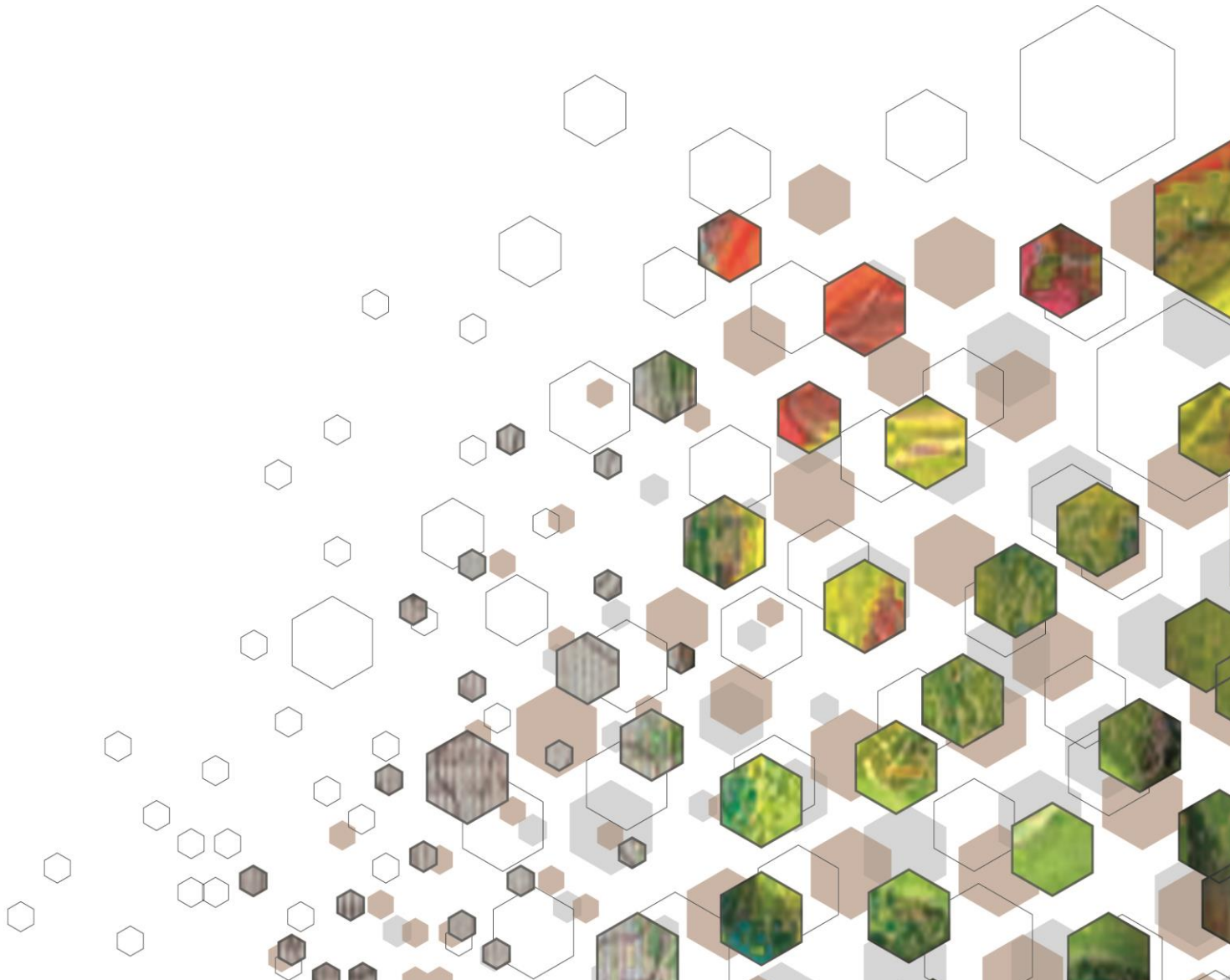




Phytochemical investigation of *Acronychia* species using NMR and LC-MS based dereplication and metabolomics approaches

Eirini Kouloura

National and Kapodistrian University of Athens
University Paris Descartes

Phytochemical investigation of
Acronychia species using NMR and
LC-MS based dereplication and
metabolomics approaches

Eirini Kouloura

November 2014

PhD thesis under joint supervision

National and Kapodistrian University of Athens, Department of Pharmacy

University Paris Descartes, Department of Pharmacy

Examining committee

Professor Alexios-Leandros SKALTSOUNIS¹

Professor Sylvie MICHEL²

Professor Jean-Luc WOLFENDER³

Associate Professor Roxani TENTA⁴

Researcher Agnes CHARTIER⁵

Professor Sofia MITAKOU¹

Professor Emmanuel MIKROS⁶

Professor Marie-Christine LALLEMAND²

Associate Professor Gregory GENTA-JOUE²

¹ Department of Pharmacognosy and Natural Products Chemistry, Faculty of Pharmacy, National and Kapodistrian University of Athens

² Department of Pharmacognosy, UMR/CNRS 8638, Faculty of Pharmacy, University Paris Descartes

³ School of Pharmaceutical Sciences, EPGL, University of Geneva, University of Lausanne

⁴ Department of Nutrition and Dietetics, Charokopeio University, Athens, Greece

⁵ Institute of Organic and Analytical Chemistry (ICOA), University of Orleans-CNRS

⁶ Department of Pharmaceutical Chemistry, Faculty of Pharmacy, National and Kapodistrian University of Athens

Thesis summary

Medicinal plants constitute an unfailing source of compounds (natural products – NPs) utilised in medicine for the prevention and treatment of various diseases. The introduction of new technologies and methods in the field of natural products chemistry enabled the development of high throughput methodologies for the chemical composition determination of plant extracts, evaluation of their properties and the exploration of their potentials as drug candidates. Lately, metabolomics, an integrated approach incorporating the advantages of modern analytical technologies and the power of bioinformatics has been proven an efficient tool in systems biology. In particular, the application of metabolomics for the discovery of new bioactive compounds constitutes an emerging field in natural products chemistry.

In this context, *Acronychia* genus of Rutaceae family was selected based on its well-known traditional use as antimicrobial, antipyretic, antispasmodic and anti-inflammatory therapeutic agent. Modern chromatographic, spectrometric and spectroscopic methods were utilised for the exploration of their metabolite content following three basic axis: a) phytochemical investigation, identification of secondary metabolites and evaluation of their biological properties, b) development of analytical methods for identification of acetophenones (chemotaxonomic markers of the genus) and dereplication strategies for the chemical characterisation of extracts and c) application of metabolomic methodologies (LC-MS & NMR) for comparative analysis (between different species, origins, organs), chemotaxonomic studies (between species) and compound-activity correlations. Bioinformatics and sophisticated statistical tools were employed especially towards the latter methodology. In particular:

The application of various analytical and chromatographic techniques (LC-PDA, -ELSD, -HRMS, SFC-UV, HRNMR; chiral separation, FCPC) enabled the phytochemical exploration and isolation of numerous NPs (alkaloids, lignans, terpenoids). Among them, several acetophenones, an important and interesting class for the genus (*Acronychia*-type acetophenones-AtA) from *A. pedunculata* were isolated and identified. Their particular structural characteristics compelled their detailed and unambiguous structural investigation. In

addition, the evaluation of their pharmacological properties, including antimicrobial, cytotoxic (against human tumour cell lines) and anti-inflammatory activity, of all isolated AtA was assessed *in vitro*.

The small number of *Acronychia*-type acetophenones (AtA) despite their important pharmacological profile, the limited information regarding their identification as well as the low number of species investigated so far, led to the development of an identification and dereplication strategy for further exploration of AtA in complex mixtures. In particular, a UPLC-HRMS & HRMSⁿ method was developed and applied for the detailed and accurate identification thereof. Fragmentation patterns and certain ion motifs enabled the construction of decision trees for AtA identification and a MS nomenclature scheme was suggested. Moreover, this methodology was utilised for the analysis of different *A. pedunculata* extracts enabling the dereplication of known AtA and the discovery of potentially new ones. This approach, having taken advantage of state of the art analytical platforms and concepts, could be incorporated for the investigation of other complex mixtures and plant extracts.

NMR and UPLC-HRMS-based metabolomics (metabolic profiling) were used for the investigation of various *Acronychia* species and organs, from different geographical origin. Dereplication tools were also employed for the characterization of their chemical profiles. Statistical models were developed for the comparison of the different extracts revealing organ- and species-specific biomarkers. Moreover, discrimination models allowed the identification of certain compounds giving new insight into chemotaxonomic issues. Specific strategies were established for the identification of the revealed biomarkers including statistical integration of datasets obtained from both platforms. Statistical correlation of the analytical and pharmacological data resulted in the development of a methodology for tracking bioactive metabolites in extracts without any prior purification. This approach provides a novel tool for the drug discovery in natural products field.

Overall, the combination of these different approaches comprises a complete and comprehensive study of *Acronychia* species and important knowledge was obtained concerning their metabolite composition and pharmacological profile. Metabolomics were proved a promising tool in modern pharmacognosy and phytochemistry while better insight into the utilisation of new approaches in natural products drug discovery process was obtained.

Acknowledgements

This study was performed under joint supervision in the University of Athens and the University Paris Descartes. This context reinforced the strong collaboration between the laboratories and gave the chance for new collaborations. I would like to take the opportunity to acknowledge all collaborators involved in and many people contributed directly or indirectly for the accomplishment of this thesis. Particularly,

I would like first to thank all the members of my examining committee Prof. Alexios-Leandros Skaltsounis, Prof. Sylvie Michel, Associate Prof. Roxani Tenta, Prof. Jean-Luc Wolfender, Prof. Sofia Mitakou, Prof. Emmanuel Mikros, Prof. Marie-Christine Lallemand, Associate Prof. Gregory Genta-Jouve, and Researcher Agnes Chartier for agreeing to participate in this procedure, for their time and expertise.

I would also like to thank the members of my scientific board Prof. Alexios-Leandros Skaltsounis, Prof. Sylvie Michel and Prof. Sofia Mitakou for their advices, their encouragement and the transmission of their scientific knowledge during all these years. I really appreciate their interest on my scientific work and the confidence they gave to me from the very beginning.

I sincerely would like to thank Prof. Alexios-Leandros Skaltsounis for his support, scientific guidance and interest throughout all the years of my postgraduate studies. Prof. Alexios-Leandros Skaltsounis provided me the opportunity to work as a member of his worldwide acknowledged scientific team and be financially supported. During these years he encouraged me to be trained in state of the art analytical techniques as part of the laboratory equipment and to be involved in different projects. I feel really grateful for motivating me participating in international projects and through them get seconded in international laboratories and attend outstanding scientific congresses. I deeply want to thank him for respecting and supporting our initiative to organize the first congress of Postgraduate Students in Pharmaceutical Sciences held in Athens. I would like also to acknowledge him for been for me a great example of worldwide acknowledged scientist.

Acknowledgements

I truly would like to acknowledge Prof. Sylvie Michel for accepting to jointly supervise my thesis, her scientific interest and suggestions to accomplish this research. I am grateful for welcoming me at the laboratory of Pharmacognosy in Paris and her assistance during my stay in Paris.

I feel really grateful to Prof. Sofia Mitakou for being advising me and supporting me during all the years of my postgraduate studies. I would really want to thank her for all her prosperous assistance to overcome difficulties and accomplish the double PhD procedures between Athens and Paris. I would like to thank her for her persistence and patience to establish inexistent administrative schemes. I would like to express my gratitude for revising my thesis and her helpful comments.

I feel deeply grateful to Dr. Maria Halabalaki, my scientific supervisor, for her constant inspiration and guidance during all the years of my post graduate studies. Maria introduced me to the world of science, encouraged me in each step of this route and contributed significantly to my scientific emancipation. I would like to thank her for transmitting me her knowledge in natural products chemistry and metabolomics and in the same time giving me freedom to explore these fields. I deeply appreciate her contributions of time, insightful ideas, and fundraising to make my PhD an experience productive and stimulating. Apart from the scientific advices, Maria was always encouraged me and taught me to consider opportunities regarding next steps contributing significantly in my professional maturity. I would like to acknowledge her for motivating me to participate in important conferences and get seconded to international laboratories. Her advices and our collaboration, all these years, will be a point of reference for me.

During my thesis, I had the great pleasure of working with Prof. Emmanuel Mikros. I feel grateful for his devotion of time to discuss about NMR experiments and molecular mechanics stimulations performed throughout this thesis and his effective guidance. I really appreciate his contribution for being seconded to Imperial College for two months.

I feel indebted to Prof. Marie-Christine Lallemand for her precious knowledge and advices she provided me at the beginning of my postgraduate studies and her interest on my work during all these years. I would like also to thank her for her prosperous assistance to handle administrative procedures.

I would like to acknowledge Associate Prof. Gregory Genta-Jouve for assisting me with bioinformatics. I want to thank him for his efficient advices and his prompt interventions, even from distance, to provide solutions.

During this thesis, I had the great pleasure to work with NMR experts and get thorough knowledge from them. I would like to express my gratitude to Dr. Dimitra Benaki who significantly contributed to the accomplishment of the NMR experiments, for her commitment and kind interest. I'm also grateful to Dr. Sarantos Kostidis for his important contribution at the beginning of this thesis regarding NMR experiments.

I feel grateful to Dr. Vasillios Myrianthopoulos for the devotion of time to perform molecular mechanics stimulations and docking studies described throughout this thesis, his kind assistance and helpful comments on the manuscript.

I would like to acknowledge Dr. Mark Litaudon from Institut de Chimie des Substances Naturelles (ICSN/CNRS) for providing us all *Acronychia* plant materials utilized during this thesis.

I would like to thank all collaborators for their commitment to undertake the *in vitro* pharmacological evaluation assays performed during this thesis. I feel grateful to Prof. Richard Jove and Prof. Sangkil Nam from City of Hope for the accomplishment of the cytotoxic assays on human tumour cell lines summarized also in a publication. I would like to acknowledge Prof. Simon Gibbons and Dr. Proma Khondkar from University College London (UCL) for providing important insights into the antibacterial activity of several natural compounds. I really appreciate the contribution of Prof. Oliver Werz, Dr. Carlo Pergola and Mrs Verena Krauth from University of Jena for the anti-inflammatory assays they performed and the great collaboration we had for the development of metabolomics platform to track bioactivity in complex mixtures.

I would like to thank Associate Prof. Jane Hubert for her commitment to perform the ^{13}C NMR experiments for all *Acronychia* extracts at the University of Reims. I really appreciate her enthusiasm and passion for this project and the great collaboration and interaction we had during the last year of this thesis.

I deeply appreciate the assistance of Mr Kostas Tsarhopoulos for his advices and guidance concerning applications and technical issues using LTQ-Orbitrap apparatus.

Acknowledgements

I feel grateful to the secretary team of the laboratory, Foteini Kapsali, Katerina Makropoulou, Chrysa Lemonaki and Chrysa Lionaki for their kindness and effective assistance during these years.

During these years, I had the great fortune of meeting and interacting with outstanding scientists and personalities. The impact of each of them on my scientific and personal development was significant and I owe them many thanks.

I would like to thank Assistant Prof. Evangelos Gikas for being a rich pole of knowledge and being always willing for transmitting his passion for many scientific topics. I really appreciate his assistance during these years in mass spectrometry applications and troubleshooting and his instant solutions in computer issues.

I feel grateful to Senior Lecturer Dr. Hector Keun for welcoming me to his group at Imperial College and offering me the opportunity to spend a formative period interacting with exceptional scientists. I would like to thank Dr. Alexis Siskos for the great collaboration we had during these two months and for introducing me to the group and everyday life so agreeably. I would also like to thank Esther Vizcaino and Marcelo Segura for their generous advices and the fruitful interaction we had. I feel thankful to the entire group for being open and friendly to me.

I would like to thank Prof. Maique Biavatti for welcoming us to her group at the Federal University of Santa Catarina and involving us in many interesting projects and activities. I would like also to acknowledge Prof. Eloir Schenkel and Prof. Miriam Falkenberg for their kindness and interest. I feel really grateful to all the people I get to know at UFSC for their friendship, great conversations and unforgettable experiences they offered me.

At the beginning of my postgraduate studies I had the great fortune of meeting and becoming close friend with Nikos Lemonakis, Eliza Chaita, Maria Lalioti, Katerina Koutsogiannopoulou, Giorgos Papaeustatiou, Vincent Brieudes and Tina Kounadi. Until now they provide me their great friendship, moral support, constant encouragement and unforgettable social moments that made this period of PhD a unique period in my life.

Specifically, I would like to thank Nikos for being a close companion to this common route for sharing good and difficult moments and transmitting me always his positive energy. I feel grateful of having shared the training period in Brazil together!

I would also like to thank Eliza for being all the years of my postgraduate studies close to me and for representing the “voice of reason” for me during challenging moments.

During the last years, I had the chance to collaborate closely with brilliant persons who became also valuable friends. I would like to thank Dr. Katerina Termentzi, Dr. Panagiotis Stathopoulos, Job Tchoumtchoua and Eirini Baira for their friendship, support and their willingness to help. I will never forget our fruitful discussions!

I would like to acknowledge Job for our collaboration to accomplish a book chapter and for being an inspiration of ‘following your dreams’ to me!

I really appreciate the close collaboration I had with talented people in order to accomplish various scientific projects. I feel grateful to Eirini Danika and Solomon Amoah for having spent unforgettable moments together, for their friendship, kindness and joy they offered me during our collaboration.

I would like to thank some close friends Virginia Kukula, Christelle Lemus and Nicola Gaboriaud I had the chance to make during these years for being always beside me, for their interest, kindness and support.

I would like also to acknowledge all past and present group members that I had the pleasure to work with or alongside.

I deeply would like to thank my dearest friends Marianna, Klairi and Marili for our lifelong friendship, their love, emotional support and understanding all these years. I feel grateful to have found the sisters I never had!

I would like to thank my family for their love and encouragement during these years. Specifically, I owe a great thanks to my grandmother for her unconditional love, her support and her advice to enjoy all the aspects of life. I feel really grateful to my parents, Pinelopi and Giannis, who raised me with love and provided me a supportive environment in each step of my life. I appreciate their passion to offer me an excellent education and all the sacrifices they have made for me. I feel also grateful to Eleni for being always next to me, for her love and care. I would like also to thank my beloved family members Mairi, Lena, Thanasis and Nikos and my cousins Giannis, Augi, Aristeidis, Periklis and Kyroula for being a source of affection and support. I feel really grateful to my godchild Dimitrakis for his affection and for offering so much joy in my life!

Acknowledgements

Lastly, I would like to thank enormously Roberto for being part of my life all these years and for standing next to me and supporting all my decisions. I really appreciate his sunny optimism, patience and affection as well as the uncountable trips he has done to stay the most possible close to me.

Thank you,

Eirini Kouloura

Table of Contents

Thesis summary	i
Acknowledgements	iii
CHAPTER 1	1
Introduction	3
1. Modern techniques in natural products chemistry	3
2. <i>Acronychia pedunculata</i> (L.) Miq.	5
2.1. Botanical and chemotaxonomic characteristics	5
2.2. Traditional use of <i>Acronychia</i> species	6
2.3. Phytochemical profile of <i>Acronychia</i> species	7
3. Current objectives	8
Experimental	9
1. Chemicals and instrumentation	9
2. Plant Material	10
3. Extraction of <i>A. pedunculata</i>	10
4. Analytical profiling for detection of acetophenone dimers	11
5. Targeted isolation of acetophenone dimers	11
6. Enantiomer screening of acetophenone dimers	14
7. Isolation of alkaloids	15
8. Conformational analysis of acetophenone dimers	15
9. Antibacterial activity	16
10. Evaluation of cytotoxic activity against human tumor cell lines	16
10.1. Cell Lines and Culture	16
10.2. Cell Viability Assays	16
11. Anti-inflammatory activity	17

Table of contents

11.1.	Expression and Purification of 5-LO.....	17
11.2.	Cell-free and cellular LO assays	17
11.3.	mPGES-1 assay	18
11.4.	Docking on 5-Lipoxygenase (5-LO).....	18
	Results and Discussion	19
1.	Targeted isolation of acetophenone dimers	19
1.1.	Analytical profiling of <i>Acronychia</i> extracts using different detectors	19
1.2.	Preparative isolation of acetophenone dimers	21
2.	Structure elucidation of acetophenone dimers.....	22
3.	Conformational analysis of acetophenone dimers.....	30
4.	Enantiomer resolution screening of <i>Acronychia</i> -type acetophenones.....	36
5.	Isolation of <i>Acronychia</i> -type acetophenone diastereomers by SFC	40
6.	Antibacterial activity of <i>Acronychia</i> type acetophenones.....	44
7.	Cytotoxic activity of <i>Acronychia</i> type acetophenones against human tumour cell lines.....	45
8.	Anti-inflammatory activity of <i>Acronychia</i> type acetophenones	46
	Conclusion.....	51
	APPENDIX 1	53
	CHAPTER 2.....	103
	Introduction	105
1.	Background	105
2.	Multistage mass spectrometry in natural products.....	106
3.	Multistage mass spectrometry of AtA compounds.....	107
	Experimental.....	109
1.	Standards and reagents	109
2.	Sample preparation	109
3.	Mass spectrometry and data handling	109
	Results and Discussion	113

1. Structure elucidation of AtA using multistage HRMS	113
1.1. Fragmentation pattern analysis for AtA in ESI(+)	116
1.1.1. Key features for identification of AtA in ESI(+)	124
1.2. Fragmentation pattern analysis for AtA in ESI(-)	125
1.2.1. Key features for identification of AtA in ESI(-)	133
1.3. Comparison of ESI and APCI ionization of AtA.....	134
2. UHPLC-ESI(-)-HRMS ⁿ analysis of <i>Acronychia</i> extracts.....	135
3. Reproducibility of MS ⁿ spectra	141
Conclusion.....	143
APPENDIX 2	145
CHAPTER 3.....	159
Introduction	161
1. Background	161
2. Metabolomics	162
3. Plant metabolomics workflow	165
3.1. Sample preparation	165
3.1.1. Harvesting of plant material.....	166
3.1.2. Extraction procedure.....	166
3.1.3. Sample preparation for analysis.....	167
3.2. Data acquisition.....	168
3.3. Data preprocessing.....	169
3.3.1. NMR data preprocessing.....	170
3.3.2. LC-MS preprocessing	170
3.4. Data pretreatment	171
3.4.1. Data filtering	172
3.4.2. Missing values	172
3.4.3. Normalization, scaling	173

Table of contents

3.5. Data analysis	174
3.6. Biomarker identification	176
4. The case study of <i>Acronychia</i> species	176
Experimental.....	179
1. Chemicals and instrumentation	179
2. Harvesting of plant material.....	180
3. Extraction of plant material	181
4. Sample preparation for analysis	181
4.1. NMR sample preparation	182
4.2. UPLC-HRMS sample preparation.....	182
5. NMR data acquisition and spectral processing.....	183
6. UPLC-HRMS data acquisition.....	183
6.1. UPLC conditions	183
6.2. HRMS conditions	184
7. NMR data preprocessing.....	185
8. UPLC-HRMS preprocessing.....	186
8.1. Raw data preparation	186
8.2. Peak detection/ Annotation of isotope and adduct peaks	186
9. NMR data filtering / Noise reduction.....	186
10. UPLC-HRMS data filtering / Noise reduction	187
11. UPLC-HRMS missing values estimation	187
12. Normalization and scaling	188
13. Data analysis	188
14. Biomarker identification	188
15. Anti-inflammatory evaluation of <i>Acronychia</i> extracts	189
15.1. Sample preparation	189
15.2. <i>In vitro</i> 5-LO cell free assay	189

Results and Discussion.....	191
1. NMR and UPLC-HRMS plant metabolomics workflow.....	191
1.1. Sample selection and preparation.....	193
1.2. NMR metabolomics platform	194
1.2.1. Sample preparation for NMR acquisition	194
1.2.2. NMR acquisition.....	195
1.2.3. NMR data preprocessing/ pretreatment	197
1.2.4. NMR dataset evaluation	197
1.3. UPLC-HRMS metabolomics platform.....	199
1.3.1. UPLC-HRMS data acquisition.....	199
1.3.2. UPLC-HRMS run sequence.....	200
1.3.3. UPLC-HRMS data preprocessing	201
1.3.4. UPLC-HRMS data evaluation	203
1.3.5. UPLC-HRMS data pretreatment.....	208
1.3.6. UPLC-HRMS dataset evaluation.....	210
2. UPLC-HRMS based metabolite identification and dereplication in different <i>Acronychia</i> extracts	212
3. ¹³ C NMR based identification of secondary metabolites and dereplication in <i>Acronychia</i> extracts	221
4. Investigation of <i>Acronychia</i> species taxonomy using NMR and LC-MS based metabolomics approaches	226
4.1. Classification of <i>Acronychia</i> samples using NMR based metabolomics	228
4.2. Classification of <i>Acronychia</i> samples using UPLC-ESI(+)-HRMS based metabolomics.....	232
4.2.1. Biomarkers identification.....	235
4.3. Discrimination of <i>A. laurifolia</i> and <i>A. pedunculata</i> species.....	243
5. Statistical integration of different metabolomics techniques for the identification of metabolites.....	248
5.1. Integration of NMR and MS datasets of <i>Acronychia</i> extracts samples using sPLS.....	250

Table of contents

5.2. Identification of metabolites based on the sPLS model of NMR and MS datasets	251
6. Discovery of 5-LO inhibitors from <i>Acronychia</i> samples using UPLC-HRMS based metabolomics and PLS regression.....	255
6.1. PLS and OPLS regression model for prediction of 5-LO inhibition.....	257
6.2. Identification of bioactive compounds by PLS and OPLS analysis	261
Conclusion.....	265
APPENDIX 3	269
References	285

CHAPTER

Targeted Isolation and Pharmacological
Evaluation of *Acronychia*-type Acetophenones

Introduction

1. Modern techniques in natural products chemistry

Natural products chemistry research has been extremely evolved during the last years mainly due to the introduction of new technologies. Nowadays, the application of new technologies in each step of the phytochemical investigation process is resulted in more automated, high throughput and comprehensive experimental conditions. In particular, phytochemical investigation involves mainly the extraction of plant material, the profiling of plant extracts, the isolation and purification of natural products and the structural elucidation of pure isolated compounds. The profiling of plant extracts is an important step in the overall process which is significantly facilitated by the introduction of hyphenated techniques in natural products chemistry field (Wolfender et al. 2006, Sarker et al. 2012). The hyphenation of high resolution separation techniques [e.g. high performance liquid chromatography (HPLC), ultra performance liquid chromatography (UPLC)] with advanced detection systems [e.g. mass spectrometry (MS), nuclear magnetic resonance (NMR), ultraviolet (UV) detectors] has allowed the profiling of the complex plant extracts prior to any isolation step. Therefore, profiling of crude extracts leads to the qualitative and quantitative estimation of the metabolite composition and the structural characterization of contained metabolites by the combination of online spectroscopic data (Wolfender et al. 2010). This information is of crucial importance for the efficiency of the isolation procedure since tedious isolation of the same natural products, time and expenses consumption can be easily circumvented. In addition, in cases of targeted isolation studies a detailed qualitative and quantitative assessment of the starting material is essential for the scheduling of the separation step.

Despite the tremendous development of analytical techniques providing strong evidences for the identification of natural products in complex mixtures, there is still the need of isolation of pure natural products. The achievement of several amounts of highly pure compounds constitutes an essential prerequisite for the complete structural elucidation of new natural products by further spectroscopic analysis and for their pharmacological evaluation using *in vitro* or *in vivo* experiments (Seger et al. 2013). The

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

isolation step traditionally is performed using liquid-solid chromatography techniques such as column chromatography operating in ambient or medium or high pressure. However, the combination of these techniques with other orthogonal techniques based on different separation mechanisms could be utilized for the more efficient recovery of pure natural products.

In particular, liquid–liquid partition techniques possess the uniqueness of using no-solid stationary phase, while separation is achieved between two immiscible liquid phases (Ito et al. 1970). In this case, the separation is based on the partition coefficient of each compound between the two phases and eventually closely eluting compounds by conventional chromatographic column may possess totally different partition coefficients which results in their efficient separation. Another important advantage of this technique is that eliminates the irreversible adsorptive loss of samples onto the solid support matrix observed in the conventional chromatographic column (Pauli et al. 2008). Therefore, the application of liquid–liquid partition principle provides a fast and efficient technique for the isolation of natural products.

Another alternative technique for preparative isolation of natural products is supercritical fluid chromatography (SFC). This technique is based on the application of a supercritical fluid, most commonly CO₂, as a mobile phase in combination with one or more polar organic solvents. Supercritical fluids have densities and dissolving capacities similar to those of certain liquids, but lower viscosities and better diffusion properties (Taylor 2009). Consequently, improved resolution of compounds comparing to classical liquid-solid chromatography techniques is observed permitting efficient separation in terms of velocity and mass capacity (White et al. 2005). In addition, the different selectivity and elution order of compounds using SFC renders it a good complementary technique to reversed phase HPLC. It is worth noting that SFC is proved a valuable technique for the separation isomers and enantiomers, and structurally related compounds (White 2005, Taylor 2010).

New technologies have contributed extremely in natural product chemistry research in each step of this long procedure. During this study, the application of different state of the art techniques was performed for the detection, targeted isolation and unambiguous structural elucidation of *Acronychia*-type acetophenones from *Acronychia pedunculata*.

2. *Acronychia pedunculata* (L.) Miq.

2.1. Botanical and chemotaxonomic characteristics

Acronychia pedunculata is one of the 48 species of *Acronychia* genus (Bayly et al. 2013). *Acronychia* genus belonging to the Rutaceae family is represented by scrubs, small trees or trees widely distributed in the Indochina peninsula, in eastern Australia, and the islands of the western Pacific Ocean. The principal botanical features of *Acronychia* genus have been described by Hartley (Hartley 1974). The leaves are opposite monofoliolate, trifoliolate or unifoliolate, having leaflets entire, pinnately veined and articulated at the base. Inflorescences are axillary, paniculate, subcorymbose often reduced to one flower. The flowers are bisexual with 4 sepals, distinct or connate basally, usually imbricate and persistent. The petals are 4 distinct, valvate, usually white narrowly triangular becoming reflexed, deciduous or rarely semipersistent in fruit. The stamens are 8 as long as the petals, the antesealous slightly longer than the antepetalous. The filaments are flattened ending up to 2 anthers. The ovaries are tetralocular or rarely octalocular, with or without fissures between the locules, scarcely differentiated stigmas from the style with two ovules in each locus. The fruits are tetralocular or rarely octalocular represented mainly by a drupe, with or without septicidal fissures. The fruit is so characteristic that Hartley mentioned that could be used to distinguish *Acronychia* from the other genera of Rutaceae. The epicarp is semifleshy spongy-crustaceous or woody when dry with or without evident mesocarp and cartilaginous to pergamentaceous endocarp. Finally, the seeds are usually ellipsoid and shiny.

Acronychia pedunculata is a widely spread species distributed in rainforests of India, Sri Lanka, Indonesia, Malaysia, Philippines, Taiwan and southern mainland China (Figure 1). The botanical characteristics that distinguish *A. pedunculata* from the other species are the petioles which are usually longer and probably its name derives from this since "pedunculata" is a Latin word meaning "slender stalked". In addition, the leaflets are variable in shape but not suborbicular, the ovary and disc are presented entirely pubescent and inflorescences are often more than 10 cm long. However, Hartley denotes the synonymy confusion of *A. pedunculata* due to its variance according to the geographical distribution with the most dominant synonyms those of *Jambolifera pedunculata* Vahl., *A. laurifolia* Bl. and *A. apiculata* Miq. This variance is reflected in the gradation in disc size and

consequently in fruit size occurring mainly from India to China, Malaysia and Borneo (Hartley 1974). In particular in Malaysia Burkill referred to the species as *Acronychia laurifolia* Blume (*Acronychia pedunculata* (L) Miq.) (Burkill 1966). Chinese botanists in national and regional Floras called the species *Acronychia pedunculata* (L) Miq. (*Jambolifera pedunculata* L.) (Chun et al. 1965). Lecomte, in the 'Flore générale de l'Indochine' mentioned *Acronychia laurifolia* Bl. and a number of synonyms among them *Acronychia pedunculata* (L) Miq.; *Cyminosma pedunculata* DC.; *Jambolifera pedunculata* Vahl. (Lecomte 1907-1951). For further confusion Wang separates the species, ascribing to *Acronychia pedunculata* to the monsoon rain forests of Hainan Island and southern Yunnan alone (Wang 1961). Regarding the abovementioned literature data is clear that the nomenclature and synonymy of this species is still a confusing issue however, during this study the name *A. pedunculata* will be used as this botanical name is proposed from Hartley and Chinese flora.

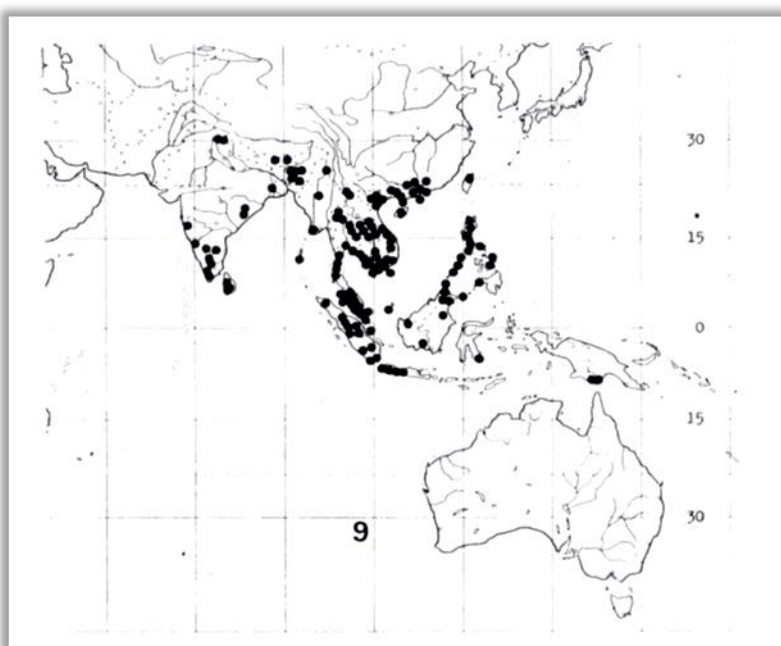


Figure 1: Distribution map of *Acronychia pedunculata* (L.) Miq.

2.2. Traditional use of *Acronychia* species

Acronychia species have been traditionally used in the eastern world from food condiments and salad ingredients to therapeutics in folk medicine. In particular, many parts of the plant including the roots, stems, leaves, and fruits of certain species in this genus have

been used for centuries in eastern traditional medicine for the treatment of asthma, cough, diarrhea, itchy skin, pain, rheumatism, scales, sores, and ulcers and also for their antihemorrhagic, antipyretic, and aphrodisiac activities (Rahmani et al. 1996). Biological evaluation of extracts of *A. pedunculata* has shown significant antiplasmodial (Horgen et al. 2001), antibacterial (Jayasinghe et al. 2006), and antifungal (Rodrigo et al. 2007) activities as well as cytotoxic effects for several cancer cell lines (Horgen et al. 2001) confirming in some extent the traditional use of this genus. However, the anti-inflammatory activity hinted beneath its traditional use is not confirmed up to date. Moreover, the essential oil obtained from flowers and leaves has been employed in cosmetics (Epifano et al. 2013).

2.3. Phytochemical profile of *Acronychia* species

Among the 48 *Acronychia* reported species, only 12 species have been investigated concerning their phytochemical profile. Nevertheless, a number of diverse secondary metabolites have been isolated from the studied species. In particular, several compounds belonging to alkaloids, acetophenones, flavonoids, phenolic acids, lignans, coumarins, steroids, and triterpenes have been reported. The most abundant chemical category found in the majority of *Acronychia* species is alkaloids belonging to quinolone or acridone basic structures (Lamberton et al. 1953). Predominantly, from *A. pedunculata* quinolone alkaloids and particularly furoquinoline derivatives have been isolated (Cui et al. 1999). Another relatively abundant chemical category in *Acronychia* genus is prenylated acetophenones. Prenylated acetophenone monomers and dimers have been reported from *A. pedunculata*, *A. trifoliolata* and *A. vestita*. Specifically, the presence of acetophenones dimers is uniquely reported from *Acronychia* genus indicating their value as chemotaxonomic markers of the genus (Adersen et al. 2007). Biological interest in prenylated acetophenones has focused on their antioxidant (Su et al. 2003), cytotoxic (Wu et al. 1989), and anti-inflammatory (Pathmasiri et al. 2005) activities, while the acetophenone dimers have been assessed for cytotoxicity against numerous cancer cell lines, with acrovestone reported to exhibit significant cytotoxicity (Wu et al. 1989, Oyama et al. 2003). Despite their potential chemotaxonomic and biological importance, only a small number of acetophenone dimers (4 derivatives) have been isolated and biologically evaluated from the genus *Acronychia*.

3. Current objectives

The objective of the current study was the detection of the acetophenone dimer derivatives in *A. pedunculata* extracts by means of metabolite profiling using various hyphenated analytical platforms and their targeted isolation utilizing orthogonal chromatographic techniques such as fast centrifuge partition chromatography (FCPC), reversed phase chromatography and supercritical fluid chromatography (SFC) techniques. In total, seven acetophenone dimer derivatives were obtained, among them three natural products. Furthermore, a detailed conformational analysis was intended since acetophenone dimers exhibit particular structural characteristics leading to a dynamic conformational equilibrium in solution and thus, incomprehensive NMR data. For this purpose, NMR studies of acetophenones dimers over a range of different temperatures were combined to molecular mechanics calculation. In addition, the presence of chiral centres in acetophenone dimers implied their enantioselective resolution using chiral normal phase chromatography. In order to ascribe the pharmacological profile of this group of compounds based on the literature data concerning the traditional use of *Acronychia* genus as therapeutic, *in vitro* evaluation of potential activities was assessed. Different assays were implemented to define the activity of these compounds involving antimicrobial activity, cytotoxic activity against human tumour cell lines and anti-inflammatory activity.

Experimental

1. Chemicals and instrumentation

Analytical HPLC-PDA analysis was performed on a Thermo Finnigan apparatus equipped with a PDA Spectra System UV6000LP. Data acquisition and processing was performed using Chromquest 4.1 software. HPLC-ELSD analysis was performed on a Waters instrument equipped with a Waters 600E pump and an Alltech 2000 ES Evaporative Light Scattering Detector (ELSD) controlled by Waters Empower Pro Software. HPLC-ESI-TOF-HRMS chromatograms were obtained by a micrOTOF ESI-MS system (Bruker Daltonics) and data acquisition and processing was performed using Bruker HyStar 3.0 software. HPLC-APCI-Orbitrap-HRMS chromatograms were acquired on a Hybrid LTO-Orbitrap instrument (Thermo Finnigan) controlled by Xcalibur software version 2.0.7. Calculations of logP values, as measure of molecular hydrophobicity, for the estimation of the chromatographic behaviour of potential contained metabolites were performed by MarvinSketch 5.3.3 software.

Fast centrifugal partition chromatography (FCPC) was performed using a CPC Kromaton with a 1000 mL column and a Laboratory Alliance pump with a pressure safety limit of 50 bar. A manual sample injection valve was used to introduce the samples into the column, with the rotation adjusted at 800 rpm and the flow rate held at 20 mL/min. Semipreparative HPLC was performed on a Thermo Finnigan apparatus equipped with a UV Spectral System UV2000 using an Ascentis RP-8 C8 (250 × 10 mm i.d.; 5 µm) (Discovery Supelco) column. Super critical fluid chromatography (SFC) was performed on a Thar SFC instrument (Pittsburgh, PA) including a fluid delivery module with high pressure pumps for CO₂ and modifier delivery, an autosampler, a column oven for temperature control, a makeup pump, a 2298 Photodiode Array (PDA) detector, an automated back pressure regulator (ABPR) for column pressure controlling, and a fraction collector. Data acquisition and processing was performed using Mass lynx 4.1.

Optical rotations were obtained on a Perkin-Elmer 341 polarimeter. Nuclear magnetic resonance (NMR) spectra were recorded on a Bruker Advance III 600 spectrometer

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

operating at 600 and 150 MHz for ^1H and ^{13}C , respectively, equipped with a 5 mm BBI probe and using CDCl_3 (Aldrich) as solvent. Chemical shifts (δ) are expressed in ppm with reference to the residual CHCl_3 solvent signals ($\delta \text{ H } 7.26/\delta \text{ C } 77.0$). The 2D NMR experiments (COSY, HSQC, HMBC, and NOESY) were performed using standard Bruker microprograms.

Ultra gradient grade MeOH (Carlo Erba), H_2O (filtered) and glacial acetic acid (Fisher optima ACS grade) were used for reversed phase HPLC separations. HPLC gradient grade (Fisher Scientific) cyclohexane (cHex), isopropanol (IPA), Ethyl acetate (EtOAc), Methyl tert-butyl ether (MTBE) and Ethanol (EtOH) were utilized for normal phase HPLC. Solvents used for extraction, FCPC were of analytical grade or technical grade and purified by distillation. For SFC CO_2 of 99.99% purity and HPLC gradient grade (Fisher Scientific) acetonitrile (ACN), methanol (MeOH) and isopropanol (IPrOH) were utilized.

2. Plant Material

The trunk bark of *Acronychia pedunculata* was collected in the dense rainforest of Mersing, Johore State, Malaysia, in April 1999. The plant was identified by botanist T. Leong Eng. A voucher specimen (KL-4882) has been deposited at the Herbarium of the Forest Research Institute, Kepong, Malaysia.

3. Extraction of *A. pedunculata*

The dried and pulverized trunk bark of the plant (2.4 kg) was extracted with Et_2O ($2 \times 5 \text{ L}$ of Et_2O , 48 h per extraction) at room temperature. After concentration under reduced pressure, 224.2 g of the first Et_2O extract were obtained. The plant residue after the first extraction with Et_2O was alkalinized with 10% NH_4OH and extracted successively with Et_2O , CH_2Cl_2 , and MeOH ($3 \times 5 \text{ L}$, each solvent). The Et_2O extract (22.5 g) obtained after alkalinization was partitioned with 10% HCl ($3 \times 150 \text{ mL}$) until the negative reaction of the aqueous phase using Mayer's reagent and then alkalinized to pH 8–9 with 28% NH_4OH and extracted with CH_2Cl_2 ($6 \times 150 \text{ mL}$) using Mayer's reagent to control the process. An organic fraction (20.3 g) and an aqueous fraction (452 mg) rich in alkaloids were obtained after this procedure.

4. Analytical profiling for detection of acetophenone dimers

Analytical profiling of the Et₂O extract was performed before separation procedures using identical reversed phase chromatographic conditions and different detectors. In particular, an Ascentis RP-8 C₈ (250 × 4.6 mm i.d.; 5 µm) (Discovery Supelco) column was applied for all aforementioned analysis and a mobile system containing MeOH and H₂O+2% acetic acid were used for HPLC separations. A gradient elution program was applied starting from 70% up to 100% MeOH in 90 min, following 10 min of isocratic elution (100% MeOH), returning back to initial conditions in 5 min and re-equilibrating with 70% MeOH for 5 min. The flow rate was set at 1 mL/min and volumes of 20 µL were injected from working solution of 10 mg/mL of all obtained extracts. The detection was performed using 4 different detectors. PDA detector was set to record UV absorption range of 200–600 nm and three different UV channels at 254, 280 and 365 nm were chosen to monitor the run. ELSD detector was operated at 80 °C with a N₂ flow rate at 2.1 L/min in order to vaporize the mobile phase. In addition, the option “impactor off” was selected since there was no suspicion of volatile compounds in *A. pedunculata* extracts and the gain option was set at 4. HPLC-ESI-TOF-HRMS chromatograms were acquired at a mass range of *m/z* 150–1500 in positive and negative mode. Nitrogen was used as nebuliser gas, at 2 bar and as dry gas at 9 L/min and 240 °C and spray voltage was set at 4.5 kV. Internal mass calibration of each analysis was performed by the infusion of 1% sodium formate in isopropanol:water 5 mM sodium hydroxide, 1:1 (v/v), at a gradient time of 110 min using a diverter valve. HPLC-APCI-Orbitrap-HRMS chromatograms were acquired also in positive and negative mode. In APCI vaporizer temperature was set at 350 °C, discharge current at 5 µA, capillary temperature at 275 °C, capillary voltage at 3V and tube lens at 90V. Full scans were acquired in profile mode at 30000 (FWHM) resolution and a mass range of *m/z* 200–1000.

5. Targeted isolation of acetophenone dimers

After the detection of potential acetophenone dimers by profiling the Et₂O extract using multiple detectors, the initial fractionation of the extract was performed by FCPC. Concerning FCPC procedure, the most critical decision step is the selection of the suitable two-phase system. Briefly, the selection of the two-phase system was performed based on

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

already described two-phase systems in the literature (Table I) by applying the 'shake-flask' method (Ruey-Shiuan et al. 1995). This involved testing of the solubility of the analytes in the two-phase system and the volume ratio formed between two phases. Subsequently, the partition coefficients of the targeted compounds were measured by reversed phase HPLC as described by Marston et al. applying the elution method used for the analytical profiling of the extract (Marston et al. 2006). After the evaluation of the results, system No 9 was found the most appropriate for the separation of this mixture of compounds.

Table I: Two phase solvent system tested for the selection of the appropriate system for the separation of acetophenone dimers

System No	n-Hex	EtOAc	MeOH	H ₂ O
1	2	3	1	1
2	6	3	2	5
3	5	3	3	5
4	4	6	3	3
5	1	1	1	1
	n-Hex	EtOAc	Acetone	H ₂ O
6	1	1	1	1
	EtOAc	n-Butanol	H ₂ O	
7	2	1	3	
	n-Hept	EtOAc	MeOH	H ₂ O
8	4	1	4	1
9	10	1	10	1
10	5	4	5	4
11	5	1	5	1
12	10	3	10	3
	n-Hept	EtOAc	MeCN	H ₂ O
13	10	1	8	1
14	8	3	8	1
15	10	2	7	1

In particular, 15g of the extract were fractionated with a two-phase solvent system composed of n-heptane–ethyl acetate–methanol–water (10:1:10:1), using the organic phase first as mobile phase. The separation afforded 80 fractions of 50 mL each. After the collection of the 60 initial fractions, the apparatus was switched to the descending mode, and another 20 fractions were collected. Two fractions out of the eighty obtained contained in high purity the two major metabolites of the Et₂O extract, acrovestone (4,

243.2 mg; fraction 13) and acrofolione A (**6**, 471.8 mg; fraction 71), while the other metabolites were isolated as mixtures. The purity of all fractions was estimated using reversed phase HPLC-PDA at 280 nm (Figure A 1). The same conditions as for analytical profiling were applied using working solutions of 5 mg/mL for FCPC fractions and 1 mg/mL for pure compounds. Fraction 13 was found to contain 87% acrovestone, while fraction 71, 92% acrofolione A. Moreover, yellow crystals of acrovestone (**4**, 182.5 mg) precipitated from fraction 13.

In order to isolate all prenylated acetophenone dimers present in the extract, FCPC fractions containing mixtures of acetophenones were purified further by semipreparative HPLC using an elution program of a 90 min linear gradient from 70% to 100% MeOH, then 10 min pure MeOH, 1 min back to initial conditions, and 9 min for re-equilibration (70% MeOH) with a flow rate of 5 mL/min. Separation of a portion (80 mg, 10 mg per injection) of fraction 8 afforded acropyrone (**1**, 10.7 mg), and similarly a portion of fractions 17 and 18 (120 mg, 10 mg per injection) was subjected to semipreparative HPLC to obtain acropyranol A (**2**, 7.3 mg) and acrovestenol (**5**, 8.7 mg). Acropyranol B (**3**, 3.2 mg) and acrofolione B (**7**, 8.5 mg) were isolated from fraction 28 (100 mg, 10 mg per injection), while an additional quantity of acropyranol B (**3**, 2.8 mg) was isolated from fractions 33–39 (50 mg, 10 mg per injection). Finally, acronyline (**8**, 26.0 mg), a prenylated acetophenone monomer, was isolated as transparent crystals from fraction 69.

Acropyrone (**1**): yellowish oil; $[\alpha]_D^{25}$ 0 (c 1, CHCl₃); UV (MeOH) $\lambda_{\max}$ (log ϵ) 214 (4.13), 226 (4.15), 289 (4.12), and 333 (3.67, sh) nm; ¹H NMR (CDCl₃, 600 MHz) and ¹³C NMR (CDCl₃, 150 MHz), see Table A 1: NMR spectroscopic data of Acropyrone (**1**) at 47 °C; APCI(+)-HRMS m/z 575.2610 [M+Na]⁺ (**3**), 553.2788 [M+H]⁺ (calcd for C₃₂H₄₁O₈, 553.2807) (**18**), 319.1905 (100), 303.1594 (16), 235.0966 (15).

Acropyranol A (**2**): yellowish oil; $[\alpha]_D^{25}$ 0 (c 1, CHCl₃); UV (MeOH) $\lambda_{\max}$ (log ϵ) 210 (4.47), 229 (4.42), 292 (4.37), 340 (3.98, sh) nm; ¹H NMR (CDCl₃, 600 MHz) and ¹³C NMR (CDCl₃, 150 MHz), see Table A 3; APCI(+)-HRMS m/z 571.2906 [M+H]⁺ (calcd for C₃₂H₄₃O₉, 571.2913) (**3**), 319.1907 (100), 253.1073 (22).

Acropyranol B (**3**): yellowish oil; $[\alpha]_D^{25}$ 0 (c 1, CHCl₃); UV (MeOH) $\lambda_{\max}$ (log ϵ) 209 (4.38), 231 (4.36), 298 (4.29), 329 (4.06, sh) nm; ¹H NMR (CDCl₃, 600 MHz) and ¹³C NMR (CDCl₃, 150 MHz), see Table A 5; APCI(+)-HRMS m/z 571.2897 [M+H]⁺ (calcd for C₃₂H₄₃O₉, 571.2913) (**52**), 319.1904 (100), 253.1070 (14).

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

Acrovestone (**4**): yellow amorphous crystals; $[\alpha]_D^{25}$ 0 (c 1, CHCl₃); UV (MeOH) λ_{max} (log ϵ) 211 (4.55), 231 (4.50), 295 (4.40), 340 (4.19, sh) nm; ¹H NMR (CDCl₃, 600 MHz) and ¹³C NMR (CDCl₃, 150 MHz), see Table A 7; APCI(+)-HRMS m/z 555.2955 [M+H]⁺ (calcd for C₃₂H₄₃O₈, 555.2963) (25), 319.1908 (100), 237.1124 (14).

Acrovestenol (**5**): yellowish oil; $[\alpha]_D^{25}$ 0 (c 1, CHCl₃); UV (MeOH) λ_{max} (log ϵ) 210 (4.29), 231 (4.21), 299 (4.15), 335 (3.99, sh) nm; ¹H NMR (CDCl₃, 600 MHz) and ¹³C NMR (CDCl₃, 150 MHz), see Table A 9; APCI(+)-HRMS m/z 571.2906 [M+H]⁺ (calcd for C₃₂H₄₃O₉, 571.2913) (4), 319.1910 (100), 235.0968 (15).

Acrofolione A (**6**): yellowish oil; $[\alpha]_D^{25}$ 0 (c 1, CHCl₃); UV (MeOH) λ_{max} (log ϵ) 211 (4.45), 231 (4.47), 293 (4.41), 335 (4.01, sh) nm; ¹H NMR (CDCl₃, 600 MHz) and ¹³C NMR (CDCl₃, 150 MHz), see Table A 11; APCI(+)-HRMS m/z 571.2900 [M+H]⁺ (calcd for C₃₂H₄₃O₉, 571.2913) (4), 335.1856 (23), 319.1905 (100), 253.1074 (17).

Acrofolione B (**7**): yellowish oil; $[\alpha]_D^{25}$ 0 (c 1, CHCl₃); UV (MeOH) λ_{max} (log ϵ) 210 (4.45), 232 (4.47), 291 (4.40), 334 (4.06, sh) nm; ¹H NMR (CDCl₃, 600 MHz) and ¹³C NMR (CDCl₃, 150 MHz), see Table A 13; APCI(+)-HRMS m/z 571.2907 [M+H]⁺ (calcd for C₃₂H₄₃O₉, 571.2913) (13), 319.1908 (100), 253.1073 (11).

Further purification of diastereomer derivatives was performed using SFC. In particular, separation of acrofolione A (**6**) diastereomers, which was isolated in adequate quantity for further preparative purification by SFC, was achieved on a Viridis Silica 2-Ethylpyridine column (150 × 10 mm i.d.; 5 μ m) (Waters). The mobile phase was composed of supercritical CO₂ modified by MeOH and the flow rate was set at 15 mL/min. The elution program started with an initial conditioning step of 5% MeOH for 1 min, then a gradient step from 5% to 40% of MeOH in 6 min and an isocratic step of 1 min of 40% MeOH were followed to conclude with 1 min step returning to the initial condition and 1 min of re-equilibration. A working solution of 5 mg/mL was prepared and the injection volume was 200 μ L.

6. Enantiomer screening of acetophenone dimers

The occurrence of chiral centres in acetophenone dimer structures lead to the development of a screening strategy in order to determine the number of different enantiomers and estimate their relative ratio. Therefore, enantiomer separation of isolated acetophenone dimers was carried out using multiple immobilized polysaccharide-based

chiral stationary phases (CSPs). Specifically, analytical columns (250 × 4.6 mm i.d.; 5 µm) with different chiral selectors CHIRALPAK IA, CHIRALPAK IB, CHIRALPAK IC and CHIRALPAK ID were applied. Different isocratic elution methods were tested in a context of an enantiomer separation screening strategy which are quoted in section 4 of results and discussion part. All compounds were prepared in solutions of 100 µg/mL diluted in Hexane/Isopropanol (99/1). For all tested methods injection volumes of 10 µL were applied, the temperature of the column oven was set at 25 °C and the flow rate at 1 mL/min. The detection was performed using a UV absorption range of 200–600 nm and three different UV channels at 254, 280 and 365 nm were chosen to monitor the run.

7. Isolation of alkaloids

The alkaloid rich fraction obtained after the alkalinization was submitted to flash chromatography using CH₂Cl₂–MeOH (100:0 to 30:70) gradient solutions, which afforded four furoquinoline alkaloids, dictamnine (**9**, 8.5 mg), pteleine (**10**, 3.9 mg), evolitrine (**11**, 26.6 mg), and kokusaginine (**12**, 32.0 mg). Dictamnine and pteleine were isolated for the first time from this species.

8. Conformational analysis of acetophenone dimers

Conformational analyses of all acetophenone dimers were performed at the Molecular Mechanics (MM) level using the mixed Low mode/Monte Carlo algorithm (10000 steps, AMBER* forcefield) and an implicit GB/SA chloroform solvent model as implemented in Macromodel v. 9 software (Mohamadi et al. 1990). Resulting conformations were clustered using XCluster software. Boltzmann normalized populations of all dominant conformers were determined for each compound using the MM energies calculated at 273K. Critical to the conformation analysis of the acetophenone dimers was the relative arrangement of the isopentyl and isoprenyl chains leading to the discrimination between two major conformational states (Figure 8).

9. Antibacterial activity

Different *Staphylococcus aureus* strains, *Bacillus subtilis*, *Streptococcus pneumoniae*, *Escherichia coli*, *Klebsiella pneumoniae*, *Proteus*, *Pseudomonas aeruginosa*, *Salmonella typhi* bacteria strains were used to assess the antibacterial activity of acrovestone (**4**) and acrofolione A (**6**). All strains were cultured on nutrient agar (Oxoid) prior to determination of minimum inhibitory concentration (MIC). Cation-adjusted Mueller–Hinton broth (MHB; Oxoid), containing Ca^{2+} (20 mg/L) and Mg^{2+} (10 mg/L), was used for susceptibility tests. Bacterial inocula equivalent to the 0.5 McFarland turbidity standard were prepared in normal saline and diluted to give a final inoculum density of 5×10^5 cfu/mL. Test compounds were dissolved in DMSO before dilution into MHB for use in MIC determinations. The inoculum (125 μL) was added to all wells and the microtitre plate was incubated at 37 °C for 24 h. The MIC was recorded as the lowest concentration at which no bacterial growth was observed (Gibbons et al. 2000). Norfloxacin was used a positive control.

10. Evaluation of cytotoxic activity against human tumor cell lines

10.1. Cell Lines and Culture

Human A2058 melanoma and DU145 prostate cancer cell lines were obtained from the American Type Culture Collection (ATCC). Cells were cultured in RPMI 1640 medium supplemented with 10% heat-inactivated FBS and 1% penicillin/streptomycin. Normal human dermal fibroblast (NHDF) cells were purchased from Lonza. Cells were cultured in DMEM media containing 10% FBS.

10.2. Cell Viability Assays

Cell viability assays were performed as described previously (Liu et al. 2011). Briefly, cells were seeded onto 96-well plates at a density of 5000 cells per well. After overnight incubation, cells were treated for 48 h with the isolated compounds **1–7** or DMSO as the vehicle control. MTS Reagent (CellTiter 96 AQueousOne Solution Cell Proliferation Assay;

Promega) was added to each well according to the manufacturer's instructions. Absorbance was monitored at 490 nm using a microplate reader (Bio-Rad). Cell viability (%) was normalized to the vehicle control. Each experiment was performed in triplicate or quadruplicate. Sorafenib was used as positive control. IC₅₀ values against DU145 and A2058 cells were 5.1 ± 0.7 and 3.8 ± 0.9 μ M, respectively.

11. Anti-inflammatory activity

11.1. Expression and Purification of 5-LO

5-LO was expressed in *E. coli* BL21 (DE3) cells, transformed with pT3-5LO, and purification of 5-LO was performed as described previously (Fischer et al. 2003). Thus, *E. coli* were collected by centrifugation ($7,700 \times g$ for 15 min), lysed with 50 mM triethanolamine/HCl, pH 8.0, 5 mM ethylenediaminetetraacetate (EDTA), 60 μ g/mL soybean trypsin inhibitor (STI), 1 mM phenylmethylsulphonyl fluoride (PMSF), 1 mM DTT and 1 mg/mL lysozyme, homogenized by sonication (3×15 sec) and centrifuged at $10,000 \times g$ for 15 min and then at $40,000 \times g$ for 70 min at 4 °C. The supernatant was then applied to an ATP-agarose column (Sigma; Deisenhofen, Germany). Partially purified 5-LO was immediately used for activity assays.

11.2. Cell-free and cellular LO assays

For determination of 5-LO activity in cell-free assays, samples of partially purified human 5-LO (1 mL, in PBS buffer containing 0.1 % glucose and 1 mM EDTA) were incubated 10 min at 4 °C with vehicle (0.1% DMSO, control) or test compounds, pre-warmed for 30 sec at 37 °C and 2 mM CaCl₂ and the indicated concentrations of AA were added. The reaction was stopped after 10 min at 37 °C by addition of 1 mL ice-cold methanol and 30 μ L of 1 N HCl, 200 ng prostaglandin B₁ and 500 μ L of PBS were added.

For assays of intact cells, freshly isolated neutrophils (5×10^6) or monocytes (2×10^6) were resuspended in 1 mL PGC buffer. After pre-incubation with vehicle or compounds for 10 min, LO product formation was started by addition of the respective stimuli, as indicated. The reaction was stopped as indicated for purified 5-LO.

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

Formed 5-LO metabolites, 12(S)-H(P)ETE and 15(S)-H(P)ETE were extracted and analysed by HPLC as described (Werz 2002). 5-LO products include LTB₄ and its all-trans isomers, and 5(S)-H(P)ETE.

11.3. mPGES-1 assay

Preparations of A549 cells and determination of mPGES-1 activity was performed as described (Koeberle et al. 2008). In brief, cells were treated with 2 ng/mL IL-1 β for 72 h at 37 °C, 5% CO₂, harvested, sonicated and homogenized (homogenization buffer: 0.1 M potassium phosphate buffer, pH 7.4, 1 mM PMSF, 60 μ g/mL STI, 1 μ g/mL leupeptin, 2.5 mM glutathione, and 250 mM sucrose). The homogenate was centrifuged at 10,000 $\times$ g for 10 min and 174,000 $\times$ g for 1 h at 4 °C, and the resulting pellet (microsomal fraction) was resuspended in 1 mL homogenization buffer, and the total protein concentration was determined. Microsomal membranes were diluted in potassium phosphate buffer (0.1 M, pH 7.4) containing 2.5 mM glutathione. Test compounds or vehicle were added, and after 15 min at 4 °C reaction (100 μ L total volume) was initiated by addition of 20 μ M PGH₂. After 1 min at 4°C, the reaction was terminated using stop solution (100 μ L; 40 mM FeCl₂, 80 mM citric acid, and 10 μ M 11 β - PGE₂ as internal standard), followed by solid-phase extraction and analysis of PGE₂ by HPLC as described previously (Koeberle et al. 2008).

11.4. Docking on 5-Lipoxygenase (5-LO)

The crystal structure of the stable S663D mutant of 5-LO in complex with the natural substrate arachidonic acid (pdb code: 3V99) was downloaded from RCSB Protein Data Bank and was used for docking calculations. The protein was prepared using the PPrep module of Maestro. All isolated *Acronychia*-type acetophenones were built and prepared using the LigPrep module of Maestro. Docking of the aforementioned compounds was performed in the active site of the enzyme utilizing Glide SP algorithm, a preconstructed grid of the protein binding pocket and default settings. Docked poses were evaluated in term of binding affinity using the Glidescore scoring function (Friesner et al. 2004, Halgren et al. 2004).

Results and Discussion

1. Targeted isolation of acetophenone dimers

In the context of targeted isolation of acetophenone dimers, *Acronychia pedunculata* trunk barks were selected according to the literature data which suggests the presence of this group of compounds and alkaloid compounds as main chemical categories. In a first step, a specific extraction protocol for the isolation of alkaloids separately from the rest contained secondary metabolites was followed. In particular, this protocol involved the extraction of the plant material with diethyl ether (Et₂O), the alkalization of the plant residue and the successive extraction with Et₂O, CH₂Cl₂ and MeOH. Due to the high nonpolar nature of acetophenone dimers, the Et₂O extract, presumably containing acetophenone dimers, was partitioned using HCl solution and after a second alkalization step an organic and an aqueous fractions were obtained, rich in acetophenone and alkaloids, respectively (see Experimental Section 3 for further details).

1.1. Analytical profiling of *Acronychia* extracts using different detectors

In order to achieve thorough information concerning the metabolite composition of the obtained acetophenone rich extract, the profiling of the extract was performed by reversed phase HPLC hyphenated with a number of diverse detectors (Figure 2). The nonpolar nature of Et₂O extract as well as the *in silico* calculations of logP values as a measure of molecular hydrophobicity for known acetophenone dimers (6.3- 8.1) implied the utilization of a mobile phase with high organic solvent content (see Experimental Section 4 for further details). This enabled the effective separation of the contained metabolites and the collection of valuable information concerning the metabolite composition of *A. pedunculata* extract based on the chromatographic, spectral and spectrometric features.

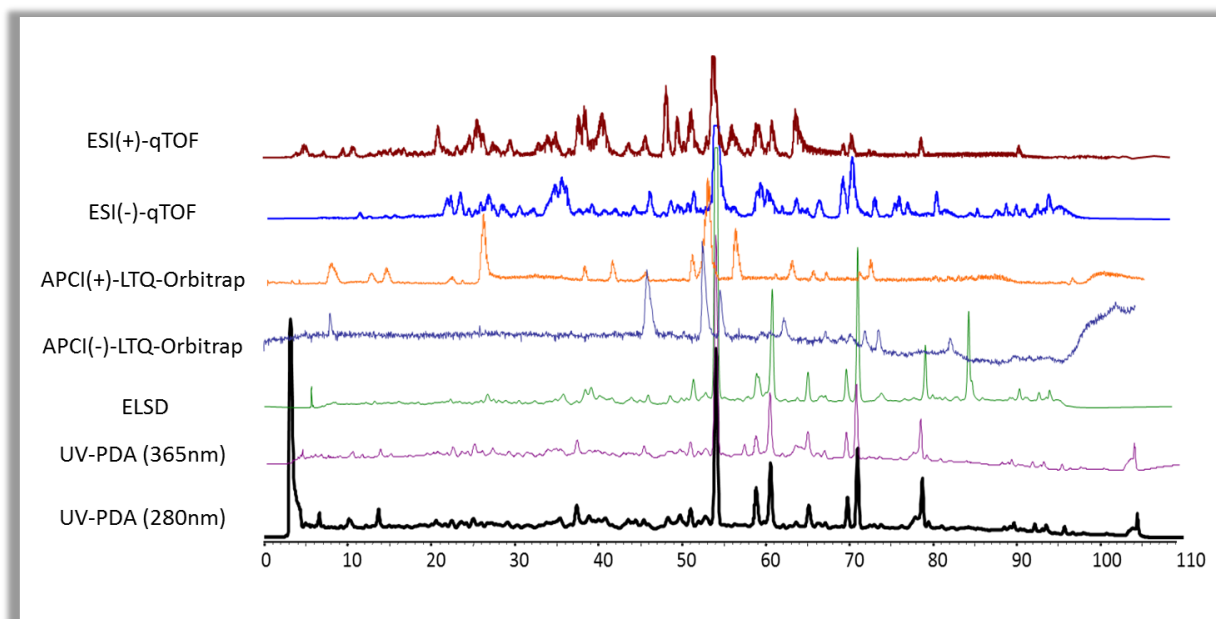


Figure 2: HPLC based chromatograms of Et₂O extract obtained from trunk barks of *A. pedunculata* using a number of diverse detectors

In particular, the analysis of Et₂O extract of *A. pedunculata* by HPLC-PDA revealed the presence of a group of compounds with characteristic absorption maxima at approximately 240 and 290 nm and a shoulder at 320- 350 nm. In total, seven major peaks presented this spectral features and comparison with the literature data confirmed the potential presence of acetophenone dimers (Figure A 2) (Wu et al. 1989, Oyama et al. 2003, Pathmasiri et al. 2005). The analysis of the Et₂O extract using ELSD detector presented similar chromatographic appearance to the chromatograms obtained by HPLC-PDA analysis. The analysis by ELSD is based on the scattering of the light beam by the different eluting compounds, thus the detection of compounds lacking chromophores is possible (Vervoort et al. 2008). In this context, only one additional peak at 85.5 min was observed denoting the presence of a highly unpolar compound lacking chromophores on its structure. This is probably attributed to a triterpene compound such as b-sitosterol (Rahmani et al. 1996) or b-amyrin (De Silva et al. 1991) previously reported in the literature.

Furthermore, Et₂O extract of *A. pedunculata* was analysed by HPLC hyphenated with mass spectrometers (MS). These analyses enabled the extraction of spectrometric characteristics from the ionized molecules in the extract. In addition, the high resolution MS analyses obtained by both TOF and Orbitrap analysers allowed the estimation of elemental compositions (ECs) and RDB eq. values of the contained metabolites contributing

significantly to the detection of compounds of interest. The ionization of molecules was performed using different type of atmospheric pressure ionization (API) sources. Specifically, TOF analyser was equipped with an electrospray ionization (ESI) source leading to the ionization of a number of compounds in both positive and negative mode. HPLC-ESI($\pm$)-TOF analysis facilitated the ionization of compounds eluting between 20 and 50 minutes which did not exhibit spectrometric characteristics potentially attributed to acetophenone dimers. Moreover, the interface of atmospheric pressure chemical ionization (APCI) source with Orbitrap analyser was applied for the analysis of the extract under study and led mainly to the ionization of more unpolar compounds. Interestingly, the extraction of mass spectra from the HPLC-APCI(+)-Orbitrap chromatogram, revealed common spectrometric characteristics for the seven compounds with similar absorption maxima (Figure A 3). In particular, a base peak in source characteristic fragment ion at m/z 319.1905 denoted the presence of compounds belonging to the same chemical group. The molecular ions of all seven peaks potentially corresponding to acetophenone dimers were detected and the elemental compositions (ECs), RDB eq. values and Δm (ppm) between the theoretical and measured values were calculated for each of these peaks. The exploration of their molecular ions unravelled the occurrence of five isomer compounds with a proposed molecular formula of $C_{32}H_{42}O_9$. The rest two compounds exhibited slightly different molecular formulas calculated for $C_{32}H_{42}O_8$ and $C_{32}H_{40}O_8$. Overall, comparing spectrometric and literature data, all seven compounds could be potentially attributed to acetophenone dimers.

The combination of the above mentioned findings, obtained from the analysis of Et_2O extract of *A. pedunculata* by HPLC hyphenated with multiple detectors, and literature data reinforced the initial assumption suggesting the occurrence of acetophenone dimers in this extract.

1.2. Preparative isolation of acetophenone dimers

After the profiling of the Et_2O extract of *A. pedunculata* by HPLC hyphenated with multiple detectors, seven major possible acetophenone dimers were detected. The isolation of these compounds in preparative scale was aimed in order to obtain adequate quantity for efficient structure elucidation and pharmacological assessment of the purified and potentially new acetophenone dimers.

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

As a first fractionation step of the Et₂O extract, FCPC technique was selected in order to achieve efficient separation of the contained acetophenone dimer derivatives with good recovery, in short time and consuming rational quantity of solvents. The two phase system applied for the separation of the Et₂O extract of *A. pedunculata* resulted in the isolation of two acetophenone dimers acrofolione A (**6**) and acrovestone (**4**) in high purity and important quantities while the rest derivatives were isolated in mixtures and further purified by semi-preparative HPLC (see Experimental Section 5 for further details). At this point, it is worth noting that FCPC enabled the separation of compounds that would not be possible to separate by semi-preparative HPLC. In particular, compounds closely eluted in the analytical reversed phase HPLC chromatogram, with difference of one to two minutes such as acropyranol B (**2**), acrofolione B (**7**) and acrovestone (**4**), acrovestenol (**5**) were resolved separately by FCPC. Thus, the combination of FCPC technique with semi-preparative reversed phase HPLC was proved highly efficient for the preparative isolation of acetophenone dimers from a complex mixture.

2. Structure elucidation of acetophenone dimers

Acetophenone dimers or *Acronychia*-type acetophenones may be considered as a particular group of acetophenones exhibiting specific structural features. They are polyhydroxylated, fully substituted derivatives consisting of two aromatic rings linked to an isopentyl chain. Apart from the hydroxy groups present, isoprenyl, acetyl, and methoxy units compose their common substituents, also additional rings derived from the isoprenyl moiety after cyclization have been observed. The presence of multiple hydroxy groups on these molecules results in the formation of extended inter and intra hydrogen bonds (Wu et al. 1989). All the dimers reported so far as well as the new natural products described in the current study (compounds **1-3**) present structural differences only in one aromatic ring (ring A) while the B aromatic ring and the isopentyl chain remain constant (Figure 3). Moreover, *Acronychia*-type acetophenones exhibit an extended degree of similarity and symmetry regarding the substitution of the two aromatic rings. For instance, in acrovestone (**4**), which represents a model compound for *Acronychia*-type acetophenones, the two aromatic rings differ only in the presence of a methyl group.

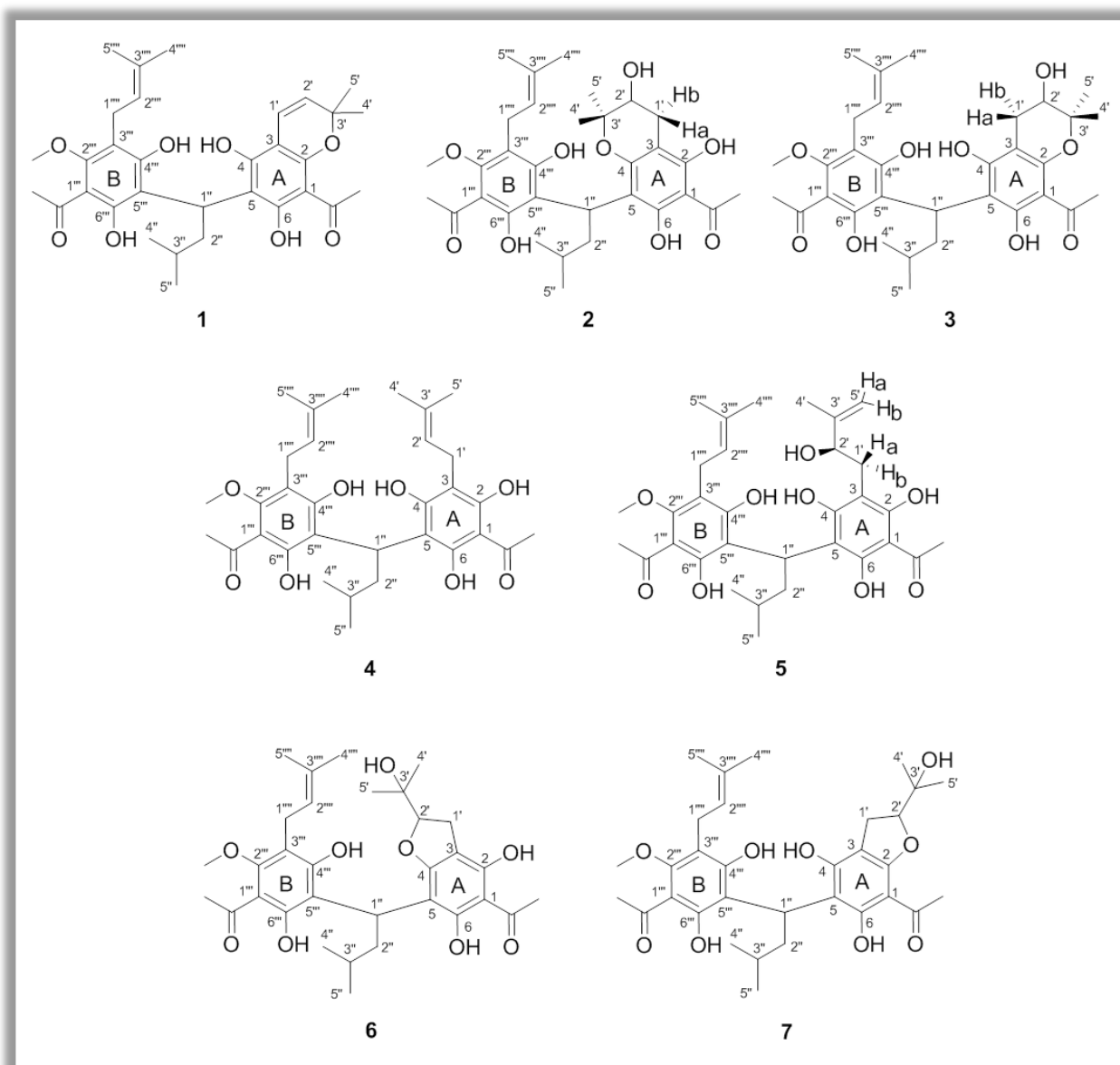


Figure 3: *Acronychia*-type acetophenones isolated from the Et₂O extract of *A. pedunculata*

These structural peculiarities together with the presence of numerous rotamers complicate their structure elucidation using NMR spectroscopy. Specifically, the lack of signals in the aromatic region, signal overlapping, and poorly resolved and broad peaks in the NMR spectra complicate their identification and have resulted in inaccurate structural conclusions. In order to overcome these difficulties, NMR experiments over a broad range of temperatures (from 0 to 52 °C) of the isolated compounds were performed supporting the presence of conformational rotamers equilibrium in solution. Therefore, the experiments recorded at 47 °C (320 K) were utilized for their structure elucidation. The fast interconversion of rotamers in the NMR time scale due to increased temperature resulted in

the improvement of peak shape of several proton signals with more comprehensive multiplicity, leading consequently to better resolution, facilitating the structure elucidation of acetophenone dimers considerably (see Results and Discussion 3 for further details). However, at 320 K the resolution of hydroxy protons was not possible and the NMR data obtained from experiment performed at 273K (0 °C) were utilized to obtain valuable information from NOESY and HMBC spectra.

Compound 1 was isolated as optically inactive yellowish oil and, therefore, characterized as a racemic mixture. The UV spectrum in MeOH showed characteristic absorption maxima for an acetophenone dimer at 214, 226, 289, and 333 (sh) nm. Its molecular formula was deduced as $C_{32}H_{40}O_8$ from the APCI(+)-HRMS data, implying 13 degrees of unsaturation (Figure A 4). The high-resolution mass spectrum revealed a pseudomolecular ion at m/z 553.2790 $[M+H]^+$ (calcd for 553.2796) as well as a fragment ion at m/z 319.1904, which is characteristic of all such isolated dimers. Based on its accurate mass ($\Delta m = 0.0419$ ppm), the proposed elemental composition, and the ring double-bond equivalent value, this ion corresponds to a fragment derived by the cleavage of the basic acetophenone skeleton at C-5 (see Chapter 2 for further details). As an *Acronychia*-type acetophenone, compound 1 exhibited two fully substituted aromatic rings connected to an isopentyl chain. Despite the lack of signals in the aromatic region of the ^1H NMR spectrum, the presence of these rings was determined by characteristic signals corresponding to several deshielded quaternary carbons in the HMBC spectrum as well as from its accurately measured molecular mass and the degree of unsaturation evident. Following the common structural pattern of an *Acronychia*-type acetophenone, ring B appeared identical to known compounds, while ring A was assigned as 1- [5,7-dihydro-2,2-dimethyl-2H-1-benzopyran-8-yl]ethanone (IUPAC nomenclature). For all isolated dimers, the typical ring B substituents are a methoxy group in an *ortho* position to an isoprenyl and an acetyl group, in a *meta* position to two hydroxyl groups, and in a *para* position to the characteristic isopentyl chain connecting the two aromatic rings.

Specifically, the ^1H NMR spectrum of 1 (Table A 1, Figure A 5) displayed signals ascribable to a characteristic isopentyl chain. The deshielded methine H-1" signal between the two aromatic rings resonated at δ_H 4.73 as a triplet ($J = 7.7$ Hz), and the methylene H-2" was observed at δ_H 2.17 as a multiplet, correlating with the corresponding carbon atoms at δ_C 28.3 and 39.2, respectively, as indicated from the HSQC spectrum (Figure A 7). The methine H-3" occurred as a multiplet at δ_H 1.41, while the protons of the two methyl groups,

H-4'' and H-5'', resonated together as a broad peak at δ_H 0.88. The carbon at δ_C 26.8 was attributed to C-3'', and the signal at δ_C 22.5 to C-4'' and C-5'', due to their correlations with the corresponding protons in the HSQC spectrum. Moreover, the COSY experiment confirmed the sequence of the protons of the isopentyl chain. The position of the isopentyl chain between the two aromatic rings was determined by cross-peak correlations of H-1'' with the downfield-shifted quaternary aromatic carbons at δ_C 158.6, 160.7, and 162.4 in the HMBC spectrum, which were assigned as C-4, C-6/6'', and C-4'', respectively (Figure A 8).

