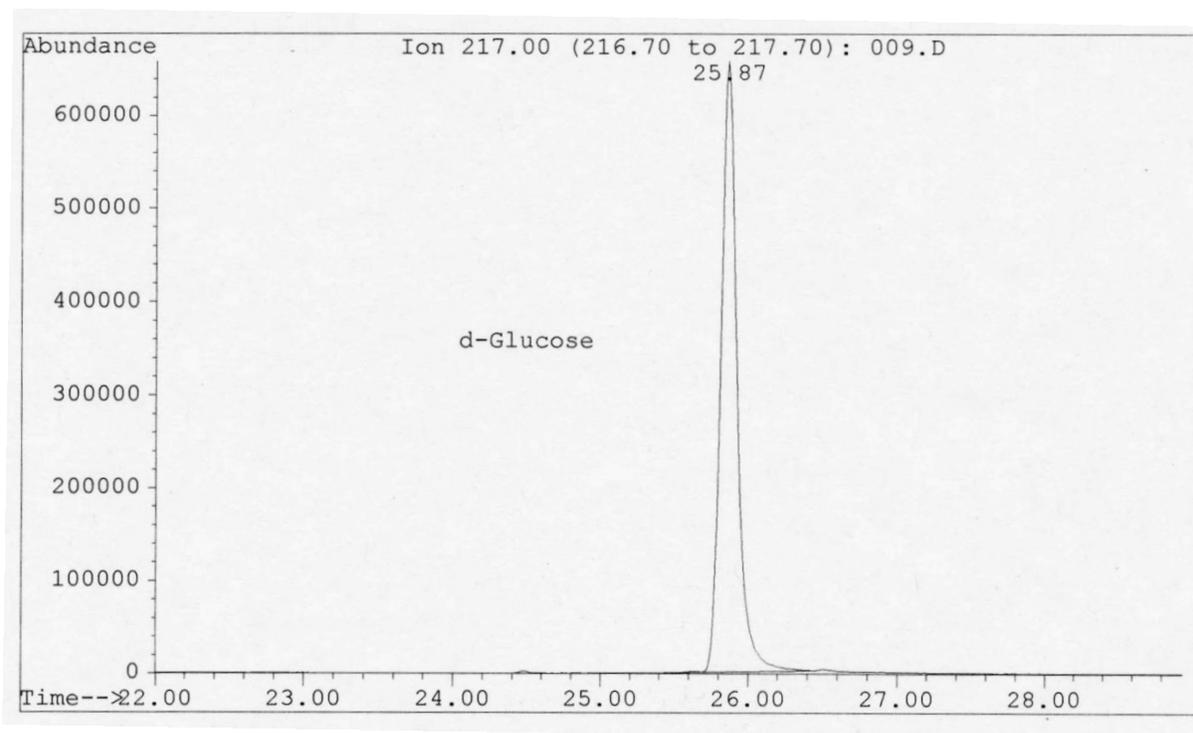


Figure S61: GC-MS chromatogram of derivatized D-glucose.

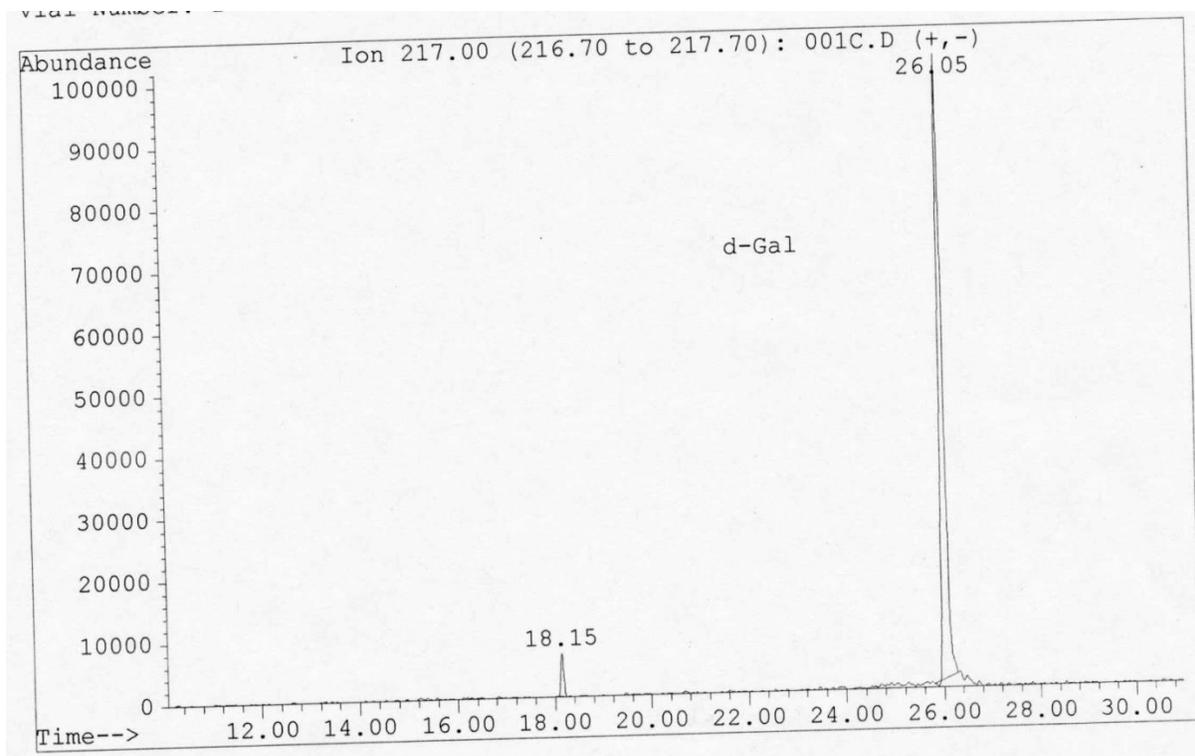


Figure S62: GC-MS chromatogram of derivatized D-galactose.

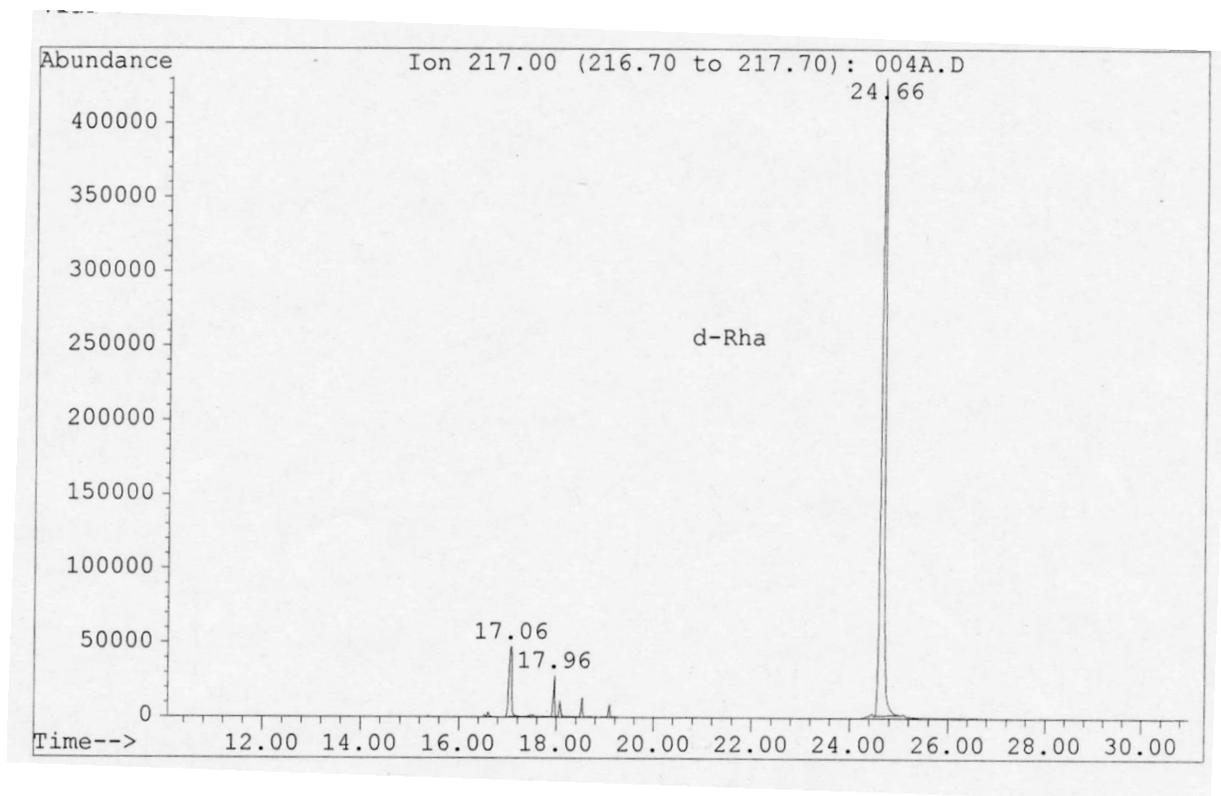


Figure S63: GC-MS chromatogram of derivatized D-rhamnose.

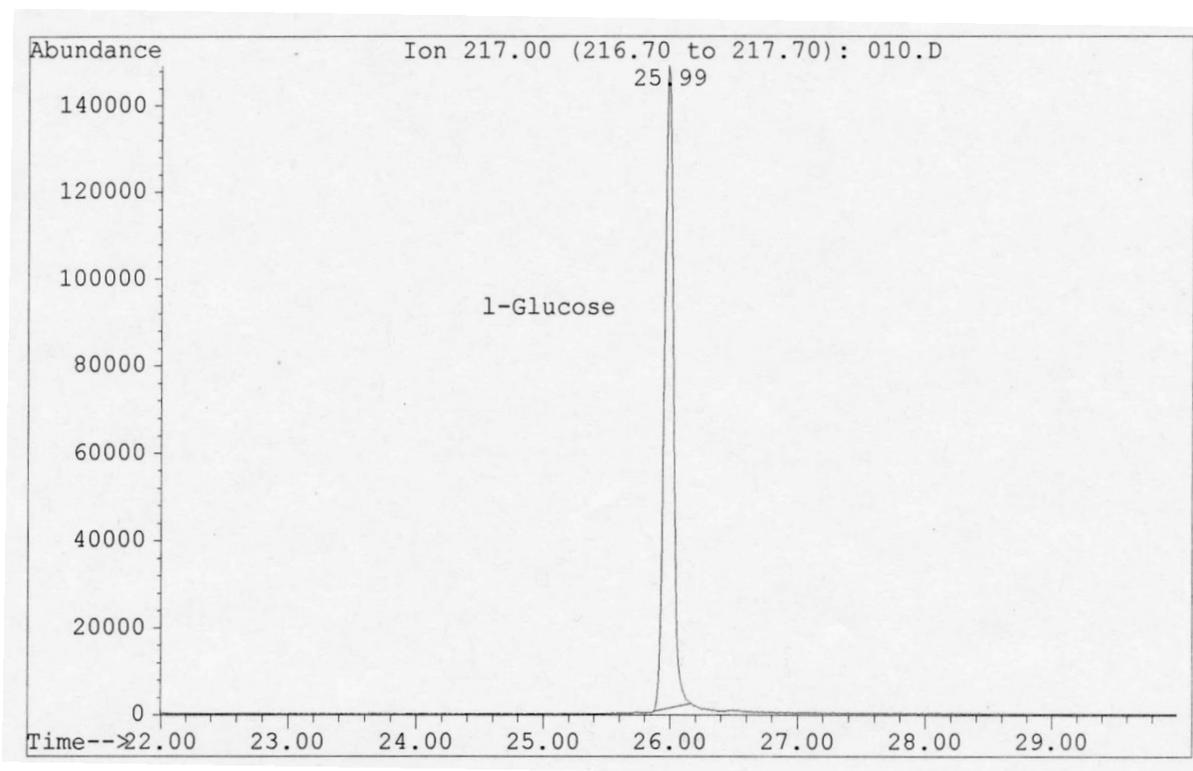


Figure S64: GC-MS chromatogram of derivatized L-glucose.

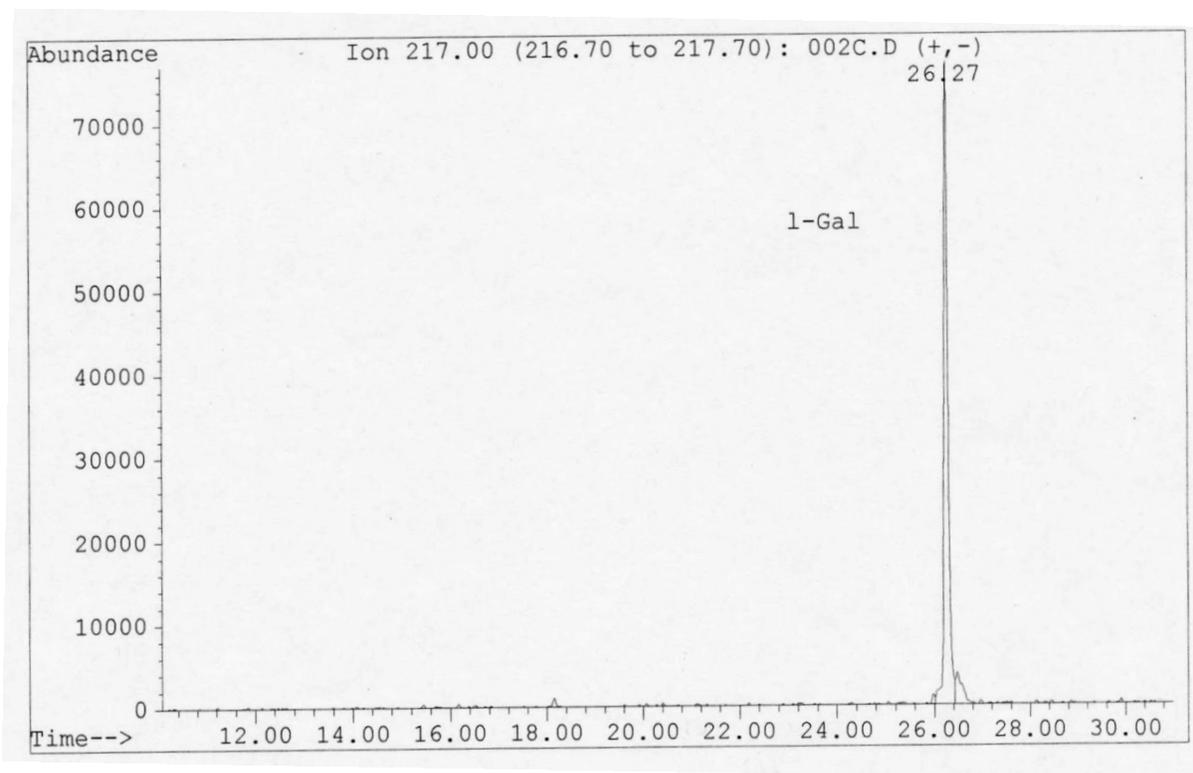


Figure S65: GC-MS chromatogram of derivatized L-galactose.

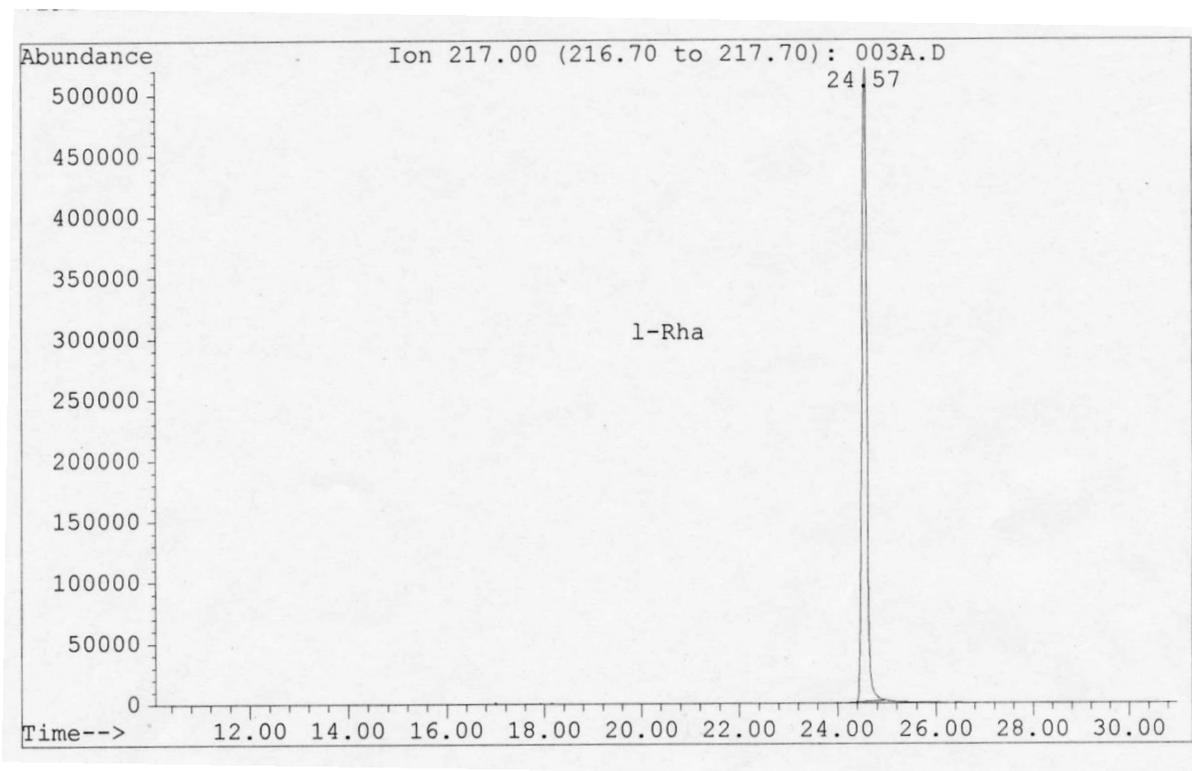


Figure S66: GC-MS chromatogram of derivatized L-rhamnose.

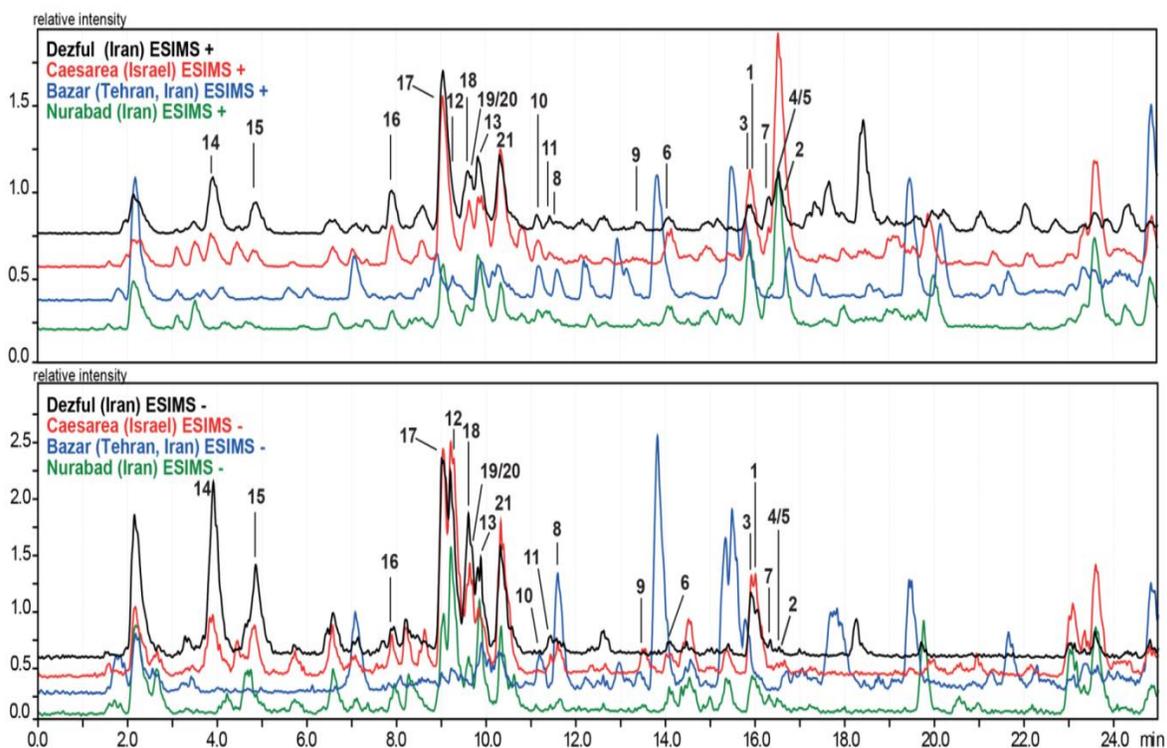


Figure S67: HPLC-ESIMS traces recorded in positive and negative ion mode of *Ziziphus spina-christi* of four different origins. Peak numbering designates compounds 1-21.

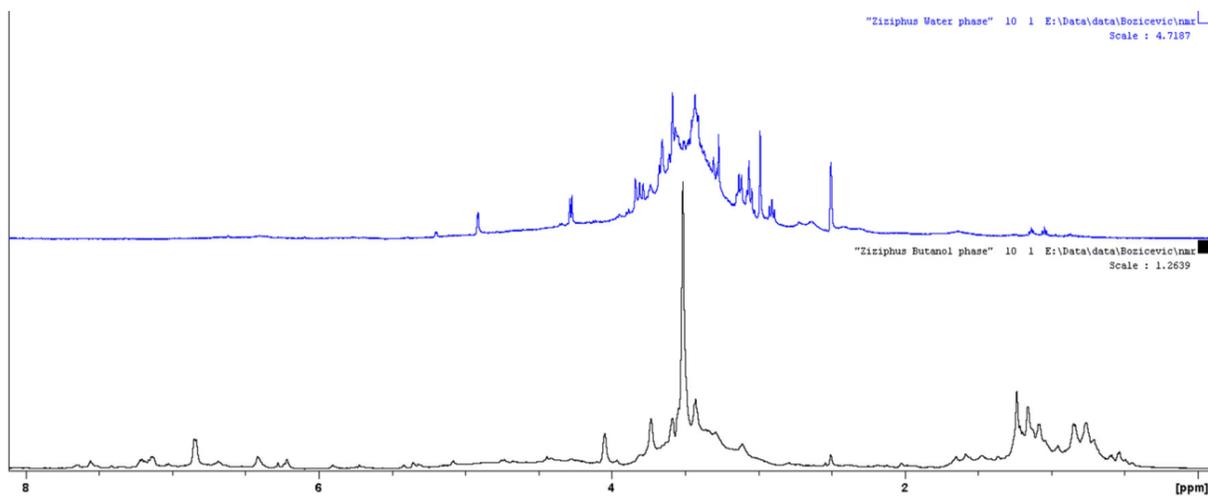
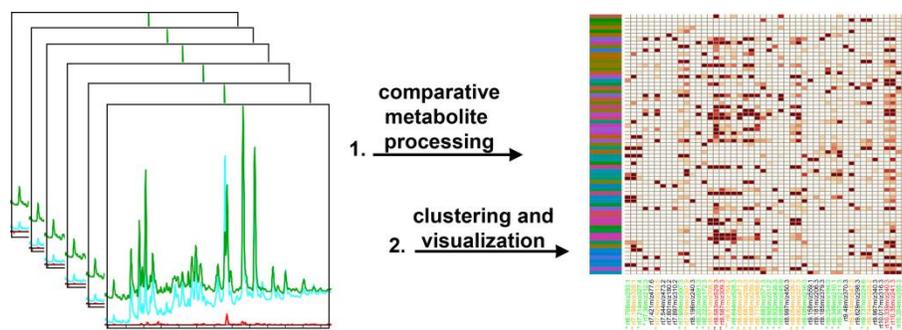


Figure S68: ¹H-NMR spectrum of water and *n*-BuOH phases (500 MHz, DMSO-d₆).

3.2 Automated comparative metabolite profiling of large LC-ESIMS datasets in an ACD/Labs add-in, and data clustering on a new open-source web platform FreeClust

Alen Bozicevic, Maciej Dobrzynski, Hans De Bie, Eliane Garo, Frank Gafner, Matthias Hamburger

Anal. Chem. **2017**, 89, 12682–12689. doi: 10.1021/acs.analchem.7b02221.



A two-steps protocol comprising a comparative metabolite profiling tool integrated in ACD/Labs, and a web platform FreeClust designed for clustering and visualization of the results in 2D heatmaps was developed. This versatile two-steps data mining workflow provides valuable add-in processing options applicable in different research fields such as natural products and early drug discovery research.

Preparation of honey samples (solid phase extraction on C18 EC cartridges), HPLC profiling, and preparation of the figures were my contribution. Writing of the manuscript draft was my contribution with the support of Eliane Garo. Development of the custom pop-up form in the ACD/MS Workbook Suite was performed by Hans De Bie and colleagues with the contribution of Eliane Garo and mine. The development of the on-line clustering platform FreeClust was done by Maciej Dobrzynski. Eliane Garo and I participated in decision making regarding the implementation of statistical functions and design of FreeClust.

Alen Bozicevic

Automated Comparative Metabolite Profiling of Large LC-ESIMS Data Sets in an ACD/MS Workbook Suite Add-in, and Data Clustering on a New Open-Source Web Platform FreeClust

Alen Božičević,[†] Maciej Dobrzyński,[‡] Hans De Bie,^{||} Frank Gafner,[#] Eliane Garo,[†] and Matthias Hamburger^{*,†}

[†]Division of Pharmaceutical Biology, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland

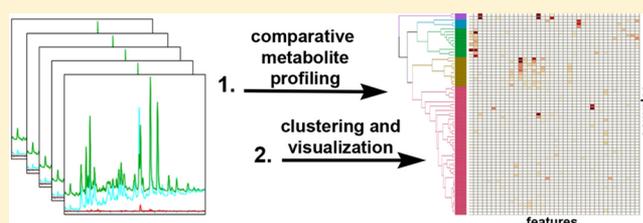
[‡]Institute of Cell Biology, University of Bern, Baltzerstrasse 4, 3012 Bern, Switzerland

^{||}Advanced Chemistry Development, Inc., 8 King Street East Suite 107, Toronto, Ontario M5C, Canada

[#]Mibelle Biochemistry, Mibelle AG, Bolimattstrasse 1, 5033 Buchs, Switzerland

Supporting Information

ABSTRACT: The technological development of LC-MS instrumentation has led to significant improvements of performance and sensitivity, enabling high-throughput analysis of complex samples, such as plant extracts. Most software suites allow preprocessing of LC-MS chromatograms to obtain comprehensive information on single constituents. However, more advanced processing needs, such as the systematic and unbiased comparative metabolite profiling of large numbers of complex LC-MS chromatograms remains a challenge. Currently, users have to rely on different tools to perform such data analyses. We developed a two-step protocol comprising a comparative metabolite profiling tool integrated in ACD/MS Workbook Suite, and a web platform developed in R language designed for clustering and visualization of chromatographic data. Initially, all relevant chromatographic and spectroscopic data (retention time, molecular ions with the respective ion abundance, and sample names) are automatically extracted and assembled in an Excel spreadsheet. The file is then loaded into an online web application that includes various statistical algorithms and provides the user with tools to compare and visualize the results in intuitive 2D heatmaps. We applied this workflow to LC-ESIMS profiles obtained from 69 honey samples. Within few hours of calculation with a standard PC, honey samples were preprocessed and organized in clusters based on their metabolite profile similarities, thereby highlighting the common metabolite patterns and distributions among samples. Implementation in the ACD/Laboratories software package enables ulterior integration of other analytical data, and *in silico* prediction tools for modern drug discovery.



High performance and ultrahigh performance liquid chromatography coupled to mass spectrometry (HPLC/UHPLC-MS/HRMS) are widely used for the analysis of different kinds of complex samples, such as foods, environmental samples, biological fluids, and natural product extracts. LC-MS is a highly versatile and powerful analytical tool, because it enables analysis of trace components, qualitative and quantitative analyses of single compounds in complex mixtures, as well as metabolite fingerprinting.^{1,2} Significant advances in LC-MS hardware have been achieved over the past 25 years with respect to sensitivity, robustness, reliability, automation, and speed of data acquisition. Large sets of complex samples are measured nowadays routinely by LC-MS, thereby producing large data sets in a relatively short period of time.^{3–5}

In recent years efforts have been made in the development of software algorithms and tools to efficiently mine data from the wealth of data generated by LC-MS.⁶ Different software suites are used for the preprocessing of raw chromatograms, which typically includes peak picking and chromatogram deconvolu-

tion.^{7,8} Most instrument vendors provide software enabling data acquisition and processing in one place. However, data processing is limited to their proprietary data format. Free software suites, such as XCMS and MZmine 2, are powerful alternatives for expert users who do not need major technical support. Platforms such as ACD/Spectrus offer a flexible interface which accepts most file formats and is appropriate for users working with different analytical techniques (i.e., NMR, IR, LC/UV/MS) and instruments from different vendors. However, most of these software suites only support part of the processing workflow, and users have to rely on different tools and algorithms to perform more advanced tasks, such as peak identification and dereplication, or comparison of different LC-MS profiles, clustering, and data visualization.

Received: June 9, 2017

Accepted: October 31, 2017

Published: October 31, 2017

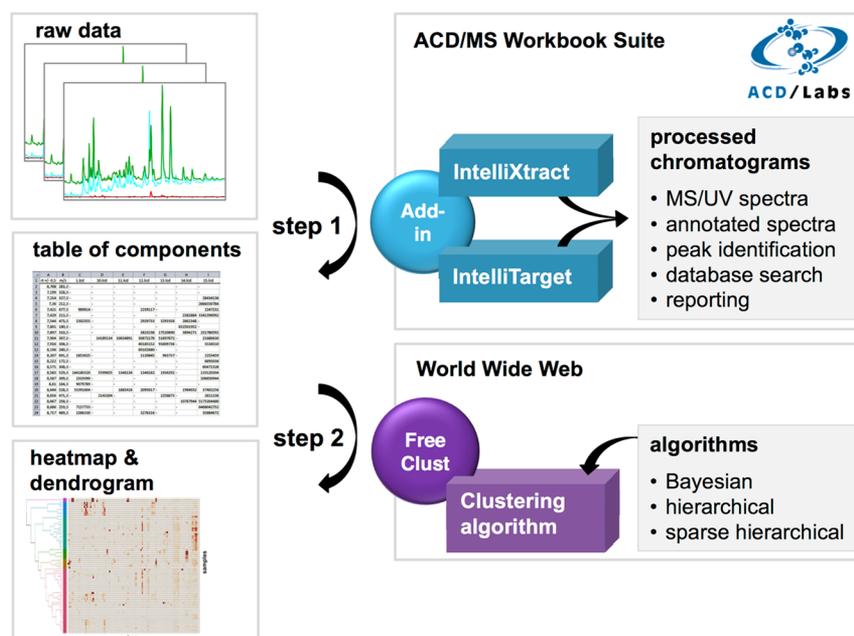


Figure 1. Schematic representation of the proposed two steps in the LC-ESIMS data mining workflow. Step 1: LC-ESIMS chromatograms are loaded into the add-in tool that is using the IntelliXtract (IX) and IntelliTarget (ITA) algorithms. Results are presented as a matrix table (table of components). The processed chromatograms are saved as separate Spectrus files and are accessible in ACD/MS Workbook Suite for additional processing, databasing, and reporting. Step 2: Matrix table is loaded into the FreeClust web application for clustering and graphical representation. Results are displayed as heatmap for clustering. Bayesian, hierarchical, and sparse hierarchical clustering algorithms are available.

Peak identification and dereplication is a major challenge, especially in the field of natural product research. Hence, a number of dereplication workflows have been developed to address this important issue.^{7–9} The Global Natural Product Social Molecular Network (GNPS), a free platform used by more than 10 000 users, is probably one of the most successful examples, as it streamlines the dereplication effort through a common leverage of raw and processed MS/MS data.¹⁰ The ACD/Laboratories platform also provides interesting tools for dereplication via in-house and commercial MS and NMR spectral database searches.

Comparative processing of large groups of LC-MS data sets for the purpose of pattern identification and sample classification remains another major challenge in the field.^{11–16} Such types of advanced interrogation of large data sets require (i) additional cross-comparison of individual chromatograms and (ii) the identification of metabolite patterns and subsequent data clustering. The free metabolomics platforms such as XCMS¹⁷ and MZmine 2,¹⁸ used for preprocessing, offer some basic postprocessing tools, such as PCA, and could be used for comparative analyses. They are creating so-called mapped metabolic networks, used to identify biomarkers to support the diagnostic and monitoring of various diseases.^{19–21} They were also used to highlight gene-to-metabolite networks in plants and allowed the elucidation of a specific natural product biosynthesis pathway.^{22,23} Similar metabolic mapping approaches have been recently developed for planar chromatography using rTLC.²⁴ The web server MetaboAnalyst 3.0 supports an extensive range of tools for data analysis and can be used for more advanced statistical evaluations.²⁵ However, it is more appropriate for expert users. Because of the diversity of the processing algorithms and tools, the user often has to implement different software suites in his workflow. This can be a major bottleneck for users, especially for nonexperts in the field of metabolomics.

We implemented ACD/Spectrus in our natural product discovery setting, because this software suite provides an optimal platform for processing and evaluation of all our different analytical data in a single environment. We here propose a two step data mining approach for the comparative processing of large groups of LC-MS data. In an attempt to develop a workflow that is straightforward and user-friendly, a first processing step consists of a customized add-in tool developed within the ACD/Spectrus platform. The second processing step includes a new open access web-based application clustering tool called FreeClust. FreeClust was developed in the R programming language, and its interactive design allows for various modes of visualization to facilitate data interpretation. This two step data mining approach was then tested with a set of HPLC-ESIMS chromatograms originating from 69 honey samples.