The isoprenyl unit of **1** was deduced from the presence of a characteristic spin system consisting of a methylene proton at δ_H 3.29 that appeared as a broad singlet (H-1''') correlating with a carbon (HSQC spectrum) at δ_C 22.9. Also, an olefinic proton appeared as a broad triplet at δ_H 5.19 ($J = 6.5$ Hz, H-2'''), correlating with a carbon at δ_C 123.0 along with two methyl groups at δ_H 1.69 (3H, s, H-4''') and 1.76 (3H, s, H-5''') correlating with carbon signals at δ_C 25.7 and 17.9, respectively. The HMBC spectrum revealed a cross-peak correlation between H-1''' and an olefinic quaternary carbon at δ_C 131.4 and was assigned therefore as C-3'''. Moreover, useful NOE correlations were observed between H-1''' and H-5''' signals, as well as between those of H-2''' and H-4''', and were used to define the relative orientation of the two methyl groups (Figure A 6). In the NOESY spectrum were also observed correlations between the protons of the methoxy group at C-2''' at δ_H 3.71 (3H, s, CH₃O-2''') and the H-1''' and H-5''' protons. The position of the isoprenyl unit on ring B was defined through the HMBC correlation of H-1''' with two downfield shifted quaternary carbons at δ_C 160.0 (^3J) and 162.4 (^3J) assigned as C-2''' and C-4'''. Likewise, the correlation of the CH₃O-2''' protons with C-2''' (^3J) observed in the HMBC spectrum and the correlation of H-1'' with C-4''' (^3J) and C-6''' (^3J) revealed the positions of the ring B substituents. Finally, a NOE correlation of the methoxy protons CH₃O-2''' with the protons at δ_H 2.70 (3H, s, CH₃CO-1''') and a HMBC correlation of the CH₃CO-1''' protons with C-1''' (^3J) at δ_C 108.0 supported the presence of an acetyl group in an *ortho* position to the CH₃O-2'''.

An additional 2,2-dimethyl-2H-pyran ring attached to ring A in **1** was deduced by the two characteristic doublets at δ_H 6.64 (H-1') and 5.43 (H-2') in the ¹H NMR spectrum with a coupling constant of 9.9 Hz, typical of the olefinic protons of a pyran ring. Correlations of these protons with two carbon atoms at δ_C 117.0 and 124.9, respectively, were evident in the HSQC spectrum. Also characteristic were the signals of the H-4' and H-5' methyl groups that resonated at δ_H 1.47 as a broad singlet integrating for six protons and correlating with a carbon atom at δ_C 27.9 assigned as C-4'/C-5'. Moreover, the presence of a quaternary

carbon at δ_C 78.0 correlating with H-1' (^3J), H-2' (^2J), H-4' (^2J), and H-5' (^2J) in the HMBC spectrum resulted in its assignment as C-3'. The assignment of the fusion of the ring to C-2/C-3 was confirmed unambiguously through NOE correlation of the protons of the methyl groups (H-4' and H-5') with the methyl protons of the acetyl group at δ_H 2.68 (3H, s, CH₃CO-1). Moreover, two pairs of singlets were observed at δ_H 15.92/16.10 and 15.61/15.75 in the ^1H NMR spectrum (Figure A 10) corresponding to the downfield shifted OH-6 and OH-6'' signals, due to the presence of hydrogen bonds with the carbonyl group of the acetyl moieties CH₃CO-1 and CH₃CO-1'', respectively. Furthermore, OH-4 and OH-4'' resonated at δ_H 9.22/9.24 (two singlets) and at δ_H 10.08 (singlet), respectively. It is important to note that the signals corresponding to the hydroxy groups were observed only in the ^1H NMR spectrum recorded at 0 °C. Overall, compound **1** could be proposed structurally as 1-[6-[1-[3-acetyl-2,6-dihydroxy-4-methoxy-5-(3-methylbut-2-en-1-yl)phenyl]-3-methylbutyl]-5,7-dihydroxy-2,2-dimethyl-2H-1-benzopyran-8-yl]-ethanone and was given the trivial name acropyrone.

Compound **2** was also obtained as yellowish optically inactive oil. Its APCI(+)-HRMS provided the molecular formula C₃₂H₄₂O₉ based on the pseudomolecular ion [M+Na]⁺ at m/z 593.2725 (calcd for 593.2721) and the typical fragment of *Acronychia*-type acetophenones at m/z 319.1907, implying 12 degrees of unsaturation (Figure A 4). UV absorption maxima in MeOH at 210, 229, 292, and 340 (sh) nm revealed the presence of a typical *Acronychia*-type acetophenone structure.

The ^1H NMR data of **2** (Table A 3) were closely related to those of **1**. Apart from the signals corresponding to the constant part of the molecule (ring B), an additional 3-hydroxy-2,2-dimethyl-3,4-dihydro-2H-pyrano ring was observed at ring A, characterized by a substitution pattern that is now being reported for the first time among *Acronychia*-type acetophenone derivatives. Specifically, in the ^1H NMR spectrum (Figure A 13), signals corresponding to a methylene group were observed at δ_H 2.92 as a doublet of doublets (J = 16.9/4.8 Hz, H-1'a) and at δ_H 2.67 as a multiplet (H-1'b). The signal of a methine group (H-2') resonated at δ_H 3.86 as a triplet with a coupling constant of 4.8 Hz also was observed. Finally, the protons of the H-4' and H-5' methyl groups were observed as two singlets at δ_H 1.48 and 1.53, respectively. The correlations in the HSQC spectrum of H-1'a/H-1'b and H-2' with the carbons at δ_C 25.7 and 68.7, respectively, confirmed the position of the additional OH group at C-2' (Figure A 15). The HMBC correlations of H-1'a/H-1'b (^3J) and H-2' (^2J) with a quaternary carbon at δ_C 80.6 led to the assignment of this carbon as C-3' (Figure A 16).

Carbon atoms C-4' and C-5' at δ_C 24.4 and 21.6, respectively, were defined through cross-peak correlations observed between them and the H-4' and H-5' methyl protons, in the HSQC spectrum. Furthermore, the positions of the methyl groups at C-4' and C-5' were determined from their HMBC correlations with C-2' (^3J) and C-3' (^2J). Similarly to **1**, the OH-6''' and OH-6 groups were observed significantly deshielded as two single peaks at δ_H 14.04 and 15.65 in the ^1H NMR spectrum at 0 °C (Figure A 18), while the OH-2', OH-4, and OH-4''' signals were observed at δ_H 8.37, 9.20, and 9.86 as broad singlet peaks, respectively. Finally, the 3-hydroxy-2,2-dimethyl-3,4-dihydro-2H-pyrano ring was demonstrated to be fused at the C-3/C-4 positions, as there was no NOE correlations of the protons of the methyl groups (H-4' and H-5') with the methyl protons of CH_3CO -1 that resonated at δ_H 2.70 (3H, s) (Figure A 14). Due to the small isolated quantity of **2**, the absolute configurations of C-2' and C-1'' could not be defined. However, useful NOE correlations of the protons of the methyl groups (H-4' and H-5') with H-1'a and H-1'b, respectively, determined their relative orientation. Compound **2** (acropyranol A) was identified therefore as 1-[8-[1-[3-acetyl-2,6-dihydroxy-4-methoxy-5-(3-methylbut-2-en-1-yl)phenyl]-3-methylbutyl]-3,5,7-trihydroxy-2,2-dimethyl-3,4-dihydro-2H-1-benzopyran-6-yl]ethanone.

Compound **3** was isolated as optically inactive yellowish oil, and its UV absorption maxima (MeOH), observed at 209, 231, 298, and 329 (sh) nm, were typical for an *Acronychia*-type acetophenone. Its molecular formula was determined as $C_{32}H_{42}O_9$, implying 12 degrees of unsaturation, based on the APCI(+)-HRMS data, such as its pseudomolecular ion $[M+H]^+$ at m/z 571.2897 (calcd for 571.2902) and the typical fragment of *Acronychia*-type acetophenones at m/z 319.1905 (Figure A 4). In exhibiting a different retention time at the HPLC-DAD chromatogram (Figure A 1) and the same mass spectrometric data as **2**, compound **3** was concluded to be a structural isomer of **2**.

Indeed, the NMR data (Table A 5) of **3** were found to be very closely related to those of **2**. As for compound **2**, all the signals corresponding to the typically substituted ring B were detected and determined from the NMR spectra (^1H NMR, HSQC, HMBC). Moreover, characteristic signals corresponding to the 3-hydroxy-2,2-dimethyl-3,4-dihydro-2H-pyrano ring were also observed. However, differences in the chemical shifts of specific protons (^1H NMR, Figure A 21) and carbons (HMBC, Figure A 24) were evident. In particular, the methylene protons of the additional ring resonated at δ_H 2.86 (1H, m, H-1'a) and 2.63 (1H, m, H-1'b) and correlated to a carbon signal at δ_C 26.3, while the protons of the methine H-2' appeared at δ_H 3.78 (1H, m) and correlated to a carbon at δ_C 68.7 (HSQC, Figure A 23).

Finally, the signals of the characteristic H-4' and H-5' methyl protons were observed at δ_H 1.35 (3H, s) and 1.38 (3H, s) and correlated with their corresponding carbon atoms at δ_C 24.9 and 21.8 in the HSQC spectrum. Characteristic also were the HMBC correlations of H-1'b (^3J), H-4' (^2J), and H-5' (^2J) with a quaternary carbon at δ_C 78.4, leading to the assignment of this carbon as C-3'. Similar to compound **1**, hydroxy group signals for OH-6 and OH-6''' were observed as two pairs of downfield shifted singlets at δ_H 15.59/15.72 and 16.05/16.22 in the ^1H NMR spectrum at 0 °C (Figure A 26), while signals for OH-4 and OH-4''' resonated at δ_H 9.24/9.29 and 10.07/10.15, respectively. It is important to note the difference in chemical shifts of the hydroxy groups between compounds **1**, **3**, and **2** (^1H NMR). Specifically, OH-6 and OH-6''' of compounds **1** and **3** appeared more shielded at δ_H 15.5–16.5 as two pairs of double peaks, while in compound **2** they are observed at δ_H 14.04 and 15.65 as singlets. This difference indicates the opposite fusion pattern of the additional ring of **1** and **3** when compared to **2**. Moreover, NOE correlations between protons H-4', H-5' of the methyl groups and the protons of the CH₃CO-1 group at δ_H 2.66 (3H, s) confirmed the position as well as the fusion of this ring at the C-2 and C-3 positions (Figure A 22). The latter correlation was absent in the NOESY spectrum of **2**, illustrating the different fusion profile. Similar to compound **2** and due to the small quantity of compound **3** isolated, the absolute configurations at C-2' and C-1'' were not deduced, but the NOE correlations observed between the H-4' and H-1'a protons as well as the H-5' and H-1'b facilitated the determination of their relative orientation. Thus, compound **3** (acropyranol B) was assigned as 1-[6-[1-[3-acetyl-2,6-dihydroxy-4-methoxy-5-(3-methylbut-2-en-1-yl)phenyl]-3-methylbutyl]-3,5,7-trihydroxy-2,2-dimethyl-3,4-dihydro-2H-1-benzopyran-8-yl]ethanone.

Compounds **4** and **5** were isolated as optically inactive *Acronychia*-type acetophenones possessing aliphatic chain on ring A of their basic structure. Particularly, compound **4** was identified as acrovestone (Table A 7, Figure A 29-Figure A 36) and **5** as acrovestenol (Table A 9, Figure A 37-Figure A 44) deduced from the UV, APCI(+)-HRMS (Figure A 4) and NMR spectra (^1H NMR, NOESY, HSQC, HMBC). Moreover, compounds **6** and **7** were isolated as optically inactive derivatives corresponding to acrofolione A (Table A 11, Figure A 45-Figure A 52) and acrofolione B (Table A 13, Figure A 53-Figure A 60) structures, respectively possessing an additional 2-(2-hydroxypropan-2-yl)-2,3-dihydro-1-furan ring fused at different positions on the parent structure. Compounds **6** and **7** have been reported previously in a phytochemical study of *Acronychia trifoliolata* (Oyama et al. 2003). However, their structural differentiation was not evident regarding the position of

the additional furan ring. More specifically, similarly to acropyranol A (**2**) and acropyranol B (**3**), compounds **6** and **7** are structural isomers differing only at the fusion position of the 2-(2-hydroxypropan-2-yl)-2,3-dihydro-1-furan ring occurring at either C-2/C-3 or C-4/C-3 of ring A. According to the present study, the position of this ring could be deduced by the correlation of the protons of the methyl groups of the 2-(2-hydroxypropan-2-yl)-2,3-dihydro-1-furan ring with those of the acetyl group of ring A, as observed in the NOESY spectrum. In particular, a NOE correlation was observed between H-4' and H-5' protons of methyl groups at δ_H 2.51 (3H, s) and 2.25 (3H, s) and the protons of the acetyl group at δ_H 2.64 (3H, s, CH₃CO-1) of ring A in the spectrum of **7** (Figure A 54); this correlation was not observed in the analogous spectrum of **6** (Figure A 46). The same difference was evident from the NOESY spectra of compounds **2** and **3**, indicating the position of the hydroxypyran ring (Figure A 14, Figure A 22). Thus, this characteristic cross-peak correlation may be used as a diagnostic signal not only for the specific pair of isomers but generally for the determination of the position of the additional ring in *Acronychia*-type acetophenone derivatives (Figure 4).

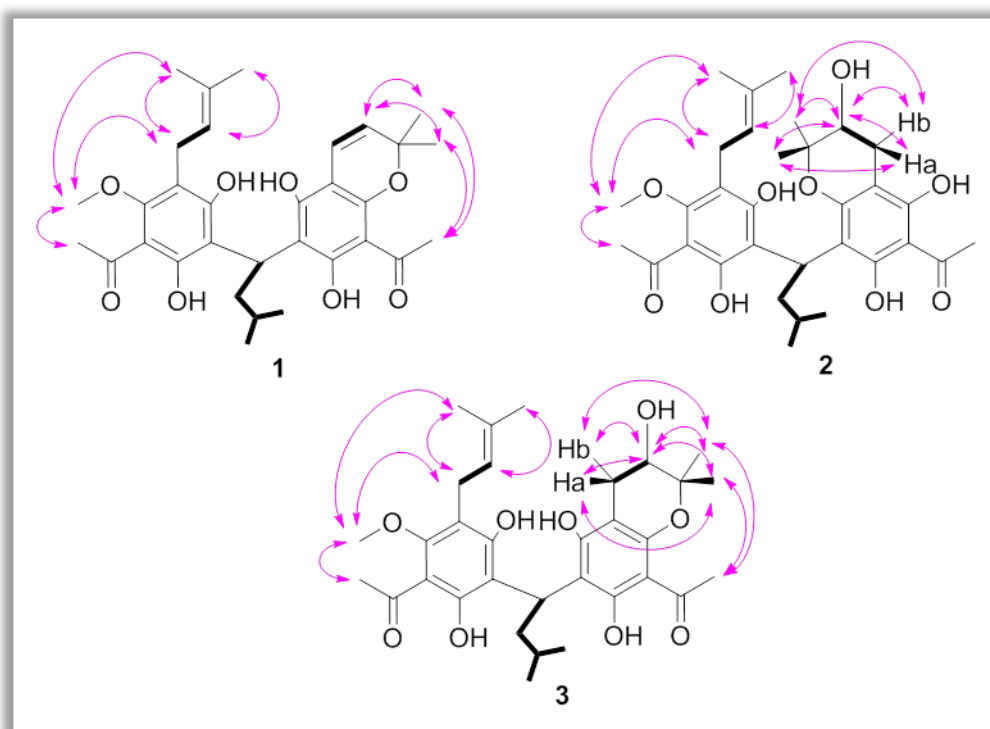


Figure 4: COSY (bold lines) and key NOE correlations (arrows) for compounds 1–3

3. Conformational analysis of acetophenone dimers

Due to their abovementioned structural characteristics, *Acronychia*-type acetophenones presented complicated NMR spectra at room temperature which hampered the structural elucidation of these molecules. In particular, concerning ^1H NMR spectra broad, unresolved or slightly resolved peaks and multiple signals corresponding to the same protons were observed suggesting the presence of rotational conformers (rotamers) in CDCl_3 solution. The duplication of several ^1H NMR signals was attributed to the rotation about specific C-C bonds. In order to investigate these phenomena and facilitate the structure elucidation procedure for this kind of molecules, dynamic NMR experiments in variable temperatures were performed along with molecular mechanics calculations to prove the hypothesis that conformational isomers are interchanged via rotation.

Initially, acrovestone (**4**) the model compound of this category was investigated. Variable temperature ^1H NMR experiments revealed changes in the signals of protons belonging to the isopentyl and the isoprenyl chains indicating a possible rotation of the molecule about $\text{C1}''\text{-C5}'''$ and $\text{C3-C1}' / \text{C3}'''\text{-C1}''''$ (Figure 5).

Specifically, signals of all isopentyl chain protons were affected from the variable temperatures applied during the experiment. At room temperature, the multiplicity of peaks could not be resolved and the majority of signals were observed as broad peaks. By cooling down the solution at $0\text{ }^\circ\text{C}$ two principal rotamers seemed to appear. Characteristic were the signals of $\text{H-1}''$ and $\text{H-2}''$ resonating at approximately δ_{H} 4.75 and 2.20, respectively, which were duplicated, indicating the presence of two rotational conformers in a ratio of 1:1. The same behaviour was observed for the signals attributed to $\text{H-4}''$ and $\text{H-5}''$ methyl protons at δ_{H} 0.88 which were detected as two doublets of 6.3 Hz coupling constant. In general, at low temperatures, the two individual groups of chemically equivalent nuclei exchanged by the intramolecular process of rotation were detected due to their slow interconversion in the NMR time scale. Increasing the temperature from 0 to $52\text{ }^\circ\text{C}$ in CDCl_3 solution single, well resolved peaks were observed thus suggesting the fast interconversion between the two rotamers in the NMR time scale and the presence of averaged ^1H NMR signals. Particularly, $\text{H-1}''$ was resonated as one triplet at δ_{H} 4.75 of 7.3 Hz coupling constant and $\text{H-4}''$ and $\text{H-5}''$ were detected at δ_{H} 0.88 as a doublet of 6.6 Hz. Moreover, signals corresponding to methylene $\text{H-2}''$, were observed as two broad peaks at $52\text{ }^\circ\text{C}$ suggesting a slower interconversion rate for $\text{H-2}''$ at that temperature.

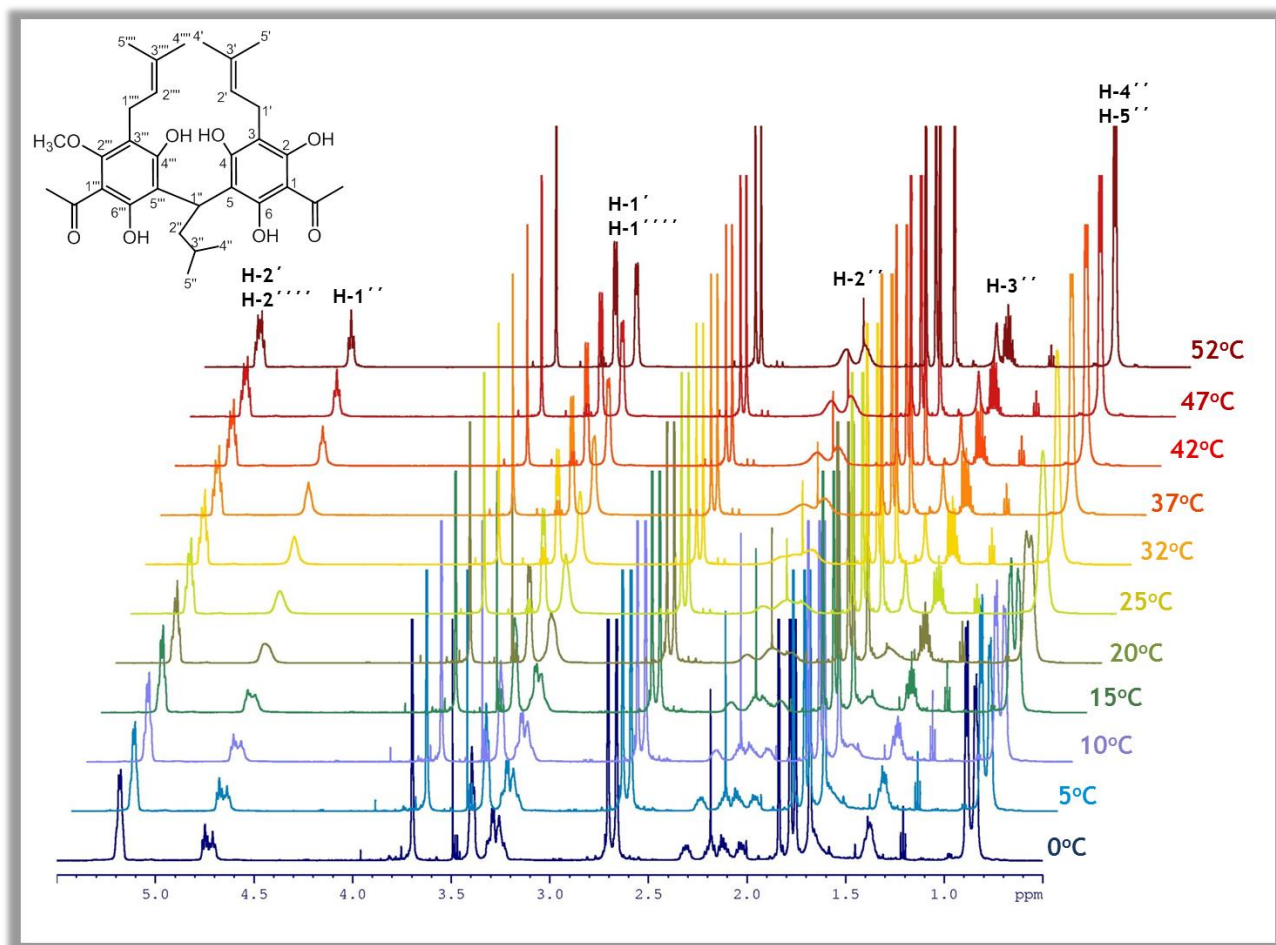


Figure 5: Variable temperature ^1H NMR spectra of acrovstone (4) in CDCl_3 , δ_{H} 0-6

Furthermore, rotation of the isoprenyl chains about $\text{C3-C1}'$ and $\text{C3}'''\text{-C1}''''$ was suggested by the respective protons as detected at 0 and 52 °C (Figure 5). As a matter of fact, the signals corresponding to $\text{H-1}'$, $\text{H-1}'''$ at δ_{H} 3.42 and 3.31, respectively, observed as two doublets at 52 °C, were splitted into a number of doublets at 0 °C. Concerning $\text{H-2}'$ and $\text{H-2}'''$, a faster interconversion rate was deduced from the well resolved triplet of triplets at δ_{H} 5.23 and 5.21, respectively, at 52 °C which were coalesced and broadened moving towards 0°C.

Similarly, hydroxyl protons of the molecule presented interesting behaviour in variable temperature NMR experiments (Figure 6). At low temperatures, duplicated, well resolved peaks were detected corresponding to hydroxyl protons facilitating their characterization. In case of hydroxyl protons significant differences in terms of chemical resonances between the signals corresponding to the same protons were observed (from 70-510 Hz) indicating a slow exchange. Therefore, the coalescence of the majority of hydroxyl signals (apart from

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

OH-6''') was observed at high temperatures ($\sim 40^\circ\text{C}$) comparing with the signals corresponding to the protons of the isoprenyl or isoprenyl chains ($\sim 20^\circ\text{C}$).

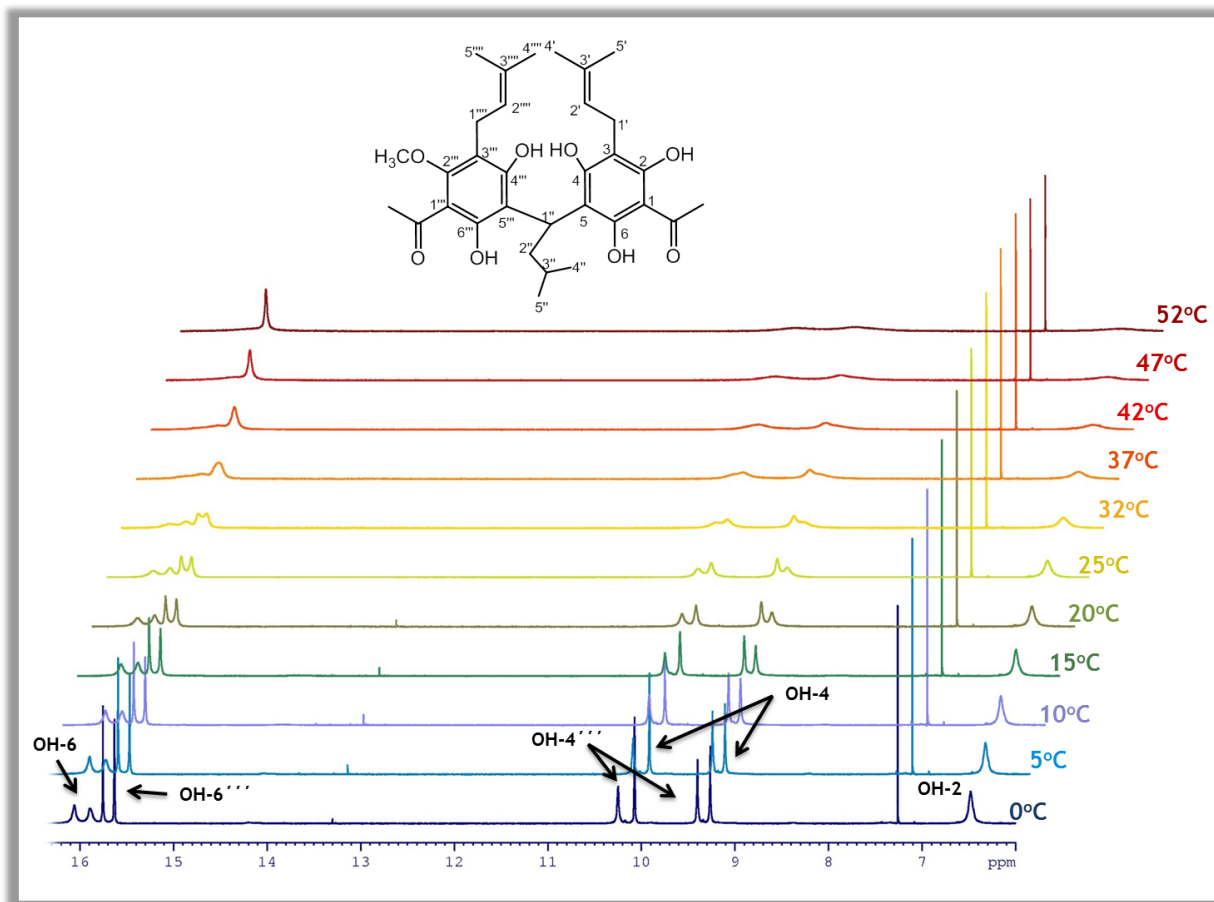


Figure 6: Variable temperature ^1H NMR spectra of acrovestone (**4**) in CDCl_3 , δ_{H} 6-16

A convenient method for the determination of rotamers is by Nuclear Overhauser Effect Spectroscopy (NOESY). During this experiment, the protons under chemical exchange showed a negative cross peak and thus indicated the presence of rotamers. In particular, NOESY spectrum of acrovestone (**4**) at 0°C revealed interesting rotational exchange peaks which were used for the assignment of the signals corresponding to the different individual rotamers (Figure 7, Figure A 35). Obviously, these peaks were absent in the NOESY spectrum at 47°C since signals corresponding to only one conformer were detected (Figure A 30). Unfortunately, the distance between the isoprenyl chains (significantly greater than $5\text{ \AA}$) did not permit the generation of diagnostic NOE cross peaks for the determination of the relative orientation of the isoprenyl chains on the molecule of the individual rotamers.

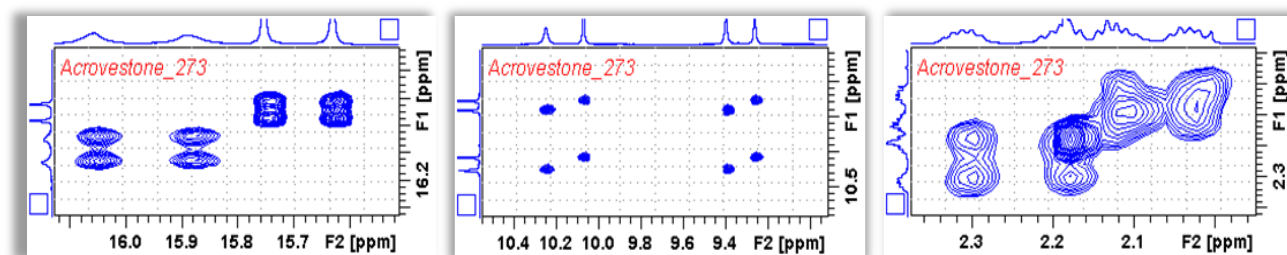


Figure 7: NOESY snapshots at 0 °C of representative rotational exchange peaks

The assignment of the different rotamers was performed at 0 °C using the HMBC spectrum (Figure A 36) and specifically three key carbon atoms that were observed in slightly diverse chemical resonances while all the rest carbon atoms were identical for both rotational conformers. In details, C-2'' was observed at δ_C 39.3 and 39.5 for the two different rotamers and correlated with H-1'' at δ_H 4.71 and 4.74 respectively. Moreover, the cross peaks between C-5 (δ_C 108.4 or 109.0) and H-2'' and OH-4 as well as C-5''' (δ_C 113.4 or 113.0) and H-2'', OH-4''' and OH-6''' unravelled the individual rotamer signals.

Conformational isomers exist in a dynamic equilibrium, where the relative free energies of isomers determine the population of each isomer and the energy barrier of rotation determines the rate of interconversion between isomers (Morris et al. 1996). In order to estimate the barriers of rotation about C1''-C5''' the rate of exchange (k_e) between the two states was calculated using the Gutowski-Holm equation (Equation 1) (Gutowsky et al. 1956). Moreover, the free energy of activation for the rotation ($\Delta G^\ddagger$) was calculated at the coalescence temperature using equation 2 (Akhmedov et al. 2004):

$$k_e = 2.22 \Delta\nu \quad (\text{Equation 1})$$

$$\Delta G^\ddagger = 4.576 T_c [(10.32 + \log(T_c/k_e))] \quad (\text{Equation 2})$$

where $\Delta\nu$ is the frequency difference between the two signals of the different conformers corresponding to the same proton and T_c is the coalescent temperature for a specific signal. In ^1H NMR spectrum (Figure 5, 298 K) of acrovestone, the chemical shift difference for the two resonance signals of H-1'' is 52 Hz, corresponding to $k_e = 294 \text{ s}^{-1}$ and $\Delta G^\ddagger 14.9 \text{ kcal mol}^{-1}$ at T_c 294 K. Correspondingly, $\Delta G^\ddagger$ for OH-6''' was calculated for $15.1 \text{ kcal mol}^{-1}$ at T_c 311 K. Therefore, a rotation barrier of $\sim 15 \text{ kcal mol}^{-1}$ was proposed for acrovestone rotamers.

In parallel, the assessment of the different rotamers was performed by molecular mechanics calculation using MacroModel software in order to determine the more

energetically stable conformers of acrovestone. Calculations afforded two distinct conformational states corresponding to two energy minima (Figure 8). The relative energy of the second conformer (Figure 8, green molecule) with respect to the global minimum (Figure 8, violet molecule) was $0.11 \text{ kcal mol}^{-1}$. Furthermore, their populations were estimated at 273K using Boltzmann distribution, revealing a ratio 1:1 between the two conformers. These findings demonstrated a good agreement between the molecular mechanics simulations and experimental NMR data.

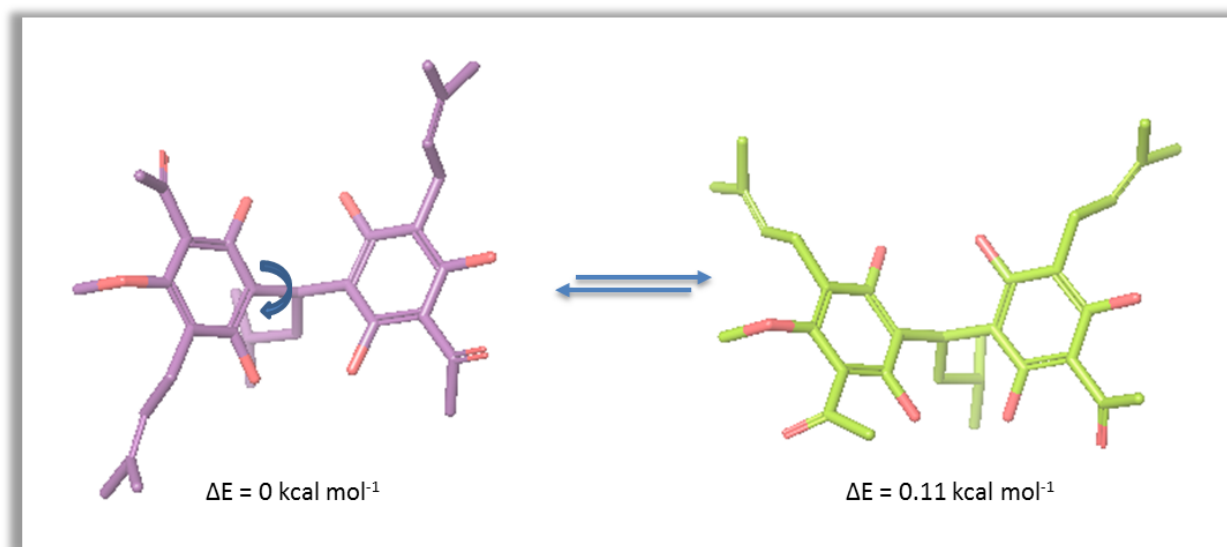


Figure 8: Major rotamers of acrovestone observed in NMR data at 273 K and confirmed by molecular mechanics calculations; global minimum structure is presented in violet and the second more stable structure is presented in green

Moreover, as aforementioned, signals suggesting isoprenyl group rotation were observed in variable temperature ^1H NMR experiments (Figure 5). Therefore, in order to visualize possible rotation about C3-C1' and C3'''-C1''', superposition of the energetically more favored conformers calculated by molecular mechanics simulations was performed separately for molecules with different isoprenyl orientations. Indeed, different proposed conformations for isoprenyl chains derived from calculations confirmed the initial hypothesis (Figure 9).

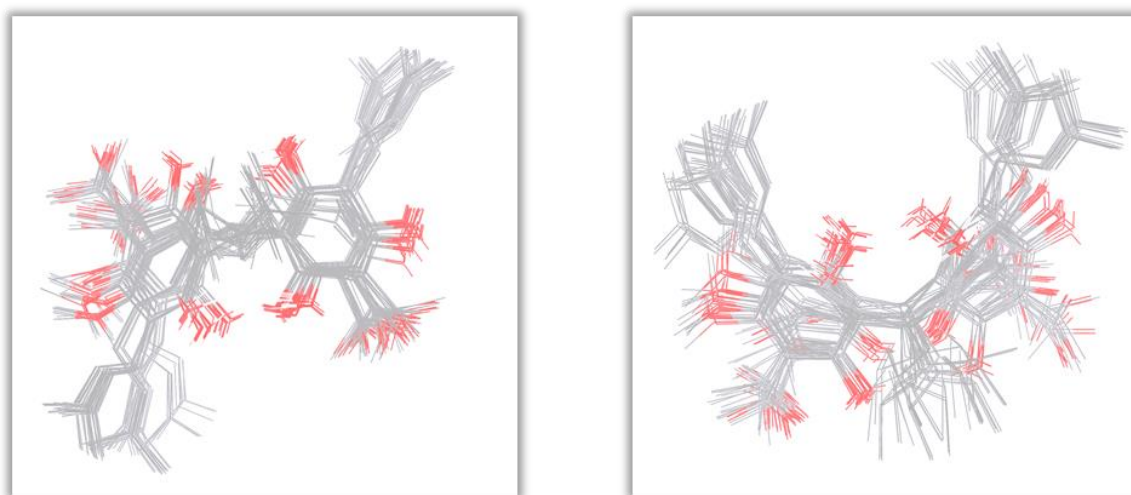


Figure 9: Superimposed energetically more favored conformers calculated by molecular mechanic simulations are separated according to the different isoprenyl orientations into two clusters

Likewise, NMR spectra of all isolated *Acronychia*-type acetophenones recorded at 0 °C revealed the presence of rotational isomers. Duplicated peaks concerning mainly signals corresponding to the protons of the isopentyl chain and the additional ring as well as hydroxy protons were assigned to the individual rotamers (Table A 2/ 4/ 6/ 8/ 10/ 12/ 14). Towards this effort, key carbon signals (C-5, C5'') were essential for the attribution of the duplicated peaks to the respective rotational conformers. In parallel, Boltzmann populations of all dominant conformers substantiated the occurrence of rotational conformers at 0 °C. However, in the case of acropyranol A (**2**) and acrofolione A (**6**) a different behavior was observed. Despite the detection of multiple hydroxyl protons at 0 °C, the rest of the signals in the ¹H NMR spectrum were observed similar to the respective ones observed at 47 °C suggesting a limited rotation of these molecules. This finding was also confirmed by molecular mechanics calculation. The conformation of the additional ring presenting the same orientation with the isoprenyl chain was favoured in both compounds. A conformation resulting from the rotation of molecules **2** and **6** about C1''-C5''' presented a relative energy difference of 7.62 and 7.74 kcal mol⁻¹, respectively, corresponding to populations smaller than 3%. Therefore, the rotation of acropyranol A (**2**) and acrofolione A (**6**) about C1''-C5''' seems to be almost absent while limited rotations of the isoprenyl chain and isopentyl chain were observed explaining the experimental ¹H NMR spectroscopic data (Figure 10).

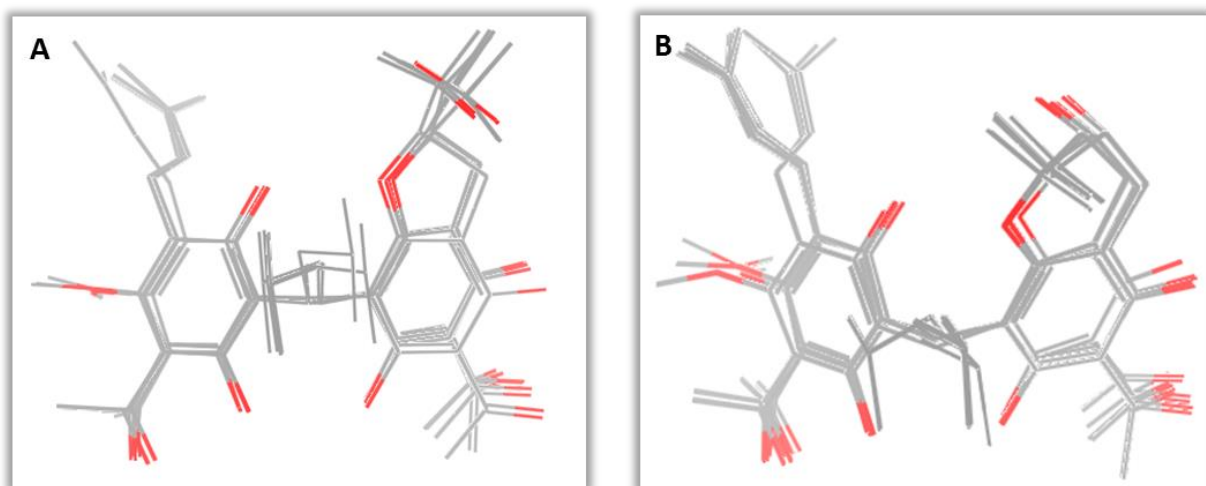


Figure 10: Energetically more favored conformers of A. Acrofolione A and B. Acropyranol A as calculated by molecular mechanics simulations

4. Enantiomer resolution screening of *Acronychia*-type acetophenones

Acronychia-type acetophenones purified by FCPC and semi-preparative HPLC were isolated as optically inactive racemic mixtures. These molecules are comprised of two aromatic rings connected with an isopentyl chain leading to the formation of a chiral carbon (C-1''). Moreover, some derivatives possessing an additional ring on aromatic ring A of the core structure presented a second chiral centre supporting the potential occurrence of diastereomers. The resolution of different diastereomer compounds was not obtained in reversed phase chromatography. However, in the ^1H NMR spectra of acropyranol A (**2**) and acrofolione A (**6**), at 47 °C, the presence of non-equivalent signals attributed to the protons of the additional ring denoted the presence of diastereomers (Table A 3, Table A 11). Therefore, the development of a methodology for the enantiomer and diastereomer separation was compulsory.

The application of chiral stationary phases was an essential prerequisite to achieve enantiomer separation. In this context, the immobilized polysaccharide-derived Chiral Stationary Phases (CSPs) were applied since they provide a number of advantages comparing to other CSPs. In particular, different chiral materials immobilized on 5 μm silica gel are designed for extended solvent compatibility (comparing to non-immobilized) offering broad chiral recognition abilities, high chromatographic efficiency and excellent reproducibility (Zhang et al. 2008). The screening strategy for enantiomer resolution

involved the screening of all derivatives using four different immobilized chiral columns and mixtures of different organic solvents. The four different columns were composed of immobilized tris(3,5-dimethylphenylcarbamate) of amylose (CHIRALPAK IA), tris(3,5-dimethylphenylcarbamate) of cellulose (CHIRALPAK IB), tris(3,5-dichlorophenylcarbamate) of cellulose (CHIRALPAK IC) and tris(3-chlorophenylcarbamate) of amylose (CHIRALPAK ID). The structural and stereochemical differences in these polymeric chiral selectors are estimated to recognise specific classes of solutes, sometimes with a common overlapping area. Therefore, the screening of the different immobilized materials is highly recommended. As far as the mobile phase is concerned, immobilized columns give the possibility to apply organic solvents of various natures. In practice, the selection of the mobile phase is based mainly on the chiral recognition criteria, but also on considerations concerning the physical and chemical natures of the solute (solubility, hydrophobicity, chemical and stereochemical stability, etc.).

The method development in such cases is performed based on a screening approach of different organic solvent mixtures and then optimization of the solvent ratio according to the compound behaviour. Briefly, according to Zhang et al. alkane–lprOH mixtures, alkane–EtOH mixtures and alkane–MtBE–EtOH mixtures were tested initially. In a second step, alkane–EtOAc mixtures and finally the addition of an acidic additive were also attempted. The abovementioned method development strategy gave rise to a number of systems assayed that are summarized in Table II.

Table II: Isocratic elution systems tested during method development for enantioselective separation

Solvent mixtures	Hex / IPrOH	Columns	Hex/Et OH	Columns	MTBE/Et OH	Columns	Hex/EtO Ac	Columns
Starting condition	80:20	IA, IB, IC, ID	80:20	IA, IB, IC, ID	98:2	IA, IB, IC, ID	70:30	IA, IB, IC, ID
<i>Optimization</i>	99:1	IC, ID	99:1	ID	90:10	IB, ID	90:10	IA, IC, ID
<i>Optimization</i>	95:5	IC, ID	97:3	IA, IB, IC, ID	85:15	IC, ID	80:20	IA
<i>Optimization</i>	90:10	IC, ID	95:5	IA, IB, IC, ID	80:20	IA, IB, IC, ID	60:40	IB
<i>Optimization</i>	85:15	IA	90:10	ID				
<i>Optimization</i>	75:25	IB						
<i>Optimization</i>	60:40	IB						

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

Solvent mixtures	Hex/MTBE/ EtOH	Columns	Hex/MTBE/ EtOH+0.5%F.A	Columns	Hex/EtOH+0.5%F.A	Columns
Optimization	90:9:1	IA, ID	93:5:2	IA	96:4	IA
Optimization	90:8:2	IB	94:4:2	IA, ID	97:3	IA
Optimization	91:7:2	IA	96:2:2	IA	98:2	IA
Optimization	92:6:2	IA	80:18:2	ID	99:1	
Optimization	91:6:3	IA	85:13:2	ID		
Optimization	50:45:5	IB				

As concluded from the method development table, the enantiomer/ diastereomer resolution of *Acronychia*-type acetophenones was a painful task. The majority of the solvent mixtures did not achieve any enantiomer/ diastereomer separation. Uniquely, mixtures of Hex with small percentage of EtOH+0.1% F.A using CHIRALPAK IA led to a successful characterization of the isolated racemic mixtures (Figure 11). The enantioselective separation of *Acronychia*-type acetophenones using CHIRALPAK IA and a mobile phase consisting of Hex/ EtOH+0.1% F.A (95/5- 99/1) confirmed the presence of enantiomers and diastereomers of the different derivatives.

Specifically, concerning acrovestone (**4**) the two enantiomers were observed using Hex/ EtOH+0.1% F.A (99/1); however, baseline resolution could not be obtained. Enantiomers of acropyrone (**1**) were not resolved probably due to its high hydrophobicity leading to a retention time of 4.2 min using Hex/ EtOH+0.1% F.A (99/1) as mobile phase. The occurrence of one chiral centre in latter optically inactive compounds implied the presence of two enantiomers while the rest compounds possessing two chiral carbons were suspected for the presence of either two enantiomers or four isomers. Poor resolution of acrovestenol (**5**) isomers observed using Hex/ EtOH+0.1% F.A (98/2) suggested the presence of two enantiomers which was in good accordance with the NMR data indicating absence of non-equivalent signals and thus absence of diastereomers. The chiral chromatographic separation of Acropyranol A (**2**) revealed the presence of two enantiomers, however, the NMR data denoted the presence of diastereomers which were possibly co-eluted. The presence of four Acropyranol B (**3**), Acrofolione A (**6**) and Acrofolione A (**6**) isomers was observed during the analysis of these compounds. The findings concerning Acrofolione A (**6**) were in total agreement with the NMR data however, concerning diastereomers of Acropyranol B (**3**) and Acrofolione B (**7**) probably the close chemical resonances of their NMR signals did not lead to their observation by NMR.

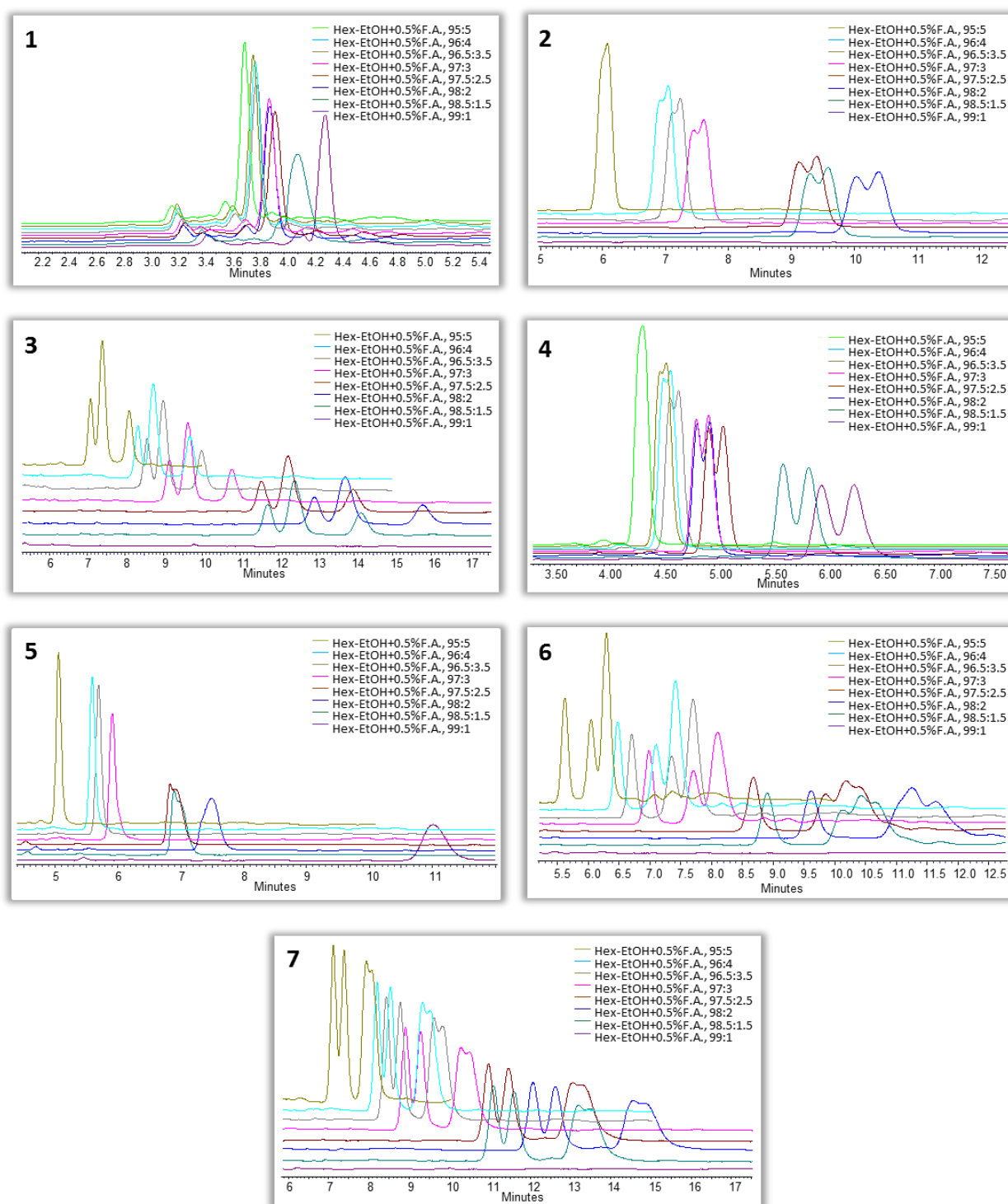


Figure 11: HPLC chromatograms at 280 nm obtained from the analysis of *Acronychia*-type acetophenones using CHIRALPAK IA CSP; acropyrone (1), acropyranol A (2), acropyranol B (3), acrovestone (4), acrovestenol (5), acrofolione A (6) and acrofolione B (7)

Overall, concerning *Acronychia*-type acetophenones the determination of diastereomers, enantiomers and rotamers was a complicated task due to structural

peculiarities of these compounds. In such cases, in order to tackle these issues combination of different spectroscopic, chromatographic and computational techniques need to be applied and combined to get a better insight.

5. Isolation of *Acronychia*-type acetophenone diastereomers by SFC

The determination of diastereomer presence in some *Acronychia*-type acetophenone samples led to the development of a strategy for the separation of mixtures of diastereomers. Among the available techniques, achiral supercritical fluid chromatography (SFC) has been proven the most efficient for diastereomeric resolution (Ebinger et al. 2013). The low viscosity and the high diffusivity of the mobile phase in SFC allows higher flow rates and lower pressure drops comparing to HPLC and thus, significant improvement in terms of speed and efficiency is obtained (Pinkston et al. 2006). Therefore, a SFC method development protocol was set for the screening of different stationary phases and operating conditions in analytical scale with an ultimate goal the scaling of the method in semipreparative conditions for the isolation of adequate quantities of diastereomers. Briefly, a typical method development procedure in SFC includes primary the screening of different stationary phases, then the screening of diverse co-solvents and the alteration of temperature and pressure conditions for the improvement of the peak resolution. Moreover, additives can be finally utilized for further improvement of peak shape. The transfer of the method in semi-preparative and preparative SFC conditions usually improves drastically the resolution of the peaks since the same particle size is applied with a considerably increased flow rate.

In the current study, two different achiral stationary phases were tested (Silica and 2-ethyl pyridine) in combination with three diverse co-solvents (MeCN, MeOH, IPrOH) applying a standard gradient elution program of 10 min consisting of 1min conditioning with 5% co-solvent, 6 min of gradient elution up to 40% of co-solvent, 1 min of 40% co-solvent, return back to initial conditions in 1 min and 1 min of 5% co-solvent re-equilibration. The flow rate was set at 5 mL/min allowing fast and effective chromatographic performance. Moreover, increase in oven temperature and pressure conditions was tested to assess the impact on the chromatographic behaviour. The aforementioned conditions resulted to the screening of acrofolione A (**6**) diastereomers using ten different methods in analytical scale (Table III).

Table III: Ten different SFC methods assayed during method development

No	Column	Solvent	Temp. (°C)	Pres. (bar)
1	2EthSil	MeCN	40	125
2	Sil	MeCN	40	125
3	2EthSil	MeOH	40	125
4	Sil	MeOH	40	125
5	2EthSil	lprOH	40	125
6	Sil	lprOH	40	125
7	2EthSil	MeOH	60	125
8	2EthSil	lprOH	60	125
9	2EthSil	MeOH	40	150
10	2EthSil	MeOH	40	150

The obtained chromatograms are illustrated in **Figure 12** representing the impact of different stationary phases and chromatographic conditions on chromatographic behaviour. In particular, 2-ethyl pyridine silica column had a significantly superior selectivity for acrofolione A (**6**) than silica column. Among the three co-solvents tested, MeOH was preferred as sharper peaks in parallel with slight diastereomer separation were obtained. Elevated temperature and pressure conditions led to co-elution of diastereomers. Therefore, method 3 was selected for scaling up to semi-preparative separation conditions.

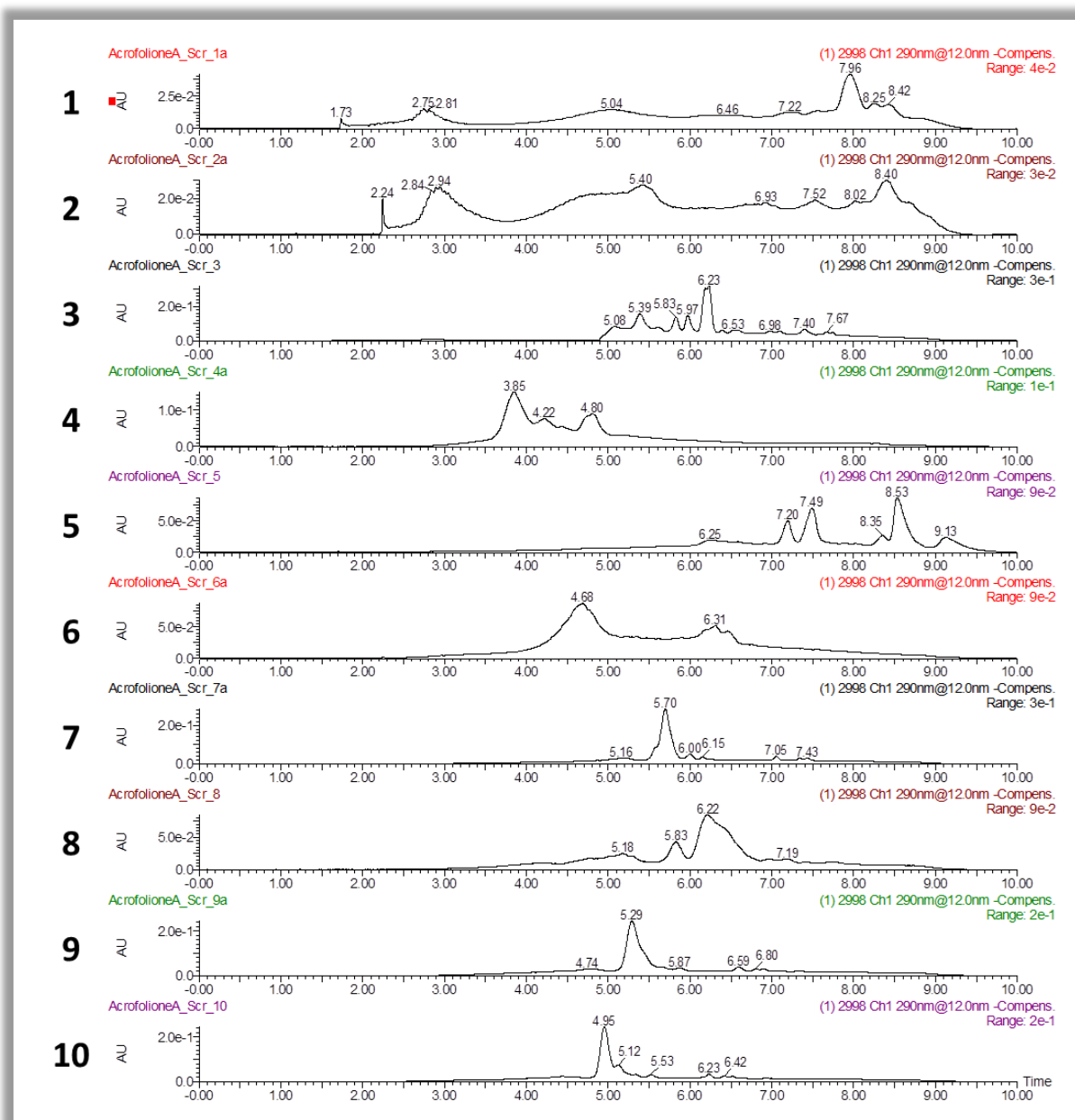


Figure 12: SFC chromatograms at 290 nm obtained from the method development procedure for the diastereomer separation of Acrofolione A (6)

In semi-preparative scale the same stationary phase was applied for the separation of acrofolione A diastereomers. **Figure 13** illustrates the chromatogram obtained by injection of 1 mg of mixture of diastereomers in semipreparative column. The semipreparative separation scaled perfectly, although the analytical conditions were not optimized to maximize throughput. The resolution of diastereomers was considerably improved comparing to analytical scale allowing collection of high purity compounds (**Figure 13**). However, the loading of the column with greater quantity was prohibitive due to the

obtained resolution. Therefore, ten consecutive injections were performed. Diastereomers 1 and 2 were collected using fraction collector, concentrated and analysed by NMR (47 and 0 °C). The separation produced two fractions with high diastereomeric purity.

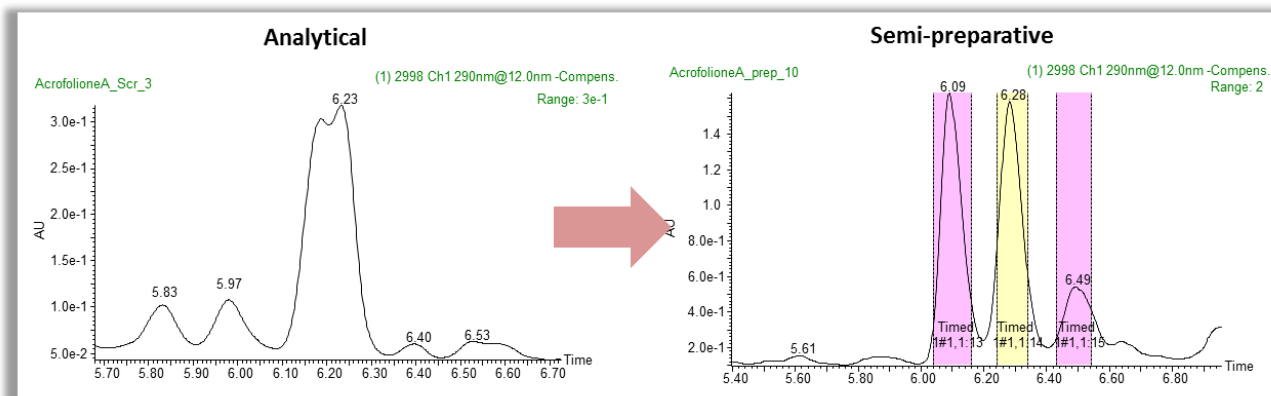


Figure 13: Analytical and semi-preparative chromatograms obtained by the SFC analysis of acrofolione A

Comparative examination of ^1H NMR data of the different diastereomers 1 and 2 revealed small differences in chemical shifts for the majority of protons (Figure 14). Shifting was mainly observed in protons belonging to the additional 2-(2-hydroxypropan-2-yl)-2,3-dihydro-1-furan ring and H-2''' of the isoprenyl chain. Moreover, the relative configuration of the different diastereomers led to slightly altered J -couplings representing the different distance and dihedral angles between specific protons. However, the relative orientation of H-1'' and H-2' could not be established as the distance between the two protons did not permit the generation of NOE cross peaks. Furthermore, as expected HMBC and NOE spectra of the two diastereomers presented similar cross peaks (Table A 15).

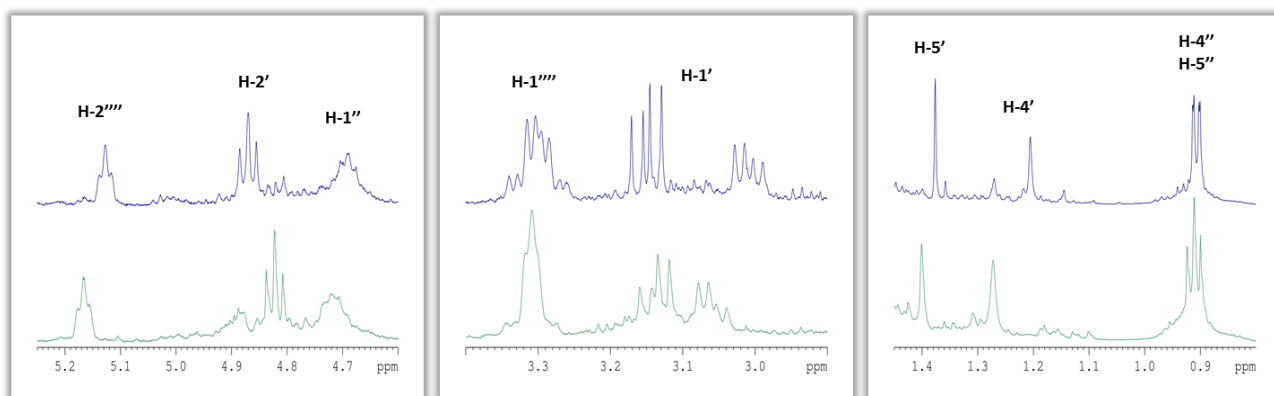


Figure 14: Different regions of ^1H NMR spectra obtained from acrofolione A₁ (green) and acrofolione A₂ (blue) at 47 °C

6. Antibacterial activity of *Acronychia* type acetophenones

The traditional use of *Acronychia* genus as antimicrobial agent and for the treatment of skin infections and diseases related to respiratory system implied the potential existence of antibacterial agents. Therefore, two representative compounds acrovestone (**4**) and acrofolione A (**6**) were tested for their antibacterial activity against different *Staphylococcus aureus* strains resistant in various antibiotics, *Bacillus subtilis*, *Streptococcus pneumoniae*, *Escherichia coli*, *Klebsiella pneumoniae*, *Proteus*, *Pseudomonas aeruginosa*, *Salmonella typhi* bacteria strains. Acrovestone demonstrated significant antibacterial activity against all *Staphylococcus aureus* strains and *Bacillus subtilis*. The minimum inhibitory concentrations (MICs) were calculated for the two *Acronychia*-type acetophenones and compared to the respective ones of the positive control norfloxacin (Table IV).

Table IV: Minimum inhibitory concentrations (MICs) of acrovestone and acrofolione A comparing to the positive control norfloxacin

Name of Organism	Type	Acrovestone (mg/L)	Acrofolione A (mg/L)	Norfloxacin (mg/L)
<i>S. aureus</i> (XU212)	Gram(+)	2	32	16
<i>S. aureus</i> (SA-1199B)	Gram(+)	2	64	32
<i>S. aureus</i> (RN-4220)	Gram(+)	0.5	32	0.5
<i>S. aureus</i> (EMRSA-15)	Gram(+)	2	64	0.5
<i>S. aureus</i> (EMRSA-16)	Gram(+)	4	64	0.25
<i>S. aureus</i> (NCTC-8532)	Gram(+)	1	16	0.5
<i>B. subtilis</i> (Z11)	Gram(+)	1	64	0.25
<i>S. pneumoniae</i>	Gram(+)	>128	>128	2
<i>E. coli</i> (O127:H6)	Gram(-)	>128	>128	0.25
<i>K. pneumoniae</i> (NCTC-9633)	Gram(-)	>128	>128	0.25
<i>Proteus</i>	Gram(-)	>128	>128	>128
<i>P. aeruginosa</i>	Gram(-)	>128	>128	>128
<i>S. typhi</i> (SL 1344)	Gram(-)	>128	>128	1

In particular, concerning XU212 (resistant to tetracycline) and SA-1199B (resistant to ciprofloxacin) strains, acrovestone showed 8 and 16 fold decreases in MIC, respectively, comparing to the positive control norfloxacin. These findings are of great importance suggesting acrovestone as a new extremely potent antibacterial agent. Acrofolione A found

to exhibit moderate antibacterial activity against *S. aureus* strains and *B. subtilis* while both compounds were found inactive against gram(-) bacteria that were assayed.

7. Cytotoxic activity of *Acronychia* type acetophenones against human tumour cell lines

The cytotoxicity of the isolated *Acronychia*-type acetophenones (**1–7**) was examined using an MTS assay against human DU145 prostate cancer and A2058 melanoma cells. Cell viability (%) was normalized to the vehicle control (Table A 16). All acetophenones tested inhibited differentially cell viabilities for both cell lines. Compounds **1–3** and **7** displayed relatively weak cytotoxicity, while compounds **4–6** showed substantial cytotoxicity for both cell lines, as shown in Table V. Interestingly, compounds **4** and **5** exhibited the most potent activity, with IC₅₀ values of 0.38 and 2.8 μ M, among the compounds tested against A2058 melanoma cells (Table V). Among the *Acronychia*-type acetophenones, compounds **4** and **5** were the most effective for DU145 cells, with IC₅₀ values of 0.93 and 2.7 μ M, respectively. All compounds were also assayed for their cytotoxicity against normal human dermal fibroblast (NHDF) cell line (Table V). They were found to be inactive and exhibited IC₅₀ values of >5 μ M, thereby suggesting these compounds may be selective to tumor cells. These data suggest also that the presence of a short aliphatic, hydrophobic chain such as an isoprenyl (**4**) or modified isoprenyl moiety (**5**) at the C-3 position of the ring A enhances cytotoxicity against both the tumor cell lines used. In contrast, the presence of an additional ring seems to reduce the cytotoxicity of the acetophenone dimers (**1**, **2**, **3**, and **7**) investigated. Moreover, the presence of an additional ring fused at the C-3 and C-4 positions (ring A) seems to enhance the cytotoxic activity compared to their isomers having the additional ring fused at the C-2 and C-3 positions (**2** vs **3** and **6** vs **7**). It is worth noting the different activity profile of the two isomers **6** and **7** against the two cancer cell lines used. This difference indicates that the relative position of the additional ring on the basic *Acronychia*-type acetophenone skeleton contributes significantly to the cytotoxic activity.

Table V: IC₅₀ Values of Compounds **4–6** against Two Human Tumor and a Normal Cell Lines; ^a Cells were treated with compounds in a dose-dependent manner for 48 h, then, the MTS assay was used to assess cell viability ^bData are expressed as means $\pm$ SD

compound	IC ₅₀ value (μM) ^a		
	DU145	A2058	NHDF
4	0.93 ± 0.07 ^b	0.38 ± 0.04	> 5.0
5	2.7 ± 0.5	2.8 ± 0.3	> 5.0
6	> 5.0	4.2 ± 0.6	> 5.0
sorafenib	5.1 ± 0.7	3.8 ± 0.9	

8. Anti-inflammatory activity of *Acronychia* type acetophenones

Acronychia genus has been mainly utilized in traditional medicine for the treatment of asthma, cough and rheumatism designating an important anti-inflammatory activity (Rahmani et al. 1996). Until today the anti-inflammatory properties of *Acronychia* constituents and specifically *Acronychia*-type acetophenones have not been assessed with the exception of a publication referring to the moderate COX-2 inhibitory effect of acrovestone (**4**) and acrovestenol (**5**) (Pathmasiri et al. 2005). In this context, the estimation of the anti-inflammatory activity of all *Acronychia* type acetophenones isolated during this study was performed by *in vitro* assessment of their inhibitory effects on 5-lipoxygenase (5-LO) and microsomal prostaglandin E₂ synthase-1 (mPGES-1).

5-LO is a key enzyme in the biosynthesis of leukotrienes (LTs) catalysing the initial transformation of arachidonic acid (AA). LTs are pivotal pro-inflammatory mediators associated strongly with the occurrence of asthma and other inflammation related diseases such as chronic inflammation, atherosclerosis and tumorigenesis (Rådmark et al. 2007). Currently, leukotriene antagonists are used in the treatment of asthma demonstrating the importance of this pharmacological target. 5-LO is an iron containing redox active enzyme and its natural substrate is AA. Generally, 5-LO inhibitors are categorized according to the inhibitory modes of action into three main groups. The first involves redox-active 5-LO inhibitors, acting by reducing the 5-LO active site iron from the active ferric state to the inactive ferrous state. The second concerns iron ligand inhibitors which have been reported to chelate the active site iron. The third group are non redox-type inhibitors that compete with AA for binding to 5-LO, lacking redox properties (Pergola et al. 2010). Recent developments in anti-inflammatory compounds promote those with dual action, possessing higher anti-inflammatory efficacy accompanied by reduced number and severity of side effects. In this context, initially, 5-LO in combination with cyclooxygenase (COX) inhibition

was targeted leading to greater efficiency and lower gastric toxicity (Laufer et al. 1994). More recently, the combination of 5-LO with microsomal prostaglandin E_2 synthase-1 (mPGES-1) inhibition was proposed as a promising category of anti-inflammatory agents that hamper cardiovascular toxicity observed in the previous case (Koeberle et al. 2008).

In order to comprehensively explore 5-LO inhibition capacity of all isolated *Acronychia*-type acetophenones both cell-free and intact cell assays were performed. Moreover, the estimation of mPGES-1 inhibition was considered important to assess a potential dual action of this group of compounds. For the estimation of 5-LO activity zileuton, which is approved for asthma treatment (Wenzel et al. 1996), was used as a positive control whereas mPGES-1 inhibition capacity was compared to MK886 (Rouzer et al. 1990). Interestingly, all *Acronychia*-type acetophenones exhibited important inhibition of 5-LO and mPGES-1 with some compounds presenting IC_{50} values significantly close to the positive controls (Table VI).

Table VI: Inhibition of 5-LO and mPGES-1 activity by natural *Acronychia*-type acetophenones and control inhibitors; ^a values are means $\pm$ SE, $n = 3 - 4$; n.d.: not determined; ^b 5-LO, cell-free: isolated human recombinant 5-LO; ^c 5-LO, intact cells: intact human PMNL stimulated with 2.5 μ M A23187 ionophore; ^d mPGES-1, cell-free: microsomal preparations of IL-1 β stimulated A549 cells

compound	5-LO		mPGES-1
	IC_{50} [μ M] ^a		IC_{50} [μ M] ^a
	cell-free ^b	intact cells ^c	cell-free ^d
acopyrone (1)	6 $\pm$ 0.6	> 10	4.2 $\pm$ 0.2
acopyranol A (2)	2 $\pm$ 0.5	5.0 $\pm$ 0.6	1.9 $\pm$ 0.2
acopyranol B (3)	7.3 $\pm$ 1.8	5.3 $\pm$ 0.7	2.7 $\pm$ 0.1
acrovestone (4)	2.7 $\pm$ 0.2	2.3 $\pm$ 0.6	1.1 $\pm$ 0.02
acrovestenol (5)	2.5 $\pm$ 0.3	6.3 $\pm$ 2	1.1 $\pm$ 0.02
acrofolione A (6)	5.3 $\pm$ 1.3	7.3 $\pm$ 0.7	1.3 $\pm$ 0.2
acrofolione B (7)	> 10	4.5 $\pm$ 1	1.1 $\pm$ 0.1
zileuton	0.6 $\pm$ 0.1	1.7 $\pm$ 0.7	n.d.
MK886	n.d.	0.03 $\pm$ 0.01	2.5 $\pm$ 0.5

In particular, concerning mPGES-1 inhibition capacity, the majority of tested compounds displayed IC_{50} values of 2-fold magnitude lower comparing to the positive

control (MK866). Acropyranol A (**2**) demonstrated potent inhibition of 5-LO in cell-free (IC_{50} 2 μ M) and intact cell assay (IC_{50} 5 μ M) while important was also the mPGES-1 inhibition (IC_{50} 1.9 μ M). Moreover, interesting inhibitory activity of both 5-LO and mPGES-1 enzymes was observed from acrovestenol (**5**). However, among them, acrovestone (**4**) represented the most potent dual inhibitor with IC_{50} values of 2.7 μ M in cell-free and 2.3 μ M in intact cell 5-LO assays along with an IC_{50} value of 1.1 μ M in cell-free mPGES-1 assay (Figure 15). These results suggest acrovestone as an important natural dual inhibitor potentially applicable for the treatment of inflammation related diseases.

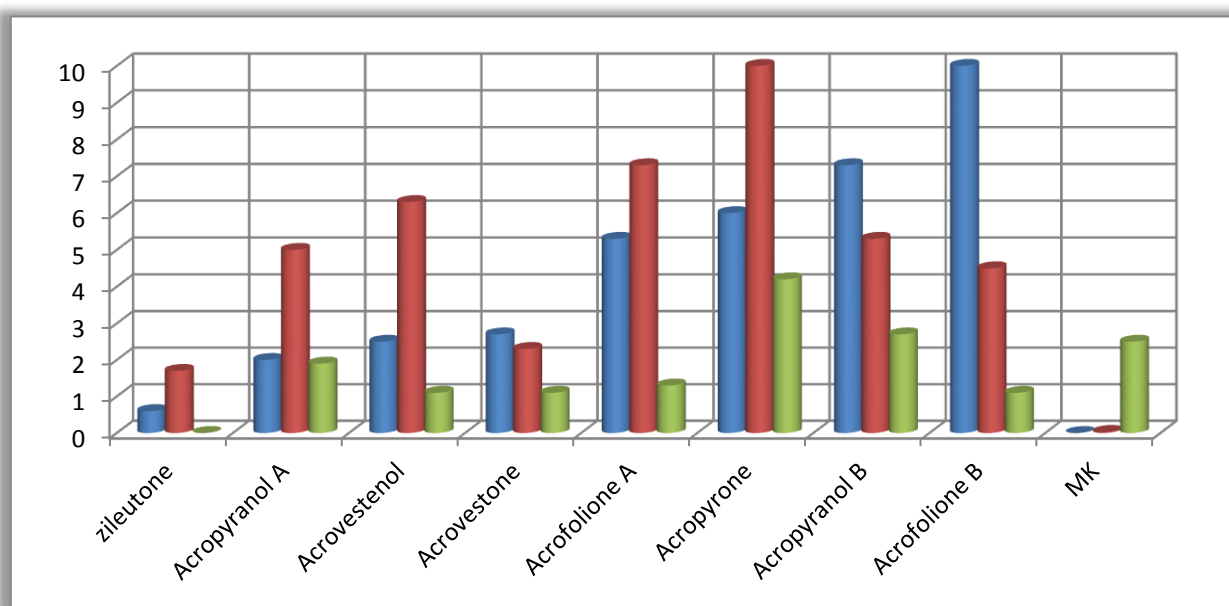


Figure 15: Comparative histogram representation of 5-LO and mPGES-1 inhibition activity by *Acronychia*-type acetophenones

Overall, this potency magnitude observed for *Acronychia*-type acetophenones is really promising for natural compounds (Pergola et al. 2010) confirming the traditional use of *Acronychia* genus for asthma treatment. Furthermore, these findings indicate a potentially new scaffold for development of new dual 5-LO/ mPGES-1 inhibitors.

Towards an effort to explore the way *Acronychia*-type acetophenones interact and inhibit 5-LO, docking calculations were performed using the structure S663D stable 5-LO mutant in complex with the natural substrate, arachidonic acid. The interaction mode of acrovestone with 5-LO is illustrated in Figure 16. Acrovestone binds 5-LO in the active site of the enzyme, forming a number of hydrophobic interactions as well as hydrogen bonds with residues of the catalytic site. However, no chelation with iron was suggested by docking. All

Acronychia-type acetophenones were found to bind 5-LO with highly similar interactions. This provides a first insight into the structure activity relationship (SAR) of 5-LO inhibition by *Acronychia*-type acetophenones suggesting possible synthetic modifications on the acetophenone dimer lead.

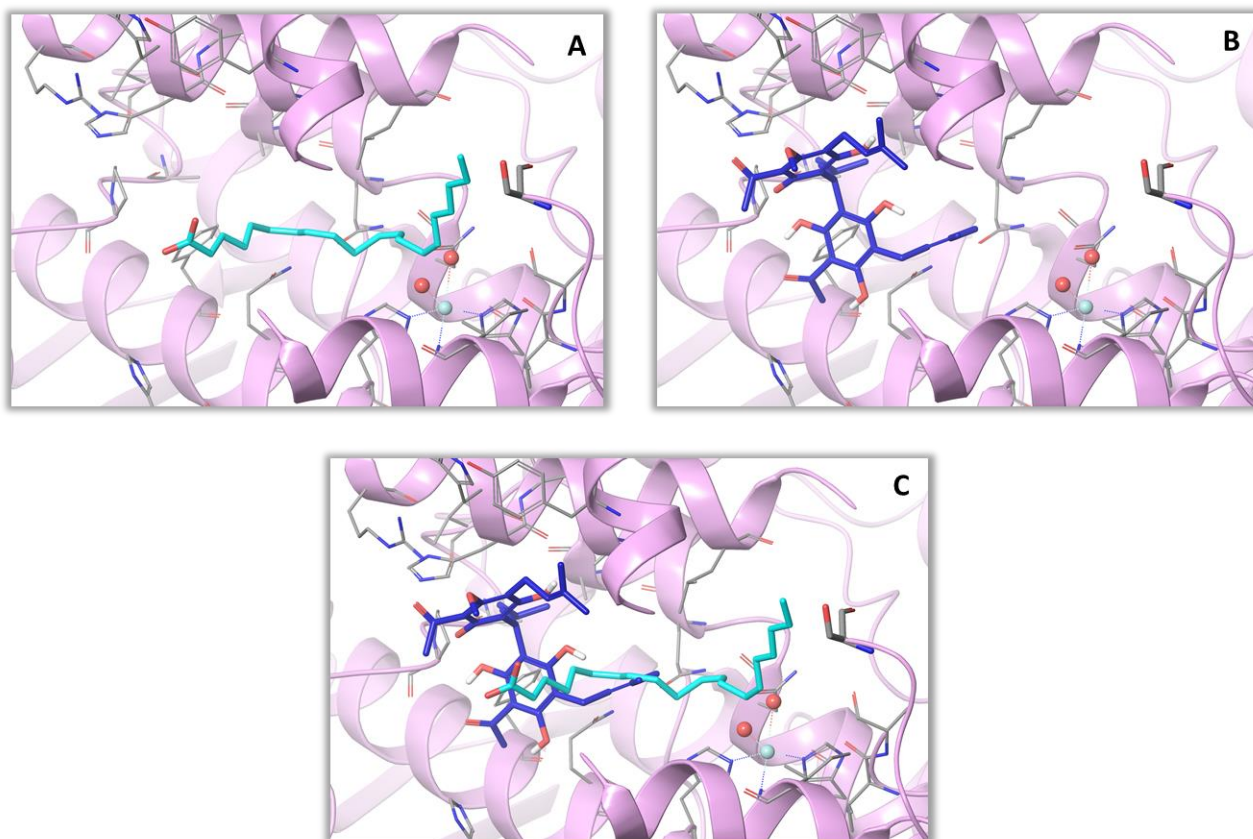


Figure 16: A. Cocrystal structure of AA bound to 5-LO (pdb code 3V99), B. docking results of acrovestone bound to 5-LO and C. superposition of AA and acrovestone in the active site of 5-LO; the protein is depicted in a ribbon representation in purple color while the catalytic site iron and adjacent water molecules are depicted as blue and red spheres, respectively

Conclusion

In this chapter, the targeted isolation of a specific category of compounds, *Acronychia*-type acetophenones is described from trunk barks of *Acronychia pedunculata*. Following a special extraction protocol, *Acronychia*-type acetophenones were extracted separately from alkaloid compounds, also present in the plant material. The application of multiple analytical hyphenated techniques for the detection of selected metabolites conducted to a following isolation step in an efficient and rapid manner. Moreover, separation and purification of targeted compounds was performed by combining orthogonal chromatographic techniques (FCPC, semiprep-HPLC) leading to the isolation of seven *Acronychia*-type acetophenones, among them three new natural products, in a rapid and efficient way.

The structural peculiarities of this category of secondary metabolites led to the application of different methodologies for the detection of conformational rotamers present in dynamic equilibrium in solution. Initially, variable temperature NMR studies were performed allowing the definition of rotational barrier energies of the observed conformers. Emphasis was given at low temperature experiments to determine rotamer presence and spectral characteristics of individual conformer. In parallel, Boltzmann normalized populations of all dominant conformers were determined for each compound using the molecular mechanics energies calculated at the same temperature with NMR experiments (0 °C) leading to a thorough description of conformational rotamers. Furthermore, resolution of enantiomers and diastereomers of *Acronychia*-type acetophenones was successfully performed by normal phase chiral chromatography allowing the complete characterization of these compounds. In case of acrofolione A the two diastereomers were isolated in high purity using SFC selecting the method following a rapid screening protocol.

The pharmacological evaluation of the isolated *Acronychia*-type acetophenones was based on the traditional use of the plant material. A significant antibacterial activity of acrovestone against a number of *Staphylococcus* strains supported the traditional use of *Acronychia* for skin and respiratory infections. Moreover, the assessment of cytotoxicity of all isolated compounds against prostate and melanoma human cancer cells revealed an interesting activity of acrovestone and acrovestenol while absence of cytotoxicity was

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

observed in normal cell lines. Finally, the anti-inflammatory activity designated from the utilization of the plant material for the treatment of asthma and rheumatism was assessed by the inhibition capacity of 5-LO and mPGES-1 enzymes indicating acrovestone, acropyranol A and acrovestenol the most potent compounds among *Acronychia*-type acetophenones.

APPENDIX

1

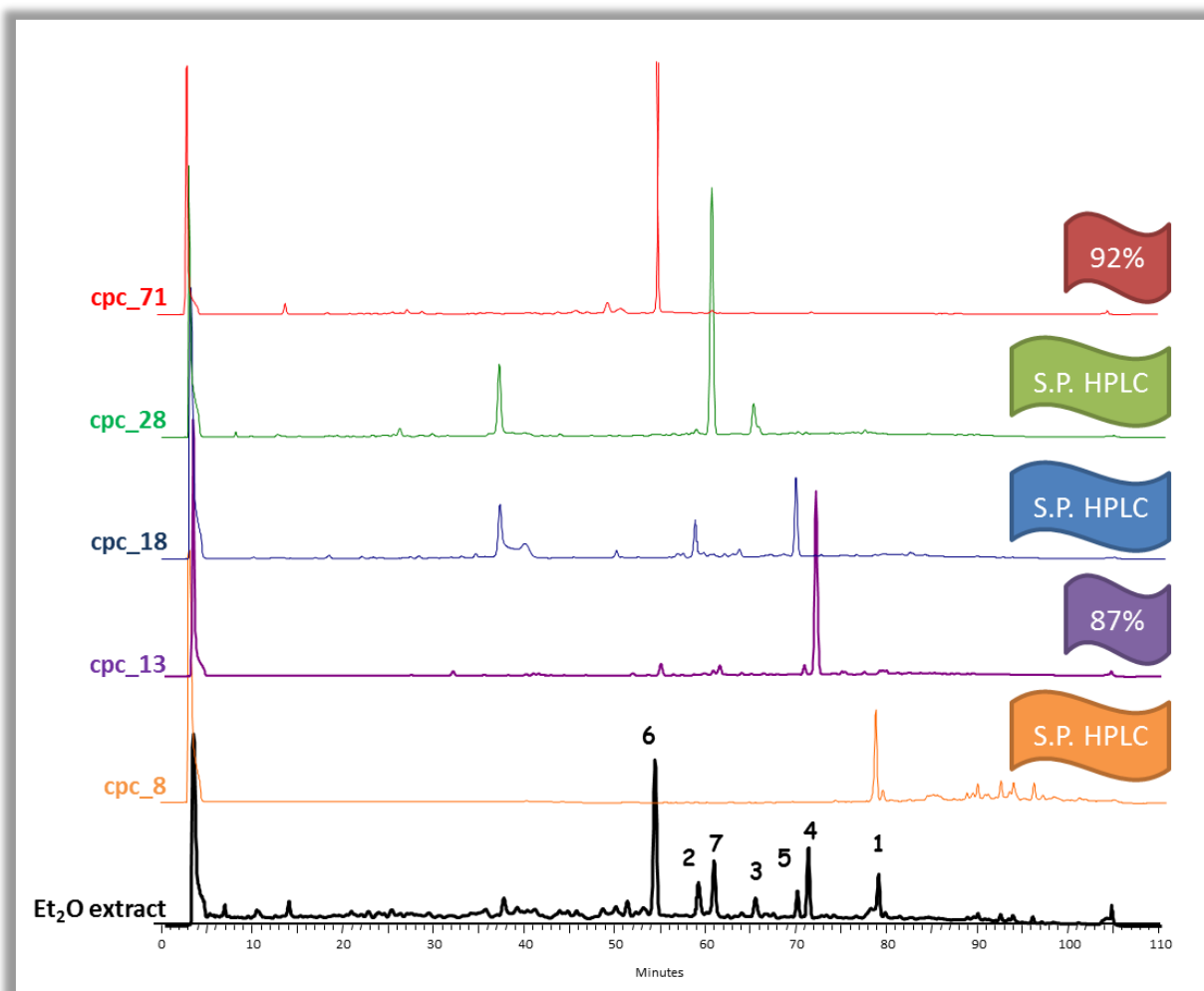


Figure A 1: HPLC chromatograms at 280 nm of selected FCPC fractions obtained by the fractionation of Et₂O extract of *A. pedunculata*

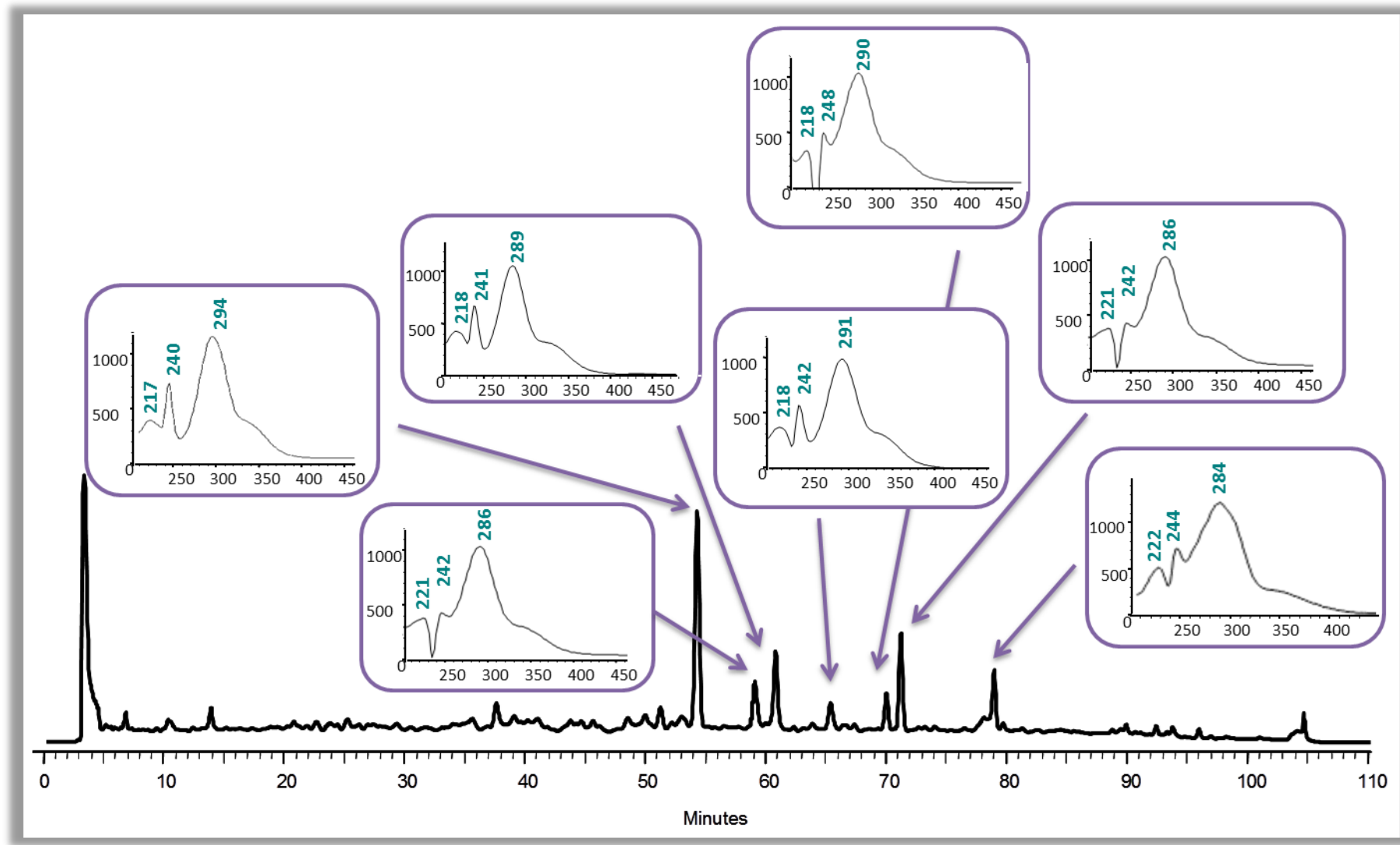


Figure A 2: HPLC-PDA chromatogram at 280 nm of Et₂O extract obtained from trunk barks of *Acronychia pedunculata* and the extracted UV spectra of the major peaks

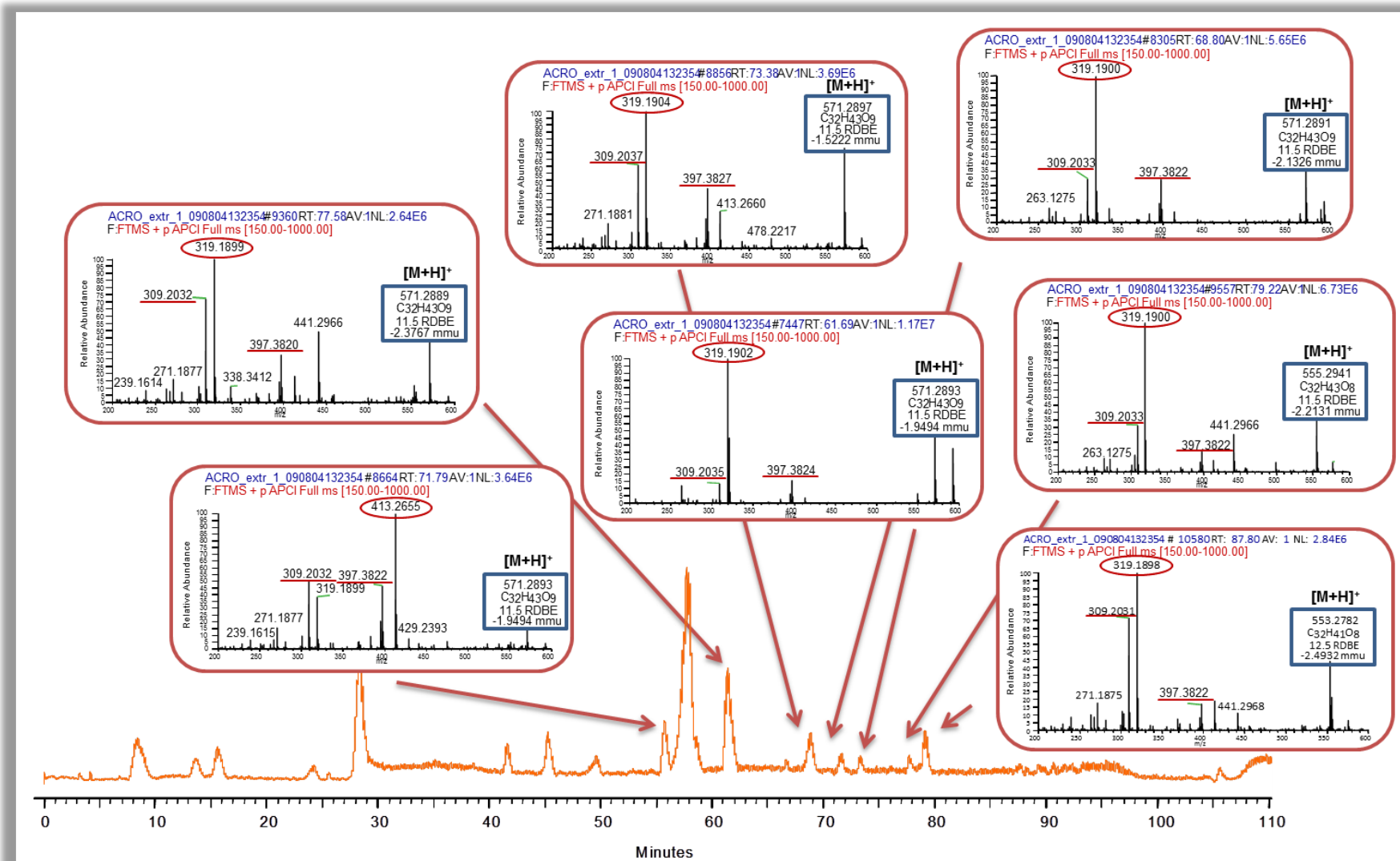


Figure A 3: Total Ion current (TIC) chromatogram obtained from the HPLC-APCI(+)-Orbitrap analysis of Et₂O extract of *A. pedunculata* and the extracted full scan HRMS spectra of the peaks presented common spectrometric features

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

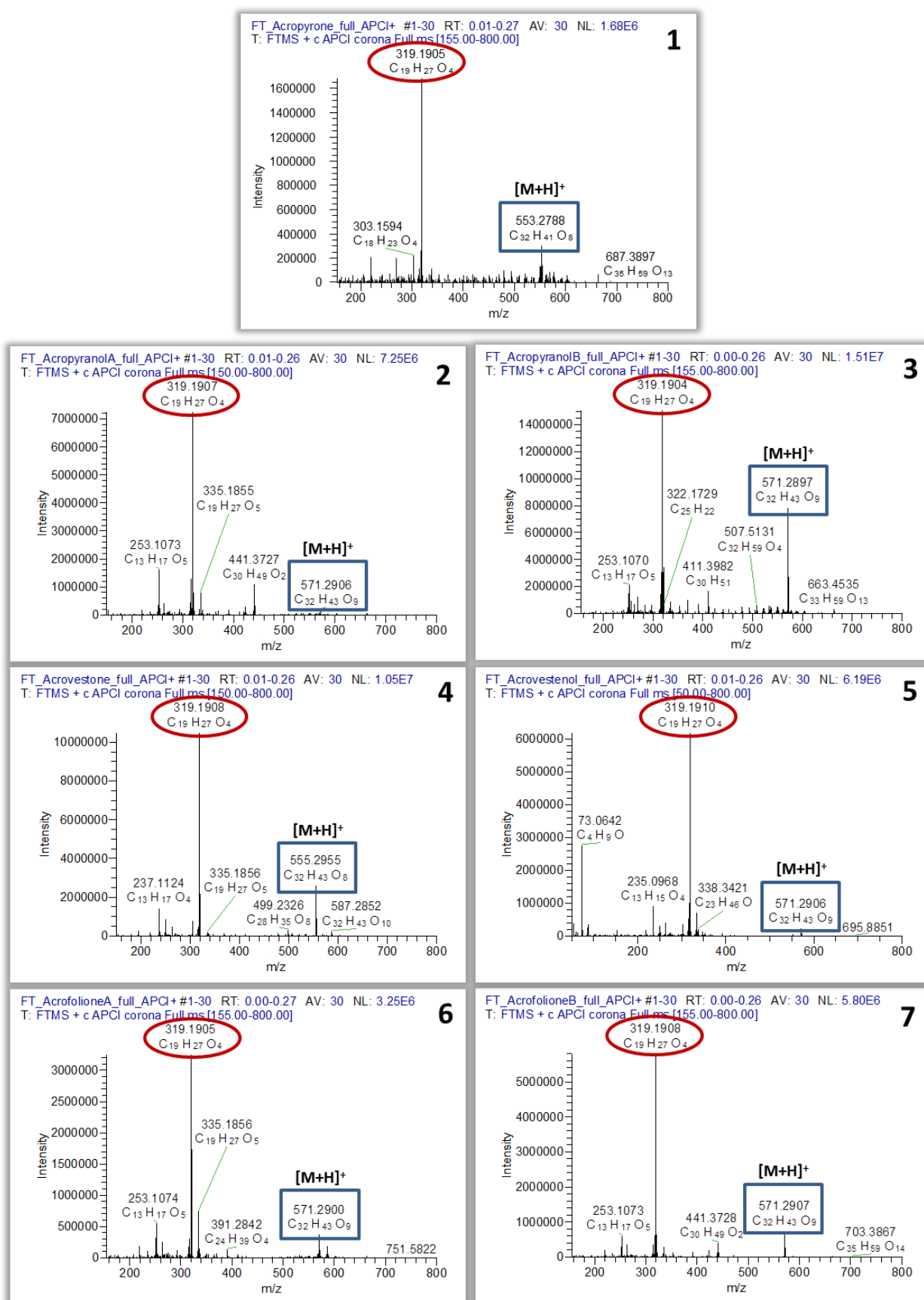
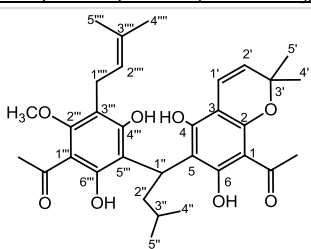


Figure A 4: APCI(+)-HRMS full scan spectra of isolated compounds 1-7

Table A 1: NMR spectroscopic data of Acropyrone (1) at 47 °C

NMR spectroscopic data (600 MHz, 47 °C CDCl ₃) for Acropyrone				
				
No	¹ H (J, Hz)	¹³ C	HMBC	NOESY
1		104.9		
2		155.8		
3		103.5		
4		158.9		
5		108.9/112.9		
6		161.0		
1'	6.66 (d, 10.1)	117.3	2, 3, 4, 3', 4', 5'	2'
2'	5.43 (d, 10.1)	124.8	2, 3, 3', 4', 5'	1', 4', 5'
3'		78.0		
4'	1.48 s	28.0	2', 3', 5'	2', MeCO-1
5'	1.48 s	28.0	2', 3', 4'	
1''	4.74 (t, 7.7)	28.5	4, 5, 6, 2'', 3'', 4'', 5'', 6''	2'', 3'', 4'', 5''
2''	2.22 m	39.4	4'', 5''	1'', 4'', 5''
	2.14 m			
3''	1.43 m	27.2	1'', 2'', 4'', 5''	1'', 2'', 4'', 5''
4''	0.88 (d, 5.2)	22.5	2'', 3'', 5''	1'', 2'', 3''
5''	0.88 (d, 5.2)	22.5	2'', 3'', 4''	
1'''		108.2		
2'''		160.0		
3'''		116.7		
4'''		162.4		
5'''		108.9/112.9		
6'''		161.0		
1''''	3.30 (d, 4.8)	23.2	3''', 2''', 3''''	2''', 5''', MeO
2''''	5.20 (tt, 6.5/1.6)	123.2	3''', 1''', 4''', 5''''	1''', 4''', MeO
3''''		131.4		
4''''	1.69 s	25.6	2''', 3''', 5''''	2''''
5''''	1.77 s	17.8	2''', 3''', 4''''	1''', MeO
MeO	3.71 s	62.5	2'''	1''', 5''', MeCO-1'''
MeCO-1	2.68 s	32.5	1, MeCO-1	4', 5'
MeCO-1		203.9		
MeCO-1'''	2.70 s	30.4	1'', MeCO-1'''	MeO
MeCO-1'''		204.3		
OH-4				
OH-6				
OH-4'''				
OH-6'''				

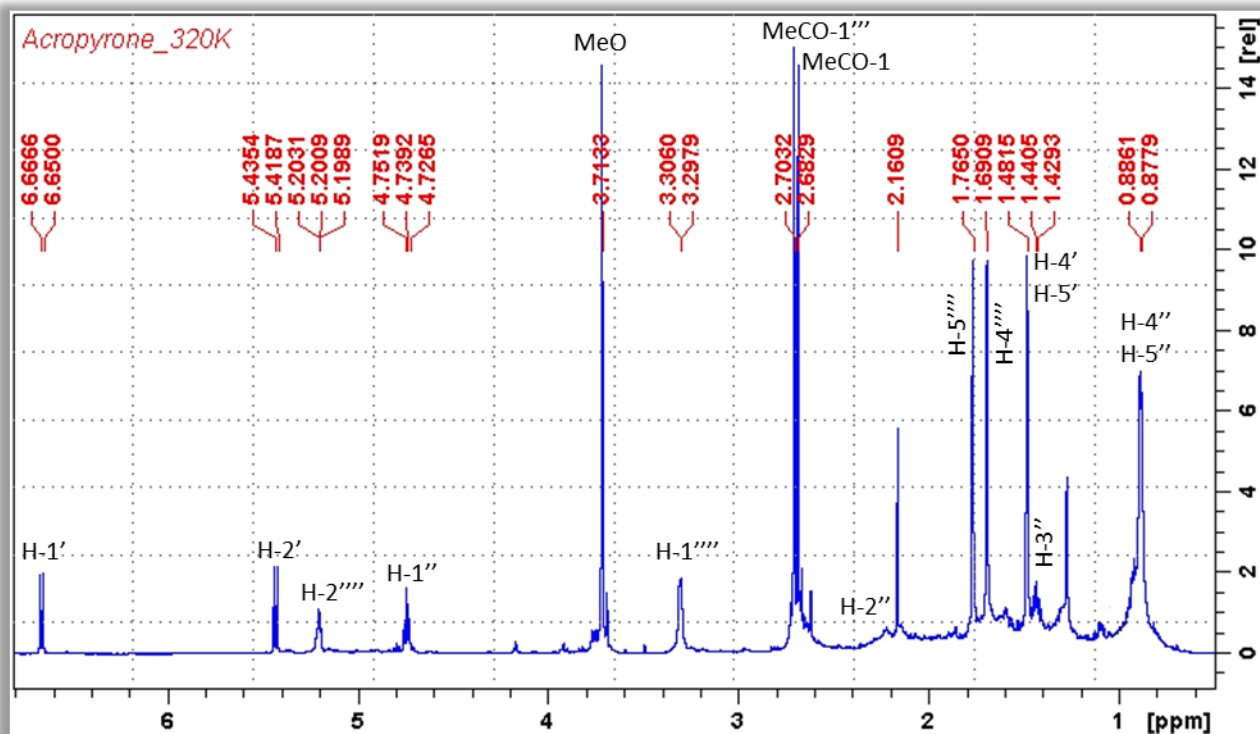


Figure A 5: ^1H NMR (600 MHz, CDCl_3 , 47 $^\circ\text{C}$) spectrum of Acropyrone (1)

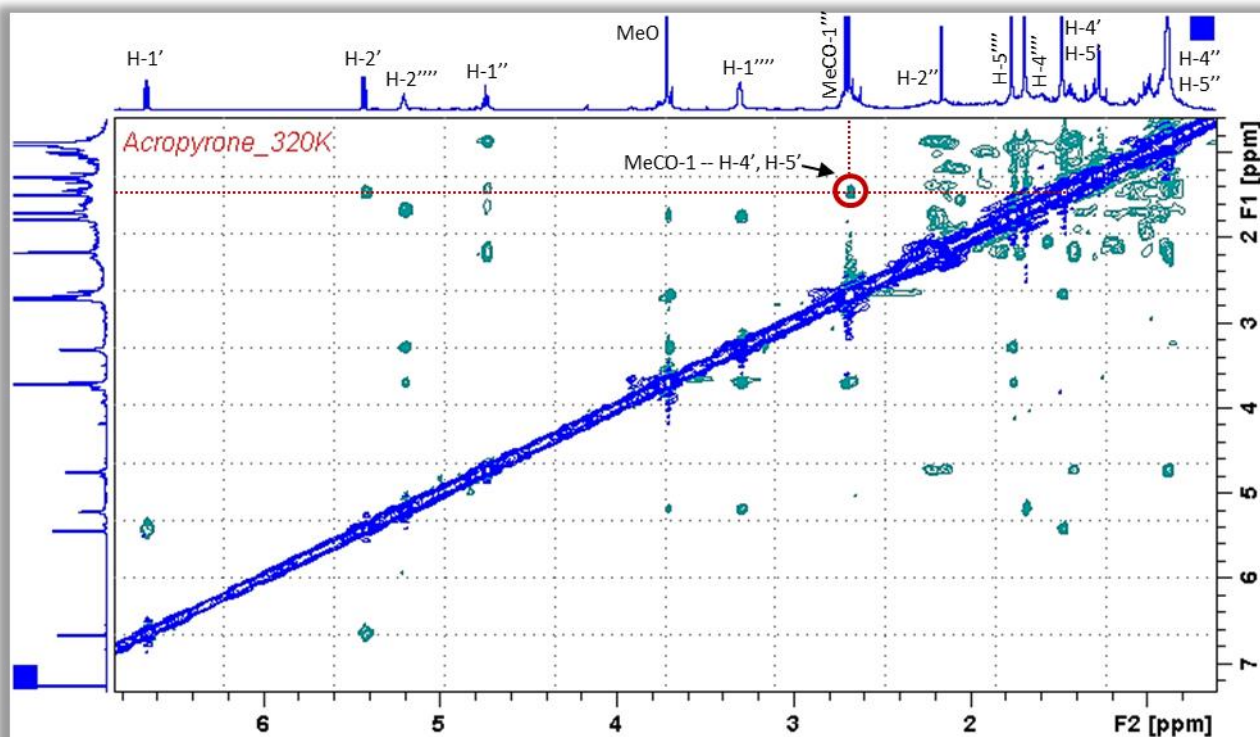
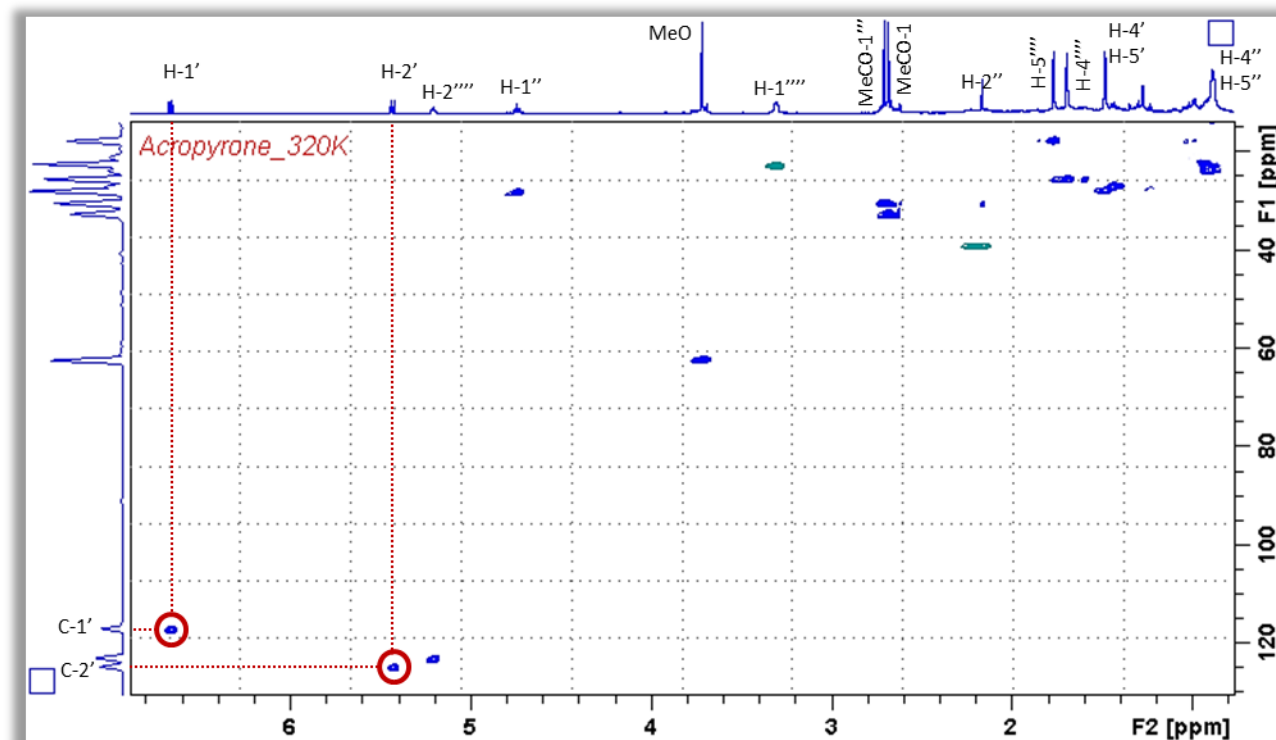
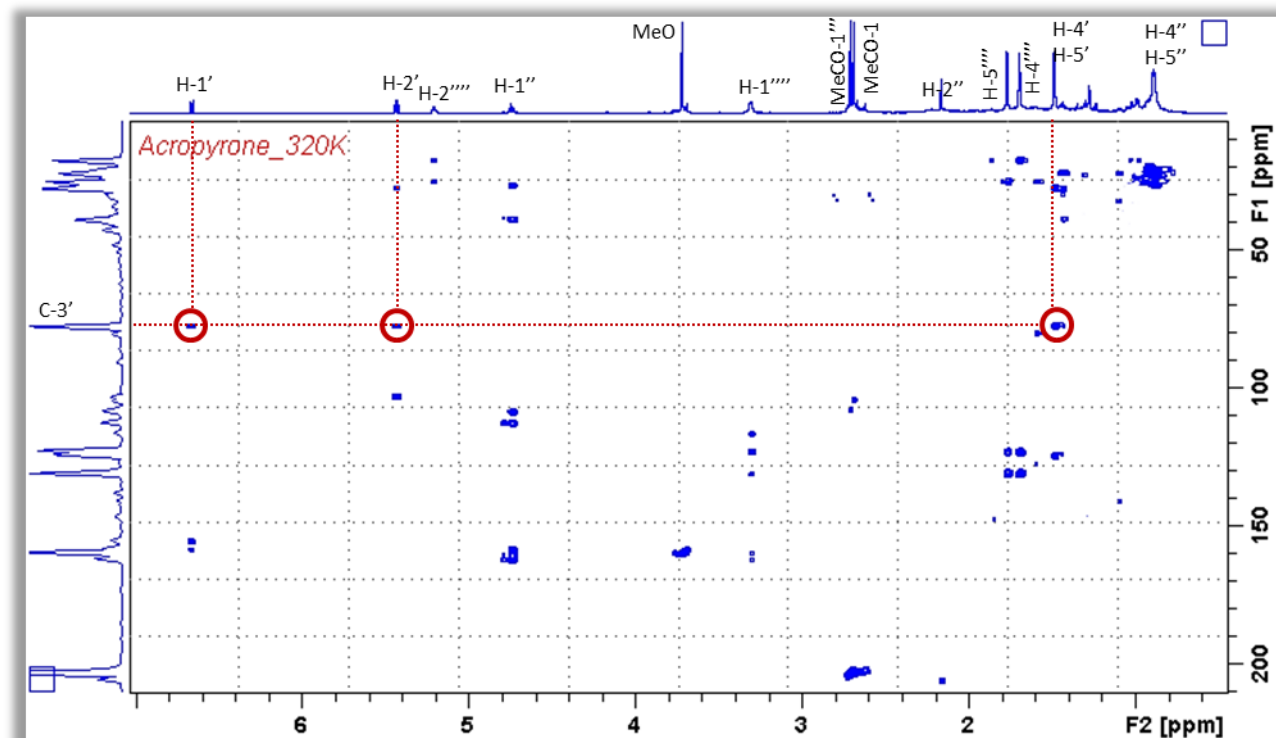
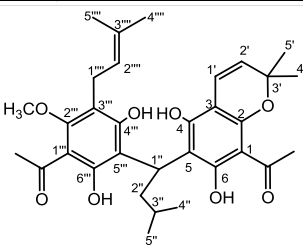


Figure A 6: NOESY (600 MHz, CDCl_3 , 47 $^\circ\text{C}$) spectrum of Acropyrone (1)

Figure A 7: HSQC (600 MHz, CDCl₃, 47 °C) spectrum of Acropyrone (1)Figure A 8: HMBC (600 MHz, CDCl₃, 47 °C) spectrum of Acropyrone (1)

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

Table A 2: NMR spectroscopic data of Acropyrone (1) at 0 °C; different rotamers' signals are assigned in red and blue

NMR spectroscopic data (600 MHz, 0 °C CDCl ₃) for Acropyrone				
				
No	¹ H (J, Hz)	¹³ C	HMBC	NOESY
1		104.8		
2		155.7		
3		103.4		
4		158.7		
5		108.5		
		108.0		
6		161.3		
1'	6.62 (d, 10.0)	116.9	2, 3, 4, 3', 4', 5'	2'
2'	5.44 (d, 10.0)	124.9	2, 3, 3', 4', 5'	1', 4', 5'
3'		78.1		
4'	1.48 s	27.8	2', 3', 5'	2', MeCO-1
5'	1.45 s	27.9	2', 3', 4'	2', MeCO-1
1''a	4.71 (t, 7.3)	27.8	4, 5, 2'', 3'', 4'', 5'', 6, 6''	4'', 5'', OH-4, OH-4''
	4.70 (t, 7.3)			
2''a	2.19 m	39.5		
	2.30 m	39.1		
2''b	2.13 m	39.5	5, 1'', 3'', 4'', 5'', 5''	1'', 4'', 5'', OH-4, OH-4''
	2.04 m	39.1		
3''	1.40 m	26.6	2'', 4'', 5''	2'', 4'', 5''
4''	0.89 (d, 6.2)	22.4	2'', 3'', 5''	1'', 2'', 3''
5''	0.84 (d, 6.5)	22.3	2'', 3'', 4''	1'', 2'', 3''
1'''		108.1		
2'''		159.9		
3'''		116.8		
4'''		162.2		
5'''		112.5		
		113.1		
6'''		160.8		
1''''	3.33brm	22.8	2''', 3''', 4''', 2''', 3''''	2''', 5''', MeO
2''''	5.17 brt	122.8	3''', 1''', 4''', 5''''	1''', 4''', MeO
3''''		131.8		
4''''	1.68 s	25.7	3''', 2''', 3''', 5''''	2''''
5''''	1.75 s	17.8	3''', 2''', 3''', 4''''	1''', MeO
MeO	3.70 s	62.6	2''	1''', 2''', 5''', MeCO-1''
MeCO-1	2.68 s	32.9	1, MeCO-1	4', 5', OH-6
MeCO-1		203.4		
MeCO-1'''	2.70 s	30.8	1''', MeCO-1'''	MeO, OH-6'''
MeCO-1'''		204.2		
OH-4	10.11		2, 3, 4, 5	1'', 2'', OH-6'''
	9.27		2, 3, 4, 5	
OH-6	16.11		1, 5, 6	1'', 3'', 4'', 5'', MeCO-1, OH-4'''
	15.94		1, 4, 5	
OH-4'''	9.26		2''', 3''', 4''', 5'''	1'', 2'', OH-6
	10.11		2''', 3''', 4''', 5'''	
OH-6'''	15.78		1''', 5''', 6'''	1'', 3'', 4'', 5'', MeCO-1''', OH-4
	15.65		1''', 2''', 5'''	

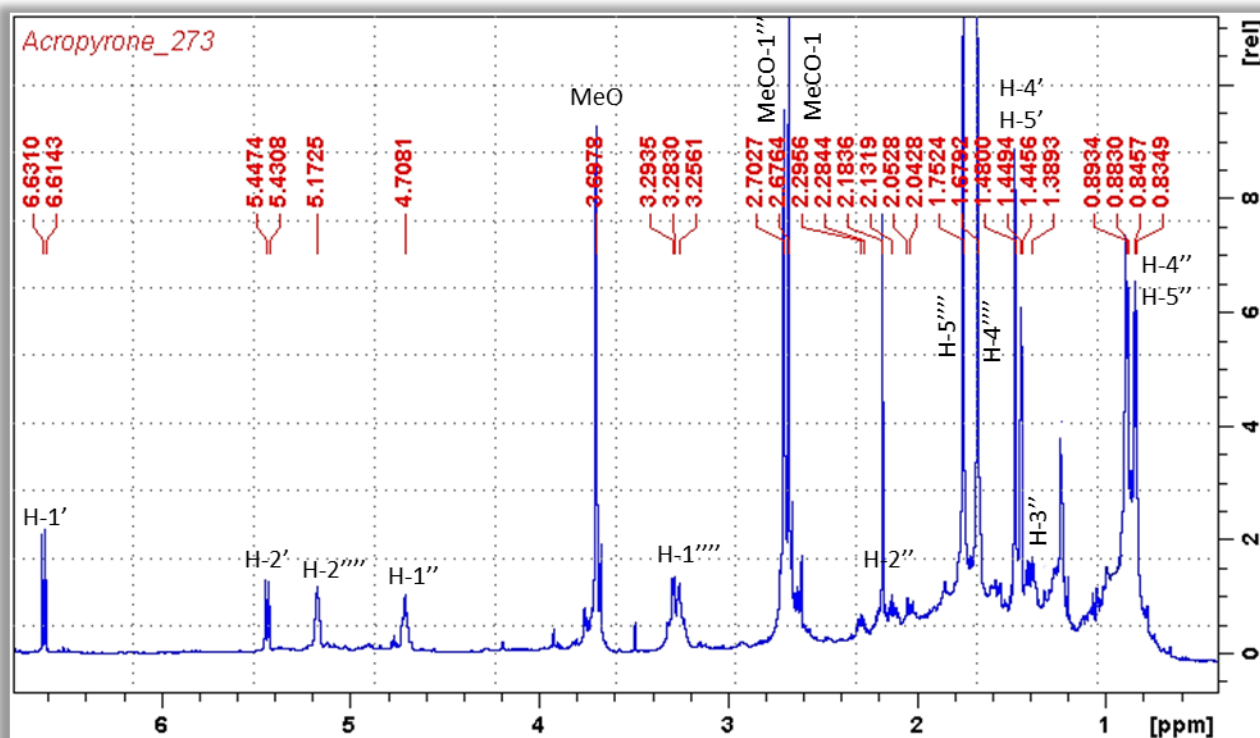


Figure A 9: ^1H NMR (600 MHz, CDCl_3 , 0°C) spectrum of Acropyrone (1) δ_{H} 0-7

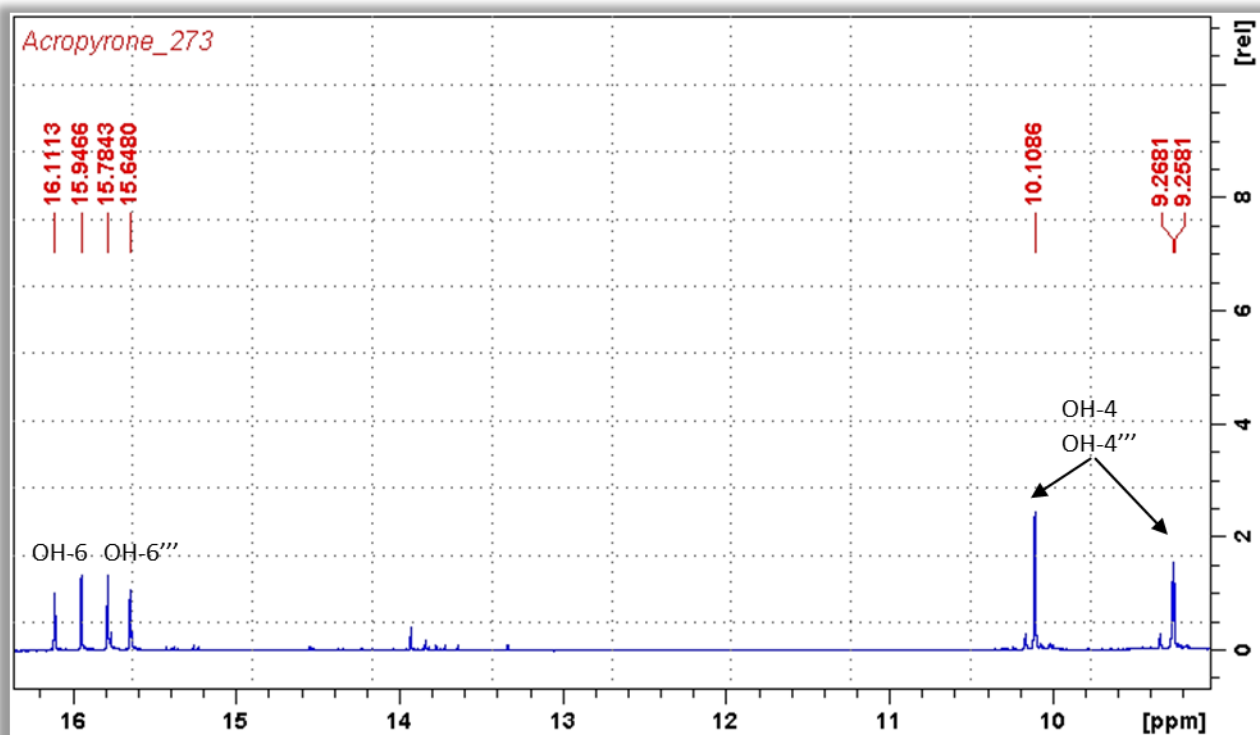


Figure A 10: ^1H NMR (600 MHz, CDCl_3 , 0°C) spectrum of Acropyrone (1) δ_{H} 9-16.2

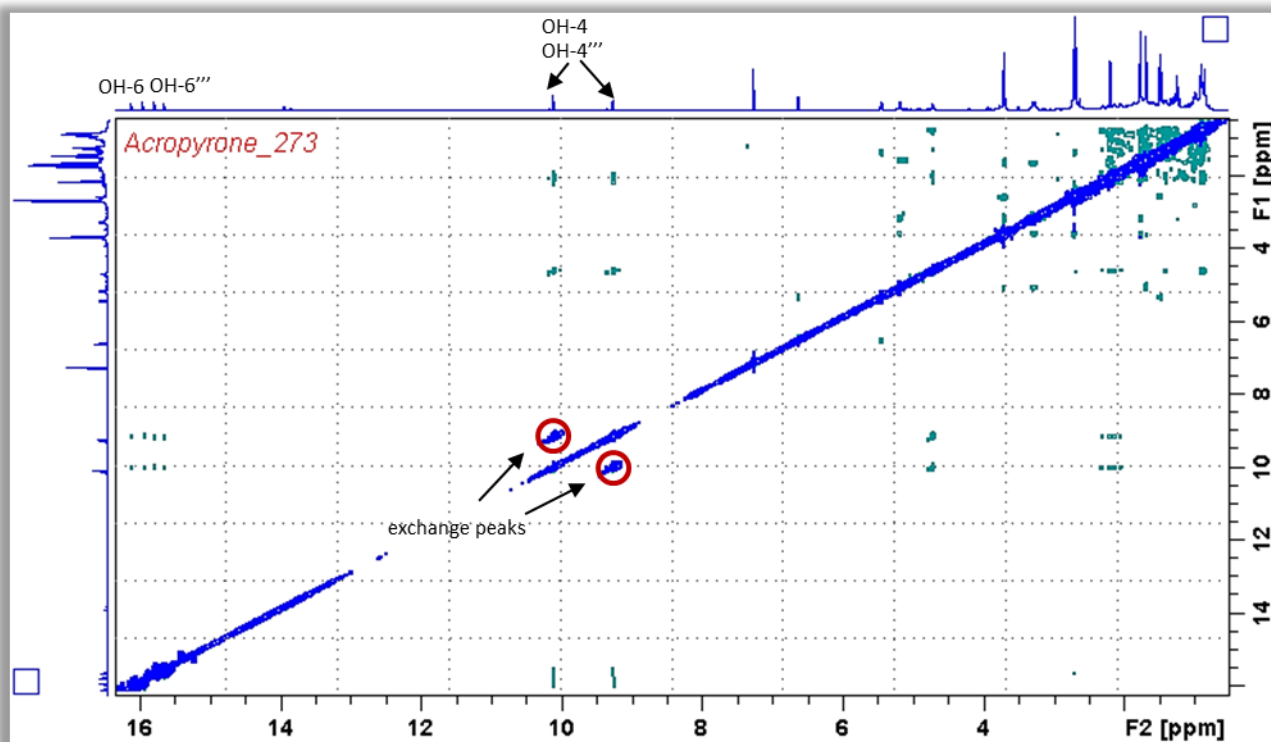


Figure A 11: NOESY (600 MHz, CDCl_3 , 0 °C) spectrum of Acropyrone (1)

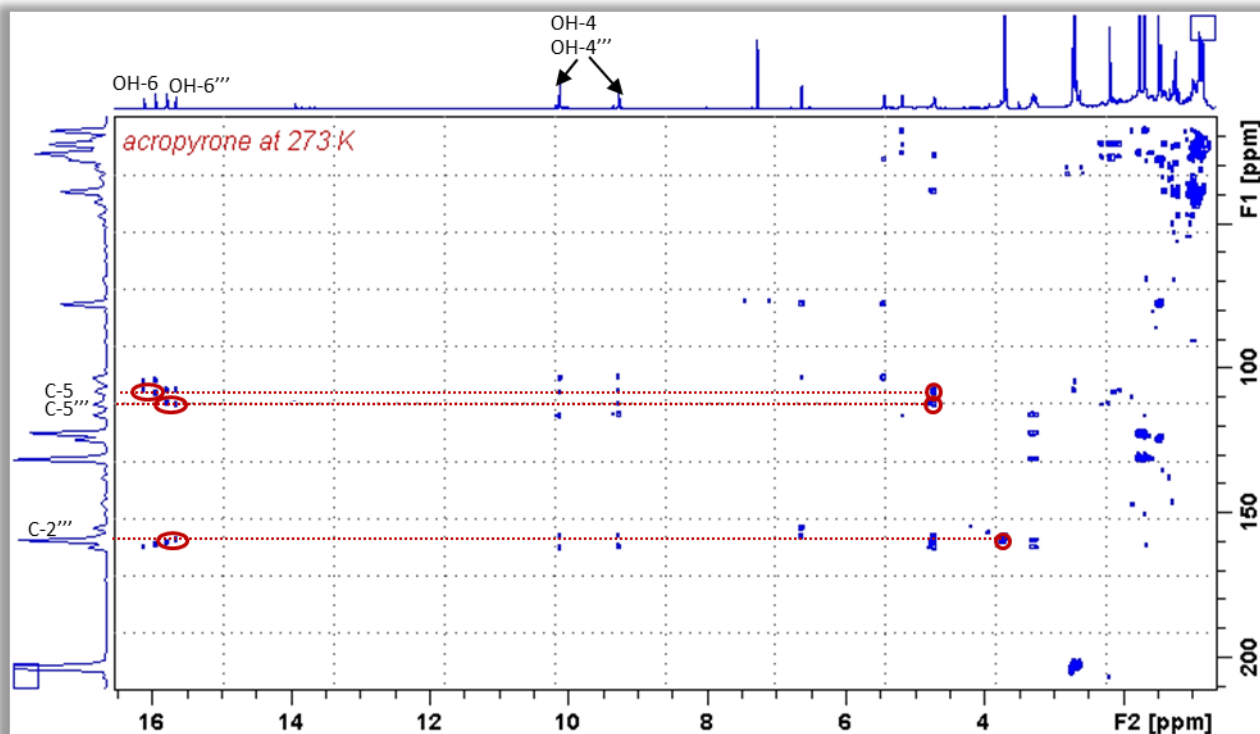


Figure A 12: HMBC (600 MHz, CDCl_3 , 0 °C) spectrum of Acropyrone (1)

Table A 3: NMR spectroscopic data of Acropyranol A (2) at 47 °C; different diastereomers' signals are assigned in grey

NMR spectroscopic data (600 MHz, 47 °C CDCl ₃) for Acropyranol A				
No	¹ H (J, Hz)	¹³ C	HMBC	NOESY
1		106.7		
2		162.5		
3		98.7		
4		155		
5		106.9		
6		159.7		
1'a	2.94 (dd, 17.1/ 4.8)	25.7	2, 3, 4, 5, 2', 3'	1'b, 2', 4'
	2.92 (dd, 17.1/ 4.6)			
1'b	2.68 m			1'a, 2', 5'
2'	3.86 (t, 5.3)	68.7	3, 1', 4', 5'	1'a, 1'b, 4', 5'
	3.85 (t, 5.5)			
3'		80.7		
4'	1.49 s	24.4	2', 3', 5'	1'a, 2'
	1.47 s			
5'	1.53 s	21.3	2', 3', 4'	1'b, 2'
	1.51 s			
1''	4.75 brs	28.3	6, 2'', 4'', 5'', 6''	2'', 3'', 4'', 5''
2''a	2.31 m	40	5, 1'', 3'', 4'', 5'', 5''	2''b, 3'', 4'', 5''
2''b	1.99 brs			
3''	1.41 m	27	1'', 2'', 4'', 5''	1'', 2''a, 4'', 5''
4''	0.90 (d, 6.6)	22.6	2'', 3'', 5''	1'', 2''a, 2''b, 3''
5''	0.89 (d, 6.6)	22.3	2'', 3'', 4''	
1'''		108.4		
2'''		160.2		
3'''		116.3		
4'''		162		
5'''		112.8		
6'''		159.7		
1''''	3.30 (d, 6.8)	22.9	2''', 3''', 4''', 2'''', 3''''	2''', 5''', MeO
2''''	5.16 (tt, 6.8/1.3)	123	3''', 1''', 4''', 5''''	1''', 4''', MeO
3''''		131.3		
4''''	1.67 s	25.5	2''', 3''', 5''''	2''''
5''''	1.77 s	17.8	2''', 3''', 4''''	1''', MeO
MeO	3.73 s	62.6	2'''	1''', 2''', 5''', MeCO-1'''
MeCO-1	2.70 s	33.4	1, MeCO-1	
MeCO-1		204.7		
MeCO-1'''	2.72 s	30.5	1''', MeCO-1'''	MeO
MeCO-1'''		203.8		
OH-2	14.0 s		1, 2, 3	MeCO-1, 1'
OH-6				
OH-2'				
OH-4'''				
OH-6'''	15.60 s		1''', 5''', 6'''	MeCO-1''', 1''

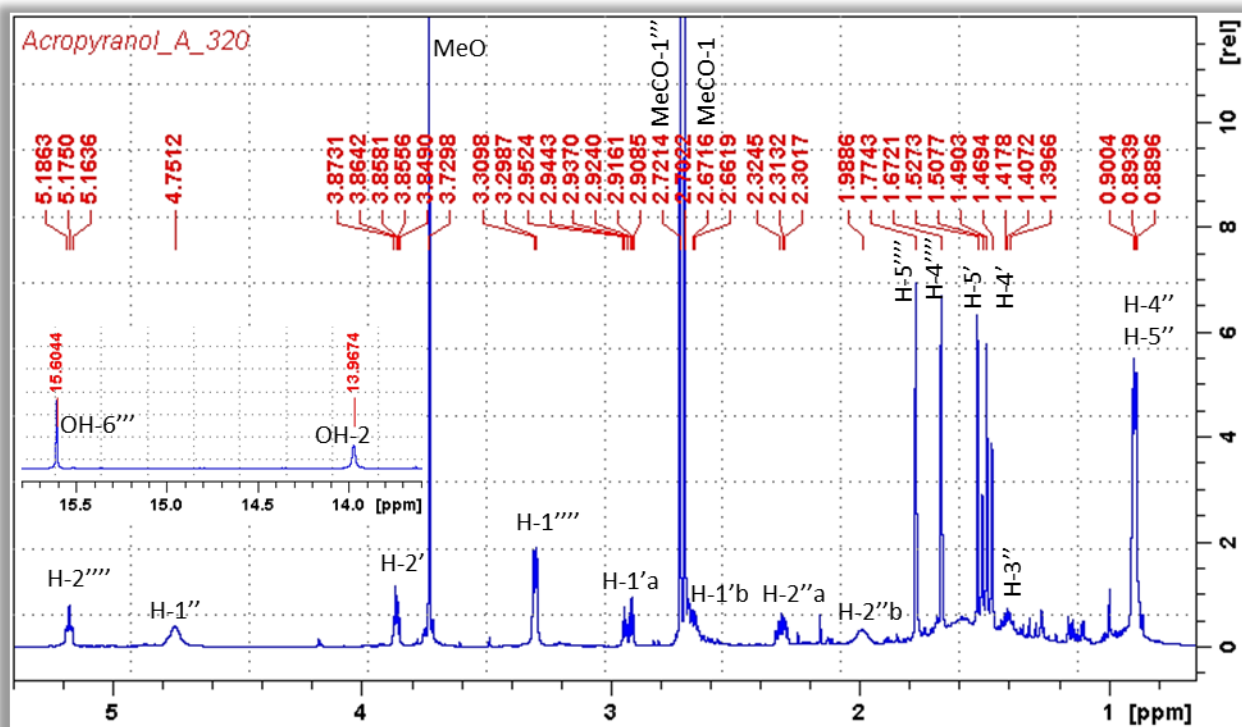


Figure A 13: ^1H NMR (600 MHz, CDCl_3 , 47 °C) spectrum of Acropyranol A (2)

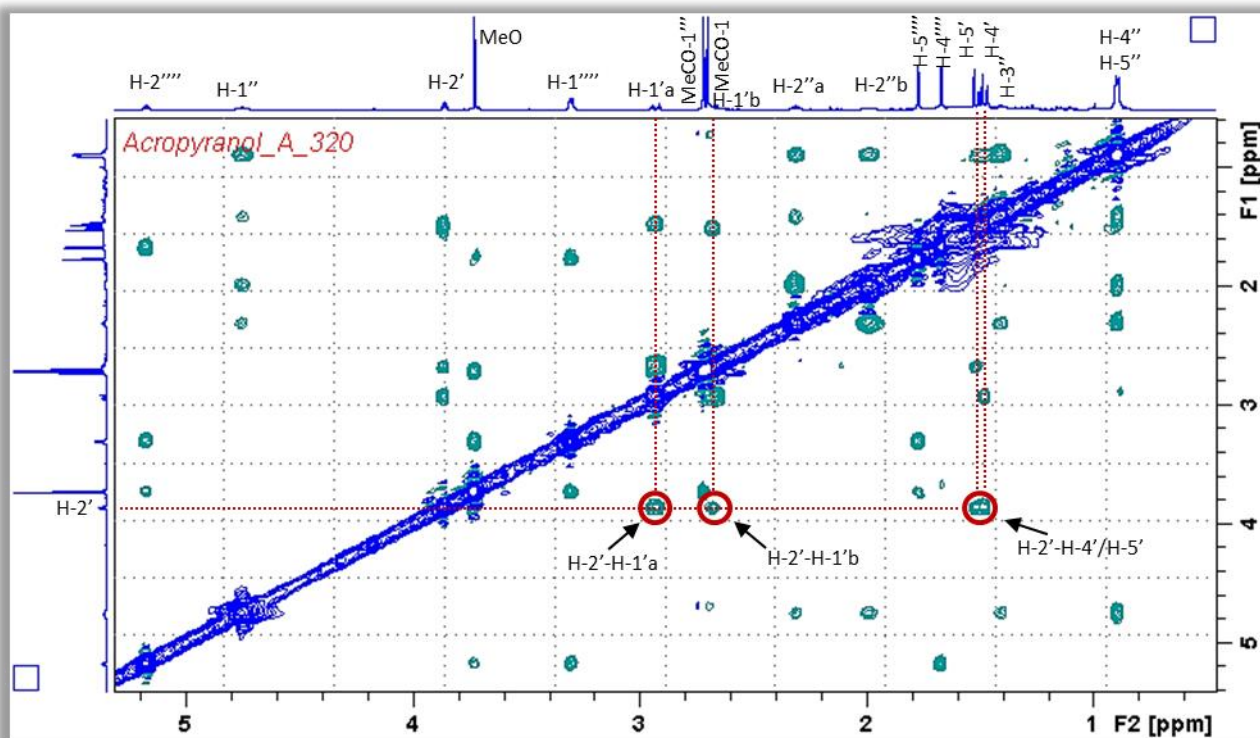
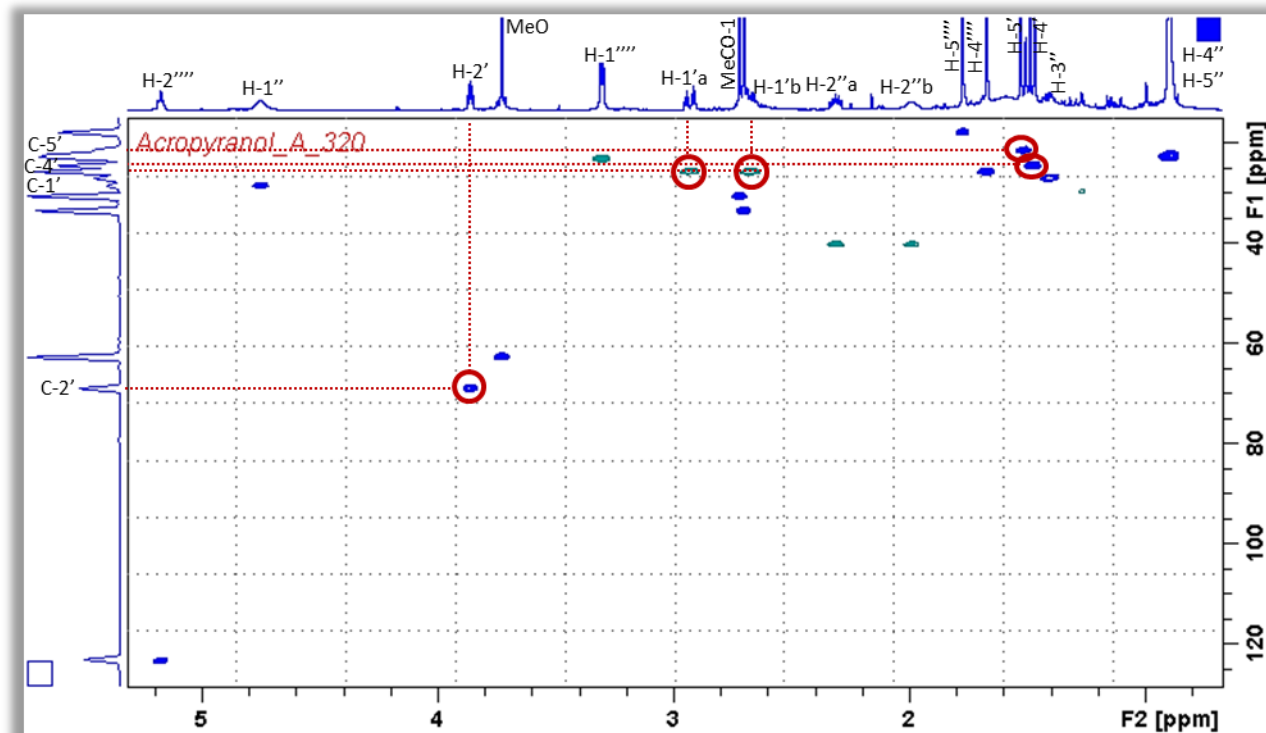
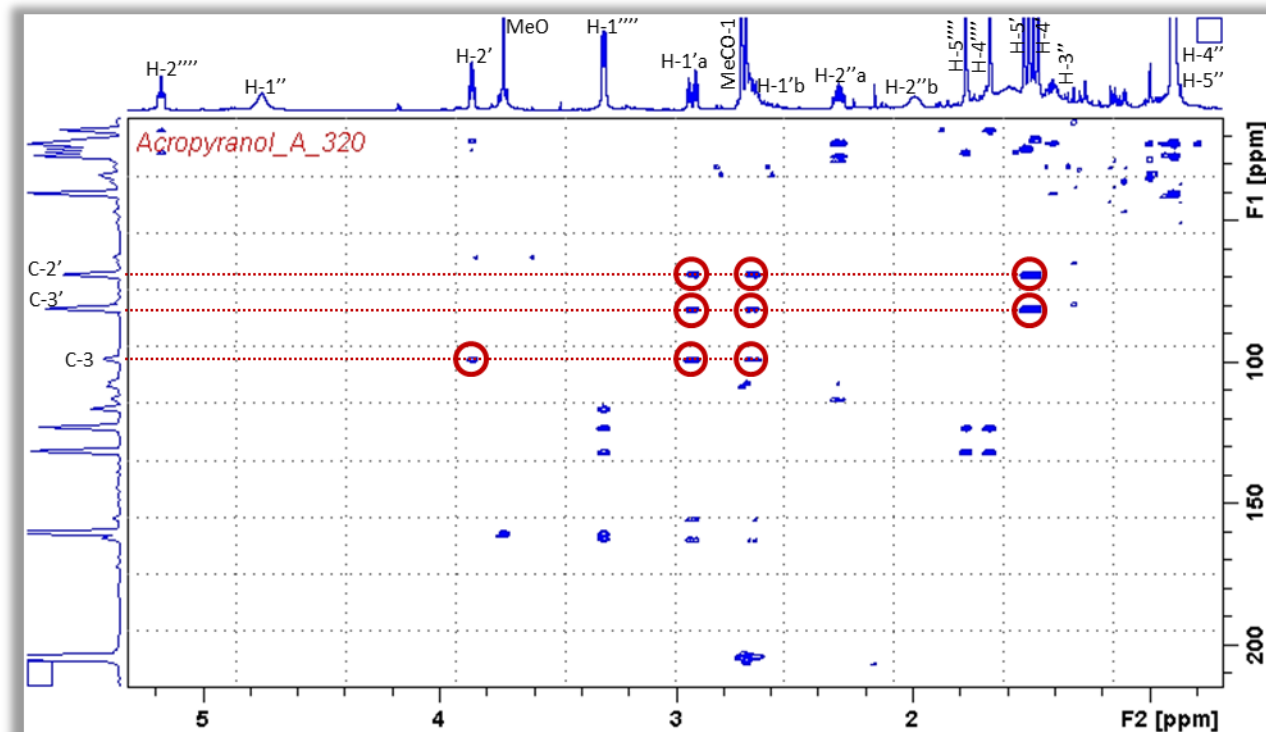
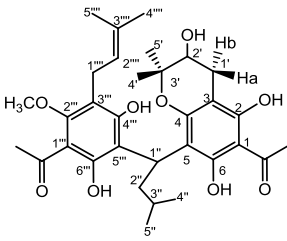


Figure A 14: NOESY (600 MHz, CDCl_3 , 47 °C) spectrum of Acropyranol A (2)

Figure A 15: HSQC (600 MHz, CDCl₃, 47 °C) spectrum of Acropyranol A (2)Figure A 16: HMBC (600 MHz, CDCl₃, 47 °C) spectrum of Acropyranol A (2)

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

Table A 4: NMR spectroscopic data of Acropyranol A (2) at 0 °C; different rotamers' signals are assigned in red and blue

NMR spectroscopic data (600 MHz, 0 °C CDCl ₃) for Acropyranol A				
				
No	¹ H (J, Hz)	¹³ C	HMBC	NOESY
1		106.5		
2		162.5		
3		98.2		
4		155.3		
5		105.9		
6		159.8		
1'a	2.90 brs	25.3	2, 3, 4, 5, 2', 3'	1'b, 2', 4'
1'b	2.64 m			1'a, 2', 5'
2'	3.87 (t, 5.2)	68.3	3	1'a, 1'b, 4', 5'
3'		80.6		
4'	1.46 s	24.2	2, 3, 5'	1'a, 2'
	1.47 s			
5'	1.53 s	20.9	2, 3, 4'	
	1.51 s			1'b, 2'
1''	4.79 brs	27.8	4, 5, 6, 2'', 4''', 5''', 6'''	2'', 3'', 4'', 5'', OH-6, OH-4'''
	4.65 brs			
2''a	2.33 m		5, 1'', 3'', 4'', 5'', 5'''	2''b, 3'', 4'', 5'', OH-6, OH-4'''
2''b	1.96 brs	39.6		2''a, 4'', 5''
	1.81 brs			
3''	1.36 m	26.5	1'', 2'', 4'', 5''	1'', 2''a, 4'', 5''
4''	0.91 s	22.9	2'', 3'', 5''	1'', 2''a, 2''b, 3''
5''	0.85 s	21.8	2'', 3'', 4''	
1'''		108.0		
2'''		159.8		
3'''		116.0		
4'''		161.7		
5'''		112.5		
6'''		159.6		
1''''	3.28 brm	22.8	2''', 3''', 4''', 2''''', 3''''	2''''', 5''''', MeO
2''''	5.15 brs	122.5	3''', 1''', 4''', 5''''	1''''', 4''''', MeO
3''''		131.3		
4''''	1.65 s	25.7	2''''', 3''''', 5''''	2''''
5''''	1.76 s	17.8	2''''', 3''''', 4''''	1''''', MeO
MeO	3.71 s	62.6	2'''	1''''', 2''''', 5''''', MeCO-1'''
MeCO-1	2.69 s	33.6	1, MeCO-1	OH-2, OH-6
MeCO-1		204.7		
MeCO-1'''	2.72 s	30.9	1''', MeCO-1'''	MeO, OH-6'''
MeCO-1'''		203.8		
OH-2	14.16		1, 2, 3	MeCO-1, 1'
	14.09		1, 2, 3	
OH-6	9.90/ 9.88		1	MeCO-1, 1'', 2'', OH-6'''
	9.19/ 9.22		1	
OH-2'				
OH-4'''	9.26/ 9.28		3''', 4'''	4', 1'', 2''
	8.4		3'''	
OH-6'''	15.72/ 15.71		1'''	MeCO-1''', OH-6
	15.66/ 15.67		1''', 5''', 6'''	

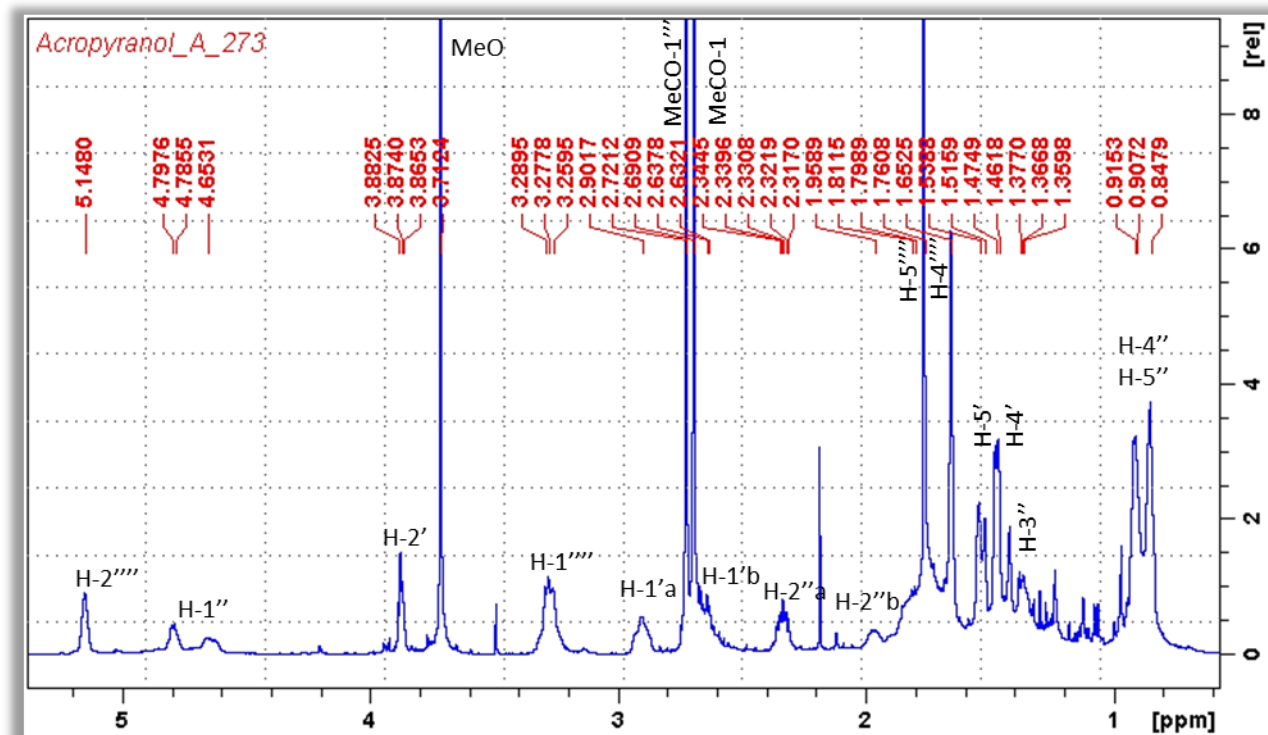


Figure A 17: ^1H NMR (600 MHz, CDCl_3 , 0 $^\circ\text{C}$) spectrum of Acropyranol A (2), δ_{H} 0-5.5

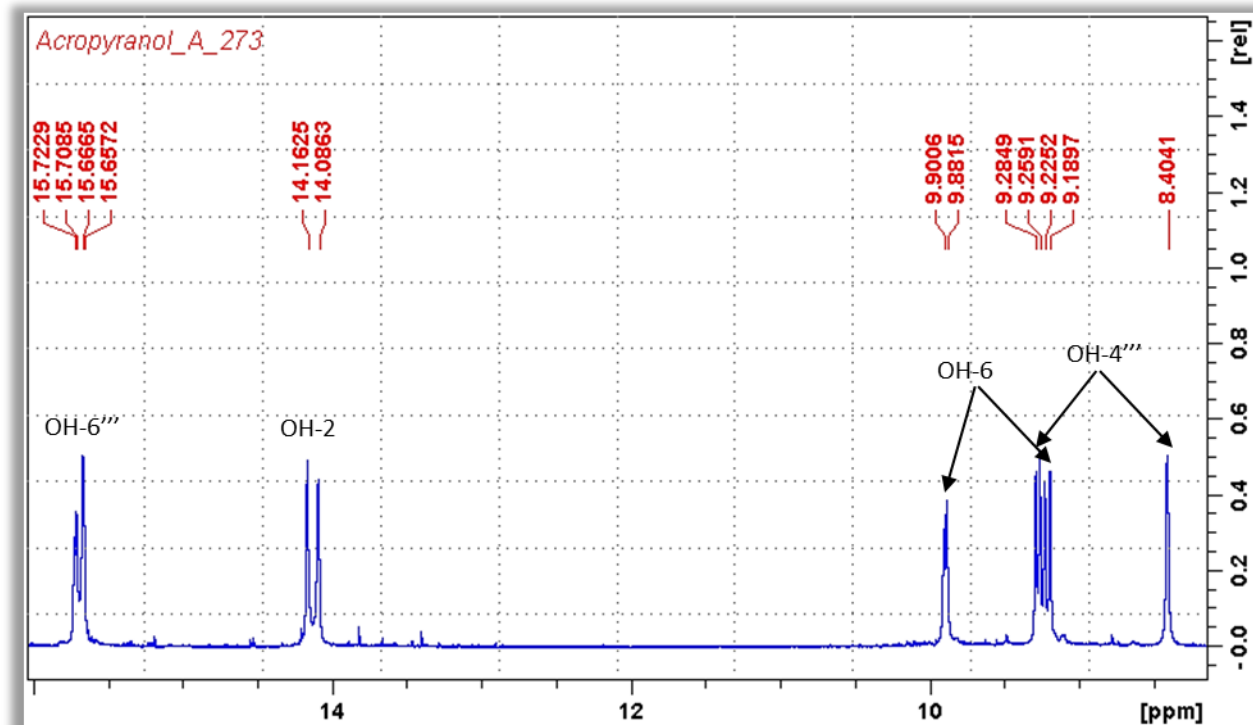


Figure A 18: ^1H NMR (600 MHz, CDCl_3 , 0 $^\circ\text{C}$) spectrum of Acropyranol A (2), δ_{H} 8-16

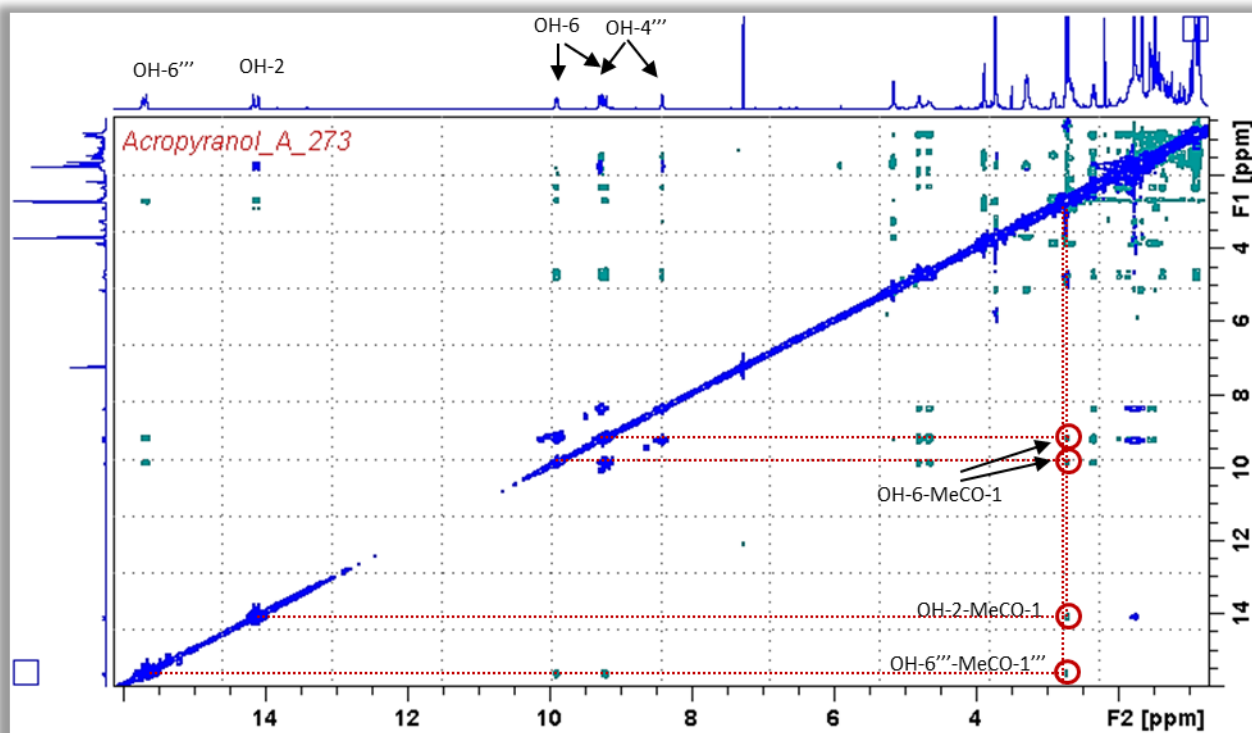


Figure A 19: NOESY (600 MHz, CDCl₃, 0 °C) spectrum of Acropyranol A (2)

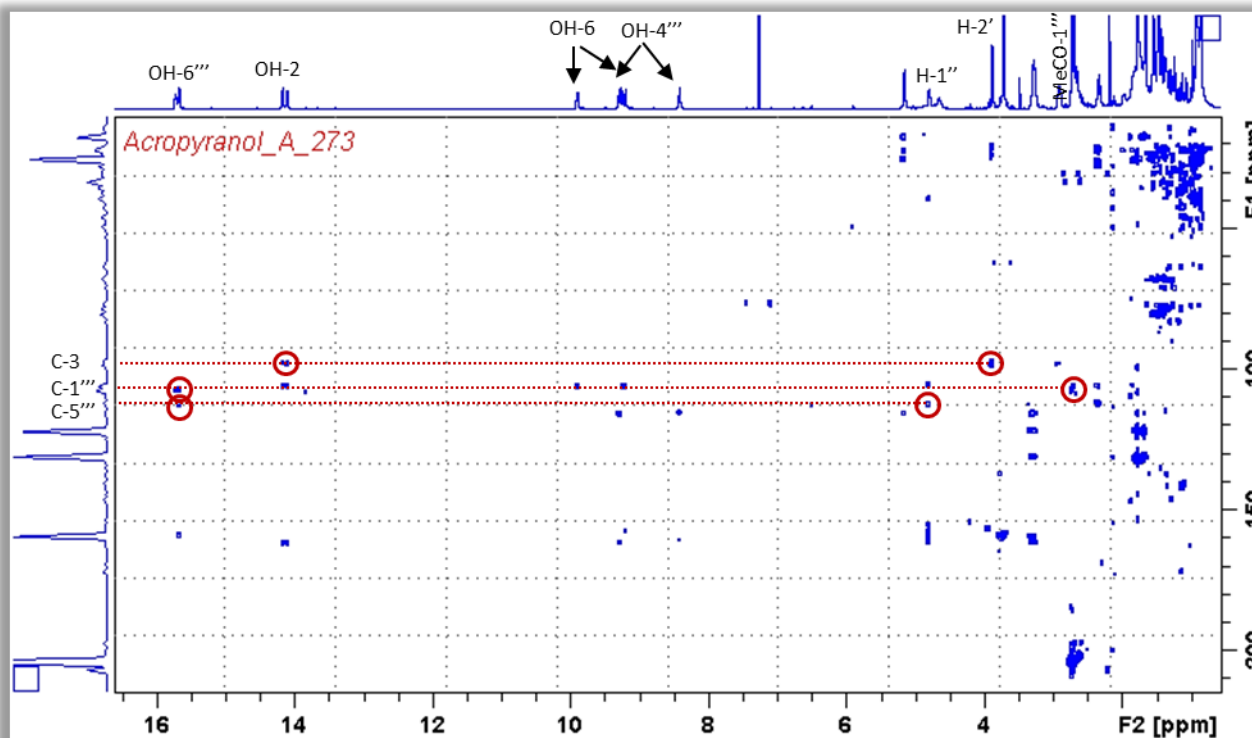
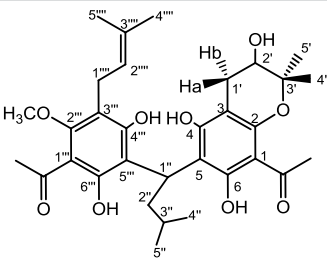