EXPERIMENTAL SECTION

LC-MS Separation. LC-MS analysis was performed on a Shimadzu Prominence HPLC system composed of a LC-20 AD binary pump, CTO-20AC column thermostat, SDP-M20A PDA detector, CBM-20A system controller, and coupled to an LCMS-8030 Triple Quadrupole Mass spectrometer equipped with an electrospray ionization source (ESI). Conditions were as follows: capillary voltage, 4.5 V; desolvation line temperature, 250 °C; heat block temperature, 500 °C; drying gas (nitrogen) flow, 15 L min⁻¹; nebulizing gas (nitrogen) flow, 3 L min⁻¹. For data acquisition in both positive and negative ionization modes, full-scan (160–1500) was carried out with 6000 u/sec scan speed and 0.150 s per event time. LC separation was carried out on a C₁₈ SunFire column (3.5 μm, 3 × 150 mm i.d., Waters) equipped with a guard column (3 × 20 mm i.d.). HPLC solvents were kept in Teflon FEP bottles (Nalgene, ThermoFisher) with 5 μm PTFE in-line filters (Vici). The mobile phase consisted of water (A) and acetonitrile (B),

both containing formic acid (0.1%, v/v), and the flow rate was set to 0.5 mL min⁻¹. The injection volume was 2 μ L. The following gradient was used: 0–2 min 5% B; 2–27 min 5–100% B; 27–32 min 100% B. The equilibration time between each injection was 7.0 min. Data acquisition was performed with LabSolutions software (Shimadzu). Data were processed and analyzed with ACD/Spectrus Processor (ACD/Laboratories) on an Intel Core i7, 2600 CPU at 3.40 GHz personal computer. For the add-in tool, the main parameters used were: Peak Number, 5; tR Threshold, 0.5. For IntelliXtract (IX) Options: ¹²C/¹³C error (%), value 10; CODA, True; Peak S/N Threshold, 6; fwhm (scans), 0.1; Convert Profile to Centroid; Data Analysis Region, Start tR 6 min and End tR 20 min; For IntelliTarget (ITA) Options: Peak S/N Threshold, 200; Peak Area Threshold (counts), 500. Tables S-2 and S-3 provide the entire list of parameters used.

Preparation of Honey Samples. Honey (3.0 g) was dissolved in 20 mL of distilled water (Millipore, Merck) at room temperature (Table S-1). The C₁₈ ec cartridges (Chromabond, 3 mL/500 mg) were eluted with the aid of a SPE vacuum manifold (Visiprep). Samples were loaded onto water preconditioned cartridges, and dried on the vacuum manifold. Cartridges were first rinsed with 3 \times 5 mL distilled water to remove sugars, then eluted with 4 mL HPLC grade methanol (Macron) to obtain a sugar-free fraction. Solvent was removed using a GeneVac EZ-2 plus centrifugal evaporator, and dried residues redissolved in extra pure dimethyl sulfoxide (Scharlau) at a final concentration of 5 mg mL⁻¹.

RESULTS AND DISCUSSION

The two-step data mining workflow was developed for comparative metabolite profiling of large groups of LC-ESIMS chromatograms (step 1), and data clustering and visualization (step 2) (Figure 1). The comparative metabolite profiling is a preprocessing step using a customized add-in tool developed in the ACD/MS Workbook Suite software. The tool manages ACD/IntelliXtract (ACD/IX) and ACD/IntelliTarget (ACD/ITA) modules in sequence to process the loaded chromatograms. ACD/IX uses a generic component finding algorithm called COmponent Detection Algorithm (CODA)²⁶ to reduce noise and background signals in mass spectra, and thereby streamlines the interpretation of data. Chromatograms are deconvoluted in a comprehensive way to enable the identification of trace components and coeluting peaks. Single peaks in a chromatogram are defined by their retention time (RT) and ion mass (m/z), and treated as input values (RT- m/z). Peak recognition in the ACD/IX module takes into account ESI traces recorded in both positive and negative ionization mode, if available. All relevant information for spectral interpretation, such as isotopes, ¹²C/¹³C ratios, adduct ions, multimers, neutral losses, fragment ions, and [M + H]⁺ or [M - H]⁻ ion assignments are extracted by ACD/IX and added to the input values. This provides then a so-called input list for each individual chromatogram. The comparative metabolite profiling add-in tool is then used to compare features (metabolites) between samples and combine them into one major input list. Peaks with same RT- m/z values are considered as duplicates and are, therefore, eliminated at this step. The ACD/ITA module uses this list to check the presence of each individual compound in all data sets. A table of components is then generated with the abundance, expressed as area under the curve (AUC), of detected compounds in all chromatograms.

The open access web-based application FreeClust was designed for clustering and visualization of the results obtained in step 1. The Excel spreadsheet obtained from step 1 is loaded into the FreeClust application which allows for data presentation utilizing hierarchical, sparse hierarchical, and Bayesian clustering methods. Outliers are identified in a histogram to help the user in defining the optimal clustering method. In addition, the importance score factor of each metabolite within the group of data sets is calculated to indicate its weight in the clustering. A color-code is used to show ion abundance in each sample. FreeClust is freely accessible at <http://bioz-lcms-chromclust.bioz.unibas.ch:3838/shiny-freeclust/>.

Data Acquisition. Chromatographic data acquired on HPLC-ESIMS and UHPLC-ESIMS systems are opened in ACD/MS Workbook Suite. Various file formats, such as “.lcd”, “.d”, “.raw”, and “.spectrum” are supported, thus enabling data import from instruments of major manufacturers. Functions for automated baseline subtraction and MS signal normalization were not implemented in the present version of the comparative batch processing tool. These steps were performed manually.

Comparative Metabolite Profiling (Data Mining Step 1). The customized tool developed in ACD/MS Workbook Suite is managed through a pop-up form which allows users to load multiple files to be processed (Figure S-1). This add-in tool is using ACD/IX and ACD/ITA algorithms. Parameters in ACD/IX, ACD/ITA, as well as in the customized pop-up form can be optimized. ACD/IX and ACD/ITA include settings such as detection of adduct ions, peak picking (S/N, peak minimal height, peak area count), or peak recognition parameters (Table S-2, S-3). Furthermore, the following parameters can be selected in the pop-up form for an optimal processing:

- ESI⁺/ESI⁻ defines the ion polarity to be used in the data mining workflow.
- tR Threshold defines the resolution of retention time. Two peaks with the same m/z will be considered as the same input if their RT difference is within the RT threshold.
- no. peaks (number of peaks) designates the maximal number of most abundant peaks per sample to be processed.

If the no. peaks entered is bigger than the total number of peaks identified based on S/N value (ACD/IX option), the latter value will be considered for analysis. The optimal no. peaks is selected empirically and may vary depending on the batch composition. Use of this parameter is illustrated by two theoretical examples (Figure 2). The relative abundance of molecular ions present in each sample is represented by different shades of blue. The batch in the first example comprises two samples which contain four and five peaks, respectively (Figure 2A). Compounds identified in each sample are all different which results in a total of 9 peaks in the batch. If no. peaks is set to 2, the total number of peaks found in the batch would be 4 (peaks A2, A3, B5, and B9). The remaining 5 peaks would not be integrated. If no. peaks is set to 4, all peaks in sample A and peaks B5 and B7–B9 would be integrated. Peak B6 would not be considered although being more abundant than peak A1.

The batch in the second example contains six samples with a total of ten different peaks, and varying numbers of peaks per

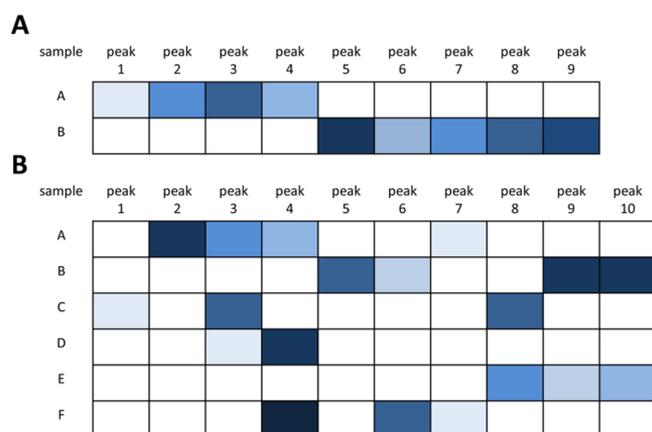


Figure 2. Schematic representation of relative peak abundances (shades of blue) in different chromatograms (samples). Example comprising two samples and nine different compounds (A). Example comprising six samples and ten different compounds (B).

sample (between 2 and 4 per sample) (Figure 2B). Only peak 1 is specific to sample C, while all other peaks occur in several samples. If no. peaks is set to 2, peaks A2, A3, B9, B10, C3, C8, D3, D4, E8, E10, F4, and F6 would be integrated. Thus, peaks 1 and 7 would not be considered for data mining. If the no. peaks is increased to 3, all peaks would be integrated.

Save Into Spectrus Format. By activating this option, processed chromatograms are saved as separate files and become accessible in ACD/MS Workbook Suite for cross-checking of results, and for additional processing (extraction of UV and MS data), reporting, and databasing.

Table of Components. The comparative metabolite profiling tool automatically delivers results assembled as a matrix table in an Excel spreadsheet. Ions intensities below a defined threshold are represented as a dash, and are treated as 0 (zero) for statistical analysis (Figure 3).

	A	B	C	D	E	F	G	H	I	J	BS
1	rt +/- 0.5	m/z	1.lcd	2.lcd	3.lcd	4.lcd	5.lcd	6.lcd	7.lcd	8.lcd	71.lcd
2	6.708	283.3	-	-	2497141	2139886	3135742	-	-	-	2732824.3
3	7.199	328.3	-	-	-	-	-	-	-	-	373815552
4	7.214	327.3	-	-	-	-	3883635	-	-	-	12440916
5	7.36	212.3	-	-	66179900	24111734	11332565	-	391383360	138551888	-
6	7.421	477.3	989814	-	1144676	-	-	-	-	-	-
7	7.429	213.3	-	-	80301408	177329664	162665888	-	465924608	128368016	-
8	7.544	473.3	1502355	1959474	-	-	-	-	-	-	-
9	7.801	180.3	-	-	-	-	-	-	-	-	-
10	7.897	310.3	-	-	4132200	26791150	-	-	-	-	-
11	7.904	307.3	-	-	94874424	30154762	3104939	-	-	3437850	-
12	7.916	308.3	-	-	17922508	140469120	55545088	-	-	-	-
13	8.196	240.3	-	-	-	-	-	-	-	3340705	59419224
14	8.207	691.3	1853025	4298403	1713947	9817492	-	7713136	2932791	6599761	651405
15	8.222	172.3	-	-	1144676	-	267621776	-	7593150	-	-
16	8.571	308.3	-	2704010	3114362	131288272	-	833858880	-	42414424	-
17	8.583	529.3	144180320	126409880	63494228	67073340	5802333	891800768	182291632	822185792	1246383
18	8.587	309.3	2319390	39968836	-	152328336	-	1257021312	90370964	908352512	-
85	18.596	379.3	-	983753216	-	-	-	-	-	-	-

Figure 3. Example of a table of components obtained after processing in ACD/MS Workbook Suite. Individual peaks are defined by RT (column A) and m/z values (column B), and sorted by their RT. Columns represent the different samples. Sample names are reported in row 1. The matrix elements represent the abundance of molecular ions of single compounds in each sample, expressed as AUC.

Data Clustering Using FreeClust (Data Mining Step 2).

A web-application called FreeClust was developed in the R programming language (free R software package *Shiny*²⁷). It is independent from the ACD/Laboratories platform and provides an interactive and freely accessible environment for statistical evaluation and data visualization of large data sets (Figure S-2). Different data sets formats (matrix table or text files) can be uploaded on FreeClust. Data sets are clustered

within FreeClust using Bayesian, hierarchical, or sparse hierarchical clustering methods.

Loading Data in FreeClust. Text files or Excel spreadsheets exported to CSV files in Excel are recognized by FreeClust. For text files, values should be listed in columns separated by spaces, commas, or semicolons. Missing values are described in the input file as “NA”, “-”, or an empty space. For matrix tables, feature names and sample names should be listed in first columns and first row, respectively. If feature information is described in more than one column, the concatenation function in excel should be used.

Group of Data Sets. One data set is defined by measurement values, also designated as observations or features (e.g., metabolites defined by RT- m/z) recorded for one sample. A group of data sets includes, for example, all values obtained after comparative metabolite profiling. This group can be large in terms of sample size (high number of samples compared to the total number of observations recorded) or feature size (high number of features observed compared to the total number of samples).

Basic Processing. The app offers basic data processing, such as data trimming, clipping, and rescaling. In data trimming, outliers are excluded from clustering (values below or above a threshold are treated as missing values). In data clipping, outliers are limited but not excluded from clustering (values below or above the threshold are converted to the threshold value). Rescaling normalizes features to allow their comparison in case they were measured on different scales or in different units. It is performed independently on every feature (mean value is subtracted from each individual feature value, and the difference is further divided by the standard deviation). Values treated as missing are not affected by trimming nor clipping.

Hierarchical Clustering. It provides best results with groups of data sets characterized by large sample size. This method is considering all features for clustering. It builds hierarchy of clusters using standard R functions *dist* and *hclust*.²⁸ Different metrics (e.g., Euclidean, Maximum, Manhattan), as well as linkage criteria (e.g., complete-linkage clustering), are used to calculate pairwise distances or dissimilarities between feature patterns in samples.

Sparse Hierarchical and Bayesian Clustering. These methods provide best results when applied on groups of data sets characterized by a large feature size. Only a selected number of features are considered here for clustering. Feature importance is expressed as a score factor. A color-code and asterisks are used to display the importance of each feature within the group (below the heatmap). Sparse and Bayesian clustering algorithms use different methods to determine the importance of features.

Sparse hierarchical clustering builds a hierarchy of clusters using R package *sparcl*.²⁹ The algorithm quantifies the contribution of a particular feature, that is, its importance, to clustering by giving a weight between 0 and 1 to each feature. For example, if a feature has the same value across all samples, its contribution for building clusters will be nil, and its weight will be assigned to zero.

Bayesian clustering builds clusters using the R package *bclust*.³⁰ Feature importance is quantified here by the log Bayes factor (log ratio of the likelihood probability). For example, a negative value of importance implies negative evidence that a feature participates in the optimal clustering. To apply *bclust*, missing values need to be converted to zeroes, and this data adjustment can be performed in FreeClust.

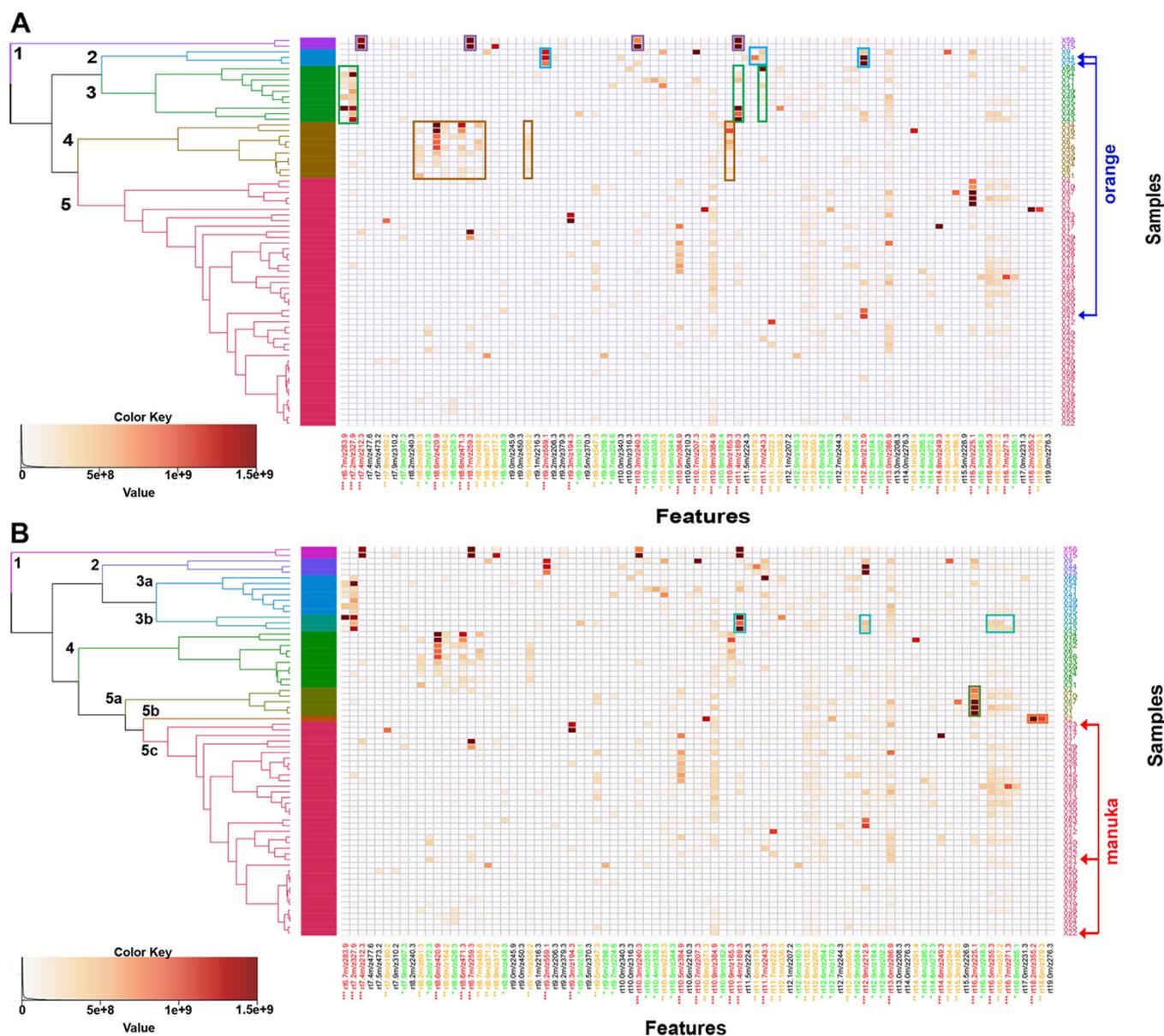


Figure 4. Dendrogram and 2D heatmap of a group of data sets obtained from the comparative metabolite profiling of 69 honey samples. The dendrogram is shown on the left-hand-side of the heatmap, and clusters are highlighted by different colors. Color shades of individual cells in the heatmap correspond to the relative abundance of molecular ions. Sample identity is displayed on the y axis (right-hand-side of the heatmap). Features (RT and m/z values) are represented on the x-axis. The importance score of each feature is indicated by asterisks and colors (black, green, yellow, and red). Bayesian clustering was applied, and samples were classified in 5 (A) and 8 (B) clusters, respectively. Characteristic metabolite patterns in specific clusters are highlighted with colored frames. Red and blue arrows are highlighting data sets obtained from same honey type, namely, orange honeys (in blue) and manuka honeys (in red).

Heatmaps. Feature values are visualized as heatmaps by using different color shades (e.g., color darkness is directly proportional to metabolite abundance). An interactive plot allows to highlight individual values in the heatmap. A zoom-in option enables the exploration of large data sets (Figure S-3).

Dendrogram. The cluster tree or dendrogram is presented on the left-hand-side of the heatmap. It shows how individual samples are grouped in clusters. For easy read-out, clusters can be highlighted using distinct colors. The number of clusters to be highlighted can be modified accordingly.

Practical Example. The two-step data mining workflow was tested with a set of HPLC-ESIMS chromatograms obtained from the analysis of 69 honey samples. Data were acquired in a full scan mode on a triple quadrupole instrument. The

calculation time needed for comparative metabolite profiling on a standard PC was 8.3 h, and results comprising a total of 84 features, were automatically assembled in a matrix table in Excel. Due to the large variability of honey samples, only the most abundant metabolites were considered in the comparison analysis. The number of peaks was arbitrarily set to 5. In addition, each processed chromatogram was saved as a Spectrus file (“Save Into Spectrus Format”) to serve for extraction of additional spectroscopic data. The Excel table containing the results was then loaded in FreeClust and presented as a 2D heatmap after applying Bayesian clustering (Figure 4).

The 2D heatmap provides an overview of the metabolite distribution within the data sets. By using the interactive interface, different clusters were readily identified. Distinct

metabolite patterns for clusters 1–4 can be seen in the heatmap and are highlighted in Figure 4A with the corresponding cluster color. Cluster 1 containing samples X15 (chestnut honey) and X56 (black locust honey) was characterized by four highly abundant metabolites. In contrast, cluster 5 did not show obvious similarities in metabolite patterns. The clustering level was therefore increased to 8 (Figure 4B). Cluster 5 was thereby subdivided into three groups (5a to 5c). This level adjustment also affected cluster 3 which was further subdivided into two groups (3a and 3b).

Processing of the individual Spectrus files provided additional relevant information on single constituents. The four highly abundant metabolites in cluster 1 (RT 7.4, 8.7, 10.8, and 11.4 min, respectively) were reported as XIC (Figure 5A, 5B). Supplementary spectroscopic data indicated very similar UV spectra, and the molecular masses were deduced from the MS⁺ and MS⁻ spectra as 211.3, 258.3, 239.3, and 188.3, respectively (Figure 5C). The UV spectra and MS fragmentation pattern recorded for the four metabolites thus suggested similar structures.

The metabolite profiles of honeys of same floral origin (according to labeling by provider) were compared. Interestingly, three samples of orange honey were grouped in 2 different clusters, namely 2 and 5 (X25, X44, and X47, highlighted in Figure 4A). Samples X25 and X44 were characterized by a rather similar metabolite pattern and were thus classified in cluster 2. Sample X47 was classified in cluster 5, and only one metabolite (RT 12.9 min, m/z 212.9) was present in all three samples. Three samples of manuka honey were grouped in the same cluster (5c) suggesting similar metabolite patterns for these three samples (samples X21–X23 highlighted in Figure 4B). Clustering levels could be further increased to unravel small differences between metabolite patterns.

The step 1 of our workflow enabled the extraction of additional structural information from MS spectra, as shown for the peak RT 10.5 min with m/z 384.9. A molecular mass of 362 could be deduced from the ions m/z 361.1 $[M - H]^-$ and m/z 384.9 $[M + Na]^+$. Fragment ions at m/z 344.9 $[M - OH]^+$, m/z 327.0 $[M - 2OH]^+$, m/z 308.9 $[M - 3OH]^+$, and m/z 290.9 $[M - 4OH]^+$ suggested the presence of four hydroxy groups (Figure S-4A). As for peak at RT 13.0 min with m/z 286.9, a mass of 264 was deduced from the molecular ions m/z 264.9 $[M + H]^+$ and m/z 263.2 $[M - H]^-$. The ion at m/z 286.9 was identified as the sodium adduct. The fragment ion at m/z 246.9 $[M - OH]^-$ was indicative of a loss of a hydroxy group (Figure S-4B).

As ACD/MS Workbook Suite is using a deconvolution algorithm, coeluting peaks can be identified. In sample 51, a TIC peak eluting at RT 16.7 min was deconvoluted into three XIC peaks with m/z 255, m/z 257, and m/z 271 (Figure S-5). With the support of mass spectra recorded in positive and negative ionization modes, molecular masses of 254, 256, and 270 were deduced (data obtained from Spectrus files, not shown here).

CONCLUSIONS

Our proposed two-step data mining workflow enables (i) comparison of metabolite profiles of complex LC-MS chromatograms in an automated and systematic way and (ii) clustering and visualization of the results in a 2D heatmap for data interpretation.

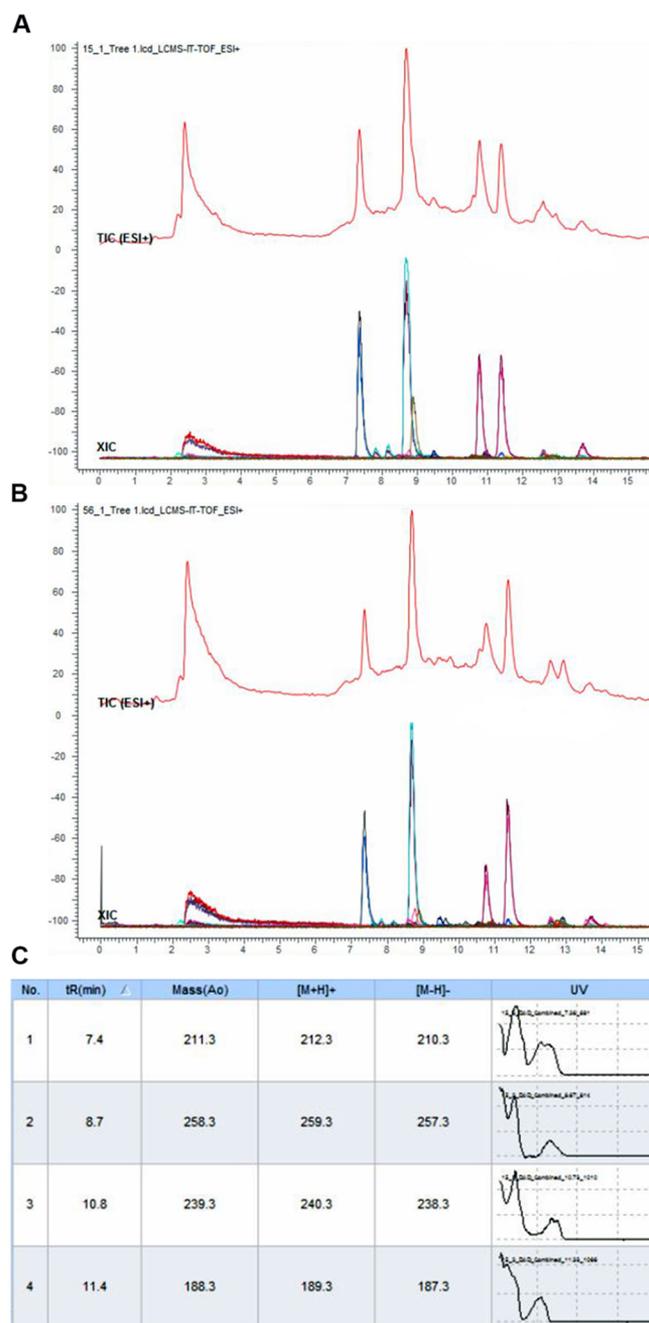


Figure 5. Information extracted from individual Spectrus files in ACD/MS Workbook Suite. Chromatograms obtained from sample X15 (chestnut, A) and sample X56 (black locust, B) is shown. The total ion chromatogram (TIC) is displayed in red, and extracted ion chromatograms (XIC) are overlaid using different colors. Additional spectroscopic information, such as negative- and positive-ion mass, molecular mass, and UV spectrum are available in a table (C).

Developed within the ACD/Laboratories software platform, the add-in tool (step 1) provides a simple way to compare large numbers of LC-MS data sets. Moreover, the comprehensive results can be fully integrated into a multidisciplinary workflow. Relevant structural information can be obtained from a single peak in an individual chromatogram for dereplication, metabolite characterization or databasing work. This tool was developed to work with ACD/IX. A new version of the algorithms, ACD/IX2, was released in July 2017, and provides additional features (e.g., capabilities for processing of HRMS,

GC-MS, and APCI data). We are currently modifying the add-in tool to operate with ACD/IX2.

The web application FreeClust (step 2) can cluster and display the results in intuitive 2D heatmap graphs for easy and unbiased data interpretation. The user-friendly design of FreeClust allows analytical chemists to perform statistical analysis on their data without the need of expert knowledge in advanced statistical tools. Moreover, the 2D heatmap has been designed as an interactive interface that allows users to quickly identify differences between clusters. The user can access specific values by simply moving the cursor to the cell of interest. FreeClust was developed as an open-access web application which is independent from the ACD/Laboratories environment. It can be used with data sets obtained from various sources and, therefore, fulfills needs not only in the fields of phytochemistry and phenotyping,⁷ but also in other areas such as molecular biology.

The proposed two-step data mining workflow can be applied in various fields of research where metabolite profiles of numerous samples need to be compared. In the field of natural products this includes, for example, the quality control of botanicals, analysis of chemical traits in plant breeding, or prioritization/selection of crude extracts in drug discovery.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.analchem.7b02221](https://doi.org/10.1021/acs.analchem.7b02221).

Newly developed add-in tool in ACD/MS Workbook Suite, FreeClust web application platform, heatmap zoom, MS spectra from Spectrus files, peak deconvolution in ACD/Laboratories, information on 69 honey samples used in this study, ACD/Laboratories IntelliXtract parameters table, and ACD/Laboratories IntelliTarget parameters table (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: matthias.hamburger@unibas.ch. Tel.: +41 61 207 14 25. Fax: +41 61 207 14 74.

ORCID

Matthias Hamburger: [0000-0001-9331-273X](https://orcid.org/0000-0001-9331-273X)

Notes

The authors declare the following competing financial interest(s): H.D.B. is a permanent employee at Advanced Chemistry Development, Inc.

■ ACKNOWLEDGMENTS

We thank Dr. Gerd Rheinwald for valuable discussions.

■ REFERENCES

- (1) Lu, L.; Wang, J.; Xu, Y.; Wang, K.; Hu, Y.; Tian, R.; Yang, B.; Lai, Q.; Li, Y.; Zhang, W.; Shao, Z.; Lam, H.; Qian, P.-Y. *Sci. Rep.* **2015**, *4*, 6537.
- (2) Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass Spectrom. Rev.* **2007**, *26*, 51–78.
- (3) Puyana, M.; Pawlik, J.; Blum, J.; Fenical, W. *Rev. Bras. Farmacogn.* **2015**, *25*, 592–599.
- (4) Higgs, R. E.; Zahn, J. A.; Gygi, J. D.; Hilton, M. D. *Appl. Environ. Microbiol.* **2001**, *67*, 371–376.

- (5) Siddiqui, M. J.; Ismail, Z.; Saidan, N. H. *Pharmacogn. Mag.* **2011**, *7*, 92–96.
- (6) Castillo, S.; Gopalacharyulu, P.; Yetukuri, L.; Orešič, M. *Chemom. Intell. Lab. Syst.* **2011**, *108*, 23–32.
- (7) Chervin, J.; Stierhof, M.; Tong, M. H.; Peace, D.; Hansen, K. Ø.; Urgast, D. S.; Hammer Andersen, J.; Yu, Y.; Ebel, R.; Kyeremeh, K.; Paget, V.; Cimpan, G.; Van Wyk, A.; Deng, H.; Jaspars, M.; Tabudravu, J. N. *J. Nat. Prod.* **2017**, *80*, 1370–1377.
- (8) El-Elimat, T.; Figueroa, M.; Ehrmann, B. M.; Cech, N. B.; Pearce, C. J.; Oberlies, N. H. *J. Nat. Prod.* **2013**, *76*, 1709–1716.
- (9) Olivon, F.; Grelier, G.; Roussi, F.; Litaudon, M.; Touboul, D. *Anal. Chem.* **2017**, *89*, 7836–7840.
- (10) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, M.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, C.; Melnik, A. V.; Meehan, M. J.; Liu, W.-T.; Crüsemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderón, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C.-C.; Yang, Y.-L.; Humpf, H.-U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; Boya P, C. A.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodriguez, A. M.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P. M.; Phapale, P.; Nothias, L. F.; Alexandrov, T.; Litaudon, M.; Wolfender, J. L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D. T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Muller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. O.; Pogliano, K.; Linnington, R. G.; Gutierrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N. *Nat. Biotechnol.* **2016**, *34*, 828–837.
- (11) Wolfender, J.-L.; Marti, G.; Thomas, A.; Bertrand, S. *J. Chromatogr. A* **2015**, *1382*, 136–164.
- (12) Jonsson, P.; Bruce, S. J.; Moritz, T.; Trygg, J.; Sjöström, M.; Plumb, R.; Granger, J.; Maibaum, E.; Nicholson, J. K.; Holmes, E.; Antti, H. *Analyst* **2005**, *130*, 701–707.
- (13) LaMarche, B. L.; Crowell, K. L.; Jaitly, N.; Petyuk, V. A.; Shah, A. R.; Polpitiya, A. D.; Sandoval, J. D.; Kiebel, G. R.; Monroe, M. E.; Callister, S. J.; Metz, T. O.; Anderson, G. A.; Smith, R. D. *BMC Bioinf.* **2013**, *14*, 49–63.
- (14) De Vos, R. C. H.; Moco, S.; Lommen, A.; Keurentjes, J. J. B.; Bino, R. J.; Hall, R. D. *Nat. Protoc.* **2007**, *2*, 778–791.
- (15) Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J. *Nat. Rev. Drug Discovery* **2015**, *14*, 111–129.
- (16) Quinn, R. A.; Nothias, L.-F.; Vining, O.; Meehan, M.; Esquenazi, E.; Dorrestein, P. C. *Trends Pharmacol. Sci.* **2017**, *38*, 143–154.
- (17) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779–787.
- (18) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. *BMC Bioinf.* **2010**, *11*, 395.
- (19) Pietiläinen, K. H.; Sysi-Aho, M.; Rissanen, A.; Seppänen-Laakso, T.; Yki-Järvinen, H.; Kaprio, J.; Orešič, M. *PLoS One* **2007**, *2*, e218.
- (20) Laaksonen, R.; Katajamaa, M.; Päivä, H.; Sysi-Aho, M.; Saarinen, L.; Junni, P.; Lütjohann, D.; Smet, J.; Van Coster, R.; Seppänen-Laakso, T.; Lehtimäki, T.; Soini, J.; Orešič, M. *PLoS One* **2006**, *1*, e97.
- (21) Kind, T.; Tolstikov, V.; Fiehn, O.; Weiss, R. H. *Anal. Biochem.* **2007**, *363*, 185–195.
- (22) Rischer, H.; Orešič, M.; Seppänen-Laakso, T.; Katajamaa, M.; Lammertyn, F.; Ardiles-Diaz, W.; Van Montagu, M. C.; Inzé, D.; Oksman-Caldentey, K. M.; Goossens, A. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 5614–5619.