Figure A 20: HMBC (600 MHz, CDCl₃, 0 °C) spectrum of Acropyranol A (2)

Table A 5: NMR spectroscopic data of Acropyranol B (3) at 47 °C

NMR spectroscopic data (600 MHz, 47 °C CDCl ₃) for Acropyranol B				
				
No	¹ H (J, Hz)	¹³ C	HMBC	NOESY
1		104.9		
2		154.8		
3		100		
4		161.9		
5		108.7		
6		161.6		
1'a	2.88 (dd, 17.2/5.0)	26.3	2, 3, 4, 2', 3'	1'b, 2', 4'
1'b	2.62, m		2, 3, 4, 2', 3'	1'a, 2', 5'
2'	3.79, brs	68.7		1'a, 1'b, 4', 5'
3'		78.4		
4'	1.37, s	24.9	2', 3', 5'	1'a, 2'
5'	1.40, s	21.8	2', 3', 4'	1'b, 2'
1''	4.75 (t, 7.5)	28.2	5, 6, 2'', 3'', 5'', 6''	2'', 3'', 4'', 5''
2''	2.16 brs	39.4		1'', 3'', 4'', 5''
3''	1.42 m	27	1'', 2'', 4'', 5''	1'', 2'', 4'', 5''
4''	0.88 (brd, 5.3)	22.4	2'', 3'', 5''	1'', 2'', 3''
5''	0.88 (brd, 5.3)	22.4	2'', 3'', 4''	
1'''		108.2		
2'''		160.2		
3'''		116.6		
4'''		162.5		
5'''		112.9		
6'''		161.6		
1''''	3.30 brd	23.1	2''', 3''', 4''', 2'''', 3''''	2''', 5''', MeO
2''''	5.20 (t, 6.4)	123.2	3''', 1''', 4''', 5''''	1''', 4''''
3''''		131.4		
4''''	1.68 s	25.5	3''', 2''', 3''', 5''''	2''''
5''''	1.76 s	17.8	3''', 2''', 3''', 4''''	1''', MeO
MeO	3.70 s	62.4	2'''	1''', 5''', MeCO-1'''
MeCO-1	2.66 s	32.6	1, MeCO-1	
MeCO-1		203.6		
MeCO-1'''	2.70 s	30.3	1''', MeCO-1'''	MeO
MeCO-1'''		204.1		
OH-4				
OH-6				
OH-2'				
OH-4'''				
OH-6'''				

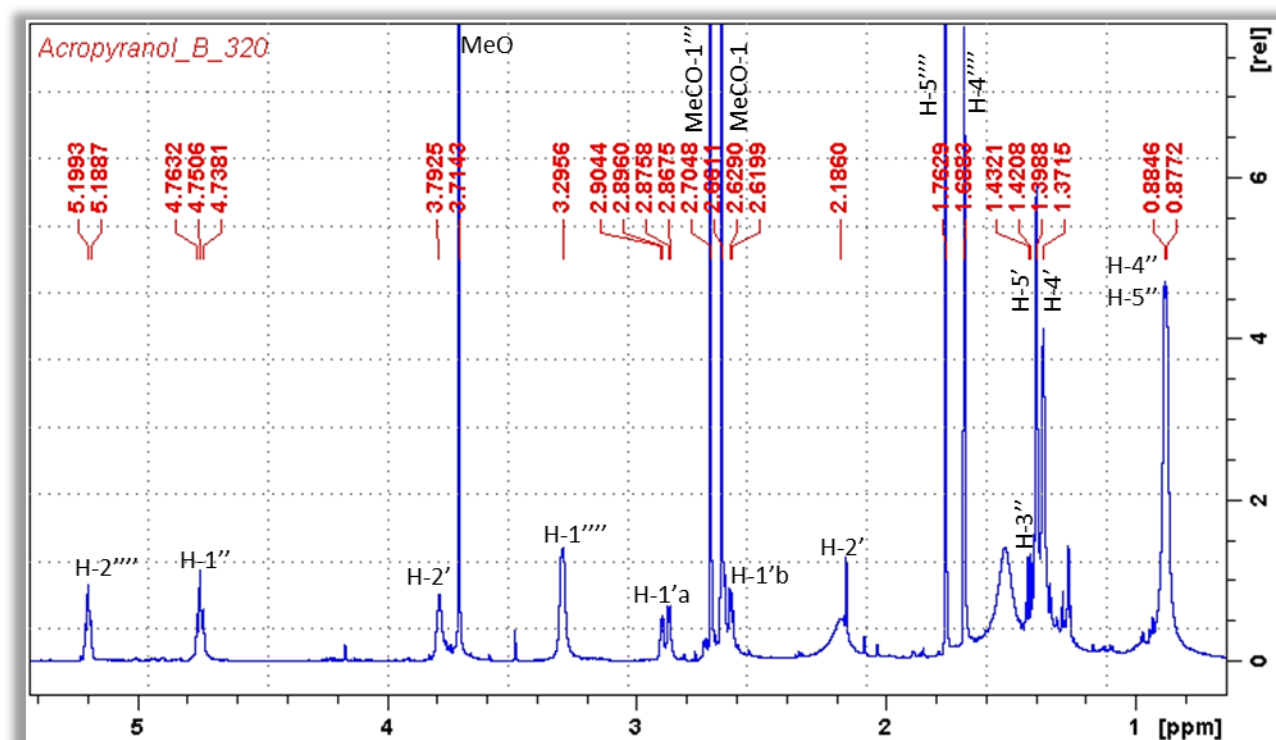


Figure A 21: ^1H NMR (600 MHz, CDCl_3 , 47 $^\circ\text{C}$) spectrum of Acropyranol B (3)

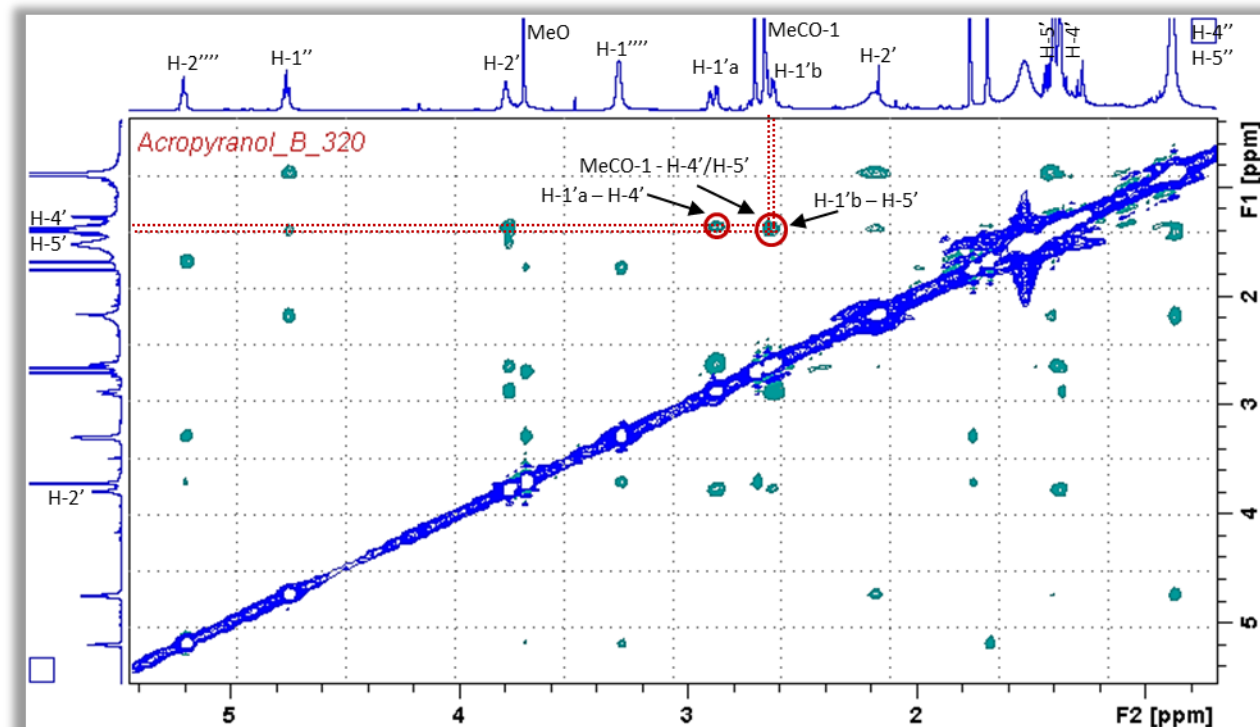
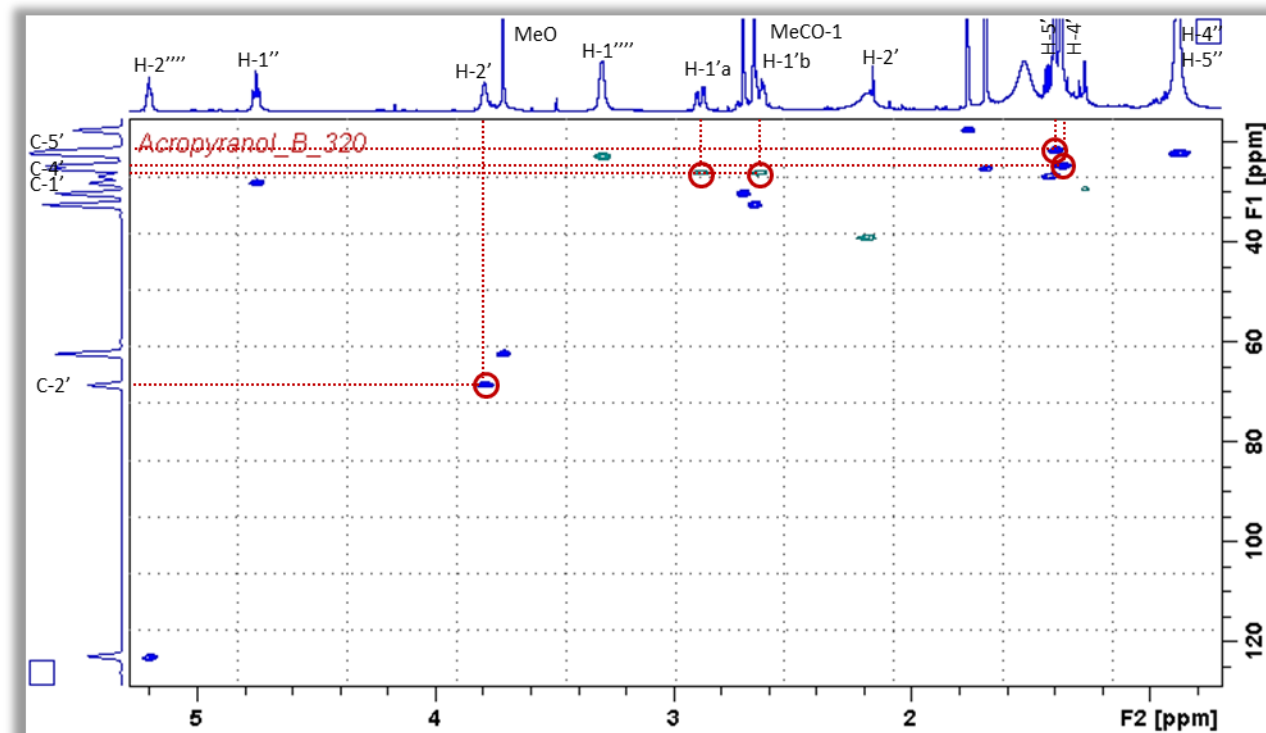
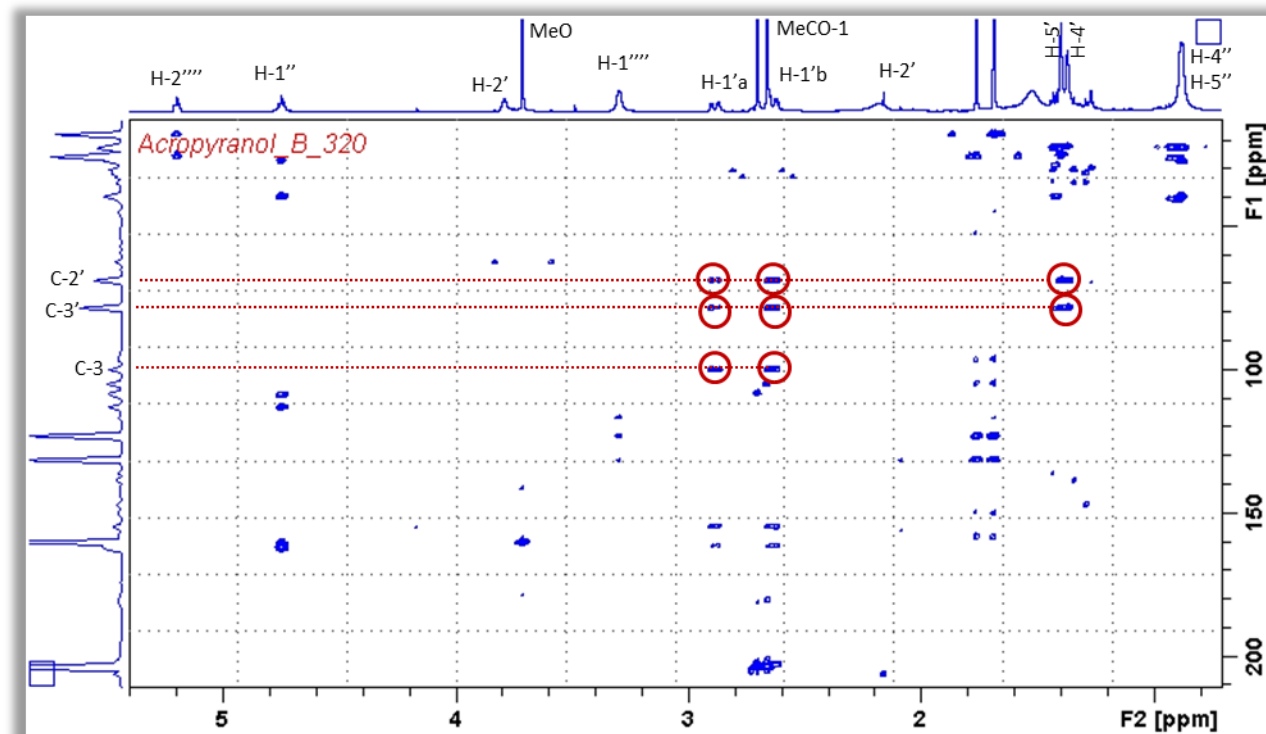
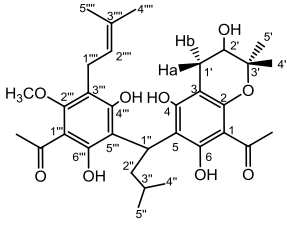


Figure A 22: NOESY (600 MHz, CDCl_3 , 47 $^\circ\text{C}$) spectrum of Acropyranol B (3)

Figure A 23: HSQC (600 MHz, CDCl_3 , 47 °C) spectrum of Acropyranol B (3)Figure A 24: HMBC (600 MHz, CDCl_3 , 47 °C) spectrum of Acropyranol B (3)

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

Table A 6: NMR spectroscopic data of Acropyranol B (3) at 0 °C; different rotamers' signals are assigned in red and blue

NMR spectroscopic data (600 MHz, 0 °C CDCl ₃) for Acropyranol B				
				
No	¹ H (J, Hz)	¹³ C	HMBC	NOESY
1		104.5		
2		154.5		
3		99.3		
4		161.3		
5		107.8		
		108.4		
6		161.4		
1'a	2.83, m	26.1	2, 3, 4, 2', 3'	1'b, 2', 4'
1'b	2.63, m			1'a, 2', 5'
2'	3.78 (t, 4.8)	68.2	3, 5'	1', 2', 4', 5'
	3.82 (t, 4.4)			
3'		78.1		
4'	1.31, s	24.6	2', 3', 5'	1'a, 2'
	1.37, s			
5'	1.38, s	22.2	2', 3', 4'	1'b, 2'
	1.42, s			
1''	4.71 (t, 7.7)	27.9	5, 6, 2'', 3'', 5'', 6''	2'', 3'', 4'', 5'', OH-4, OH-4''
	4.74 (t, 7.8)			
2''a	2.26 m	38.4		
	2.07 m	38.4	5, 3'', 4'', 5'', 5''	1'', 3'', 4'', 5'', H-4, OH-4''
2''b	2.14 m	39.0		
3''	1.39 m	26.7	1'', 2'', 4'', 5''	1'', 2''a, 4'', 5''
4''	0.84 brm	22.4	2'', 3'', 5''	1'', 2'', 3''
5''	0.89 brm	22.4	2'', 3'', 4''	
1'''		107.8		
2'''		159.7		
3'''		116.3		
4'''		162.3		
5'''		112.7		
		112.2		
6'''		160.4		
1''''	3.27 brm	22.8	2''', 3''', 4''', 2''', 3''''	2''', 5''', MeO
2''''	5.17 brs	122.7	3''', 1''', 4''', 5''''	1''', 4''', MeO
3''''		131.4		
4''''	1.68 s	25.6	2''', 3''', 5''''	2''''
5''''	1.75 s	17.9	2''', 3''', 5''''	1''', MeO
MeO	3.70 s	62.5	2'''	1''', 5''', MeCO-1'''
MeCO-1	2.66 s	33.0	1, MeCO-1	OH-6
MeCO-1'		203.7		
MeCO-1'''	2.70 s	30.8	1''', MeCO-1'''	MeO, OH-6'''
MeCO-1''''		203.9		
OH-4	9.27/ 9.28			
	10.10/ 10.11		2, 3, 4, 5	1'', 2'', OH-6'''
OH-6	16.24/ 16.26			
	16.08/ 16.09		1, 5, 6, MeCO-1	1'', MeCO-1, OH-4'''
OH-2'				
OH-4'''	10.18/ 10.21			
	9.33/ 9.35		2''', 3''', 4''', 5''''	1'', 2'', OH-6
OH-6'''	15.63			
	15.77		1''', 5''', 6''', MeCO-1'''	1'', 2''b, MeCO-1''', OH-4

Acropyranol_B_273

16.2558
16.2394
16.0957
16.0753
15.7652
15.6318
10.2144
10.1849
10.1051
10.0964
9.3535
9.3259
9.2843
9.2670

OH-6
OH-6'''
OH-4'''
OH-4

[ppm]

75

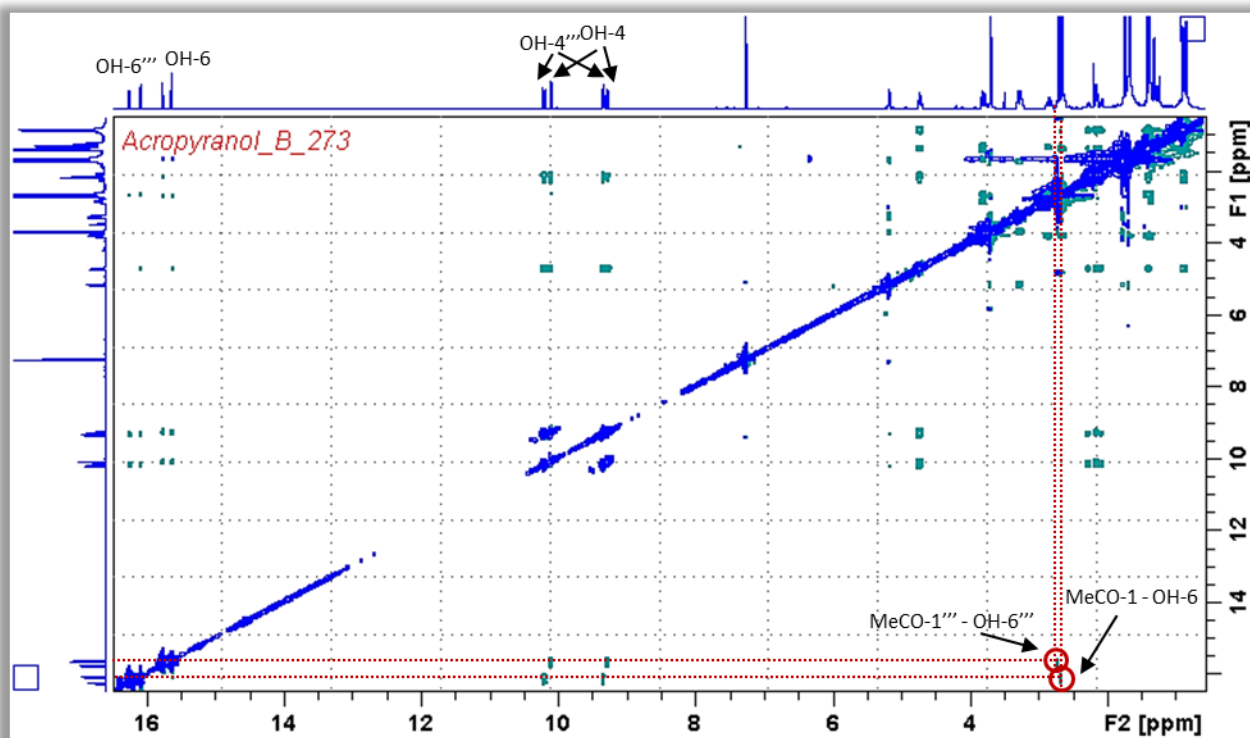


Figure A 27: NOESY (600 MHz, CDCl₃, 0 °C) spectrum of Acropyranol B (3)

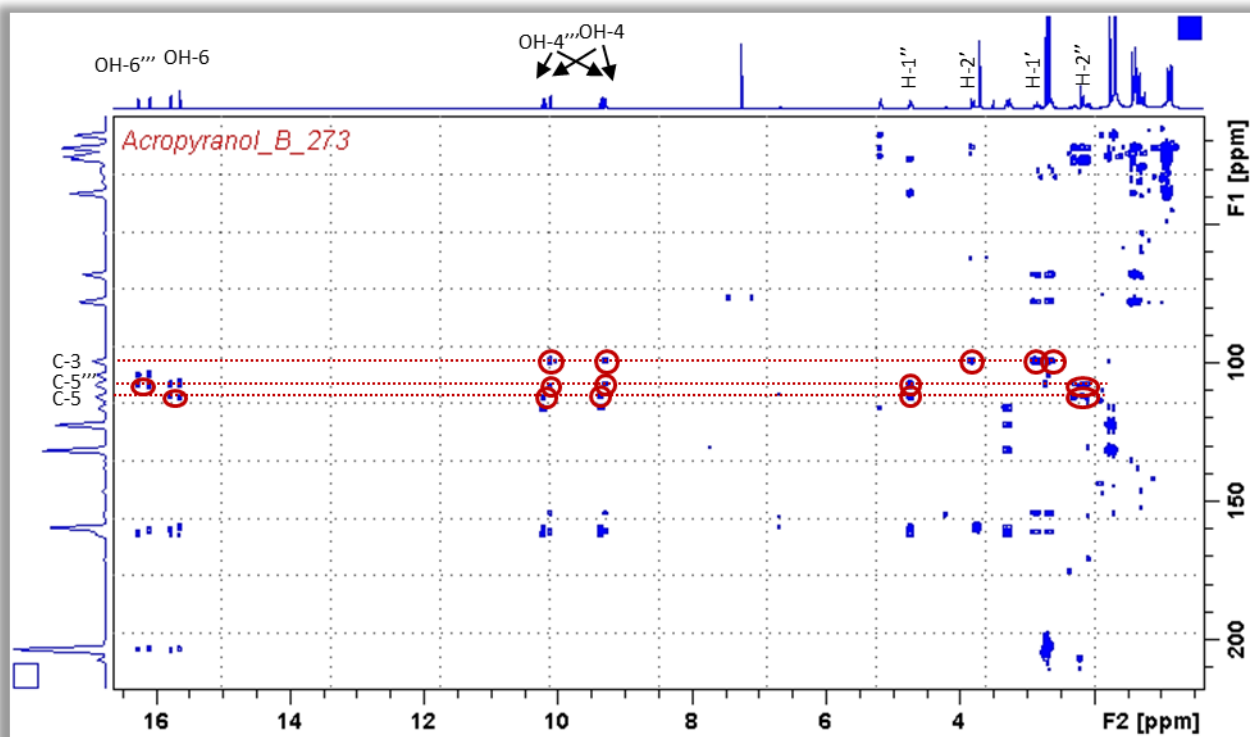
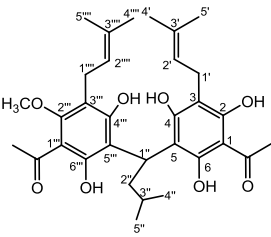


Figure A 28: HMBC (600 MHz, CDCl₃, 0 °C) spectrum of Acropyranol B (3)

Table A 7: NMR spectroscopic data of Acrovestone (4) at 47 °C

NMR spectroscopic data (600 MHz, 47 °C CDCl ₃) for Acrovestone				
				
No	¹ H (J, Hz)	¹³ C	HMBC	NOESY
1		104.7		
2		158.2		
3		106.4		
4		160.8		
5		108.7		
6		159.1		
1'	3.42 (d, 7)	22.2	2, 3, 4, 2', 3', 4', 5'	2', 5'
2'	5.23 (tt, 1.3/7)	121.6	3, 1', 4', 5'	1', 4'
3'		136.6		
4'	1.79 s	25.8	3, 2', 3', 5'	2'
5'	1.84 s	17.7	3, 2', 3', 4'	1'
1''	4.75 (t, 7.3)	28.7	4, 5, 2'', 3'', 4''', 5'''	2'', 3'', 4'', 5''
2''	2.15 brs 2.24 brs	39.5		1'', 3'', 4'', 5''
3''	1.42 (m, 6.6)	27.1	1'', 2'', 4'', 5''	1'', 2'', 4'', 5''
4''	0.88 (d, 6.6)	22.5	2'', 3'', 5''	1'', 2'', 3''
5''	0.88 (d, 6.6)	22.5	2'', 3'', 4''	
1'''		108.2		
2'''		160.3		
3'''		116.7		
4'''		162.6		
5'''		113.2		
6'''		160.3		
1''''	3.31 (d, 6.4)	23.0	2''', 3''', 4''', 2''''', 3''''', 4''''	2''', 5''', MeO
2''''	5.21 (tt, 1.5/6.4)	123.1	3''', 1''', 4''', 5''''	1''', 4''''
3''''		131.7		
4''''	1.69 s	25.6	3''', 2''', 3''', 5''''	2''''
5''''	1.77 s	17.9	3''', 2''', 3''', 4''''	1''', MeO
MeO	3.72 s	62.5	2''''	1''', 5''', MeCO-1''''
MeCO-1	2.68 s	32.5	1, MeCO-1	
MeCO-1		204.1		
MeCO-1'''	2.71 s	30.4	1''', MeCO-1'''	MeO
MeCO-1'''		204.1		
OH-2				
OH-4				
OH-6				
OH-4'''				
OH-6'''				

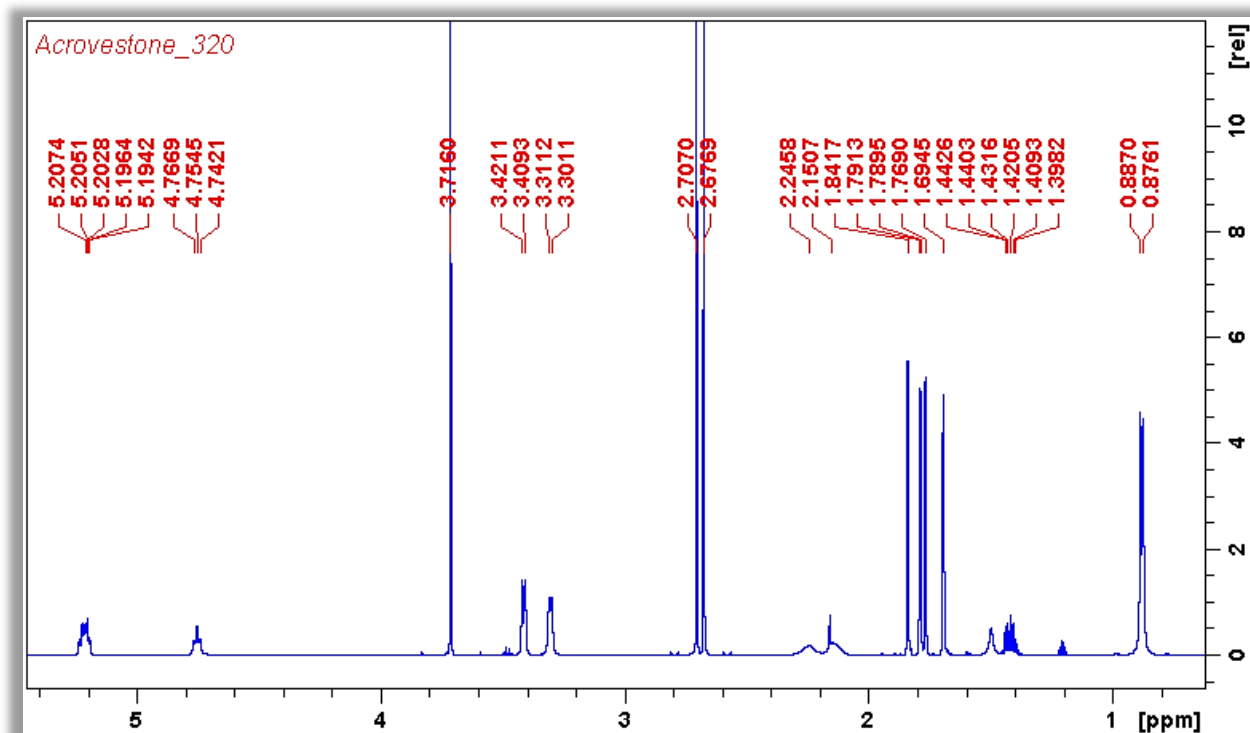


Figure A 29: ¹H NMR (600 MHz, CDCl₃, 47 °C) spectrum of Acrovestone (4)

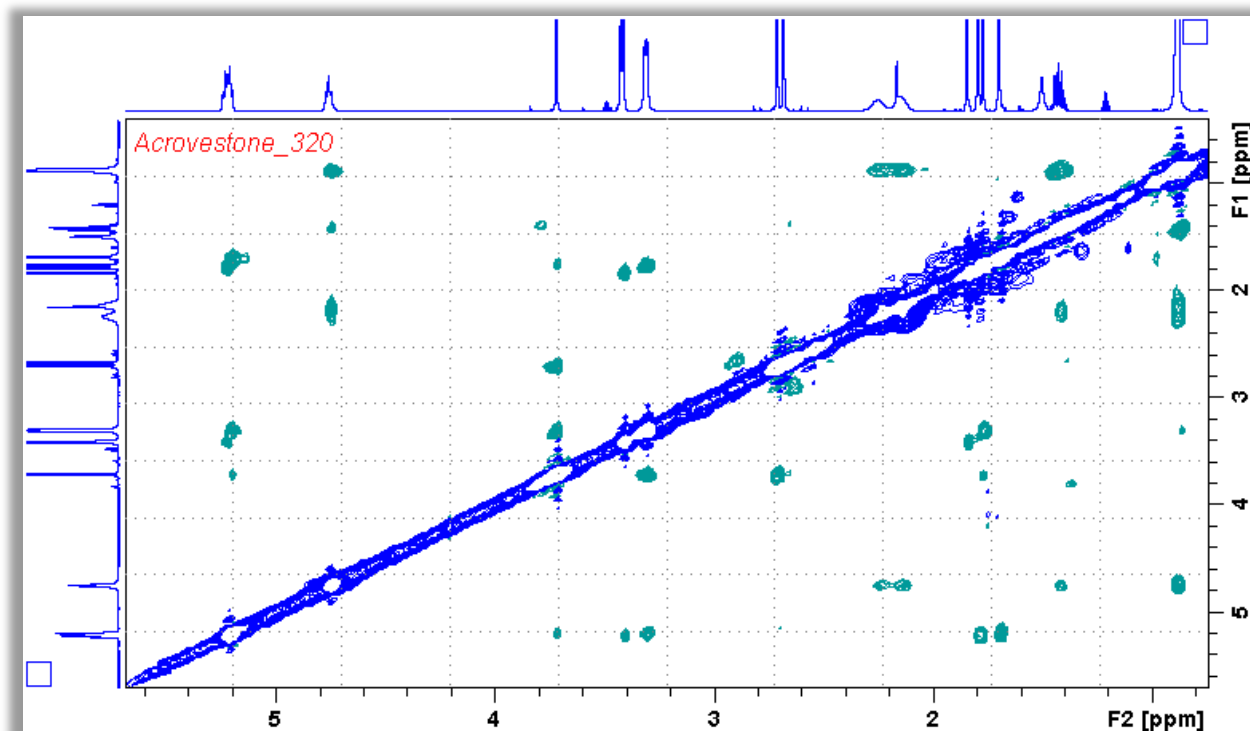


Figure A 30: NOESY (600 MHz, CDCl₃, 47 °C) spectrum of Acrovestone (4)

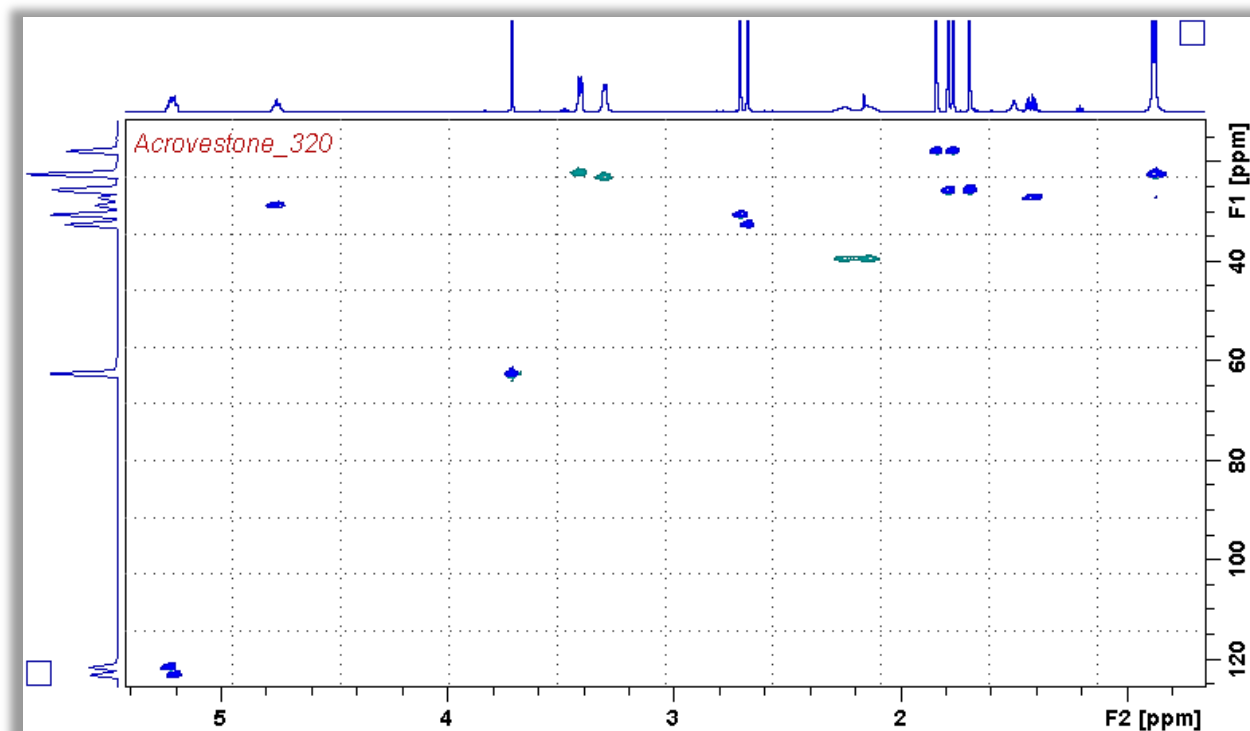


Figure A 31: HSQC (600 MHz, CDCl₃, 47 °C) spectrum of Acrovestone (4)

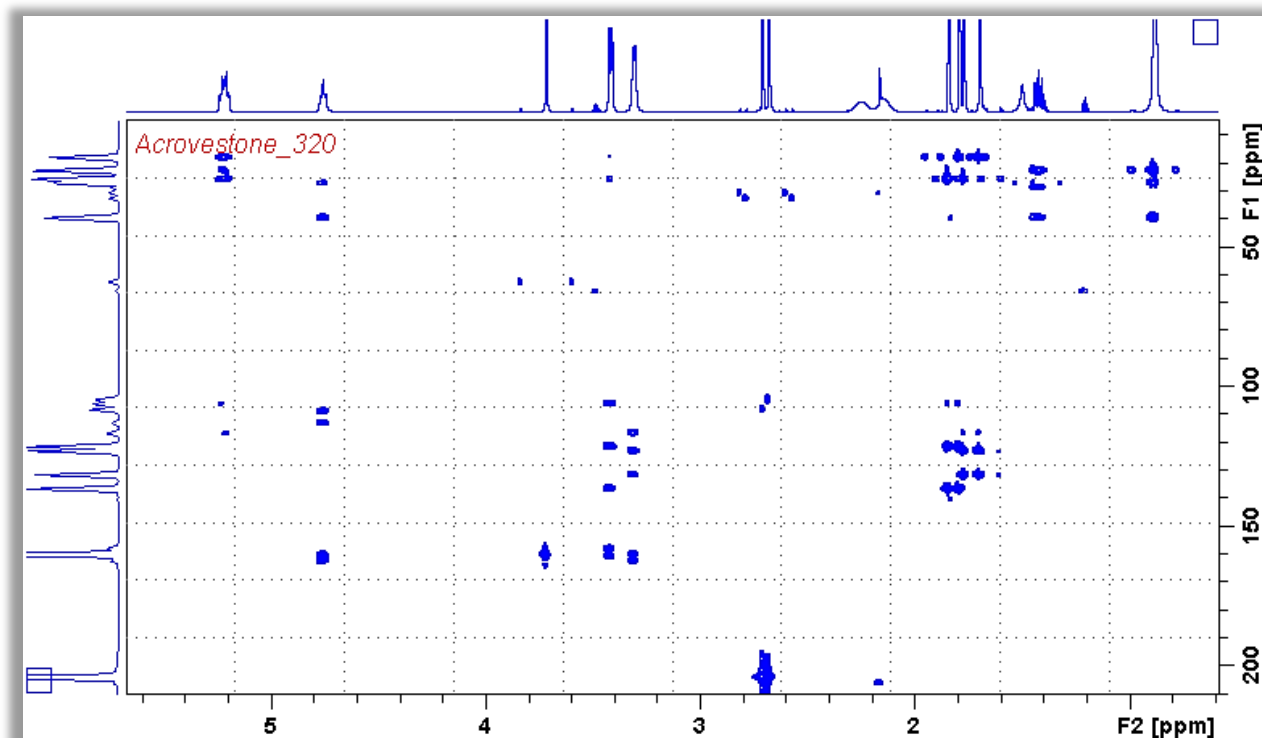
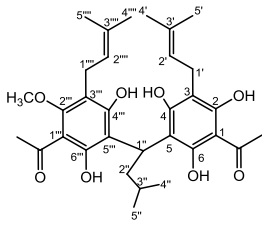


Figure A 32: HMBC (600 MHz, CDCl₃, 47 °C) spectrum of Acrovestone (4)

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

Table A 8: NMR spectroscopic data of Acrovestone (4) at 0 °C; different rotamers' signals are assigned in red and blue

NMR spectroscopic data (600 MHz, 0 °C CDCl ₃) for Acrovestone				
				
No	¹ H (J, Hz)	¹³ C	HMBC	NOESY
1		104.4		
2		158.1		
3		106.1		
4		160.9		
5		108.4		
		109.0		
6		159.1		
1'	3.40 brd 3.39 (d, 6.6)	22.1	2, 3, 4, 2', 3'	2', 5', OH-2, OH-4
2'	5.19 brt	121.5	3, 1', 4', 5'	1', 4', OH-2, OH-4
3'		137.0		
4'	1.78 s	25.9	3, 2', 3', 5'	2'
5'	1.84 s	17.9	3, 2', 3', 4'	1', OH-2
1''	4.71 (t, 7.7) 4.75 (t, 7.7)	28.2	4, 5, 2'', 3'', 4'', 5''	2'', 3'', 4'', 5'', OH-4, OH-4'', OH-6, OH-6''
2''a	2.19 m	39.3		
	2.31 m	39.5		
2''b	2.12 m	39.3	5, 1'', 3'', 4'', 5'', 5''	1'', 3'', 4'', 5'', OH-4, OH-4''
	2.03 m	39.5		
3''	1.38 (m, 3.1)	26.7	1'', 2'', 4'', 5''	1'', 2'', 4'', 5''
4''	0.89 (d, 6.4)	22.5	2'', 3'', 5''	1'', 2'', 3''
5''	0.84 (d, 6.2)	22.4	2'', 3'', 4''	1'', 2'', 3''
1'''		107.9		
2'''		160.0		
3'''		116.6		
4'''		162.6		
5'''		113.4		
		113.0		
6'''		160.3		
1''''	3.27 brm	23.0	2''', 3''', 4''', 2''', 3''''	2''', 5''', MeOH, OH-4'''
2''''	5.18 brt	122.9	3''', 1''', 4''', 5''''	1''', 4''', MeOH, OH-4'''
3''''		132.1		
4''''	1.68 s	25.8	3''', 2''', 3''', 5''''	2''''
5''''	1.76 s	18.0	3''', 2''', 3''', 4''''	1''', MeO
MeO	3.70 s	62.7	2''''	1''', 2''', 5''', MeCO-1'''
MeCO-1	2.66 s	32.9	1, MeCO-1	OH-2, OH-6
MeCO-1		204.3		
MeCO-1'''	2.71 s	30.8	1''', MeCO-1'''	MeOH, OH-6'''
MeCO-1'''		204.5		
OH-2	6.48 s			1', 2', 5', MeCO-1
OH-4	9.27 s 10.07 s		2, 3, 4, 5	1', 2', 1'', 2'', OH-6'''
OH-6	15.89 s 16.06 s			1'', 2'', 3'', OH-4''', MeCO-1
OH-4'''	10.25 s 9.40 s		3''', 4''', 5''''	1'', 2'', 1''', 2''', OH-6
OH-6'''	15.63 s 15.75 s		1''', 5''', 6'''	1'', 2'', 3'', OH-4, MeCO-1'''

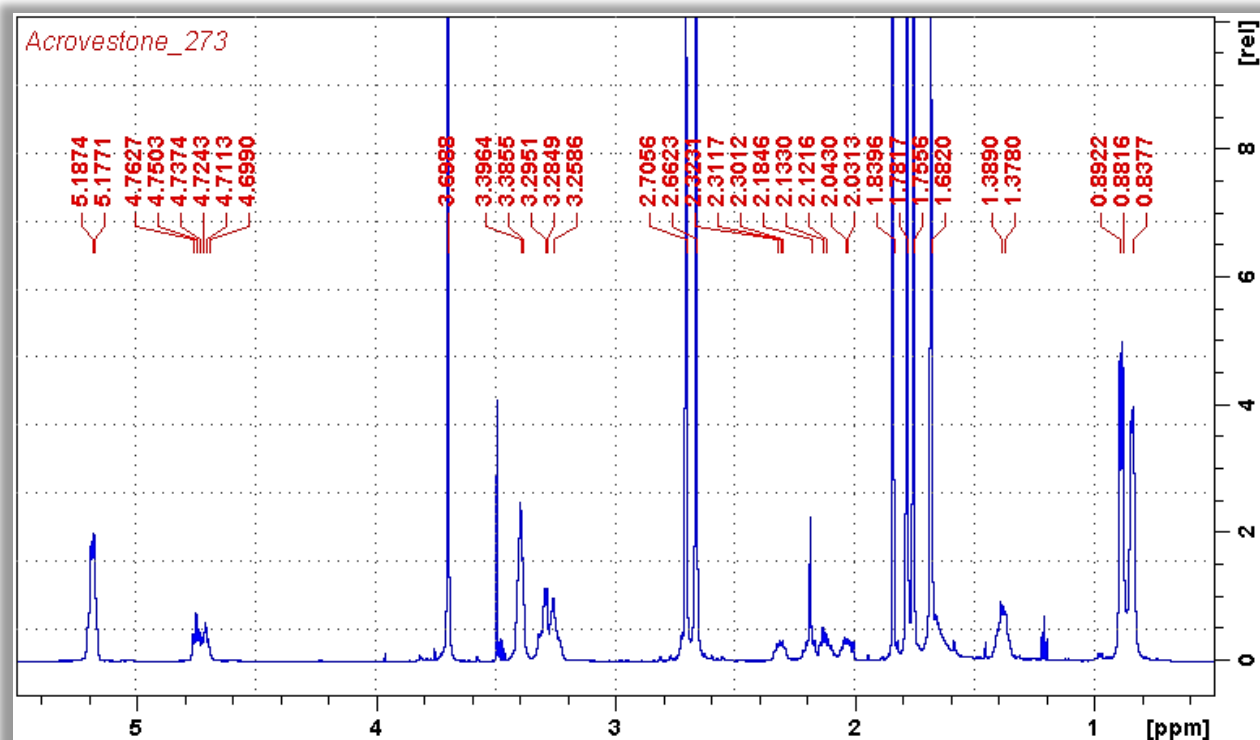


Figure A 33: ^1H NMR (600 MHz, CDCl_3 , 0°C) spectrum of Acrovestone (4), δ_{H} 0-5.5

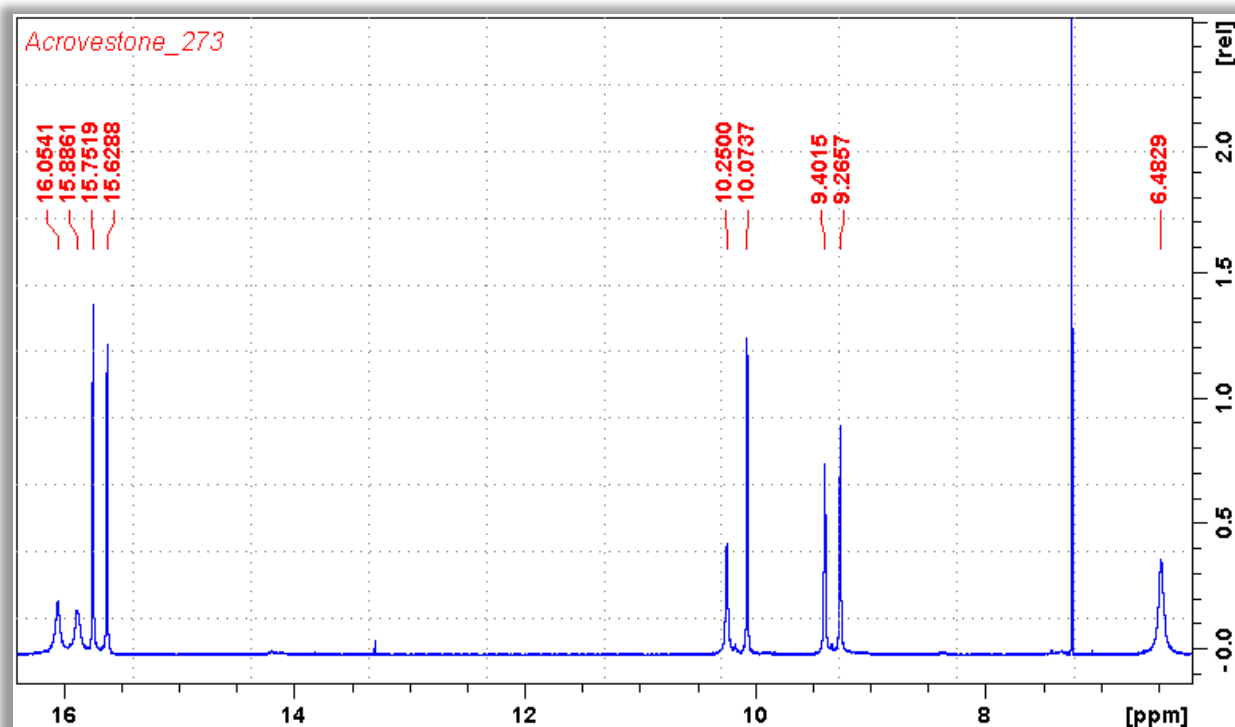


Figure A 34: ^1H NMR (600 MHz, CDCl_3 , 0°C) spectrum of Acrovestone (4), δ_{H} 6-16.2

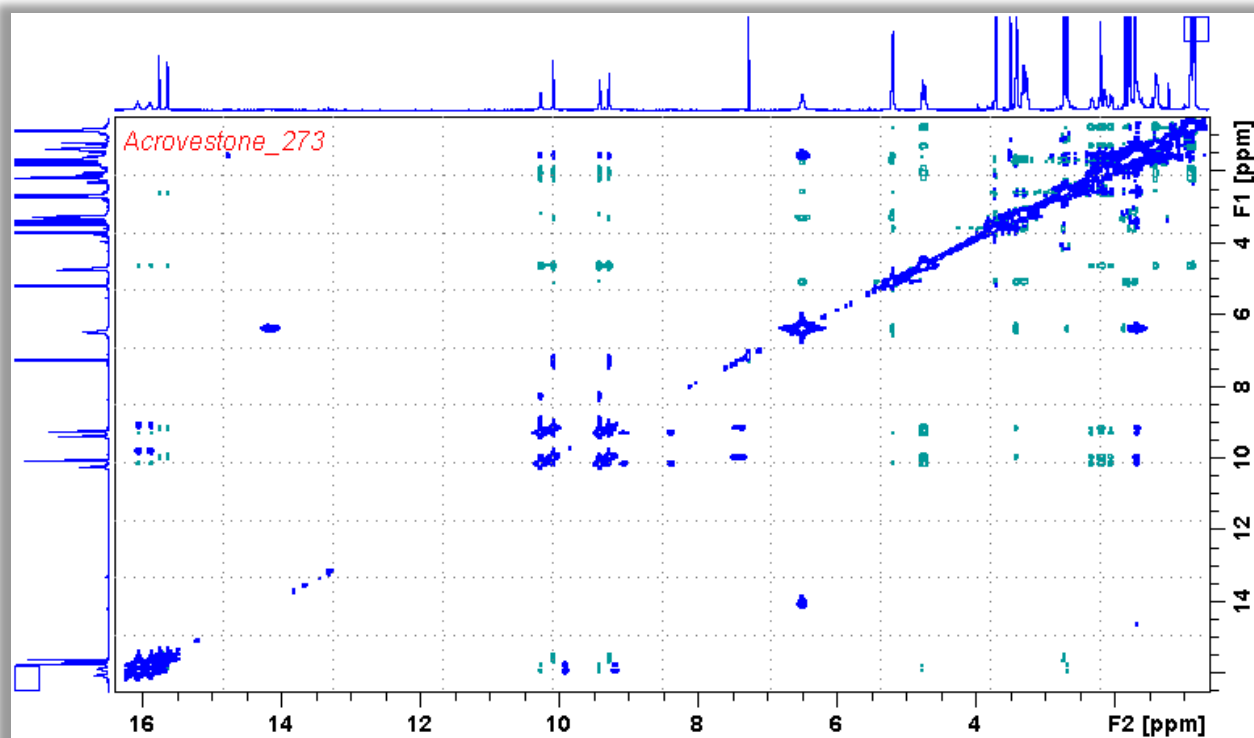


Figure A 35: NOESY (600 MHz, CDCl₃, 0 °C) spectrum of Acrovestone (4)

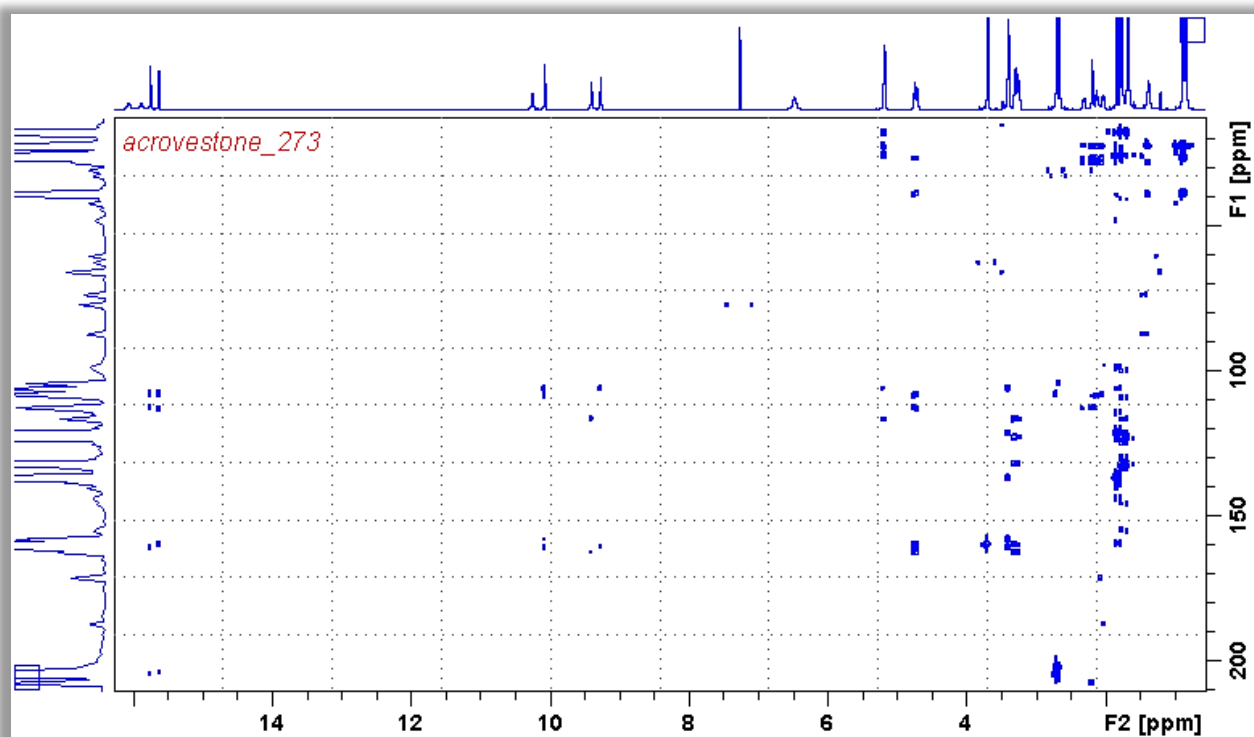


Figure A 36: HMBC (600 MHz, CDCl₃, 0 °C) spectrum of Acrovestone (4)

Table A 9: NMR spectroscopic data of Acrovestenol (5) at 47 °C

NMR spectroscopic data (600 MHz, 47 °C CDCl ₃) for Acrovestenol				
No	¹ H (J, Hz)	¹³ C	HMBC	NOESY
1		105.5		
2		159.7		
3		106.9		
4		161.8		
5		108.5		
6		161.1		
1'a	3.15 brs	29.4		1'b, 2', 4'
1'b	2.74 m			1'a
2'	4.33 (t, 6.5)	78.2	3, 1', 3', 4', 5'	1'a, 4', 5'a
3'		146.7		
4'	1.86 s	18.5	2', 3', 5'	1'a, 2', 5'b
5'a	5.05 brs	110.0	2', 4'	2', 5'b
5'b	4.88 brs			4', 5'a
1''	4.75 (t, 7.0)	28.8	4, 5, 2'', 3'', 4''', 5'''	2'', 3'', 4'', 5''
2''a	2.13 brs	39.6	4'', 5'', 5'''	1'', 4'', 5''
2''b	2.26 m		5, 1'', 3'', 4'', 5'', 5'''	
3''	1.42 m	27.1	1'', 2'', 4'', 5''	1'', 2'', 4'', 5''
4''	0.88 (d, 5.6)	22.5	2'', 3'', 5''	1'', 2'', 3''
5''	0.88 (d, 5.6)	22.5	2'', 3'', 4''	
1'''		108.2		
2'''		160.2		
3'''		116.8		
4'''		162.8		
5'''		113.2		
6'''		160.6		
1''''	3.31 (d, 5.6)	23.1	2''', 3''', 4''', 2''''', 3''''	2''''', 5''''', MeO
2''''	5.21 (t, 6.5)	123.3	3''', 1''''', 4''''', 5''''	1''''', 4''''', MeO
3''''		131.3		
4''''	1.69 s	25.8	3''', 2''''', 3''''', 5''''	2''''
5''''	1.77 s	17.9	3''', 2''''', 3''''', 4''''	1''''', MeO
MeO	3.71 s	62.6	2'''	1''''', 2''''', 5''''', MeCO-1'''
MeCO-1	2.71 s	30.6	1, MeCO-1	
MeCO-1		204.1		
MeCO-1'''	2.72 s	32.8	1''''', MeCO-1'''	MeO
MeCO-1'''		204.1		
OH-2				
OH-4				
OH-6				
OH-4'''				
OH-6'''				

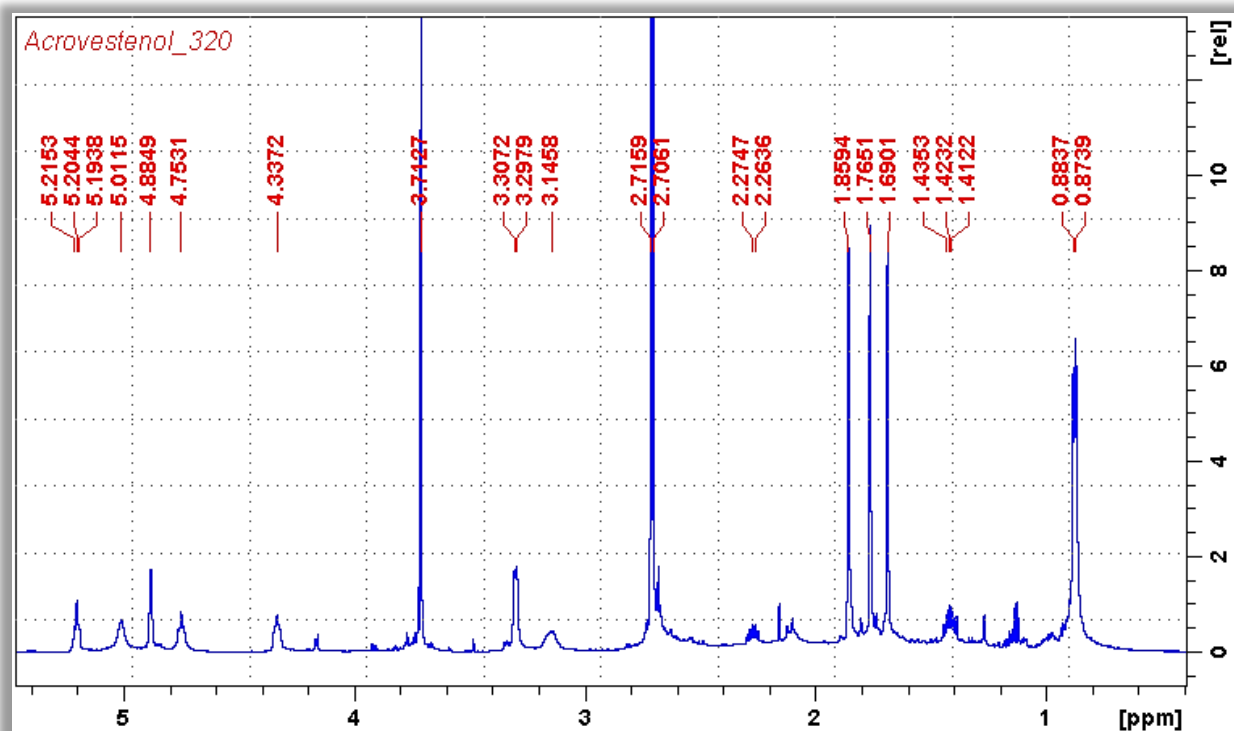


Figure A 37: ^1H NMR (600 MHz, CDCl_3 , 47 °C) spectrum of Acrovestenol (5)

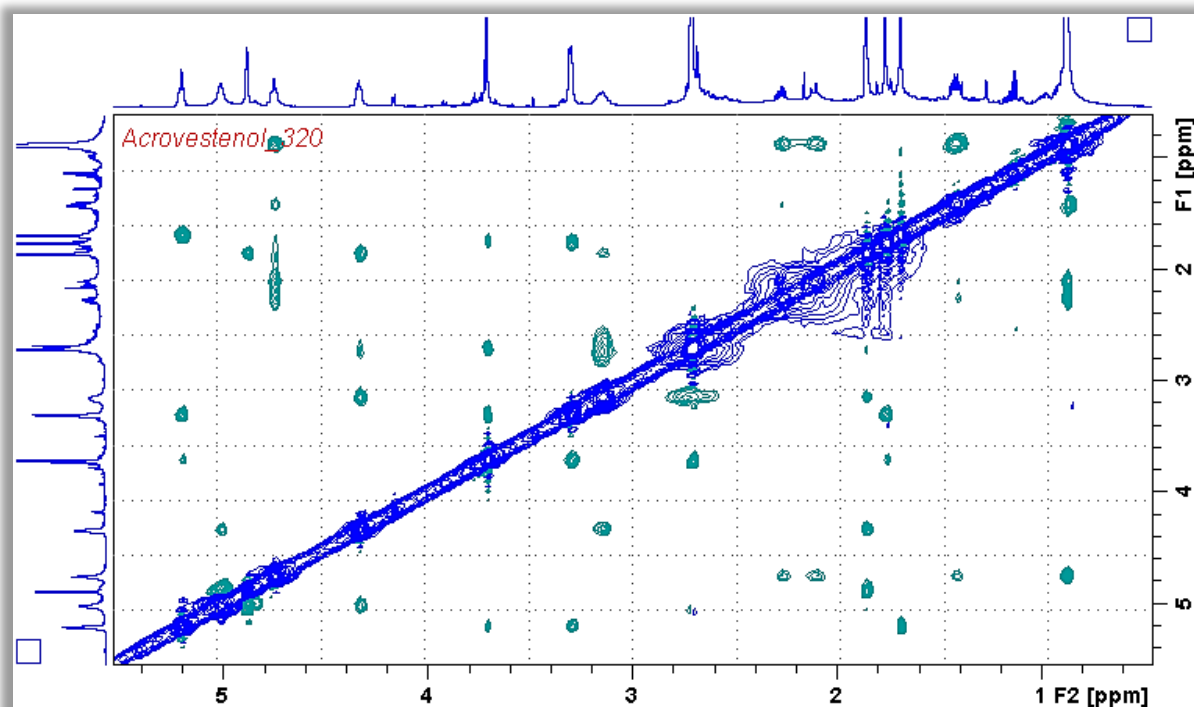


Figure A 38: NOESY (600 MHz, CDCl_3 , 47 °C) spectrum of Acrovestenol (5)

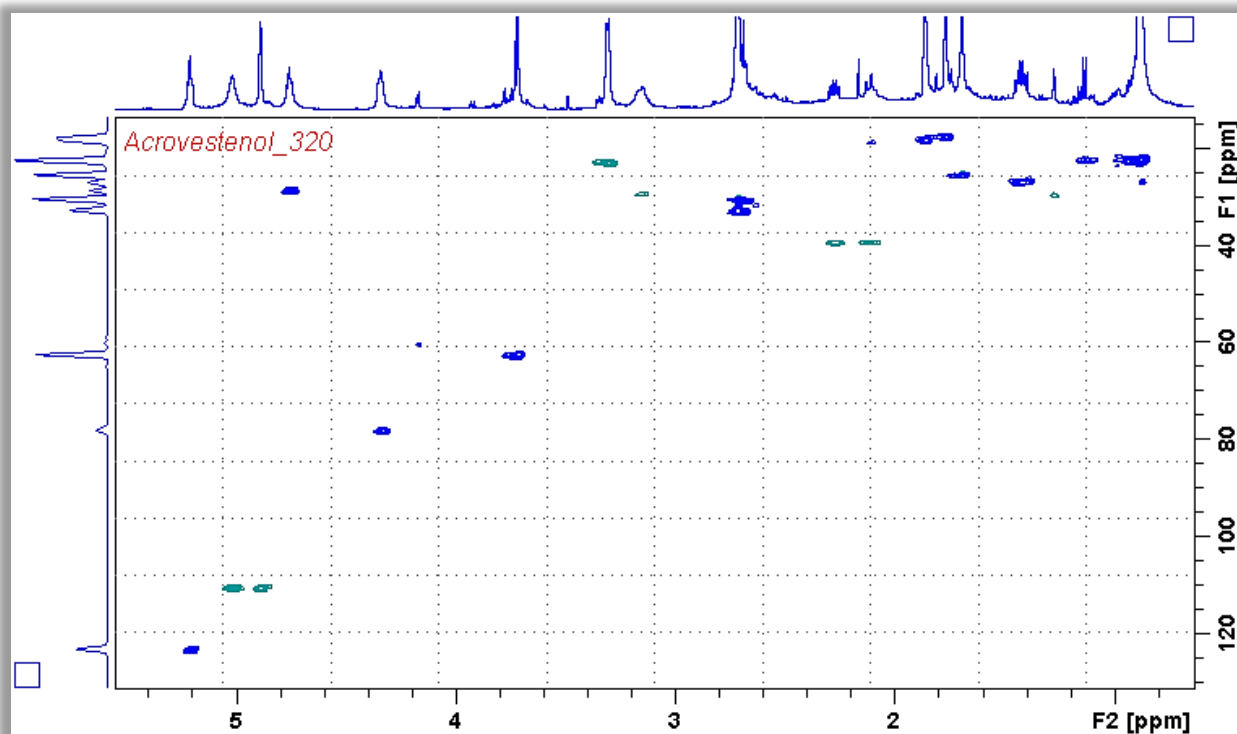


Figure A 39: HSQC (600 MHz, CDCl₃, 47 °C) spectrum of Acrovestenol (5)

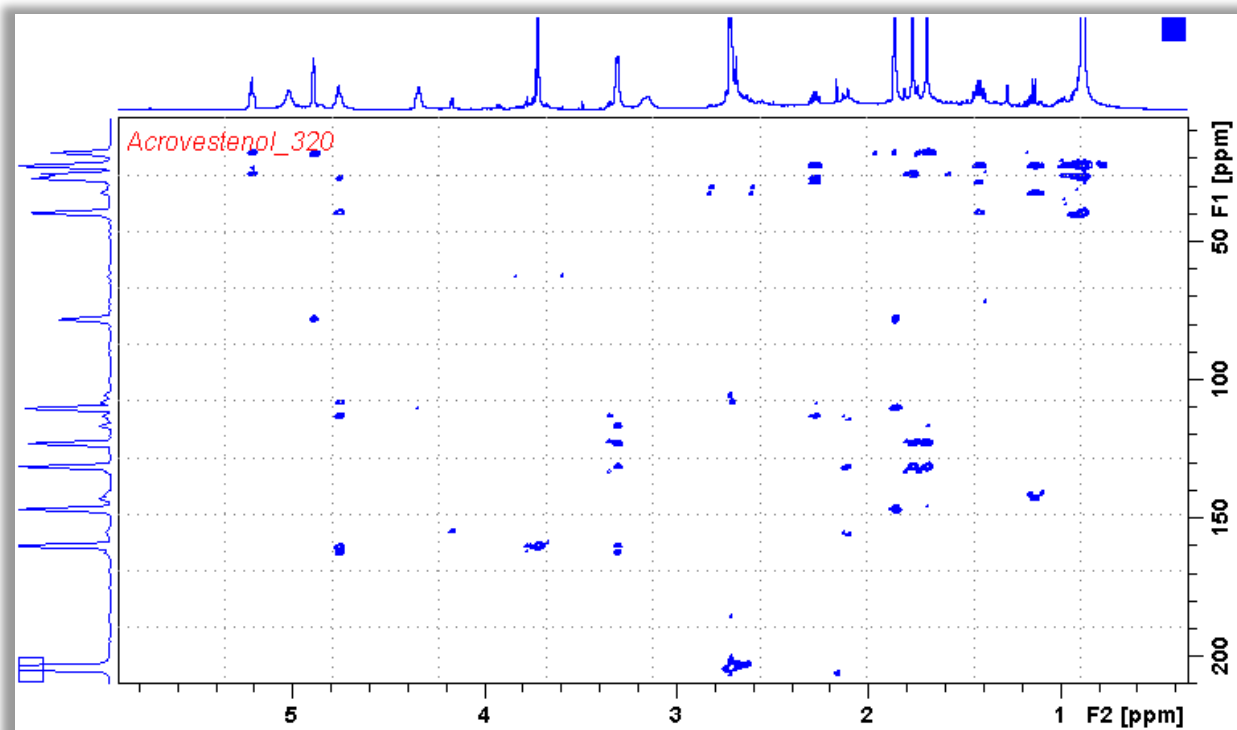
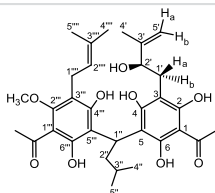


Figure A 40: HMBC (600 MHz, CDCl₃, 47 °C) spectrum of Acrovestenol (5)

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

Table A 10: NMR spectroscopic data of Acrovestenol (5) at 0 °C; different rotamers' signals are assigned in red and blue

NMR spectroscopic data (600 MHz, 0 °C CDCl ₃) for Acrovestenol				
				
No	¹ H (J, Hz)	¹³ C	HMBC	NOESY
1		105.5		
2		159.7		
3		106.9		
4		161.8		
5		108.0		
		108.5		
6		161.1		
1'a	3.07 brt	29.2	2, 3, 4, 2'	1'b, 2', 4'
	3.18 brd	29.6	2, 3, 4	
1'b	2.73 m	29.2	2'	1'a
	2.48 m	29.6	2, 3, 4, 2'	
2'	4.29 brs	78.0	3, 1', 4', 5'	1'a, 4', 5'a
		78.8		
3'		147.0		
4'	1.84 s			
	1.86 s	18.5	2', 3', 5'	1'a, 2', 5'b
5'a	4.97 (d, 8.8)	110.3		2', 5'b
	5.08 brs	110.6	2', 3', 4'	
5'b	4.85 brs	110.3		4', 5'a
	4.89 brs	110.6	3', 4'	
1''	4.70 m	28.1	5, 2'', 3'', 4'', 5'', 6''	2'', 3'', 4'', 5'', OH-4, OH-4'', OH-6
	4.74 m			
2''a	2.26 m	39.1		
	2.42 m	38.8	5, 1'', 3'', 4'', 5'', 5'''	1'', 4'', 5''
2''b	1.98 m	39.1		
	1.91 m	38.8		
3''	1.40 brs	26.5	2'', 4'', 5''	4'', 5''
4''	0.88 (d, 6.7)	22.6	2'', 3'', 5''	1'', 2'', 3''
5''	0.84 (d, 6.7)	22.6	2'', 3'', 4''	1'', 2'', 3''
1'''		108.2		
2'''		160.2		
3'''		116.8		
4'''		162.8		
5'''		113.6		
		113.0		
6'''		160.6		
1''''	3.27 brm	22.9	2''', 3''', 4''', 2'''', 3''''	2''', 5''', MeOH
2''''	5.17 brt	122.9	3''', 1''', 4''', 5''''	1''', 4''', MeOH
3''''		132.1		
4''''	1.68 s	25.5	3''', 2''', 3''', 5''''	2''''
5''''	1.75 s	18.0	3''', 2''', 3''', 4''''	1''', MeO
MeO	3.69 s	62.7	2''	1''', 2''', 5''', MeCO-1'''
MeCO-1	2.70 s	32.9	1, MeCO-1	OH-2, OH-6
MeCO-1		204.3		
MeCO-1'''	2.70 s	30.8	1''', MeCO-1'''	MeOH, OH-6'''
MeCO-1'''		204.5		
OH-2	9.36/ 9.40		1, 2	1', 2', 5', MeCO-1
	9.40/ 9.43			
OH-4	9.94/ 9.96		3, 4	1'', OH-6'''
	9.13/ 9.18			
OH-6	16.00/ 16.02		1, 5, 6	1'', 2'', OH-4''', MeCO-1
	16.17/ 16.19			
OH-4'''	10.33/ 10.36		3''', 4''', 5'''	1'', OH-6
	9.48/ 9.52			
OH-6'''	15.56/ 15.61		1''', 5''', 6'''	OH-4, MeCO-1'''
	15.69/ 15.72			

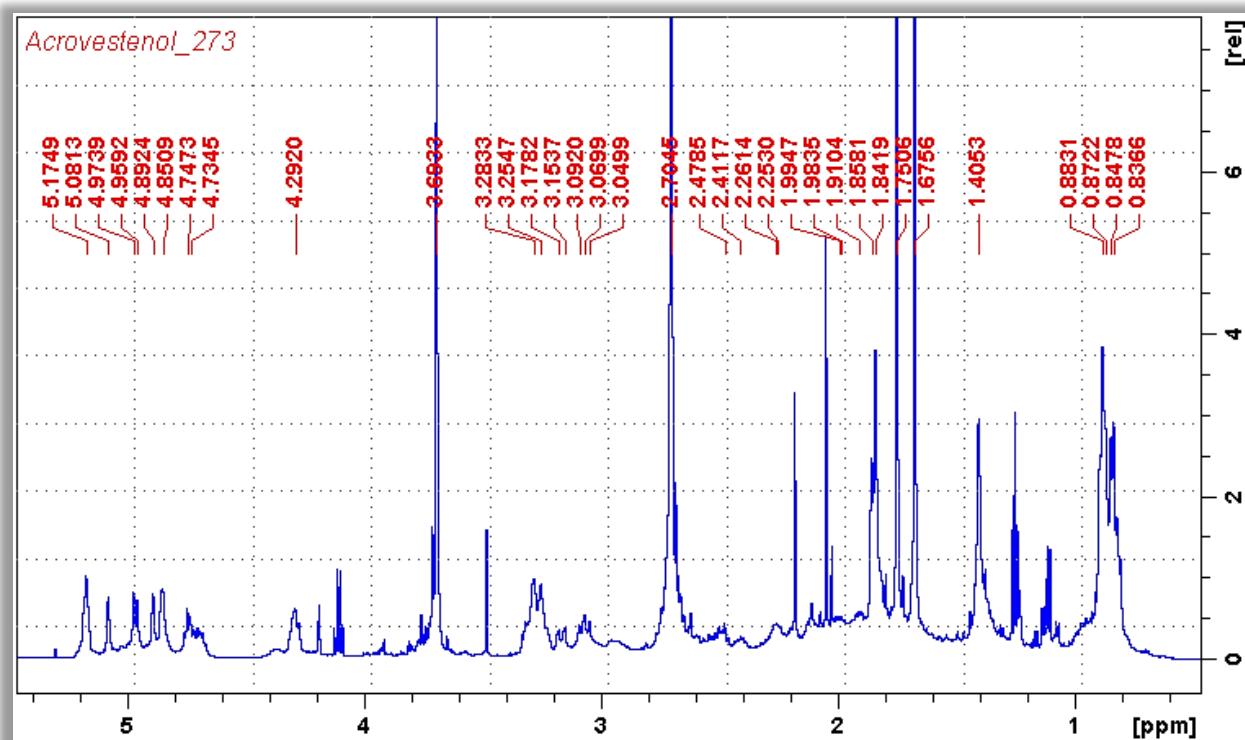


Figure A 41: ^1H NMR (600 MHz, CDCl_3 , 0 $^\circ\text{C}$) spectrum of Acrovestenol (5), δ_{H} 0-6

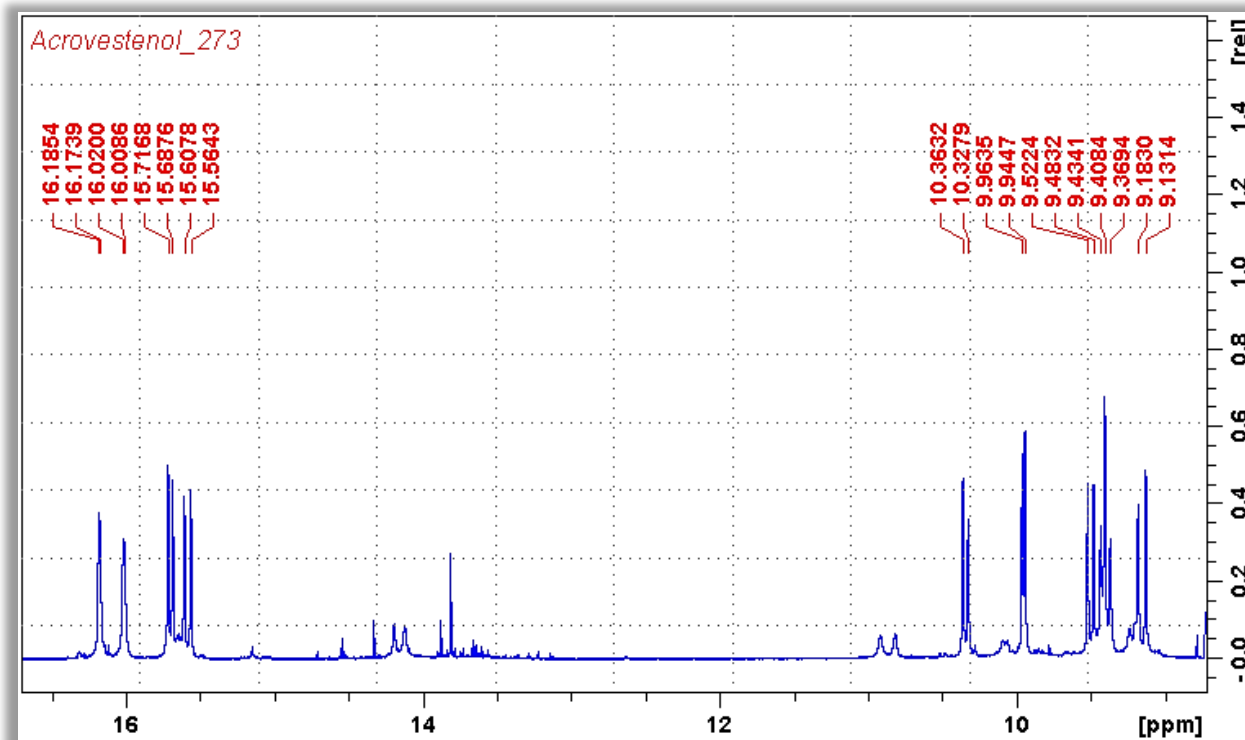


Figure A 42: ^1H NMR (600 MHz, CDCl_3 , 0 $^\circ\text{C}$) spectrum of Acrovestenol (5), δ_{H} 8.5-16.5

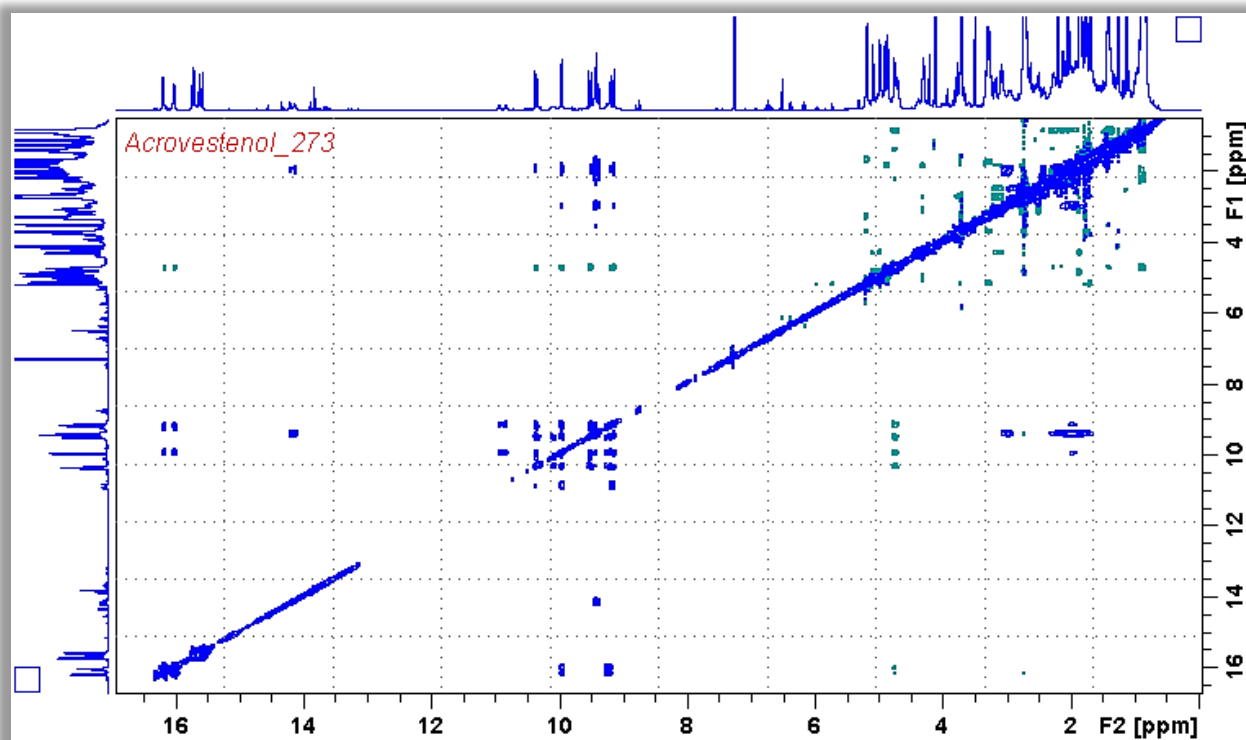


Figure A 43: NOESY (600 MHz, CDCl₃, 0 °C) spectrum of Acrovestenol (5)

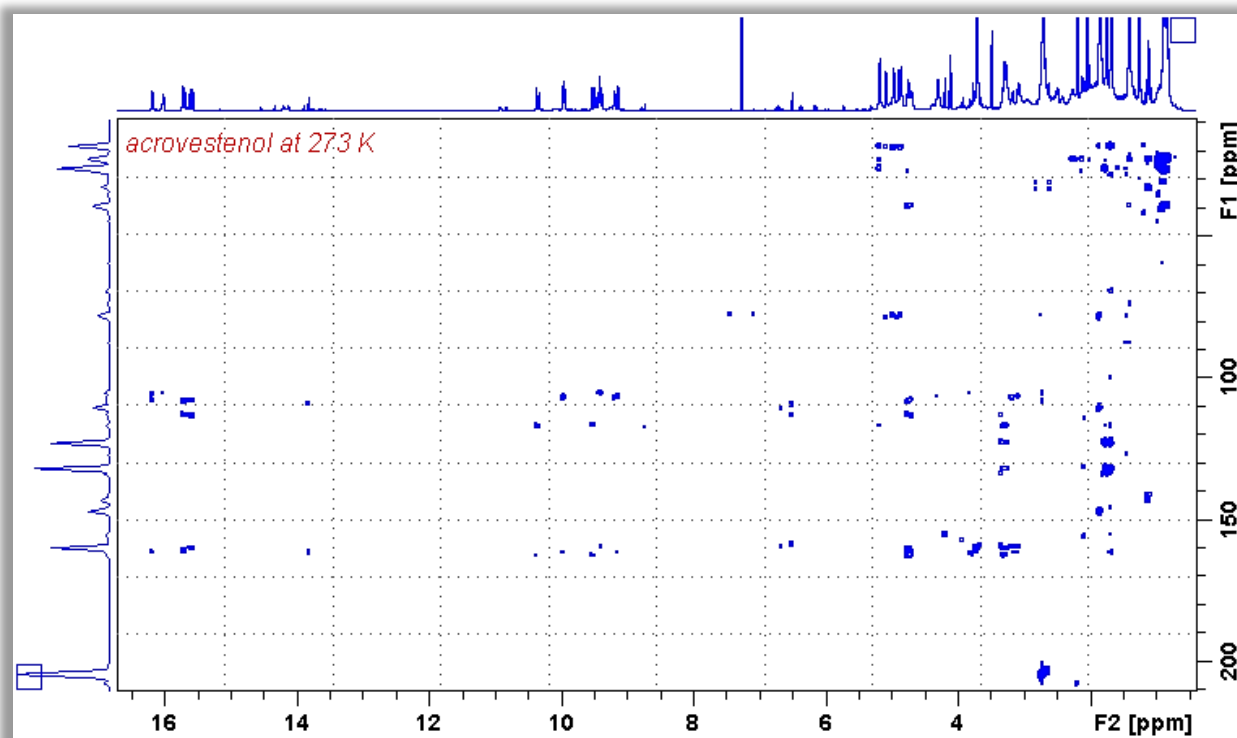
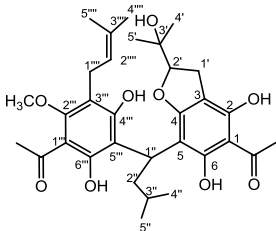


Figure A 44: HMBC (600 MHz, CDCl₃, 0 °C) spectrum of Acrovestenol (5)

Table A 11: NMR spectroscopic data of Acrofolione A (6) at 47 °C; different diastereomers' signals are assigned in grey

NMR spectroscopic data (600 MHz, 47 °C CDCl ₃) for Acrofolione A				
				
No	¹ H (J, Hz)	¹³ C	HMBC	NOESY
1		107.3		
2		162.1		
3		104.5		
4		159.4		
5		103.3		
6		159.9		
1'a	3.03 m	27.3	2', 3'	1'b, 4', 5'
1'b	3.14 m			1'a, 2'
2'	4.87 (t, 8.8)	92.9		1', 4', 5'
	4.82 (t, 8.9)	92.5	4', 5'	
3'		71.7		
4'	1.21 s	23.6	2', 3', 5'	1'a, 2', 5'
	1.27 s	24.2		
5'	1.38 s	25.8	2', 3', 4'	1'a, 2', 4'
	1.39 s			
1''	4.70 brs	28.6	2'', 3''	2'', 3'', 4'', 5''
2''a	2.04 m	39.3	5, 1'', 3'', 4'', 5'', 5'''	1'', 2''b, 3'', 4'', 5''
2''b	2.24 m			1'', 2''a, 3'', 4'', 5''
3''	1.43 m	26.9	1'', 2'', 4'', 5''	1'', 2'', 4'', 5''
4''	0.90 (d, 6.3)	22.3	2'', 3'', 5''	2'', 3'', 5''
	0.90 (d, 6.5)			
5''	0.92 (d, 6.3)		2'', 3'', 4''	2'', 3'', 4''
	0.91 (d, 6.5)			
1'''		108.8		
2'''		160.0		
3'''		116.4		
4'''		160.5		
5'''		113.0		
6'''		160.1		
1''''	3.30 brm	22.8	3''', 4''', 2''', 3''''	2''', 5''', MeO
2''''	5.13 (t, 6.7)	123.0	3''', 1''', 4''', 5''''	2''', 4''''
	5.16 (t, 6.7)			
3''''		131.9		
4''''	1.69 s	25.4	2''', 3''', 5''''	2''''
5''''	1.77 s	17.5	2''', 3''', 4''''	1''', MeO
MeO	3.73 s	62.4	2'''	1''', 5''', MeCO-1'''
MeCO-1	2.69 s	33.2	1, MeCO-1	
MeCO-1		202.0		
MeCO-1'''	2.72 s	30.4	1''', MeCO-1'''	MeO
MeCO-1'''		204.1		
OH-2				
OH-6				
OH-3'				
OH-4'''				
OH-6'''	15.40 s		1''', 5''', 6'''	1'', 4'', 5'', MeCO-1'''

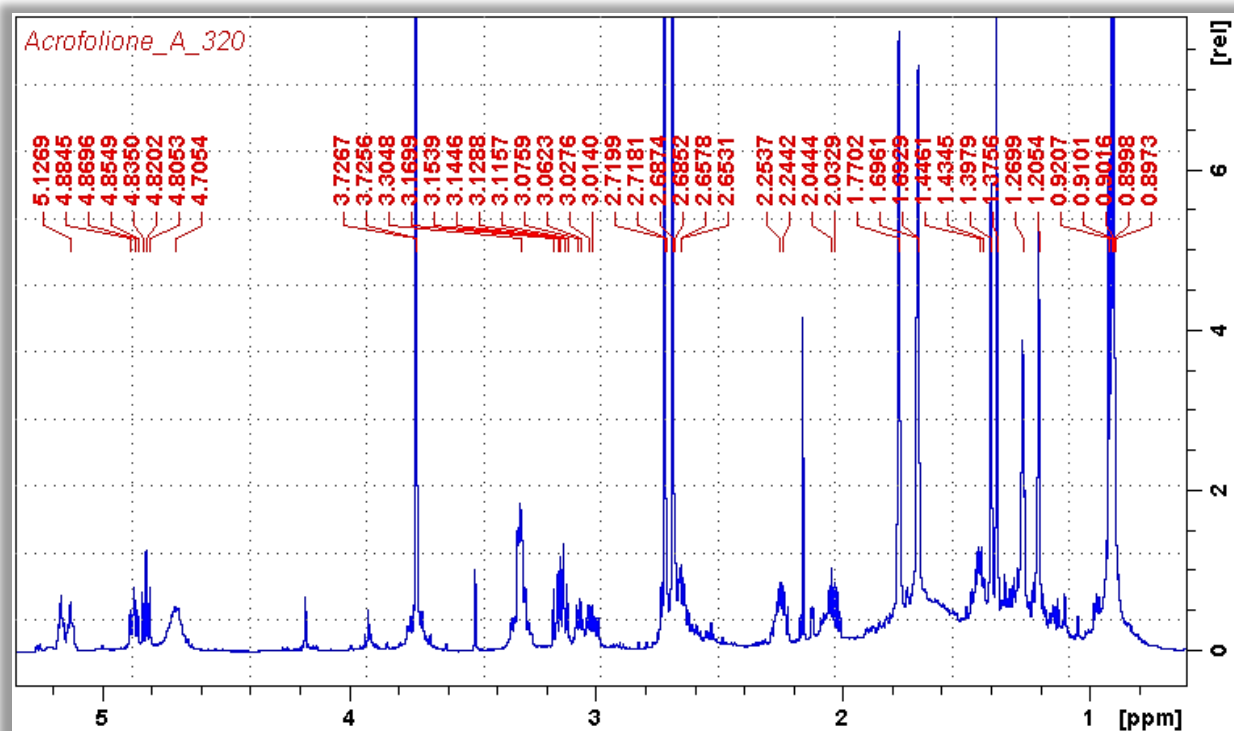


Figure A 45: ^1H NMR (600 MHz, CDCl_3 , 47 °C) spectrum of Acrofolione A (6)

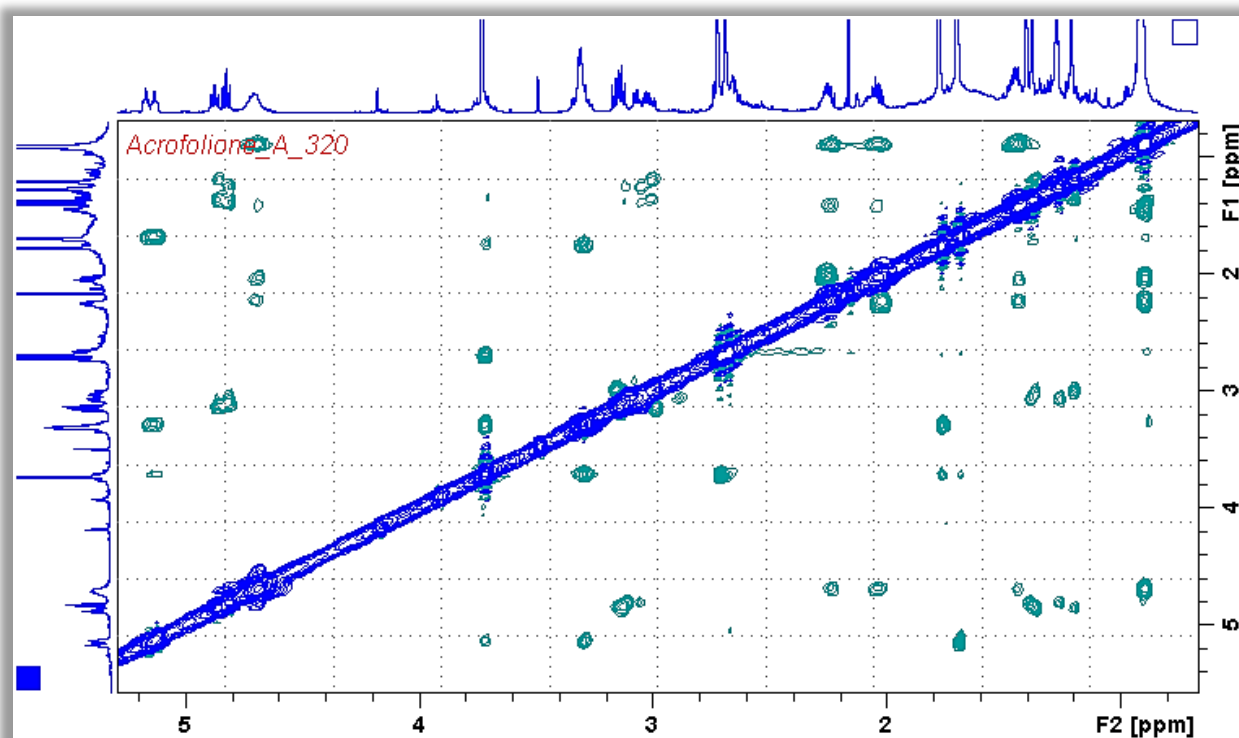


Figure A 46: NOESY (600 MHz, CDCl_3 , 47 °C) spectrum of Acrofolione A (6)

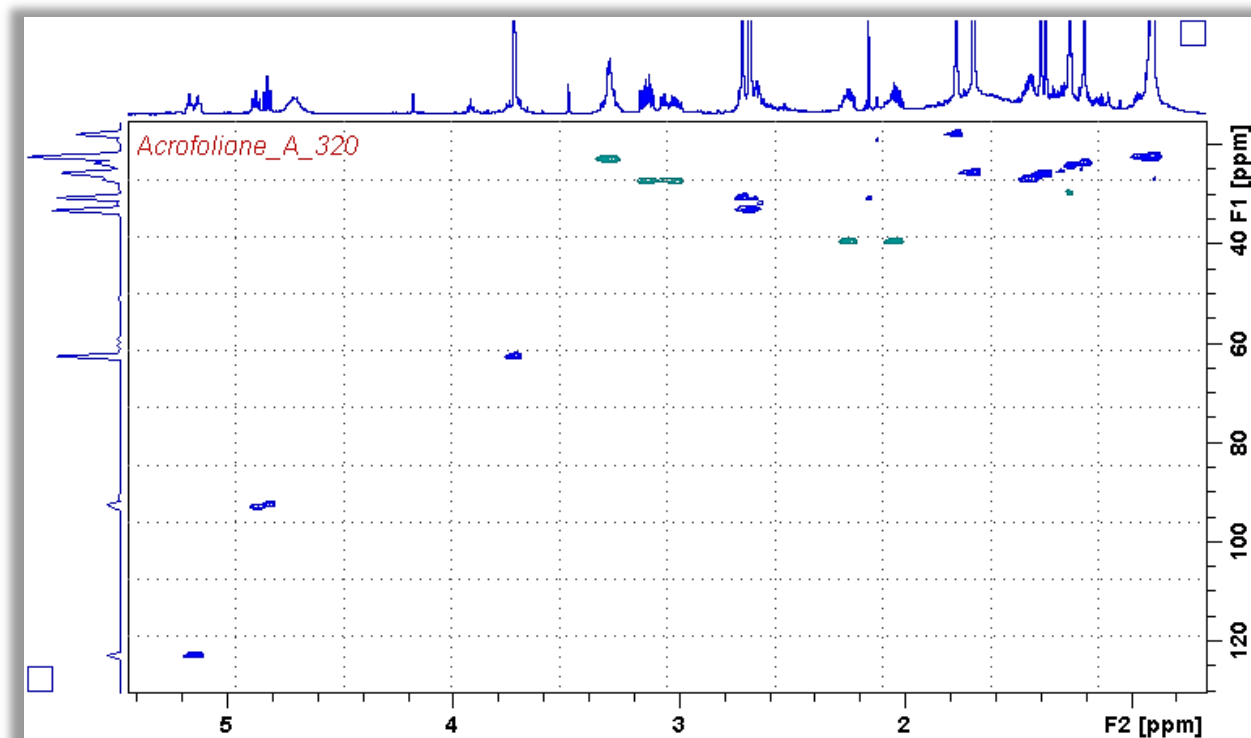


Figure A 47: HSQC (600 MHz, CDCl₃, 47 °C) spectrum of Acrofolione A (6)

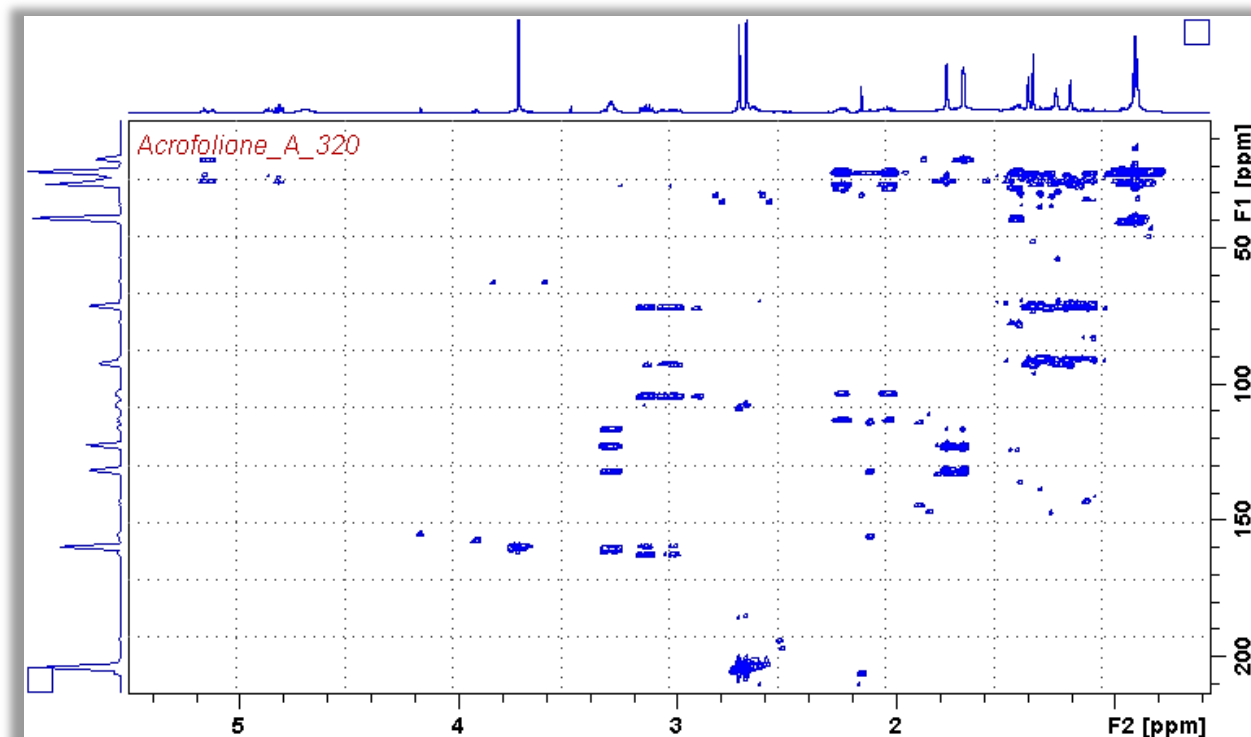
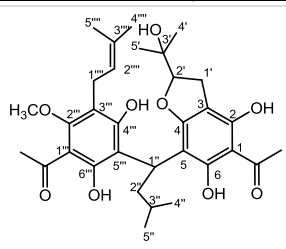


Figure A 48: HMBC (600 MHz, CDCl₃, 47 °C) spectrum of Acrofolione A (6)

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

Table A 12: NMR spectroscopic data of Acrofolione A (6) at 0 °C; different rotamers' signals are assigned in red and blue

NMR spectroscopic data (600 MHz, 0 °C CDCl ₃) for Acrofolione A				
				
No	¹ H (J, Hz)	¹³ C	HMBC	NOESY
1		107.0		
2		162.1		
3		104.5		
4		159.4		
5		103.3		
6		159.9		
1'a	3.11 brs			2', 5'
1'b	3.16 (dd, 15.2/ 9.8)	27.1	2, 3, 4, 3'	2', 4', 5'
	2.95 brs			
2'	4.92 brs	93.0		1'b, 5'
	4.84 brs	92.4		
3'		71.7		
4'	1.11 brs	22.7	2', 3', 5'	1', 2'
5'	1.36 s	25.6	2', 3', 4'	
1''	4.70 brs	27.3		2'', 4'', 5'', OH-6, OH-4'''
2''a	2.01 m	39.0	5, 3'' 4'', 5'', 5'''	1'', 2'', 3'', 4'', 5'', OH-4'''
2''b	2.24 brs			
3''	1.42 m	26.5	1'', 2'', 4'', 5''	1'', 2'', 4'', 5'', OH-4'''
4''	0.88 (d, 6.5)	22.3	2'', 3'', 5''	2'', 3''
5''	0.88 (d, 6.5)		2'', 3'', 4''	2'', 3''
1'''		108.4		
2'''		159.8		
3'''		116.4		
4'''		160.5		
5'''		112.7		
6'''		160.1		
1''''	3.29 brm	22.7	3''', 2''', 2''', 3''''	2''', 5''', MeOH
2''''A	5.06 brs	122.7		1''', 4''', MeOH
2''''B	5.14 brs			
3''''		132.0		
4''''	1.67 s	25.5	2''', 3''', 5''''	2''''
5''''	1.75 s	17.8	2''', 3''', 4''''	1''', MeO
MeO	3.71 s	62.5	2'''	1''', 2''', 5''', MeCO-1'''
MeCO-1	2.68 s	33.6	1, MeCO-1	OH-2, OH-6
MeCO-1		204.8		
MeCO-1'''	2.72 s	31.0	1''', MeCO-1'''	MeO, OH-6'''
MeCO-1'''		204.2		
OH-2	13.82 s			MeCO-1
	13.92 s			
OH-6	9.00 s			1'', MeCO-1
	9.73 s			
OH-3'				
OH-4'''	7.91 s			
	8.64 s			1'', 2'', 3''
OH-6'''	15.48 s			MeCO-1''', OH-4
	15.64 s			

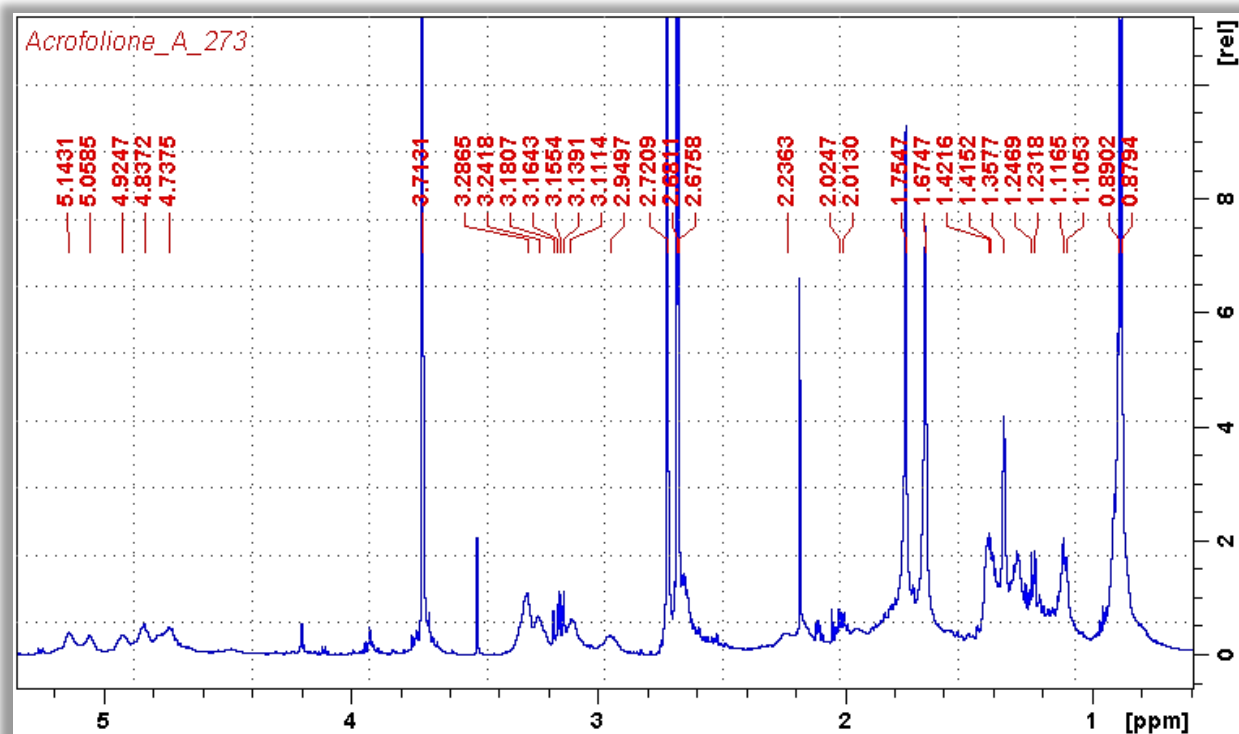


Figure A 49: ^1H NMR (600 MHz, CDCl_3 , 0 °C) spectrum of Acrofolione A (6), δ_{H} 0-6

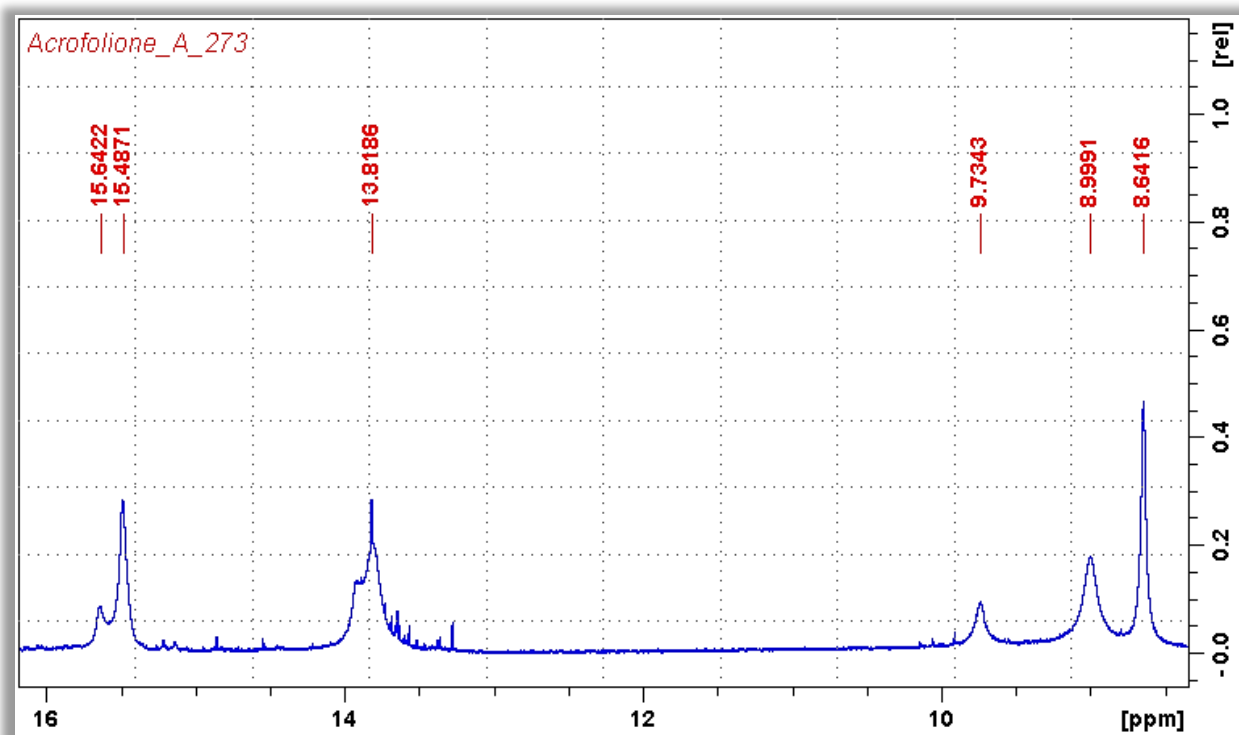


Figure A 50: ^1H NMR (600 MHz, CDCl_3 , 0 °C) spectrum of Acrofolione A (6), δ_{H} 8-16

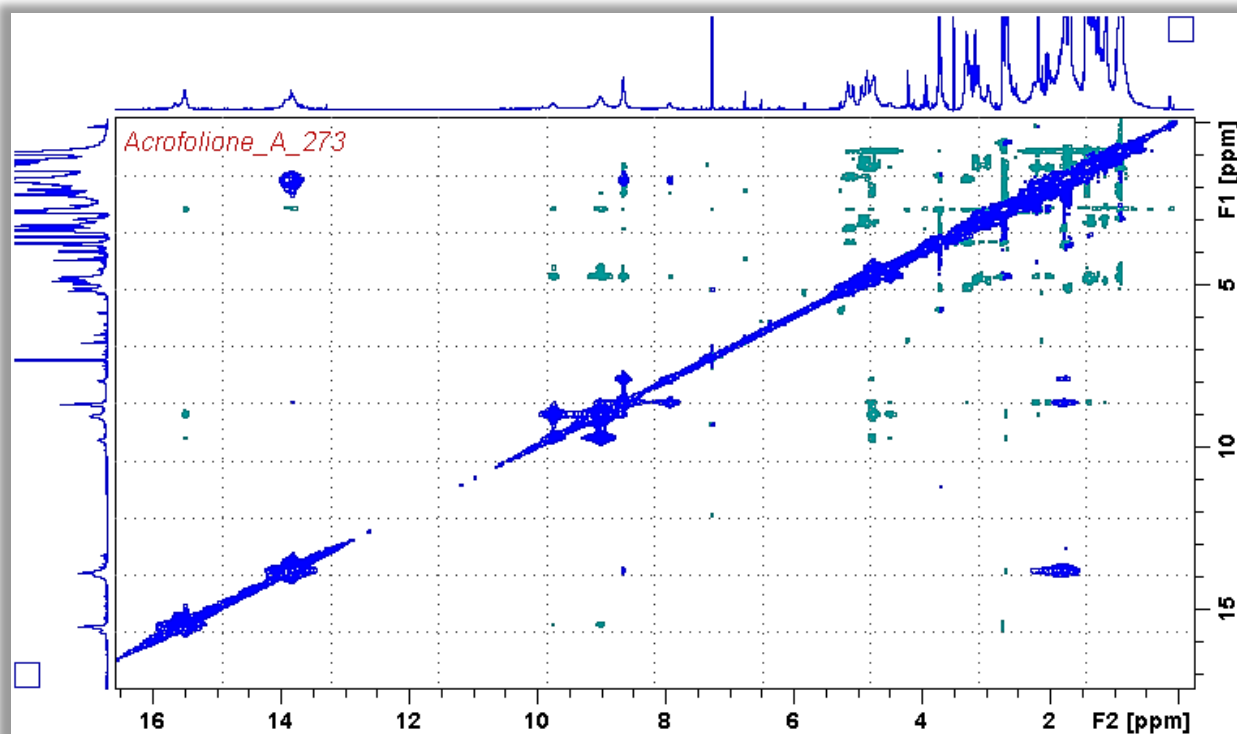


Figure A 51: NOESY (600 MHz, CDCl₃, 0 °C) spectrum of Acrofolione A (6)

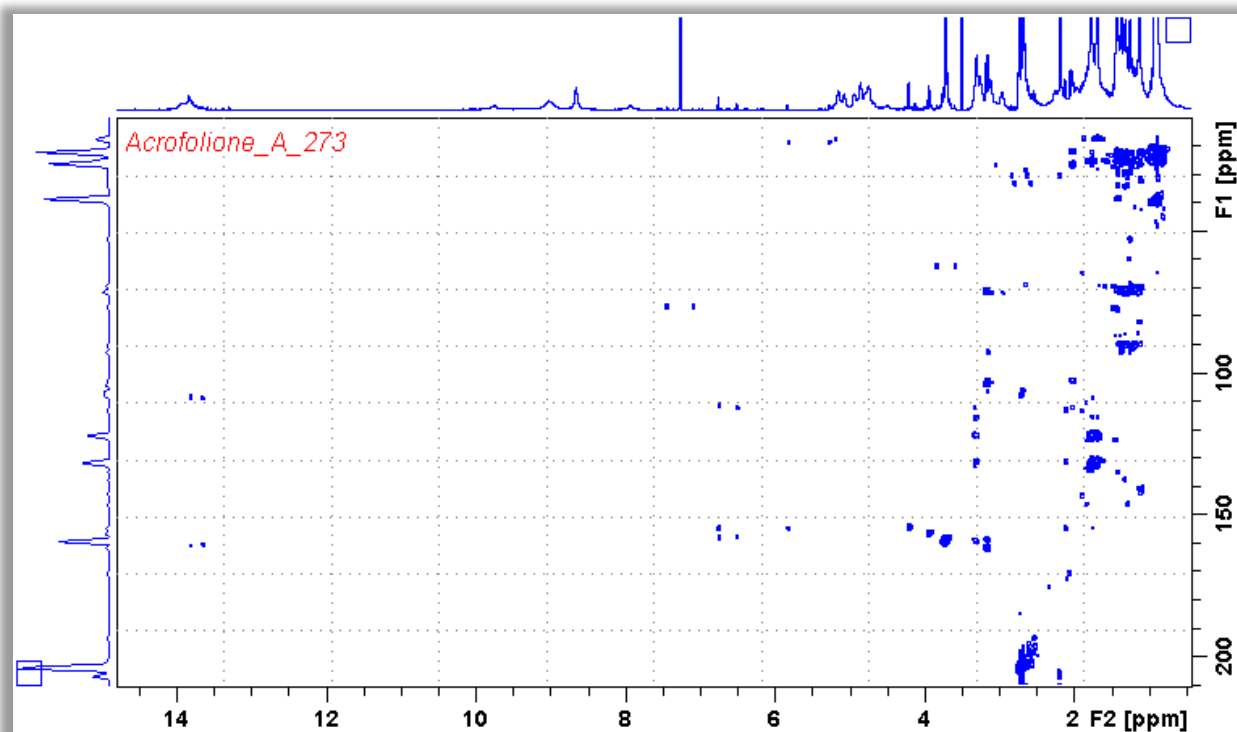


Figure A 52: HMBC (600 MHz, CDCl₃, 0 °C) spectrum of Acrofolione A (6)

Table A 13: NMR spectroscopic data of Acrofolione B (7) at 47 °C

NMR spectroscopic data (600 MHz, 47 °C CDCl ₃) for Acrofolione B				
No	¹ H (J, Hz)	¹³ C	HMBC	NOESY
1		101.1		
2		161.3		
3		106.5		
4		159.5		
5		109.2		
6		159.9		
1'	3.05 brs	27.4	2', 3'	2', 4'
2'	4.73 (t, 9.3)	92.0	4'	1', 4', 5', MeCO-1
3'		71.7		
4'	1.25 s	24.5	2', 3', 5'	1', 2'
5'	1.35 s	26.1	2', 3', 4'	1', 2', MeCO-1
1''	4.72 (t, 7.7)	28.3	5, 2'', 3'', 5'''	2'', 3'', 4'', 5''
2''	2.22 brs	39.5	4'', 5''	1'', 3'', 4'', 5''
	2.14 brs			
3''	1.43 m	27.0	2'', 4'', 5''	1'', 2'', 4'', 5''
4''	0.89 brs	22.5	2'', 3'', 5''	2'', 3'', 5''
5''	0.89 brs	22.5	2'', 3'', 4''	2'', 3'', 4''
1'''		108.3		
2'''		160.3		
3'''		116.7		
4'''		162.6		
5'''		113.2		
6'''		160.9		
1''''	3.30 (d, 4.8)	23.2	2''', 3''', 2'''', 3''''	2''', 5''', MeO
2''''	5.20 (t, 6.5)	123.3	3''', 1''', 4''', 5''''	2''', 4''', MeO
3''''		131.7		
4''''	1.69 s	25.6	2''', 3''', 5''''	2''''
5''''	1.76 s	17.9	2''', 3''', 4''''	1''', MeO
MeO	3.71 s	62.5	2'''	1''', 2''', 5''', MeCO-1'''
MeCO-1	2.63 s	30.6	1, MeCO-1	5'
MeCO-1		202.0		
MeCO-1'''	2.70 s	30.4	1''', MeCO-1'''	MeO
MeCO-1'''		204.1		
OH-4				
OH-6				
OH-3'				
OH-4'''				
OH-6'''				

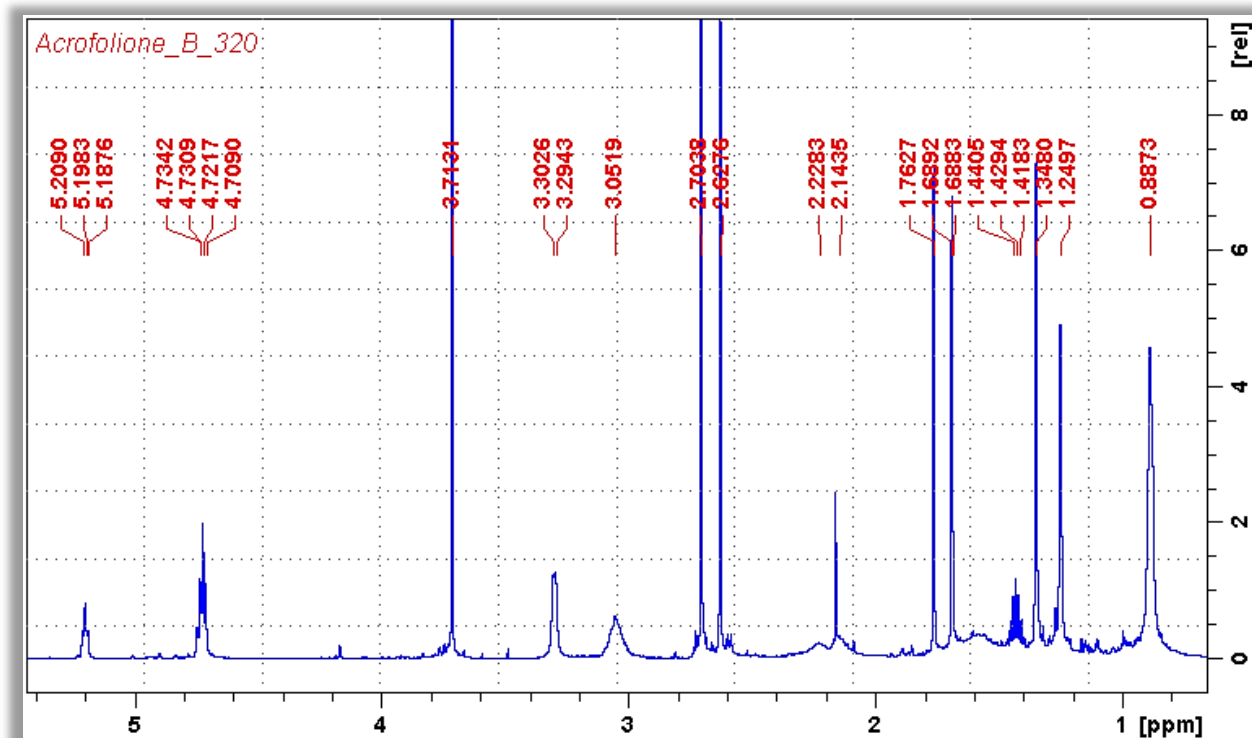


Figure A 53: ^1H NMR (600 MHz, CDCl_3 , 47 °C) spectrum of Acrofolione B (7)

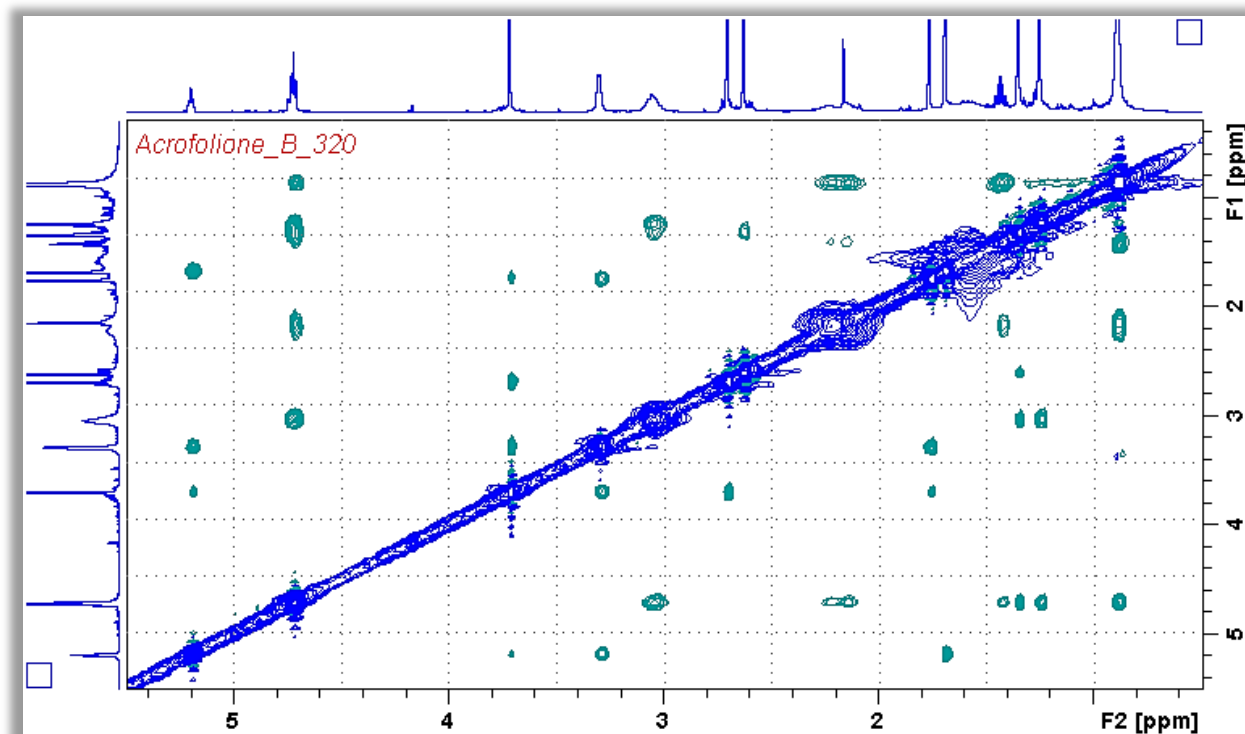


Figure A 54: NOESY (600 MHz, CDCl_3 , 47 °C) spectrum of Acrofolione B (7)

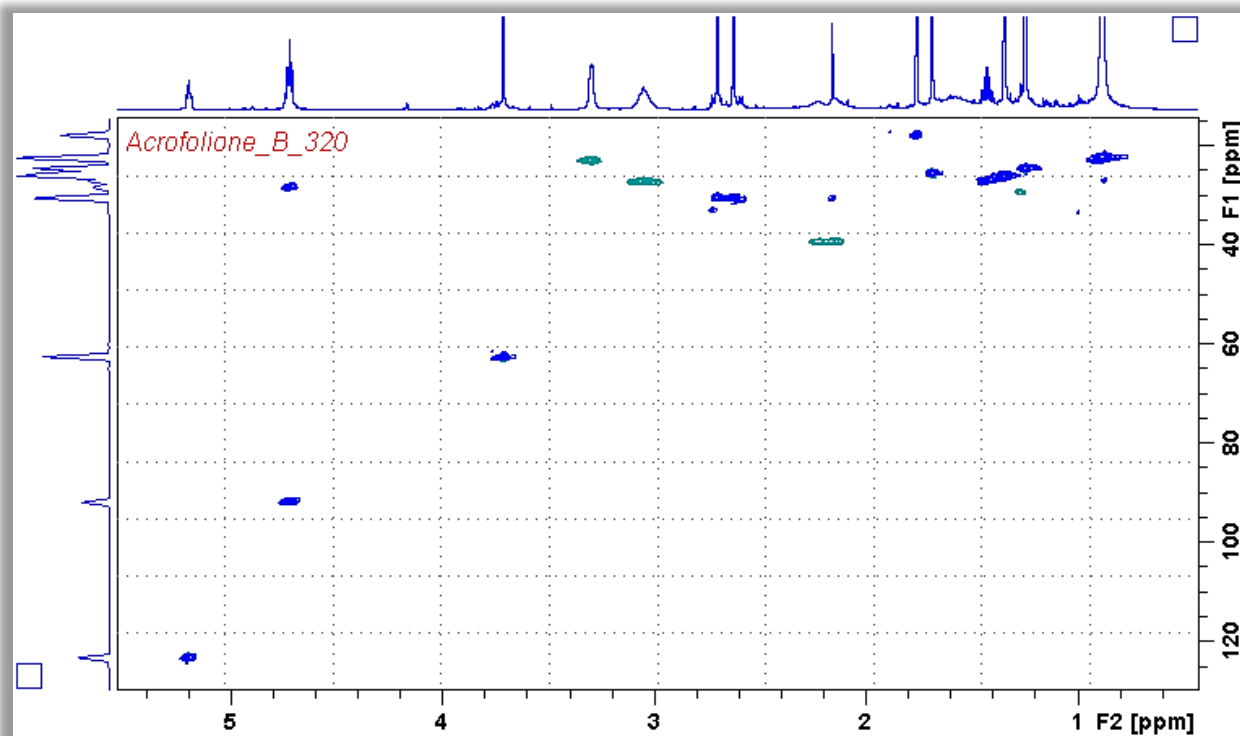


Figure A 55: HSQC (600 MHz, CDCl₃, 47 °C) spectrum of Acrofolione B (7)

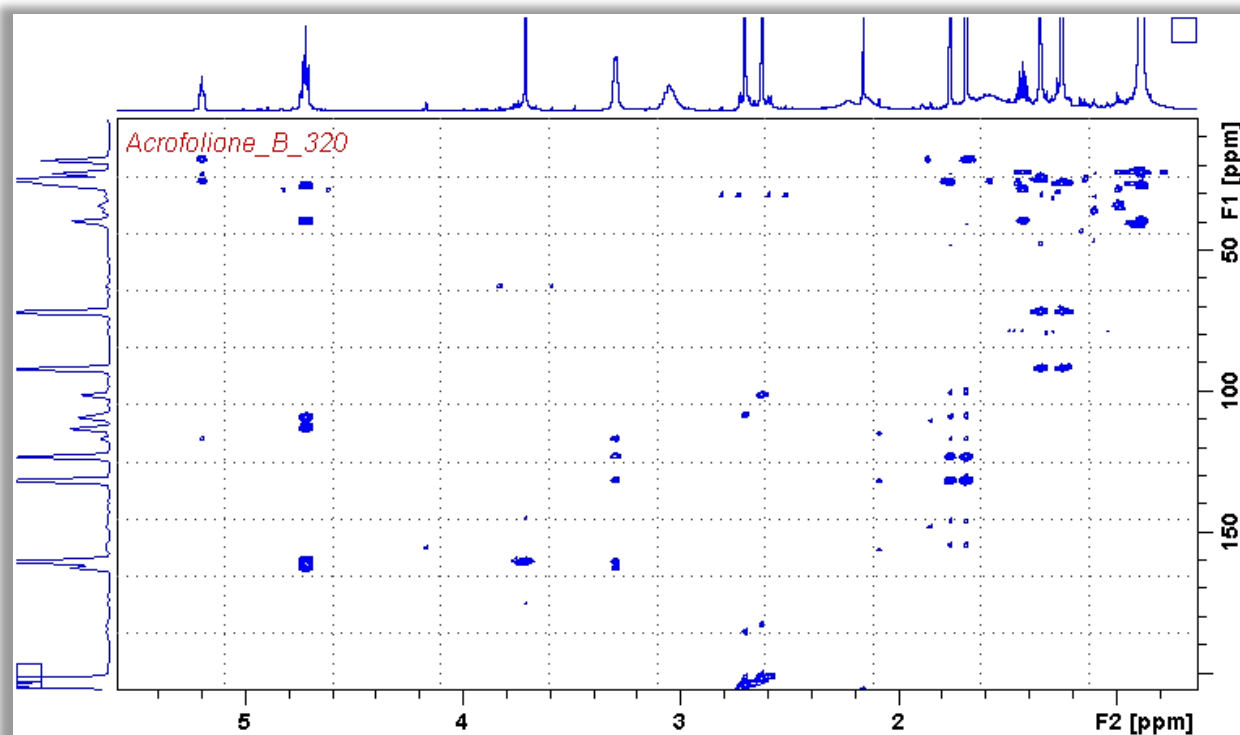
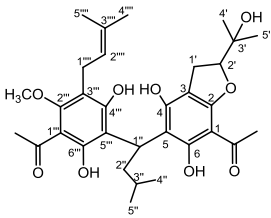


Figure A 56: HMBC (600 MHz, CDCl₃, 47 °C) spectrum of Acrofolione B (7)

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

Table A 14: NMR spectroscopic data of Acrofolione B (7) at 0 °C; different rotamers' signals are assigned in red and blue

NMR spectroscopic data (600 MHz, 0 °C CDCl ₃) for Acrofolione B				
				
No	¹ H (J, Hz)	¹³ C	HMBC	NOESY
1		101.1		
2		161.3		
3		106.5		
4		159.5		
5		108.1		
		109.1		
6		159.9		
1'	3.08 brs 3.00 brs	27.1	2, 3, 4, 3'	2', 4', 5'
2'	4.76 (t, 9.2) 4.75 (t, 8.8)	91.6	4', 5'	1', 4', 5'
3'		71.8		
4'	1.25 s 1.21 s	24.2	2', 3', 5'	1', 2', MeCO-1
5'	1.37 s 1.34 s	26.3	2', 3', 4'	1', 2', MeCO-1
1''	4.70 brs 4.69 brs	27.9	4, 5, 6, 2'', 3'', 4''', 5''', 6''' 4, 5, 6, 2'', 3'', 4''', 5''', 6'''	2'', 3'', 4'', 5'', OH-4, OH-4''', OH-6, OH-6'''
2''a	2.33 m 2.18 m	38.7 39.3	5, 1'', 3'', 4'', 5'', 5'''	1'', 3'', 4'', 5'', OH-4, OH-4'''
2''b	2.00 m 2.11 m	38.7 39.3	5, 1'', 3'', 4'', 5'', 5'''	1'', 3'', 4'', 5'', OH-4, OH-4'''
3''	1.41 brs	26.8	1'', 2'', 4'', 5''	1'', 4'', 5''
4''	0.89 (d, 6.6)	22.6	2'', 3'', 5''	1'', 2'', 3''
5''	0.84 (d, 6.2)	22.4	2'', 3'', 4''	
1'''		108.1		
2'''		159.9		
3'''		116.6		
4'''		162.3		
5'''		113.0		
		112.3		
6'''		160.9		
1''''	3.31 brm	22.9	2''', 3''', 4''', 2'''', 3''''	2''', 5''', MeO, OH-4''''
2''''	5.17 brs	122.8	3''', 1''', 4''', 5''''	1''', 4''', OH-4''''
3''''		116.6		
4''''	1.68 s	25.8	2''', 3''', 5''''	2''''
5''''	1.75 s	18.1	2''', 3''', 4''''	1''', MeO
MeO	3.69 s	62.7	2''''	1''', 5''', MeCO-1''''
MeCO-1	2.63 s	30.9	1, MeCO-1	5', OH-6
MeCO-1		202.1		MeO, OH-6'''
MeCO-1'''	2.72 s	30.8	1''', MeCO-1''''	
MeCO-1''''		204.2		
OH-4	9.24/ 9.29 10.10/ 10.15		2, 3, 4, 5 2, 3, 4, 5	1'', 2'', OH-6''' 1'', 2'', OH-6'''
OH-6	15.34/ 15.36 15.19/ 15.19		1, 2, 5 1, 5, 6	1'', MeCO-1, OH-4'''' 1'', MeCO-1, OH-4''''
OH-3'				
OH-4'''	10.00/ 10.03 9.15/ 9.18		2''', 3''', 4''', 5''' 2''', 3''', 4''', 5'''	1'', 2'', 1''', 2''', OH-6 1'', 2'', 1''', 2''', OH-6
OH-6'''	15.64/ 15.67 15.79/ 15.81		1''', 2''', 5''' 1''', 4''', 5''', 6'''	1'', MeCO-1''', OH-4 1'', MeCO-1''', OH-4

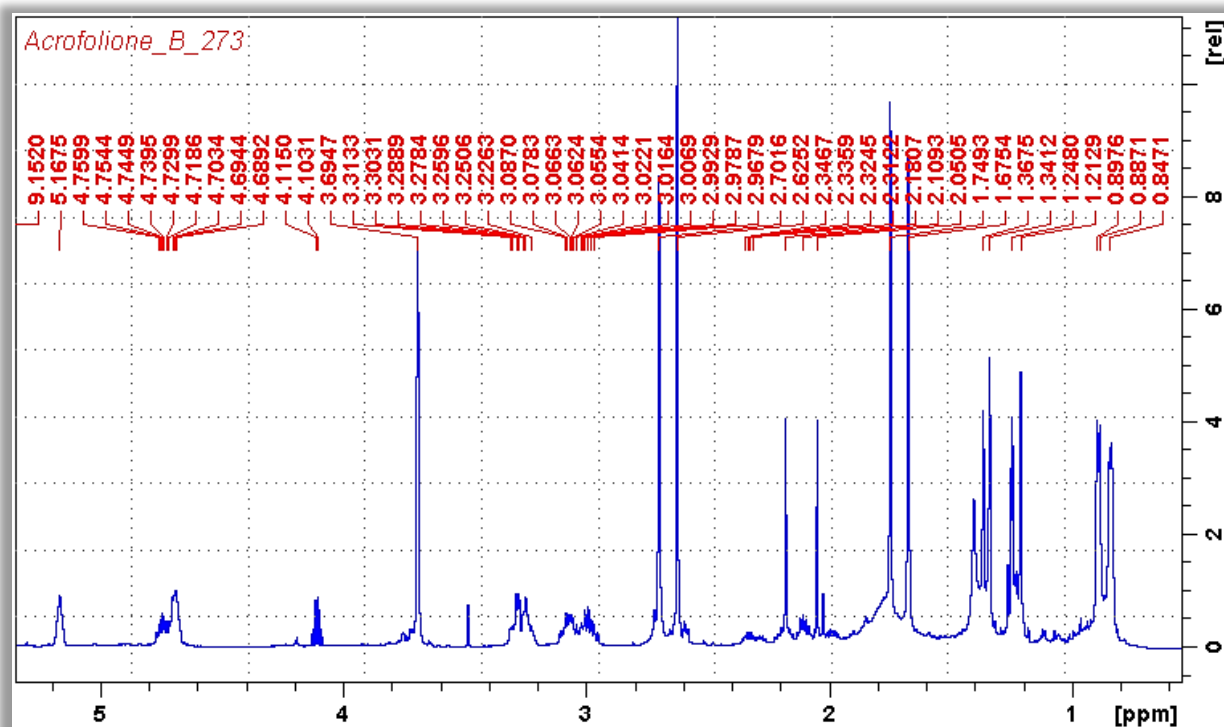


Figure A 57: ^1H NMR (600 MHz, CDCl_3 , 0 °C) spectrum of Acrofolione B (7), δ_{H} 0-6

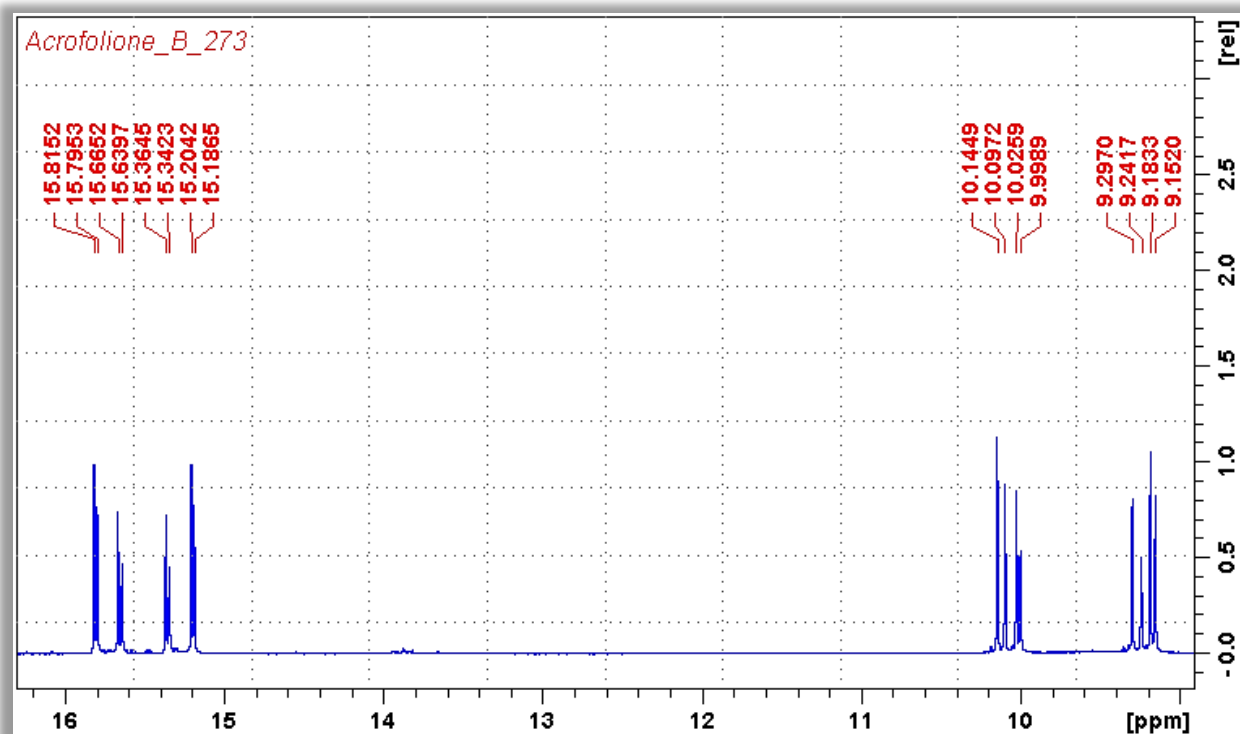


Figure A 58: ^1H NMR (600 MHz, CDCl_3 , 0 °C) spectrum of Acrofolione B (7), δ_{H} 9-16.2

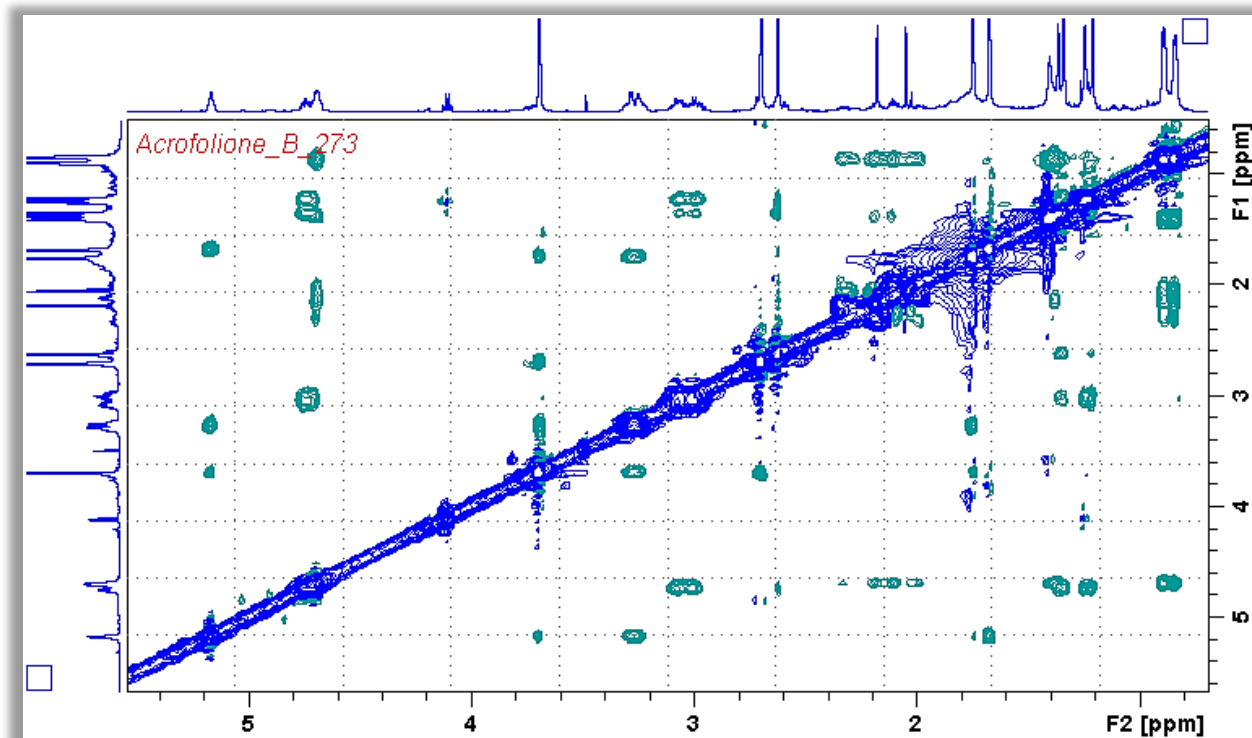


Figure A 59: NOESY (600 MHz, CDCl₃, 0 °C) spectrum of Acrofolione B (7)

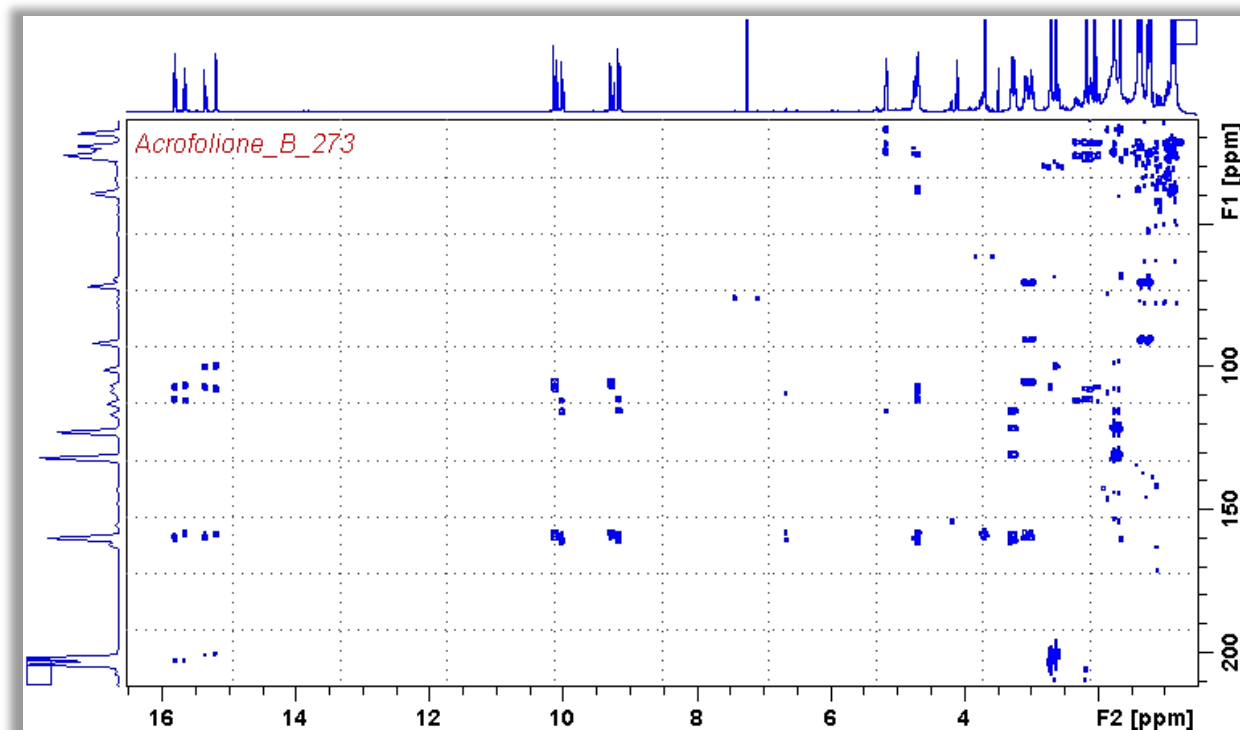
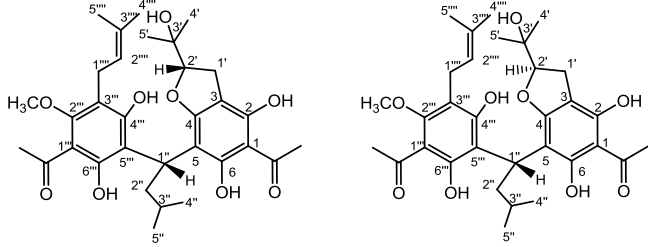


Figure A 60: HMBC (600 MHz, CDCl₃, 0 °C) spectrum of Acrofolione B (7)

Table A 15: NMR spectroscopic data of Acrofolione A (6) diastereomers at 47 °C; different ^1H and ^{13}C NMR data are observed but did not attributed to specific diastereomers while HMBC and NOESY data were identical

NMR spectroscopic data (600 MHz, 47 °C CDCl_3) for Acrofolione A diastereomers						
						
No	Diastereomer 1		Diastereomer 2		HMBC	NOESY
	^1H (J, Hz)	^{13}C	^1H (J, Hz)	^{13}C		
1		107.3		107.3		
2		162.1		162.1		
3		104.5		104.5		
4		159.4		159.4		
5		103.3		103.3		
6		159.9		159.9		
1'a	3.06 (dd, 7.8/ 15.2)	27.3	3.06 (dd, 8.5/ 15)	27.3	2', 3'	1'b, 4', 5'
1'b	3.15 (dd, 9.5/ 15.2)		3.14 (dd, 9.6/ 15)			1'a, 2'
2'	4.87 (t, 9)	92.6	4.82 (t, 8.9)	93.0		1', 4', 5'
3'		71.7		71.7		
4'	1.21 s	24.2	1.27 s	24.2	2', 3', 5'	1'a, 2', 5'
5'	1.38 s	25.8	1.39 s	25.8	2', 3', 4'	1'a, 2', 4'
1''	4.69 brt	28.6	4.72 brs	28.6	2'', 3''	2'', 3'', 4'', 5''
2''a	2.03 m	39.3	2.06 brm	39.3	5, 1'', 3'', 4'', 5'', 5'''	1'', 2''b, 3'', 4'', 5''
2''b	2.26 brm		2.24 m			1'', 2''a, 3'', 4'', 5''
3''	1.43 m	26.9	1.43 m	26.9	1'', 2'', 4'', 5''	1'', 2'', 4'', 5''
4''	0.91 (dd, 1.4/ 6.6)	22.3	0.90 (t, 6.8)	22.3	2'', 3'', 5''	2'', 3'', 5''
5''	0.90 (dd, 1.4/ 6.6)		0.90 (t, 6.8)		2'', 3'', 4''	2'', 3'', 4''
1'''		108.8		108.8		
2'''		160.0		160.0		
3'''		116.4		116.4		
4'''		160.5		160.5		
5'''		113.0		113.0		
6'''		160.1		160.1		
1''''	3.30 brm	22.8	3.30 brm	22.8	3''', 4''', 2''', 3''''	2''', 5''', MeO
2''''	5.13 (t, 6.7)	122.8	5.16 (t, 6.7)	123.0	1''', 4''', 5''''	2''', 4''''
3''''		131.9		131.9		
4''''	1.69 s	25.4	1.69 s	25.4	2''', 3''', 5''''	2''''
5''''	1.77 s	17.5	1.77 s	17.5	2''', 3''', 4''''	1''', MeO
MeO	3.73 s	62.4	3.73 s	62.4	2'''	1''', 5''', MeCO-1'''
MeCO-1	2.69 s	33.2	2.69 s	33.2	1, MeCO-1	
MeCO-1		202.0		202.0		
MeCO-1'''	2.72 s	30.4	2.72 s	30.4	1''', MeCO-1'''	MeO
MeCO-1'''		204.1		204.1		
OH-2						
OH-6						
OH-3'						
OH-4'''						
OH-6'''	15.40 s		15.40 s		1''', 5''', 6'''	1'', 4'', 5'', MeCO-1'''

Chapter 1: Targeted isolation and pharmacological evaluation of AtA

Table A 16: Inhibitory activities of compound 1-7 on DU145 prostate cancer and A2058 melanoma tumor cells; data are expressed as means $\pm$ SD

Cell viability (% control) at 10 μ M		
Compound	DU145	A2058
1	63 $\pm$ 6*	56 $\pm$ 8
2	40 $\pm$ 6	33 $\pm$ 6
3	55 $\pm$ 9	54 $\pm$ 8
4	21 $\pm$ 2	21 $\pm$ 2
5	26 $\pm$ 2	31 $\pm$ 1
6	26 $\pm$ 2	33 $\pm$ 6
7	77 $\pm$ 9	69 $\pm$ 8

CHAPTER

LC-MS based structural elucidation and
dereplication of AtA

Introduction

1. Background

Acronychia-type acetophenones (AtA) are considered as a characteristic chemical group of natural products found in the genus *Acronychia* of Rutaceae family (Adersen et al. 2007). Their biological interest is focused mainly on their significant cytotoxicity against several human tumor cell lines (Wu et al. 1989, Oyama et al. 2003) and anti-inflammatory activity (Pathmasiri et al. 2005). Up to date a small number of derivatives have been reported exhibiting high potential as anticancer agents against human melanoma and prostate tumor cell lines (Kouloura et al. 2012). The cytotoxicity of these compounds seems to be strongly associated to the different structural elements of the core structure of AtA modulating their biological properties.

Structurally, AtA are fully substituted phloroglucinol dimers consisting of two aromatic rings connected with an isopentyl chain. Naturally occurred derivatives are characterized by a standard substituted ring (ring B) while the other aromatic ring (ring A) is bearing different substituents or additional rings resulting to structurally similar derivatives or isomers (Figure 1). Moreover, they are highly symmetric with the ability to form extended intra- and inter-molecular hydrogen bonds and they are characterized by the presence of different enantiomers, diastereomers and rotamer. This particular nature of AtA renders their identification challenging and until now, limited data have been reported for the structure elucidation thereof. A previous study presenting the x-ray crystallography of acrovestone (**1**), a model AtA compound, is available (Wu et al. 1989) but mainly the structure elucidation of these compounds has been performed using NMR techniques (Pathmasiri et al. 2005, Kouloura et al. 2012). However, high temperature NMR (47°C) acquisitions and NOE experiments are highly required for the elucidation of isomers (Kouloura et al. 2012). According to our knowledge, the analysis of AtA via mass spectrometric techniques has not been reported up to now while MS data regarding the phloroglucinol derivatives in general are almost absent. Therefore, the application of modern mass spectrometric methods for the investigation and identification of AtA is of great interest.

2. Multistage mass spectrometry in natural products

Mass spectrometry, especially coupled with atmospheric pressure ionization (API) sources (ESI: electrospray ionization and APCI: atmospheric pressure chemical ionization) have been applied widely for the structural investigation of natural products (Tian et al. 2003, Cuyckens et al. 2004, Wang et al. 2004, Liu et al. 2007, Bankefors et al. 2011). In addition, the introduction of multistage mass spectrometry (MS^n) approach in combination with dissociation methods such as collision induced dissociation (CID) facilitates significantly the identification process of secondary metabolites (Niessen 2000). This approach leads to the generation of sequential fragmentation spectra also called spectral trees or ion trees, where the number of the successive MS levels is depending on the ionization efficiency and the concentration of the metabolites under study. Additionally, the use of mass analyzers enabling high resolution measurements facilitates the determination of the elemental composition of both pseudomolecular and fragment ions. Thus, their application in structural studies results in deeper and thorough conception of the fragmentation mechanism of the compounds under investigation (van der Hooft et al. 2010).

Generally, the characterization of fragment ions observed in the different MS^n spectra and the elucidation of fragmentation patterns is a demanding process while limited mass spectral fragmentation databases in API conditions are available. Furthermore, although there are numerous studies dealing with the characterization of fragment ions produced by ESI-MS/MS studies in diverse compounds, limited data are available concerning general fragmentation rules regarding API sources (Holčapek et al. 2010, Weissberg et al. 2011). For instance Mass Frontier software which was used in the current study contains a large fragmentation library with a number of established fragmentation mechanisms in EI & CI conditions and associated structures. In this context, the hybrid linear ion trap-Orbitrap mass spectrometer (LTO-Orbitrap) has been proven a powerful tool for the development of methodologies for sensitive and accurate structure elucidation of small compounds (Dunn et al. 2008, Hooft et al. 2012, Zhang J Fau - Huang et al. 2012, Blaas et al. 2013, Qiu et al. 2013). The linear trap analyzer offers the capability of producing MS^n spectra combining the advantage of high sensitivity and fast scan speed (Dear et al. 2006). On the other hand, the Orbitrap analyzer ensures high resolution and mass accuracy close to Fourier-transform ion cyclotron resonance (FT-ICR) (Zubarev et al. 2013) providing high mass accurate measurements for both precursor and product ions commonly less than 3ppm (Lim et al.

2007, Qiu et al. 2013). The combination of the aforementioned analyzers provides important advantages offering new insights intriguing structural issues such as the identification of isomers. Additionally, LC-HRMSⁿ techniques facilitate drastically the structure identification procedure offering orthogonally chromatographic information. Especially, in the field of natural products chemistry this has been proven significantly useful enabling the profiling of complex mixtures such as natural extracts. In particular, an ion tree-based strategy could be considered as a method of choice for the systematic detection and structural determination of extracts' constituents. The main advantage of this strategy is the dereplication of several chemical groups in different mixtures without prior isolation (Wolfender et al. 2000, Kind et al. 2010).

3. Multistage mass spectrometry of AtA compounds

In the present study, a systematic analysis of AtA using a LTQ-Orbitrap platform applying both ESI and APCI ionization sources was intended in order to establish a methodology for the unambiguous structure elucidation of AtA by multistage HRMS. Therefore, a detailed study on the fragmentation mechanisms of these compounds under different ionization conditions was thoroughly investigated. For this purpose, the detection of key fragment ions that could be utilized for the identification of this specific category of compounds and the discrimination of different derivatives, even isomers was aimed. Since a limited number of AtA is reported up to day despite their important pharmacological importance, a UHPLC-ESI(-)-MSⁿ method was developed for the dereplication of known and the identification of possibly new *Acronychia*-type acetophenones in *Acronychia* extracts.

Experimental

1. Standards and reagents

The reference AtA investigated in this study (Figure 1) have been previously isolated from the trunk bark of *Acronychia pedunculata* and unambiguously identified by spectroscopic methods (Kouloura et al. 2012). The purity of all compounds was determined by RP-HPLC-PDA and NMR and found >89%. MS grade methanol (MeOH) and water (H₂O) were purchased by Merck (Germany) and formic acid (FA) from Sigma-Aldrich (Germany).

2. Sample preparation

Stock solutions of all reference compounds were prepared in MeOH (1 mg/mL). Further dilutions were performed in order to obtain the working samples of 10 µg/mL (ESI) and 20 µg/mL (APCI) used in the analysis by direct infusion method. A standard mixture of the reference compounds (10 µg/mL) was also prepared. For the preparation of the *Acronychia* extracts a previous method was applied (Kouloura et al. 2012). Briefly, an aliquot of 2 g of plant material was extracted using a two steps procedure. Initially, the plant material was extracted with Et₂O (A) (2 × 50mL, 24h per extraction) and the Et₂O extract was obtained. Then the plant residue was alkalized with 10% NH₄OH and extracted successively with Et₂O, CH₂Cl₂ and MeOH (2 × 50mL of each solvent) in order to exclude alkaloids. Therefore, the enriched Et₂O, CH₂Cl₂ and MeOH extracts were also obtained. All the extracts were diluted in methanol to obtain solutions of 100 µg/mL. All samples were stored at -20 °C prior to analysis.

3. Mass spectrometry and data handling

The HRMSⁿ analysis of AtA by direct infusion method and the LC-MS analysis were performed using an UHPLC-LTO-Orbitrap platform (Thermo Finnigan, San Jose, CA, USA).

Chapter 2: LC-MS based structural elucidation and dereplication of AtA

For infusions, ESI and APCI ionization sources, in both modes were used while for the LC-MS analysis only ESI(-) was incorporated. For all acquisitions only the Orbitrap analyzer was employed. In ESI direct infusion method all samples were subjected to analysis using an infusion pump at a flow rate of 5 $\mu\text{L}/\text{min}$, in a mass range of m/z 50-800 and injection time at $10^5 \mu\text{s}$. The resolution was set at 30000 (FWHM) and the data were acquired in profile mode. 1 microscan was used for the full scan and 2 microscans for HRMSⁿ acquisitions. For each acquisition 30 scans were averaged into a single profile spectrum. In positive mode, spray voltage was set at 3.5 V, capillary temperature at 350 °C, capillary voltage at 10 V and tube lens at 95 V. In negative mode, spray voltage was set at 3.5 V, capillary temperature at 300 °C, capillary voltage at -50 V and tube lens at -48 V. Nitrogen was used as sheath and auxiliary gas at flows of 12 and 6 arbitrary units, respectively, in both modes.