- (23) Zou, W.; Tolstikov, V. V. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 1312–1324.
- (24) Fichou, D.; Ristivojević, P.; Morlock, G. E. *Anal. Chem.* **2016**, *88*, 12494–12501.
- (25) Xia, J.; Sinelnikov, I. V.; Han, B.; Wishart, D. S. *Nucleic Acids Res.* **2015**, *43*, W251–W257.
- (26) Windig, W.; Phalp, J. M.; Payne, A. W. *Anal. Chem.* **1996**, *68*, 3602–3606.
- (27) Oksanen, J.; Guillaume Blanchet, F.; Friendly, M.; Kindt, R.; Legendre, P.; McGlenn, D.; Minchin, P. R.; O'Hara, R. B.; Simpson, G. L.; Solymos, P.; Stevens, M. H. H.; Szoecs, E.; Wagner, H. *Vegan, Community Ecology Package*, 2016. <https://cran.r-project.org/web/packages/vegan/index.html>.
- (28) R Development Core Team. *R, A language and environment for statistical computing*, 2011. <http://www.R-project.org/>.
- (29) Witten, D. M.; Tibshirani, R. *J. Am. Stat. Assoc.* **2010**, *105*, 713–726.
- (30) Nia, V. P.; Davison, A. C. *J. Stat. Softw.* **2012**, *47*, 1–22.

Supporting Information

Automated comparative metabolite profiling of large LC-ESIMS datasets in an ACD/MS Workbook Suite add-in, and data clustering on a new open-source web platform FreeClust

Alen Božičević,[†] Maciej Dobrzyński,[‡] Hans De Bie,[‡] Frank Gafner,[#] Eliane Garo,[†] and Matthias Hamburger^{*†}

[†]Division of Pharmaceutical Biology, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland

[‡]Institute of Cell Biology, University of Bern, Baltzerstrasse 4, 3012 Bern, Switzerland

[‡]Advanced Chemistry Development, Inc., 8 King Street East Suite 107, Toronto, Ontario M5C, Canada

[#]Mibelle Biochemistry, Mibelle AG, Bolimattstrasse 1, 5033 Buchs, Switzerland

E-mail addresses:

Corresponding author: matthias.hamburger@unibas.ch

Co-authors: alen.bozicevic@unibas.ch, maciej.dobrzynski@izb.unibe.ch, hans.debie@acdlabs.com, frank.gafner@mibellegroup.com, eliane.garo@unibas.ch.

Table of Contents

Figure S-1. The add-in tool in ACD/MS Workbook Suite.

Figure S-2. The FreeClust web application platform.

Figure S-3. The “zoom in” function for the heatmap in FreeClust.

Figure S-4. Extracted MS spectra from Spectrus files.

Figure S-5. Peak deconvolution in ACD/Labs.

Table S-1. Commercial honey samples.

Table S-2. ACD/Labs IntelliXtract (ACD/IX) parameters.

Table S-3. ACD/Labs IntelliTarget (ACD/ITA) parameters.

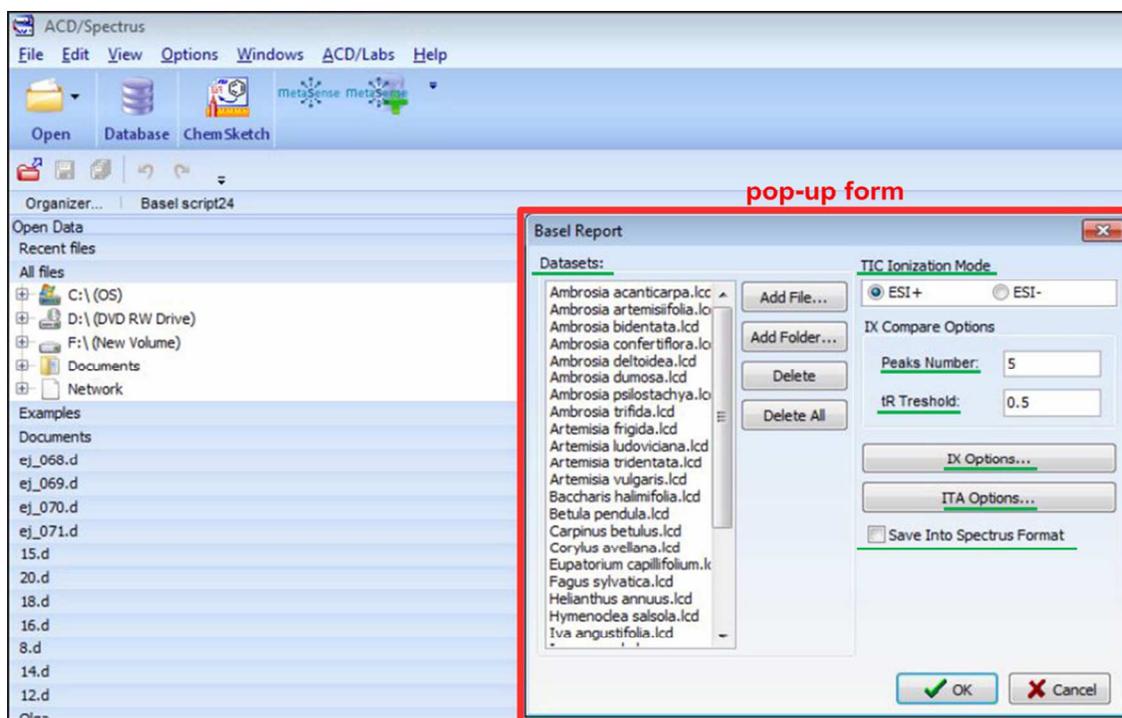


Figure S-1. The customized add-in tool developed in ACD/MS Workbook Suite is managed through the pop-up form. Items underlined in green are described in the main text.

Clustering algorithms

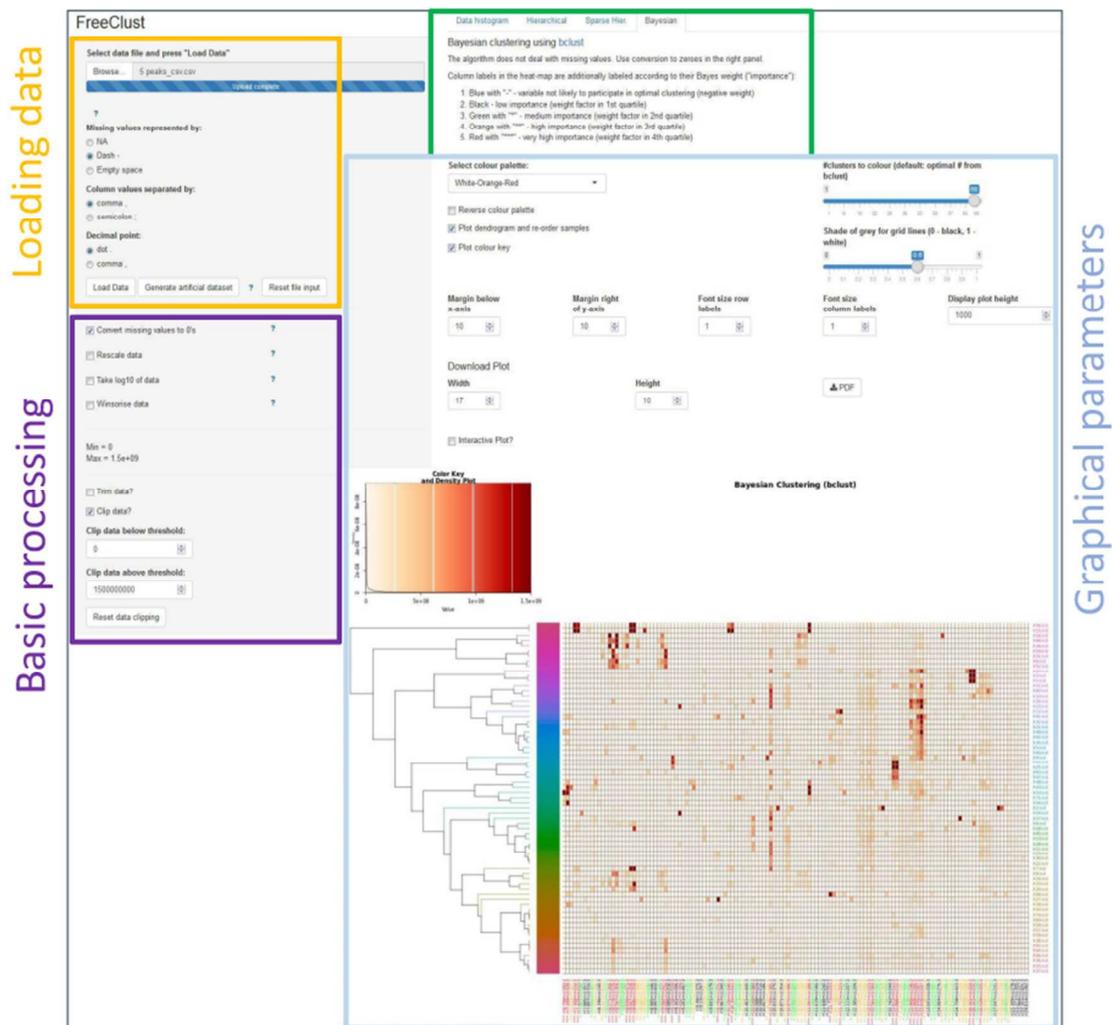


Figure S-2. The FreeClust web application platform.

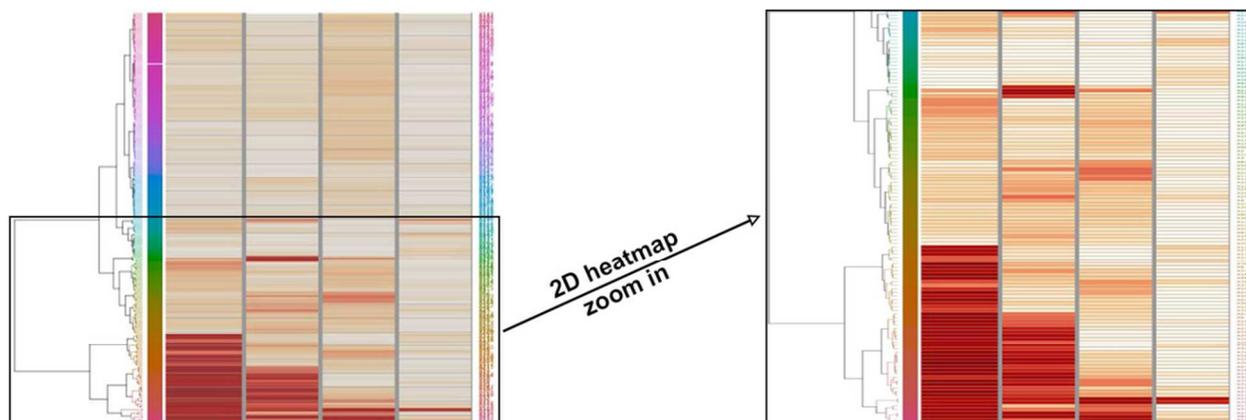


Figure S-3. In the interactive plot, the user can zoom-in into the 2D heatmap. The region of interest will then be automatically expanded within the existing frame to explore the 2D heatmap as shown above.

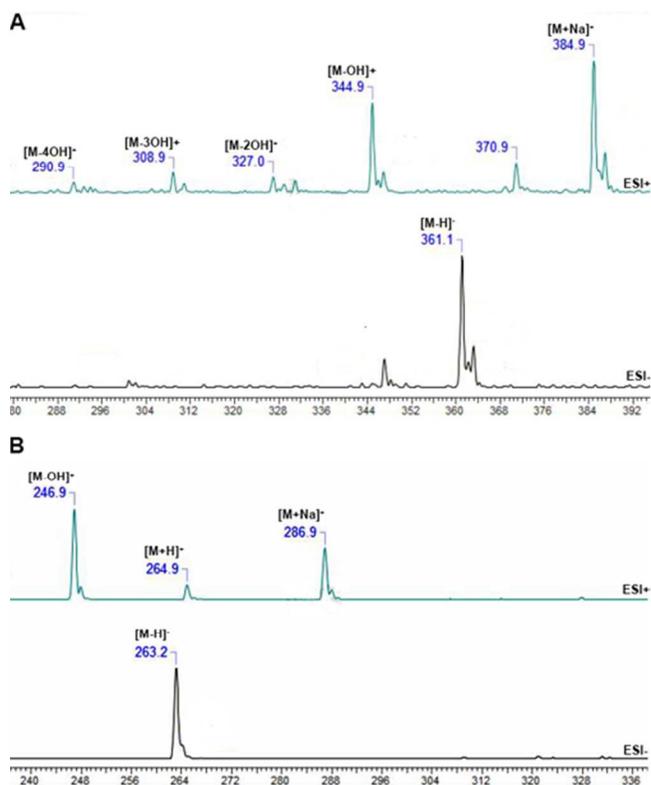


Figure S-4. Extracted MS spectra from Spectrus files with MS+ shown in teal and MS- in black. An example is shown with peaks m/z 384.9 at RT 10.5 min (A) and m/z 286.9 at RT 13.0 min (B).

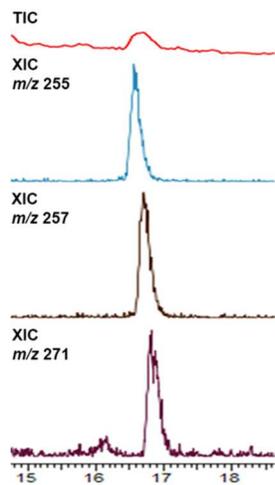


Figure S-5. TIC and XIC of analytes in sample 51 eluting between 16.5 and 17.0 min.

Table S-1. Honey samples selected for this study.

<u>Index</u>	<u>Labeling</u>	<u>Origin</u>	<u>Price Fr./kg</u>	<u>weight [g]</u>	<u>Tara1 [ug]</u>	<u>Tara2 [ug]</u>	<u>Netto extract</u>
1	Acacia	Serbia	13,8	3,0	2310,1	2313,6	3,5
2	Acacia	Romania	43,6	3,0	2302,4	2308,2	5,8
3	Acacia	Italy	76,3	3,0	2299,1	2301,2	2,1
4	Alpine flowers	Italy	76,3	3,0	2322,6	2326,6	4,0
5	Mountain	Spain	39,6	3,0	2304,5	2308,8	4,3
6	Mountain	Switzerland	35,6	3,0	2332,7	2346,3	13,6
7	Mountain	Switzerland	39,6	3,0	2293	2297,0	4,0
8	Flowers	Switzerland	46	3,0	2289,4	2296,5	7,1
9	Flowers	Tasmania	97,5	3,0	2283,9	2287,7	3,8
10	Borage	New Zealand	31,8	3,0	2300,5	2302,1	1,6
11	British summer	UK	48	3,0	2329,4	2331,4	2,0
12	Eucalyptus	Spain	31,6	3,0	2295,2	2298,5	3,3
13	Fleur de café	Guatemala	52,3	3,0	2321,8	2324,1	2,3
14	Heather	UK	52,25	3,0	2319,8	2323,2	3,4
15	Chestnut	Italy	39,6	3,0	2320,7	2327,5	6,8
16	Chestnut	Italy	76,3	3,0	2325,9	2332,1	6,2
17	Clover	New Zealand	39,8	3,0	2305,6	2308,1	2,5
18	Lavender	France	47,6	3,0	2355,2	2357,6	2,4
19	Tilia	Europe	32,25	3,0	2316,6	2330,4	13,8

Table S-1. continued.

20	Almond	Spain	35,6	3,0	2287,9	2290,0	2,1
21	Manuka	New Zealand	46	3,0	2303,8	2309,7	5,9
22	Manuka	New Zealand	80,7	3,0	2308	2317,0	9,0
23	Manuka	New Zealand	127,2	3,0	2322,2	2326,5	4,3
24	Millefiori	Italy	76,3	3,0	2311,6	2316,6	5,0
25	Orange	Italy	39,6	3,0	2325,1	2328,2	3,1
26	Pyrenees	Spain	35,6	3,0	2360,2	2363,3	3,1
27	Rosemary	Portugal	39,6	3,0	2322,2	2324,9	2,7
28	Norway spruce	Germany	39,6	3,0	2285,5	2290,7	5,2
29	Flowers	Mexico/Chile		3,0	2291	2294,1	3,1
30	Thymus	New Zealand	97,5	3,0	2288,7	2292,0	3,3
31	Forest	Switzerland	46	3,0	2293,2	2303,8	10,6
32	Forest	Italy	35,6	3,0	2309,7	2314,0	4,3
33	Mountain (Pro Montagna)	Switzerland	40	3,0	2279,4	2291,3	11,9
34	Oak forest	Spain	18,8	3,0	2293,6	2299,8	6,2
35	Max Havelaar	Middle-, South America	11,2	3,0	2283	2286,4	3,4
36	Sunflower	Europe	12,4	3,0	2300,2	2302,9	2,7
37	Forest	Europe	28,35	3,0	2278,3	2284,0	5,7
38	Sunflower	Europe	28,35	3,0	2282,9	2286,5	3,6
39	Summer flower	Europe	28,35	3,0	2287,6	2292,4	4,8
40	Lavender	Spain	10,4	3,0	2299,6	2302,6	3,0
41	Max Havelaar	Mexico/Chile	12,4	3,0	2288,8	2292,9	4,1
42	Flowers liquid	Middle-, South America	5,5	3,0	2297,8	2300,9	3,1
43	Flowers dense	Middle-, South America	5,5	3,0	2327	2329,8	2,8
44	Orange	Italy	10,4	3,0	2327,3	2330,4	3,1

Table S-1. continued.

45	Rosemary	Spain	10,4	3,0	2278,8	2282,2	3,4
46	Tilia	Middle-, East Europe	10	3,0	2304,5	2314,6	10,1
47	Orange	Middle-, South America	10	3,0	2343,5	2345,9	2,4
48	Flowers	EU, South America, Asia	5,5	3,0	2302,5	2304,5	2,0
49	Organic	Mexico/Chile	11,6	3,0	2304,8	2308,1	3,3
50	Honey	Bulgaria		3,0	2288,1	2289,6	1,5
51	Bee	Leimental (Switzerland)		3,0	2314,1	2316,5	2,4
52	Organic forest	Biel-Benken (Switzerland)		3,0	2300,9	2310,9	10,0
53	Forest Khmum Prey	Cambodia		3,0	2304,1	2310,7	6,6
54	Forest	Cambodia		3,0	2326,1	2335,2	9,1
55	Rape		17	3,0	2328,4	2334,2	5,8
56	Robinia		17	3,0	2302,3	2304,0	1,7
57	Tilia		17	3,0	2312,1	2322,4	10,3
58	Summer flower		17	3,0	2303,7	2306,5	2,8
59	Korn flower		15	3,0	2313,7	2318,2	4,5
60	Bhutan		17	3,0	2308,5	2310,9	2,4
63	Döttinger flower	Switzerland		3,0	2299,4	2301,2	1,8
64	Döttinger flower	Switzerland		3,0	2299,9	2306,9	7,0
65	Döttinger flower	Switzerland		3,0	2291,6	2296,8	5,2
66	Aarau	Switzerland		3,0	2321,3	2325,0	3,7
67	Eigenmann mountain	Switzerland		3,0	2300,1	2305,0	4,9
68	Farm	Ethiopia		3,0	2282,1	2288,5	6,4
69	Uniball	Iran		3,0	2322,5	2333,8	11,3
70	Khansar nemoneh	Iran		3,0	2283,8	2286,1	2,3
71	Tsirang	Bhutan		3,0	2296,9	2302,3	5,4

Table S-2. ACD/Labs IntelliXtract (ACD/IX) parameters for optimization of peak recognition and integration.

Parameter	Value
<input checked="" type="checkbox"/> 12C/13C Error (%)	
<input type="checkbox"/> Value	10
<input type="checkbox"/> A + 2	
<input checked="" type="checkbox"/> Percent	30
<input checked="" type="checkbox"/> Abundance Ratio Error (%)	50
<input checked="" type="checkbox"/> Sulfur	True
<input checked="" type="checkbox"/> Sulfur Min	0
<input checked="" type="checkbox"/> Sulfur Max	3
<input checked="" type="checkbox"/> Chlorine	True
<input checked="" type="checkbox"/> Chlorine Min	0
<input checked="" type="checkbox"/> Chlorine Max	3
<input checked="" type="checkbox"/> Bromine	True
<input checked="" type="checkbox"/> Bromine Min	0
<input checked="" type="checkbox"/> Bromine Max	3
<input type="checkbox"/> Additional M Ion Correction (%)	
<input checked="" type="checkbox"/> Value	5
<input checked="" type="checkbox"/> Apply Ion Presence Detection 12C/13C Filter	
<input type="checkbox"/> Apply	False
<input type="checkbox"/> 12C/13C (Pass)	True
<input type="checkbox"/> 12C/13C (Fail)	False
<input type="checkbox"/> 12C/13C (None)	False
<input checked="" type="checkbox"/> Apply Ion Presence Detection for	
<input type="checkbox"/> Apply	True
<input type="checkbox"/> Molecular Ions	True
<input type="checkbox"/> Adduct Ions	True
<input type="checkbox"/> Primary Fragment Ions	True
<input type="checkbox"/> Multimer Ions	False
<input type="checkbox"/> Unconfirmed 12C Ions	False
<input type="checkbox"/> Unconfirmed 13C Ions	False
<input type="checkbox"/> Unassigned Ions	False
<input type="checkbox"/> Label Parent Ion Modifications	False
<input checked="" type="checkbox"/> Apply Ion Presence Detection Threshold	
<input type="checkbox"/> Apply	True
<input type="checkbox"/> Abundance Type	Height
<input type="checkbox"/> Abundance Units	%
<input type="checkbox"/> Value	1

Table S-2. continued.

Parameter	Value
<input checked="" type="checkbox"/> Auto Neutral Loss Determination	
<input type="checkbox"/> Apply	True
<input type="checkbox"/> Display Accuracy (Decimal Places)	2
<input checked="" type="checkbox"/> Calculate the Following Mass Values	
<input checked="" type="checkbox"/> Peak Top	True
<input checked="" type="checkbox"/> Peak Top Window Type	% of Peak FWHM
<input checked="" type="checkbox"/> Peak Top Window Value	100
<input checked="" type="checkbox"/> Left Shoulder	False
<input checked="" type="checkbox"/> Exclude the First Scan	True
<input checked="" type="checkbox"/> Right Shoulder	False
<input checked="" type="checkbox"/> Exclude the Last Scan	True
<input checked="" type="checkbox"/> Entire Peak	False
<input checked="" type="checkbox"/> Exclude the First and Last Scans	True
<input checked="" type="checkbox"/> Apply Intensity Based Weighting	True
<input checked="" type="checkbox"/> CODA	
<input checked="" type="checkbox"/> Apply	True
<input checked="" type="checkbox"/> Peak Minimum FWHM	Scans
<input checked="" type="checkbox"/> Value	0
<input checked="" type="checkbox"/> Peak Minimum Height (% of max)	0
<input checked="" type="checkbox"/> Peak S/N Threshold	6
<input checked="" type="checkbox"/> Peak Area Threshold (counts)	0
<input checked="" type="checkbox"/> Pick the Number of Most Abundant Peaks	False
<input checked="" type="checkbox"/> Number of Peaks	1
<input checked="" type="checkbox"/> Apply Smoothing	False
<input checked="" type="checkbox"/> Method	FFT
<input checked="" type="checkbox"/> Shape of FFT	Fix
<input checked="" type="checkbox"/> Fix V1(%)	5
<input checked="" type="checkbox"/> Fix V2(%)	50
<input type="checkbox"/> Polynomial	1
<input type="checkbox"/> Points	3
<input type="checkbox"/> Enhancement	True
<input type="checkbox"/> Baseline Correction	False
<input type="checkbox"/> Box Half Width (Scans)	5
<input type="checkbox"/> Noise Factor	1
<input type="checkbox"/> MCQ Value	1
<input type="checkbox"/> Abundance Value (% of TIC)	0

Table S-2. continued.

Parameter	Value
<input checked="" type="checkbox"/> COMPARE keeps certain Peaks	
<input checked="" type="checkbox"/> Peak Tagged as Molecular Ions	True
<input checked="" type="checkbox"/> Peaks with other textual descriptors	True
<input checked="" type="checkbox"/> Classified as Good	True
<input checked="" type="checkbox"/> Classified as Review	True
<input checked="" type="checkbox"/> Classified as Poor	True
<input checked="" type="checkbox"/> Compare Parameters	
<input type="checkbox"/> tR Window (min)	0.2
<input type="checkbox"/> Abundance	Area
<input type="checkbox"/> Abundance Ratio	5:1
<input type="checkbox"/> Isotope Cluster Tolerance	True
<input type="checkbox"/> Centre of Gravity (%)	50
<input type="checkbox"/> Cluster Dispersion (%)	50
<input type="checkbox"/> Boolean Comparison	False
<input type="checkbox"/> Component MC Summing	
<input checked="" type="checkbox"/> Apply	False
<input checked="" type="checkbox"/> Component Window	
<input type="checkbox"/> CW Type	Scans
<input type="checkbox"/> Value	0.1
<input type="checkbox"/> FWHM Deviation (%)	20
<input checked="" type="checkbox"/> Components Analysis	
<input type="checkbox"/> Apply	True
<input type="checkbox"/> Default Peak Width	0.05
<input type="checkbox"/> Default Peak Function	Gauss
<input type="checkbox"/> Lorentz Fraction	1
<input type="checkbox"/> Vary position within	0.1
<input type="checkbox"/> Max width limit	10
<input type="checkbox"/> Min width limit	0.001
<input type="checkbox"/> Tolerance	3
<input type="checkbox"/> Max number of iteration	150
<input type="checkbox"/> Lock Width	True
<input type="checkbox"/> Lock Height	True
<input type="checkbox"/> Fixed Width	False
<input type="checkbox"/> Parameter	0.05
<input type="checkbox"/> Max Number of Peaks per MC	15
<input type="checkbox"/> Max Number of Co-eluted Peaks per Region	15

Table S-2. continued.

Parameter	Value
<input checked="" type="checkbox"/> Convert Profile to Centroid	
<input type="checkbox"/> Apply	True
<input type="checkbox"/> Mass Values to Center	Top
<input type="checkbox"/> Abundance to Center	Area
<input type="checkbox"/> Peak Picking Minimum FWHM (Points)	3
<input type="checkbox"/> Abundance Threshold (%)	1
<input type="checkbox"/> Centroid %	80
<input checked="" type="checkbox"/> Apply Smoothing	False
<input checked="" type="checkbox"/> Polynomial	1
<input checked="" type="checkbox"/> Window Width (Points)	3
<input checked="" type="checkbox"/> Data Analysis Region	
<input checked="" type="checkbox"/> Full	False
<input checked="" type="checkbox"/> Start tR	5
<input checked="" type="checkbox"/> End tR	20
<input type="checkbox"/> Extract Pure Component Spectra	
<input checked="" type="checkbox"/> Apply	False
<input checked="" type="checkbox"/> Extract Selected Components Only	True
<input checked="" type="checkbox"/> Directory	
<input checked="" type="checkbox"/> File Name	
<input checked="" type="checkbox"/> Action if File Name Exists	AutoName
<input type="checkbox"/> HRMS Data Processing Approach	
<input checked="" type="checkbox"/> Apply	False
<input checked="" type="checkbox"/> Type of Mass	Peak Top
<input checked="" type="checkbox"/> Mass Value Window (Da, +/-)	0.01
<input type="checkbox"/> Ionization	
<input type="checkbox"/> Type	Auto
<input type="checkbox"/> Isotope Spacing Tolerance in Spectrum (Da)	
<input checked="" type="checkbox"/> Value	0.5
<input checked="" type="checkbox"/> Mass Values	Accurate
<input checked="" type="checkbox"/> Type of Mass	Peak Top
<input type="checkbox"/> M Ion Error (%)	
<input checked="" type="checkbox"/> Value	20
<input type="checkbox"/> Mass Accuracy	
<input type="checkbox"/> Units	Da
<input type="checkbox"/> Value	0.5
<input type="checkbox"/> MC Display (Da)	0.5
<input checked="" type="checkbox"/> MC Calculation (Da)	1
<input checked="" type="checkbox"/> m/z Values	Averaged
<input type="checkbox"/> Mass Unity Range Shift	0.3

Table S-2. continued.

Parameter	Value
<input type="checkbox"/> Mass Accuracy for Peak Labeling	
<input checked="" type="checkbox"/> Type of Mass	Peak Top
<input type="checkbox"/> Mass Accuracy	0.5
<input type="checkbox"/> Missing Isotopes Search	
<input checked="" type="checkbox"/> Apply	True
<input checked="" type="checkbox"/> Good	True
<input checked="" type="checkbox"/> Poor	True
<input checked="" type="checkbox"/> Review	True
<input checked="" type="checkbox"/> Depth of Analysis	3
<input type="checkbox"/> Modifications used for Peak Tagging	
<input type="checkbox"/> Apply	True
<input type="checkbox"/> Peak Quality Analysis	
<input checked="" type="checkbox"/> W (Area)	20
<input checked="" type="checkbox"/> W (Height)	20
<input checked="" type="checkbox"/> W (Asymmetry)	10
<input checked="" type="checkbox"/> W (S/N)	0
<input checked="" type="checkbox"/> W (Carbon)	20
<input checked="" type="checkbox"/> W (12C/13C)	20
<input checked="" type="checkbox"/> W (Class)	10
<input checked="" type="checkbox"/> W (Uniqueness)	0
<input type="checkbox"/> Peak Quality Threshold	
<input checked="" type="checkbox"/> Apply	False
<input checked="" type="checkbox"/> Value	1
<input type="checkbox"/> Polarity	
<input type="checkbox"/> Auto	True
<input type="checkbox"/> Positive	True
<input type="checkbox"/> S/N Calculation	
<input checked="" type="checkbox"/> Parameter	Scans
<input checked="" type="checkbox"/> Scans	50
<input checked="" type="checkbox"/> tR	0.1

Table S-2. continued.

Parameter	Value
<input type="checkbox"/> Search for Adducts	
<input type="checkbox"/> Proton	1
<input type="checkbox"/> Ammonia	0
<input type="checkbox"/> Sodium	1
<input type="checkbox"/> Potassium	0
<input type="checkbox"/> Acetonitrile	0
<input type="checkbox"/> Methanol	0
<input type="checkbox"/> H ₂ O	0
<input type="checkbox"/> Lithium	0
<input type="checkbox"/> Acetate	0
<input type="checkbox"/> Formate	0
<input type="checkbox"/> TFA	0
<input type="checkbox"/> Search for Multimers	
<input checked="" type="checkbox"/> Proton	2
<input checked="" type="checkbox"/> Ammonia	0
<input checked="" type="checkbox"/> Sodium	1
<input checked="" type="checkbox"/> Potassium	0
<input checked="" type="checkbox"/> Acetonitrile	0
<input checked="" type="checkbox"/> Methanol	0
<input checked="" type="checkbox"/> H ₂ O	0
<input checked="" type="checkbox"/> Lithium	0
<input checked="" type="checkbox"/> Acetate	2
<input checked="" type="checkbox"/> Formate	2
<input checked="" type="checkbox"/> TFA	2

Table S-2. continued.

Parameter	Value
<input type="checkbox"/> Sensitivity Analysis for Area (TMC)	
<input checked="" type="checkbox"/> Apply	True
<input checked="" type="checkbox"/> Filter Value	0.05
<input type="checkbox"/> Sensitivity Analysis for FWHM (TMC)	
<input checked="" type="checkbox"/> Apply	False
<input checked="" type="checkbox"/> Filter Value	0.05
<input type="checkbox"/> Sensitivity Analysis for Height (TMC)	
<input checked="" type="checkbox"/> Apply	False
<input checked="" type="checkbox"/> Filter Value	0.05
<input checked="" type="checkbox"/> Structure Search	
<input checked="" type="checkbox"/> Apply	False
<input checked="" type="checkbox"/> Mass Values	Accurate
<input checked="" type="checkbox"/> Mass Accuracy	0.001
<input checked="" type="checkbox"/> Type of Mass	Peak Top
<input checked="" type="checkbox"/> Molecular Ions	True
<input checked="" type="checkbox"/> 12C Ions	False
<input checked="" type="checkbox"/> Unassigned Ions	False
<input checked="" type="checkbox"/> Directory	
<input checked="" type="checkbox"/> Label (User Data field)	
<input checked="" type="checkbox"/> 12C/13C (Pass)	True
<input checked="" type="checkbox"/> 12C/13C (Fail)	True
<input checked="" type="checkbox"/> 12C/13C (None)	True

Table S-3. ACD/Labs IntelliTarget (ACD/ITA) parameters for optimization of peak picking.

Parameter	Value
<input checked="" type="checkbox"/> 12C/13C Error (%)	
<input type="checkbox"/> Value	10
<input checked="" type="checkbox"/> Additional M Ion Correction (%)	
<input checked="" type="checkbox"/> Value	5
<input checked="" type="checkbox"/> Apply Ion Presence Detection 12C/13...	
<input type="checkbox"/> Apply	False
<input type="checkbox"/> 12C/13C (Pass)	True
<input type="checkbox"/> 12C/13C (Fail)	False
<input type="checkbox"/> 12C/13C (None)	False
<input checked="" type="checkbox"/> Apply Ion Presence Detection for	
<input type="checkbox"/> Apply	True
<input type="checkbox"/> Molecular Ions	True
<input type="checkbox"/> Adduct Ions	True
<input type="checkbox"/> Primary Fragment Ions	True
<input type="checkbox"/> Multimer Ions	False
<input type="checkbox"/> Unconfirmed 12C Ions	True
<input type="checkbox"/> Unconfirmed 13C Ions	False
<input type="checkbox"/> Unassigned Ions	True
<input type="checkbox"/> Label Ions with Status Not Found	True
<input type="checkbox"/> Label Parent Ion Modifications	False
<input checked="" type="checkbox"/> Apply Ion Presence Detection Thresh...	
<input type="checkbox"/> Apply	False
<input type="checkbox"/> Abundance Type	Height
<input type="checkbox"/> Abundance Units	%
<input type="checkbox"/> Value	1
<input checked="" type="checkbox"/> Baseline Correction	
<input type="checkbox"/> Baseline Correction	False
<input type="checkbox"/> Box Half Width (Scans)	5
<input type="checkbox"/> Noise Factor	1
<input checked="" type="checkbox"/> Calculate the Following Mass Values	
<input checked="" type="checkbox"/> Peak Top	True
<input checked="" type="checkbox"/> Peak Top Window Type	% of Peak FWHM
<input checked="" type="checkbox"/> Peak Top Window Value	100
<input checked="" type="checkbox"/> Left Shoulder	False
<input checked="" type="checkbox"/> Exclude the First Scan	True
<input checked="" type="checkbox"/> Right Shoulder	False
<input checked="" type="checkbox"/> Exclude the Last Scan	True
<input checked="" type="checkbox"/> Entire Peak	False
<input checked="" type="checkbox"/> Exclude the First and Last Scans	True
<input checked="" type="checkbox"/> Apply Intensity Based Weighting	True

Table S-3. continued.