For APCI acquisitions, the same infusion pump was used at a flow rate of 10 $\mu\text{L}/\text{mL}$, a mass range of m/z 50-800 and injection time at $5 \times 10^5 \mu\text{s}$. The resolution was set at 30000 (FWHM) and the data were acquired in profile mode. 1 microscan was used for full scan and 3 microscans for the HRMSⁿ acquisitions. All acquisitions were averaged from 30 scans into a single profile spectrum. In positive ionization mode vaporizer temperature was set at 350 °C, discharge current at 5 μA , capillary temperature at 275 °C, capillary voltage at 3 V and tube lens at 40 V. In negative ionization mode, vaporizer temperature was set at 350 °C, discharge current at 5 μA , capillary temperature at 275 °C, capillary voltage at -8,5 V and tube lens at -75 V. Nitrogen was used as sheath and auxiliary gas at flows of 30 and 10 arbitrary units, respectively, for both modes.

Regarding the MSⁿ experiments, the collision energy (CID) was adjusted between 16% and 35%, in both sources and modes according to the requirements of each compound. The activation time was set at 30 ms (q , 0.25). A window of 1 u was used to isolate the precursor ions. The fragmentation of all reference compounds involved four fragment levels (MS² until MS⁵). Specifically, MS² fragmentation of the pseudomolecular or adduct ions were initially performed and then MS³-MS⁵ fragmentations of the most intense ions were followed using a 5% threshold of relative intensity.

LC experiments were conducted using an Accela UHPLC system (Thermo Finnigan, San Jose, CA, USA) consisting of an Accela pump and autosampler. Chromatographic separations were carried out on a Hypersil GOLD™ C₁₈ column 150mm $\times$ 4.6 mm, 3 μm . The mobile phase was consisted of a) H₂O with 0.1% FA and b) MeOH. The elution program was composed of an initial gradient elution step of 2 min from 10% MeOH to 75% MeOH

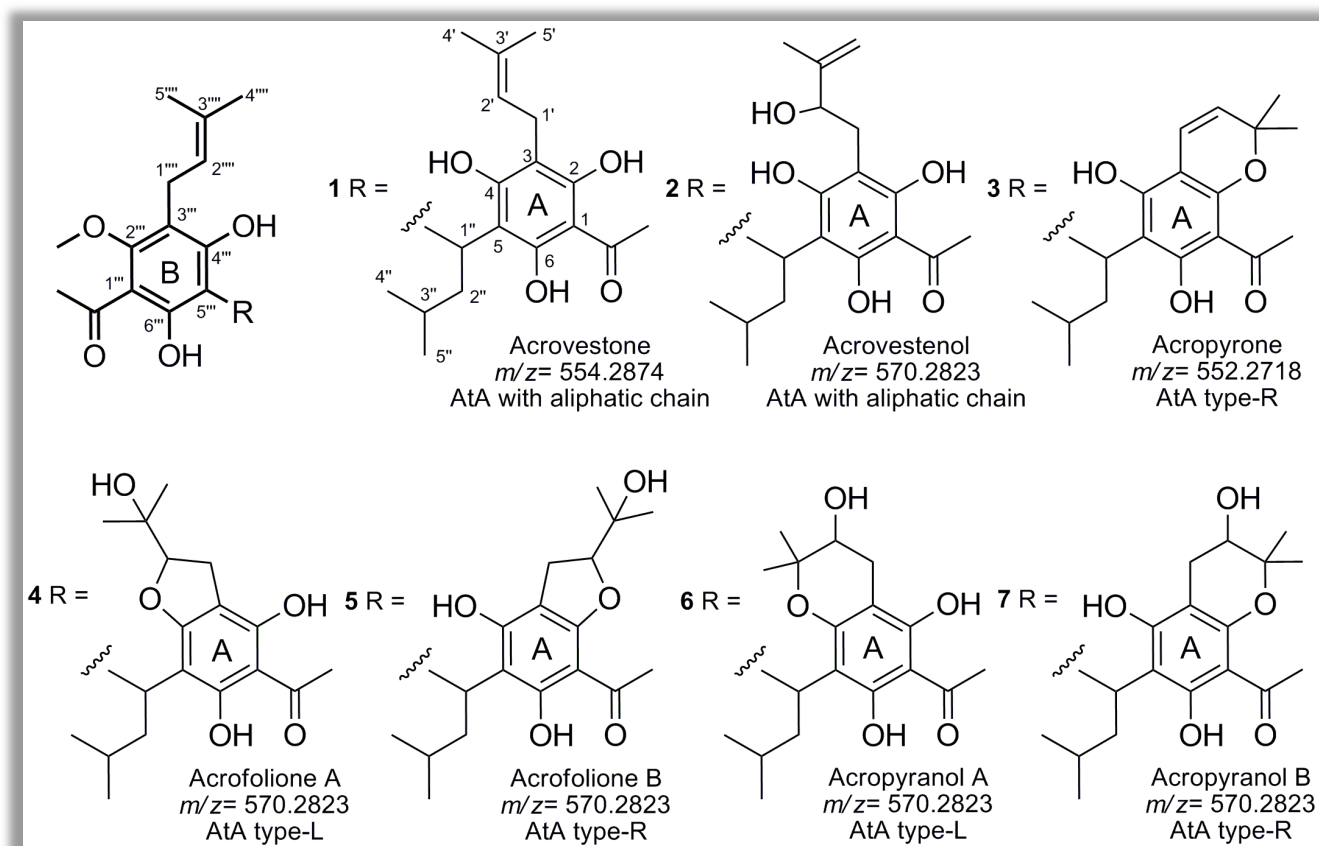
followed by 1 min of isocratic step (75% MeOH). Then a gradient elution step of 22 min from 75% MeOH to 100% MeOH and after 1 min of isocratic step (100% MeOH) were performed. A step of 1 min returning to the initial condition and finally 3 min of equilibration were followed. The flow rate was 500 $\mu\text{L}/\text{min}$. LC experiments were performed using ESI source in negative mode and the operating conditions were identical as described above apart from the flow rate of nitrogen which was used as sheath and auxiliary gas and adjusted to 40 and 20 arbitrary units, respectively. A method consisting of a full scan and a data dependent acquisition was developed. In the full scan experiments the mass range was set between m/z 50-800 and the FT resolution at 30000 (FWHM). The data dependent MS^n experiments using CID value of 22% were performed at 7500 (FWHM) mass resolution.

Data acquisition and analysis were performed using Xcalibur software version 2.0.7 (Thermo Scientific). For identification of AtA specific criteria were set. Specifically, for elemental composition (EC) prediction, C (max no. 40), H (max no. 50), O (max no. 12) and Na (max no. 1) were selected as preferred elements. Mass tolerance was set to 3 ppm for all ions (full scan and MS^n) while the RDBeq. values were restricted between 0 and 16. Using Xcalibur peak peaking algorithms, ion lists were generated together with the corresponding relative intensities, Δ values (in ppm), RDBeq. values and suggested EC for all fragments. Using Excel environment MS^n spectral trees were constructed. The generation of proposed fragmentation mechanisms and spectral trees were facilitated using Mass Frontier software (HighChem, Slovakia).

Results and Discussion

1. Structure elucidation of AtA using multistage HRMS

Acronychia-type acetophenones (AtA) are prenylated acetophenone dimers, highly symmetric and fully substituted, consisted of two aromatic rings A and B linked with an isopentyl chain between C-5 and C-5'''. Regarding all AtA reported so far, ring B, is constant, substituted with two hydroxyl (C-4''' and C-6'''), an acetyl (C-1'''), an isoprenyl (C-3''') and a methoxy group (C-2'''). In contrast, ring A is characterized by different substitution patterns; typically modified isoprenyl chain (compounds **1**, **2**) or additional rings (compounds **3-7**). According to the fusion position of the additional ring, AtA could be divided to type-L (compounds **4**, **6**) or type-R (compounds **3**, **5**, **7**) where the ring is formed at position 3 and 4 or position 2 and 3, respectively (Figure 17).



Chapter 2: LC-MS based structural elucidation and dereplication of AtA

Figure 17: The different structures of reference *Acronychia*-type acetophenones (AtA). AtA are characterized according to their substitution patterns; AtA with additional rings are annotated as type-R and type-L indicating the fusion orientation of the ring towards right or left, respectively.

During this study, all reference AtA were studied using API ionization techniques in order to investigate the ionization behavior of this chemical group. In this context, the full scan and MSⁿ high resolution spectra of all AtA compounds were acquired using ESI(±) and APCI(±). Thus, the spectral trees for all compounds were constructed with the assistance of the Excel and Mass Frontier software (Figure A 61, Figure A 62). Only in case of APCI(+) the construction of spectral trees was not possible as the molecular ion was generated in low abundance due to the extended fragmentation under these conditions. Generally, AtA upon subjected to fragmentation, regardless the applied ionization source and mode, presented characteristic fragments derived from the instant rupture of the bonds connecting the aromatic rings with the isopentyl chain. Since, there is no previous data regarding the analysis of AtA compounds or similar groups by MS, a systematic nomenclature was developed and proposed.

Mabry and Markham developed the widely used nomenclature system for flavonoids which has been evolved by Claeys and co-workers (Ma et al. 1997) while Costello and coworkers (Domon et al. 1988) have been introduced a systematic nomenclature for glycoconjugates. Following a similar rationale, a nomenclature system is proposed based on characteristic cleavages related to rings A and B. General trends employed in systematic studies investigating the fragmentation patterns of aliphatic chains by straightforward bond cleavage were taken under consideration as well (Klagkou et al. 2003). Specifically, **A** and **B** species refer to the cleavage of the corresponding aromatic rings while the subscripts **1** or **0** indicate the presence or not of the bridge isopentyl chain on the derived fragments, respectively (Figure 18). Therefore, typical product ions generated from adduct, protonated or deprotonated ions of AtA, in both modes could be annotated following the proposed nomenclature scheme (Table VII).

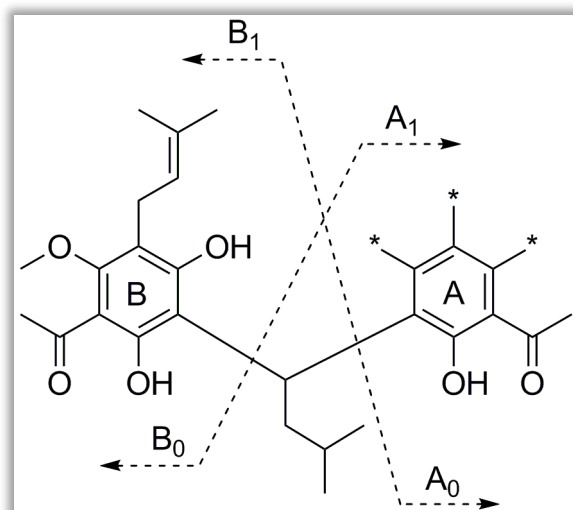


Figure 18: Basic fragmentation pattern of AtA and nomenclature of product species. *Positions of substituents in reference AtA.

Common losses and fragmentation patterns are generally observed in positive and negative mode when AtA are analyzed. For instance, the characteristic loss of C_4H_8 (-56 u) is occurred in both modes denoting the cleavage of an isoprenyl or a modified isoprenyl moiety. This cleavage as well as the loss of small neutral units, such as H_2O (-18 u), CH_4 (-16 u), C_2H_2O (-42 u), CO (-28 u) are also commonly generated from adducts, protonated or deprotonated molecule ions and fragment ions thereof. It is worth mentioning that aromatic rings remained intact under CID conditions in both ionization sources and modes which demonstrates their stability under these conditions (Klagkou et al. 2003).

Table VII: Typical product ions, in the HRMS/MS spectra of reference compounds in ESI ($\pm$), annotated according to the proposed nomenclature of AtA.

POSITIVE ION MODE			Compound	NEGATIVE ION MODE		
Fragment ion	Precursor ions (m/z)	Fragment ion (m/z)		Fragment ion	Precursor ions (m/z)	Fragment ion (m/z)
[A_1+Na] $^+$ or [$M+Na-B_0$] $^+$	593.2713	343.1511	2, 4-7	[A_1-H] $^-$ or [$M-H-B_0$] $^-$	569.2750	319.1547
	577.2766	327.1563	1		553.2803	303.1600
	575.2599	325.1405	3		551.2646	301.1441
[A_0+Na] $^+$ or [$M+Na-B_1$] $^+$	593.2713	275.0889	2, 4-7	[A_0-H] $^-$ or [$M-H-B_1$] $^-$	569.2750	251.0923
	577.2766	n.o	1		553.2803	235.0976
	575.2599	n.o	3		551.2646	233.0817
[B_1+Na] $^+$ or	593.2713	341.1719	2, 4-7	[B_1-H] $^-$ or	569.2750	317.1753

$[M+Na-A_0]^+$	577.2766		1	$[M-H-A_0]^-$	553.2803	
	575.2599		3		551.2646	
$[B_0+Na]^+$ or $[M+Na-A_1]^+$	593.2713	273.1096	2, 4-7	$[B_0-H]^-$ or $[M-H-A_1]^-$	569.2750	
	577.2766	n.o	1		553.2803	249.1130
	575.2599	273.1096	3		551.2646	

1.1. Fragmentation pattern analysis for AtA in ESI(+)

The behavior of reference AtA was initially investigated by infusion method. Based on the HRMSⁿ spectra the ion trees of compounds **1-7** were constructed. All ions reaching a 5% intensity threshold in all levels were selected for further analysis. As first general rule, it was observed that all compounds upon subjected to ESI(+) were detected as stable adduct ions with sodium $[M+Na]^+$ in full scan mode while the protonated molecular ions were completely absent under the given conditions. Similarly, all the fragment ions in MSⁿ spectra were also detected as sodium adducts. This ionization preference could be explained by the particular polyhydroxylated nature of AtA offering influential localized charge on the multiple oxygen atoms as well as their chelating ability which is proven by their structural motif and their tendency to form heterodimers (Kruve et al. 2013).

Acrovestone (**1**), a model AtA compound bearing two prenyl groups on C-3 of each aromatic ring, yielded characteristic fragment ions in MS², MS³ and MS⁴ spectra upon subjected to ESI(+) observed in all reference compounds. In particular, two characteristic fragmentation motifs were observed i) the elimination of intact aromatic rings by rupture of C-5 or C-5''' and ii) the consecutive loss of C₄H₈ (56u) indicating the fission of isoprenyl and/or isopentyl chains (Figure 19).

Thus, in MS² level the characteristic fragment ions at m/z 327.1563 ($[A_1+Na]^+$) and 341.1719 ($[B_1+Na]^+$) were arose from the loss of B₀ and A₀ moieties, respectively. Further loss of C₄H₈ groups was observed from the $[A_1+Na]^+$ and $[B_1+Na]^+$ fragment ions in the MS³ spectra yielding the fragment ions at m/z 271.0939 ($[A_1+Na-C_4H_8]^+$) and at m/z 285.1096 ($[B_1+Na-C_4H_8]^+$), respectively. The latter was also generated directly from the $[M+Na]^+$ ion in the MS² spectrum. Finally, the $[B_1+Na-C_4H_8]^+$ ion underwent a consequent neutral loss of a C₄H₈ group to produce the fragment ion at m/z 229.0471 ($[B_1+Na-C_8H_{16}]^+$) as detected in the MS³ and MS⁴ spectra (Figure 19).

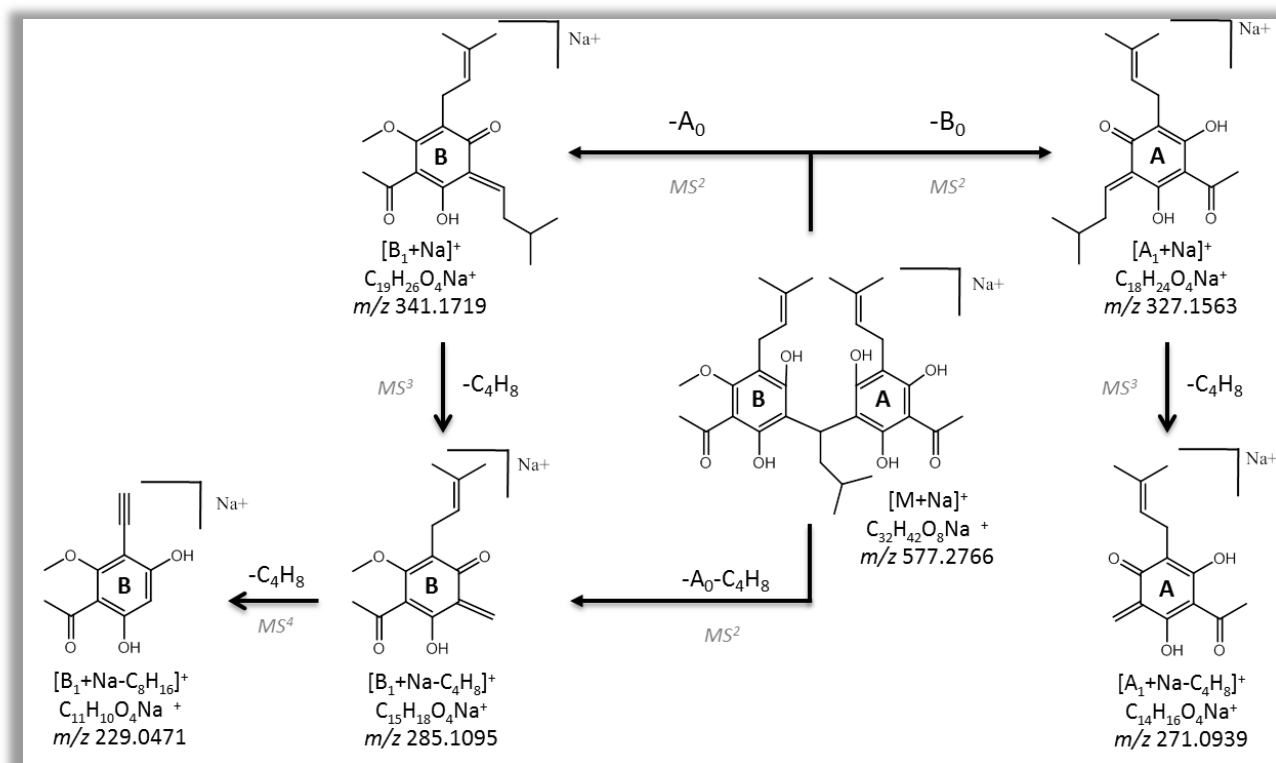


Figure 19: Proposed fragmentation pathways for acrovostone (1) in ESI(+)

Acrovostenol (2), an AtA bearing a hydroxylated modified isoprenyl chain at the position C-3 of the A ring, formed $[M+Na]^+$ adduct ion at m/z 593.2710 under ESI(+). At MS^2 , the main fragment ions occurred by the elimination of A_0 ($[B_1+Na]^+$, m/z 341.1719)], $A_0+C_4H_8$ ($[B_1+Na-C_4H_8]^+$, m/z 285.1094)], and B_0 ($[A_1+Na]^+$, m/z 343.1511)] species similarly to 1. Moreover, a fragment ion corresponding to elimination of B_1 ($[A_0+Na]^+$, m/z 275.0887) was also observed (Figure A 63).

The presence of the additional hydroxyl group on the ring A favoured the neutral loss of H_2O and C_4H_8O in MS^3 level from both $[A_1+Na]^+$ and $[A_0+Na]^+$ ions resulting to the generation of the $[A_1+Na-H_2O]^+$ (m/z 325.1409), $[A_1+Na-C_4H_8O]^+$ (m/z 271.0940), $[A_0+Na-H_2O]^+$ (m/z 257.0781) and $[A_0+Na-C_4H_8O]^+$ (m/z 203.0312) fragment ions, respectively. According to the MS^3 spectra, the loss of H_2O compared to the loss of the C_4H_8O is more preferable from $[A_1+Na]^+$ ion than from $[A_0+Na]^+$. Thus, a high abundance of the $[A_1+Na-H_2O]^+$ ion (100%) was observed while a relatively smaller abundance was recorded for the $[A_0+Na-H_2O]^+$ product ion (32%). In contrary, the $[A_1+Na-C_4H_8O]^+$ ion was observed in 32% and the $[A_0+Na-C_4H_8O]^+$ ion in 100% of relative abundance. It is important to note that the $[A_0+Na-H_2O]^+$ was not observed in any other AtA under investigation at any MS^n level and

could be considered indicative for AtA with hydroxylated side chains (total number of oxygen atoms is 9 vs 8 in model AtA 1). Regarding ring B, the cleavage of the A_0 and $A_0+C_4H_8$ groups from the $[M+Na]^+$ ion observed in the MS^2 spectrum revealed the characteristic AtA fragmentation pattern which includes the successive loss of C_4H_8 observed in the MS^2 , MS^3 and MS^4 levels forming $[B_1+Na]^+$, $[B_1+Na-C_4H_8]^+$ and $[B_1+Na-C_8H_{16}]^+$ fragment ions (Table VIII, Figure A 63).

Acropyrone (3) is an AtA derivative with an additional dimethylpyran ring on the main structure at C-2 and C-3 positions of ring A. In MS^2 spectra the typical $[B_1+Na]^+$, $[B_1+Na-C_4H_8]^+$ and $[A_1+Na]^+$ ions were observed following the general fragmentation rules of AtA. All generated ions underwent further elimination of C_4H_8 groups resulting to the formation of $[B_1+Na-C_8H_{16}]^+$ (m/z 229.0468) in MS^3 and MS^4 levels and $[A_1+Na-C_4H_8]^+$ (m/z 269.0781) in MS^3 . In addition, $[B_0+Na]^+$ (m/z 273.1093) in MS^2 together with $[B_0+Na-C_4H_8]^+$ (m/z 217.0469) in MS^3 spectra were detected resulting from the cleavage of A_1 and $A_1-C_4H_8$, respectively. These latter ions were generated in all cases of AtA with additional ring and could be considered as key fragments for the identification thereof (Table VIII, Figure A 64).

Acrofolione A (4) and acrofolione B (5) are regioisomers possessing an additional 2-(2-hydroxypropan-2-yl)-2,3-dihydro-1-furan ring fused at different positions on the parent structure which influences their fragmentation behavior. For instance, 5 (type-R) compared to 4 (type-L) presented more extended fragmentation in MS^2 , MS^3 and mainly in MS^4 spectra. More specifically, 4 revealed four main fragment ions in the MS^2 spectrum. Apart from the characteristic $[A_1+Na]^+$, $[B_1+Na]^+$ and $[B_1+Na-C_4H_8]^+$ ions of AtA, the indicative for AtA with additional ring $[B_0+Na]^+$ (m/z 273.1095) ion and $[B_0+Na-C_4H_8]^+$ (m/z 217.0473) in MS^3 were detected. All fragment ions obtained from $[M+Na]^+$ in MS^2 spectrum underwent further fragmentation resulting to the elimination of a C_4H_8 group from each precursor ion giving rise to the $[A_1+Na-C_4H_8]^+$, $[B_1+Na-C_4H_8]^+$ and $[B_1+Na-C_8H_{16}]^+$ ions, respectively. At MS^4 level only the $[B_1+Na-C_8H_{16}]^+$ ions were detected derived from the $[B_1+Na-C_4H_8]^+$ parent ions (Table VIII, Figure A 65).

On the other hand, 5 presented similar fragmentation pattern to 4 and all the ions detected in the MS^n spectra of 4 were also present in the MS^n spectra of 5, accordingly. However, additional ions were recorded in 5 revealing another pattern. In particular, $[A_0+Na]^+$ at m/z 275.0889 was detected in MS^2 which underwent further fragmentation (MS^3) giving rise to $[A_1+Na-C_4H_{10}O]^+$ (m/z 269.0783) and $[A_1+Na-C_8H_{16}O]^+$ (m/z 215.0313) ions (Table VIII, Figure A 66). It is obvious that the occurrence of these ions is due to the

presence of an additional hydroxyl group on ring A. These ions were not detected in the case of **3** since the lack of free hydroxyl group disables the elimination of $C_4H_{10}O$ and $C_8H_{16}O$ species. Nevertheless, their absence from the spectra of **4** designates that the specific losses are in favour when the fusion occurs between C-2 and C-3 positions (type-R).

The other pair of regioisomers, acropyranol A (**6**) and acropyranol B (**7**) differ only at the fusion position of the additional 3-hydroxy-2,2-dimethyl-3,4-dihydro-2H-pyrano ring and comprise structural isomers of **2**, **4** and **5** as well. At their MS^n spectra all specific ions indicating the AtA identity and occurrence of hydroxylated additional ring are present. Moreover, characteristic ions representing the fusion ring orientation could be observed. Specifically, the $[M+Na]^+$ adduct ion of **6** in MS^2 level revealed the $[B_1+Na]^+$, $[B_1+Na-C_4H_8]^+$, $[A_1+Na]^+$ and $[B_0+Na]^+$ fragment ions (Table 2, Figure 4). The generated product ions were further underwent C_4H_8 elimination including the $[A_1+Na]^+$ ion which, in contrary to **4**, gave rise to an additional fragment ion at m/z 271.0938 ($[A_1+Na-C_4H_8O]^+$) in the MS^3 spectrum. This fragment ion with relative intensity 10% derived from the elimination of a C_4H_8O group and is considered as characteristic fragment for acropyranols as it is present in both **6** and **7**.

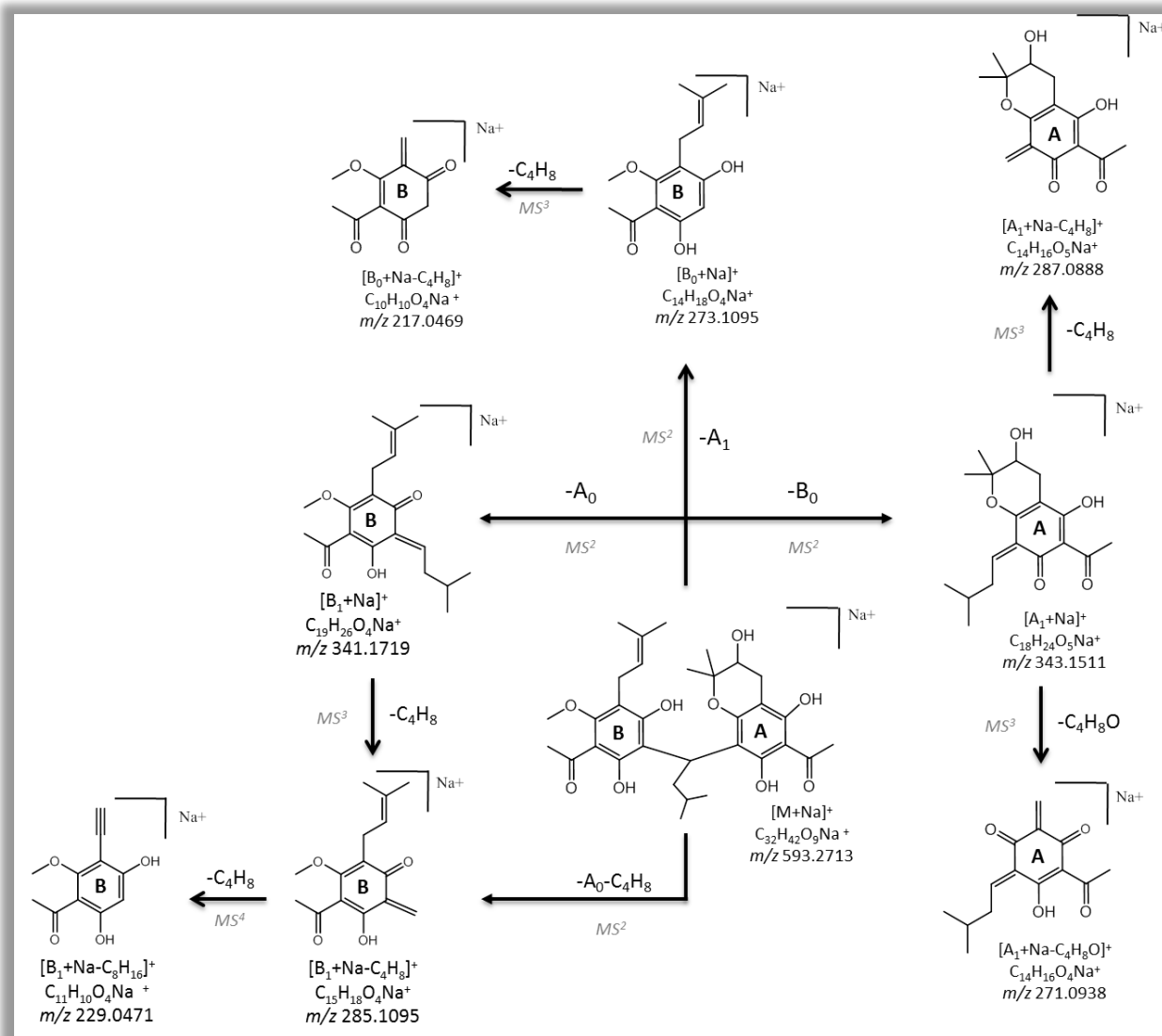


Figure 20: Proposed fragmentation pathways for acropyranol A (6) in ESI(+)

Acropyranol B 7 (type-R) in comparison to 6 (type-L) presented additional fragment ions in all MS^n spectra similarly to 5 (type-R). Specifically, in MS^2 spectrum, the $[A_0+Na]^+$ fragment ion arose at m/z 275.0888 similarly to 2 and 5 which afforded the $[A_0+Na-C_4H_8O]^+$ (m/z 203.0313) ion in the MS^3 after the elimination of a C_4H_8O group. Very informative was the fragmentation behavior of $[A_1+Na]^+$ ion which generated the key fragment ion of acropyranol compounds $[A_1+Na-C_4H_8O]^+$ (m/z 271.0938); the $[A_1+Na-C_4H_8]^+$ (m/z 287.0888) ion which was oppositely favored compared to 6 (rel. abundance 40% vs 100%) as well as the least abundant $[A_1+Na-H_2O]^+$ (m/z 325.1407) ion. Finally, similarly to 5 the fragmentation of the precursor $[A_1+Na-C_4H_8]^+$ ion generated the $[A_1+Na-C_8H_{16}O]^+$ (m/z 215.0312) and $[A_1+Na-C_4H_{10}O]^+$ (m/z 269.0783) ions derived from the

elimination of C_4H_8O and H_2O . The presence of both fragment ions indicates the characteristic fragmentation behavior of AtA type-R (Table VIII, Figure 21).

Chapter 2: LC-MS based structural elucidation and dereplication of AtA

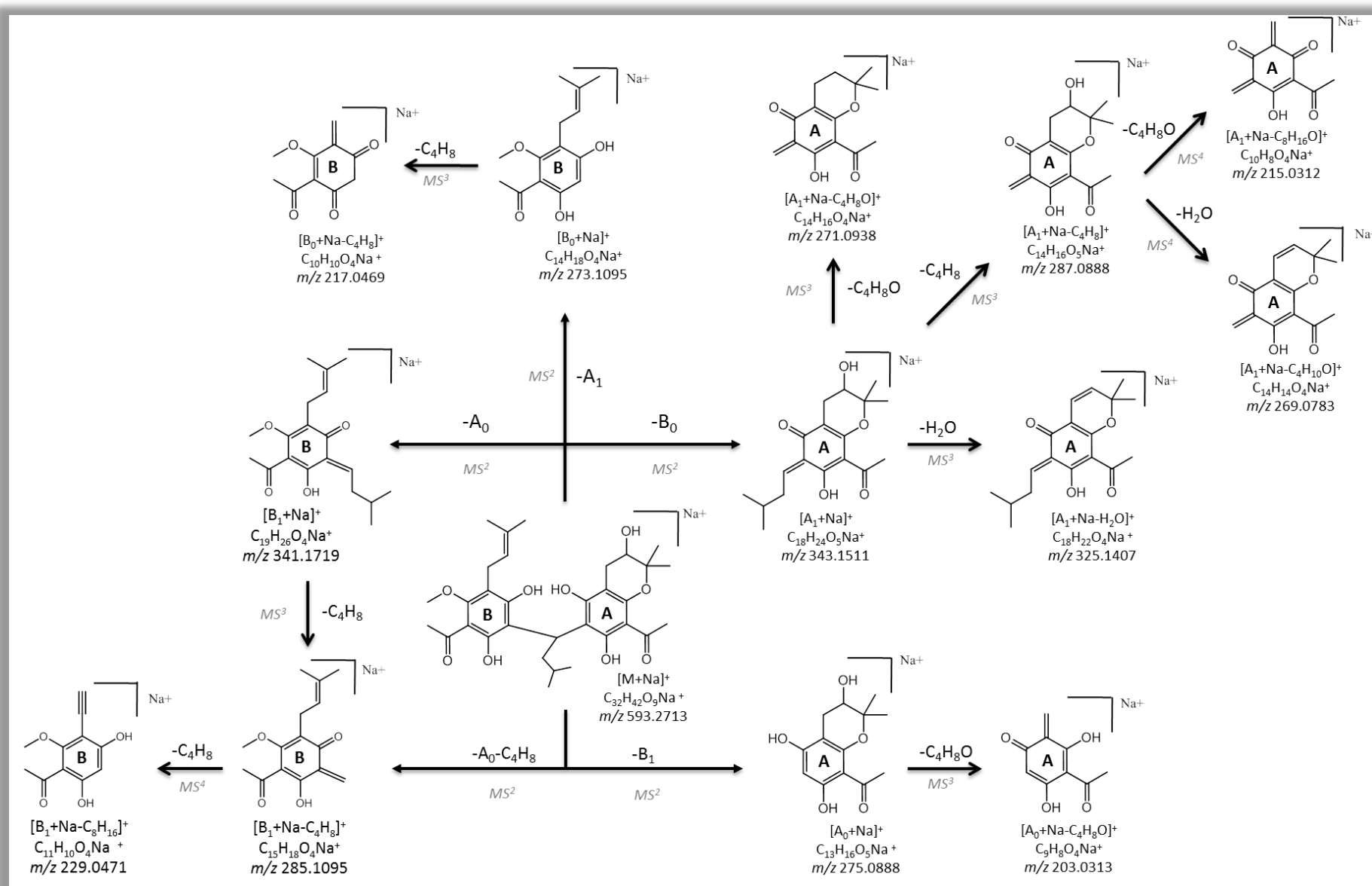


Figure 21: Proposed fragmentation pathways for acropyranol B (7) in ESI(+)

Table VIII: HRMS and HRMSn ions of reference compounds (1-7) in ESI(+); Precursor ions are in bold and highlighted

Compound	Pseudomolecular ions [M+Na] ⁺ m/z (% base peak)	Fragment ions (MS ²) m/z (% base peak)	Fragment ions (MS ³) m/z (% base peak)	Fragment ions (MS ⁴) m/z (% base peak)
Acrovestone (1)	577.2766 (29/ -0.95)	<u>341.1719 (100/ -1.27)</u> , <u>327.1563 (15/ -1.08)</u> , <u>285.1095 (6/ -0.81)</u>	<u>285.1096 (100)</u> ; 271.0939 (100); 229.0471 (100)	229.0471 (100)
Acrovestenol (2)	593.2710 (44/ -1.87)	<u>343.1511 (18/ -1.5)</u> , <u>341.1719 (100/ -1.4)</u> , <u>285.1094 (5/ -1.2)</u> , <u>275.0887 (8/ -1.11)</u>	325.1409 (100), 271.0940 (39); <u>285.1094 (100)</u> ; 229.0469 (100); 257.0781 (32), 203.0312 (100)	229.0469 (100)
Acropyrone (3)	575.2599 (13/ -2.74)	<u>341.1717 (100/ -1.72)</u> , <u>325.1405 (19/ -1.52)</u> , <u>285.1093 (5/ -1.54)</u> , <u>273.1093 (4/ -1.52)</u>	<u>285.1093 (100)</u> ; 269.0781 (100); 229.0468 (100); 217.0469 (100)	229.0468 (100)
Acrofolione A (4)	593.2715 (11/ -1.48)	<u>343.1512 (65/ -0.91)</u> , <u>341.1722 (100/ -0.35)</u> , <u>285.1095 (7/ -0.56)</u> , <u>273.1095 (18/ -0.62)</u>	287.0888 (100); <u>285.1095 (100)</u> ; 229.0471 (100); 217.0473 (100)	229.0471 (100)
Acrofolione B (5)	593.2713 (12/ -1.74)	<u>343.1513 (60/ -1.31)</u> , <u>341.1722 (100/ -0.75)</u> , <u>285.1095 (13/ -1)</u> , 275.0889 (28/ -0.85), <u>273.1097 (3/ -0.71)</u>	<u>287.0888 (100)</u> ; <u>285.1095 (100)</u> ; 229.0471 (100); 217.0473 (100)	269.0783 (100), 215.0313 (70); 229.0471 (100)
Acropyranol A (6)	593.2713 (67/ -1.25)	<u>343.1511 (12/ -1.54)</u> , <u>341.1719 (100/ -1.19)</u> , <u>285.1095 (8/ -0.97)</u> , <u>273.1095 (3/ -0.99)</u>	287.0888 (100), 271.0938 (40); <u>285.1095 (100)</u> ; 229.0469 (100); 217.0469 (100)	229.0469 (100)
Acropyranol B (7)	593.2713 (75/ -1.6)	<u>343.1511 (31/ -1.23)</u> , <u>341.1722 (100/ -0.58)</u> , <u>285.1095 (9/ -0.52)</u> , <u>275.0888 (5/ -0.49)</u> , <u>273.1096 (5/ -0.44)</u>	325.1407 (10), <u>287.0888 (40)</u> , 271.0938 (100); <u>285.1095 (100)</u> ; 229.0470 (100); 203.0313 (100); 217.0469 (100)	269.0783 (100), 215.0312 (19); 229.0470 (100)

1.1.1. Key features for identification of AtA in ESI(+)

The behavior of **1-7** under ESI(+) revealed a specific fragmentation pattern including key fragment ions and characteristic cleavages that could be used for the identification of AtA, the separation of region and structural AtA isomers as well as their detection in complex mixtures (**Figure 22**). These features are summarized below.

~ All AtA are detected as adducts ions with sodium $[M+Na]^+$ and the respective pseudomolecular ions $[M+H]^+$ are not observed in full scan HRMS.

~ In all MS levels both aromatic rings A and B remain intact.

~ $[B_1+Na]^+$ (A_0 loss) and $[B_1+Na-C_4H_8]^+$ (A_0 & C_4H_8 loss) ions in MS^2 and $[B_1+Na-C_8H_{16}]^+$ ion in MS^4 levels are always present and could be considered as diagnostic for the identification of AtA. Additionally, $[B_1+Na]^+$ (m/z 341.1719) ion is recorded as base peak in all MS^2 spectra.

~ $[A_1+Na]^+$ (B_0 loss) is present in all AtA and its relative abundance varies according to the substitution motif of ring A.

~ The formation of $[B_0+Na]^+$ (m/z 273.1096) ions arising from the loss of A_1 moiety are detected in the MS^2 level only when additional rings are present on the parent structure (**3-7**). Moreover, these ions undergo further elimination of a C_4H_8 moiety giving rise to the $[B_0+Na-C_4H_8]^+$ ions (m/z 217.0469). Both type of ions are indicative and could be used for the identification of AtA with additional ring.

~ The ions $[A_1+Na-C_4H_{10}O]^+$ and $[A_1+Na-C_8H_{16}O]^+$ at MS^4 level are recorded only when the additional hydroxylated fused ring has right orientation and could facilitate the identification of AtA type-R (**5 & 7**).

~ $[A_0+Na]^+$ (B_1 loss) ions are observed in MS^2 only in case of hydroxylated modified side chains and presence of additional type-R ring.

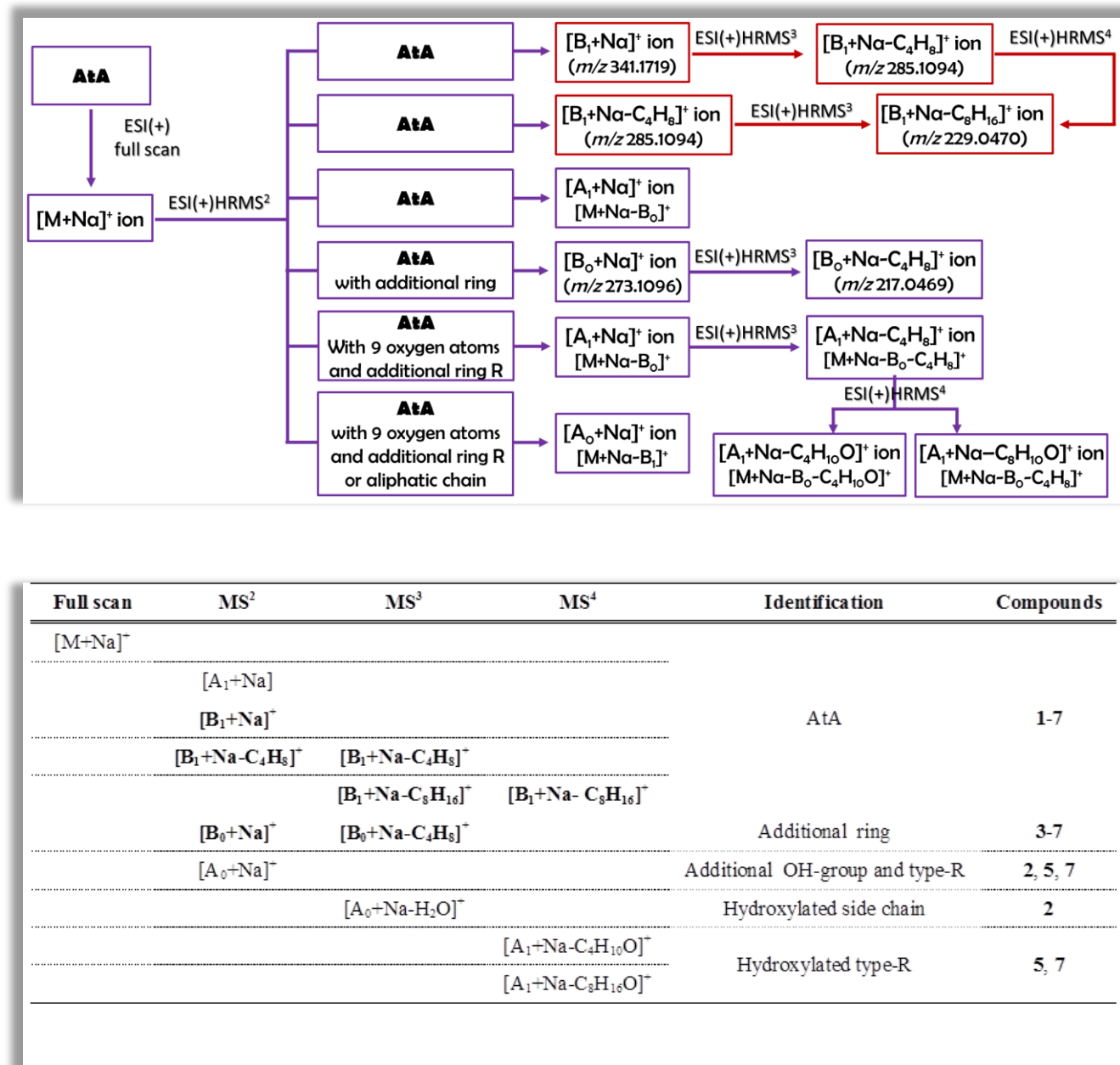


Figure 22: Decision tree and table for structural characterization of AtA in ESI (+)

1.2. Fragmentation pattern analysis for AtA in ESI(-)

In order to investigate the fragmentation behavior of AtA in negative ionization, all reference compounds were analyzed generating HRMSⁿ spectra for all ions with relative intensity greater than 5%. Generally, AtA upon subjected to ESI(-) generate mainly deprotonated ions in full scan and HRMSⁿ spectra. Compared to positive mode, more extensive fragmentation was observed presenting improved sensitivity.

Specifically, the ionization of acrovestone (**1**) under negative conditions generated the $[M-H]^-$ ion at m/z 553.2803 which was selected and further analyzed by HRMSⁿ. The MS² spectrum was dominated by three principal ions $[B_1-H]^-$, $[A_1-H]^-$ and $[A_0-H]^-$ which arose from the rupture of the C-C bonds connecting the two aromatic rings with the isopentyl chain. In detail, the abovementioned ions observed at m/z 317.1756, 303.1600 and 235.0979 were generated from the neutral loss of A₀, B₀ and B₁ moieties, respectively. It seems that these fragment ions are formed following a heterolytic fission mechanism and are stabilized through different electron delocalization (Furtado et al. 2007). The relative abundance of all the typical ions are high and comparable (100%, 75% and 99%, respectively) while the $[B_0-H]^-$ ion (m/z 249.1123) is detected in very low abundance (6%) and practically could be considered absent (Table IX). This probably indicates that the methyl group (B ring) constrains the stabilization of the derived ions and unsubstituted phloroglucinol core is required for the generation of the $[B_0-H]^-$ ion.

Moreover, the losses CH₃, C₄H₁₀ and C₅H₁₀ units from $[B_1-H]^-$ ion (m/z 317.1756) as observed in the MS³ spectrum gave rise to the $[B_1-H-CH_3]^-$ (m/z 302.1519), $[B_1-H-C_4H_{10}]^-$ (m/z 259.0973) and $[B_1-H-C_5H_{10}]^-$ (m/z 247.0973) ions, respectively. The radical ion which derives from the loss of CH₃ could be stabilized in different resonance states and in accordance to other chemical groups of natural compounds, it is highly stable (61%) (Justesen 2001). This motif is repeated in all AtA under investigation (Table IX).

Likewise, the loss of C₄H₁₀ is typical for all AtA and indicative for the occurrence of isoprenyl chain. It seems that the presence of two oxygen species in *ortho* position favors the elimination of the C₄H₁₀ unit and could probably correlated to *ortho* elimination mechanisms known in EI-MS (Gross 2011). The major $[B_1-H-C_4H_{10}]^-$ ion underwent further fragmentation revealing the fragment ions $[B_1-H-C_5H_{14}]^-$ (m/z 243.0658), $[B_1-H-C_6H_{12}O]^-$ (m/z 217.0869) and $[B_1-H-C_8H_{17}]^-$ (m/z 204.0428) in the MS⁴ spectrum deriving from the loss of CH₄, C₂H₂O and C₄H₇ units, respectively.

Regarding the $[A_1-H]^-$ ion when subjected to further fragmentation, the elimination of CO₂ was favored giving rise to the $[A_1-H-CO_2]^-$ fragment ion at m/z 259.1701 compared to smaller fragment ions derived from the cleavage of C₂H₂O and C₈H₁₆ groups as observed in the MS³ spectrum. The CO₂ loss is a characteristic in negative mode, in all MS levels and in all AtA derivatives which probably occurs through rearrangements and *ortho* elimination mechanisms. Finally, the fragmentation of the $[A_0-H]^-$ ion is consisted of the loss of CO₂ moiety giving rise to the fragment ion at m/z 191.1079 in the MS³ spectrum. The complete

fragmentation pattern of **1** is illustrated in Figure 23 while all the derived fragments in MS²-MS⁴ levels, the corresponding precursor ions, together with their relative abundances are presented in Table IX.

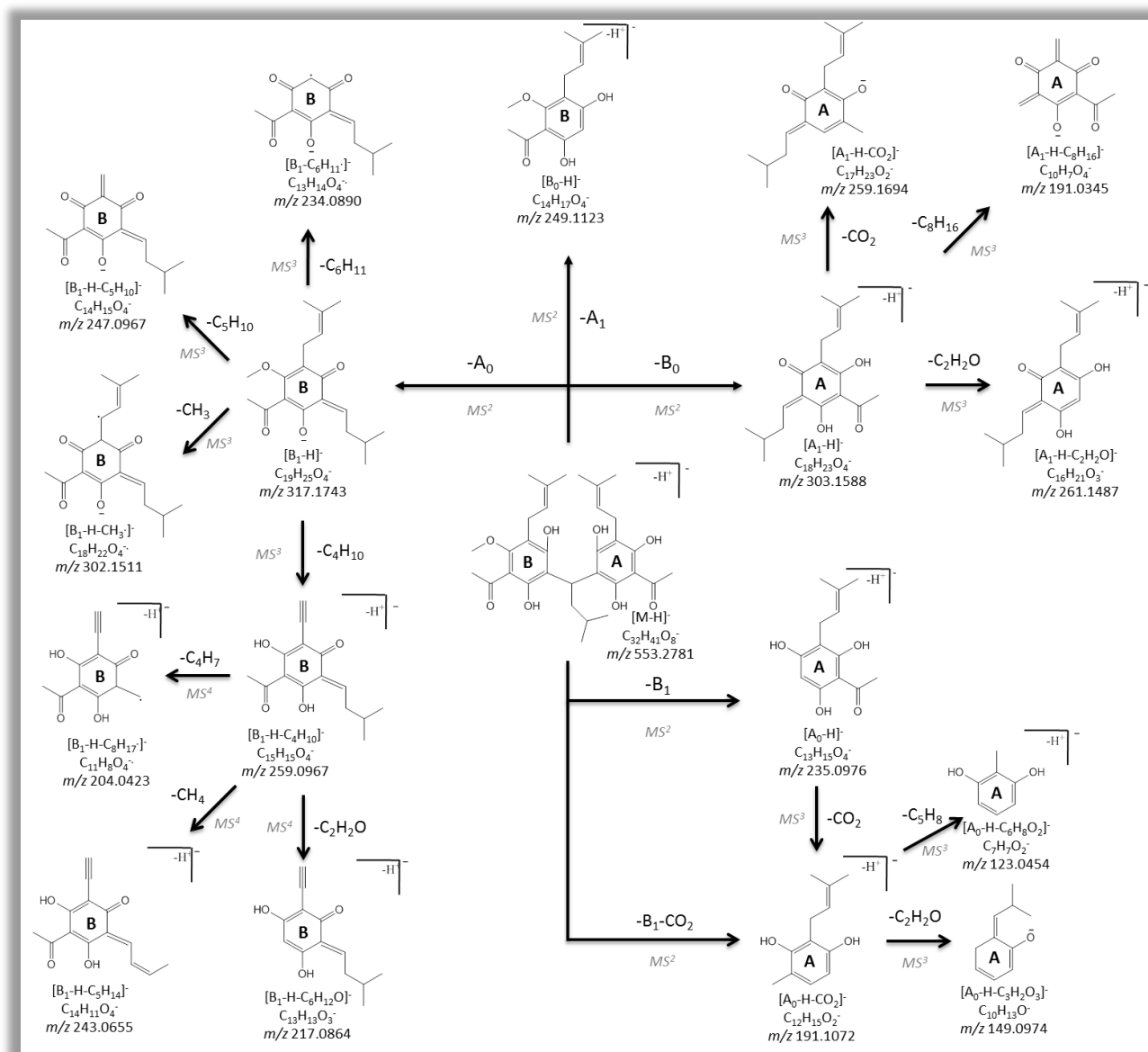


Figure 23: Proposed fragmentation pathways for acrovestone (1) in ESI(-)

In acrovostenol (**2**) compared to acrovestone (**1**), the presence of the additional hydroxyl group on the isoprenyl unit at ring A led to an enhanced ionization in negative conditions. The fragmentation of the $[M-H]^-$ ion observed in full scan acquisition resulted in the generation of the characteristic $[B_1-H]^-$, $[A_1-H]^-$ and $[A_0-H]^-$ fragment ions at m/z 317.1754, 319.1546 and 251.0923 respectively with the last ion detected as BP (Table IX, Figure A 67).

In MS^3 and MS^4 the fragmentation pattern for the $[B_1-H]^-$ ions is similar to **1** while differences are observed regarding the more stable $[A_1-H]^-$ and $[A_0-H]^-$ ions. It is worth noting that extensive fragmentation is taking place especially for these ions indicating that hydroxylated side chains are more prone to fission. For instance, neutral losses of H_2O , C_4H_6O and C_4H_8O moieties were detected providing the $[A_1-H-H_2O]^-$ (m/z 301.1442), $[A_1-H-C_4H_6O]^-$ (m/z 249.1130) and $[A_1-H-C_4H_8O]^-$ (m/z 247.0974) fragment ions in the MS^3 spectrum with $[A_1-H]^-$ as precursor ion. The fragmentation of $[A_0-H]^-$ generated $[A_0-H-H_2O]^-$ (m/z 233.0817), $[A_0-H-C_2H_2O]^-$ (m/z 209.0818) and $[A_0-H-C_4H_8O]^-$ (m/z 179.0351) in the MS^3 spectrum. Among them, characteristic is the loss of H_2O in MS^3 giving rise to both $[A_1-H-H_2O]^-$ and $[A_0-H-H_2O]^-$ which are detected as base peaks. This loss is detected only in **2** and it is indicative for AtA having an additional hydroxyl group on the isoprenyl chain. Finally, these ions undergo further cleavages of C_2H_2O and CO_2 moieties giving rise to the $[A_1-H-C_2H_4O_2]^-$ (m/z 259.1338), $[A_1-H-CH_2O_3]^-$ (m/z 257.1545), $[A_0-H-C_2H_4O_2]^-$ (m/z 191.0714) and $[A_0-H-CH_2O_3]^-$ (m/z 189.0921) as observed in the MS^4 spectra.

The negative ionization of acropyrone (**3**) generated the less abundant $[M-H]^-$ ion, among the other compounds under study, as observed at the full scan acquisition. Acropyrone as all reference compounds presented the $[B_1-H]^-$, $[A_1-H]^-$ and $[A_0-H]^-$ key fragment ions in MS^2 detected at m/z 317.1753, 301.1441 and 233.0817, respectively. Additionally, the $[B_0-H]^-$ ion (m/z 249.1130) was observed with relative intensity 33%. This ion all **3-7** compounds and could be consider as indicative for AtA compounds with additional ring. In MS^3 level, the fragmentation scheme of the $[B_1-H]^-$ ion was identical to all reference compounds characterized by the loss of CH_3 , C_4H_{10} , and C_5H_{10} units. The same losses were observed from $[B_0-H]^-$ ion generating the $[B_0-H-CH_3]^-$ (m/z 234.0896), $[B_0-H-C_4H_{10}]^-$ (m/z 191.0353) and $[B_0-H-C_5H_{10}]^-$ (m/z 179.0351) ions respectively as observed in the MS^3 . Finally, the fragmentation of the $[A_0-H]^-$ ion led to the ions at m/z 191.0715, 189.0922 and 187.0765 attributed to the $[A_0-H-C_2H_2O]^-$, $[A_0-H-CO_2]^-$ and $[A_0-H-CH_2O]^-$ ions respectively while it seems that the dimethylpyran ring remains intact (Table IX, Figure A 68).

Acrofolione A (**4**) and acrofolione B (**5**) presented similar fragmentation motif in all levels however more extended compared to other AtA. Specifically, apart from the expected predominant $[A_1-H]^-$, $[B_1-H]^-$ and $[A_0-H]^-$ ions, the $[B_0-H]^-$ fragment ion at m/z 249.1129, indicative for additional ring, was also observed. The $[A_1-H]^-$ ion was more abundant in the MS^2 spectrum of acrofolione A comparing to acrofolione B (20% in **4** vs 87% **5**) (Table IX).

This difference in the intensity of $[A_1-H]^-$ ions could be possibly attributed to steric hindrance considerations (Sawada et al. 1992). Moreover, some minor ions at m/z 259.0972, 193.0505 and 191.0349 attributed to the $[B_1-H-C_4H_{10}]^-$, $[A_0-H-C_3H_6O]^-$ and $[B_0-C_4H_{10}]^-$ ions respectively were also present. In the MS^2 level the only difference between the two isomers regards the presence of the relatively small abundant $[A_0-H-C_4H_8O]^-$ or $[B_0-H-C_5H_{10}]^-$ fragment ion (m/z at 179.0350) in acrofolione A (**4**) and the $[A_1-H-C_3H_6O]^-$ ion at m/z 261.1129 in acrofolione B (**5**).

For both compounds, further fragmentation of the $[A_1-H]^-$ ion led to the generation of a main fragment ion at m/z 261.1130 corresponding to the $[A_1-H-C_2H_2O]^-$ ion. Moreover, the key fragment ions of the $[B_1-H]^-$ ion corresponding to the loss of CH_3 , C_4H_{10} and C_5H_{10} moieties were observed at m/z 302.1520, 259.0975 and 247.0976 as in the previously mentioned AtA. The same losses were detected for the $[B_0-H]^-$ fragment ion yielding three highly abundant ions at m/z 234.0897, 191.0351 and 179.0351, respectively. Finally, the fragmentation of the $[A_0-H]^-$ ion revealed the most abundant $[A_0-H-C_3H_6O]^-$ ion at m/z 193.0506 corresponding to the cleavage of C_3H_6O and the small abundant $[A_0-H-C_4H_8O]^-$ ion at m/z 179.0351 as observed in the MS^3 spectrum (Table IX, Figure A 69, Figure A 70).

Acropyranol A (**6**) and acropyranol B (**7**), like the majority of reference compounds presented the principal $[A_1-H]^-$, $[B_1-H]^-$, $[A_0-H]^-$ and $[B_0-H]^-$ fragment ions observed at m/z 319.1546, 317.1754, 251.0923 and 249.1130 in the MS^2 spectra. Additionally, fragment ions at m/z 259.0974 and 179.0351 were observed derived from the loss of C_4H_{10} and C_4H_8O units in both isomers. Only acropyranol B yielded the ions at m/z 247.0974 and 191.0350 attributed to the $[A_1-H-C_4H_8O]^-$ or $[B_1-H-C_5H_{10}]^-$ and $[B_0-H-C_4H_{10}]^-$ ions, respectively.

Similarly to the other pair of isomers **4** & **5**, the $[A_1-H]^-$ ion was more abundant in the MS^2 spectrum of acropyranol B comparing to acropyranol A (5% in **6** vs 48% **7**). The former ion was further fragmented giving rise to the $[A_1-H-C_4H_8O]^-$ fragment ion observed at m/z 247.0974 in the MS^3 spectrum. Regarding the $[B_1-H]^-$ ion, it underwent further fragmentation giving rise to the key fragment ions at m/z 302.1519, 259.0972 and 247.0973 observed in the MS^3 spectra of both acropyranol compounds. In addition, the cleavage of a C_4H_8O moiety from $[A_0-H]^-$ ion was observed in the MS^3 spectrum resulting in the generation of the relatively most abundant $[A_0-H-C_4H_8O]^-$ fragment ion at m/z 179.0351. Finally, only in the case of acropyranol A the $[B_0-H]^-$ ion yielded the fragment ions

Chapter 2: LC-MS based structural elucidation and dereplication of AtA

at m/z 234.0896, 191.0351 and 179.0351 corresponding to the cleavage of CH_3 , C_4H_{10} and C_5H_{10} moieties as observed in the MS^3 spectrum (Figure 24, Figure 25).

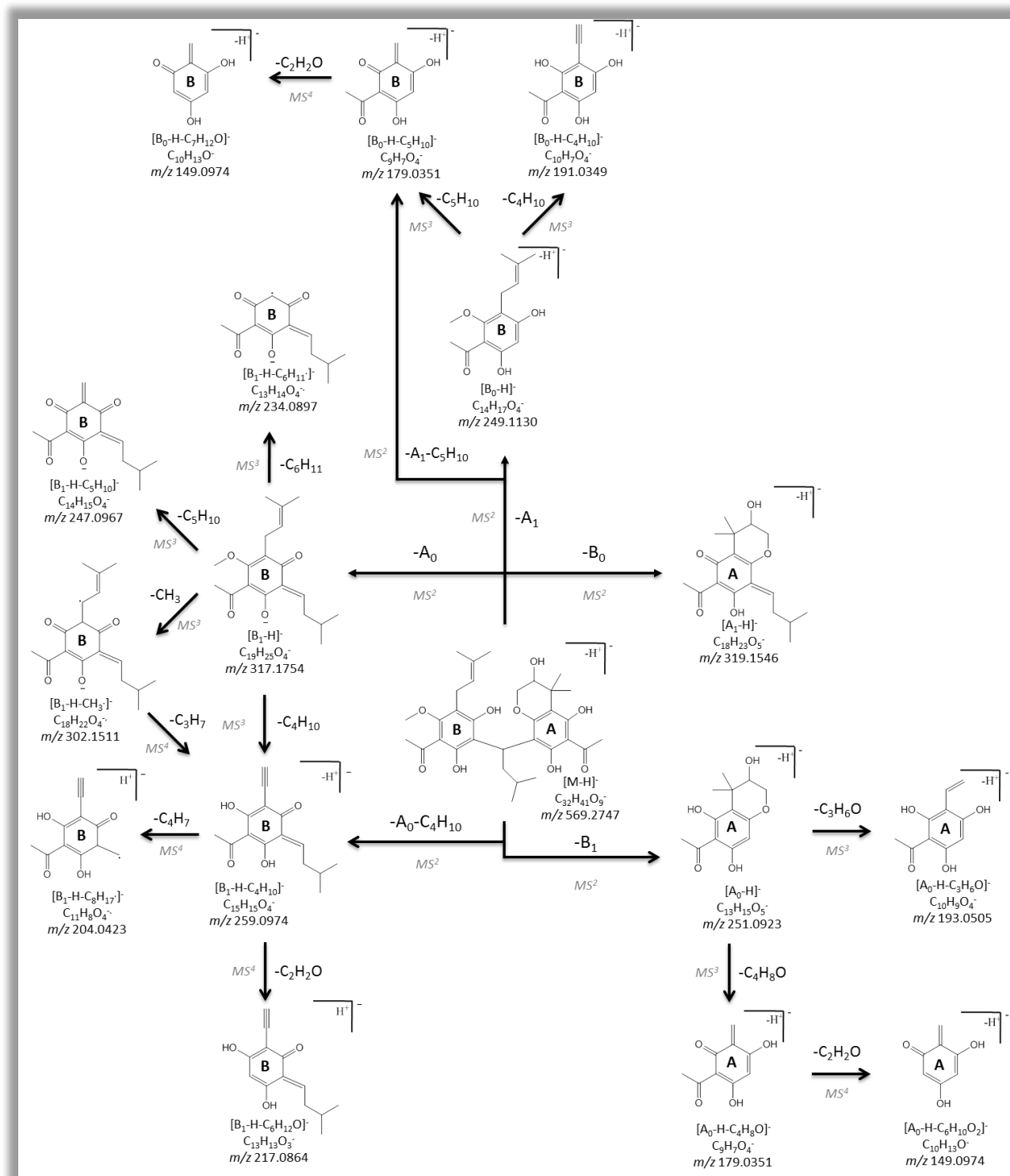


Figure 24: Proposed fragmentation pathways for acropyranol A (6) in ESI (-)

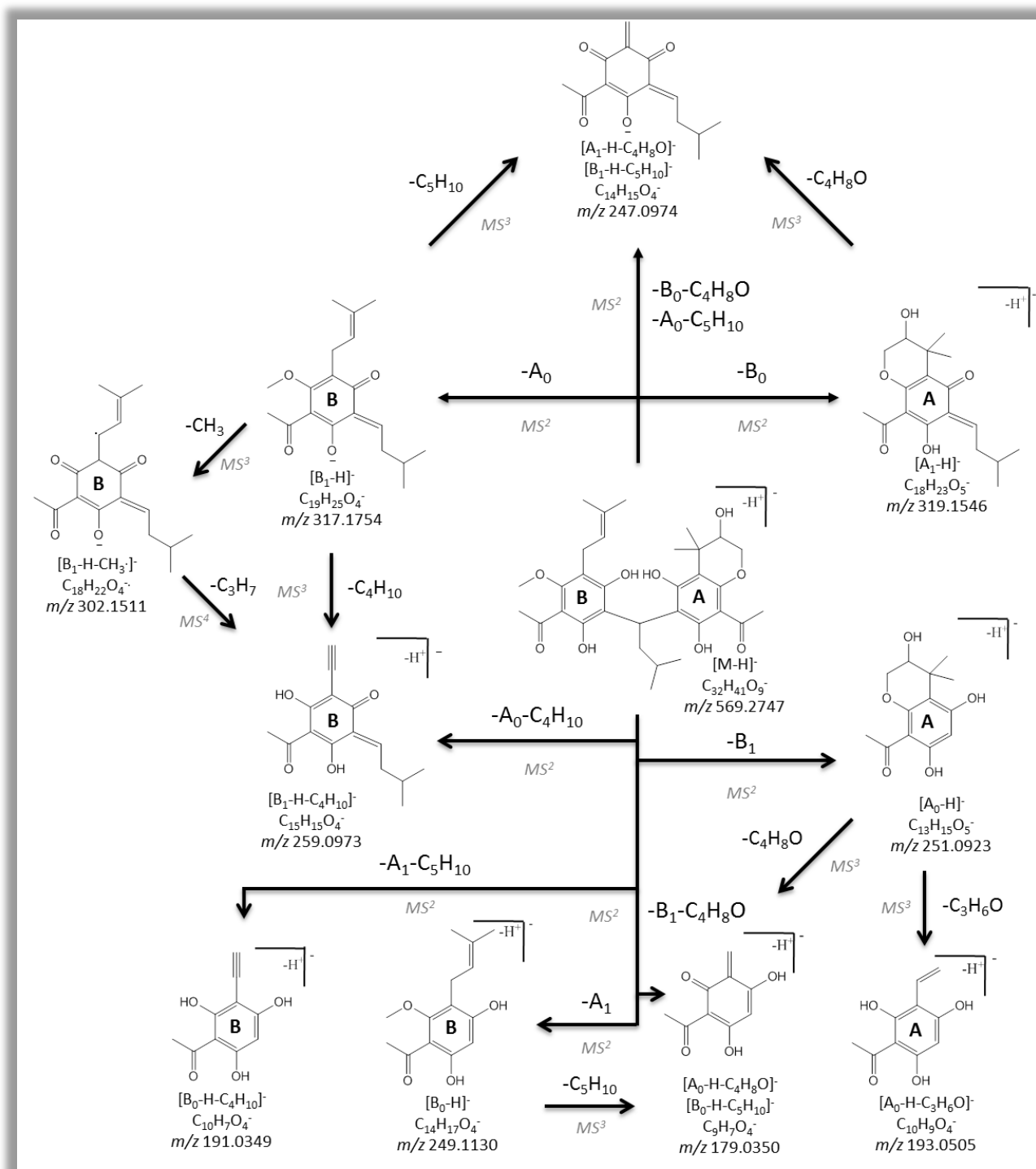


Figure 25: Proposed fragmentation pathways for acropyranol B (7) in ESI (-)

Chapter 2: LC-MS based structural elucidation and dereplication of AtA

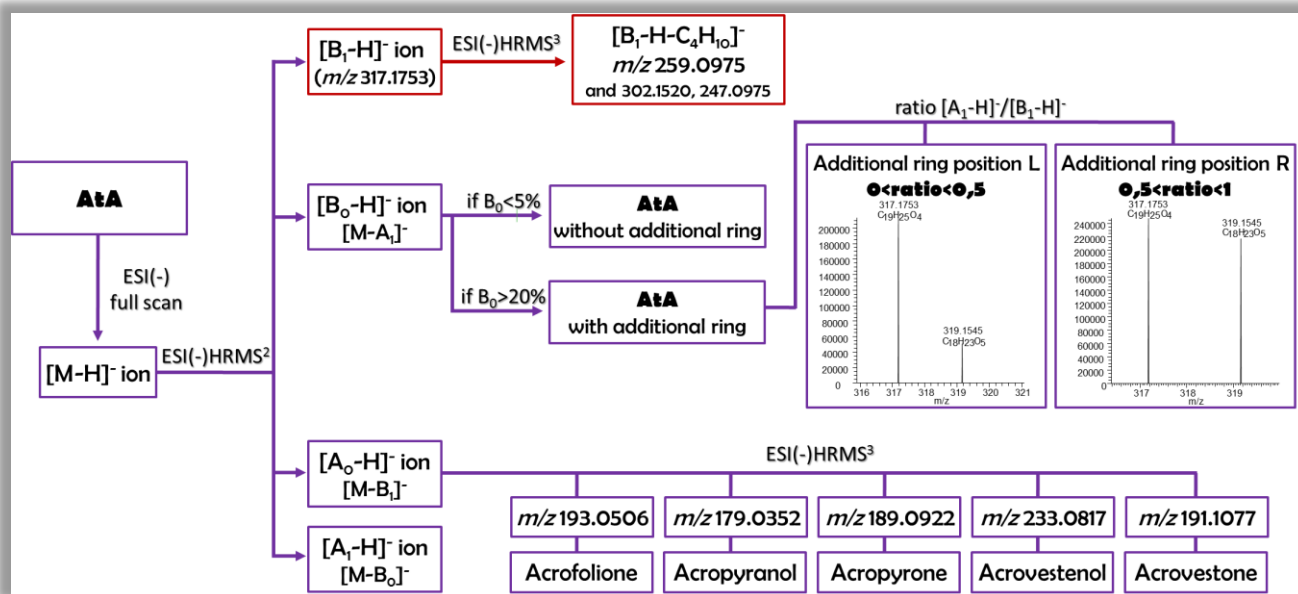
Table IX: HRMS and HRMSn ions of reference compounds (1-7) in ESI(-); Precursor ions are in bold and highlighted

Compound	Pseudomolecular ions [M-H] ⁺ m/z (% base peak/delta ppm)	Fragment ions (MS ²) m/z (% base peak/ delta ppm)	Fragment ions (MS ³) m/z (% base peak)	Fragment ions (MS ⁴) m/z (% base peak)
Acrovestone (1)	553.2781 (35/ -1.48)	317.1743 (100/ -0.44) , 303.1588 (75/ -0.27) , 249.1123 (6/ 0.13), 235.0976 (99/ 0.24) , 191.1072 (5/ 0.57)	302.1511 (61), 259.0967 (100) , 247.0967 (33), 234.0890 (13); 261.1487 (42), 259.1694 (100), 191.0345 (28); 191.1072 (100)	243.0655 (56), 217.0864 (98), 204.0423 (100); 149.0972 (84), 123.0453 (100)
Acrovestenol (2)	569.2748 (29/ -1.44)	319.1546 (42/ -1.68) , 317.1754 (18/ -1.5) , 251.0923 (100/ -0.94) , 233.0817 (12/ -0.85)	301.1442 (100) , 249.1130 (18) , 247.0974(55) ; 302.1520 (50) , 259.0974 (100) , 247.0974 (34) , 234.0896 (11); 233.0817 (100) , 209.0818(13), 179.0351 (28) ; 191.0714 (30) , 189.0921(100) , 187.0765 (23), 147.0818 (13)	259.1338 (67), 257.1545 (100), 255.1389 (26); 207.1027 (100), 205.1235 (91); 203.1078 (77); 259.0974 (100); 244.0738 (20), 243.0661 (53), 217.0869 (100), 204.0428 (82); 205.0870 (100), 203.1076 (38); 191.0714 (28), 189.0921 (100), 187.0765 (22); 137.0247 (53); 147.0818 (100); 187.0765 (38), 147.0818 (100)
Acropyrone (3)	551.2646 (19/ -0.88)	317.1753 (54/ -1.31) , 301.1441 (54/ -1.09), 249.1130 (33/ -0.89) , 233.0817 (100/ -0.78)	302.1519 (27), 259.0974 (100), 247.0974 (10); 234.0896 (43), 191.0353 (83), 179.0351 (75); 191.0715 (26), 189.0922 (100), 187.0765 (20)	
Acrofolione A (4)	569.2747 (67/ -1.61)	319.1545 (20/ -1.86) , 317.1753 (86/ -1.8) , 259.0972 (5/ -1.52), 251.0922 (52/ -1.4) , 249.1129 (100/ -1.31) , 193.0505 (5/ -0.52), 191.0349 (12/ -0.43), 179.0350 (6/ -0.19)	261.1130 (100); 302.1520 (45) , 259.0974 (100), 247.0975 (31), 234.0896 (7); 193.0506 (100) , 179.0351 (8); 234.0897 (76) , 191.0351 (90), 179.0351 (100)	259.0974 (100); 151.0403 (18), 149.0611 (76); 191.0351 (100); 137.0247 (100)
Acrofolione B (5)	569.2746 (54/ -1.44)	319.1545 (87/ -1.8) , 317.1753 (100/ -1.58) , 261.1129 (5/ -1.22) , 259.0972 (6/ -1.13) , 251.0922 (27/ -1.22) , 249.1129 (29/ -1.08) , 193.0506 (5/ -0.23), 191.0349 (5/ -0.13)	261.1131 (100) , 247.0975 (7); 302.1522 (49) , 259.0975 (100) , 247.0976 (34) , 234.0897(10); 219.1026 (86), 217.1233 (100); 217.0870(73), 204.0428 (41); 193.0507 (100) , 179.0351 (7); 234.0896 (84) , 191.0350 (94), 179.0351 (100)	245.0819 (12), 219.1026 (100), 217.1234 (96); 259.0976 (100); 243.0662 (28), 217.0870 (48), 204.0429 (45); 205.0871 (100); 151.0403 (41), 149.0611 (100); 191.0350 (100), 179.0351 (16); 137.0248 (100)
Acropyranol A (6)	569.2750 (21/ -1.06)	319.1547 (5/ -1.38), 317.1754 (100/ -1.34) , 259.0974 (5/ -0.73), 251.0923 (42/ -0.79) , 249.1130 (22/ -0.74) , 179.0351 (5/ 0.48)	302.1521 (49) , 259.0975 (100) , 247.0975 (34), 234.0896 (10), 191.0351 (5); 193.0507 (23), 179.0352 (100); 191.0351 (48), 179.0351 (70), 234.0896 (61)	259.0975 (100); 217.0870 (38), 204.0429 (31)
Acropyranol B (7)	569.2748 (41/ -1.41)	319.1546 (48/ -1.68) , 317.1754 (100/ -1.52) , 259.0973 (5/ -1.17), 251.0923 (21/ -0.98) , 249.1130 (35/ -1.05) , 247.0974 (5/ -0.89), 191.0350 (5/ -0.1), 179.0350 (5/ -0.17)	247.0974 (100); 302.1522 (45) , 259.0975 (100), 247.0976 (24); 179.0352 (100); 179.0351 (36)	259.0974 (100)

1.2.1. Key features for identification of AtA in ESI(-)

During the study of AtA under ESI(-) characteristic fragment ions were generated that could be used for the detection and structural characterization of these compounds (Figure 26). Specifically:

- ~ All compounds generate pseudomolecular ions $[M-H]^-$ ions when ionized under these conditions.
- ~ In all MS levels both A and B aromatic rings remain intact.
- ~ $[B_1-H]^-$ (A_0 loss), $[A_1-H]^-$ (B_0 loss), and $[A_0-H]^-$ (B_1 loss) fragment ions are observed in all reference compounds and could be used as diagnostic ions for AtA identification. $[B_1-H-C_4H_{10}]^-$ (m/z 259.0972), $[B_1-CH_3]^-$ (m/z 302.1519) and $[B_1-H-C_5H_{10}]^-$ (m/z 247.0973) ions observed in the MS^3 spectra of $[B_1-H]^-$ of all AtA compounds could be also utilized for their selective identification.
- ~ The presence of the $[B_0-H]^-$ ion with relative abundance $\geq 20\%$ is characteristic for AtA with additional ring (**3-7**). The $[B_0-H]^-$ ion undergoes further fragmentation generating $[B_0-H-C_5H_{10}]^-$ ion at m/z 179.0351 in the MS^3 spectrum and could be also utilized.
- ~ The ratio $[A_1-H]^-/[B_1-H]^-$ of relative intensities could be used for determination of the fusion position in AtA isomers with additional ring. Specifically, in AtA type-L, the ratio $[A_1-H]^-/[B_1-H]^-$ is calculated < 0.5 while in AtA type-R, the same ratio is < 0.5 .
- ~ The identification of the different AtA substitution patterns in ring A can be performed by examining the product ions of the $[A_0-H]^-$ ion in the MS^3 spectra.



Full scan	MS ²	MS ³	MS ⁴	Identification	Compounds
[M-H] ⁻					
		[B ₁ -H-CH ₃] ⁻			
	[B ₁ -H] ⁻	[B ₁ -H-C ₄ H ₁₀] ⁻		AtA	1-7
		[B ₁ -H-C ₅ H ₁₀] ⁻			
	[A ₁ -H] ⁻				
	[A ₀ -H] ⁻				
	[B ₀ -H] ⁻	[B ₀ -H-C ₅ H ₁₀] ⁻		Additional ring at A	3-7
	[A ₁ -H] ⁻ /[B ₁ -H] < 0.5			Type-L	4,6
	[A ₁ -H] ⁻ /[B ₁ -H] > 0.5			Type-R	5,7

Figure 26: Decision tree and table for characterization of AtA in ESI (-)

1.3. Comparison of ESI and APCI ionization of AtA

As already mentioned, both ESI and APCI sources were assessed for the analysis of AtA compounds. Despite the fact that APCI is not extensively discussed in this manuscript, some interesting observations were made that is worth to briefly mention. Both ESI and APCI resulted to charged analytes in full scan but a different ionization behavior of AtA was evidenced. In specific, in APCI extensive in-source fragmentation of the pseudomolecular ions was observed. Especially in positive mode all full scan spectra were dominated by the

$[B_1+H]^+$ ion (m/z 319.1904) (Figure A 71) while the pseudomolecular ion was present in very low intensity or was completely absent. This fact opposed the generation of ion trees and the further investigation of AtA in APCI(+). On the other hand, in APCI(-) the fragmentation was less extensive although it was significantly higher comparing to ESI(-). Nonetheless, it enabled the generation of MS^n trees for all compounds. This difference in ionization behavior of AtA between the two sources could be correlated with their relative high lipophylicity (Gabelica et al. 2005). Regarding the fragmentation motif of AtA in higher levels using APCI(-) was found similar to ESI(-) in all reference compounds.

2. UHPLC-ESI(-)-HRMSⁿ analysis of *Acronychia* extracts

As thoroughly discussed in the previous sections, ion tree-based methodologies were proven significantly useful for the characterization of AtA, in both ESI negative and positive mode. Certain fragmentation patterns and decision trees could be used for the accurate and selective identification of AtA compounds. In order to validate this approach but also to generate a strategy for dereplication of AtA in complex mixtures a UHPLC-ESI(-)-HRMSⁿ methodology was developed and applied. For the generation of ion trees a data dependent method utilizing the dynamic exclusion (DE) principle was incorporated. DE is widely applied in dereplication studies and is quite useful when highly background ions are present such as in plant extracts (Zhang et al. 2009, Andrews et al. 2011).

Specifically, three extracts deriving from the trunk bark of *Acronychia pedunculata* were analyzed and screened for the presence of the reference compounds while emphasis was given on the detection of unknown derivatives. For the selection of optimal ionization mode a standard mixture of the seven reference compounds was analyzed in both modes and sensitivity aspects were considered (Figure A 72). Despite the fact that both modes displayed high sensitivity and could afford key fragment ions, ESI(-) was finally selected since it provided better signal to noise ratio and thus minor compounds could be detected more easily.