Parameter	Value
<input checked="" type="checkbox"/> Component Window	
<input type="checkbox"/> CW Type	% Peak width FWHM
<input type="checkbox"/> Value	30
<input type="checkbox"/> FWHM Deviation (%)	50
<input checked="" type="checkbox"/> Components Analysis	
<input type="checkbox"/> Apply	False
<input type="checkbox"/> Default Peak Width	0.05
<input type="checkbox"/> Default Peak Function	Gauss
<input type="checkbox"/> Lorentz Fraction	1
<input type="checkbox"/> Vary position within	0.1
<input type="checkbox"/> Max width limit	10
<input type="checkbox"/> Min width limit	0.001
<input type="checkbox"/> Tolerance	3
<input type="checkbox"/> Max number of iteration	150
<input type="checkbox"/> Lock Width	True
<input type="checkbox"/> Lock Height	True
<input type="checkbox"/> Fixed Width	False
<input type="checkbox"/> Parameter	0.05
<input type="checkbox"/> Max Number of Peaks per MC	15
<input type="checkbox"/> Max Number of Co-eluted Peaks per Region	15
<input checked="" type="checkbox"/> Convert Profile to Centroid	
<input type="checkbox"/> Apply	False
<input type="checkbox"/> Mass Values to Center	Top
<input type="checkbox"/> Abundance to Center	Area
<input type="checkbox"/> Peak Picking Minimum FWHM (Points)	3
<input type="checkbox"/> Abundance Threshold (%)	1
<input type="checkbox"/> Centroid %	80
<input checked="" type="checkbox"/> Apply Smoothing	False
<input checked="" type="checkbox"/> Polynomial	1
<input checked="" type="checkbox"/> Window Width (Points)	3
<input checked="" type="checkbox"/> Data Analysis Region	
<input type="checkbox"/> Full	True
<input type="checkbox"/> Start tR	0
<input type="checkbox"/> End tR	100
<input checked="" type="checkbox"/> Ionization	
<input type="checkbox"/> Type	Auto
<input checked="" type="checkbox"/> Isotope Pattern Calculation by Formula	
<input type="checkbox"/> Relative Intensity Threshold (%)	5
<input type="checkbox"/> Number of Most Abundant Peaks per Sat...	1

Table S-3. continued.

Parameter	Value
<input checked="" type="checkbox"/> Isotope Spacing Tolerance in Spectru...	
<input checked="" type="checkbox"/> Value	0.5
<input checked="" type="checkbox"/> Mass Values	Accurate
<input checked="" type="checkbox"/> Type of Mass	Peak Top
<input checked="" type="checkbox"/> M Ion Error (%)	
<input checked="" type="checkbox"/> Value	20
<input checked="" type="checkbox"/> Mass Accuracy for Peak Labeling	
<input checked="" type="checkbox"/> Type of Mass	Peak Top
<input type="checkbox"/> Mass Accuracy	0.5
<input checked="" type="checkbox"/> Modifications used for Peak Tagging	
<input type="checkbox"/> Apply	True
<input type="checkbox"/> Peak Picking	
<input type="checkbox"/> Peak Minimum FWHM	Scans
<input type="checkbox"/> Value	0
<input type="checkbox"/> Peak Minimum Height (% of max)	0
<input type="checkbox"/> Peak S/N Threshold	6
<input type="checkbox"/> Peak Area Threshold (counts)	0
<input type="checkbox"/> Pick the Number of Most Abundant ...	True
<input type="checkbox"/> Number of Peaks	4
<input checked="" type="checkbox"/> Peak Quality Threshold	
<input checked="" type="checkbox"/> Apply	False
<input checked="" type="checkbox"/> Value	1
<input checked="" type="checkbox"/> Polarity	
<input type="checkbox"/> Auto	True
<input type="checkbox"/> Positive	True
<input checked="" type="checkbox"/> S/N Calculation	
<input checked="" type="checkbox"/> Parameter	Scans
<input checked="" type="checkbox"/> Scans	50
<input checked="" type="checkbox"/> tR	0.1

Table S-3. continued.

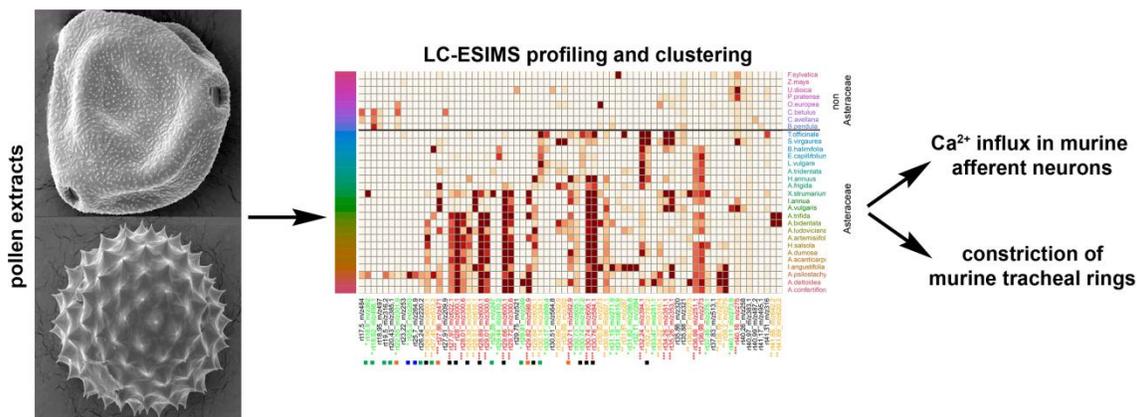
Parameter	Value
<input checked="" type="checkbox"/> Search for Adducts	
<input type="checkbox"/> Apply	False
<input type="checkbox"/> Proton	1
<input type="checkbox"/> Ammonia	0
<input type="checkbox"/> Sodium	1
<input type="checkbox"/> Potassium	0
<input type="checkbox"/> Acetonitrile	0
<input type="checkbox"/> Methanol	0
<input type="checkbox"/> H ₂ O	0
<input type="checkbox"/> Lithium	0
<input type="checkbox"/> Acetate	1
<input type="checkbox"/> Formate	1
<input type="checkbox"/> TFA	1
<input checked="" type="checkbox"/> Search for Multimers	
<input checked="" type="checkbox"/> Apply	False
<input checked="" type="checkbox"/> Proton	2
<input checked="" type="checkbox"/> Ammonia	0
<input checked="" type="checkbox"/> Sodium	1
<input checked="" type="checkbox"/> Potassium	0
<input checked="" type="checkbox"/> Acetonitrile	0
<input checked="" type="checkbox"/> Methanol	0
<input checked="" type="checkbox"/> H ₂ O	0
<input checked="" type="checkbox"/> Lithium	0
<input checked="" type="checkbox"/> Acetate	2
<input checked="" type="checkbox"/> Formate	2
<input checked="" type="checkbox"/> TFA	2

Table S-3. continued.

Parameter	Value
<input type="checkbox"/> Smoothing	
<input type="checkbox"/> Apply Smoothing	False
<input type="checkbox"/> Method	FFT
<input type="checkbox"/> Shape of FFT	Fix
<input type="checkbox"/> Fix V1(%)	5
<input type="checkbox"/> Fix V2(%)	50
<input type="checkbox"/> Polynomial	1
<input type="checkbox"/> Points	3
<input type="checkbox"/> Enhancement	True
<input checked="" type="checkbox"/> Structure Search	
<input checked="" type="checkbox"/> Apply	False
<input checked="" type="checkbox"/> Mass Values	Accurate
<input checked="" type="checkbox"/> Mass Accuracy	0.001
<input checked="" type="checkbox"/> Type of Mass	Peak Top
<input checked="" type="checkbox"/> Molecular Ions	True
<input checked="" type="checkbox"/> 12C Ions	False
<input checked="" type="checkbox"/> Unassigned Ions	False
<input checked="" type="checkbox"/> Directory	
<input checked="" type="checkbox"/> Label (User Data field)	
<input checked="" type="checkbox"/> 12C/13C (Pass)	True
<input checked="" type="checkbox"/> 12C/13C (Fail)	True
<input checked="" type="checkbox"/> 12C/13C (None)	True
<input type="checkbox"/> Target Analysis	
<input type="checkbox"/> Mass Accuracy for MC Generation (Da)	0.5
<input type="checkbox"/> Type of Mass	Peak Top
<input type="checkbox"/> Mass Accuracy for Assignment (Da)	0.5
<input type="checkbox"/> FFQ Threshold	0

3.3 Secondary metabolites in allergic plant pollen samples modulate afferent neurons and murine tracheal rings

Alen Bozicevic, Maria De Mieri, Christina Nassenstein, Silke Wiegand, Matthias Hamburger
J. Nat. Prod. **2017**, *80*, 2953–2961. doi: 10.1021/acs.jnatprod.7b00495.



Pollen originating from thirty plant species were analyzed by HPLC coupled to PDA, ESIMS, and ELSD detectors, and off-line NMR spectroscopy. Polyamine conjugates, flavonoids, and sesquiterpene lactones (STLs) were identified. Selected pollen extracts and pure compounds were tested in murine afferent neurons and in murine tracheal preparations. Electrophilic STLs increased the intracellular Ca²⁺ concentration in 15%, 32%, and 37% of cinnamaldehyde responsive neurons, respectively. In organ bath experiments, only STLs induced a weak dilatation of naïve tracheas, and strongly lowered the maximal methacholine-induced tracheal constriction.

Extraction of plant material, phytochemical profiling, isolation and purification, structure elucidation, writing of the manuscript, and preparation of the figures were my contribution. Interpretation of analytical data for structural elucidation (mass spectrometry, NMR) was a contribution of both Maria De Mieri and me. The organ bath experiments, the Ca²⁺-imaging and organ bath experiments, and data interpretation were contributions of Dr. Christina Nassenstein and Ms. Silke Wiegand. After training, I was able to perform a part of Ca²⁺-imaging experiments and elaboration of data.

Alen Bozicevic

Secondary Metabolites in Allergic Plant Pollen Samples Modulate Afferent Neurons and Murine Tracheal Rings

Alen Božičević,[†] Maria De Mieri,^{†,‡} Christina Nassenstein,^{‡,§} Silke Wiegand,^{‡,§} and Matthias Hamburger^{*,†,‡}

[†]Division of Pharmaceutical Biology, University of Basel, Klingelbergstrasse 50, CH-4056 Basel, Switzerland

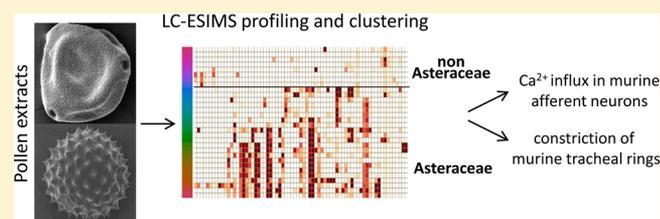
[‡]Institute of Anatomy and Cell Biology, Justus-Liebig-University, Aulweg 123, D-35385 Giessen, Germany

[§]Deutsches Zentrum für Lungenforschung (DZL), Aulweg 130, D-35392 Giessen, Germany

Supporting Information

ABSTRACT: Plant pollens are strong airborne elicitors of asthma. Their proteinaceous allergens have been studied intensively, but little is known about a possible contribution of pollen secondary metabolites to the nonallergic exacerbation of asthma. Pollen samples originating from 30 plant species were analyzed by HPLC coupled to PDA, ESIMS, and ELSD detectors and off-line NMR spectroscopy. Polyamine conjugates, flavonoids, and sesquiterpene lactones were identified.

Polyamine conjugates were characteristic of all Asteraceae species. The presence of sesquiterpene lactones in Asteraceae pollen varied between species and pollen lots. All plant pollen, including those from non-Asteraceae species, contained to some extent electrophiles as determined by their reaction with *N*-acetyl-L-cysteine. Selected pollen extracts and pure compounds were tested in murine afferent neurons and in murine tracheal preparations. Tetrahydrofuran extracts of *Ambrosia artemisiifolia* and *Ambrosia psilostachya* pollen and a mixture of sesquiterpene lactones coronopilin/parthenin increased the intracellular Ca²⁺ concentration in 15%, 32%, and 37% of cinnamaldehyde-responsive neurons, respectively. In organ bath experiments, only the sesquiterpene lactones tested induced a weak dilatation of naïve tracheas and strongly lowered the maximal methacholine-induced tracheal constriction. A tetrahydrofuran extract of *A. psilostachya* and coronopilin/parthenin led to a time-dependent relaxation of the methacholine-precontracted trachea. These results provide the first evidence for a potential role of pollen secondary metabolites in the modulation of the tracheal tone.



Asthma is a chronic inflammatory disorder of the airways associated with airway hyper-responsiveness¹ and is one of the chronic diseases with the highest mortality.² Two recent global studies estimated that asthma affects between 200 and 300 million children and adults worldwide.^{3,4} The incidence varies considerably between countries and ranges from 1% to 16%.⁵ Asthma triggers are very diverse in nature and origin and can be classified into those nonallergic and allergic. Nonallergic triggers include exercise, emotional conditions, cigarette smoke, and a cold temperature. Certain medicines, viral infections, animal dander, dust mites, molds, and plant pollen are known allergic triggers, capable of inducing an inflammatory process resulting in hypersensitivity reactions.⁶ Plant pollens are known to be strong airborne elicitors of asthma in humans, and the role of proteinaceous pollen allergens in asthma, as well as their structures, has been studied extensively.^{7–10} In contrast, little is known regarding a potential contribution of pollen secondary metabolites in the nonallergic exacerbation of asthma symptoms.

Secondary metabolites in pollen and their role in vegetative and generative cells have been investigated previously. Besides the widely studied antioxidative properties of phenolic substances in bee pollen,^{11–14} the biosynthesis of free polyamines and their role in pollen embryogenesis^{15,16} and in

other developmental functions have been of major interest.^{17,18} Furthermore, primary and secondary pollen metabolites, such as carbohydrates, amino acids, lipids, hormones, glutathione, and polyamines, were found to be directly involved in the adaptation processes to environmental conditions such as a high temperature.¹⁹

Besides the well-established role of pollen proteins in the inflammatory and allergic airway hyper-responsiveness and tone modulation, there is so far no evidence for a possible involvement of pollen phytochemicals in the exacerbation of nonallergic asthma symptoms. The transient receptor potential cation channel A1 (TRPA1) is a ubiquitous protein highly expressed in sensory neurons in skin and airways.^{20–22} This Ca²⁺-selective channel can be activated through covalent (e.g., cinnamaldehyde) or noncovalent (e.g., menthol) mechanisms, rendering it a possible target for structurally diverse molecules.²³ TRPA1 has been associated with activation of afferent C-fibers and airway hyper-responsiveness.^{24,25} On the other hand, contraction of myocytes in the trachea has been linked to the endogenous free polyamines spermidine and

Received: June 9, 2017

Published: November 7, 2017

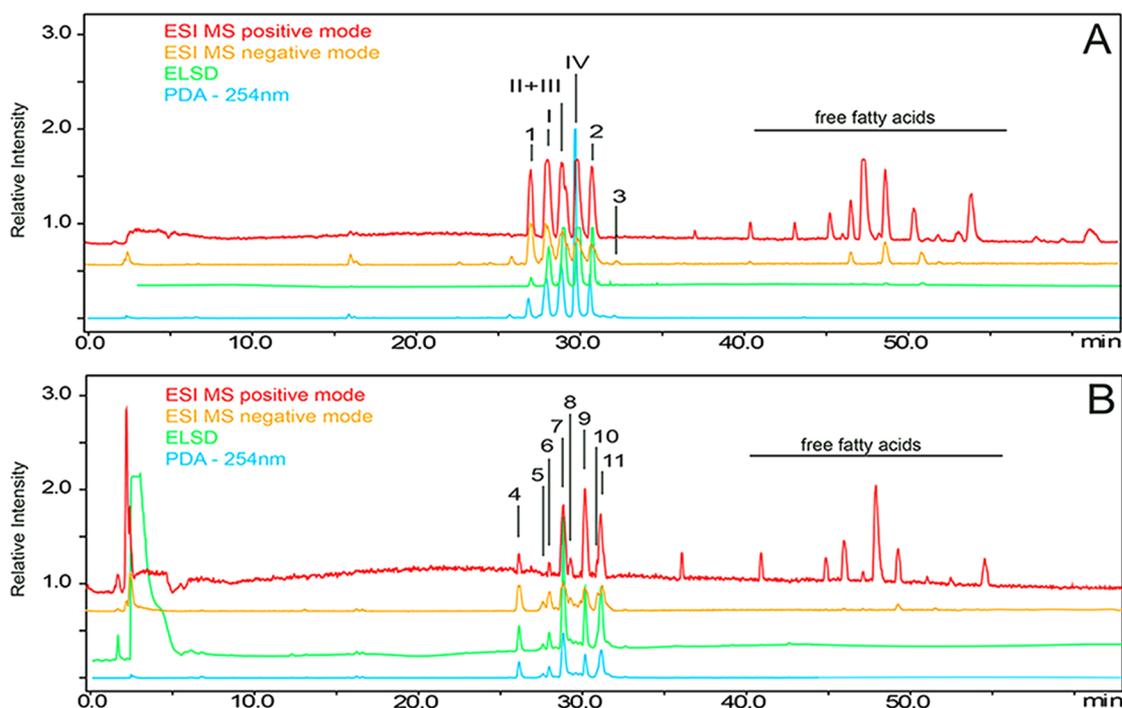


Figure 1. HPLC profiles of extracts from pollen of *A. artemisiifolia* (lot Aa1). Profiles of THF (A) and MeOH (B) extracts were recorded with PDA, ELSD, and ESIMS in positive- and negative-ion modes. Peak numbering designates unambiguously identified compounds 1–11, and I–IV designate peaks for which only partial structural information was obtained.

spermine directly acting on airway smooth muscle cells.²⁶ The polyamine-mediated activation of the type I phosphatidylinositol 4-phosphate 5-kinase (PIP5K) gamma isoform induces the release of sarcoplasmic Ca^{2+} to the cytosol, leading to myosin phosphorylation and contraction of smooth muscle fibers. Relaxation of trachea has been linked to the activation of bitter taste receptors (T2Rs) expressed in ciliated epithelial and smooth muscle cells in the airways.²⁷

To address a possible contribution of pollen secondary metabolites to the clinical exacerbation of the pollen protein-mediated allergic response, a phytochemical profiling was performed of plant pollen from species known for their varying degrees of allergenicity.²⁸ Pollen from species of 11 genera of the family Asteraceae was selected, including the highly allergenic *Ambrosia artemisiifolia* L. and *Ambrosia psilostachya* DC., and from eight genera belonging to plant families other than the Asteraceae. This constituted pollen from a total of 30 plant species. For *A. artemisiifolia*, six different lots of pollen (Aa1–Aa6), as well as samples from flowers and leaves, were analyzed, and for *A. psilostachya* five lots of pollen (Ap1–Ap5) were analyzed. All pollen samples were from commercial suppliers of allergens for the manufacturing of pharmaceutical products for hyposensitization. Phytochemical profiling of secondary metabolites was performed by means of HPLC-PDA-ESIMS-ELSD and supported by targeted isolation of selected compounds and off-line microprobe NMR analysis. Sesquiterpene lactones (STLs) and other electrophiles in pollen extracts were quantified by reaction with *N*-acetyl-L-cysteine. Finally, extracts of the highly allergenic *A. artemisiifolia* and *A. psilostachya* pollen, selected STLs, and polyamine conjugates were tested in murine afferent neurons for their effect on Ca^{2+} influx measured by calcium imaging and for constriction of murine tracheal rings in organ bath experiments.

RESULTS AND DISCUSSION

Comparison of HPLC-ESIMS Profiles of Pollen Extracts. Pollen samples originating from 30 plant species (Table S1, Supporting Information) were defatted with *n*-hexane and extracted exhaustively with tetrahydrofuran (THF) and MeOH. The THF extracts were analyzed by HPLC coupled in series to PDA, ESIMS (positive- and negative-ion modes), and ELSD detectors (Figure S1, Supporting Information). HPLC-ESIMS data were automatically analyzed using ACD/Spectrus Processor software. The analysis consisted of data deconvolution followed by a parallel comparative processing. In a second step, the data table was loaded into a Web application designed for data clustering and visualization, and data were presented as a 2D heat map (Figure S2, Supporting Information).²⁹ The most striking observation from this data set was linked to peaks eluting between t_R 27.0 and 32.4 min (indicated with black squares in Figure S2 and Table S2, Supporting Information) that were detected only in pollen from Asteraceae species. Peaks designated with orange squares showed UV-vis absorption spectra with λ_{max} 240–350 nm and molecular ions that were suggestive of phenolic glycosides (Table S3, Supporting Information). Peaks labeled with blue and green squares, in contrast, showed a $\lambda_{\text{max}} < 240$ nm and molecular ions in the range of m/z 250–300 that were indicative of STLs. For example, the extract of *A. psilostachya* pollen showed at t_R 25.3 and 25.7 min two major HPLC peaks with molecular ions m/z $[\text{M} + \text{H}]^+$ 263.0 and 264.9 (λ_{max} 233 and 230, respectively) (Figure S2 and Table S3, Supporting Information). Based on this preliminary comparison, a targeted isolation of metabolites was conducted. HPLC chromatograms of *A. artemisiifolia* (sample Aa1) THF and MeOH extracts (Figure 1) and an *A. psilostachya* (sample Ap1) THF extract (Figure S1, Supporting Information) served as a basis for a targeted small-scale purification of characteristic major compounds.

Major Phytochemical Observations. THF and MeOH extracts of *A. artemisiifolia* pollen afforded 11 compounds that were identified by HRESIMS and 1D and 2D NMR data.

The new compound **1** was obtained as an amorphous white solid, and a molecular formula of $C_{34}H_{38}N_3O_7$ was assigned on the basis of a molecular ion of m/z 600.2400 $[M + H]^+$ in the HRESIMS (calcd for $C_{34}H_{37}N_3O_7$, 600.2409). 1D and 2D NMR spectra of **1** (Figures S3–S6, Supporting Information) were similar to those of previously reported $N^1,N^5(Z,Z)-N^{10}(E)$ -tri-*p*-coumaroylspermidine (**2**).³⁰ However, the six olefinic protons in **1** (δ_H 6.47, d, $J = 12.6$ Hz, H-7'; δ_H 6.37, d, $J = 12.6$ Hz, H-7''; δ_H 6.47, d, $J = 12.6$ Hz, H-7''') exhibited vinylic coupling constants indicative of a *cis* configuration. Furthermore, the HMBC correlations of H-2'' (δ_H 6.86, br s) and H-6'' (δ_H 6.69) with C-3'' (δ_C 145.5) were indicative of a caffeoyl moiety attached to N-5 (Table 1). Thus, compound **1** was established as $N^1,N^{10}(Z,Z)$ -di-*p*-coumaroyl- $N^5(Z)$ -caffeoyl-spermidine.

Table 1. 1H and ^{13}C NMR Spectroscopic Data for Compound **1** (DMSO- d_6 , 500 MHz for 1H and 125 MHz for ^{13}C NMR)

position	type	1H [ppm]	^{13}C [ppm] ^a
1 and 10	NH	7.96 and 7.89 (m) ^b	
2 and 9	CH ₂	3.15 and 3.03 (m) ^b	38.0, 37.7, 36.2, 35.7
3, 7, and 8	CH ₂	1.27–1.77 (m)	28.0, 26.8, 26.3, 26.0, 25.6, 24.2
4 and 6	CH ₂	3.19–3.36 (m)	47.3, 45.4, 43.5, 42.0
1'/1''	C		125.9
2'/2''	CH	7.58 and 7.60 (each d, $J = 8.6$ Hz)	131.4
6'/6''	CH	6.71 (d, $J = 8.6$ Hz)	114.4
5'/5''	CH		115.5
4'/4''	C		157.5
7'/7''	CH	6.47 (d, $J = 12.6$ Hz)	135.9
8'/8''	CH	5.79 (d, $J = 12.6$ Hz)	120.7
9'/9''	C		165.8
1''	C		126.6
2''	CH	6.86 (br s)	115.5
3''	C		145.5 ^d
4''	C		144.3 ^d
5''	CH	6.68 ^c	115.0
6''	CH	6.69 ^c	120.2
7''	CH	6.37 and 6.33 (each d, $J = 12.6$ Hz)	131.8
8''	CH	5.88 and 5.85 (each d, $J = 12.6$ Hz)	120.5
9''	C		167.9

^a ^{13}C NMR chemical shifts were extracted from HSQC and HMBC spectra. ^bAssigned by 2D NOESY. ^cSignals partially overlapped. ^dAssigned on the basis of ^{13}C NMR shifts predicted by ACD/Laboratories software.

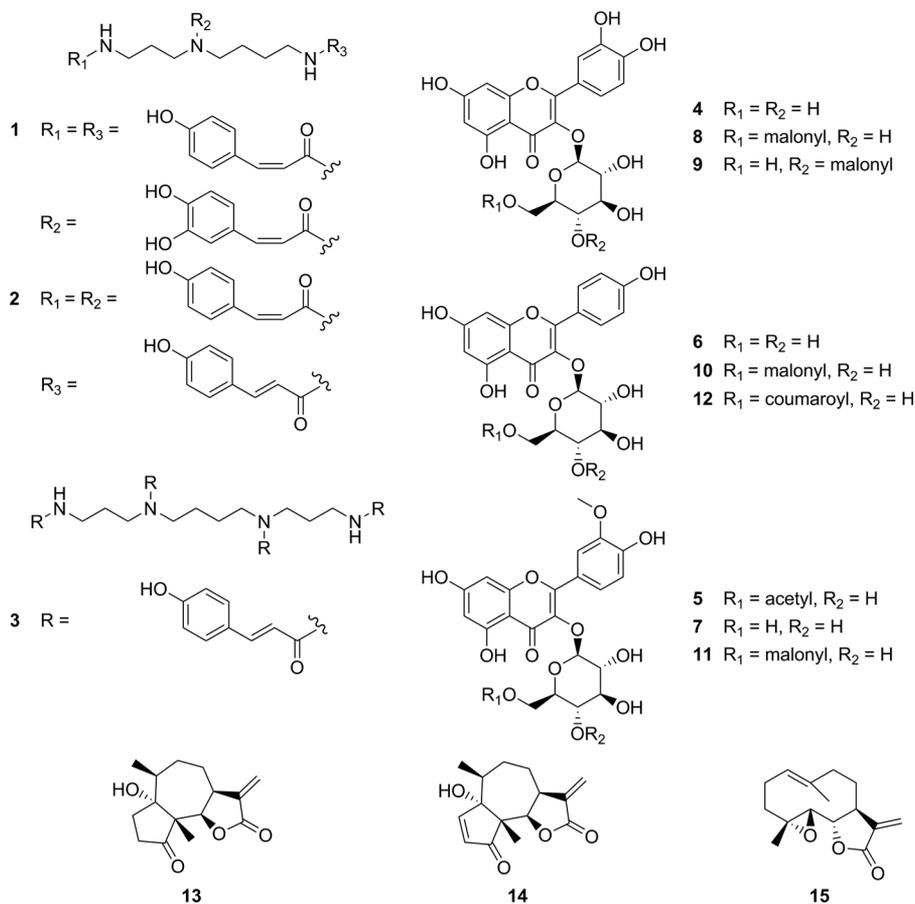
Compound **9** was obtained as an amorphous yellow solid. A molecular formula of $C_{24}H_{23}O_{15}$ was assigned on the basis of a molecular ion at m/z 551.1071 $[M + H]^+$ in the HRESIMS spectrum (calcd for $C_{24}H_{22}O_{15}$, 551.0959). 1D and 2D NMR spectra (Figures S7–S10, Supporting Information) were virtually superimposable to those of compound **4** (Table S6, Supporting Information) and to quercetin-3-*O*- β -D-glucopyranoside.³¹ The presence of a malonyl residue was indicated by HMBC correlations of a methylene (δ_H 3.34, s) with quaternary carbons C-1'' (δ_C 166.9) and C-3'' (δ_C 168.0). Its

attachment at C-4 of the glucopyranoside moiety was suggested by a downfield shift of glucose C-4 (δ_C 72.0 in **9**, vs δ_C 69.9 in **4**) and confirmed by the HMBC correlation of Glc-4 (δ_H 4.64, dd, $J = 9.5, 9.0$ Hz) with C-1'' (δ_C 166.9) (Table 2). Thus, compound **9** was established as quercetin 3-(4''-*O*-malonyl)- β -glucopyranoside (**9**).

Besides the polyamine conjugate **1**, the known polyamine derivatives $N^1,N^5(Z,Z),N^{10}(E)$ -tri-*p*-coumaroylspermidine (**2**) and $N^1,N^5,N^{10},N^{14}(E,E,E,E)$ -tetra-*p*-coumaroylspermine (**3**) were identified (Tables S4 and S5, Supporting Information). Their NMR and MS data were in good agreement with previously reported values.^{30,31} Additional HPLC peaks I–IV (Figure 1) corresponded to isomers of **1** and **2** and were tentatively assigned by comparison of online spectroscopic data with those of **1** and **2** (Table 1; Table S2, Supporting Information). Thus, the peaks with molecular ions m/z 584 (m/z 606 $[M + Na]^+$), m/z 600 (m/z 622 $[M + Na]^+$), and m/z 787 could be assigned to compounds **1**–**3** and their isomers (I–IV). It thus appears that conjugated polyamines occurred in all Asteraceae pollen, but were absent in pollen from other plant families (Figure S2, Supporting Information). From the MeOH extract of *A. artemisiifolia* several flavonoid glycosides (**4**–**11**) were identified by comparison of their NMR and MS data with reported values.^{32–36} They were determined to be derivatives of quercetin, kaempferol, and isorhamnetin (Table S3; Tables S6–S13, Supporting Information). However, STLs were absent in the THF or MeOH extracts prepared from five different lots of *A. artemisiifolia* pollen. From the THF extract of *A. pilostachya* pollen (Ap1) the two major peaks (m/z 263.0 $[M + H]^+$; m/z 264.9 $[M + H]^+$) that corresponded to the dominant peak in the ELSD trace ($t_R = 25.3$ min) were isolated (Figure S1, Supporting Information). The two compounds were obtained as a mixture (70:30) and were identified as the known sesquiterpene lactones coronopilin (**13**) and parthenin (**14**) (Tables S14 and S15, Supporting Information).^{37,38} In addition, the minor kaempferol glycoside **12** (not visible in Figure S2, Supporting Information) was obtained.³³ The absence of sesquiterpene lactones in *A. artemisiifolia* was confirmed by a comparison of HPLC profiles of pollen with those recorded with THF extracts from the aerial parts. The HPLC-ESIMS profiles of flowers and leaves showed peaks with molecular ions m/z 243.1 $[M + H]^+$, m/z 265.1 $[M + H]^+$, m/z 232.2 $[M + H]^+$, and m/z 234.1 $[M + H]^+$ and UV spectra ($\lambda_{max} \approx 230$ nm) that were indicative of additional sesquiterpene lactones (Figures S11 and S13, Supporting Information).

Non-Asteraceae Pollen Contain Electrophiles. Given the electrophilic nature of many sesquiterpene lactones including coronopilin (**13**) and parthenin (**14**), electrophiles in all pollen samples and in leaves and flowers of *A. artemisiifolia* were quantified with the aid of a validated spectrophotometric assay.³⁹ Electrophilic metabolites in pollen extracts were covalently trapped with *N*-acetyl-L-cysteine (NAC). The excess of NAC was treated with Ellman's reagent to form 5-thio-2-nitrobenzoate dianion (TNB²⁻), which was quantified by spectrophotometry at 412 nm. The concentration of electrophiles was expressed as parthenolide (**15**) (Figures S15–S17, Supporting Information). Since phenolic compounds like curcumin have been shown to form adducts with cysteine,⁴⁰ quercetin, quercetin 3-*O*- β -glucoside, caffeic acid, and **1** as typical representatives of pollen secondary metabolites were tested for reactivity with NAC. Monitoring of adduct formation by HPLC-ESIMS showed that only **15** formed an

Chart 1



adduct with NAC (Figure S18, Supporting Information). Extracts of *A. artemisiifolia* pollen (six lots; Table S16, Supporting Information), leaves (two lots), and flowers (two lots), *A. psilostachya* pollen (five lots, Table S17, Supporting Information), and pollen extracts from other Asteraceae and non-Asteraceae plant species were compared. Concentrations of electrophiles in *A. artemisiifolia* leaves and flowers were 10- to 20-fold higher compared to six different lots of pollen (Aa1–Aa6; Figure S19, Supporting Information). The five lots of *A. psilostachya* pollen (Ap1–Ap5) showed the highest concentration of electrophiles among all pollen samples that were analyzed. However, concentrations of coronopilin (13) and parthenin (14) varied considerably between these lots. In sample Ap1 they were the two dominant compounds of this type. Ap5 had the highest concentration of electrophiles, but 13 was present only in traces and 14 was not detected. In Ap2, compounds 13 and 14 were absent (Figure 3, Figure S20, Supporting Information). Compared to *A. psilostachya*, pollen from other *Ambrosia* and *Artemisia* species and from other genera of the Asteraceae showed considerably lower concentrations of electrophiles. Interestingly, pollen from *Phleum pratense*, *Zea mays*, *Fagus sylvatica*, and *Urtica dioica* (all non-Asteraceae) contained electrophiles at concentrations comparable to some Asteraceae pollen (Figure S19, Supporting Information). The limited amounts of pollen from these species precluded a purification and structural characterization of electrophiles.

THF extracts from pollen of *A. artemisiifolia* (Aa1) and *A. psilostachya* (Ap1), the mixture of coronopilin (13) and parthenin (14), and a mixture of conjugated polyamines (1–

3, and I–IV) were tested in murine afferent neurons and tracheal ring preparations.

Electrophiles Increase $[Ca^{2+}]_i$ in Murine Afferent Neurons. The intracellular calcium concentration $[Ca^{2+}]_i$ was measured with the aid of calcium imaging after stimulation of murine afferent neurons with *A. artemisiifolia* and *A. psilostachya* pollen extracts and with a mixture of compounds 13 and 14. Conjugated polyamines were not soluble under the assay conditions and were not tested. At the final DMSO concentration (0.1%), which was also used as a vehicle control, only a low number of cinnamaldehyde-responsive (CA+; 8%, $n = 107$) and cinnamaldehyde-unresponsive (CA–; 2%, $n = 90$) neurons showed an increase in $[Ca^{2+}]_i$. *A. psilostachya* extract and compounds 13 and 14 increased $[Ca^{2+}]_i$ in 32% ($n = 207$) and 37% ($n = 372$) of CA+ neurons ($p < 0.05$ compared to vehicle control) and only in 7% ($n = 197$) or 6% ($n = 311$) of CA– neurons, respectively (not significant compared to vehicle control). The *A. artemisiifolia* extract increased $[Ca^{2+}]_i$ in only 15% ($n = 144$) of CA+ neurons (not significant compared to vehicle control and CA– neurons) (Figure 2). For the two pollen extracts tested, the number of activated neurons appeared to correlate with the concentration of electrophiles.

Sesquiterpene Lactones Induce Relaxation of Murine Tracheal Rings. Next, we studied the effect of *A. psilostachya* extract, a mixture of coronopilin (13) and parthenin (14), and a mixture of conjugated polyamines (1–3 and I–IV) on murine tracheal ring preparations. Tracheal tone in naïve and methacholine-constricted tracheal rings was recorded in response to vehicle (0.5% DMSO), extract, and compounds. In naïve tracheal rings *A. psilostachya* extract and conjugated

Table 2. ^1H and ^{13}C NMR Spectroscopic Data for Compound 9 (DMSO- d_6 , 500 MHz for ^1H and 125 MHz for ^{13}C NMR)

position	type	^1H [ppm]	^{13}C [ppm]
2	C		156.7
3	C		133.2
4	C		177.4
5	C		161.2
6	CH	6.21 (br s)	98.8
7	C		164.3
8	CH	6.41 (br s)	93.7
9	C		156.4
10	C		104.0
1'	C		121.1
2'	CH	7.61 (d, $J = 2.2$ Hz)	116.3
3'	C		144.9
4'	C		148.6
5'	CH	6.87 (d, $J = 8.5$ Hz)	115.3
6'	CH	7.56 (dd, $J = 8.5, 2.2$ Hz)	121.5
1''	C		166.9
2''	CH ₂	3.34 (s)	40.6
3''	C		168.0
Glc-1	CH	5.53 (d, $J = 7.6$ Hz)	101.1
Glc-2	CH	3.40 ^a	74.4
Glc-3	CH	3.51 (dd, $J = 9.0, 9.0$ Hz)	74.4
Glc-4	CH	4.64 (dd, $J = 9.5, 9.0$ Hz)	72.0
Glc-5	CH	3.37 ^a	75.0
Glc-6	CH ₂	3.43 (dd, $J = 11.8, 2.0$ Hz) 3.25 (dd, $J = 11.8, 6.0$ Hz)	60.6

^aSignals partially overlapped.

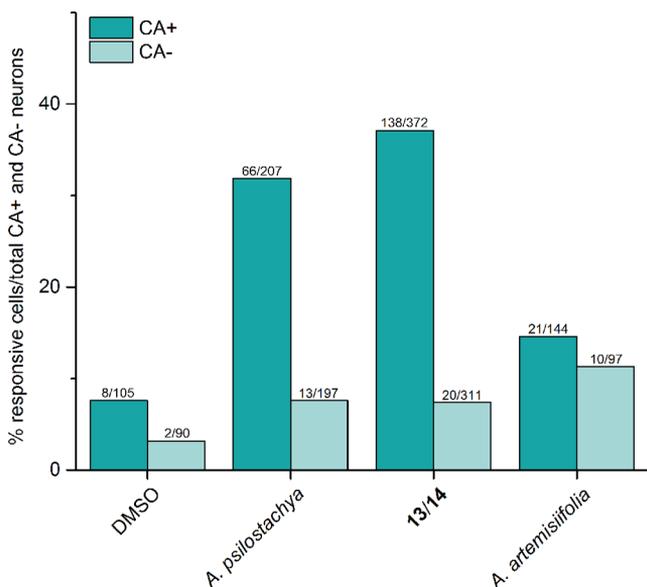


Figure 2. Stimulation of murine dorsal root ganglia neurons with pollen extracts and sesquiterpene lactones. *A. psilostachya* THF (lot Ap1) extract and coronopilin (13)/parthenin (14) increased the $[\text{Ca}^{2+}]_i$ in 32% and 37% of CA⁺ neurons, respectively. *A. artemisiifolia* THF (lot Aa1) extract induced an increase of $[\text{Ca}^{2+}]_i$ in 15% of CA⁺ neurons. Cinnamaldehyde (CA) was used as positive control. CA⁺ and CA⁻ designate cinnamaldehyde-responsive and -unresponsive neurons, respectively. DMSO (0.5% in assay) was used as vehicle control.

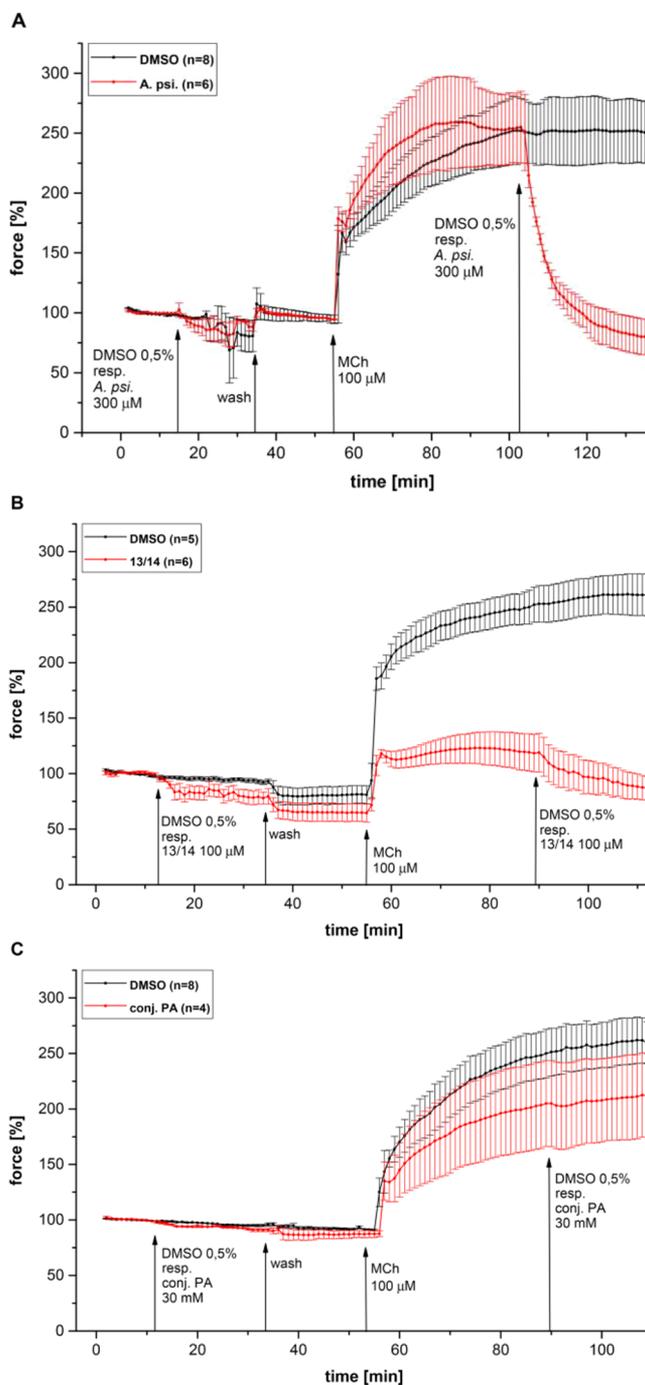


Figure 3. Mean cumulative time-response curves for the constrictive response of murine tracheal rings in organ bath experiments to *A. psilostachya* THF (lot Ap1) extract (A), coronopilin (13)/parthenin (14) (B), and conjugated polyamines 1–3 and I–IV (C). DMSO (0.1% in assay) was used as vehicle control, and methacholine was used to induce constriction of tracheal rings.

polyamines had no significant effect on tracheal tone (Figures 3A and C, Figure 4A). In contrast, 13/14 induced significant relaxation ($p = 0.017$; Figure 3B). After washing, the tracheal rings were constricted with methacholine (muscarinic receptor agonist known to induce a strong tracheal constriction). Interestingly, the magnitude of constriction in the 13/14-pretreated tracheal rings was significantly lower compared to control ($p < 0.05$). Methacholine-constricted tracheal rings were treated with either *A. psilostachya* extract (Figure 3A), 13/

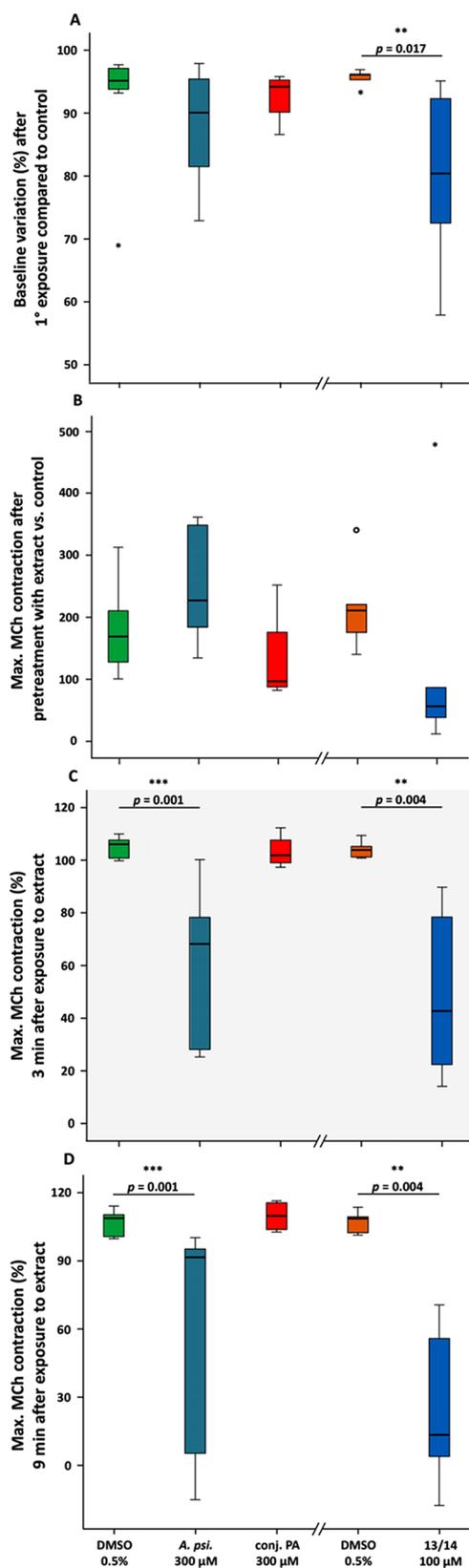


Figure 4. Effect of pollen extracts and isolated compounds on naïve and methacholine-constricted murine tracheal rings. All values were normalized to the maximal response to methacholine (MCh). The pretreatment of tracheal rings with conjugated polyamines 1–3 and I–IV (conj. PAs; 30 mM, $n = 4$) and *A. psilostachya* THF extract (lot Ap1; 300 µM, $n = 6$) did not significantly differ from the negative control (0.1% DMSO). The mixture of coronopilin (13)/parthenin

Figure 4. continued

(14) (100 µM, $n = 6$) induced weak but significant relaxation in naïve tracheal rings ($p = 0.017$) (A). The maximal MCh-induced tracheal constriction remained unaltered after pretreatment with conj. PAs (30 mM, $n = 4$) and *A. psilostachya* (300 µM, $n = 6$), and was comparable to pretreatment with 0.1% DMSO. In contrast, a significant ($* = p < 0.05$) decrease of the maximal MCh-induced constriction was observed by pretreatment of naïve tracheal rings with coronopilin (13)/parthenin (14) (B). The second exposure of MCh-precontracted tracheal rings with *A. psilostachya* ($p = 0.001$) and coronopilin (13)/parthenin (14) ($p = 0.004$), at time points 3 min (C) and 9 min (D), resulted in a time-dependent relaxation.

14 (Figure 3B), or conjugated polyamines (Figure 3C). Conjugated polyamines had no effect on methacholine-constricted tracheal rings (Figure 3C), while *A. psilostachya* extract and 13/14 both caused significant relaxation ($p < 0.05$).

An initial large phytochemical profiling of plant pollen was conducted, with a focus on pollen known to be allergic to varying degrees. Caffeoyl and coumaroyl conjugates of spermine and spermidine were found to be characteristic of the family Asteraceae. The compounds coronopilin (13) and parthenin (14) were detected only in the pollen of *A. psilostachya*, albeit at varying concentrations in the different lots. The six lots of *A. artemisiifolia* pollen that were analyzed herein did not contain STLs present in the leaves and flowers. Concentrations of electrophiles were found to be consistently low among the *A. artemisiifolia* pollen lots analyzed in our study, but were 15- to 20-fold higher in flowers and leaves. These findings are in contrast to earlier studies reporting a high concentration of STLs in pollen of *A. artemisiifolia* collected in Italy.⁴¹ Differing STL profiles of *A. artemisiifolia* collected in the US and in Europe suggest the occurrence of distinct chemotypes^{41,42} that might also comprise accessions with low concentrations of STLs in pollen. In turn, the concentrations of electrophiles in pollen of other species in the Asteraceae were variable but consistently lower than in *A. psilostachya*. Interestingly, metabolites with electrophilic properties were also detected in pollen from non-Asteraceae species, and their concentrations were comparable to those found in pollen of most species in the Asteraceae. The nature of these electrophilic compounds remains to be investigated.

Testing of pollen extracts and pure compounds for increase of $[Ca^{2+}]_i$ in murine afferent neurons and for constriction in murine tracheal rings led to somewhat conflicting results. TRPA1-mediated Ca^{2+} influx in afferent neurons was proportional to the concentration of electrophiles in *A. artemisiifolia* and *A. psilostachya* pollen extracts, and the rise of $[Ca^{2+}]_i$ would be expected to induce tracheal constriction.^{24,43} However, *A. psilostachya* and the mixture of coronopilin (13) and parthenin (14) induced relaxation in both methacholine-constricted and native tracheal rings. This effect could possibly be due to an activation of bitter receptors (T2Rs), via a Ca^{2+} -mediated activation of the large conductance Ca^{2+} -dependent K^+ channel (BK_{Ca}) channel and subsequent hyperpolarization and smooth muscle relaxation.^{27,44,45} The relaxation of methacholine-constricted trachea caused by these compounds could possibly also be linked to the disruptive character of electrophiles on synaptic or neuromuscular junction proteins (electrophilic neurotoxicants).⁴⁶ Further experiments are needed to elucidate the molecular mechanism behind these biological observations. Polyamine conjugates did not affect the tracheal tone, in

contrast to the known effects of free polyamines spermidine and spermine.^{28,47} Given the high enzymatic activity in airway epithelial tissue,⁴⁸ a release of physiologically relevant concentrations of free polyamines from polyamine conjugates is conceivable but requires further investigation. Also, the chemical nature of electrophiles in the pollen of plants other than in the Asteraceae and their possible activities need to be explored.

EXPERIMENTAL SECTION

General Experimental Procedures. NMR spectra were recorded on a Bruker AVANCE III 500 MHz spectrometer (Bruker BioSpin) equipped with a 1 mm TXI microprobe (¹H detected experiments) or a 5 mm BBO probe (for ¹³C NMR spectra) operating at 500 (¹H) and 125 MHz (¹³C), respectively. All experiments were carried out at 291.15 K. Chemical shifts are reported as δ values (ppm) with the residual solvent signal as internal reference, J in Hz. Standard pulse sequences of the software package Topspin 3.0 were used. HPLC-PDA-MS-ELSD profiling was carried out on an HPLC system consisting of a degasser, binary high-pressure mixing pump, column thermostat, PDA detector (SDP-M20A) with thermostated UV cell, and a triple quadrupole LCMS-8030 with ESI interface coupled via T split (all Shimadzu). Both positive- and negative-ion HPLC-ESIMS chromatograms were recorded simultaneously in the range of m/z 160 to 1500; capillary voltage: 4.5 V; scan speed: 6000 u/s; event time: 0.150 s. The second line of the T-splitter was connected to a model 3300 ELSD (Alltech). ELSD settings were as follows: nitrogen flow: 2.5 L/min, temperature: 55 °C, detector gain: 8. Data acquisition and analysis was performed with LabSolutions software (Shimadzu). HRESIMS were recorded with a microTOF ESIMS system (Bruker Daltonics) connected to a 1100 series HPLC (Agilent). Mass spectra were recorded in the range m/z 160 to 1500 in the positive-ion mode with the aid of microTOF control software 1.1 (Bruker Daltonics). Mass calibration was performed with a solution of 0.1% formic acid in 2-propanol–water (1:1) containing 5 mM NaOH. HPLC grade MeOH, MeCN (Scharlau), and H₂O (EASY-pure II, Barnstead, Dubuque) were used for HPLC separations. A 1100 series instrument with PDA detector (Agilent) was used for semipreparative isolation and purification. HPLC solvents used for analytical separation contained 0.1% HCOOH (Scharlau). HPLC samples were dissolved in DMSO (Scharlau). DMSO-*d*₆ and MeOH-*d*₄ (100 atom %D) for NMR experiments were purchased from Amar Chemicals. Sephadex LH-20 (GE Healthcare) column chromatography was performed using a B-688 pump (Büchi) and a SuperFrac fraction collector (Pharmacia Biotech). For extraction and column chromatography, technical grade solvents were used after distillation. Analytical scale extraction of pollen material was performed by pressurized liquid extraction using an ASE 200 instrument with attached solvent controller (Dionex). Preparative scale extraction was performed by maceration. Spectrophotometric measurements were performed with a Lambda 35 instrument (PerkinElmer). Samples were analyzed in standard 1 cm cuvettes.

Plant Material. Pollen material was purchased from Allergon AB (Ängelholm, Sweden) and GREER Laboratories (Lenoir, NC, USA), and voucher specimens (969–1010) were deposited at the Department of Pharmaceutical Biology, University of Basel, Switzerland. A complete list of pollen and sample-related information are given in [Supporting Information](#), Tables S1, S16, and S17. *A. artemisiifolia* leaves and flowers, collected on the outskirts of Novara (Italy), were provided by Prof. Giovanni Appendino, Università degli Studi del Piemonte Orientale, Italy. Voucher specimens designated with the codes 1008, 1009 (flowers), and 1010, 1011 (leaves) are deposited at the Department of Pharmaceutical Biology, University of Basel.

HPLC Separation. Analytical scale separation was performed at 40 °C on a SunFire C₁₈ (150 × 3.0 mm; 3.5 μ m; Waters) column equipped with a guard column (20 × 3.0 mm) with 0.1% HCOOH in H₂O (solvent A) and 0.1% HCOOH in MeOH (solvent B) as mobile phase. The following gradient profile was used: 0–2 min 5% B isocratic, 2–50 min 5–100% B, 52–62 min 100% B. The flow rate was

0.4 mL/min, and sample injection volume was 2 μ L. Semipreparative separations were carried out at 25 °C on a SunFire C₁₈ (150 × 10 mm; 5 μ m; Waters) column equipped with a precolumn (10 × 10 mm). Gradients of water and MeCN were used. The flow rate was 4 mL/min.

Extraction and Isolation. Dried *A. artemisiifolia* leaves and flowers were ground with a ZM 1 ultracentrifugal mill. Aliquots (300 mg) of *A. artemisiifolia* pollen, leaves, and flowers, *A. psilostachya* pollen, and pollen samples from other species were separately packed into steel cartridges and extracted by pressurized liquid extraction. Samples were sequentially extracted with *n*-hexane, THF, and MeOH (3 × 5 min cycles per solvent, 120 °C, 70 bar). Extracts were dried on a rotary evaporator. For preparative scale extraction, *A. artemisiifolia* pollen (Aa1) (95 g) was extracted at room temperature sequentially with *n*-hexane, THF, and MeOH (3 × 8 h per solvent, 500 mL each). *n*-Hexane, THF, and MeOH extracts were dried on a rotary evaporator, yielding 10.8 g (11.3%), 3.0 g (3.2%), and 17.0 g (17.9%), respectively. THF (2.8 g) and MeOH (12.0 g) extracts of *A. artemisiifolia* were both separated on a Sephadex LH-20 column (7 × 100 cm i.d.) eluted with MeOH at a flow rate of 1.5 mL/min. Fractions of 15 min were collected. A total of 10 fractions (1–10) were combined based on TLC analysis. Fractions T5 (232 mg) and T6 (283 mg) from the THF extract contained conjugated polyamines, and fractions M4 (441 mg), M5 (290 mg), and M6 (215 mg) from the MeOH extract consisted mainly of flavonoids. All fractions were submitted to semipreparative HPLC. Fraction T5 afforded compound 1 (3.1 mg; t_R = 27.0 min), compound 2 (4.2 mg; t_R = 28.4 min), and mixture of polyamine isomers I–IV (10.4 mg). Compound 3 (2.7 mg; t_R = 32.4 min) was obtained from fraction T6. Compounds 4 (6.5 mg; t_R = 26.1 min), 8 (4.0 mg; t_R = 29.6 min), and 9 (6.3 mg; t_R = 30.1 min) were purified from fraction M4, and compounds 5 (4.6 mg; t_R = 27.2 min) and 10 (2.1 mg; t_R = 30.5 min) from fraction M5. Compound 6 (4.3 mg; t_R = 27.7 min), 7 (5.2 mg; t_R = 29.1 min), and 11 (4.9 mg; t_R = 30.9 min) were isolated from fraction M6. Compound 12 (0.9 mg; t_R = 30.5 min) and a mixture of 13 and 14 (1.5 mg; t_R = 25.6 min) were isolated from the *A. psilostachya* (Ap1) THF extract (12.4 mg) obtained from 970 mg of pollen. Retention times given for compounds 1–14 refer to the analytical HPLC conditions given above.

*N*¹,*N*¹⁰(*Z,Z*)-*di-p-Coumaroyl-N*⁵(*Z*)-*caffeoylspermidine* (1): white, amorphous solid; $[\alpha]_D^{22}$ –37 (c 0.8, DMSO); UV (DMSO) λ_{max} (log ϵ) 272 (3.85) nm; ¹H NMR (DMSO, 500 MHz) and ¹³C NMR (DMSO, 125 MHz), see [Table 1](#); HRESIMS m/z 600.2400 [M + H]⁺ (calcd for C₃₄H₃₇N₃O₇, 600.2409).

Quercetin 3-(4"-O-malonyl)- β -glucopyranoside (9): yellow, amorphous solid; $[\alpha]_D^{22}$ –79 (c 0.7, MeOH); UV (DMSO) λ_{max} (log ϵ) 260 (3.62), 352 (3.88) nm; ¹H NMR (DMSO, 500 MHz) and ¹³C NMR (DMSO, 125 MHz), see [Table 2](#); HRESIMS m/z 551.1071 [M + H]⁺ (calcd for C₂₄H₂₂O₁₅, 551.0959).

Spectrophotometric Quantification of Electrophiles. *N*-Acetyl-L-cysteine and parthenolide were from Sigma-Aldrich, and 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB or Ellman's reagent), triethylamine (TEA), and anhydrous sodium phosphate (HNa₂PO₄) from Fluka. Sodium dihydrogen phosphate (H₂NaPO₄) was from Merck. As a source of cysteine residues, a stock solution of *N*-acetyl-L-cysteine in MeOH (0.429 mg/mL) was prepared daily. The stock solution of Ellman's reagent was obtained by dissolving 16 mg of DTNB in 1 mL of sodium phosphate buffer (pH 8; 74 mg of NaH₂PO₄·H₂O and 1.68 g of anhydrous Na₂HPO₄ in 100 mL of distilled water). TEA was used undiluted. For the calibration curve, a stock solution of parthenolide in MeOH was prepared at a concentration of 1.13 mg/mL. Five different concentrations of parthenolide (15) (ranging from 4.55 to 45.5 μ M) were prepared by mixing 200 μ L of cysteine stock solution, 40 μ L of TEA, and incubation of 10, 20, 30, 50, and 100 μ L of parthenolide stock solution for 40 min at room temperature. At the end of the incubation, 30 μ L of Ellman's reagent was added, and the solution was diluted with MeOH to 10.0 mL (reaction described in Figure S15, [Supporting Information](#)). A blank solution was prepared by mixing 200 μ L of MeOH, 40 μ L of TEA, and 30 μ L of Ellman's reagent and diluted to 10.0 mL with MeOH. The absorbance was measured at 412 nm

against the blank. The calibration curve $y = -0.0047x + 0.8318$ showed a correlation coefficient (R^2) of 0.9972 (Figure S16, Supporting Information) and was used for the calculation of % electrophiles in the extract. For the determination of electrophiles in pollen extracts, the following procedure was used: *n*-hexane, THF, and MeOH extracts were dissolved in DMSO at a concentration of 25 mg/mL and filtered if needed. The samples were prepared by mixing 200 μ L of cysteine stock solution and 40 μ L of TEA and incubating 20 μ L of extract (MeOH as control) for 40 min, at room temperature. After the incubation time, 30 μ L of Ellman's reagent was added, and the mixture was diluted with MeOH to 10.0 mL. The concentration of electrophiles in the extracts was calculated with the formula described in Figure S17, Supporting Information.

Measurement of Intracellular Ca^{2+} Increase. The intracellular $[\text{Ca}^{2+}]_i$ measurements were performed with dissociated jugular/nodose neurons. The coverslip was loaded with Fura-2 AM (8 μ m) in L-15 medium containing 20% fetal bovine serum and incubated for 30 min at 37 °C. The coverslip was placed in a Biopetechs Delta T (Butler) that was superfused with Locke's solution [at 37 °C; composed of (mM): 136 NaCl, 5.6 KCl, 1.2 MgCl_2 , 2.2 CaCl_2 , 1.2 NaH_2PO_4 , 14.3 NaHCO_3 , and 10 dextrose; pH 7.3–7.4] for 20 min before the experiment by an infusion pump (8 mL/min). Changes in intracellular $[\text{Ca}^{2+}]_i$ were measured by digital microscopy (Leica DMS5000B) equipped with an X-Cite 200DC illumination system (Waltham) for ratiometric recording of single cells. A field of cells was monitored by sequential dual excitation, 360 and 380 nm, and the analysis of the image ratios used methods previously described. Ratio images were acquired every second. Superfused buffer was stopped 20 s prior to test material applications. In each experiment, the cells on the coverslip were exposed to vehicle (0.1% DMSO or EtOH), cinnamaldehyde (100 μ M), capsaicin (1 μ M), and KCl (80 mM) for 1 min each. Between each stimulus, the cells were washed with fresh buffer for at least 2 min for the cells to recover prior to the addition of the second stimulus. Each set of images for the Ca^{2+} measurements also included a bright field image of the field of cells under study. Cells that had an average diameter (long and short axis) of over 15 μ m were analyzed. Those cells that failed to respond to capsaicin were considered healthy neurons only if they responded to KCl with a rapid rise in Ca^{2+} .

Assessment of Tracheal Ring Constriction. Tracheal rings were isolated from male and female C57BL/6 mice (age range: 10 to 16 weeks), respectively. Until use, the segments were stored at +4 °C in MEM gassed with a 95% O_2 –5% CO_2 gas mixture. Stainless steel wire hooks were inserted from each side into the tracheal segments, connected to an isometric force transducer (Hugo Basile), placed in an isolated organ bath filled with 15 mL of MEM, and oxygenated with a 95% O_2 –5% CO_2 gas mixture at 37 °C. Tension due to isometric changes in airway tone was recorded, processed, and analyzed by LabChart 7 software (AD Instruments). The experiment started after equilibrating the tracheas for at least 30 min in the bath at a tension of 0.5 g. Airway constriction was measured after application of DMSO (vehicle, 0.5%), *A. psilostachya* THF extract (300 μ M), 1–3, and I–IV (30 mM), or a mixture of compounds 13 and 14 (100 μ M), respectively, which were allowed to incubate for at least 20 min. After washing, methacholine (100 μ M) was added. After maximal constriction was reached, DMSO (vehicle, 0.5%), *A. psilostachya* THF extract (300 μ M), 1–3, and I–IV (30 mM), or the mixture of 13 and 14 (100 μ M), respectively, was added again, and the smooth muscle contraction was measured.

All animal experiments were performed in accordance with the German animal welfare law and had been declared to the Animal Welfare Officer of the University of Giessen (Registration No. 426 M). The animal housing facility was licensed by the local authorities. The methods used to euthanize the animals humanely were consistent with the recommendations of the AVMA Guidelines for the Euthanasia of Animals.

Statistical Analysis. Values in the text and figures are expressed as the arithmetic means \pm the standard error of the mean (SEM) from experiments with tracheas from 4 to 10 independent mice. Increasing isometric smooth muscle tone corresponds to increasing isometric

force expressed as force expressed in grams (g) and converted to %. Data were evaluated statistically in an analysis of variance and the Kruskal–Wallis or Mann–Whitney–U post-test, depending on the number of groups that were compared. Differences in the abundance of neurons responding to cinnamaldehyde or pollen extracts, respectively, were calculated by using the chi-square test. A difference was considered statistically significant when the probability value p was below 0.05 ($p < 0.05$).

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jnatprod.7b00495.

HPLC profiles of pollen, leaves, and flowers extracts; NMR spectra (1D and 2D) of the new compounds 1 and 9; tables with NMR and HRESIMS data of known compounds 2–8 and 10–14; calibration curve and formula used for spectroscopic quantification; table containing information on pollen origin (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: matthias.hamburger@unibas.ch. Tel: +41-61-207-1425. Fax: +41-61-207-1474.

ORCID

Maria De Mieri: 0000-0001-5567-2072

Matthias Hamburger: 0000-0001-9331-273X

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Prof. Dr. Giovanni Appendino, University Amedeo Avogadro, Novara, Italy, for generous provision of *A. artemisiifolia* leaf and flower samples. Calcium imaging and organ bath experiments were funded by the German Center for Lung Research (Deutsches Zentrum für Lungenforschung, DZL).

■ REFERENCES

- (1) Wenzel, S. E. *Nat. Med.* **2012**, *18*, 716–725.
- (2) Yach, D.; Hawkes, C.; Gould, C. L.; Hofman, K. J. *J. Am. Med. Assoc.* **2004**, *291*, 2616–2622.
- (3) The Global Asthma Report 2014. <http://www.globalasthmareport.org/2011/about/about.php> (accessed April 29, 2017).
- (4) Institute for Health Metrics and Evaluation (IHME). <http://ghdx.healthdata.org/gbd-results-tool> (accessed April 29, 2017).
- (5) Masoli, M.; Fabian, D.; Holt, S.; Beasley, R. *Allergy* **2004**, *59*, 469–478.
- (6) Galli, S. J.; Tsai, M.; Piliponsky, A. M. *Nature* **2008**, *454*, 445–454.
- (7) Guterma, J.; Bewersdorff, M.; Traidl-Hoffmann, C.; Ring, J.; Mueller, M. J.; Behrendt, H.; Jakob, T. *J. Allergy Clin. Immunol.* **2007**, *120*, 293–299.
- (8) Gilles, S.; Fekete, A.; Zhang, X.; Beck, I.; Blume, C.; Ring, J.; Schmidt-Weber, C.; Behrendt, H.; Schmitt-Kopplin, P.; Traidl-Hoffmann, C. *J. Allergy Clin. Immunol.* **2011**, *127*, 454–461.
- (9) Weber, R. W. *J. Allergy Clin. Immunol.* **2003**, *112*, 229–239.
- (10) Verdino, P. *Clin. Rev. Allergy Immunol.* **2006**, *30*, 73–95.
- (11) Rzepecka-Stojko, A.; Stojko, J.; Kurek-Górecka, A.; Górecki, M.; Kabała-Dzik, A.; Kubina, R.; Moździerz, A.; Buszman, E. *Molecules* **2015**, *20*, 21732–21749.
- (12) Strohl, M. J.; Seikel, M. K. *Phytochemistry* **1965**, *4*, 383–399.