In order to mine AtA derivatives in the extracts, BP and extracted ion chromatograms (XICs) in full scan, MS^2 and MS^3 levels for key ions were generated. Specifically, in MS^2 level the $[B_1-H]^-$ ion at $m/z = 317.1753$ and in MS^3 the $[B_1-H-C_4H_{10}]^-$ ($m/z = 259,0975$) were selected according to the AtA identification decision tree (Figure 26). The profiling of the

extracts using the XICs, retention time (Rt), relative intensities of fragment ions, RDBeq. values, Δm values and suggested ECs for deprotonated molecules lead to the detection of 35 AtA with the majority being potential new derivatives (Figure 27, Table X). Furthermore it is important to note that the accuracy of the method was proven quite high as it is illustrated from the Δm values ($0.27 \leq \Delta m \leq 1.43$ ppm). These data are included in Table 4 and hereafter, the identification process as well as the characterization of representative new derivatives will be briefly discussed.

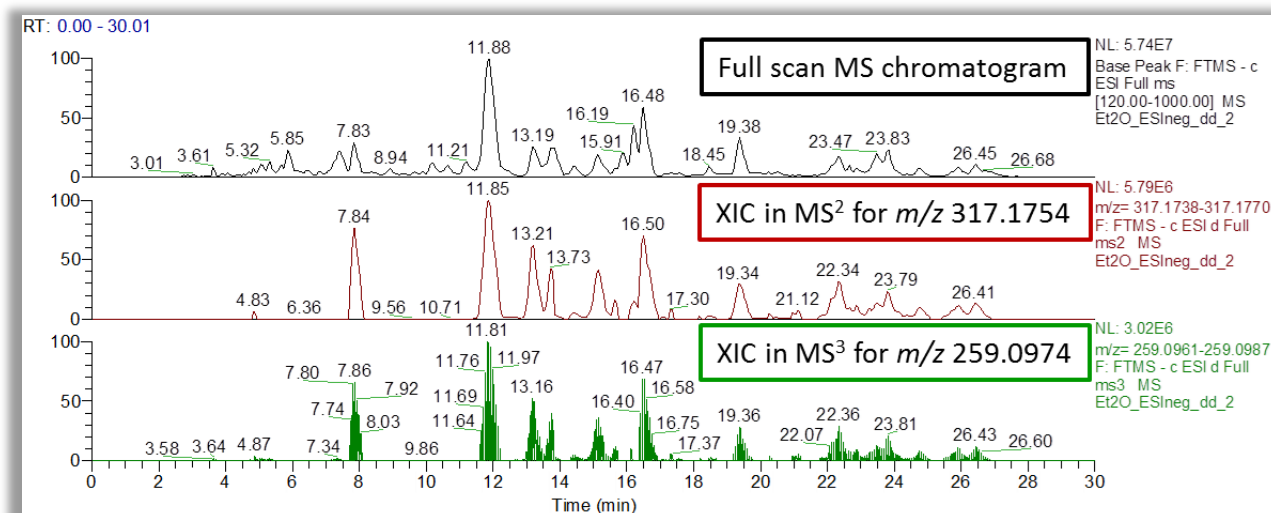


Figure 27: Peak base chromatogram and extracted ion chromatograms (XIC) of Et₂O (A) extract

Starting from the more straightforward structures, the peak detected at 17.07 min (*peak 12*) presented a $[M-H]^-$ ion at m/z 555.2961 with an elemental composition of $C_{32}H_{43}O_8^-$ and 11.5 RDBeq. value. Based on the ESI(-) decision tree, all the ions suggesting AtA structure were present while the $[B_0-H]^-$ ion was not detected indicating the absence of an additional ring (figure 8). Based on the direct comparison with acrovestone (**1**), *peak 12* could be easily attributed to dihydroacrovestone. Likewise, the two peaks observed at 16.79 min (*peak 11*) and 14.72 (*peak 28*) presented 1 degree of unsaturation less and the same fragmentation pattern with acrovestenol (**2**); thus both could be identified as dihydroacrovestenol isomers. Moreover, a compound eluted at 14.15 min (*peak 10*) presented a $[M-H]^-$ ion at m/z 569.2753 with an elemental composition of $C_{32}H_{41}O_9^-$ and 12.5 RDBeq. value while no $[B_0-H]^-$ ion was detected. Presenting the same structural and spectrometric features with **2**, could be considered as acrovestenol isomer. The different relative intensities of the key ions $[A_0-H-H_2O]^-$ and $[A_0-H-C_2H_2O]^-$ in MS³ level verify this suggestion.

Additionally, *peak 9* resolved in 13.59 min presented an EC of $C_{32}H_{43}O_{10}^-$ as revealed from the $[M-H]^-$ ion observed at m/z 587.2856. This peak meets all the criteria to be an AtA with an additional side chain bearing two OH groups. Moreover, it presents one degree less of unsaturation (11.5 vs 12.5). This motif suggests a hydroxyl dihydroacrovestenol derivative.

The analysis of the MS^2 spectrum of *peak 33* ($C_{32}H_{39}O_8^-$, m/z 551.2649) reveals all the key fragment ions for AtA while based on EC and RDBeq. values corresponds to an acropyrone isomer. The $[B_0-H]^-$ fragment ion with relative intensity of 32% and a $[A_1-H]^-/[B_1-H]^-$ ratio of 0.05 indicating the presence of an AtA type-L derivative. Interestingly, in case of *peak 33* two additional fragment ions were observed in the MS^2 spectrum at m/z 235.0974 (BP) corresponding to the $[A_0-H+2H]^-$ fragment ion and at m/z 303.1597 attributed to the $[A_1-H+2H]^-$ ion. The occurrence of these ions in *peak 33* comparing to acropyrone (**3**) implies the instability of the additional ring with left orientation.

Peak 8 detected at 4.75 min providing a deprotonated molecular ion at m/z 585.2695 and an EC of $C_{32}H_{41}O_{10}^-$. Considering the EC of **2**, **4-7** ($C_{32}H_{41}O_9^-$) the presence of an additional OH group could be suggested. Moreover, the presence of the $[B_0-H]^-$ ion (54%) indicates additional ring while the ratio of $[A_1-H]^-/[B_1-H]^- = 0.4$ leads to the assumption that it regards an hydroxylated AtA type-L. Similarly, *peak 29* with an elemental composition of $C_{37}H_{49}O_9^-$ (m/z 637.3376) and presenting a $[B_0-H]^-$ fragment ion with 53% relative intensity and a $[A_1-H]^-/[B_1-H]^-$ ratio of 0.53 is deduced as an AtA with an additional ring type-R. Following the same rational, *peak 27* and *peak 35* revealed the same EC of $C_{33}H_{43}O_9^-$ (m/z 583.2908) and the same degree of unsaturation (12.5). With direct comparison to **2**, **4-7** ($C_{32}H_{41}O_9^-$) and the biosynthetic probabilities could be proposed that both regard methoxylated derivatives of hydroxylated AtA. The $[B_0-H]^-$ ion were present in both cases and the $[A_1-H]^-/[B_1-H]^-$ was calculated < 0.5 suggesting additional ring and AtA L-type. However, the low abundances of $[B_0-H]^-$ (17 & 8%, respectively) raise some doubts for the assignment or signify the influence of methyl group on the fragmentation motifs. Finally, *peak 14* ($R_t=15.42$) comprise an interesting case since an uncommon EC was suggested ($C_{29}H_{33}O_8^-$) while the rest of the features as well as the MS^n data are in accordance to an AtA structure. Specifically, the high intensity of $[B_0-H]^-$ ions indicates the presence of an additional ring and the $[A_1-H]^-/[B_1-H]^- = 0.92$ an AtA type-R. A dihyrdofuran ring could be a reasonable assumption taking under consideration the biosynthetic motifs in *Acronychia*.

Apart from the abovementioned compounds which comprise relative structures to the reference AtA, several other peaks satisfying the identification criteria but presenting different structural features were detected. These peaks could be divided in three different groups regarding their ECs and more specifically, AtA-C₃₇, AtA-C₄₂ and AtA-C₄₇. By straight comparison of their structural features with the reference AtA it could be easily presumed that they differ in the number of isoprenyl units (IP). Thus, group AtA-C₃₇ (*peaks 17, 32, 34, 29, 13, 16, 30 and 31*) includes all derivatives with 1 additional IP, AtA-C₄₂ (*peaks 15, 18, 19, 21, and 26*) with 2 IP and AtA-C₄₇ (*peaks 20, 22, 23, 24, and 25*) with 3 IP, on the basic AtA structure. Based on the chemical and biosynthetic possibilities could be attributed to geranyl-, farnesyl- or geranylgeranyl derivatives according to the substitution patterns and the different combinations of these units. Furthermore, the number of O atoms could be utilized for the detection of hydroxylated polyprenylated AtA (e.g. **13**) and the number of H atoms the dihydropolyprenylated AtA (e.g. **16**). It is worth noting that the elution order generally follows the trend C₃₂→C₃₇ →C₄₂ →C₄₇ verifying the occurrence of additional lipophylic units in higher Rt. Regarding their spectrometric behavior they present many similarities with AtA since most of the characteristic ions are also detected (Table 4). Nevertheless, additional ions, cleavages and patterns are observed especially regarding the geranylgeranyl derivatives. Given the fact that no reference polyprenylated AtA were available any assumption could be dangerous and it was not attempted in this manuscript.

Concerning the distribution of compounds, all reference AtA were unambiguously traced in the Et₂O(A) extract as expected and in the other two extracts after alkalization (Et₂O and CH₂Cl₂). The only exception concerns **7** which was not traced elsewhere. Generally, more similarities were observed between the Et₂O and CH₂Cl₂ extracts (Table X).

Table X: Identification of AtA compounds in three different *A. pedunculata* extracts

Peak No.	Rt (min)	[M-H] ⁻	EC	Δm (ppm)	RDBeq.	MS ²						Characterization	Extract
						[A0-H]-	[A1-H]-	[B0-H]-	[B1-H]-	Other	[A1-H] ⁻ / [B1-H] ⁻		
1	16.22	553.2802	C32H41O8	- 0.979	12.5	235.0974 (79)	303.1598 (60)	n.d.	317.1754 (91)		0.6	Acrovestone	Et2O(A), Et2O, CH2Cl2
2	15.93	569.2752	C32H41O9	- 0.696	12.5	251.0921 (100)	319.1543 (48)	n.d.	317.1752 (21)	233.0816 (12)	2.2	Acrovestenol	Et2O(A), Et2O, CH2Cl2
3	19.11	551.2648	C32H39O8	- 0.492	13.5	233.0817 (100)	301.1440 (41)	249.1129 (35)	317.1752 (64)		0.6	Acropyrone	Et2O(A), Et2O, CH2Cl2
4	11.63	569.2755	C32H41O9	- 0.274	12.5	251.0923 (47)	319.1546 (19)	249.1132 (78)	317.1755 (84)	191.0349 (12)	0.2	Acrofolione A	Et2O(A), Et2O, CH2Cl2
5	13.48	569.2754	C32H41O9	- 0.379	12.5	251.0921 (25)	319.1544 (94)	249.1129 (25)	317.1752 (100)		0.9	Acrofolione B	Et2O(A), Et2O, CH2Cl2
6	12.96	569.2752	C32H41O9	- 0.696	12.5	251.0921 (37)	n.d.	249.1129 (20)	317.1752 (100)		-	Acropyranol A	Et2O(A), Et2O, CH2Cl2
7	14.87	569.2752	C32H41O9	-0.59	12.5	251.0921 (23)	319.1544 (52)	249.1128 (29)	317.1752 (100)		0.5	Acropyranol B	Et2O(A)
8	4.75	585.2700	C32H41O10	- 0.855	12.5	267.0869 (100)	335.1492 (25)	249.1128 (54)	317.1751 (55)		0.4	OH type-L	Et2O (A)
9	13.59	587.2856	C32H43O10	- 0.904	11.5	269.1025 (100)	337.1748 (38)	n.d.	317.1751 (5)		7.6	OH-dihydroacrovestenol	Et2O(A)
10	14.15	569.2753	C32H41O9	- 0.485	12.5	251.0920 (100)	319.1542 (64)	n.d.	317.1751 (29)		2.2	Acrovestenol isomer	Et2O(A), Et2O
11	16.79	571.2911	C32H43O9	- 0.308	11.5	253.1077 (100)	321.1701 (39)	n.d.	317.1751 (21)		1.8	Dihydroacrovestenol isomer	Et2O(A)
12	17.07	555.2961	C32H43O8	- 0.363	11.5	237.1130 (89)	305.1752 (65)	n.d.	317.1752 (100)		0.6	Dihydroacrovestone	Et2O, Et2O(A)
13	18.08	637.3378	C37H49O9	-0.59	13.5	319.1543 (100)	387.2167 (33)	n.d.	317.1751 (9)		3.6	+IP	Et2O (A)
14	15.42	509.2179	C29H33O8	- 0.317	13.5	191.0349 (27)	259.0972 (92)	249.1129 (48)	317.1752 (100)		0.9	dihydrofuran type-R	Et2O(A)
15	19.97	689.4053	C42H57O8	- 0.902	14.5	371.2217 (100)	439.2841 (17)	249.1127 (45)	317.1750 (76)	259.0971 (13)	0.2	+2IP	Et2O(A), Et2O, CH2Cl2
16	20.68	619.3273	C37H47O8	-0.6	14.5	301.1434 (59)	n.d.	249.1124 (22)	317.1746 (100)	259.0967 (11)	-	+IP	Et2O (A)
17	20.83	621.3433	C37H49O8	- 0.051	13.5	303.1591 (51)	n.d.	249.1125 (11)	317.1747 (100)	259.0968 (8)	-	+IP	Et2O(A), Et2O
18	21.83	689.4051	C42H57O8	- 1.163	14.5	371.2218 (100)	439.2842 (20)	249.1128 (17)	317.1750 (87)	259.0971 (14)	0.2	+2IP	Et2O(A), Et2O, CH2Cl2

Chapter 2: LC-MS based structural elucidation and dereplication of AtA

19	22.06	689.4052	C42H57O8	- 1.076	14.5	371.2218 (100)	439.2837 (9)	249.1127 (12)	317.1750 (97)	259.0970 (14)	0.1	+2IP	Et2O(A), Et2O, CH2Cl2
20	22.29	757.4677	C47H65O8	- 1.112	15.5	439.2842 (100)	507.3469 (8)	249.1127 (26)	317.1750 (68)	259.0970 (10)	0.1	+3IP	Et2O(A), Et2O, CH2Cl2
21	22.53	689.4050	C42H57O8	- 1.337	14.5	371.2213 (100)	439.2840 (36)	249.1126 (17)	317.1749 (75)	259.0970 (13)	0.5	+2IP	Et2O(A), Et2O, CH2Cl2
22	23.11	757.4674	C47H65O8	- 1.429	15.5	439.2839 (100)	507.3466 (12)	249.1126 (24)	317.1748 (40)	259.0968 (5)	0.3	+3IP	Et2O(A), Et2O, CH2Cl2
23	23.4	757.4677	C47H65O8	- 1.112	15.5	439.2842 (100)	507.3469 (7)	249.1127 (16)	317.1750 (55)	259.0970 (8)	0.1	+3IP	Et2O(A), Et2O, CH2Cl2
24	24.38	757.4678	C47H65O8	- 0.874	15.5	439.2840 (100)	n.d.	249.1126 (10)	317.1749 (47)	259.0970 (6)	-	+3IP	Et2O(A), Et2O, CH2Cl2
25	25.55	757.4677	C47H65O8	- 1.033	15.5	439.2841 (100)	507.3468 (9)	249.1126 (8)	317.1750 (61)	259.0970 (7)	0.1	+3IP	Et2O(A), Et2O, CH2Cl2
26	26.14	689.4052	C42H57O8	- 0.989	14.5	371.2219 (54)	439.2843 (100)	249.1128 (25)	317.1751 (87)	259.0971 (14)	1.5	+2IP	Et2O(A), Et2O, CH2Cl2
27	14.5	583.2909	C33H43O9	- 0.628	12.5	265.1076 (18)	n.d	249.1127 (17)	317.1750 (100)	233.0816 (12)	-	OH-,CH3-type-L	Et2O, CH2Cl2
28	14.72	571.2910	C32H43O9	- 0.413	11.5	253.1078 (100)	321.1701 (51)	n.d.	317.1752 (34)		1.5	dihydroacrovstenol isomer	Et ₂ O
29	15.11	637.3376	C37H49O9	- 0.873	13.5	319.1543 (100)	387.2167 (39)	249.1128 (53)	317.1752 (75)		0.5	+IP	Et2O, CH2Cl2
30	18.32	619.3271	C37H47O8	- 0.794	14.5	301.1437 (45)	n.d	249.1127 (21)	317.1750 (100)	259.0970 (8)	-	+IP	Et ₂ O
31	18.55	635.3586	C38H51O8	- 0.569	13.5	317.1750 (100)	n.d.	249.1126 (1)	317.1750 (100)	593.3469 (17)	-	+IP	Et ₂ O
32	19.38	621.3428	C37H49O8	- 0.743	13.5	303.1598 (100)	371.2222 (57)	249.1130 (2)	317.1753 (45)	259.0968 (8)	1.2	+IP	Et ₂ O
33	13.27	551.2649	C32H39O8	- 0.275	13.5	233.0817 (12)	n.d.	249.1130 (32)	317.1753 (40)	235.0974 (100), 303.1597 (36)	-	L-type Acropyrone	CH ₂ Cl ₂
34	14.83	637.3377	C37H49O9	- 0.779	13.5	319.1542 (100)	387.2166 (24)	249.1127 (44)	317.1750 (77)	259.0970 (10), 303.1595 (14)	0.3	+IP	CH ₂ Cl ₂
35	16.65	583.2907	C33H43O9	- 0.936	12.5	265.1076 (4)	333.1700 (9)	249.1127 (8)	317.1750 (29)	551.2636 (100)	0.3	2 × OCH3	CH ₂ Cl ₂

3. Reproducibility of MSⁿ spectra

LTO-Orbitrap analyzer equipped with ESI or APCI source has been proven a great tool for the structure elucidation and characterization of small molecules. During this study, a multistage HRMS platform for the dereplication of known AtA and the identification of new AtA compounds was proposed. In order to assess the validity of this approach, the reproducibility of the generated MSⁿ spectra under different conditions was investigated by the comparison of the relative intensities of the main fragment ions.

In particular, the collision energy was considered a main parameter which could potentially influence the MSⁿ spectra. In this context, MSⁿ spectra were acquired at varying collision energies in both ionization modes resulted in identical fragmentation patterns. Independently of the collision energies applied, the relative intensity of the fragment ions observed in different MSⁿ spectra did not alter significantly, proving the reproducibility of the MSⁿ spectra using this instrumentation (Figure 28A). Another parameter that was assessed was the reproducibility of the MSⁿ spectra from offline and online sample introduction. Specifically, the MSⁿ spectra obtained from infusion of reference compounds, injection of standard mix solution and extract solution were compared indicating significant difference in the relative intensity of the fragment ions (Figure 28B/C). Moreover, it is important to note that the concentration levels of the reference compounds in the various solutions were different and in plant extract solutions possible coelution phenomena were occurred without introducing any variation in fragmentation patterns behavior. On the top of that, the reproducibility of the MSⁿ spectra was also observed in experiments acquired with 1 year of interval (Figure 28D/E). These results, in line with Hooft et al. support the great reproducibility of the Orbitrap analyzer (Hooft et al. 2012) and prove the robustness of this method for the detection and characterization of AtA compounds in complex mixtures such as plant extracts.

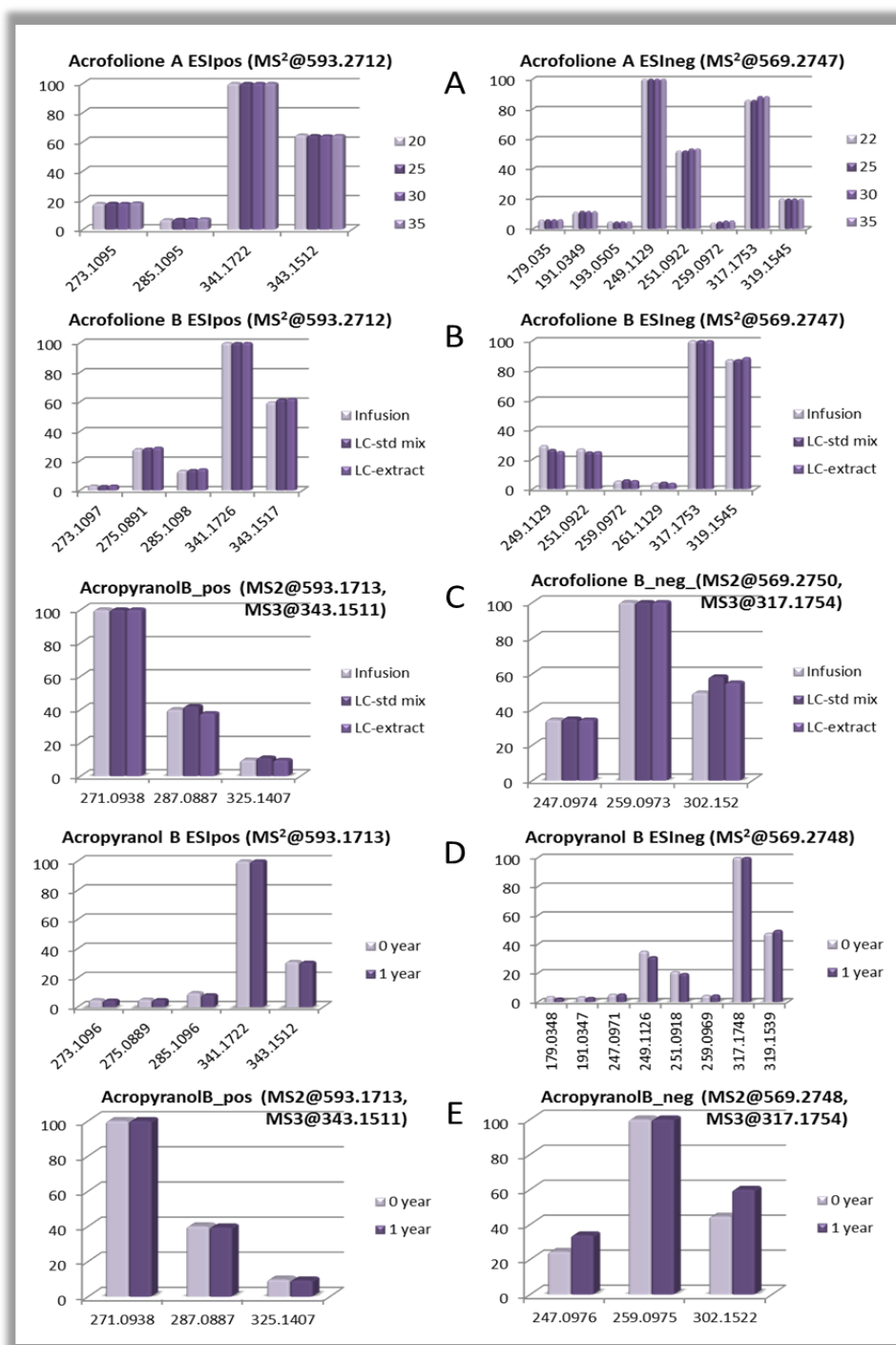


Figure 28: Reproducibility of HRMSn spectra of different AtA compounds; A: MS² spectra of Acrofolione A obtained in positive and negative ionization using different collision energies; B: MS² spectra of Acrofolione B obtained in positive and negative ionization by direct infusion, by LC-MS of the standard mixture and LC-MS of the Et₂O extract; C: MS³ spectra of Acropyranol B and Acrofolione B obtained in positive and negative ionization, respectively, by direct infusion, by LC-MS of the standard mixture and LC-MS of the Et₂O extract; D: MS² spectra of Acropyranol B obtained in positive and negative ionization with one year of interval; E: MS³ spectra of Acropyranol B obtained in positive and negative ionization with one year of interval

Conclusion

Acrorychia-type acetophenones (AtA) are fully substituted phloroglucinol dimers, a group of natural products with great biological interest. Surprisingly, there is no information available for their mass spectrometric behavior and their structure elucidation using MS. In the current study, an ion tree-based methodology was developed for the detailed mass spectrometric investigation of AtA. For all analyses, the hybrid LTQ-Orbitrap analyzer and both ESI(±) and APCI(±) ionization methods were utilized. Both sources were found efficient for the ionization of the compounds while ESI was finally selected offering better sensitivity and more comprehensive MSⁿ data. The high accuracy and resolving power of Orbitrap in full scan acquisitions but also in higher fragmentation levels (MS²-MS⁵) enabled the disclosure of characteristic cleavages and fragmentation patterns leading to the structure elucidation of AtA with high confidence. Moreover, a nomenclature scheme was proposed based on the key fragment ions of AtA. Taking advantage of the collected data an UHPLC-ESI(-)-MSⁿ method was developed for the dereplication of AtA in total extracts. 28 new possible AtA were identified amongst them several polyprenylated acetophenones which have never been referred before leading to better insight to *Acrorychia* composition.

APPENDIX

2

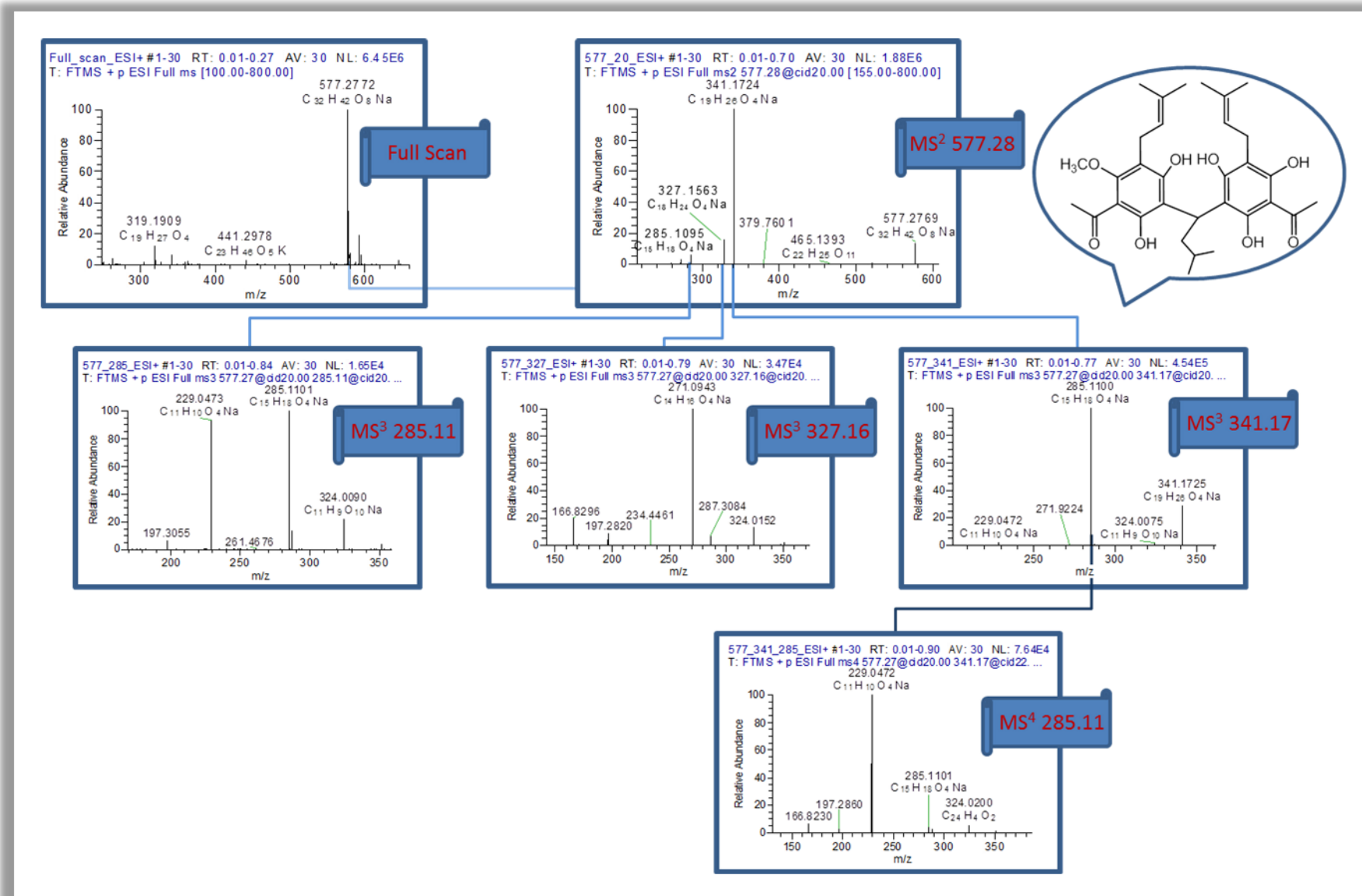
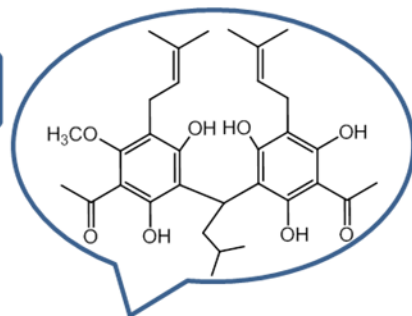


Figure A 61: Spectral Ion Tree of Acrovestone (1) in ESI positive ionization

Figure A 62: Spectral Ion Tree of Acrovestone (1) in ESI negative ionization



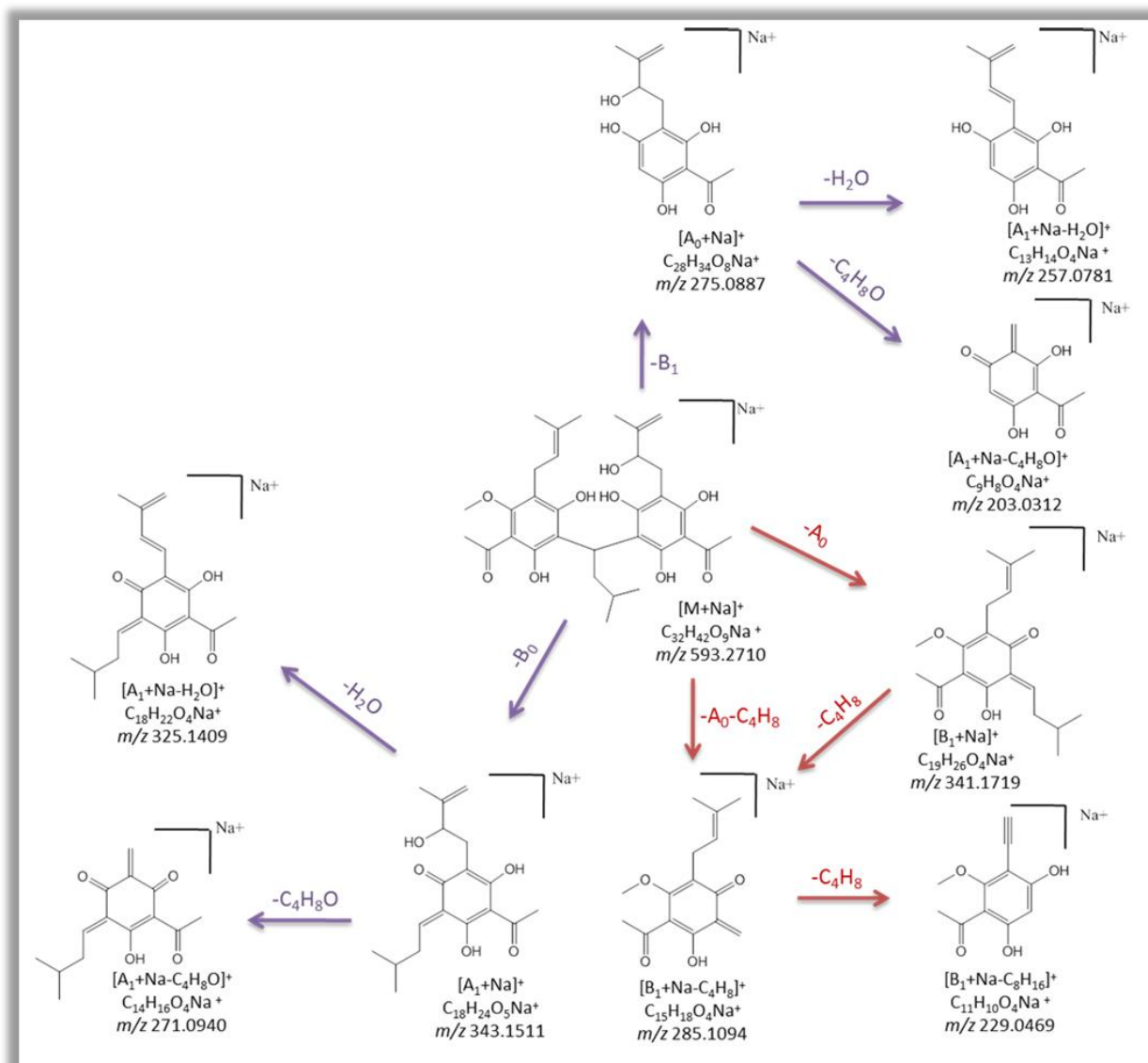


Figure A 63: Proposed fragmentation pathways for acrovestenol (2) in ESI(+)

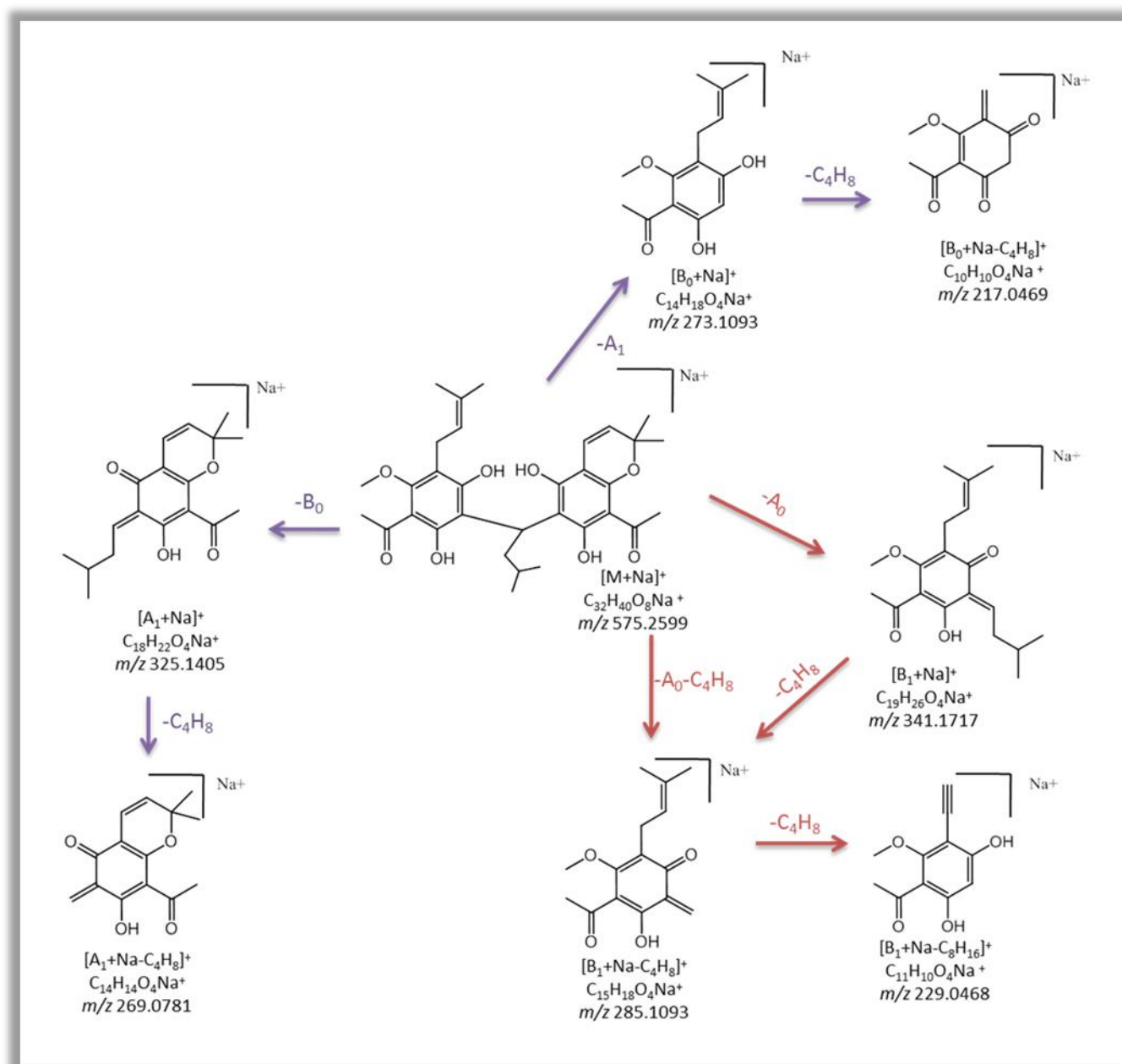


Figure A 64: Proposed fragmentation pathways for acropyrone (3) in ESI(+)

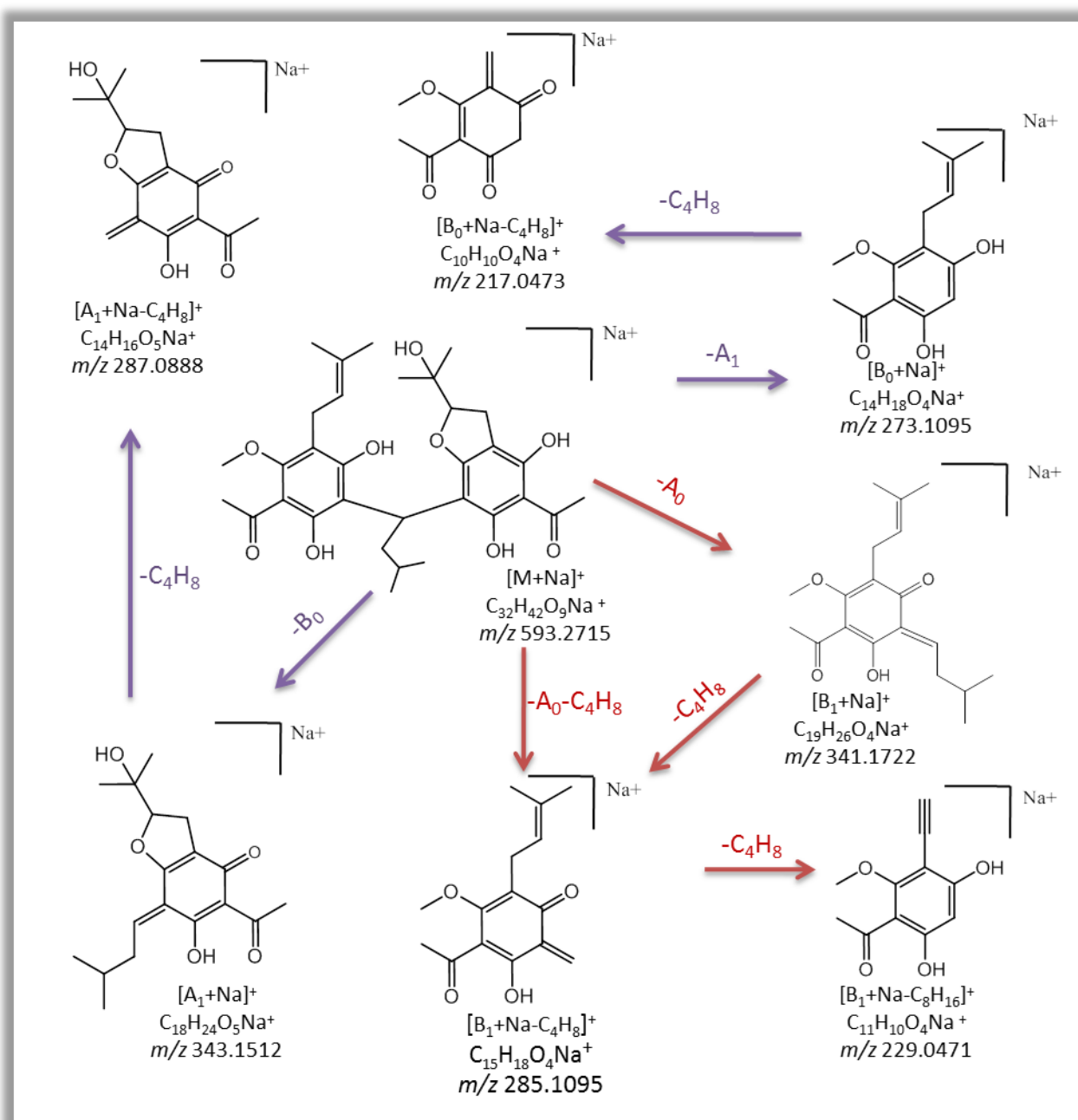


Figure A 65: Proposed fragmentation pathways for acrofolione A (4) in ESI(+)

Chapter 2: LC-MS based structural elucidation and dereplication of AtA

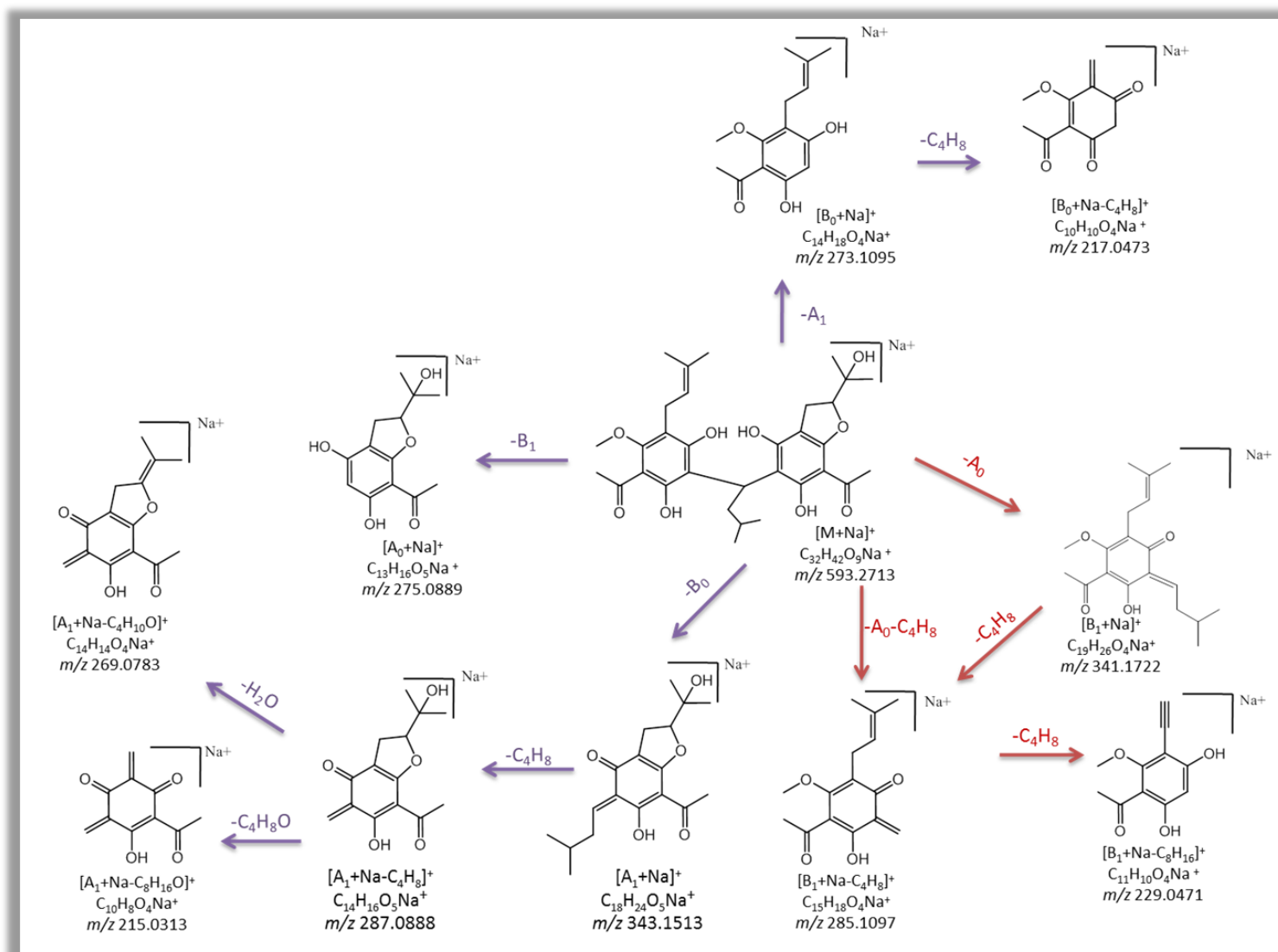
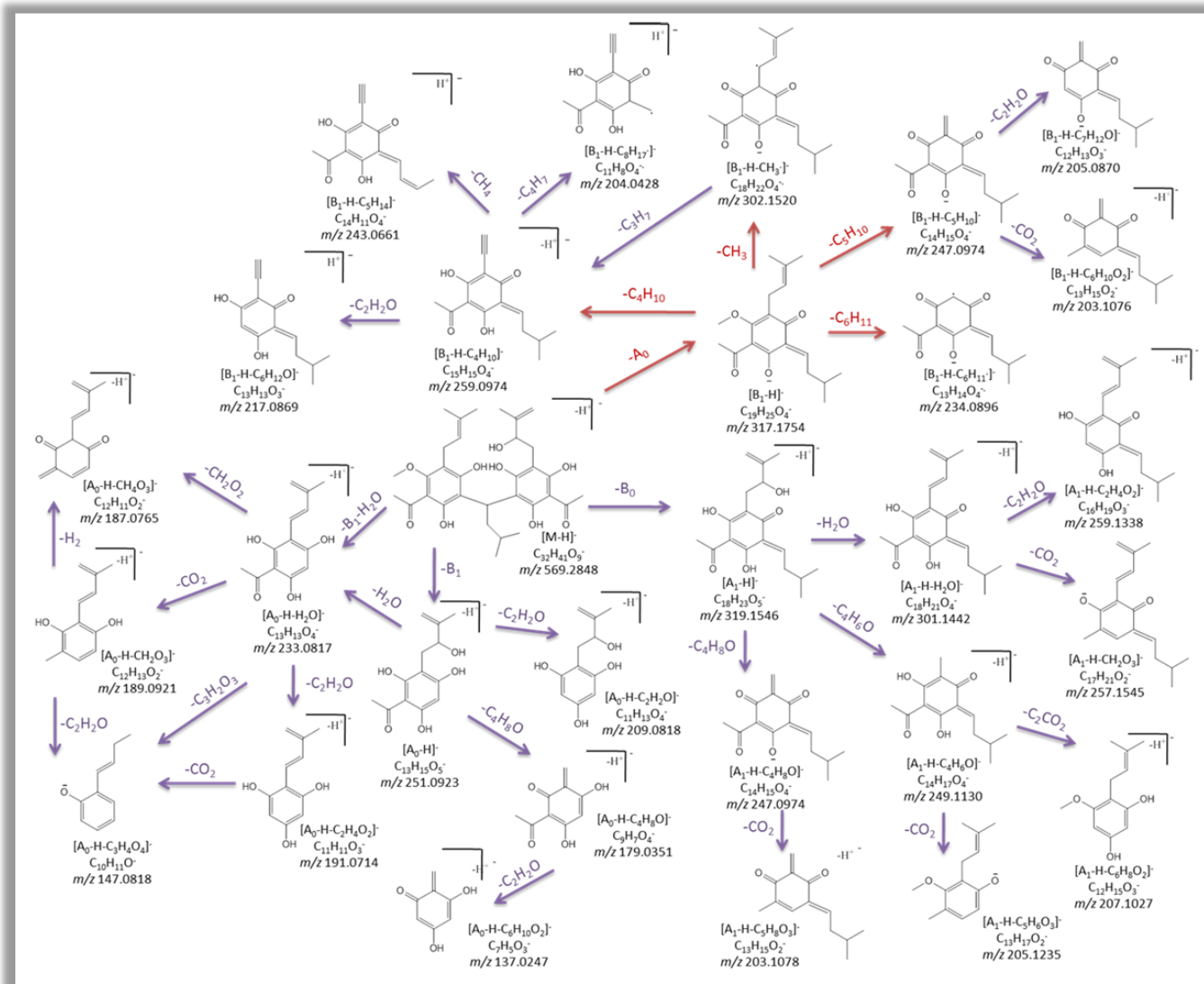


Figure A 66: Proposed fragmentation pathways for acrofolione B (5) in ESI(+)



Chapter 2: LC-MS based structural elucidation and dereplication of AtA

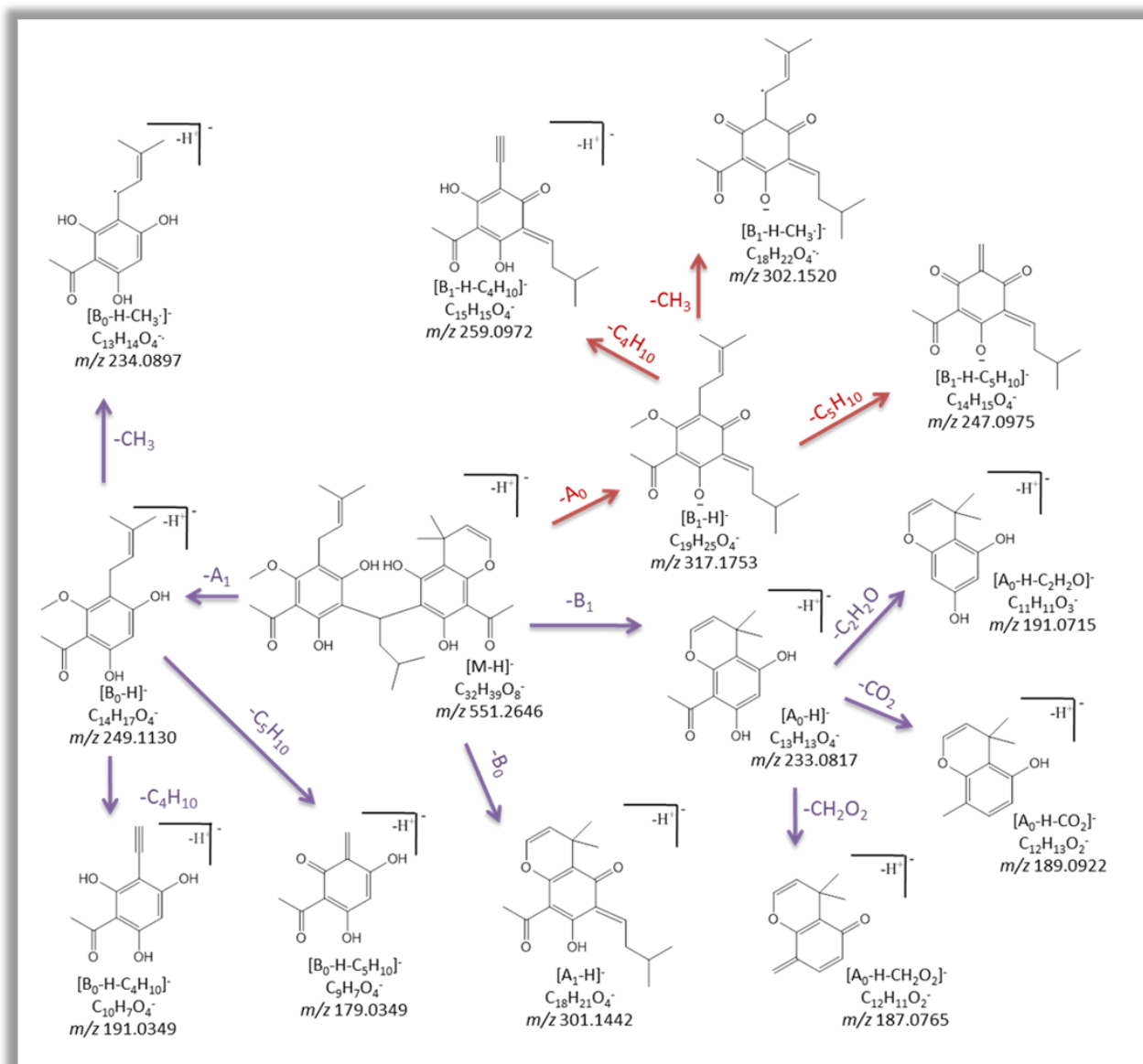


Figure A 68: Proposed fragmentation pathways for acropyronone (3) in ESI(-)

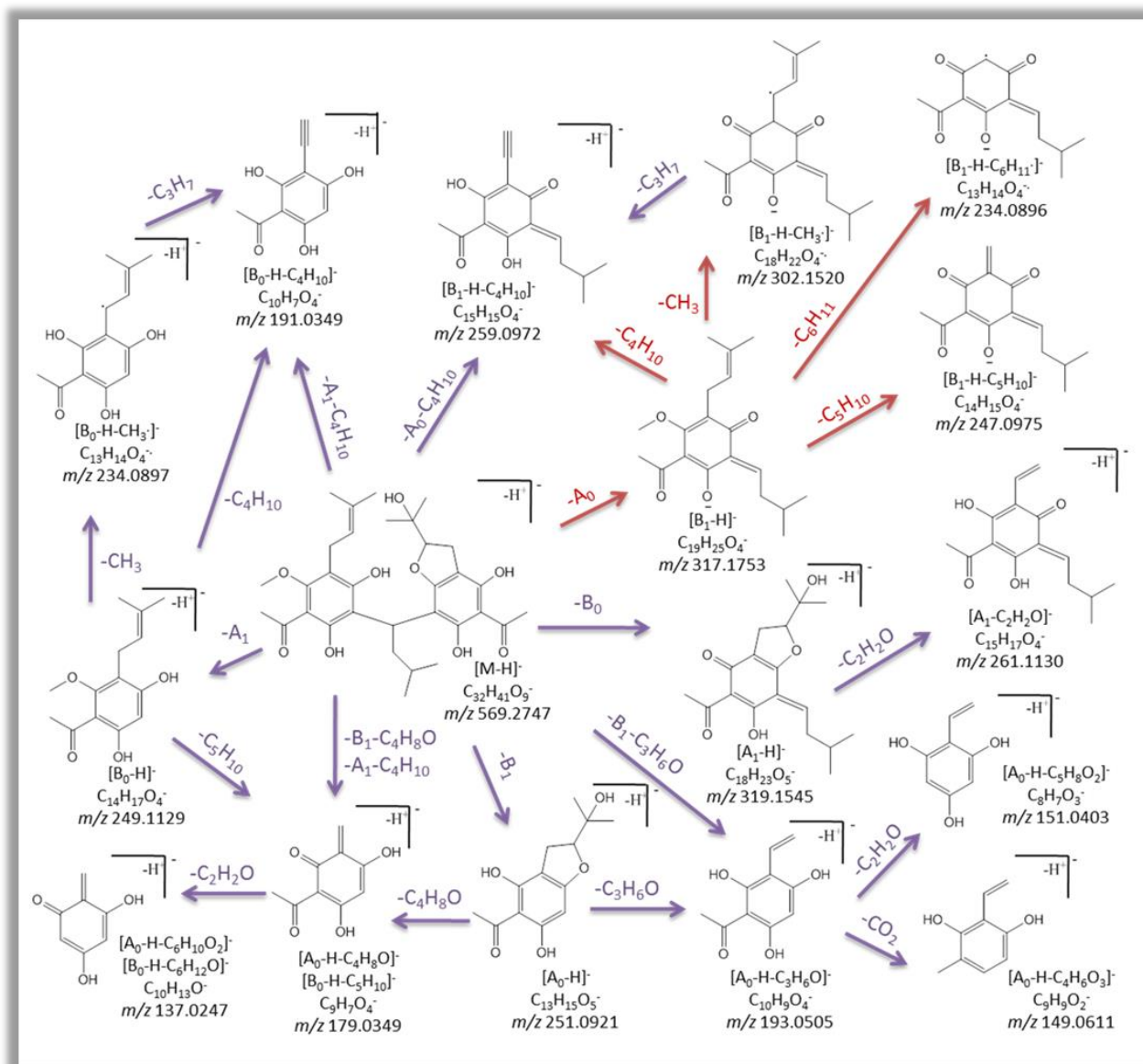


Figure A 69: Proposed fragmentation pathways for acrofolione A (4) in ESI(-)

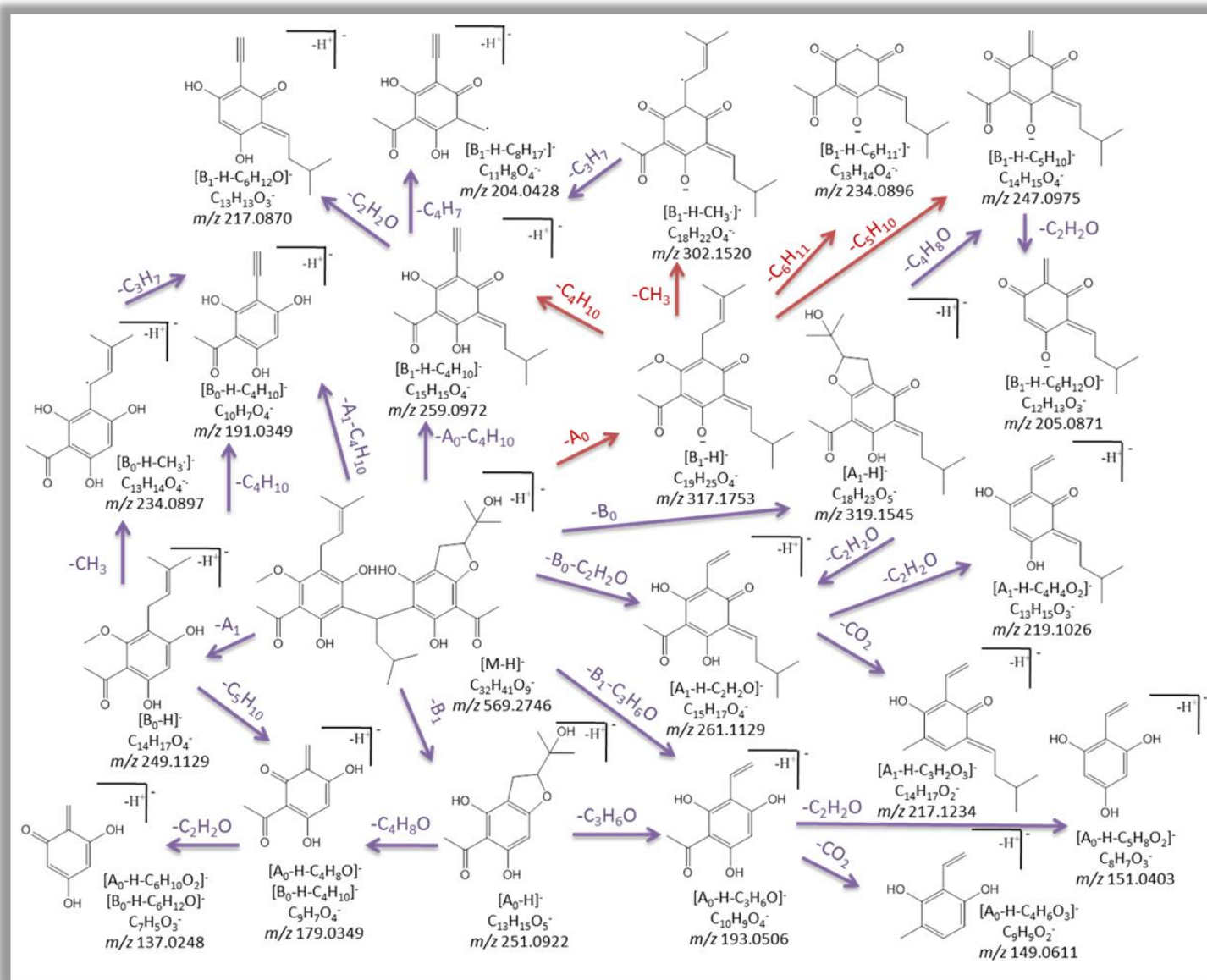
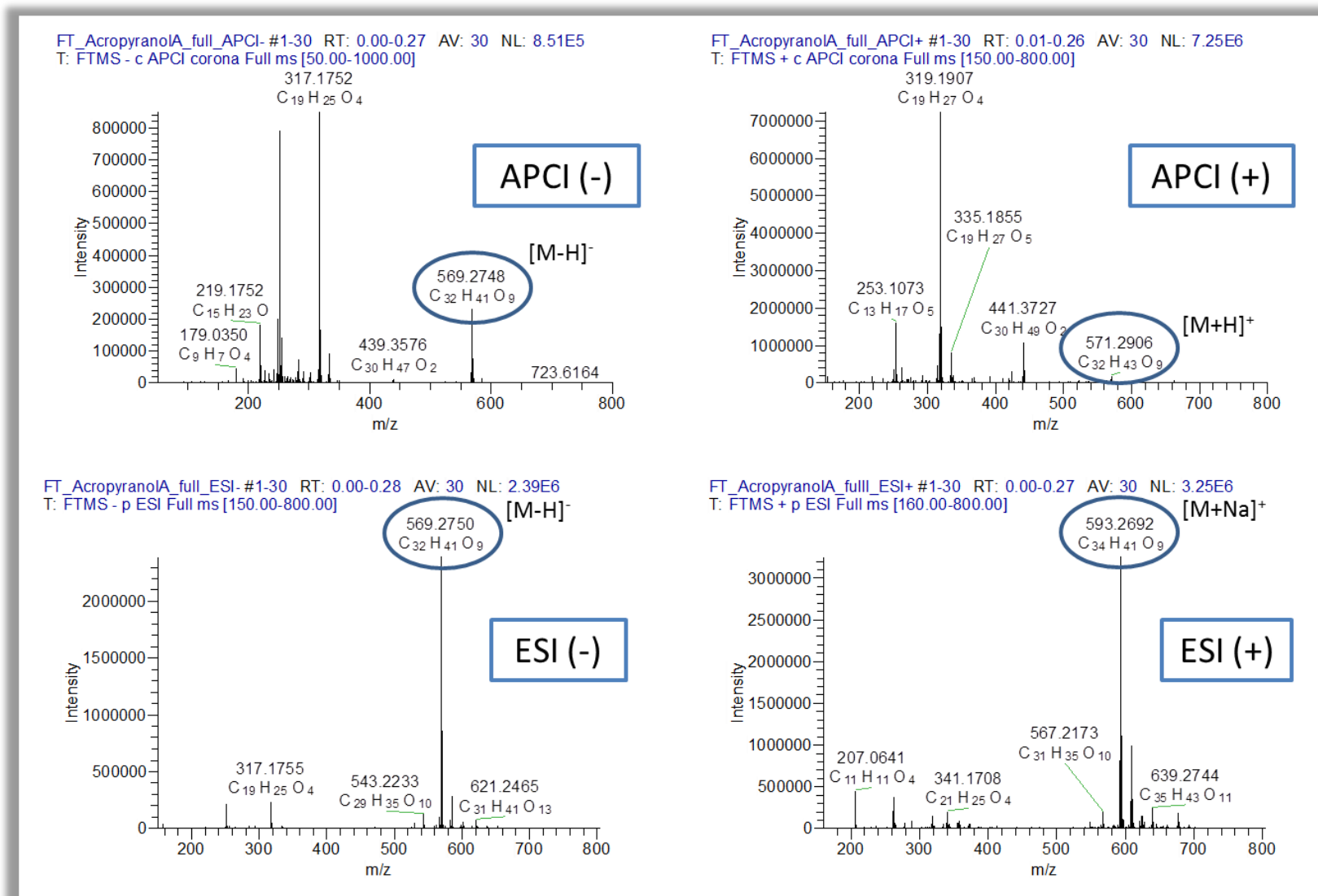


Figure A 70: Proposed fragmentation pathways for acrofolione B (5) in ESI(-)

Figure A 71: Full scan mass spectra of acropyranol A (5) in APCI ($\pm$) and ESI ($\pm$); molecular and adduct ions are highlighted

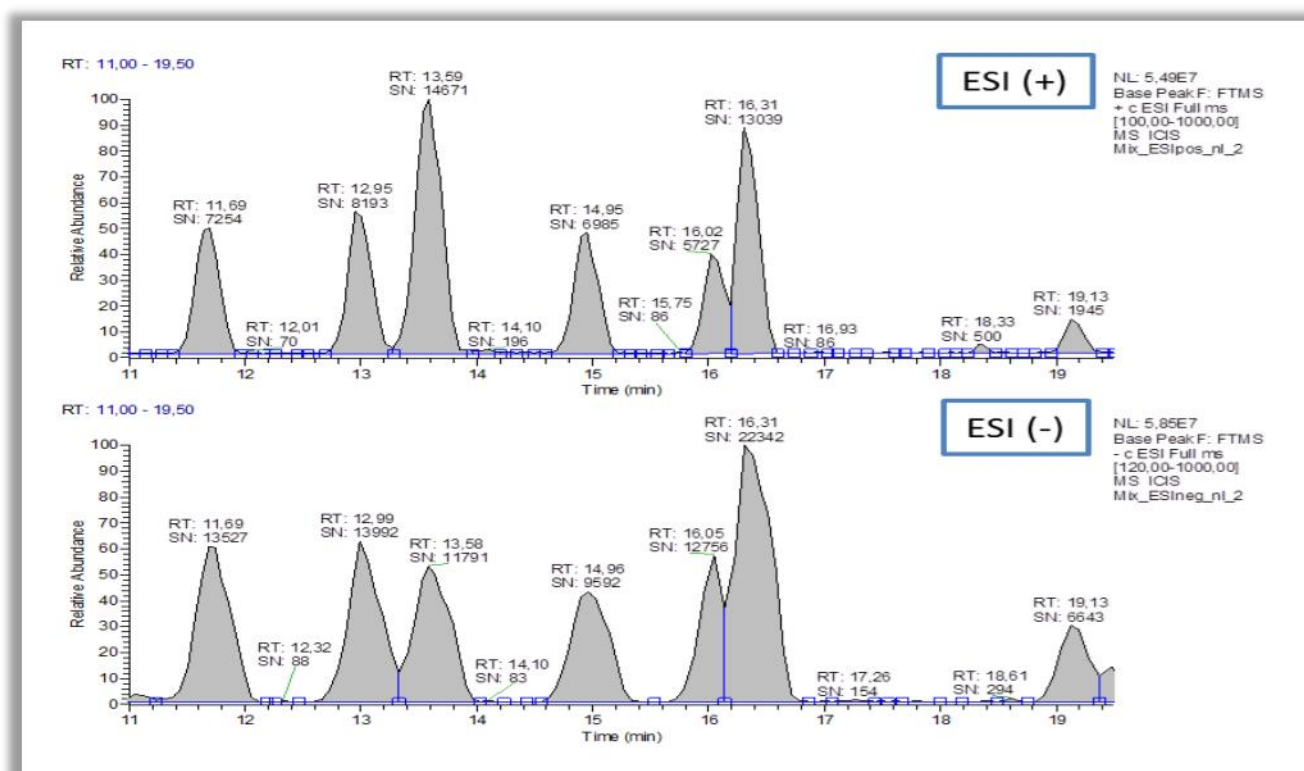


Figure A 72: Base peak chromatograms of the reference mixture acquired in ESI ($\pm$); Rt and signal to noise ratio assigned

CHAPTER

Metabolomics approaches in
Acronychia genus

Introduction

1. Background

Medicinal plants have been successfully used worldwide for centuries to treat various illnesses and diseases in different herbal preparation forms (Fabricant et al. 2001). Back to the ancient Greek times, the properties of several herbal preparations were first reported from physicians such as Hippocrates and Dioscurides. In addition, well established holistic approaches [e.g Traditional Chinese Medicine (TCM), Ayurveda] have a long tradition and are still used widely (Dias et al. 2012). In recent years, many countries in Africa, Asia and Latin America use traditional medicine as first line healthcare while the use and popularity of traditional medicine in Western world is always increasing according to WHO (Briskin 2000, Bagozzi 2003). The indications and usage of these herbal preparations are handed down orally from generation to generation consequently the scientific background on these practices is poor or absent. Nowadays, there have been targeted efforts towards the exploitation of the information derived from these traditional medicine systems. This mainly involves the identification and isolation of bioactive natural products and the exploration of their mechanisms of action which still remain unknown for the majority of them. Therefore, the drug discovery from medicinal plants constitutes an emerging field with the attempt to discover new bioactive natural products using state of art techniques (Tyler 1999, Balunas et al. 2005).

Natural products possess a predominant position as therapeutic agents throughout the years (Newman et al. 2007, Newman et al. 2012). Natural products are synthesized in nature to facilitate various physiological functions constituting a structurally diverse and complex array of compounds (Wink 2003). These unique structural features such as chiral centres, aromatic rings, complex ring systems, degree of molecule saturation, and number and ratio of heteroatoms provide high impact at the drug discovery effort (Yuliana et al. 2013). In addition, natural products have been proved to cover a larger 'chemical space' comparing to compounds designed from combinatorial chemistry that is associated to a widespread biological relevance (Dobson 2004, Larsson et al. 2007). Thus, natural products

excel comparing to synthetic chemicals from the perspective of drug lead finding in terms of molecular diversity and biological functionality (Nisbet et al. 1997). Therefore, natural products can serve not only as potential drugs but also as an initial point for the optimization of the structure in order to achieve more efficient and/or less toxic derivatives. As a matter of fact, naturally derived compounds have been the inspiration for a great number of approved drugs and drug candidates (Newman et al. 2009).

Despite the evidence that medicinal plants comprise a prolific source of lead compounds, the drug discovery process from plant material faces many challenges. Traditional drug discovery approaches usually lead to the repeated isolation of the same compounds while false positive results are often observed due to the presence of multiple bioactive compounds (synergic interactions) or the presence of highly active compounds in really small quantities (Gilbert et al. 2003). Nevertheless, the introduction of modern technological advances in the natural products drug discovery contributed significantly to circumvent these challenges. Towards this effort, the implementation of state of the art analytical techniques played a crucial role. Therefore, miniaturised and automated methodologies have been proposed providing a high impact in the reproducibility and the effectiveness of the discovery approaches of new bioactive compounds (Michel et al. 2013, Potterat et al. 2013). Despite these remarkable achievements, all these approaches are pointing in the analysis of plant extracts with targeted methods and under specific purposes disregarding the complex nature of these matrices.

2. Metabolomics

More recently, metabolomics approaches taking advantage of all these technological advances and developments have been implemented for the analysis of plant materials in a more holistic way. Metabolomics as a methodology is combining the robustness and the sensitivity of state of the art analytical techniques with the statistical power of bioinformatics and a sample through put allowing ideally the identification and quantification of every individual metabolite (Issaq et al. 2009). Metabolomics has been proven an important tool to explore biological systems providing a holistic view of the metabolites under certain biological conditions (Hall 2006). Therefore, metabolomics applied in phytochemistry field are aiming at the comprehensive and large-scale analysis of plant metabolites and thus,

provide a broad view of the metabolite composition of a given organism (Sumner et al. 2003).

Metabolomics is a reflection of genetic factors and metabolites are often regarded as the functional endpoint of a biological system (Sumner et al. 2003). Therefore, the comprehensive analysis of the metabolome of a given organism is the key point for the understanding of the biochemical status of an organism closely related to its phenotype (Fiehn 2002). The metabolome of plant organisms is consisted of primary and secondary metabolites. The primary metabolites are involved in basic functions of plants and are generally common in all species while secondary metabolites or natural products are species specific (Verpoorte et al. 2007). This is why the amount of secondary metabolites is estimated to attain the number of 200.000 metabolites (Goodacre et al. 2004). Therefore, the determination of the entire metabolome in a plant organism is a really challenging task.

A wide range of analytical platforms have been applied in metabolomic studies with the respect to analyse the complete array of metabolites including gas chromatography mass spectrometry (GC-MS) (Lisec et al. 2006), capillary electrophoresis mass spectrometry (CE-MS) (Levandi et al. 2008), fourier transform infrared spectroscopy (FT-IR) (Khairudin et al. 2014) while the most widely used are liquid chromatography hyphenated with mass spectrometry (LC-MS) and nuclear magnetic resonance spectrometry (NMR) (Weckwerth et al. 2005, Allwood et al. 2008). Each technique provides different advantages and disadvantages in terms of sensitivity, reproducibility and interpretability which have been exhaustively discussed (Sumner et al. 2003, Dunn et al. 2005, Lenz et al. 2006, Hagel et al. 2008, Wolfender et al. 2013). However, no single analytical method is sufficient to accurately survey the entire metabolome. The wide chemical diversity and range of concentration of all metabolites present in a plant material, implies the combination of multiple analytical platforms in parallel to get a better insight into the chemical composition of plant extracts (Bino et al. 2004, Moco et al. 2007). Towards this direction, recent chemometrics and bioinformatics advances promise to enhance the global understanding of plant metabolome.

In metabolomics field the analysis of multiple samples with diverse analytical techniques leads to the generation of an enormous amount of data which needs to be handled in a homogeneous way. The assistance of bioinformatics in this point is indispensable as the resulting metabolite profiles generated from multiple analytical platforms have to be translated into meaningful information (Trygg et al. 2006, Bartel et al.

2013). Therefore, independently of the analytical platform that is applied the establishment of a robust data handling workflow is necessary for the interpretation of the large amount of data (van den Berg et al. 2006). Typically, the transformation of the raw data into data matrices is performed initially in order to obtain adaptable structures for the data analysis step. These matrices are analysed applying mainly different multivariate methods in order to unravel the significant features related to the biological information. According to the study and consequently to the biological questions to be answered, a number of supervised and unsupervised multivariate methods for exploration, visualization, classification and prediction of the data are employed leading to the identification of biomarkers (van der Greef et al. 2005, Issaq et al. 2009).

Biomarkers are generated from any analytical approach as distinct features; therefore, the identification of the metabolites represented from the respective features is indispensable to obtain the biological information. For this purpose the generation of free available libraries would be of great value. Nevertheless, the great number and the diversity of the secondary metabolites hinder the achievement of a universal library of secondary metabolites. Consequently, this step remains an obstacle in the overall high throughput strategy as it is based mainly on manual handling (Moco et al. 2007). However, the identification of biomarkers as the ultimate goal of metabolomics procedure is very important as the results are translated into meaningful information.

During the last years, metabolomics have been spread out in many areas related to plant research. The unique advantages inheriting from the combination of state of the art analytical techniques with chemometrics methods using simple and throughput sample preparation steps have been exploited for diverse applications. According to the scope of study, a range of metabolite fingerprinting and profiling approaches are applied. Fingerprinting consists of a rapid holistic way for screening of the metabolic composition of an organism with the respect to compare or discriminate it from others whereas profiling is referring additionally to the identification and quantification of the contained metabolites (Hall 2006). Metabolomics, regardless the applied approach, as an untargeted, non-biased and throughput strategy has been employed successfully for chemotaxonomic studies exploring the taxonomic proximity of different plant species, origins or organs in respect to their metabolite profile (Kim et al. 2010, Safer et al. 2011, Kim et al. 2012, Porzel et al. 2014). Another important topic that metabolomics studies have contributed significantly is the quality assessment of plant medicinal products. The analysis of medicinal plant preparations

using metabolomics techniques enables the assessment of the total metabolite composition in contrast to traditional approaches that are focusing on the analysis of specific metabolites (Wang et al. 2004, Gad et al. 2013). Furthermore, the identification of bioactive compounds is an emerging field in plant metabolomics as augmented literature data are dealing with the discovery of bioactive metabolites from plant sources using metabolomics strategies (Yuliana et al. 2011, Inui et al. 2012). Towards this direction, the contribution of metabolomics strategies for drug discovery purposes is of great importance. In addition, other essential applications have been reported to take advantage of the metabolomics platforms such as exploration of metabolite variation during physiological processes (Roldan et al. 2014) and investigation of metabolite variation as a response to stress conditions or interaction with other organisms (Choi et al. 2006, Shulaev et al. 2008). Overall, the integration of metabolomics strategies in natural products chemistry field constitutes an additional powerful tool that may be exploited beneficially.

3. Plant metabolomics workflow

Plant metabolomics as every metabolomics platform is consisted of discrete stages in which special attention should be given in order to obtain reliable results from each step of this pipeline. Briefly, these stages can be summarised as sample preparation, data acquisition, data preprocessing, data pretreatment, data analysis and identification of biomarkers.

3.1. Sample preparation

Sample preparation is considered a crucial step in plant metabolomics analysis as the initial step of a downstream analysis. Since the goal of metabolomics studies is to detect the variances of the samples due to inherent biological difference, the sample preparation has to be performed in an unbiased manner, as homogeneously as possible in order to avoid induced variances from this procedure (Schripsema 2010). The sample preparation for plant metabolomics involves three basic steps: the harvesting of the plant material, the extraction of the plant material and the preparation of the samples for the analysis (Kim et al. 2010).

3.1.1. Harvesting of plant material

In the context of a metabolomics analysis, the selection of the plant materials has to be performed taking into consideration some basic criteria. The most important consideration is to ensure enough biological replicates (Kim et al. 2010). Multiple samples from different sources are a necessary prerequisite to provide enough variation to create statistically significant differences among samples. However this remains an important challenge in plant metabolomics since usually is hard to obtain multiple biological replicates (Inui et al. 2012). Another important consideration is the collection of all required metadata for the characterization of the plant material (Fiehn et al. 2007). These metadata are of valuable importance during data analysis process for classification of prediction purposes as well as the explanation of outlier behaviours.

3.1.2. Extraction procedure

The plant metabolome is comprised of thousands of different metabolites (Bino et al. 2004), thus the extraction of the entire metabolome is practically impossible. Therefore, it is widely recognized that there is no universal or optimal extraction method for plant metabolomics; hence, the chosen extraction protocol can only be regarded as good compromise that fulfils the scope of a given study (Halabalaki et al. 2014).

A number of factors affects the extraction procedure and has to be taken into account before the development of the extraction protocol. In this context, an initial factor to take into consideration is the type of the analytical method that will be used for the metabolomics study. For instance, relatively non-polar solvents are utilized when GC-MS analysis follow in order to extract the volatile and unpolar compounds, whereas more polar solvents are preferred when LC-MS or NMR techniques are applied for the metabolomics analysis (Kim et al. 2011). In addition, in case of NMR-based metabolomics studies the extraction procedure can be simplified by the direct use of deuterated NMR solvents for the extraction. This approach has the main advantage of minimizing the steps of sample preparation which may induce undesirable variations. However, the use of deuterated solvents as extraction solvents narrows the choice of solvents since a small number of deuterated solvents are commercially available (Lubbe et al. 2013). Another drawback that has been mentioned when deuterated solvents are used for the extraction of plant materials is their effect at the resolution of NMR spectra (Kim et al. 2010).

In the most cases the extraction is performed with conventional organic solvents. Undoubtedly, the choice of the extraction solvent is one of the most important aspects that affect the sample preparation. Solvents are usually selected according to their polarity and selectivity, two parameters that define the range of the metabolites that may be extracted. Other factors that have to be considered before the selection of the extraction solvent are the boiling point, the toxicity and the contaminants that may be present in a solvent. As mentioned above, the scope of the study will determine the optimal extraction solvent (Kim et al. 2010, Schripsema 2010, Halabalaki et al. 2014).