- (13) Almaraz-Abarca, N.; da Graça Campos, M.; Avila-Reyes, J. A.; Naranjo-Jimenez, N.; Herrera Corral, J.; Gonzalez-Valdez, L. S. *J. Food Compos. Anal.* **2007**, *20*, 119–124.
- (14) Serra Bonvehi, J.; Soliva Torrento, M.; Centelles Lorente, E. *J. Agric. Food Chem.* **2001**, *49*, 1848–1853.
- (15) Aloisi, I.; Cai, G.; Serafini-Fracassini, D.; Del Duca, S. *Amino Acids* **2016**, *48*, 2467–2478.
- (16) Garrido, D.; Chibi, C.; Matilla, A. *J. Plant Physiol.* **1995**, *145*, 731–735.
- (17) Falasca, G.; Franceschetti, M.; Bagni, N.; Altamura, M. M.; Biasi, R. *Plant Physiol. Biochem.* **2010**, *48*, S65–S73.
- (18) Wu, J.; Shang, Z.; Wu, J.; Jiang, X.; Moschou, P. N.; Sun, W.; Roubelakis-Angelakis, K. A.; Zhang, S. *Plant J.* **2010**, *63*, 1042–1053.
- (19) Paupière, M. J.; van Heusden, A. W.; Bovy, A. G. *Metabolites* **2014**, *4*, 889–920.
- (20) Liu, B.; Escalera, J.; Balakrishna, S.; Fan, L.; Caceres, A. I.; Robinson, E.; Sui, A.; McKay, M. C.; McAlexander, M. A.; Herrick, C. A.; Jordt, S. E. *FASEB J.* **2013**, *27*, 3549–3563.
- (21) Tóth, B. I.; Oláh, A.; Szöllosi, G.; Bíró, T. *Br. J. Pharmacol.* **2014**, *171*, 2568–2581.
- (22) Bessac, B. F.; Jordt, S. E. *Physiology* **2008**, *23*, 360–370.
- (23) Karashima, Y.; Damann, N.; Prenen, J.; Talavera, K.; Segal, A.; Voets, T.; Nilius, B. *J. Neurosci.* **2007**, *27*, 9874–9884.
- (24) Nassenstein, C.; Kevin, K.; Taylor-Clark, T.; Kollarik, M.; MacGlashan, D. M.; Braun, A.; Udem, B. J. *J. Physiol.* **2008**, *586*, 1595–1604.
- (25) Lee, L.-Y. *J. Physiol.* **2010**, *588*, 747–748.
- (26) Erle, D. J.; Sheppard, D. *J. Cell Biol.* **2014**, *205*, 621–631.
- (27) Devillier, P.; Naline, E.; Grassin-Deyle, S. *Pharmacol. Ther.* **2015**, *155*, 11–21.
- (28) Radauer, C.; Breiteneder, H. *J. Allergy Clin. Immunol.* **2006**, *117*, 141–147.
- (29) Bozicevic, A.; Dobrzynski, M.; De Bie, H.; Garo, E.; Gafner, F.; Hamburger, M. *Anal. Chem.*, submitted.
- (30) Zhao, G.; Qin, G.-W.; Gai, Y.; Guo, L.-H. *Chem. Pharm. Bull.* **2010**, *58*, 950–952.
- (31) Jung, M.; Choi, J.; Chae, H.-S.; Cho, J. Y.; Kim, Y.-D.; Htwe, K. M.; Lee, W.-S.; Chin, Y.-W.; Kim, J.; Yoon, K. D. *Molecules* **2015**, *20*, 358–365.
- (32) Wu, T.; Abdulla, R.; Yang, Y.; Aisa, H. A. *Chem. Nat. Compd.* **2008**, *44*, 370–371.
- (33) Olennikov, D. N.; Kashchenko, N. I. *Chem. Nat. Compd.* **2013**, *49*, 833–840.
- (34) Rusanov, K.; Garo, E.; Rusanova, M.; Fertig, O.; Hamburger, M.; Atanassov, I.; Butterweck, V. *Planta Med.* **2014**, *80*, 1657–1664.
- (35) Luyen, B. T. T.; Tai, B. H.; Thao, N. P.; Cha, J. Y.; Lee, H. Y.; Lee, Y. M.; Kim, Y. H. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 266–269.
- (36) Burkard, V. W.; Galensa, R.; Herrmann, K. *Phytochemistry* **1989**, *28*, 663–664.
- (37) Athar, A.; Diduck, C.; Udenigwe, C. C.; Zahid, S.; Decken, A. *ARKIVOC* **2007**, 195–203.
- (38) Das, B.; Reddy, K. R.; Ravikanth, B.; Sarma, A. V. S.; Sridhar, B. *Helv. Chim. Acta* **2008**, *91*, 1137–1143.
- (39) Salapovic, H.; Geier, J.; Reznicek, G. *Sci. Pharm.* **2013**, *81*, 807–818.
- (40) Avonto, C.; Tagliatalata-Scafati, O.; Pollastro, F.; Minassi, A.; Di Marzo, V.; De Petrocellis, L.; Appendino, G. *Angew. Chem., Int. Ed.* **2011**, *50*, 467–471.
- (41) Tagliatalata-Scafati, O.; Pollastro, F.; Minassi, A.; Chianese, G.; De Petrocellis, L.; Di Marzo, V.; Appendino, G. *Eur. J. Org. Chem.* **2012**, *2012*, 5162–5170.
- (42) Sturgeon, C. M.; Craig, K.; Brown, C.; Rundle, N. T.; Andersen, R. J.; Roberge, M. *Planta Med.* **2005**, *71*, 938–943.
- (43) Chen, J.; Sanderson, M. J. *J. Physiol.* **2017**, *595*, 3203–3218.
- (44) Liggett, S. *Trans. Am. Clin. Climatol. Assoc.* **2014**, *125*, 64–75.
- (45) Deshpande, D. A.; Wang, W. C. H.; McIlmoyle, E. L.; Robinett, K. S.; Schillinger, R. M.; An, S. S.; Sham, J. S. K.; Liggett, S. B. *Nat. Med.* **2010**, *16*, 1299–1304.
- (46) LoPachin, R. M.; Barber, D. S. *Toxicol. Sci.* **2006**, *94*, 240–255.
- (47) North, L. M.; Grasmann, H.; Khanna, N.; Inman, M. D.; Gauvreau, G. M.; Scott, J. A. *Am. J. Respir. Cell Mol. Biol.* **2013**, *48*, 694–702.
- (48) Trivedi, N. N.; Caughey, G. H. *Am. J. Respir. Cell Mol. Biol.* **2010**, *42*, 257–267.

Supporting Information

Secondary Metabolites in Allergenic Plant Pollen Modulate Afferent Neurons and Murine Tracheal Rings

Alen Bozicevic[†], Maria De Mieri[†], Christina Nassenstein[‡], Silke Wiegand[‡], Matthias Hamburger^{†,*}

[†]Division of Pharmaceutical Biology, University of Basel, Klingelbergstrasse 50, CH-4056 Basel, Switzerland

[‡]Institute of Anatomy and Cell Biology, Justus-Liebig-University, Aulweg 123, D-35385 Giessen, Germany

[†]Member of the German Center for Lung Research (Deutsches Lungenzentrum, DZL)

Email addresses:

Corresponding author: matthias.hamburger@unibas.ch

Co-authors: alen.bozicevic@unibas.ch, maria.demieri@unibas.ch,

Christina.Nassenstein@anatomie.med.uni-giessen.de, Silke.Wiegand@anatomie.med.uni-giessen.de

Contents

Figure S1. Comparative HPLC-PDA-ESIMS-ELSD analysis of THF extracts from pollen of 30 plant species.	4
Figure S2. 2D heat map of HPLC-ESIMS+ metabolite profiles.	7
Figure S3. ¹ H-NMR spectrum of compound 1	8
Figure S4. ¹ H- ¹ H COSY spectrum of compound 1	8
Figure S5. Overlay of HSQC and HMBC spectra of compound 1	9
Figure S6. 2D ¹ H- ¹ H ROESY spectrum of compound 1	9
Figure S7. ¹ H-NMR spectrum of compound 9	10
Figure S8. ¹ H- ¹ H COSY spectrum of compound 9	10
Figure S9. Overlay of HSQC and HMBC spectra of compound 9	11
Figure S10. DEPTq spectrum of compounds 9	11
Figure S11. HPLC-PDA-ESIMS-ELSD analysis of THF extract from <i>A. artemisiifolia</i> flowers.	12
Figure S12. HPLC-PDA-ESIMS-ELSD analysis of MeOH extract from <i>A. artemisiifolia</i> flowers.	12
Figure S13. HPLC-PDA-ESIMS-ELSD analysis of THF extract from <i>A. artemisiifolia</i> leaves.	13
Figure S14. HPLC-PDA-ESIMS-ELSD analysis of MeOH extract from <i>A. artemisiifolia</i> leaves.	13
Figure S15. Spectrophotometric quantification of electrophiles with <i>N</i> -acetyl-L-cysteine and Ellman's reagent.	14
Figure S16. Calibration curve for parthenolide (15).	15
Figure S17. Calculation of the % of electrophiles in extract, quantified as parthenolide (15).	15
Figure S18. Adduct formation of <i>N</i> -acetyl-L-cysteine and selected phytochemicals in plant pollen. ...	16
Figure S19. HPLC-ESIMS detection of 13 and 14 in THF extracts prepared from five different lots of <i>A. psilostachya</i> pollen.	17
Table S1. Commercially available pollen included in the study.	18
Table S2. HRESIMS and ¹ H and ¹³ C NMR Spectroscopic Data for Compound 1	19
Table S3. HRESIMS and ¹ H and ¹³ C NMR Spectroscopic Data for Compound 9	20
Table S4. HRESIMS and ¹ H and ¹³ C NMR Spectroscopic Data for Compound 2	21
Table S5. HRESIMS and ¹ H and ¹³ C NMR Spectroscopic Data for Compound 3	22
Table S6. HRESIMS and ¹ H and ¹³ C NMR Spectroscopic Data for Compound 4	23

Table S7. HRESIMS and ^1H and ^{13}C NMR Spectroscopic Data for Compound 5	24
Table S8. HRESIMS and ^1H and ^{13}C NMR Spectroscopic Data for Compound 6	25
Table S9. HRESIMS and ^1H and ^{13}C NMR Spectroscopic Data for Compound 7	26
Table S10. HRESIMS and ^1H and ^{13}C NMR Spectroscopic Data for Compound 8	27
Table S11. HRESIMS and ^1H and ^{13}C NMR Spectroscopic Data for Compound 10	28
Table S12. HRESIMS and ^1H and ^{13}C NMR Spectroscopic Data for Compound 11	29
Table S13. HRESIMS and ^1H and ^{13}C NMR Spectroscopic Data for Compound 12	30
Table S14. HRESIMS and ^1H and ^{13}C NMR Spectroscopic Data for Compound 13	31
Table S15. HRESIMS and ^1H and ^{13}C NMR Spectroscopic Data for Compound 14	32
Table S16. <i>A. artemisiifolia</i> pollen lots.	33
Table S17. <i>A. psilostachya</i> pollen lots.....	33



Figure S1. Comparative HPLC-PDA-ESIMS-ELSD analysis of THF extracts from pollen of 30 plant species.

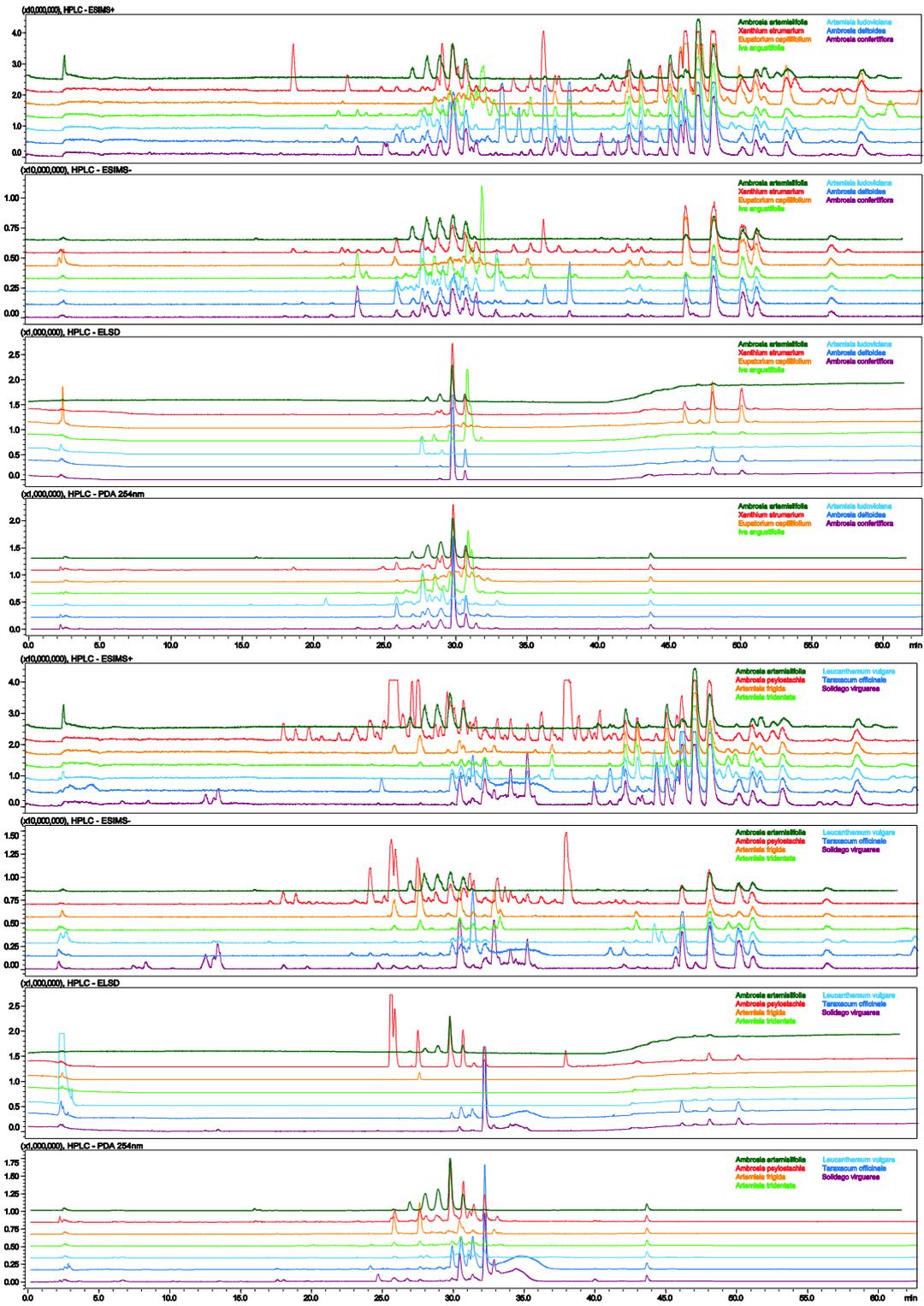


Figure S1. (continued).

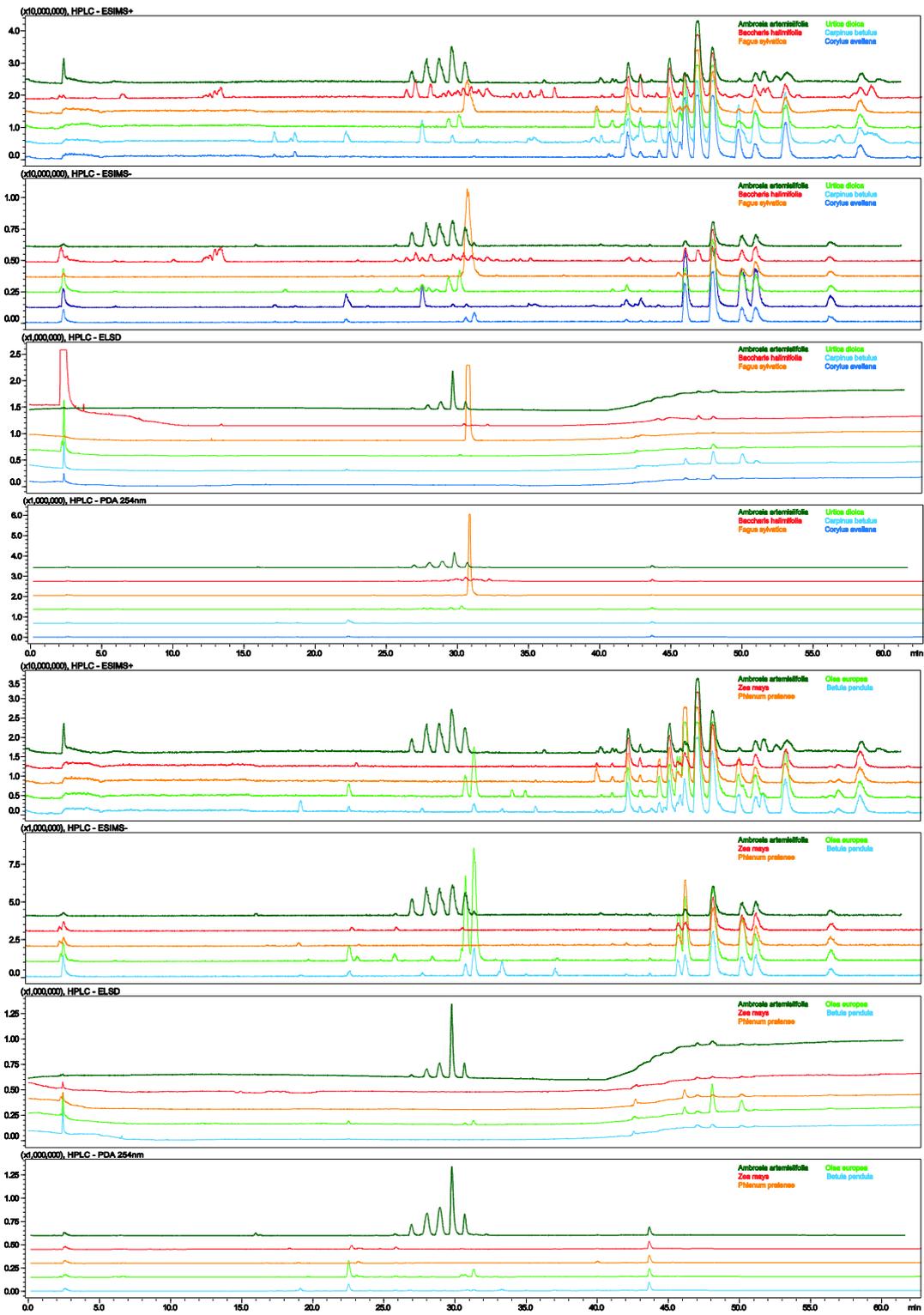


Figure S1. (continued).

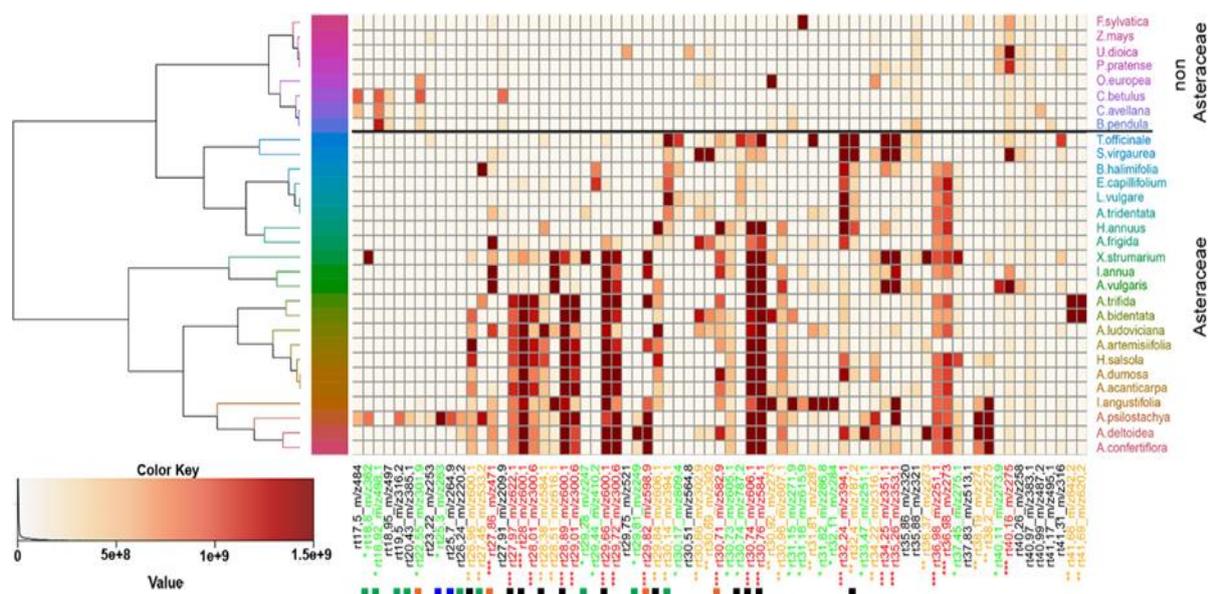


Figure S2. 2D heat map of HPLC-ESIMS+ metabolite profiles of THF extracts from *A. artemisiifolia* (lot Aa1), *A. psilostachya* (lot Ap1), and pollen from other Asteraceae and non-Asteraceae species. The x-axis represents the R_t - m/z of molecular ions. Botanical origins of pollen are listed to the right of the y-axis, and clustering is provided on the left side of the y-axis. The color-coded matrix elements represent the relative intensity of each ion. Asterisks with R_t - m/z represent the importance of each ion in the construction of clusters: no asterisk – weighting equals zero (present in a restricted number of samples), "*" – weighting factor within 0 and 0.1, "**" – weighting factor within 0.1 and 0.5, "***" – weighting factor within 0.5 and 1.0 (present in many samples). Colored squares below R_t - m/z indicate known or only partially identified metabolites. Black: conjugated polyamines; blue: coronopilin (13)/parthenin (14); green: unidentified sesquiterpene lactones; orange: partially characterized polyphenols.

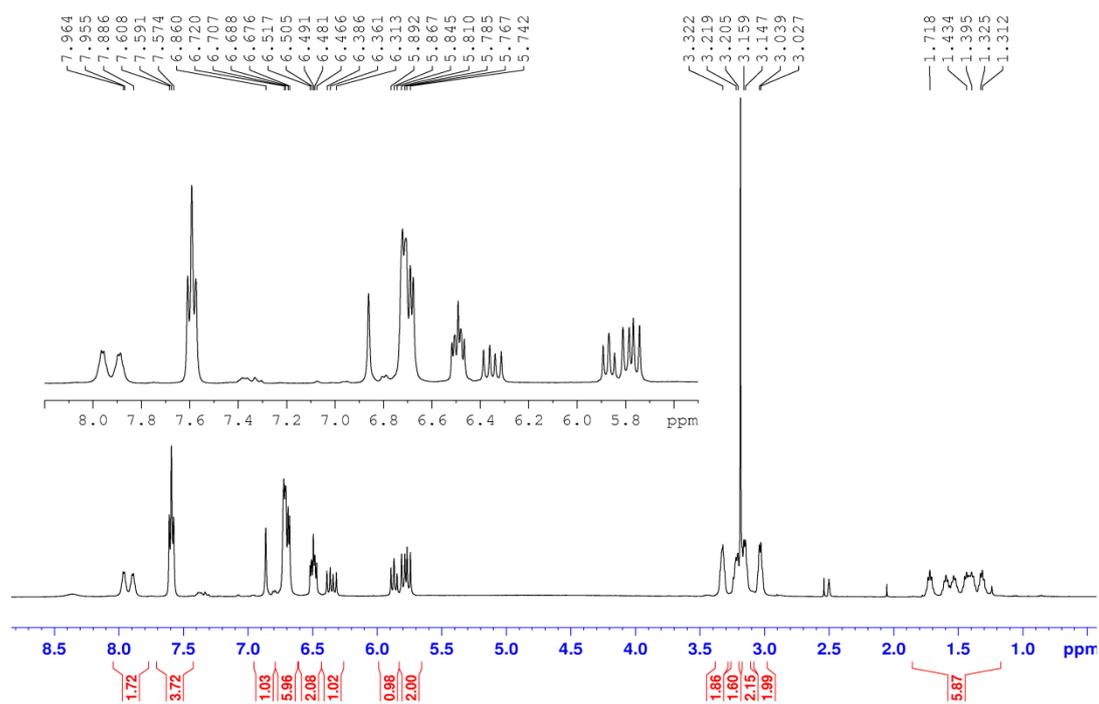


Figure S3. ^1H -NMR spectrum of compound **1** (500 MHz, $\text{DMSO-}d_6$).

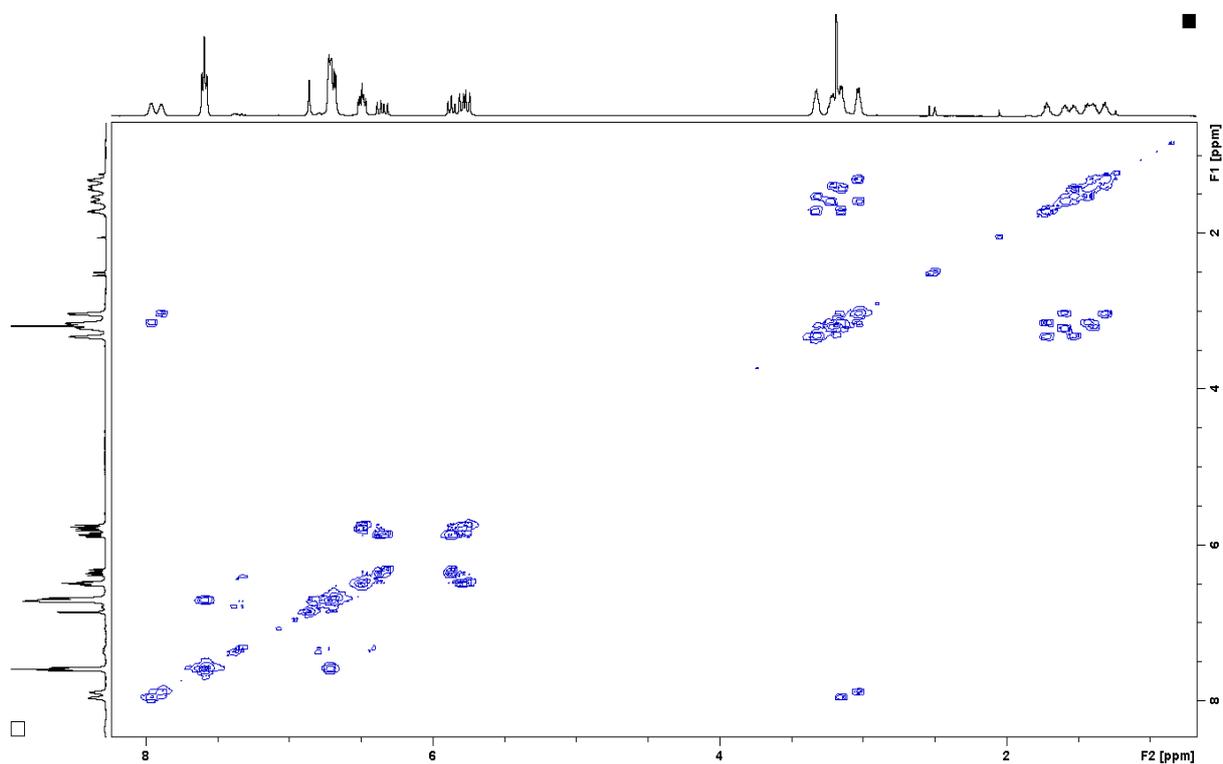


Figure S4. ^1H - ^1H COSY spectrum of compound **1** (500 MHz, $\text{DMSO-}d_6$).

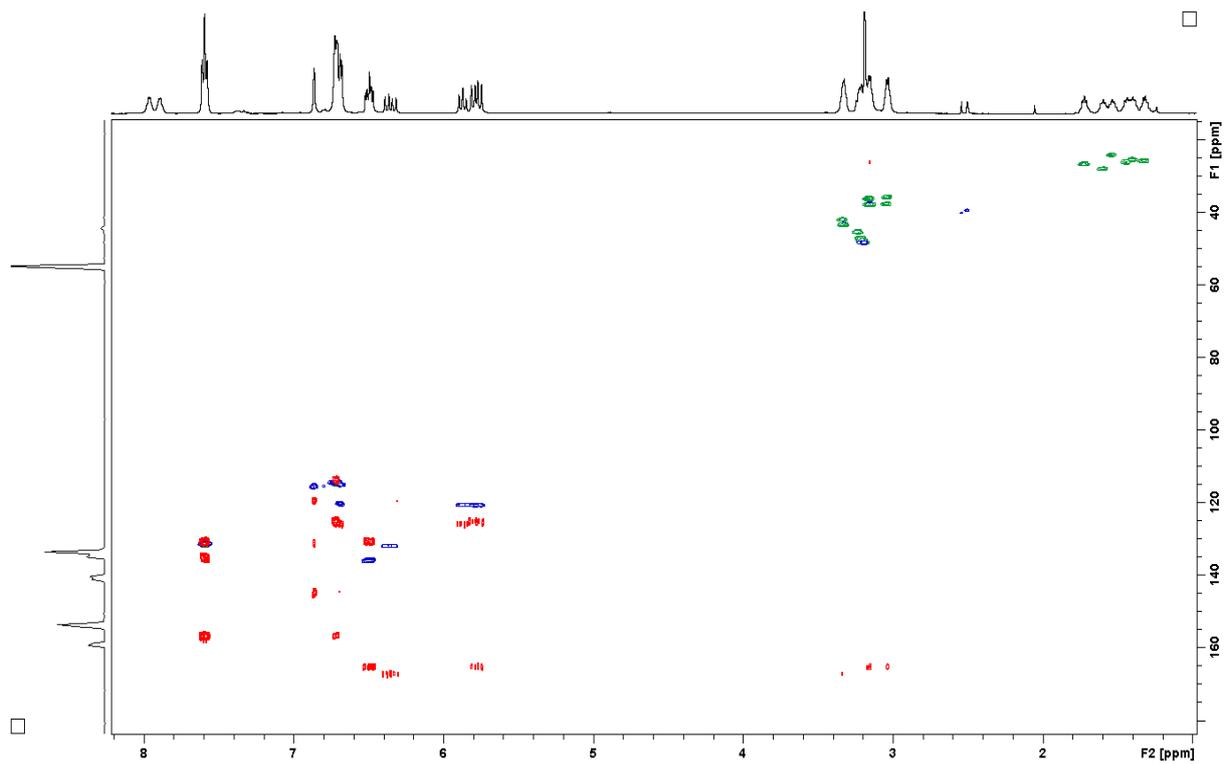


Figure S5. Overlay of HSQC (blue/green) and HMBC (red) spectra of compound **1** (500/125 MHz, DMSO- d_6).

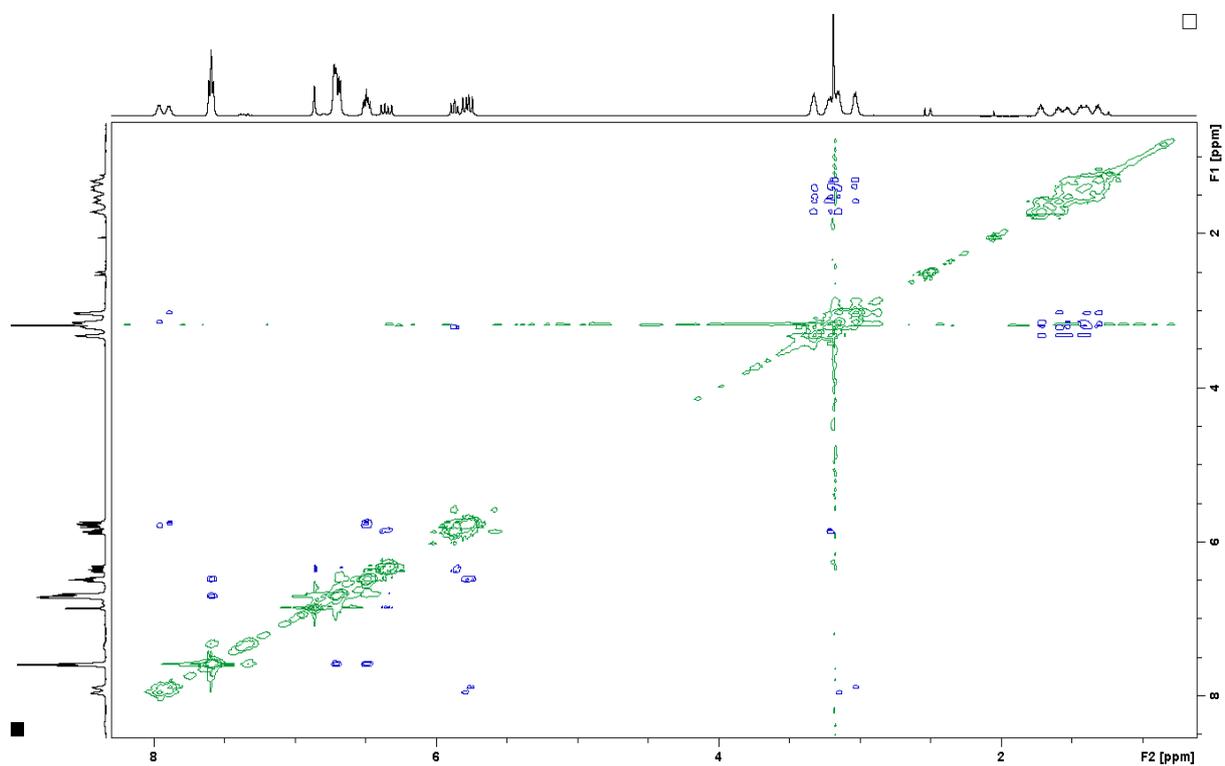


Figure S6. 2D ^1H - ^1H ROESY spectrum of compound **1** (500 MHz, DMSO- d_6).

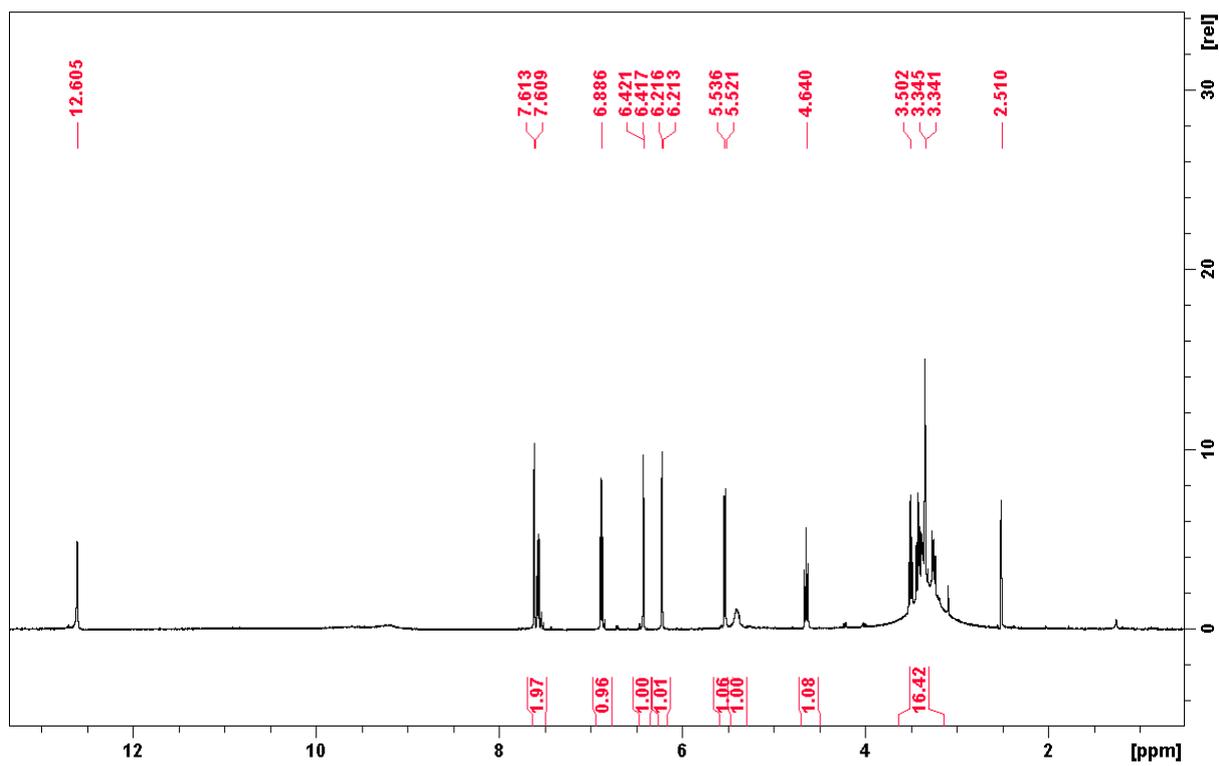


Figure S7. ^1H -NMR spectrum of compound **9** (500 MHz, $\text{DMSO-}d_6$).

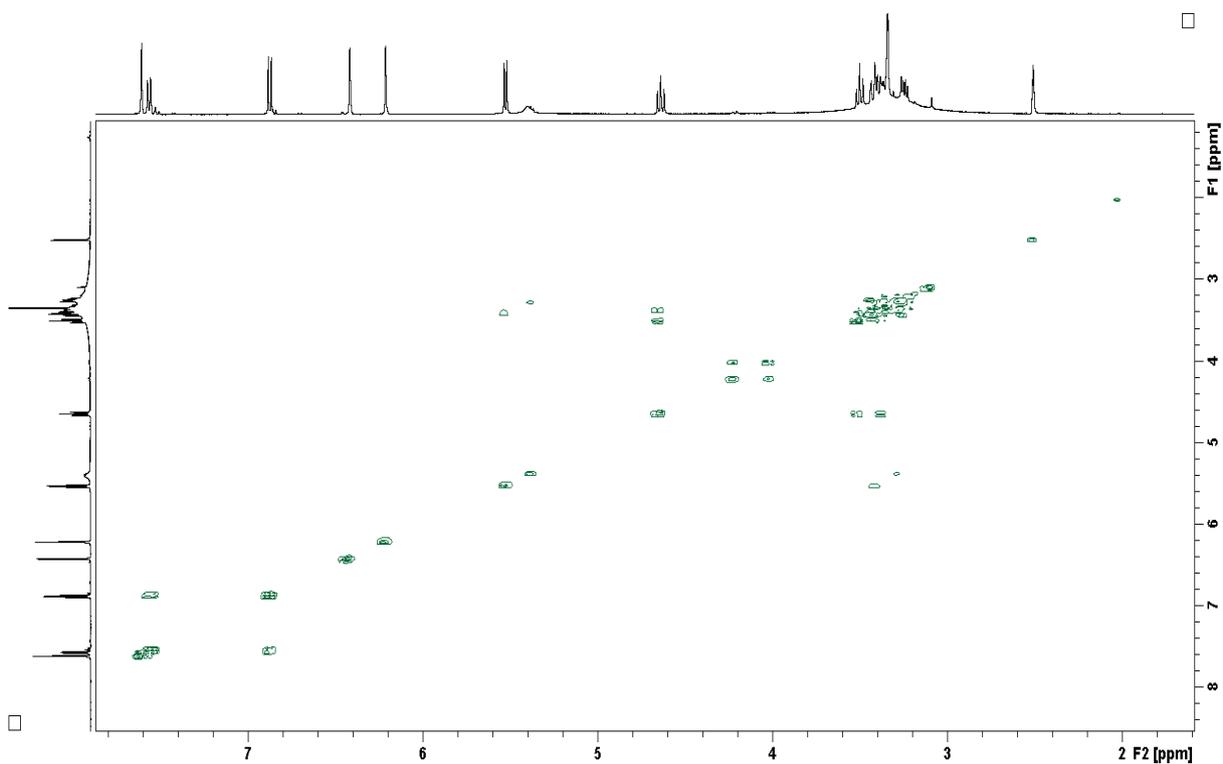


Figure S8. ^1H - ^1H COSY spectrum of compound **9** (500 MHz, $\text{DMSO-}d_6$).

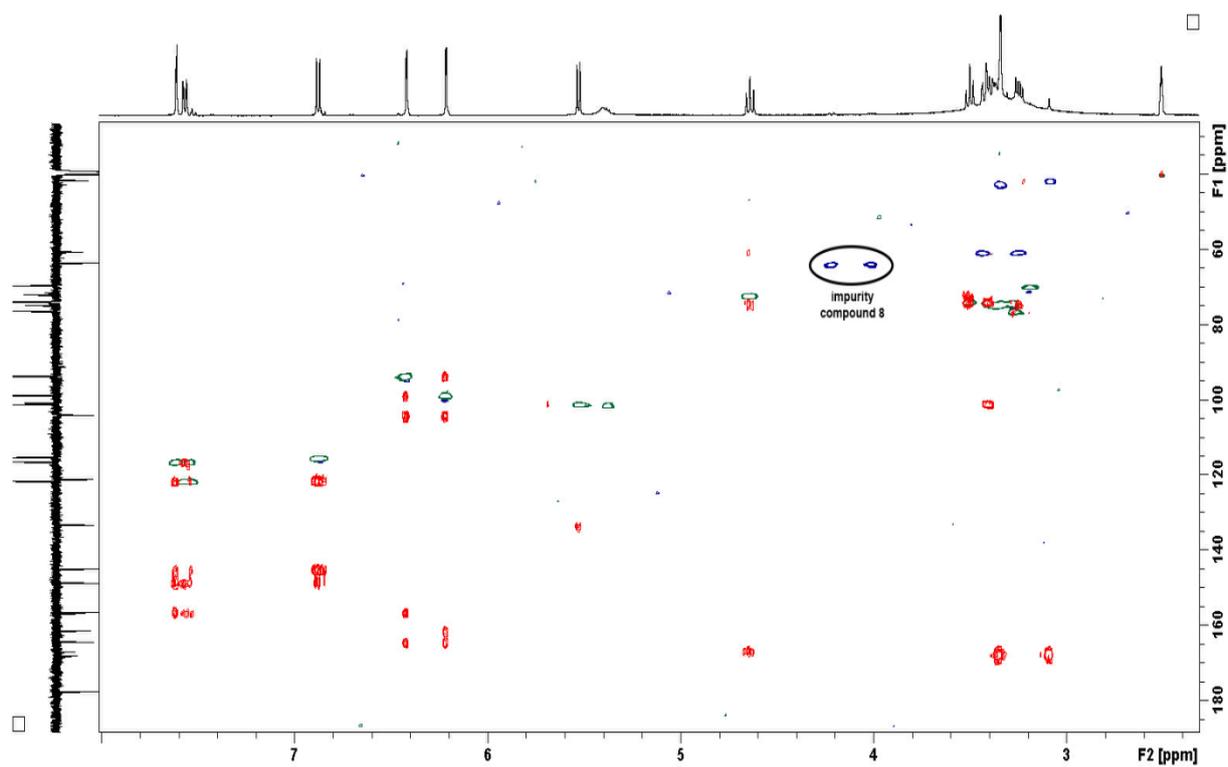


Figure S9. Overlay of HSQC (blue/green) and HMBC (red) spectra of compound **9** (500/125 MHz, DMSO-*d*₆).

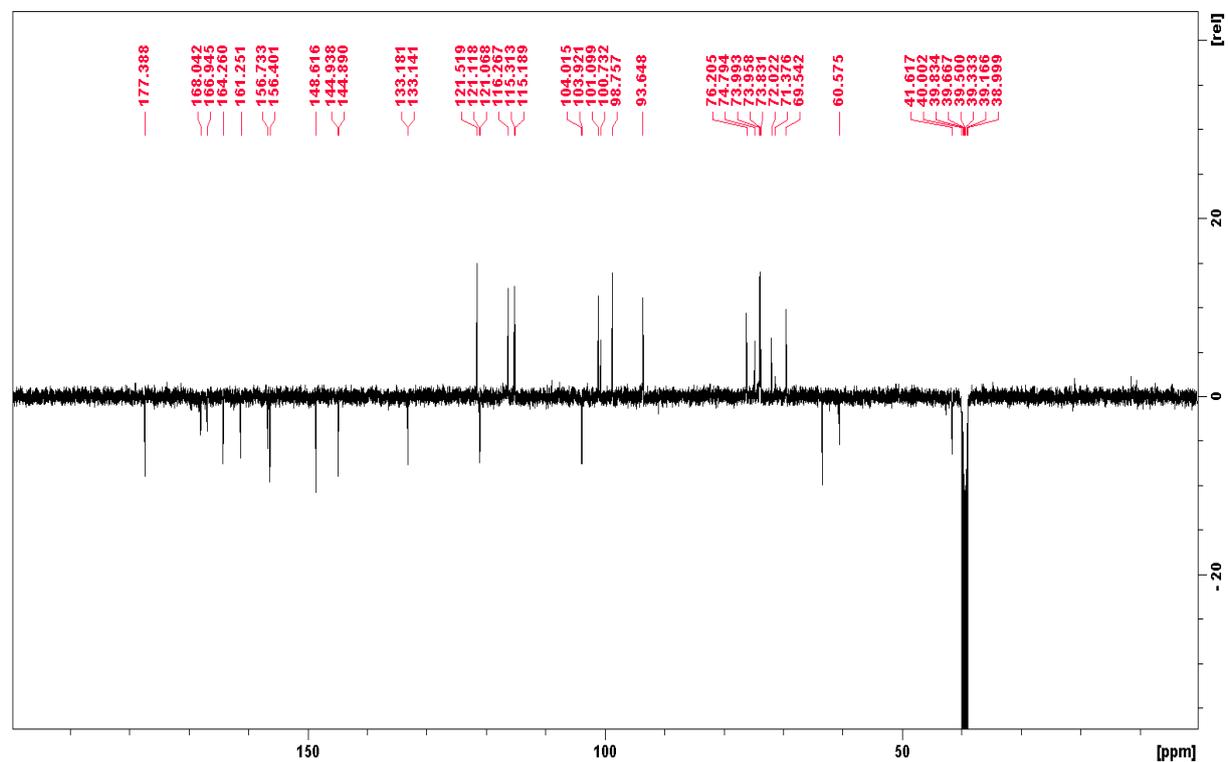


Figure S10. DEPTq spectrum of compounds **9** (500 MHz, DMSO-*d*₆).

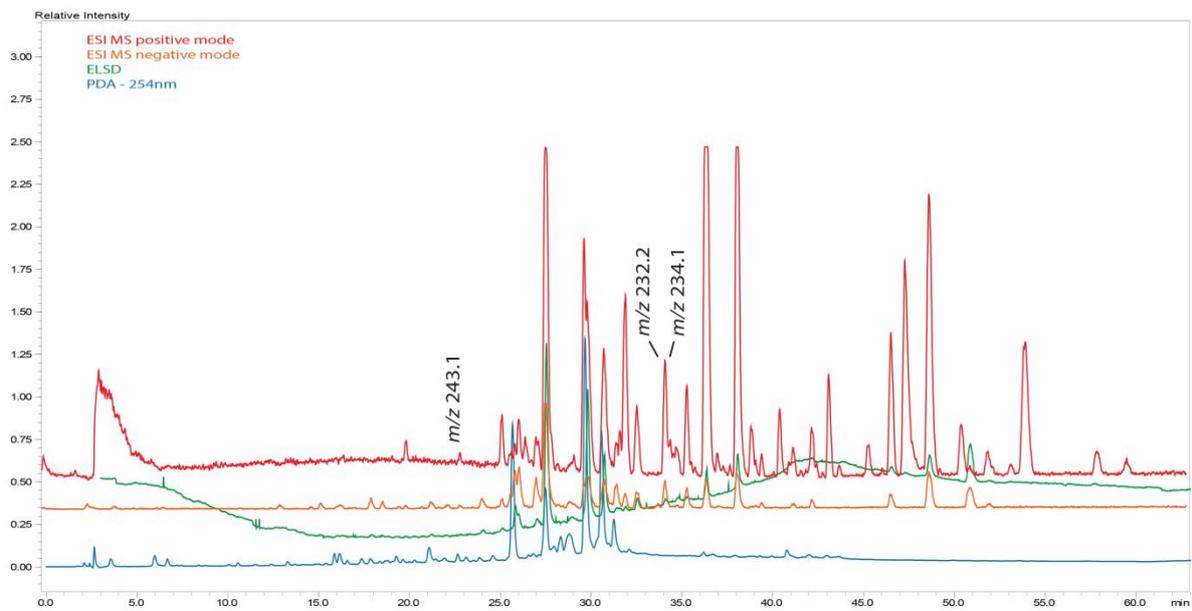


Figure S11. HPLC-PDA-ESIMS-ELSD analysis of THF extract from *A. artemisiifolia* flowers.

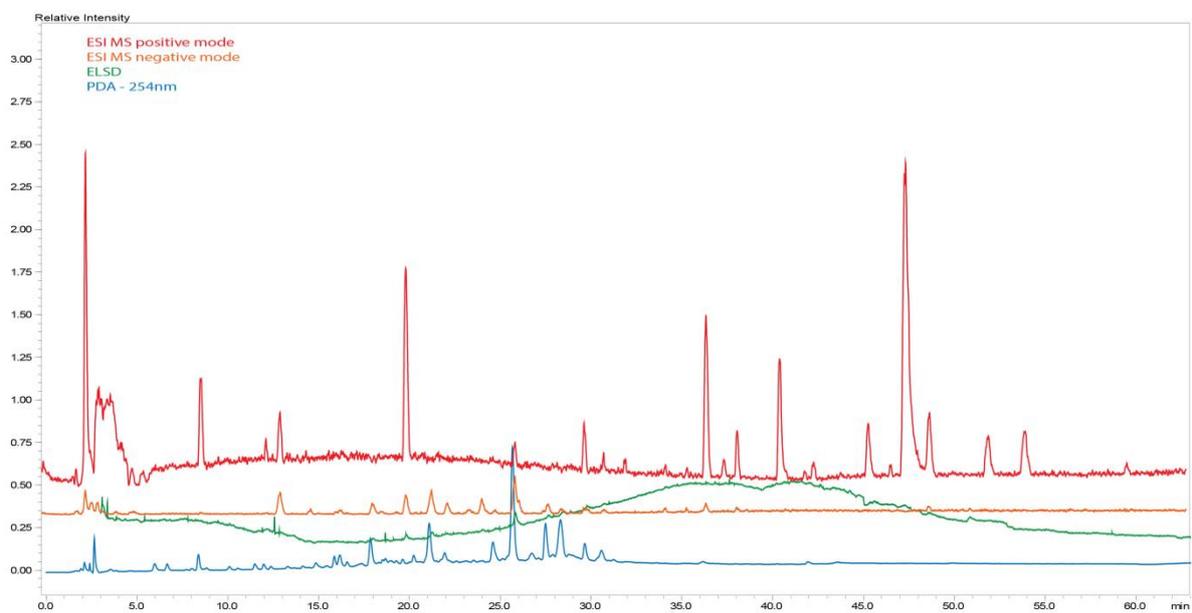


Figure S12. HPLC-PDA-ESIMS-ELSD analysis of MeOH extract from *A. artemisiifolia* flowers.

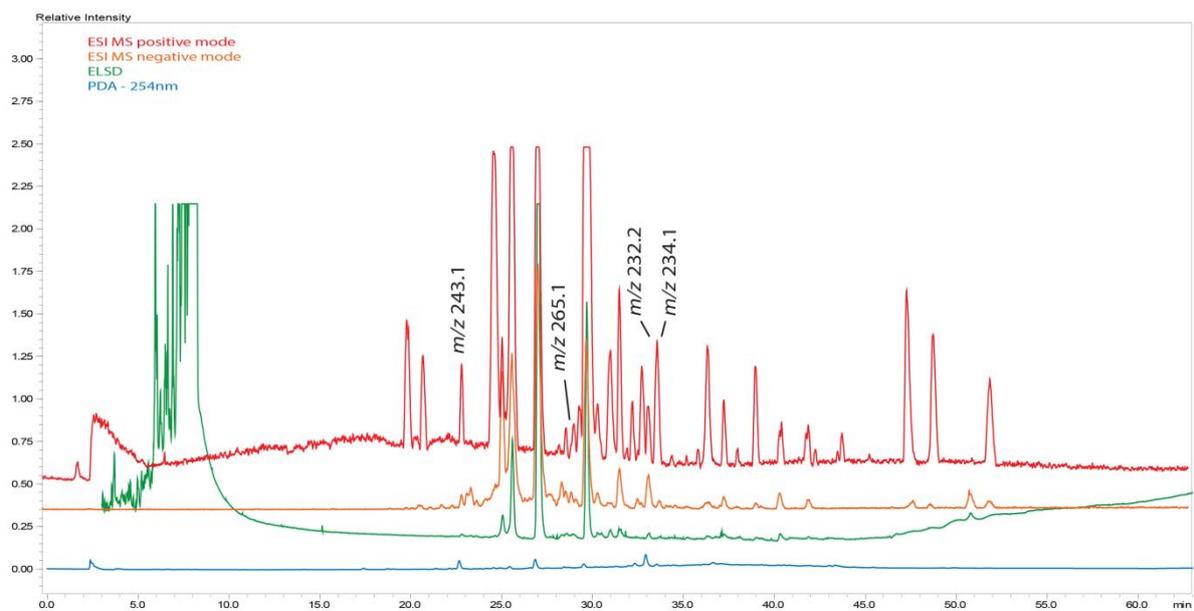


Figure S13. HPLC-PDA-ESIMS-ELSD analysis of THF extract from *A. artemisiifolia* leaves.

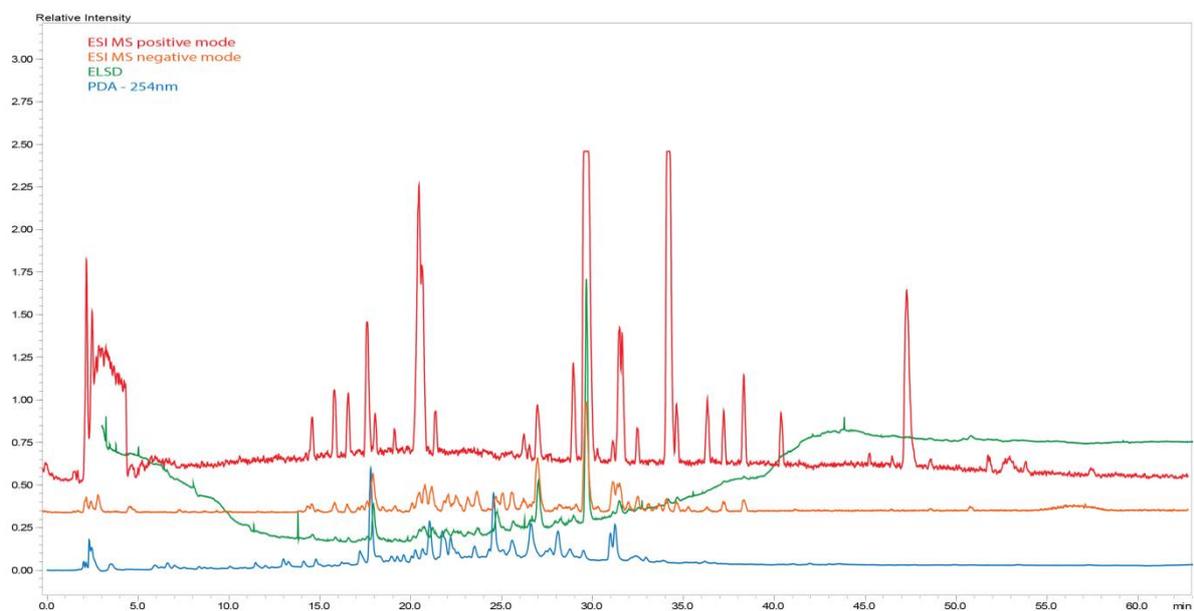


Figure S14. HPLC-PDA-ESIMS-ELSD analysis of MeOH extract from *A. artemisiifolia* leaves.

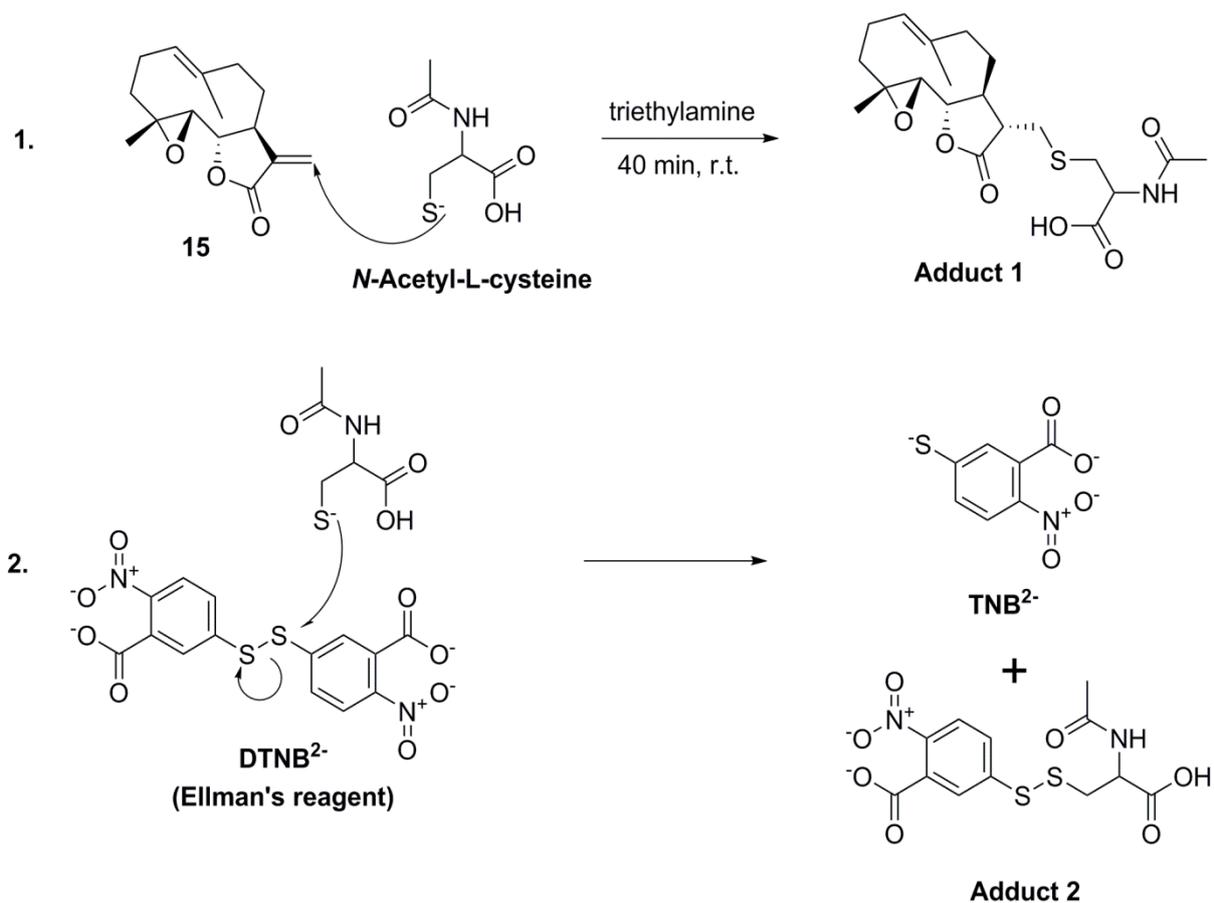


Figure S15. Spectrophotometric quantification of electrophiles with *N*-acetyl-L-cysteine and Ellman's reagent. Electrophiles (**15**) react with sulfhydryl groups (*N*-acetyl-L-cysteine) to form adducts (adduct 1) (1). Excess *N*-acetyl-L-cysteine reacts with Ellman's reagent to form the yellow TNB²⁻ ion, which can be quantified at 412 nm (2).

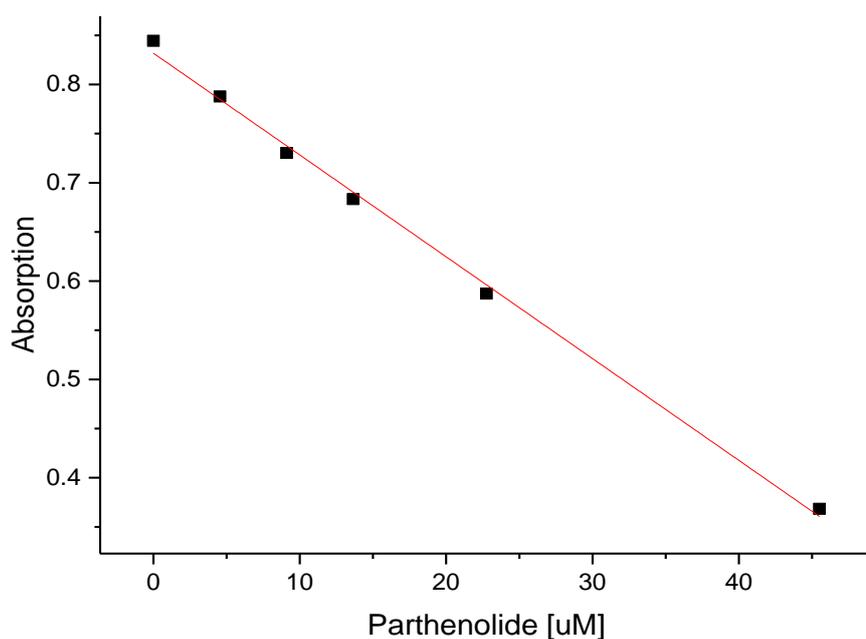


Figure S16. Calibration curve for parthenolide (15).

$$y = -0.0047x + 0.8318$$

$$y = A = A_2 - A_1$$

$$\% \text{ electrophiles}_{extr.} = \frac{x \cdot 248.3}{500}$$

Figure S17. Calculation of the % of electrophiles in extract, quantified as parthenolide (15). $y = -0.0047x + 0.8318$ is the equation obtained from the calibration curve (Figure S11). A is the absorbance ($A_2 - A_1$), where A_2 is the absorbance after the entire *N*-acetyl-L-cysteine reacted with Ellman's reagent producing the fluorescent TNB²⁻; excess of *N*-acetyl-L-cysteine is incubated with extracts and the absorbance related to the unreacted fraction of sulfhydryl groups is represented as A_1 ; 248.3 is the molecular weight of parthenolide. 500 correspond to 500 μ g of extract present in 10 mL of sample (measuring concentration); all measurements were performed at 412 nm.

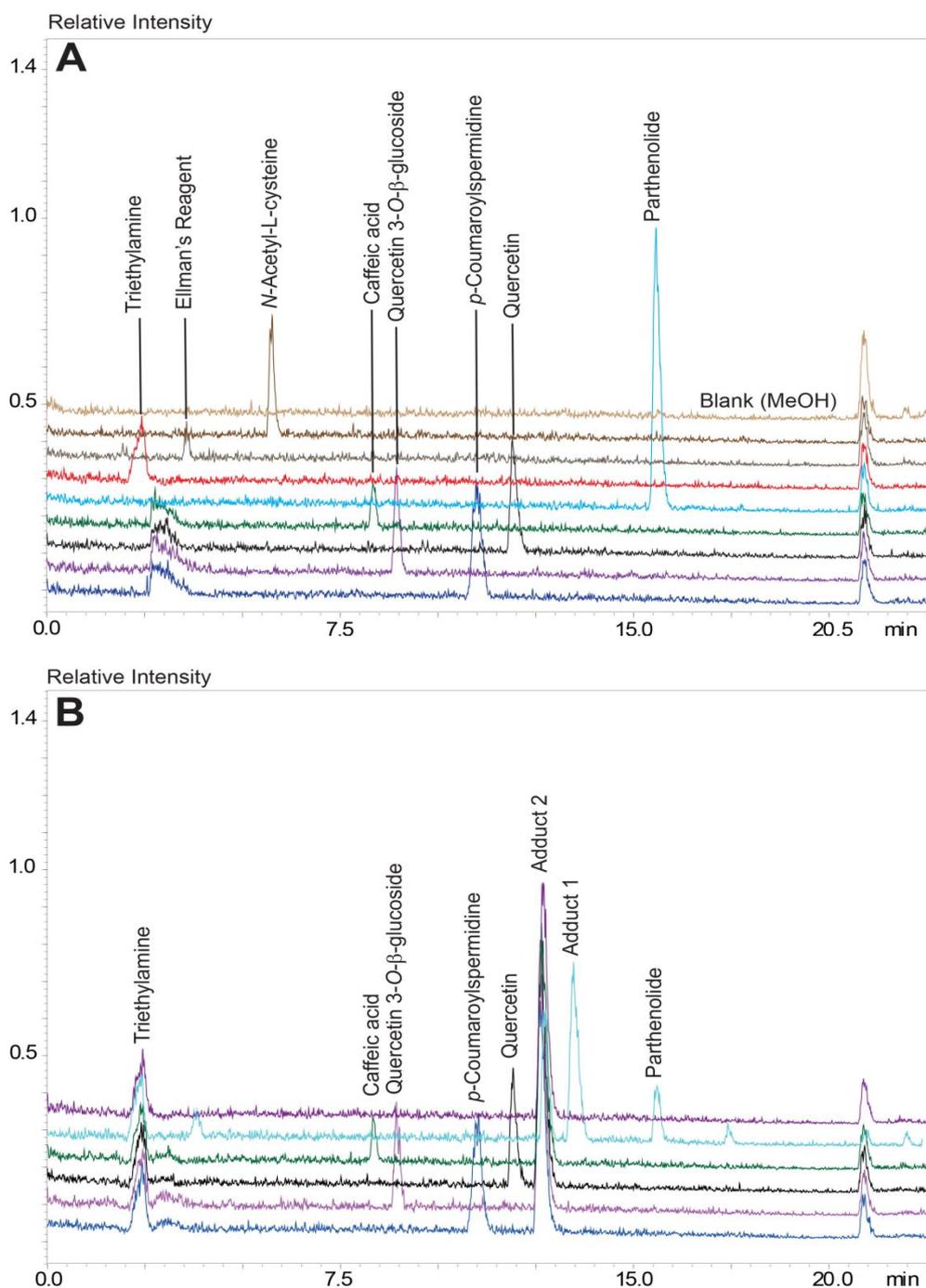


Figure S18. Adduct formation of *N*-acetyl-L-cysteine and selected phytochemicals in plant pollen, monitored by HPLC-ESIMS (positive ion mode). (A) Triethylamine, Ellman's reagent, *N*-acetyl-L-cysteine (reagents); caffeic acid, quercetin 3-*O*-β-glucoside, *p*-coumaroylspermidine, quercetin (negative controls); parthenolide (positive control) were first injected individually to determine their retention time and molecular ions. (B) HPLC-ESIMS (positive ion mode) chromatograms measured after incubating *N*-acetyl-L-cysteine with negative and positive controls, respectively. Only parthenolide reacted with *N*-acetyl-L-cysteine forming adduct 1. The unreacted fraction of *N*-acetyl-L-cysteine formed adduct 2 with Ellman's reagent. Trimethylamine was used as *N*-acetyl-L-cysteine activator.

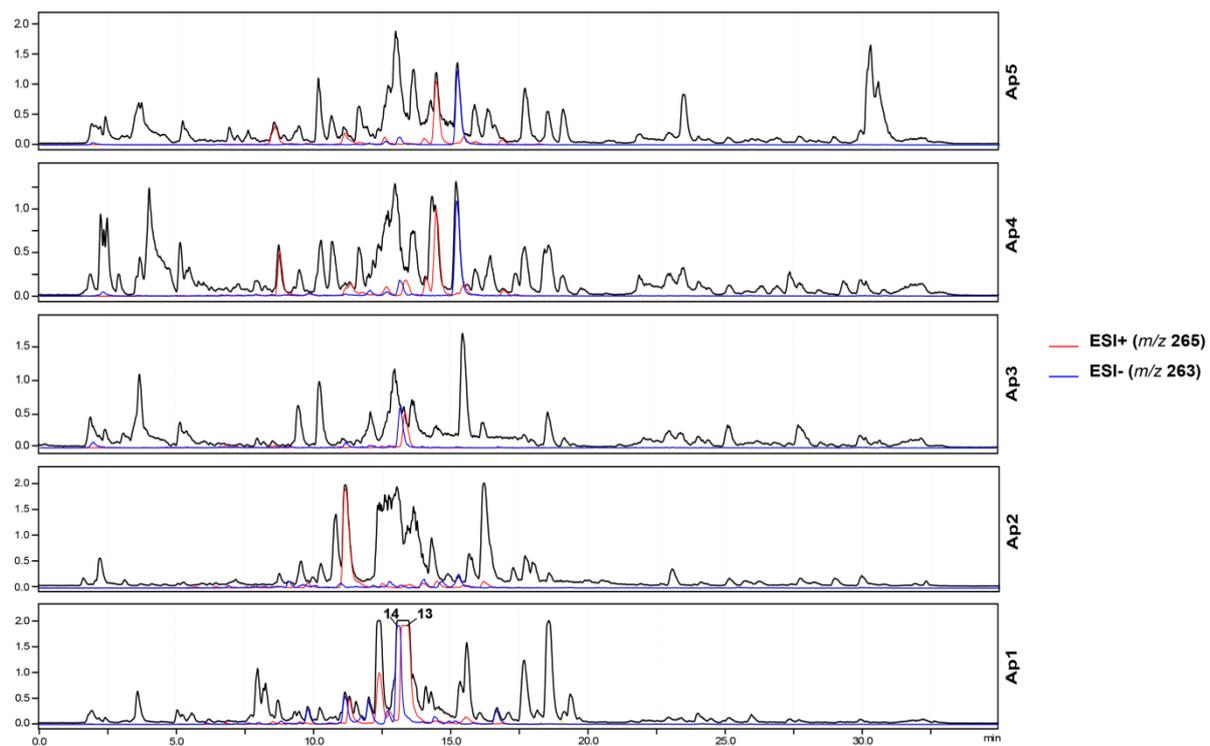


Figure S19. HPLC-ESIMS detection of **13** and **14** in THF extracts prepared from five different lots of *A. psilostachya* pollen.

Table S1. Commercially available pollen included in the study.

	plant species	lot #	supplier
1.	<i>Baccharis halimifolia</i> (Asteraceae)	P2195277	Greer
2.	<i>Hymenoclea salsola</i> (Asteraceae)	P3618767	Greer
3.	<i>Xanthium strumarium</i> (Asteraceae)	P3766835-2	Greer
4.	<i>Eupatorium capillifolium</i> (Asteraceae)	P3401503	Greer
5.	<i>Iva annua</i> (Asteraceae)	P3787377-1	Greer
6.	<i>Iva angustifolia</i> (Asteraceae)	94443	Greer
7.	<i>Artemisia ludoviciana</i> (Asteraceae)	P3397739	Greer
8.	<i>Ambrosia deltoidea</i> (Asteraceae)	P3557881-2	Greer
9.	<i>Ambrosia dumosa</i> (Asteraceae)	P3408578	Greer
10.	<i>Ambrosia trifida</i> (Asteraceae)	P3685399-6	Greer
11.	<i>Ambrosia confertiflora</i> (Asteraceae)	P3670198	Greer
12.	<i>Ambrosia bidentata</i> (Asteraceae)	P3743121	Greer
13.	<i>Ambrosia psilostachya</i> (Asteraceae)	P3406293-1	Greer
14.	<i>Ambrosia acanthicarpa</i> (Asteraceae)	P3760470	Greer
15.	<i>Artemisia frigida</i> (Asteraceae)	115587	Greer
16.	<i>Artemisia tridentata</i> (Asteraceae)	P3771025-1	Greer
17.	<i>Leucanthemum vulgare</i> (Asteraceae)	254739	Greer
18.	<i>Helianthus annuus</i> (Asteraceae)	P3716774	Greer
19.	<i>Ambrosia artemisiifolia</i> (Asteraceae)	020512201	Allergon AB
20.	<i>Artemisa vulgaris</i> (Asteraceae)	101112101	Allergon AB
21.	<i>Solidago virgaurea</i> (Asteraceae)	025912101	Allergon AB
22.	<i>Taraxacum officinale</i> (Asteraceae)	010711101	Allergon AB
23.	<i>Betula pendula</i> (Betulaceae)	012510101	Allergon AB
24.	<i>Carpinus betulus</i> (Betulaceae)	027309702	Allergon AB
25.	<i>Corylus avellana</i> (Corylaceae)	012713701	Allergon AB
26.	<i>Fagus sylvatica</i> (Fagaceae)	012813101	Allergon AB
27.	<i>Olea europea</i> (Oleaceae)	032511201	Allergon AB
28.	<i>Phleum pretense</i> (Poaceae)	011312102	Allergon AB
29.	<i>Zea mays</i> (Poaceae)	012013101	Allergon AB
30.	<i>Urtica dioica</i> (Urticaceae)	014812101	Allergon AB

Table S2. HRESIMS and ¹H and ¹³C NMR Spectroscopic Data for Compound 1 (DMSO-*d*₆, 500 MHz for ¹H, and 125 MHz for ¹³C NMR)

<i>N</i> ¹ , <i>N</i> ¹⁰ (<i>Z,Z</i>)-di- <i>p</i> -coumaroyl- <i>N</i> ⁵ (<i>Z</i>)-caffeoylspermidine		
accurate mass found	600.2400 [M+H] ⁺	
accurate mass calculated	600.2409 [M+H] ⁺	
molecular formula	C ₃₄ H ₃₇ N ₃ O ₇	
NMR data		
position	¹ H [ppm]	¹³ C ^a [ppm]
1 and 10	7.96 and 7.89 (m) ^b	-
2 and 9	3.15 and 3.03 (m) ^b	38.0, 37.7, 36.2, 35.7
3, 7 and 8	1.27-1.77 (m)	28.0, 26.8, 26.3, 26.0, 25.6, 24.2
4 and 6	3.19-3.36 (m)	47.3, 45.4, 43.5, 42.0
1'/1''	-	125.9
2'/2''	7.58 and 7.60 (each d, <i>J</i> =8.6 Hz)	131.4
6'/6''	-	-
3'/3''	6.71 (d, <i>J</i> =8.6 Hz)	114.4
5'/5''	-	-
4'/4''	-	157.5
7'/7''	6.47 (d, <i>J</i> =12.6 Hz)	135.9
8'/8''	5.79 (d, <i>J</i> =12.6 Hz)	120.7
9'/9''	-	165.8
1''	-	126.6
2''	6.86 (br s)	115.5
3''	-	145.5 ^d
4''	-	144.3 ^d
5''	6.68 ^c	115.0
6''	6.69 ^c	120.2
7''	6.37 and 6.33 (each d, <i>J</i> =12.6 Hz)	131.8
8''	5.88 and 5.85 (each d, <i>J</i> =12.6 Hz)	120.5
9''	-	167.9

^a ¹³C NMR chemical shifts were extracted from HSQC and HMBC spectra.

^b Assigned by 2D NOESY.

^c Signal partially overlapped.

^d Assigned on the basis of ¹³C NMR shifts predicted by ACD/Labs software.

Table S3. HRESIMS and ^1H and ^{13}C NMR Spectroscopic Data for Compound 9 (DMSO-*d*₆, 500 MHz for ^1H , and 125 MHz for ^{13}C NMR)

quercetin 3-<i>O</i>-(4''-<i>O</i>-malonyl)-β-glucopyranoside		
accurate mass found	551.1071 [M+H] ⁺	
accurate mass calculated	551.0959 [M+H] ⁺	
molecular formula	C ₂₄ H ₂₂ O ₁₅	
NMR data		
position	^1H [ppm]	$^{13}\text{C}^{\text{a}}$ [ppm]
2	-	156.7
3	-	133.2
4	-	177.4
5	-	161.2
6	6.21 (br s)	98.8
7	-	164.3
8	6.41 (br s)	93.7
9	-	156.4
10	-	104.0
1'	-	121.1
2'	7.61 (d, <i>J</i> =2.2 Hz)	116.3
3'	-	144.9
4'	-	148.6
5'	6.87 (d, <i>J</i> =8.5 Hz)	115.3
6'	7.56 (dd, <i>J</i> =8.5, 2.2 Hz)	121.5
1''	-	166.9
2''	3.34 (s)	40.6
3''	-	168.0
Glc-1	5.53 (d, <i>J</i> =7.6 Hz)	101.1
Glc-2	3.40 ^b	74.4
Glc-3	3.51 (dd, <i>J</i> =9.0, 9.0 Hz)	74.4
Glc-4	4.64 (dd, <i>J</i> =9.5, 9.0 Hz)	72.0
Glc-5	3.37 ^b	75.0
Glc-6	3.43 (dd, <i>J</i> =11.8, 2.0 Hz) 3.25 (dd, <i>J</i> =11.8, 6.0 Hz)	60.6

^a Signal partially overlapped.

^b Signal partially overlapped.

^c Not detected.

Table S4. HRESIMS and ¹H and ¹³C NMR Spectroscopic Data for Compound 2 (DMSO-*d*₆, 500 MHz for ¹H, and 125 MHz for ¹³C NMR)

<i>N</i> ¹ <i>N</i> ⁵ (<i>Z,Z</i>), <i>N</i> ¹⁰ (<i>E</i>)-tri- <i>p</i> -coumaroylspermidine		
accurate mass found	584.1311 [M+H] ⁺	
accurate mass calculated	584.1340 [M+H] ⁺	
molecular formula	C ₃₄ H ₃₇ N ₃ O ₆	
NMR data		
position	¹ H [ppm]	¹³ C ^a [ppm]
1	7.99 (m) ^b	-
2 and 9	3.16 and 3.06 (m) ^b	38.0, 37.8, 36.4, 36.0
3, 7 and 8	1.28-1.81 (m)	28.3, 27.0, 26.5, 26.0, 25.4, 24.1
4 and 6	3.40-3.20 (m)	47.2, 45.4, 43.5, 42.0
10	7.91 (m) ^b	-
1'	-	126.0
2'/6'	7.58 and 7.60 (each d, <i>J</i> =8.0 Hz)	131.1
3'/5'	6.74 ^c	114.5
4'	-	157.0
7'	6.50 (d, <i>J</i> =12.0 Hz)	135.8
8'	5.82/5.79 (d, <i>J</i> =12.0 Hz)	120.5
9'	-	166.0
1''	-	126.0
2''/6''	7.25 and 7.23 (each d, <i>J</i> =8.0 Hz)	129.7
3''/5''	6.74 ^c	115.1
4''	-	157.3
7''	6.45 ^c	131.6
8''	5.95/5.91 (each d, <i>J</i> =12.0 Hz)	120.6
9''	-	167.0
1'''	-	125.6
2'''/6'''	7.40 ^c	128.6
3'''/5'''	6.81 ^c	115.3
4'''	-	158.6
7'''	7.37 ^c	138.4
8'''	6.43 ^c	118.4
9'''	-	165.4

^a ¹³C NMR chemical shifts extracted from HSQC and HMBC spectra.

^b Assigned by NOESY spectra.

^c Signal partially overlapped.

Table S5. HRESIMS and ¹H and ¹³C NMR Spectroscopic Data for Compound 3 (DMSO-*d*₆, 500 MHz for ¹H, and 125 MHz for ¹³C NMR)

<i>N</i> ¹ , <i>N</i> ⁵ , <i>N</i> ¹⁰ , <i>N</i> ¹⁴ (<i>E,E,E,E</i>)-tetra- <i>p</i> -coumaroylspermine		
accurate mass found	787.3611 [M+H] ⁺	
accurate mass calculated	787.3629 [M+H] ⁺	
molecular formula	C ₄₆ H ₅₀ N ₄ O ₈	
NMR data		
position	¹ H [ppm]	¹³ C ^a [ppm]
1 and 14	8.05 and 7.93 (br m)	-
2, 4, 6, 9, 11, and 13	3.55-3.15 ^b	46.6, 45.0, 44.7, 43.5, 36.3, 36.0
3, 7, 8 and 12	1.86-1.45 ^b	29.5, 27.7, 26.5, 24.8
1'	-	125.5
2'/6'	7.50-7.45 ^b	129.3
3'/5'	6.74 ^b	115.2
4'/4'''	-	^c
7'	7.45 ^{b,d}	141.2
8'	6.90 ^{b,d}	114.4
9'	-	^c
1''	-	125.5
2''/6''	7.39 (d, <i>J</i> =8.0 Hz)	128.7
3'/5'	6.80 (d, <i>J</i> =8.0 Hz)	115.4
4'/4'''	-	158.1
7''	7.36 ^{b,d}	138.3
8''	6.44 ^{b,d}	118.4
9''	-	^c

^a ¹³C NMR chemical shifts extracted from HSQC and HMBC spectra.

^b Signal partially overlapped.

^c Not detected.

^d Assigned on the basis of ¹³C NMR shifts predicted by ACD/Labs software.

Table S6. HRESIMS and ¹H and ¹³C NMR Spectroscopic Data for Compound 4 (DMSO-*d*₆, 500 MHz for ¹H, and 125 MHz for ¹³C NMR)

quercetin 3- <i>O</i> -β-glucoside		
accurate mass found	465.0943 [M+H] ⁺	
accurate mass calculated	465.0955 [M+H] ⁺	
molecular formula	C ₂₁ H ₂₀ O ₁₂	
NMR data		
Position	¹ H [ppm]	¹³ C ^a [ppm]
2	-	156.1
3	-	133.3
4	-	177.4
5	-	161.3
6	6.20 (br s)	98.7
7	-	164.3
8	6.41 (br s)	93.5
9	-	156.4
10	-	103.9
1'	-	121.2
2'	7.58 (br s)	116.2
3'	-	144.8
4'	-	148.5
5'	6.86 (d, <i>J</i> =8.6 Hz)	115.2
6'	7.58 (dd, <i>J</i> =8.6, 2.6 Hz)	121.6
Glc-1	5.47 (d, <i>J</i> =7.0 Hz)	100.8
Glc-2	3.23 ^b	74.1
Glc-3	3.22 ^b	76.5
Glc-4	3.09 ^b	69.9
Glc-5	3.09 ^b	77.6
Glc-6	3.33 ^b 3.60 (br d, <i>J</i> =11.6)	61.0

^a ¹³C NMR chemical shifts extracted from HSQC and HMBC spectra.

^b Signal partially overlapped.

Table S7. HRESIMS and ^1H and ^{13}C NMR Spectroscopic Data for Compound 5 (DMSO-*d*₆, 500 MHz for ^1H , and 125 MHz for ^{13}C NMR)

isorhamnetin 3-O-(6''-O-acetyl)-β-glucopyranoside		
accurate mass found	521.1201 [M+H] ⁺	
accurate mass calculated	521.1217 [M+H] ⁺	
molecular formula	C ₂₄ H ₂₄ O ₁₃	
NMR data		
position	^1H [ppm]	$^{13}\text{C}^{\text{a}}$ [ppm]
2	-	156.8
3	-	^c
4	-	^c
5	-	161.2
6	6.22 (d, <i>J</i> =2.0 Hz)	98.5
7	-	164.8
8	6.46 (d, <i>J</i> =2.0 Hz)	93.4
9	-	156.8
10	-	103.8
1'	-	120.9
2'	7.83 (d, <i>J</i> =2.0 Hz)	113.0
3'	-	147.1
4'	-	149.2
5'	6.92 (d, <i>J</i> =8.2 Hz)	114.9
6'	7.45 (dd, <i>J</i> =8.2, 2.0 Hz)	122.0
7'	3.86 (s)	55.5
1''	-	170.0
2''	1.74 (s)	19.8
Glc-1	5.42 (d, <i>J</i> =8.0 Hz)	100.9
Glc-2	3.27 ^b	73.8
Glc-3	3.27 ^b	76.0
Glc-4	3.15 (m)	69.7
Glc-5	3.34 (m)	73.9
Glc-6	4.02 (dd, <i>J</i> =12.0, 6.0 Hz) 4.09 (dd, <i>J</i> =12.0, 2.0 Hz)	62.5

^a ^{13}C NMR chemical shifts extracted from HSQC and HMBC spectra.

^b Signal partially overlapped.

^c Not detected.

Table S8. HRESIMS and ^1H and ^{13}C NMR Spectroscopic Data for Compound 6 (MeOH- d_4 , 500 MHz for ^1H , and 125 MHz for ^{13}C NMR)

kaempferol -3-O- β -glucopyranoside		
accurate mass found	449.1203 [M+H] ⁺	
accurate mass calculated	449.1176 [M+H] ⁺	
molecular formula	C ₂₁ H ₂₀ O ₁₁	
NMR data		
position	^1H [ppm]	$^{13}\text{C}^a$ [ppm]
2	-	157.8
3	-	133.6
4	-	^c
5	-	161.6
6	6.33 (d, $J=2.0$ Hz)	94.3
7	-	164.5
8	6.16 (d, $J=2.0$ Hz)	99.3
9	-	156.9
10	-	104.1
1'	-	122.3
2'/6'	8.01 (pseudo d, $J=8.7, 2.0$ Hz)	131.7
3'/5'	6.87 (pseudo d, $J=8.7, 2.0$ Hz)	115.5
4'	-	160.0
Glc-1	5.17 (d, $J=7.0$ Hz)	103.8
Glc-2	3.46 ^b	75.2
Glc-3	3.44 ^b	77.5
Glc-4	3.33 (dd, $J=9.5, 9.0$ Hz)	70.9
Glc-5	3.23 (ddd, $J=9.0, 6.0, 2.0$ Hz)	77.8
Glc-6	3.69 (ddd, $J=12.0, 2.0$ Hz) 3.54 (ddd, $J=12.0, 6.0$ Hz)	62.4

^a ^{13}C NMR chemical shifts extracted from HSQC and HMBC spectra.

^b Signal partially overlapped.

^c Not detected.

Table S9. HRESIMS and ^1H and ^{13}C NMR Spectroscopic Data for Compound 7 (DMSO-*d*₆, 500 MHz for ^1H , and 125 MHz for ^{13}C NMR)

isorhamnetin 3- <i>O</i> - β -glucopyranoside		
accurate mass found	478.0987 [M+H] ⁺	
accurate mass calculated	478.1011 [M+H] ⁺	
molecular formula	C ₂₂ H ₂₂ O ₁₂	
NMR data		
position	^1H [ppm]	$^{13}\text{C}^{\text{a}}$ [ppm]
2	-	155.9
3	-	133.0
4	-	^c
5	-	160.9
6	6.22 (d, <i>J</i> =2.0 Hz)	98.5
7	-	163.9
8	6.45 (d, <i>J</i> =2.0 Hz)	93.5
9	-	156.2
10	-	103.5
1'	-	120.6
2'	7.93 (d, <i>J</i> =2.0 Hz)	113.4
3'	-	146.6
4'	-	148.3
5'	6.93 (d, <i>J</i> =8.4 Hz)	114.9
6'	7.52 (dd, <i>J</i> =8.4, 2.0 Hz)	121.8
7'	3.86 (s)	55.5
Glc-1	5.55 (d, <i>J</i> =7.0 Hz)	100.6
Glc-2	3.26 ^b	74.0
Glc-3	3.28 ^b	76.2
Glc-4	3.14 ^b	69.6
Glc-5	3.14 ^b	77.1
Glc-6	3.41 (dd, <i>J</i> =12.0, 4.0 Hz) 3.61 (dd, <i>J</i> =12.0, 2.0 Hz)	60.4

^a ^{13}C NMR chemical shifts extracted from HSQC and HMBC spectra.

^b Signal partially overlapped.

^c Not detected.

Table S10. HRESIMS and ^1H and ^{13}C NMR Spectroscopic Data for Compound 8 (DMSO- d_6 , 500 MHz for ^1H , and 125 MHz for ^{13}C NMR)

quercetin 3-O-(6''-O-malonyl)-β-glucopyranoside		
accurate mass found	551.0857 [M+H] ⁺	
accurate mass calculated	551.0959 [M+H] ⁺	
molecular formula	C ₂₄ H ₂₂ O ₁₅	
NMR data		
position	^1H [ppm]	^{13}C[ppm]
2	-	156.7
3	-	133.2
4	-	177.4
5	-	161.3
6	6.21 (br s)	98.8
7	-	164.3
8	6.41 (br s)	93.7
9	-	156.5
10	-	104.0
1'	-	121.2
2'	7.54 (d, $J=2.2$ Hz)	116.3
3'	-	144.9
4'	-	148.6
5'	6.85 (d, $J=8.8$ Hz)	115.3
6'	7.53 (dd, $J=8.8, 2.2$ Hz)	121.7
1''	-	166.7
2''	3.11 (s)	40.2
3''	-	167.9
Glc-1	5.38 (d, $J=7.3$ Hz)	101.1
Glc-2	3.30 ^a	74.0
Glc-3	3.30 ^a	76.3
Glc-4	3.18 (dd, $J=9.0, 9.0$ Hz)	69.7
Glc-5	3.35 ^a	74.0
Glc-6	4.22 (dd, $J=11.8, 2.0$ Hz) 4.01 (dd, $J=11.8, 6.0$ Hz)	63.7

^a Signal partially overlapped.

Table S11. HRESIMS and ¹H and ¹³C NMR Spectroscopic Data for Compound 10 (DMSO-*d*₆, 500 MHz for ¹H, and 125 MHz for ¹³C NMR)

kaempferol 3-O-(6''-O-malonyl)-β-glucopyranoside		
accurate mass found	535.1251 [M+H] ⁺	
accurate mass calculated	535.1110 [M+H] ⁺	
molecular formula	C ₂₄ H ₂₂ O ₁₄	
NMR data		
position	¹H [ppm]	¹³C^a[ppm]
2	-	156.2
3	-	^c
4	-	^c
5	-	160.7
6	6.21 (br s)	98.4
7	-	163.5
8	6.42 (br s)	93.3
9	-	^c
10	-	104.3
1'	-	^c
2'/6'	7.98 (pseudo d, <i>J</i> =8.7,2.0 Hz)	130.5
3'/5'	6.89 (pseudo d, <i>J</i> =8.7,2.0 Hz)	114.8
4'	-	159.0
1''	-	^c
2''	3.02 (s)	43.0
3''	-	167.5
Glc-1	5.33 (d, <i>J</i> =7.1 Hz)	101.0
Glc-2	3.23 ^b	73.8
Glc-3	3.25 ^b	75.8
Glc-4	3.16 (m)	69.2
Glc-5	3.34 (m)	73.6
Glc-6	4.15 (dd, <i>J</i> =11.8, 2.0 Hz) 4.00 (dd, <i>J</i> =11.8, 6.0 Hz)	62.6

^a ¹³C NMR chemical shifts extracted from HSQC and HMBC spectra.

^b Signal partially overlapped.

^c Not detected.

Table S12. HRESIMS and ¹H and ¹³C NMR Spectroscopic Data for Compound 11 (DMSO-*d*₆, 500 MHz for ¹H, and 125 MHz for ¹³C NMR)

isorhamnetin 3-<i>O</i>-(6''-<i>O</i>-malonyl)-β-glucopyranoside		
accurate mass found	565.1019 [M+H] ⁺	
accurate mass calculated	565.1052 [M+H] ⁺	
molecular formula	C ₂₅ H ₂₄ O ₁₅	
NMR data		
Position	¹H [ppm]	¹³C^a[ppm]
2	-	156.7
3	-	133.3
4	-	^c
5	-	161.6
6	6.21 (d, <i>J</i> =2.0 Hz)	98.4
7	-	164.8
8	6.45 (d, <i>J</i> =2.0 Hz)	93.5
9	-	156.7
10	-	104.4
1'	-	122.3
2'	7.82 (d, <i>J</i> =2.0 Hz)	113.2
3'	-	147.3
4'	-	149.3
5'	6.92 (d, <i>J</i> =8.4 Hz)	114.9
6'	7.54 (dd, <i>J</i> =8.4, 2.0 Hz)	121.8
7'	3.86 (s)	55.7
1''	-	166.8
2''	3.09 (s)	41.0
3''	-	168.0
Glc-1	5.44 (d, <i>J</i> =7.0 Hz)	101.1
Glc-2	3.28 (dd, <i>J</i> =8.5, 7.0 Hz)	74.0
Glc-3	3.31 (dd, <i>J</i> =9.0, 8.5 Hz)	76.0
Glc-4	3.17 (dd, <i>J</i> =9.0, 9.0 Hz)	69.5
Glc-5	3.39 (ddd, <i>J</i> =9.0, 6.0, 2.0 Hz)	73.7
Glc-6	4.18 (dd, <i>J</i> =12.0, 2.0 Hz) 4.09 (dd, <i>J</i> =12.0, 6.0 Hz)	63.4

^a ¹³C NMR chemical shifts extracted from HSQC and HMBC spectra.

^b Signal partially overlapped.

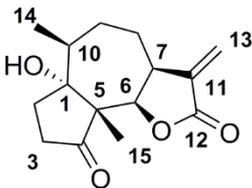
^c Not detected.

Table S13. HRESIMS and ¹H and ¹³C NMR Spectroscopic Data for Compound 12 (DMSO-*d*₆, 500 MHz for ¹H, and 125 MHz for ¹³C NMR)

kaempferol 3-O-(6''-(<i>E</i>)-<i>p</i>-coumaroyl)-β-glucopyranoside		
accurate mass found	595.1360 [M+H] ⁺	
accurate mass calculated	595.1373 [M+H] ⁺	
molecular formula	C ₃₄ H ₃₇ N ₃ O ₆	
NMR data		
position	¹H [ppm]	¹³C^a[ppm]
2	-	156.6
3	-	132.8
4	-	c
5	-	161.3
6	6.14 (d, <i>J</i> =1.8 Hz)	98.7
7	-	164.2
8	6.37 (d, <i>J</i> =1.8 Hz)	93.2
9	-	156.6
10	-	103.6
1'	-	120.6
2'/6'	7.97 (pseudo dd, <i>J</i> =8.7, 2.0 Hz)	130.3
3'/5'	6.86 (pseudo dd, <i>J</i> =8.7, 2.0 Hz)	114.7
4'	-	159.9
1''	-	124.9
2''/6''	7.34 (pseudo dd, <i>J</i> =8.5, 2.0 Hz)	129.8
3''/5''	6.79 (pseudo dd, <i>J</i> =8.5, 2.0 Hz)	115.3
4''	-	159.8
7''	7.33 (d, <i>J</i> =16.0 Hz)	144.1
8''	6.09 (d, <i>J</i> =16.0 Hz)	113.4
9''	-	166.1
Glc-1	5.43 (d, <i>J</i> =7.5 Hz)	100.7
Glc-2	3.25 (dd, <i>J</i> =8.5, 7.5 Hz)	73.9
Glc-3	3.29 (dd, <i>J</i> =9.0, 8.5 Hz)	76.1
Glc-4	3.19 (dd, <i>J</i> =9.0, 9.0 Hz)	69.8
Glc-5	3.40 (ddd, <i>J</i> =9.0, 6.0, 2.0 Hz)	74.0
Glc-6	4.28 (dd, <i>J</i> =12.0, 2.0 Hz) 4.04 (dd, <i>J</i> =12.0, 6.0 Hz)	62.7

^a ¹³C NMR chemical shifts extracted from HSQC and HMBC spectra.

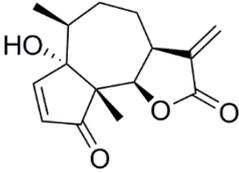
Table S14. HRESIMS and ^1H and ^{13}C NMR Spectroscopic Data for Compound 13 (DMSO- d_6 , 500 MHz for ^1H , and 125 MHz for ^{13}C NMR)

coronopilin		
accurate mass found	265.1344 [M+H] ⁺	
accurate mass calculated	265.1362 [M+H] ⁺	
molecular formula	C ₁₅ H ₂₀ O ₄	
		
NMR data^a		
Position	^1H [ppm]	^{13}C [ppm]^b
1	-	83.2
2	2.34 and 2.25 (m)	31.6
3	2.51 and 1.65 (m)	30.6
4	-	218.0
5	-	58.6
6	4.76 (d, $J=8.2$ Hz)	79.0
7	3.31 (m)	43.5
8	1.89 (m) 1.71 (m)	26.5
9	2.16 (m) 1.49 (m)	29.3
10	2.09 (m)	41.1
11	-	140.4
12	-	170.2
13	6.05 (dd, $J=2.6$ Hz) 5.69 (dd, $J=2.0$ Hz)	120.3
14	1.11 (d, $J=7.6$ Hz)	17.5
15	0.97 (s)	13.5

^a NMR chemical shifts recorded in a mixture of coronopilin (**13**) and parthenin (**14**) (70:30).

^b ^{13}C NMR chemical shifts extracted from HSQC and HMBC spectra.

Table S15. HRESIMS and ¹H and ¹³C NMR Spectroscopic Data for Compound 14 (DMSO-*d*₆, 500 MHz for ¹H, and 125 MHz for ¹³C NMR)

parthenin		
accurate mass found	263.1247 [M+H] ⁺	
accurate mass calculated	263.1271 [M+H] ⁺	
molecular formula	C ₁₃ H ₂₀ O ₄	
		
NMR data ^a		
Position	¹H [ppm]	¹³C [ppm]^b
1	-	83.0
2	7.63(d, <i>J</i> =6.0 Hz)	165.0
3	7.09(d, <i>J</i> =6.0 Hz)	130.9
4	-	210.4
5	-	58.0
6	4.85 (d, <i>J</i> =8.2 Hz)	78.0
7	3.44 (m)	43.3
8	1.83 (m) 2.09 (m)	27.4
9	2.13 (m) 1.54 (m)	29.1
10	2.10 (m)	39.3
11	-	140.8
12	-	170.2
13	6.09 (dd, <i>J</i> =2.8 Hz) 5.71 (dd, <i>J</i> =2.3 Hz)	120.3
14	1.02 (d, <i>J</i> =7.6 Hz)	16.7
15	1.10 (s)	17.4

^a NMR chemical shifts recorded in a mixture of coronopilin (**13**) and parthenin (**14**) (70:30).

^b ¹³C NMR chemical shifts extracted from HSQC and HMBC spectra.

Table S16. *A. artemisiifolia* pollen lots.

sample	pollen lot #	collection date	collection area	plant origin	supplier
Aa1	020512201	2012	Missouri, USA	cultivated	Allergon
Aa2	P3024704-3	28.08.2012	Missouri, USA	wild	Greer Labs
Aa3	P3744652-9	29.08.2012	Czech Republic	wild	Greer Labs
Aa4	P3396120-3	27.08.2013	Missouri, USA	wild	Greer Labs
Aa5	P335440-18	01.08.2008	Illinois, USA	wild	Greer Labs
Aa6	P3392522-19	20.08.2013	Illinois, USA	cultivated	Greer Labs

Table S17. *A. psilostachya* pollen lots.

sample	pollen lot#	collection date	collection area	plant origin	supplier
Ap1	P3406293-1	16.08.2013	California, USA	wild	Greer Labs
Ap2	P3801393	27.09.2013	Texas, USA	wild	Greer Labs
Ap3	P3757208	19.08.2014	Oklahoma, USA	wild	Greer Labs
Ap4	P3701938-2	06.08.2013	Arizona, USA	wild	Greer Labs
Ap5	P4129072-1	05.08.2014	Arizona, USA	wild	Greer Labs

4. CONCLUSIONS AND PERSPECTIVES

Asthma is a global burden mostly affecting developed and industrialized countries, worldwide. The variety and complexity of biological mechanisms causing asthma symptoms is considerably high and diverse, and is making the diseases still nowadays difficult to predict and to fully understand.

We investigated the role of plant pollen secondary metabolites as potential factors involved in the exacerbation of allergen-mediated asthma. This study didn't give direct answers to the question whether pollen secondary metabolites are involved in the worsening of asthma symptoms. However, we obtained relevant results regarding the modulation of airways by pollen secondary metabolites in mouse. A first aspect which still needs to be clarified is the exact pathway involved in the pollen-mediated increase of intracellular Ca^{2+} -influx. The application of cinnamaldehyde as positive control in the Ca^{2+} -imaging experiments is still not enough to discriminate the TRPA1-controlled Ca^{2+} influx from other Ca^{2+} regulating pathways (e.g. L-type calcium channels). To answer this intriguing question, *in-vitro* knock-out models of murine afferent neurons should be used in order to clarify this point. Another valid but certainly less exclusive approach would be the chemical inhibition of specific channels. Nevertheless, the caveat of this cheaper and straightforward strategy could be identified in the selection of inhibitors that are not selective enough and thus leading to undesired inhibition, or activation, of other proteins. The second aspect which calls for further experimental work are the effects induced by pollen extracts and purified compounds on a higher biological system, like murine trachea. In the specific case, it was possible to observe that the pre-treatment with sesquiterpene lactones (STLs) mixture coronopilin/parthenin was able to significantly reduce the maximal methacholine-induced constriction of tracheal preparations, indicating a possible synergistic effect of STLs on the muscarinic or other receptors. This partial relaxation was not observed when pre-treating tracheal rings with the *A. psilostachya* THF extract, containing other metabolites besides the sesquiterpene lactones coronopilin and parthenin. On the other hand, both *A. psilostachya* THF extract and the STLs mixture coronopilin/parthenin induced the relaxation of the methacholine pre-constricted trachea. These partial results are of valuable importance for developing *ad hoc* murine tracheal models, which should clarify the role of coronopilin and parthenin and identify other metabolites in the *A. psilostachya* THF extract possibly involved in the modulation of the tracheal tone.

From the phytochemical perspective, there is still need to further explore the pollen metabolome originating from different plant species. On one side, we observed strong discrepancies in the

electrophiles concentration between the pollen and the rest of the plant (leaves and flowers) in *Ambrosia artemisiifolia*. On the other side, it was interesting to observe the presence of metabolites with electrophilic properties in non-Asteraceae pollen originating from e.g. *Phleum pratense* (Poaceae) or *Urtica dioica* (Urticaceae). These are only few reasons for natural products chemists to further investigate in this area. Unfortunately, the limiting factor in the pollen phytochemical study is the availability and the stockpiling of the raw material. This aspect calls for further development and investigations on two levels: (i) development and optimization of electrophiles trapping methods for enrichment of electrophilic metabolites in limited amounts of pollen extracts; (ii) HPLC-based activity profiling of pollen extracts on cellular and/or tissue assays. These strategies would on one side facilitate the characterization of electrophiles in Asteraceae and non-Asteraceae pollen, and on the other side would help to determine if non-electrophilic secondary metabolites are involved in the modulation of specific proteins.

Our new comparative metabolite profiling tool in ACD/MS Workbook Suite is now fully operative for data obtained from low resolution LC-ESIMS systems (Shimadzu, Waters, and Agilent). The utility of this recently developed tool was demonstrated in few projects (Chapter 3.3; bioactivity profiling on *Daphne giraldii*, lymphogenesis project - Maria Teresa Faleschini). The next developmental steps will be focused to further increase the versatility of the tool, by making it usable with LC-HRMS, GC-MS, and APCI. By achieving this goal, users will be able to process large amounts of LC-MS/HRMS datasets in an automated, systematic, and comprehensive way, and simultaneously store, analyze, and database the results in a flexible computational platform like ACD/Labs.

The newly developed web application FreeClust is freely accessible on-line and can read data from text files or Excel spread sheets. This makes our application highly versatile, enabling the clustering of data originating from various instruments, (not necessarily related to LC-MS) and thus broadening its application possibilities.