Finally, other parameters of the extraction procedure such as the type of extraction, time, temperature, dissolution rate and pH have to be examined for the efficiency of the extraction method. The type of extraction is closely related to the time and the temperature of the extraction. The most frequently reported technique in metabolomics studies is the ultrasonic assisted extraction. Using this technique the diffusion of the solvent in the plant material is faster and the extraction procedure is shorter obtaining more reproducible results comparing with the classical maceration extraction. Other techniques that have been reported are the microwave assisted extraction (MAE), the supercritical fluid extraction (SFE) and the classical maceration extraction (Mushtaq et al. 2014). The dissolution rate and the pH are correlated with the interaction solvent properties and the plant material. The dissolution rate defines the time and the temperature of the extraction as well as the solvent volume that has to be added for the extraction, while the acid or basic nature of each solvent leads to the dissolution of different metabolites (Schripsema 2010).

Typically, during the extraction procedure a number of analytical replicates are prepared in order to control better the reproducibility of each step of the downstream analysis. The minimum used number is three analytical replicates while the increase of these replicates leads to a further degree of reliability of the overall methodology.

3.1.3. Sample preparation for analysis

According to the analytical platform that is intended to be applied for plant metabolomics analysis different concepts of sample preparation has to be followed. The variance in sensitivity and operation principals for diverse analytical techniques implies the use of different magnitudes of sample quantities (μg - mg magnitudes) and specific solvents (e.g LC-MS grade, deuterated solvents etc.). Thus, the sample preparation before the

analysis has to be adjusted according to the selected analytical techniques and to be executed in a homogenous manner in order to avoid induced variation in the dataset.

3.2. Data acquisition

Data acquisition involves the analytical measurement of the metabolite composition using diverse analytical platforms. A wide array of analytical platforms have been applied in plant metabolomics, among them the most popular are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). The principal of each method and the mode of operation of these two techniques is totally different leading to discrete advantages and disadvantages and hence, usually these techniques are used complementarily. Furthermore, the completely diverse instrumentation requisites of these platforms are influencing strongly each stage of the overall downstream analysis.

Nuclear magnetic resonance (NMR) spectroscopy is a valuable tool for plant metabolomics as it provides a rapid, non-destructive, high-throughput method requiring minimal sample preparation for the characterization of the metabolite composition (Hagel et al. 2008). NMR spectroscopy operates by the application of strong magnetic fields and radio frequency (RF) pulses to the nuclei of atoms provoking their promotion to high energy spin states and the detection is performed by the subsequent emission of radiation during the relaxation process (Hagel et al. 2008). NMR is the most adaptable method from the available analytical techniques to extract unambiguous information for the structural elucidation of metabolites contained in complex mixtures (Robinette et al. 2011). The major drawback of NMR is related with its low sensitivity, which is about 10^6 – 10^9 fold less than chromatography-coupled MS (Sumner et al. 2003). Although is less sensitive than MS, NMR offers an unbiased view of metabolite composition affording simultaneously quantitative information about the contained metabolites (Simmler et al. 2014).

Mass spectrometry (MS) constitutes a really valuable technique in plant metabolomics field in terms of speed, sensitivity and accuracy with the potential to identify a wide range of metabolites present in plant extracts (Dettmer et al. 2007). Mass spectrometers operate by ion formation, separation of ions according to their mass-to-charge (m/z) ratio and detection of separated ions (Dunn et al. 2005). Direct infusion mass spectrometry analysis (DIMS) provides a rapid technique for metabolite profiling of plant extracts resulting in the acquisition of a single mass spectrum for each extract. This approach relies on accurate

mass measurements and the generation of the elemental composition of each detected metabolite (Dunn et al. 2005). This technique is unable to detect isobaric metabolites while ion suppression phenomena are taking place leading to the loss of signals of several analytes (Gustavsson et al. 2001). Hyphenation of mass spectrometry with separation techniques such as gas or liquid chromatography leads to the reduction of spectra complexity due to the separation of the metabolites in time dimension. Therefore, ion suppression of the signal is reducing significantly and less complex spectra are collected corresponding to a much smaller part of metabolites accompanied with additional information on the physicochemical properties of the metabolites (Dettmer et al. 2007).

LC-MS techniques are becoming increasingly popular for the analysis of plant extract samples (Allwood et al. 2010). The main advantage of these techniques is the sensitivity which is significantly higher comparing to NMR techniques (Sumner et al. 2003, Hagel et al. 2008). In addition the fact that usually MS are hyphenated with separation techniques increase the possibilities of detection of compounds in small quantities present in complex mixtures such as plant extracts (Allwood et al. 2010). On the other hand, the main drawback of LC-MS techniques is the reproducibility issue but careful handling during data acquisition and exhaustive assessment of variability during acquisition can result in reliable data (Sangster et al. 2007). Overall, detailed information can be extracted concerning metabolite profile of plant extracts and structural data related to containing metabolites.

3.3. Data preprocessing

Data preprocessing is an essential procedure of metabolomics pipeline to transform the raw data into compatible format for data analysis. In metabolomics analysis, the raw data usually contain a large number of measurement outputs that has to be filtered and organized properly into a single data matrix in order to use the appropriate information for the data analysis (Goodacre et al. 2007). Due to the different mode of operation of metabolomics analytical platforms, raw data acquired by dissimilar platforms need particular handling using specific software designed for this purpose.

3.3.1. NMR data preprocessing

Since NMR is characterized by satisfactory reproducibility not extensive preprocessing is needed comparing with MS techniques. In NMR metabolomics data preprocessing usually includes baseline correction, alignment and binning (Smolinska et al. 2012).

According to the nature of samples acquired and the contained metabolites baseline distortion phenomena and shifted peaks appearing in the NMR datasets hamper the data analysis process. In these cases, specific algorithms for the correction of these phenomena have been developed (Smolinska et al. 2012). Concerning preprocessing algorithms for NMR data, there several software available such as Metabolab (Ludwig et al. 2011), MVAPACK (Worley et al. 2014) and Automics (Wang et al. 2009). For instance, Metabolab, a MATLAB based software package, is designed to facilitate NMR data processing by providing automated algorithms for processing series of spectra. Alternatively, when minor corrections need to be performed, manual processing using software for NMR data acquisition, such as TopSpinTM (Bruker), leads to satisfactory results.

Once the NMR data have been preceded, they have to be digitalized to numeric values for further statistical analysis. A common way to deal with NMR peaks in a very consistent way is binning (bucketing). Literally, the NMR spectrum is divided into a series of small bins (buckets) and the sum of intensities of the signals in each bin is calculated and exported into a data matrix. Binning can be used for all types of spectra, however, is more frequently used with NMR analysis. There are two different types of bins the equidistant and the non-equidistant. The advantage of the latter is that avoids peak splitting which is very usual during equidistant binning. In plant metabolomics field, usually equidistant bins of 0.04 ppm size are used to proceed with the data analysis despite the fact that this binning size reduce significantly the resolution of the analysis (Kim et al. 2010). Alternatively, smaller bins or full resolution NMR data can be utilized for the statistical analysis (Rasmussen et al. 2006).

3.3.2. LC-MS preprocessing

The preprocessing step in a LC-MS metabolomics platform is performed as mentioned above to optimize the resulting matrix of identified peaks and transform the data into a format that makes the subsequent statistical analysis easier and more robust. Since the preprocessing in LC-MS is more demanding comparing to NMR a series of software

packages have been developed for this purpose (Castillo et al. 2011), among them the most popular are XCMS (Smith et al. 2006), MZmine (Pluskal et al. 2010), MAVEN (Melamud et al. 2010), MetAlign (Lommen 2009). The preprocessing procedure can be separated into two main steps: the peak detection and the annotation of isotope and adduct peaks.

The peak detection procedure consists in the characterization of peaks in the three dimensional space (time, mass, intensity) as defined by the LC-MS data and estimation of the peak intensity. The major problem in peak detection procedure is the identification of the real peaks from the noise peaks and specifically for small abundant peaks close to noise level. According to the utilized software different algorithms are applied for peak detection. For instance, XCMS software incorporates the matched filter algorithm (Danielsson et al. 2002) and centWave algorithm (Tautenhahn et al. 2008). An important issue in the application of a given software is to understand the features and the underlying algorithms used in the software to allow optimal choice of parameters.

After the peak detection step a number of features are generated including features corresponding to isotope and adduct ions that hamper the compound identification and subsequently the biological interpretation. In order to identify the isotope and adduct ions in a data matrix inheriting from the peak detection step, specific algorithms are incorporated according to the applied software. In the case of XCMS, CAMERA software is usually combined for this purpose which has been proven efficient in isotope and adducts annotation even in high complex data (Kuhl et al. 2011).

3.4. Data pretreatment

Data pretreatment consist in the correction and refining of the dataset removing confounding variation originated from experimental inaccuracies retaining the inherent biological variation. Independently from the analytical platform applied, the preprocessing step aims to convert the raw data into numeric matrices compatible with the statistical analysis, however, before proceeding with the data analysis it is important to clean the data matrix from undesired noise peaks and focus only to the features that reflect the metabolite composition of the analysed samples. These clean data can be directly used as the input for statistical analysis; nevertheless, in the majority of cases some additional pretreatment steps are necessary to reduce the influence of disturbing factors such as noise measurements. Therefore, usually, normalization and scaling procedures are applied. Overall, pretreatment

contributes to the comparability of the different data features and constitutes an essential step before data analysis (Craig et al. 2006).

3.4.1. Data filtering

Data acquisition in metabolomics using various analytical techniques leads to the simultaneous measurements of real contained metabolites as well as noise. After the preprocessing step a general matrix is generated containing all measurements provided from the raw data including noise data. Therefore, the elimination of noise data may facilitate significantly the data analysis process. However, this step needs a level of justification in order to avoid exclusion of biologically important information.

Concerning the NMR metabolomics platform, data filtering usually consists in the elimination of the buckets corresponding to noise regions. This is performed by careful determination of noise regions in the whole dataset. LC-MS metabolomics platform is characterized by a significantly greater sensitivity than the NMR metabolomics platform which renders the method much more prone to noise detection. Therefore, during the peak picking procedure the algorithm is assumed to detect also 'false' peaks that are not representing any biological information and complicate the procedure of data analysis. This is directly comprehensible since an important number of peaks is generated from the blank samples raw data. These 'false' features are considered as the 'background' of the dataset and they are explained as features originating from the solvents, tubes, vials, or impurities. In addition, 'false' features could be characterized also peaks randomly detected by the algorithm with no consistency in the samples consequently no biological importance. Therefore, a data filtering step is an essential procedure for LC-MS based metabolomics to reduce all the aforementioned 'false' peaks in order to enhance the power of the data analysis (Hackstadt et al. 2009).

3.4.2. Missing values

Missing values in metabolomics analysis is a common phenomenon occurring as missing data in a final data matrix containing intensity numbers for a given dataset. Missing values may arise for a number of reasons including biological or/and technical reasons. An important factor for the occurrence of missing values is the applied technique. High resolution mass spectrometry techniques are much more affected from this phenomenon

than continuous techniques such as NMR (Hrydziuszko et al. 2012). In addition, the percentage of missing values in the entire dataset may range depending on the type of the metabolomics experiment. The handling of missing values is an important step in the pretreatment pipeline since it is affecting the forthcoming data analysis. Up to now two main approaches are applied in regards to the handling of missing values. A first approach suggests the exclusion of the variables containing missing values. Focusing only on the variables with no missing values could be applied to datasets with small proportion of variables that are affected by missing data. However, regarding HRMS analysis this is very rare (Xia et al. 2009) and exclusion of variables with missing data in this case would lead to reduction of statistical power. Therefore, in HRMS metabolomics analysis the most common practice is to impute missing values using appropriate algorithms (Xia et al. 2011). Imputation of missing values using appropriate algorithms provides a consistent and automated way to create data matrices compatible with data analysis process.

3.4.3. Normalization, scaling

Once the dataset is filtered and the missing values are imputed (when necessary) a critical step before data analysis is the normalization, centring and scaling procedures.

A common problem arising in metabolomics studies is the dilution issue. When dealing with a large number of samples, smaller or larger variation in sample concentrations are usually arrive. Therefore, normalization is typically applied to correct technical variation originating from sampling, sample work-up and analytical errors. The term normalization is used to express the treatment applied to each observation/ samples (also in Metaboanalyst platform). Principally this involves applying a correcting factor so that the sum of all intensities equals unity, making overall intensity scales comparable across samples. Alternatively, the normalization can be applied using an internal standard feature to normalize across all the samples or sample-specific normalization option allows to manually specify a normalization value for each sample, for instance, on the basis of dry weight (Xia et al. 2011).

Another consideration before the data analysis step is the centring and scaling of the data. Normally, the metabolome of a given plant organism issued for analysis consists metabolites presenting significant differences in concentration levels. However, the contribution of each metabolite in a given biological response may be independent of its

concentration. Therefore, a number of scaling techniques are available to scale the metabolites accordingly in order to approach better the biological information. This operation is performed on the variables of the dataset across all samples; hence it is usually referred as column wise scaling. As presumed, the scaling technique is not only dependent on the nature of the dataset but also on the statistical method chosen for analysis.

Centring may be regarded as a scaling method; however, it is usually combined with other scaling factors. Centring is a technique that is commonly applied for the treatment of metabolomics data. It converts all the concentrations to fluctuations around zero instead of around the mean of the metabolite concentrations. This permits the adjustment for differences in the offset between high and low abundant metabolites (Bro et al. 2003). Therefore, the statistical analysis of the data is not affected from the relative variation between the samples (expressed by the mean value) but it focuses only on the fluctuating part of the data increasing the fit to data (van den Berg et al. 2006).

Other scaling factors focus on the transformation of the features in order to be more comparable between them. Some available scaling features in Metaboanalyst platform are log transform, autoscaling, Pareto scaling and range scaling. Each of these techniques is treating differently the data and thus the decision of the scaling method has a direct impact on the results (van den Berg et al. 2006). For instance, in autoscaling each feature of the table is scaled so that it has unit variance using the standard deviation as scaling factor, therefore, the variables become equally important and the weights of each feature reflect their correlation. However, the main disadvantage of this technique is that 'false' peaks possibly present in the dataset become as well equally important and may obscure the interpretability of the model. Pareto scaling is similar to autoscaling but in this case the scaling factor is the square root of the standard deviation. Pareto scaling is considered in between no scaling at all and auto-scaling and gives the variable a variance equal to its standard deviation instead of unit variance. Consequently, the importance of the relatively abundant variables is reduced but it keeps the proportions more closely to reality.

3.5. Data analysis

Data analysis in metabolomics is an essential step that relates the analytical outcome with the biological information. The aim of data analysis step is the application of appropriate statistical methods in order to extract the biological important information

related to the initial scope of a given research. That includes the development of efficient and robust methods for modelling and analysis of these complicated data matrices (Trygg et al. 2006). The selection of the statistical method to apply is related from one hand to the purpose of the study and from the other hand on the nature of the dataset. Different methods are preferred for regression or classification/ discrimination purposes while the visualization capability of a method is always an important factor to consider (Steinfath et al. 2008). In metabolomics context, statistical methods should be able to unravel patterns and sources of variation in the complex datasets generated after the multistage metabolomics pipeline. Usually, due to the multidimensionality of metabolomics data, the recognition of comprehensive patterns and their interpretation has imposed the application of multivariate data analysis methods (MVDA). However, univariate or clustering statistical methods are also facilitating the analysis of metabolomics data. The application of all the aforementioned statistical methods is performed using both web based (e.g. Metaboanalyst, MeltDB, metaP-server) and commercial platforms (e.g. Simca, SAS) for data analysis (Xia et al. 2011).

A number of multivariate techniques for statistical analysis are available. Among them, the most widely used are the multivariate reduction techniques or projection techniques which are designed to overcome dimensionality problems through the compression down to simple components and pseudo variables in the form of weighted linear combinations of the original data (Liland 2011). Principal Component Analysis (PCA) and Partial Least Squares (PLS) are the most commonly used multivariate data analysis methods (Kemsley 1996).

In particular, PCA is an unsupervised method that attempts to explain the variance of the independent variables of the dataset (X dataset) by its linear transformation in principal components based on the covariance of the dataset without any previous knowledge. PCA decomposes the data into score vectors and loading vectors in a way that the variation is maximized in the first components and decreasing in the subsequent components. The scores are representing the positions of the observations in a new, rotated coordinate system and the loadings the weights for the original variables to transform them into the scores. Therefore, PCA is used as an exploratory data analysis for the visualization of inherent patterns based on the classification of the different scores and the expiration of the biomarkers responsible for this classification (Wold et al. 1987, Bartel et al. 2013).

PLS is a supervised method that attempts the relation of the data matrix containing the independent variables (X dataset) with a matrix containing the dependent variables (Y

dataset) corresponding to a given response for each sample. In PLS the decomposition of the X dataset in latent variables is performed in order to obtain the optimal prediction of the Y dataset. PLS can be used as a regression method to unravel correlated variables between X and Y datasets. In addition, PLS can be combined with Discriminant Analysis (PLS-DA) for classification purposes to uncover discrimination between predefined groups (Wold et al. 2001, Lindon et al. 2006).

3.6. Biomarker identification

All abovementioned statistical methods results in the identification of important or discriminant features (biomarkers), corresponding to metabolites related with given biological information. The biomarker identification step is the most important in the metabolomics pipeline in order to translate the data analysis result into biological knowledge. However, this step specifically in plant metabolomics field constitutes a significant bottleneck in the overall downstream analysis (Moco et al. 2007). The extremely big number of secondary metabolites in plants as well as the absence of plant specific and complete databases hampers significantly this step (Wolfender et al. 2013). The assignment of biomarkers is based on the spectroscopic or spectrometric and physicochemical characteristics depending on the applied analytical platform. The linkage between the feature information and the biological knowledge is still performed manually hindering the throughput of the overall downstream analysis. During the last years, many individual strategies have been developed to overcome this issue based on publically available databases not specified in secondary metabolites or in-house databases. Nevertheless, still a lot of work is demanded to achieve the metabolite identification of biomarkers in a high throughput way.

4. The case study of *Acronychia* species

Acronychia genus is composed of 48 species among them a great number that has not been investigated up to now. *Acronychia* species are distributed widely in Australasia and New Caledonia and they possess an important position in the eastern world as food condiments and therapeutics utilized by the traditional medicine mainly for their anti-inflammatory properties. So far, a diverse range of metabolites have been isolated from the

investigated species among them compounds possessing significant pharmacological properties regarding a number of different targets (Epifano et al. 2013) (see Chapter 1 Introduction 2 for further details).

Given this background and the previous obtained knowledge on the phytochemical and pharmacological profile of *Acronychia pedunculata* (Chapter 1, 2), the investigation of representative populations of *Acronychia* species was aimed using metabolomics strategies. In this context, different *Acronychia* samples were selected belonging to diverse species and organs and collected from six different locations in Malaysia and Vietnam in order to develop a metabolic profiling methodology for the investigation of these samples. NMR and UPLC-HRMS metabolomics platforms were developed validating every step of this downstream analysis for the phytochemical and pharmacological evaluation of *Acronychia* species. Therefore, a workflow was designed to collect a large volume of data from both metabolomics platforms for the investigation of different biological questions and their exploitation as a proof of concept for the development of specific strategies concerning particular steps of the pipeline. As initial objective, taking advantage of the high resolution UPLC-ESI(±)-LTO-Orbitrap platform the dereplication of the known compounds in investigated and unexplored *Acronychia* species was intended since the assessment of the presence of known compounds in other species would be of vital importance. Towards this goal, a ^{13}C NMR based dereplication strategy was planned in parallel to characterize *Acronychia* species. The application of metabolomics strategies was also applied to develop a statistical model for the discrimination and classification of diverse species, origins and organs utilizing unsupervised multivariate methods. The construction of this robust model enabled the identification of biomarkers responsible for the discrimination of these *Acronychia* samples and the investigation of taxonomic issues. Concerning the metabolite identification, which is a challenging task in the overall process, the combination of different strategies was proposed among them the integration of the data from the two different metabolomics platforms using a sparse PLS analysis in order to get a better insight into the structural nature of the significant metabolites. Moreover, metabolomics incorporating the predictive ability of PLS regression analysis were utilized as a tool for the prediction of bioactivity in complex plant mixtures allowing subsequently the tracking of bioactive metabolites. The statistical and structural validation of this high throughput process renders this approach a promising tool for drug discovery purposes.

Experimental

1. Chemicals and instrumentation

During this study, the extraction of *Acronychia* samples was performed using an ultrasound bath Elmasonic S100H (Elma, Germany). Evaporation of the *Acronychia* extract samples was performed with the aid of a GeneVac HT-4X EZ-2 series evaporator Lyospeed ENABLED (Genevac Ltd, UK).

All nuclear magnetic resonance (NMR) spectra, apart ^{13}C NMR, were recorded at 300 K on a Bruker Avance III 600 spectrometer operating at 600 and 150 MHz for ^1H and ^{13}C , respectively, equipped with a 5 mm BBI probe and by using CDCl_3 (Sigma-Aldrich) as deuterated solvent and hexamethyldisiloxane (HMDS) as internal standard (Sigma-Aldrich). Chemical shifts (δ) are expressed in ppm with reference to the solvent signals ($\delta\text{ H } 7.26/\delta\text{ C } 77.0$) and coupling constants in Hz. The 2D NMR experiments (JRES, COSY, HSQC and HMBC) were performed using standard Bruker microprograms. Centrifugation of the *Acronychia* extract samples for the sample preparation of the NMR samples was achieved by a Mikro 200R centrifuge (Hettich Lab Technology, Germany).

^{13}C NMR analyses were recorded at 298 K on a Bruker Avance AVIII-600 spectrometer (Karlsruhe, Germany) operating at 150 MHz for ^{13}C . The spectrometer is equipped with a cryoprobe optimized for ^1H detection and cooled ^1H , ^{13}C coils and preamplifiers, with a ^{13}C S/N calculated for 1271:1 from the standard ASTM test sample (60% C_6D_6 , 40% dioxane).

The UPLC-ESI($\pm$)-HRMS/MS experiments were carried out using a LTQ-Orbitrap apparatus (Thermo Finnigan, San Jose, CA, USA). The UPLC system is equipped with an Acquity pump and autosampler. The solvents used were LC-MS grade methanol (MeOH) and water purchased by Merck (Germany) and formic acid (FA) by Sigma-Aldrich (Germany).

2. Harvesting of plant material

In the current study, approximately 500g of dried plant materials of 20 *Acronychia* species from different species, organs and origin were provided by the French National Centre for Scientific Research (CNRS) (Gif-sur-Yvette, Paris, France). As illustrated in Table XI three different *Acronychia* species were subjected to this study, namely *Acronychia laurifolia* Bl., *Acronychia porteri* Hk.f. and *Acronychia pedunculata* (L.) Miq. from which mainly the leaves and the barks were collected separately while a sample of fruits of *Acronychia pedunculata* (L.) Miq. was also available. The plant materials were collected in two different countries, specifically 15 *Acronychia* samples were harvested in Malaysia and 5 in Vietnam. Concerning Malaysia, samples were collected in 4 distinct locations distributed from north to south of West Malaysia while in Vietnam the collection locations were situated in the north part. The plant samples were dried under shade conditions and ground into fine homogeneous powders.

Table XI: Detailed list of plant material used for the metabolomics study of *Acronychia* genus

Codes	Species	Plant Part	Country	Collection Location	Date
K-4652 (B)	<i>Acronychia laurifolia</i> Bl.	Bark	Malaysia	Ulu jelai, Pahang	20.11.1996
K-4652 (L)	<i>Acronychia laurifolia</i> Bl.	Leaves	Malaysia	Ulu jelai, Pahang	20.11.1996
K-4853 (B)	<i>Acronychia laurifolia</i> Bl.	Bark	Malaysia	Mersing, Johore	03.12.1998
K-4853 (L)	<i>Acronychia laurifolia</i> Bl.	Leaves	Malaysia	Mersing, Johore	03.12.1998
K-5445 (B)	<i>Acronychia porteri</i> Hk.f.	Bark	Malaysia	Gerik, Perak	12.07.2007
K-5445 (L)	<i>Acronychia porteri</i> Hk.f.	Leaves	Malaysia	Gerik, Perak	12.07.2007
KL-4727 (B)	<i>Acronychia laurifolia</i> Bl.	Bark	Malaysia	Mersing, Johore	26.08.1997
KL-4727 (L)	<i>Acronychia laurifolia</i> Bl.	Leaves	Malaysia	Mersing, Johore	26.08.1997
KL-4878 (B)	<i>Acronychia porteri</i> Hk.f.	Bark	Malaysia	Mersing, Johore	07.04.1999
KL-4878 (L)	<i>Acronychia porteri</i> Hk.f.	Leaves	Malaysia	Mersing, Johore	07.04.1999
KL-4882 (L)	<i>Acronychia laurifolia</i> Bl.	Leaves	Malaysia	Mersing, Johore	08.04.1999
KL-5197 (B)	<i>Acronychia laurifolia</i> Bl.	Bark	Malaysia	Jelebu, N. Sembilan	16.02.2006
KL-5197 (L)	<i>Acronychia laurifolia</i> Bl.	Leaves	Malaysia	Jelebu, N. Sembilan	16.02.2006
KL-5465 (B)	<i>Acronychia porteri</i> Hk.f.	Bark	Malaysia	Mersing, Johore	28.08.2007
KL-5465 (L)	<i>Acronychia porteri</i> Hk.f.	Leaves	Malaysia	Mersing, Johore	28.08.2007
VN-0179 (B)	<i>Acronychia pedunculata</i> (L.) Miq.	Bark	Vietnam	Chi Linh, Hai Hung	05.10.1996
VN-0179 (L)	<i>Acronychia pedunculata</i> (L.) Miq.	Leaves	Vietnam	Chi Linh, Hai Hung	05.10.1996
VN-0874 (B)	<i>Acronychia pedunculata</i> (L.) Miq.	Bark	Vietnam	Huong Khe, Ha Tinh	18.08.2001
VN-0874 (FR)	<i>Acronychia pedunculata</i> (L.) Miq.	Fruits	Vietnam	Huong Khe, Ha Tinh	18.08.2001
VN-0874 (L)	<i>Acronychia pedunculata</i> (L.) Miq.	Leaves	Vietnam	Huong Khe, Ha Tinh	18.08.2001

3. Extraction of plant material

All *Acronychia* samples were extracted in six analytical replicates using ultrasonic assisted extraction. Since our goal was to focus on the secondary metabolites of *Acronychia*, disregarding the primary metabolites eventually, the selection of EtOAc as extraction solvent was considered optimal. In addition, a compatible for both NMR- and MS-based metabolomics 'one pot' extraction was performed in order to allow a comparative analysis of the metabolite data derived from different analytical platforms. Specifically, approximately 400 mg ($\pm 5\%$) of dried plant material were weighted for all the 120 samples in eppendorf tubes. After the addition of 1.5 mL of EtOAc in the eppendorf tubes the plant materials were extracted for 30 min in ultrasound bath controlling the temperature under 40° C. Then, the extraction solutions were transferred to pre-weighted eppendorf tubes after their filtration using 0.45 nylon filters. The plant residues were added another 0.75 mL of EtOAc and extracted for 30 min more under the same conditions. The solutions from the second extraction step were added to the pre-weighted eppendorf tubes and evaporated until dry using SpeedVac apparatus.

4. Sample preparation for analysis

The 120 extracts obtained from the 20 different biological samples of *Acronychia* species were intended to be analysed by two different metabolomics platforms and to be evaluated for their anti-inflammatory activity by assessing their capacity to inhibit 5-lipoxygenase (5-LO) enzyme. In this regard, a rational separation of the obtained extracts in 3 aliquot types was designed prior to any specific sample preparation in order to obtain later solutions compatible with the specific analysis. Thus, the dilution of each extract in EtOAc was performed to obtain a stock solution of 10 mg/mL and then 700 μ L, 20 μ L and 67.5 μ L were transferred to different tubes and evaporated until dryness using SpeedVac apparatus. Consequently, dried extracts of 7 mg, 0.2 mg and 0.675 mg were obtained for the NMR analysis, UPLC-HRMS analysis and the biological evaluation, respectively, of all *Acronychia* extracts (Figure 29). Obviously, for each metabolomics platform a special sample preparation is required that will be presented in the relative sections below.

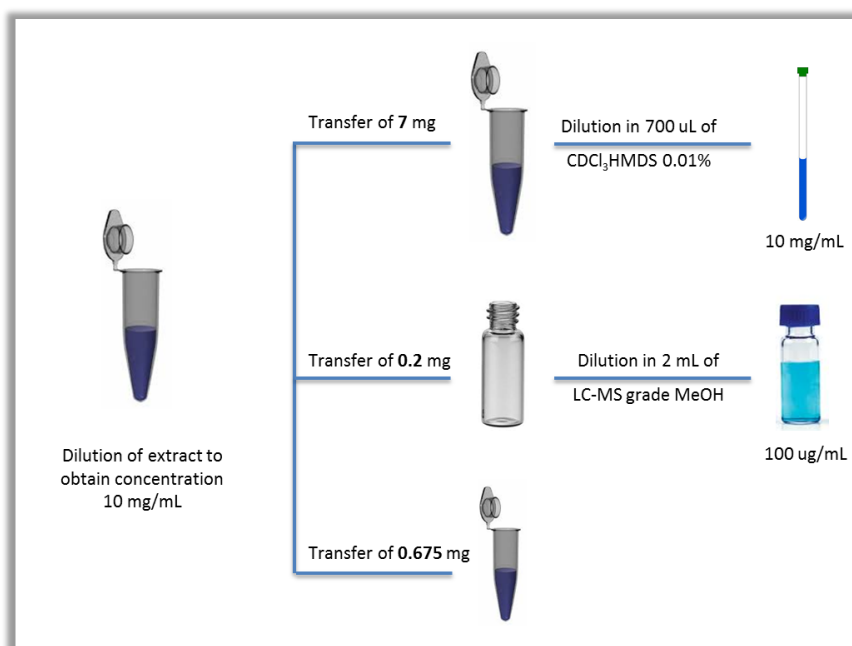


Figure 29: Schematic representation of multiple aliquots of extracts samples obtained for the sample preparation of different experiments

4.1. NMR sample preparation

All dried extracts (7 mg) were resuspended in 700 μ L of CDCl_3 0.01% hexamethyl disilane (HMDS) maintained in 0 $^{\circ}\text{C}$ during the sample preparation process to obtain final solution of 10 mg/mL. After complete dilution, using vortex and/or ultrasonication the samples were centrifuged at 15000 rpm for 5 min in order to avoid precipitant presence. Then 550 μ L of the supernatant were transferred into 5 mm diameter NMR tubes and stored in the fridge until data acquisition.

4.2. UPLC-HRMS sample preparation

As described above a part of the 120 prepared extracts (6 replicates of 20 samples) was used for the acquisition of LC-MS experiments. In particular, solutions of 100 μ g/mL were prepared by diluting 0.2 mg in 2 mL of methanol. Among the solvents that were tested for the dilution step methanol was chosen because it resulted in transparent solutions in this concentration and due to its compatibility with the LC-MS analysis. In parallel, in order to assess the quality of LC-MS data, a quality control (QC) sample was

prepared by pooling 15 μL of each extract *Acronychia* sample solution resulting in the QC sample solution of 1.8 mL totally representative of the entire sample set.

5. NMR data acquisition and spectral processing

For metabolomics analysis, ^1H NMR and Jres spectra of all 120 *Acronychia* samples were acquired within a time-interval of approximately 74 hours. Repeated control experiments in between the time frame showed no additional variation. In particular, ^1H NMR spectra were recorded using the following parameters: digital resolution 0.275 Hz/point (128k complex data points), relaxation delay = 2 s, acquisition time = 3.64 s, spectral width = 30ppm, number of transients = 256, giving a total acquisition time of 24 min and 33 sec. An exponential window function with lb = 0.3 was applied to each FID prior to Fourier transformation. 2-D J-resolved NMR spectra were acquired using 4 scans per 40 increments, which were collected into 12k data points, using spectral widths of 18028.846 Hz in F2 (chemical shift axis) and 78.123 Hz in F1 (spin-spin coupling constant axis). A 2.0 s relaxation delay was employed, giving a total acquisition time of 7 min and 38 sec. Data sets were zero-filled to 256 points in F1, and both dimensions were multiplied by sine-bell functions (SSB = 0) prior to double complex FT. J-resolved spectra tilted by 45° were symmetrized about F1, baseline corrected, and then calibrated, using TopSpin (version 3.1, Bruker). Data were exported as the 1-D projection (F2 axis) of the 2-D J-resolved spectra pJRES.

^{13}C NMR experiments were recorded using standard zgpg pulse sequence, acquisition time = 0.909 s, relaxation delay = 3 s, spectral width = 239 ppm and 4096 scans to obtain a satisfactory signal to noise (S/N) ratio. An exponential window function with EM = 1 Hz was applied to each FID prior to Fourier transformation.

6. UPLC-HRMS data acquisition

6.1. UPLC conditions

The UPLC method was developed using the QC sample as reference. The chromatographic conditions applied were a compromise between metabolite resolution,

retention time stability and sample throughput. The analysis of the samples was performed on an Acquity UPLC BEH C18 column 1.7 μm (2.1 $\times$ 150 mm,) and a gradient of 30 min was chosen in order to enable descent separation of the majority of compounds. The mobile phase was consisted of methanol and 0.1% aqueous formic acid. The gradient elution program was composed of a conditioning step followed by a steep gradient step passing from the aqueous solvent to methanol and a longer step of a flatter gradient of high methanol content due to the plethora of medium to low polarity metabolites present in the extract samples. At the end a washing step and a reconditioning step were performed in order avoid contamination and ensure stable performance. The UPLC method is presented in details in Table XII. The flow rate was set at 200 $\mu\text{L}/\text{min}$ and the sample and column temperatures were stabilized at 10°C and 30°C respectively.

Table XII: Reversed phase UPLC-MS gradient program

Time (min)	H ₂ O+0.1% F.A (%)	MeOH (%)	Flow rate ($\mu\text{L}/\text{min}$)
0	95	5	200
1	95	5	200
7	20	80	200
22	0	100	200
25	0	100	200
26	95	5	200
30	95	5	200

Column: BEH C18 column (2.1 $\times$ 150 mm, 1.7 μm) Acquity UPLC® (Waters)

Injection volume: 5 μL

Temperature: sample 10°C; column 30°C

6.2. HRMS conditions

The LTO-Orbitrap XL hybrid mass analyser used was equipped with an ESI source. For the metabolic profiling of *Acronychia* extracts, data were acquired in both positive and negative modes using a mass range of 205-1000 m/z . Full scan experiments were acquired with a resolution of 30000 in profile mode. The operating conditions of the source are presented in details in Table XIII. Moreover, tandem MS analysis of the 20 different biological samples was performed using the same conditioned applied for both positive and negative

mode incorporating a data dependent scanning event. For MS/MS experiments, the collision energy (CID) was set at 35%, the activation time at 30 ms (q , 0.25) and a window of 2 u was used to isolate the precursor ions.

Table XIII: Operating conditions of ESI in positive and negative modes

Parameters	ESI(+)	ESI(-)
Spray voltage (kV)	3.6	3.5
Sheath Gas	45	45
Auxiliary Gas	10	10
Sweep Gas	0	0
Capillary Voltage (V)	35	-35
Capillary Temperature (°C)	300	350
Tube Lens (V)	110	110

7. NMR data preprocessing

Both automated and manual processing of the NMR spectra was performed using Metabolab and TopSpin software, respectively. In particular, in both cases the data were phased and baseline corrected using the incorporated algorithms. In addition, the spectra were referenced to the internal hexamethyldisiloxane (HMDS) signal at 0.062 ppm. Both approaches lead to similar results, as observed after data analysis.

Binning was performed either directly from Metabolab software or by using the Amix 2.7 (BrukerBioSpin GmbH) software when manual processing with TopSpin 3.1 (BrukerBioSpin GmbH) was applied. In both cases, the ^1H NMR and pJRES data were binned using 0.01 and 0.04 ppm bin sizes. The resulting matrices included 120 rows corresponding to the individual samples and 1900 or 475 columns corresponding to the different variables for 0.01 and 0.04 ppm bin sizes, respectively. The ^{13}C NMR spectra were manually processed by TopSpin 3.1, aligned and binned in equidistance bins of 0.05 ppm resulting in a matrix of 120 rows corresponding to the individual samples and 889 columns corresponding to the different variables.

8. UPLC-HRMS preprocessing

During this study, XCMS package is applied for feature detection followed by annotation step using CAMERA package. Both packages are implemented in R environment and provided by bioconductor (<http://www.bioconductor.org/>).

8.1. Raw data preparation

Prior to any preprocessing step all samples were converted to centroid data in NetCDF format in order to be compatible with the XCMS software. In this file format, the data are stored as separate lists of mass/intensity pairs with each list representing one scan (Smith 2014). This procedure was accomplished using the MSConverter tool provided by proteowizard (<http://proteowizard.sourceforge.net/>).

8.2. Peak detection/ Annotation of isotope and adduct peaks

The peak detection of *Acronychia* sample dataset was performed by XCMS using the matched filter algorithm while the individual parameters were tuned according to the experimental conditions. In particular, for peak peaking slices (*step*) of m/z 0.02 mass unit wide were set, a signal-to-noise ratio (*snthresh*) of 3 and a minimum difference (*mzdiff*) of 0.005 m/z mass for peaks with overlapping retention times were defined. Moreover, a 10 s binning function (*bw*) was applied to group variables. Annotation of isotope and adduct peaks was performed using CAMERA software. Therefore, after the preprocessing step, a peak table was generated from the UPLC-ESI(+)-MS data contained 197 columns representing the individual injections and 8093 rows representing the integrated peak area. Accordingly, from the UPLC-ESI(-)-MS data a peak table of 197 columns and 2909 rows was extracted after the XCMS process.

9. NMR data filtering / Noise reduction

Concerning NMR data, the matrices inheriting from preprocessing step included 120 rows corresponding to the individual samples and 1900 or 475 columns corresponding to the different variables for 0.01 and 0.04 ppm bin sizes, respectively. With the perspective to

filter out the noise bins from the initial data matrices a careful inspection of the 120 raw data simultaneously lead to the determination of the noise regions. The bins corresponding to these noise regions were consequently excluded from both ^1H NMR and pJRES matrices. As a result the filtered data matrices contained 1394 and 308 variables for 0.01 and 0.04 bin sizes, respectively. Likewise, the ^{13}C NMR matrix was filtered excluding noise bins, the solvent signals and fatty acid spectral regions (around 29 and 30 ppm, 14.1 ppm, 31.9 ppm and 22.7 ppm) resulting in a data matrix of 855 variables.

10. UPLC-HRMS data filtering / Noise reduction

The peak list generated after the preprocessing of the UPLC-ESI(+)-HRMS data contained 197 rows representing the individual injections and 8093 columns representing the integrated peak area. Towards the elimination of 'false' peaks, the dataset was filtered initially subtracting the features corresponding to peaks present in the blank samples. Additionally, the peaks presenting a RSD > 25% in the QC samples were eliminated. Moreover, filtering of the features that did not vary significantly in the different sample groups was performed by extracting those presented p-values greater than 0.05 calculated using ANOVA in Metaboanalyst platform (www.metaboanalyst.ca). As a result, the peak table was reduced significantly in dimension accounting 4679 columns corresponding to features more relevant to the biological information.

11. UPLC-HRMS missing values estimation

The estimation of missing values was performed at the UPLC-ESI(+)-HRMS dataset by Metaboanalyst web platform importing the peak table after data filtering. In particular, missing values were calculated for 7.6% of the total values which was translated into 42458 missing values in the entire dataset. Missing values were imputed by the PPCA algorithm to continue with the downstream analysis.

12. Normalization and scaling

Normalization of all datasets inheriting from both platforms was performed using Metaboanalyst platform and scaling using both Metaboanalyst and Simca 13.0 software (Umetrics, Umea, Sweden). In particular, all datasets, independently from the analytical platform used, were normalized to the sum intensity in order to avoid possible dilution variance to influence the data analysis. Moreover, according to the scope of each data analysis, mainly Pareto scaling or autoscaling were applied to adjust scale differences between the detected features.

13. Data analysis

Univariate data analysis was performed using analysis of variance (ANOVA) for the assessment of significance in diverse cases. MVDA including PCA, PLS, PLS-DA, OPLS and OPLS-DA were performed in Simca 13.0 (Umetrics, Umea, Sweden). Moreover, PCA analysis for both initial evaluation of the datasets and data analysis purposed was performed also in Metaboanalyst platform. Hierarchical Cluster Analysis (HCA) and heatmaps based on ANOVA and combined with HCA were generated from Metaboanalyst platform. Clustering analysis of ^{13}C NMR data and heatmap generation were performed in PermutMatrix, version 1.9.3 (LIRMM, Montpellier, France). Sparse PLS (s-PLS) analysis was conducted using mixOmics package (<http://cran.r-project.org/web/packages/mixOmics/index.html>) implemented in R statistical language 3.0.3 (<http://www.r-project.org>). Specific details for statistical methods employed in this study are given in the respective sections.

14. Biomarker identification

During this study, the identification of secondary metabolites corresponding to important features extracted from both analytical platforms was performed mainly manually by matching the experimental data to an in-house database containing all isolated secondary metabolites from *Acronychia* genus. Moreover, concerning the characterization of metabolites those have not been reported from *Acronychia* genus, publically available databases were utilized for HRMS data. In particular mainly METLIN (metlin.scripps.edu),

MassBank (www.massbank.jp) and Human Metabolome DataBase HMDB (www.hmdb.ca) were used as well as MetFrag (msbi.ipb-halle.de/MetFrag) for MS/MS search.

15. Anti-inflammatory evaluation of *Acronychia* extracts

15.1. Sample preparation

An aliquot from each *Acronychia* sample particularly destined for anti-inflammatory *in vitro* biological evaluation was prepared during sample preparation procedure (see experimental 2.3 for further details). These aliquots were used to prepare 5 different concentration solutions for the *in vitro* assay. In particular, 0.675 mg of each sample was diluted in 22.5 μ L of DMSO to obtain a stock solution of 30 μ g/ mL. Then, successive dilutions with appropriate quantity of DMSO were performed to obtain solutions of 10, 3, 1, 0.3 and 0.1 μ g/mL that were used for the biological evaluation.

15.2. *In vitro* 5-LO cell free assay

5-LO was expressed in *E. coli* BL21 (DE3) cells, transformed with pT3-5LO, and purification of 5-LO was performed as described previously (Fischer et al. 2003). Thus, *E. coli* were collected by centrifugation ($7,700 \times g$ for 15 min), lysed with 50 mM triethanolamine/HCl, pH 8.0, 5 mM ethylenediaminetetraacetate (EDTA), 60 μ g/mL soybean trypsin inhibitor (STI), 1 mM phenylmethylsulphonyl fluoride (PMSF), 1 mM DTT and 1 mg/mL lysozyme, homogenized by sonication (3×15 sec) and centrifuged at $10,000 \times g$ for 15 min and then at $40,000 \times g$ for 70 min at 4 °C. The supernatant was then applied to an ATP-agarose column (Sigma; Deisenhofen, Germany). Partially purified 5-LO was immediately used for activity assays.

For determination of 5-LO activity in cell-free assays, samples of partially purified human 5-LO (1 mL, in PBS buffer containing 0.1 % glucose and 1 mM EDTA) were incubated 10 min at 4 °C with vehicle (0.1% DMSO, control) or test extracts, pre-warmed for 30 sec at 37 °C and 2 mM CaCl_2 and the indicated concentrations of AA were added. The reaction was stopped after 10 min at 37 °C by addition of 1 mL ice-cold methanol and 30 μ L of 1 N HCl, 200 ng prostaglandin B₁ and 500 μ L of PBS were added.

Chapter 3: Metabolomics approaches in *Acronychia* genus

Formed 5-LO metabolites, 12(S)-H(P)ETE and 15(S)-H(P)ETE were extracted and analyzed by HPLC as described (Werz et al. 2002). 5-LO products include LTB₄ and its all-trans isomers, and 5(S)-H(P)ETE.

Therefore, the 5-LO product formation (% of control) was calculated for all extracts assayed in five different concentrations (0.1, 0.3, 1, 3, 10 µg/mL) and the IC₅₀ values for each sample were also calculated.

Results and Discussion

1. NMR and UPLC-HRMS plant metabolomics workflow

During the last years, plant metabolomics has been developed as an emerging tool for plant sciences and natural products chemistry research (Hall 2006). The perspective of application of metabolomics strategies in plant sciences is to measure the entire number of contained metabolites, both qualitatively and quantitatively in order to obtain a complete insight into the metabolite composition under given conditions. However, the complexity of plant extracts in terms of number, concentration levels and nature of contained metabolites renders this goal unfeasible (Dunn et al. 2005). Towards this effort, the application of multiple techniques for the simultaneous measurement of the contained metabolites in a given sample is aiming to obtain a more complete view of the metabolite composition. Therefore, a growing number of studies are referring to the combination of different analytical techniques. Most commonly NMR and LC-MS approaches combining their unique advantages are applied successfully as complementary tools (Dunn 2008).

During this study, samples belonging to *Acronychia* genus were analysed using NMR and MS metabolomics platforms in parallel to provide a greater coverage of the metabolome. As aforementioned, plant metabolomics analysis involves a number of successive steps that are strongly related to the applied analytical platform. A unified scheme in Figure 30 is illustrating the common and discrete procedures that were followed during this analysis. For the sake of coherence, the sample preparation for analysis, acquisition and preprocessing procedures will be presented in accordance to the analytical platform used.

Chapter 3: Metabolomics approaches in *Acronychia* genus

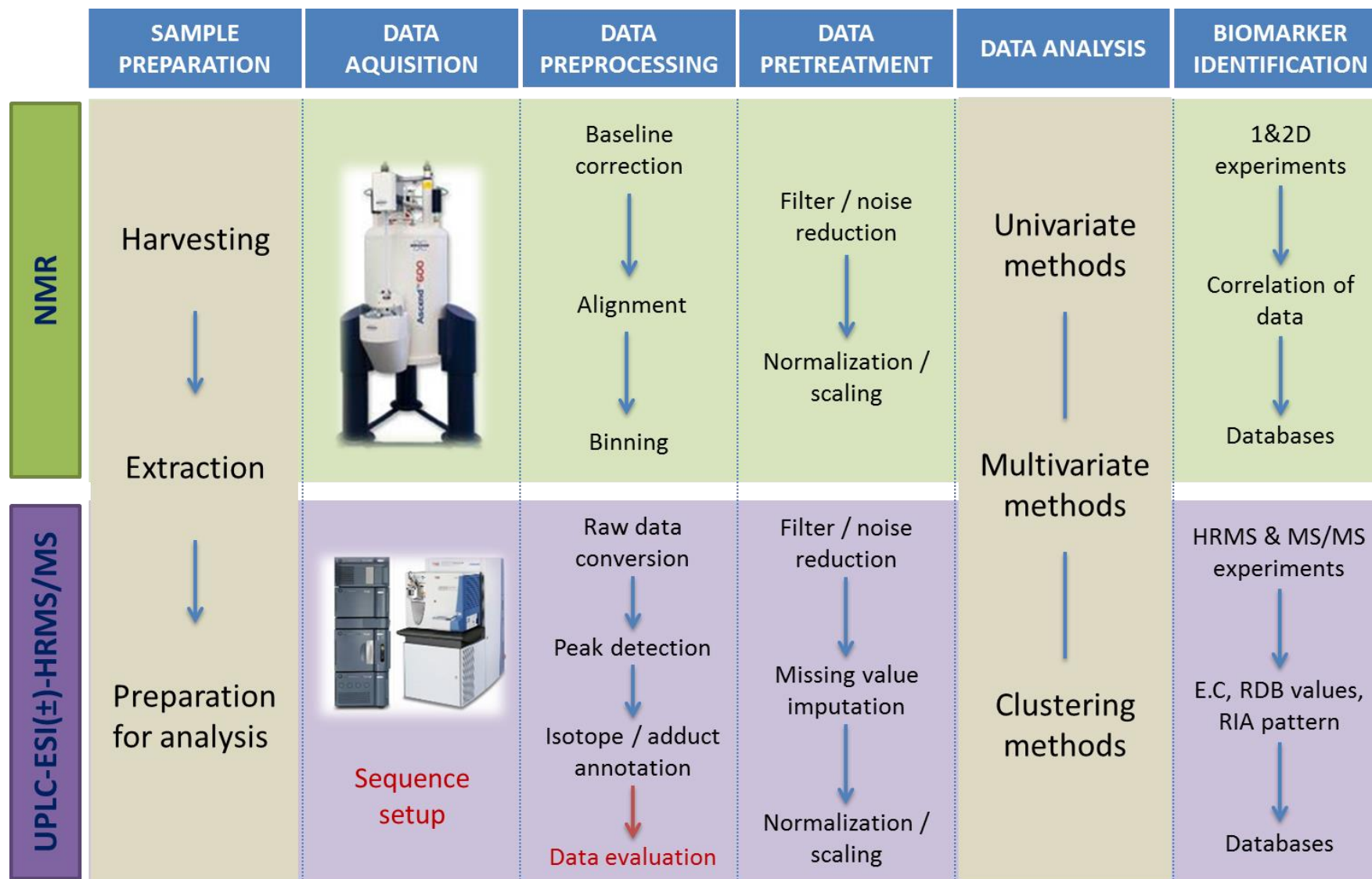


Figure 30: Schematic illustration of basic steps followed in NMR and LC-MS based metabolomics platforms

1.1. Sample selection and preparation

During this study, different *Acronychia* samples (Table XI) were selected belonging to diverse species and plant parts and collected from different locations as representative samples for the investigation of metabolite changes according to aforementioned conditions. Moreover, adequate biological replicates were accounted in order to ensure future statistical confidence. The selection of the different populations was performed considering previous literature data concerning the phytochemical investigation of *Acronychia* species, taxonomic issues and the traditional use of the plant material. Therefore, the specific dataset has been considered as a representative dataset to explore several biological questions concerning *Acronychia* species and also as a proof of concept for the development of particular methodologies related to plant metabolomics. As a critical issue in natural product chemistry, the correlation of the pharmacological activity with specific metabolites in plant extracts constituted a principal goal of this study.

After the selection of the plant material, a key point in the overall metabolomics analysis process was the determination of an extraction protocol. The parameters taken into account were from the one side the polarity of contained metabolites according to the literature data as well as the compatibility of the resulting extracts with the scheduled experiments. To allow a comparative study of the data obtained from the different analytical platforms and after pharmacological evaluation a 'one pot' extraction scheme was applied. Moreover, due to the intention to focus on the secondary metabolites of *Acronychia*, an initial extraction of all samples was performed using EtOAc and MeOH. The evaluation of the extraction efficiency by different solvents in terms of containing metabolites and reproducibility was assessed by ^1H NMR (Halabalaki et al. 2014). Both solvents resulted to reproducible extraction of metabolites from the different samples. In case of MeOH extraction, predominance of sugars, detected from the crowded signals at approximately δ_{H} 3.0- 4.0, discouraged the utilization of MeOH as extraction solvent. On the other hand, EtOAc solvent led to successful extraction of secondary metabolites, avoiding the extraction of primary metabolites (e.g sugars). Therefore, EtOAc was chosen for preparing the 120 extracts which were further aliquoted for NMR, LC-MS analysis and pharmacological evaluation (Figure 29).

1.2. NMR metabolomics platform

1.2.1. Sample preparation for NMR acquisition Before the acquisition of *Acronychia* extracts by NMR several deuterated solvents were tested in order to select the optimal. The main concerns were the investigation of the extracts' solubility and the effect of the deuterated solvent on the obtained metabolomic data. Under this prism, common deuterated solvents including Acetone- d_6 , C_5D_6 , $CDCl_3$, DMSO, MeOD and pyridine- d_5 were tested (Figure 31). As far as the solubility is concerned, the dilution of *Acronychia* extracts by Acetone- d_6 , C_5D_6 , $CDCl_3$ and DMSO resulted in transparent solutions while dilution by both MeOH and pyridine- d_5 led to precipitate formation. Moreover, the impact of the different deuterated solvents on metabolomics data was evaluated by the ^1H NMR spectra obtained after the dilution with the aforementioned solvents. Among the solvent tried in this study, $CDCl_3$ showed more promising results in terms of solubility, resolution and variety of signals. Therefore all 120 samples were prepared in $CDCl_3$ at a concentration of 10 mg/mL and acquired using ^1H NMR, 2D J-resolved and ^{13}C NMR experiments.

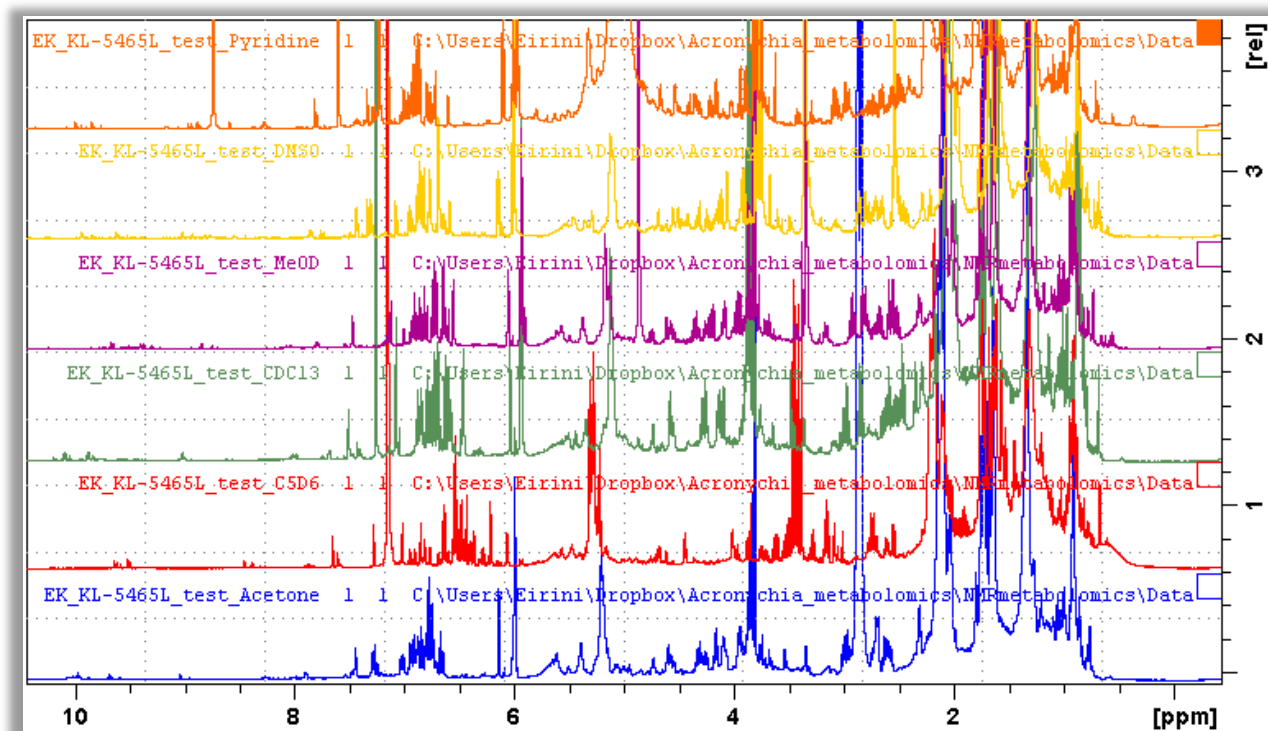


Figure 31: ^1H NMR data obtained from KL-5465L sample using different deuterated solvent during selection procedure; acetone- d_6 (blue); C_5D_6 (red); $CDCl_3$ (green); MeOD (purple); DMSO (yellow); pyridine- d_5 (orange)

1.2.2. NMR acquisition

The most commonly applied practice for NMR based metabolomics studies is the analysis of the ^1H NMR data (Kim et al. 2011). ^1H NMR methods offers the advantage of fast sample acquisition times and hence high throughput (Kim et al. 2010). Nevertheless, the effectiveness of this method is strongly dependent on the nature of extracts under study. Sometimes, as in our case, overlapping of ^1H NMR signals leads to incomprehensive peak resolution and thus hinders the robust metabolite identification. In such cases, the application of two-dimensional J -resolved (JRES) spectroscopy providing proton-decoupled projected 1D spectra (pJRES) is strongly proposed (Viant 2003).

As a matter of fact, visual inspection and comparison of the ^1H NMR spectra of the different *Acronychia* samples revealed many congested regions. In particular the overlapping in the aliphatic region resulted in a very asymmetric baseline which was difficult to handle. Therefore, projection of the 2D J -resolved spectra was performed resulting in decrease of complexity of the dataset (Figure A 73). Since JRES spectrum separates the chemical shift and spin–spin coupling data onto different axes, all protons appear as a singlet in the JRES spectra and after the projection on the axis of chemical shift many multiple peaks are converted to single peaks. For instance, Figure 32 illustrates the great improvement of the complexity of the spectra and the notable enhancement of the resolution at the aliphatic region which presented major overlapping issues. Likewise, a lot of signals were clearly resolved in the pJRES spectra.

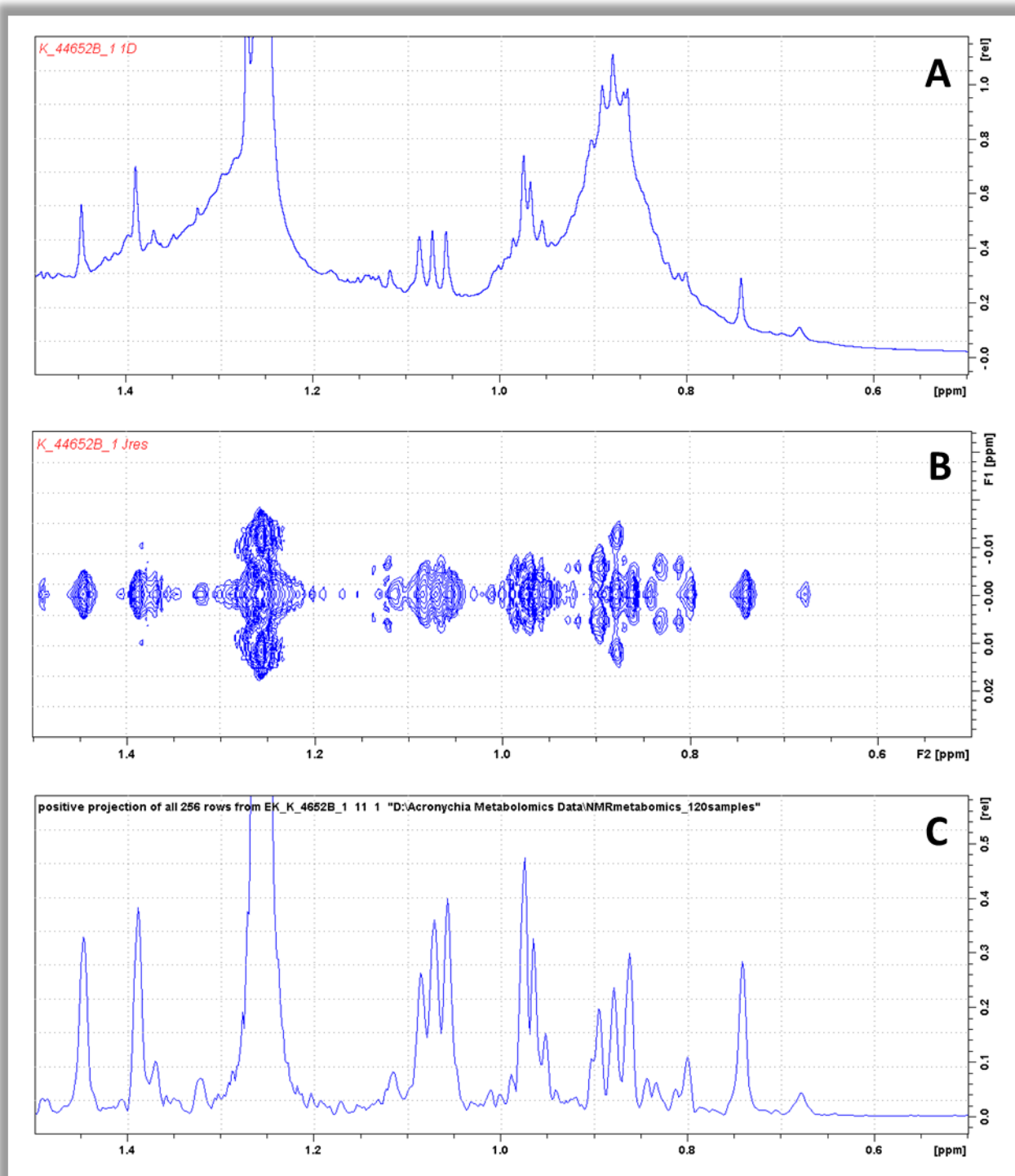


Figure 32: Example of A. ^1H NMR, B. 2D JRES and C. pJRES spectra of an *A. laurifolia* bark sample in the region of 0.5– 1.5 ppm

1.2.3. NMR data preprocessing/ pretreatment

Thereafter, a preprocessing step was performed to transform the metabolite profiles of *Acronychia* samples obtained by NMR (^1H NMR and pJRES) into data matrices in order to be used for further data analysis. For this step, both manual and automated processing of the NMR spectra was performed using TopSpin and Amix software for the first case and Metabolab software for the latter case (Figure 33). The incorporated algorithms were used for baseline correction, calibration to the internal standard and binning of equidistant size (0.01 and 0.04). The bins corresponding to noise were filtered out in order to enhance the statistical power in data analysis step. Both approaches led to similar results indicating a similar impact of the different algorithms in this particular dataset.

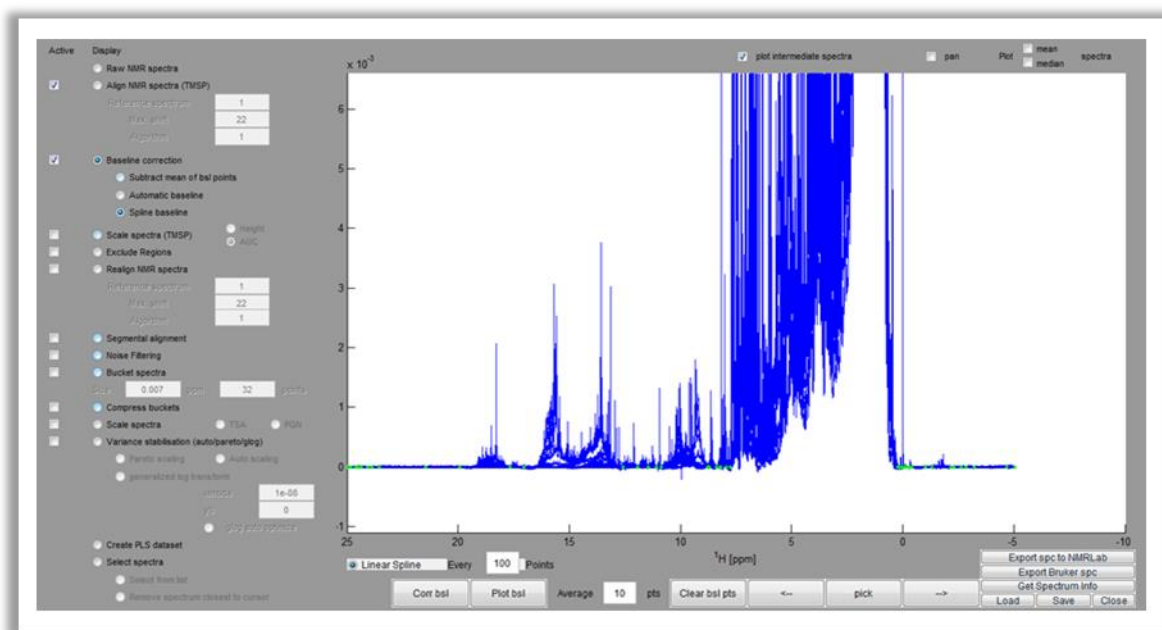


Figure 33: Snapshot from baseline correction step of the 120 ^1H NMR *Acronychia* samples using spline baseline algorithm incorporated in Metabolab software

1.2.4. NMR dataset evaluation

An initial evaluation of the NMR datasets in order to assess the reproducibility of the overall downstream analysis and the presence of possible outliers was performed before their utilization for specific applications. For this purpose, unsupervised principal component analysis (PCA) was applied for both ^1H NMR and pJRES datasets using Metaboanalyst web platform.

Chapter 3: Metabolomics approaches in *Acronychia* genus

For instance, for the evaluation of the ^1H NMR dataset, the data matrix was uploaded in Metaboanalyst platform in the appropriate format (.csv). A row wise normalization was applied by sum in which the total spectral area is assumed to be constant. Moreover, a pareto scaling was performed to make the features more comparable. By default in metaboanalyst also centring of the data is performed. Therefore the features are expressed as fluctuations round zero allowing the adjustment of low and high abundant metabolites. A characteristic figure generated during the normalization procedure illustrates some representative features before and after normalization permitting the evaluation of the normalization procedure (Figure 34).

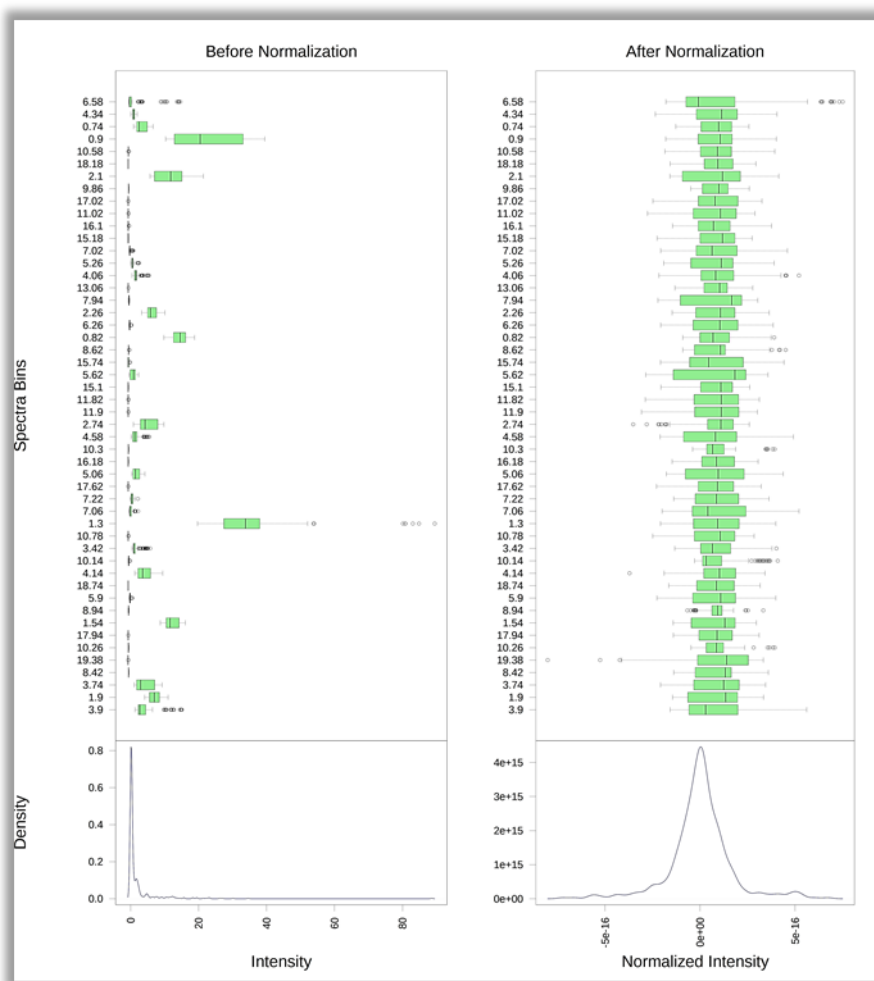


Figure 34: Data normalization view; the graph summarizes the distribution of input data values before and after normalization; the box plots on the top show the concentration distributions of individual compounds, whereas the bottom plots show the overall concentration distribution based on kernel density estimation

Thereafter, the data were analysed by principal component analysis (PCA). An eight PC model was calculated accounting for the 95.1% of the variance with PC1 explaining the

63.6% and PC2 the 17.7% of the variance. In PCA score plot (PC1 vs PC2) the samples of the same groups, coloured according to the different species and organ, were clustered well together giving an important level of confidence for the quality of the dataset (Figure 35A). Only one sample (K4853B_4) belonging to *A. laurifolia* barks group is observed out of the 95% confidence interval ellipse leading to its characterization as outlier. Moreover, the respective PCA loading plot (PC1 vs PC2) gives an indication of the features that are responsible for this classification (Figure 35B).

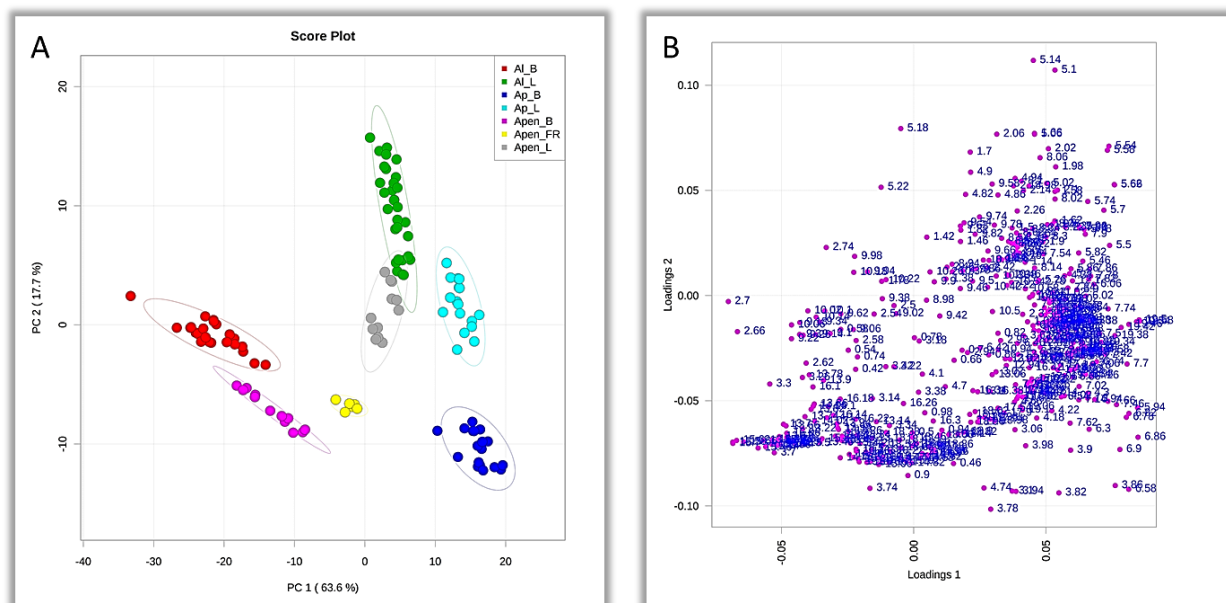


Figure 35: A. PCA score plot (PC1 vs PC2) of *Acronychia* samples obtained by the ^1H NMR dataset; *A. laurifolia* barks are coloured in red, *A. laurifolia* leaves in green, *A. porteri* barks in blue, *A. porteri* leaves in turquoise, *A. pedunculata* barks in pink, *A. pedunculata* fruits in yellow and *A. pedunculata* leaves in grey; B. PCA loading plot (PC1 vs PC2) indicating the features (bins) responsible for the classification

1.3. UPLC-HRMS metabolomics platform

1.3.1. UPLC-HRMS data acquisition

The hyphenation of two modern analytical techniques during this study offered great advantages for the analysis of *Acronychia* samples. On the one hand Acquity UPLC provides high chromatographic resolution, which results in shorter run times, narrow peak widths and an increased S/N compared to conventional HPLC. This is advantageous in metabolic profiling as a large amount of samples can be analysed rapidly allowing the detection of a greater number of metabolites. On the other hand, Orbitrap analyser

provides high mass resolution and mass accuracy over a wide dynamic range allowing the detection of metabolites with high level of confidence.

The development of the UPLC method was based on the nature of the extracts as assessed using the QC sample. The nonpolar nature of the EtOAc extracts implied the use of a reversed phase chromatography providing efficient separation of nonpolar compounds. Usually, in metabolomics applications acetonitrile (ACN) or methanol (MeOH) are applied as organic solvents of the mobile phase, therefore both solvents were tested to evaluate the chromatographic profile of the QC sample. The peak shape, peak resolution and reproducibility of the chromatographic profiles obtained by a mobile phase containing MeOH as organic solvent were significantly better comparing to the respective ones obtained by ACN. This was probably attributed to the reduced sensitivity at high ACN concentrations. For all aforementioned reasons, a mobile phase using MeOH as organic solvent and a gradient with a high percentage of MeOH were used for the acquisition of *Acronychia* extract samples.

The acquisition was performed in ESI(+) and ESI(-) after the adjustment of the conditions in both modes. The collection of both datasets is of significant importance as additional information can be obtained for the different metabolites contained in the mixtures enhancing the level of confidence in metabolite identification step.

1.3.2. UPLC-HRMS run sequence

The run sequence in LC-MS metabolomics studies has to be designed properly in order to ensure reliability of the acquired data. The first consideration is the stability of the system. In order to achieve system stability, a number of 'conditioning' samples have to be injected under the same conditions before the samples that will be subjected to analysis with a view to condition the column and ensure stable retention times and signal intensities. During this study, four blank samples were injected at the beginning of the sequence and acquired with the same method for system stability. Another important issue is the run order. During the whole run time period the MS response changes gradually as far as the LC and MS compartments are concerned. On the one hand, the aging of the column over run time leads to retention time drifts and peak broadening phenomena. On the other hand, regarding the MS instrument, the interaction of the sample with the MS source results in contamination phenomena and reduced sensitivity over the time. As presumed, this slight

change will lead to significant differences between the first and the last samples that will subsequently affect the result of the analysis. Therefore, a randomization of the different sample groups is required to ensure an equal impact on the degradation of performance over time. In addition, a common practice to enable the assessment of the data quality is the acquisition of pooled QC samples in between the extract samples in order to observe post acquisition the drifts of the LC and MS performance over time. In practice, QC samples were placed in between 5 extract samples and information regarding the retention time shift, intensity variation and mass accuracy variation was extracted (see Results and Discussion 1.3.4 for further details). Finally, blank samples were added in between 3 samples to avoid possible contamination from sample to sample. Summarizing the above information a typical run sequence list was constructed as shown in Figure 36.

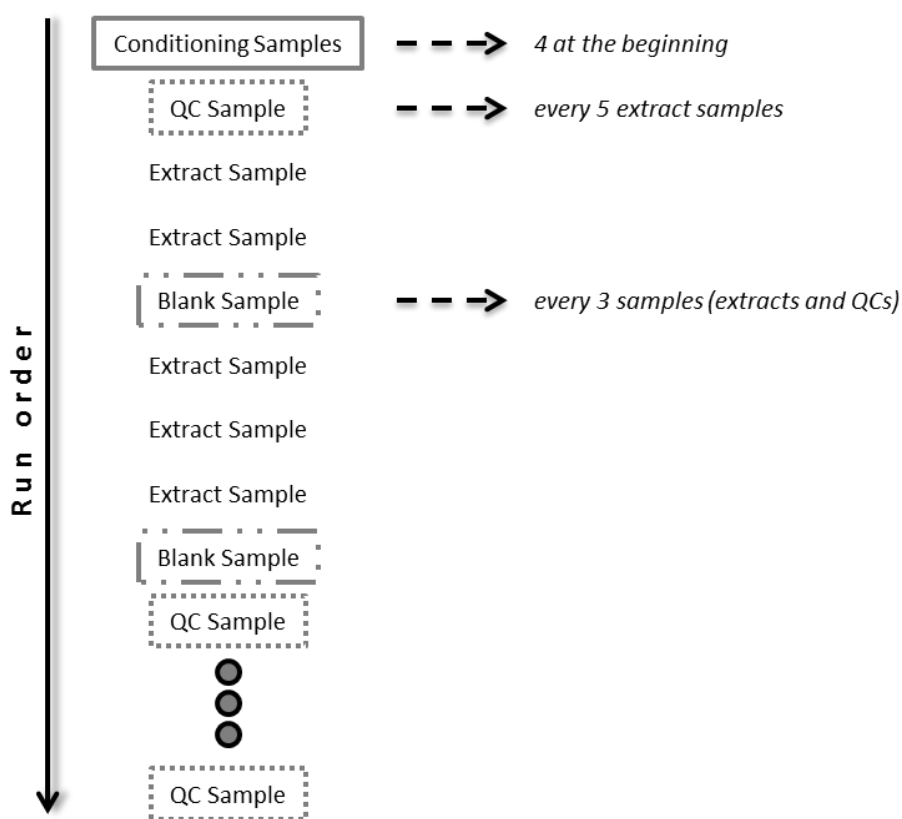


Figure 36: Representation of the concept for the construction of an LC-MS sequence run

1.3.3. UPLC-HRMS data preprocessing

In order to translate the raw data into a compatible format for data analysis, a preprocessing step was required. Among the plethora of software designed for the preprocessing of LC-MS data, the XCMS package implemented in R language was chosen. The

operation of the XCMS software is rather complex but flexible since a number of parameters may be tuned by the user. The peak detection in XCMS is performed by default using the matched filter algorithm. Matched filter algorithm has been proven to detect peaks in a reliable and robust way and to be able to discard noise and detect peaks close to noise level (Smith et al. 2006). The XCMS detection process can be split into three basic steps: peak picking, peak grouping and retention time correction followed by filling missed data.

Particularly, the matched filter algorithm divides the data into slices of mass width (defined according to the instrument used for the acquisition) and then then operates on those individual slices in the chromatographic time domain. Each slice can be represented as an extracted ion chromatogram. Then each slice is filtered with matched filtration resulting in the generation of a new chromatographic profiles accomplishing implicit background subtraction. After filtration, peaks are selected using a signal-to-noise ratio cutoff. Finally, peaks are characterized by integrating the unfiltered chromatogram between the zero-crossing points of the filtered chromatogram (Smith et al. 2006). Consequently, in the pick picking step the essential parameters to provide are the peak width (*step*) which depends on the instrument used for the acquisition, the signal to noise (*snthresh*) which depends on the background noise of the individual experiment and the minimum difference in m/z for peaks with overlapping retention times (*mzdiff*) which depends on the mass accuracy of the instrument. In this particular dataset, since the acquisition was performed using LTQ-Orbitrap apparatus providing a resolution of 30000, a peak width of 0.02 m/z mass unit was selected. By default this setting is 0.1 m/z applicable for low resolution instruments. Moreover, the mass accuracy using this apparatus in full scan mode is estimated <3 ppm for masses of 200- 1000 m/z range, thus, a minimum difference in m/z of 0.005 was applied. The signal to noise ratio was set at 3 in order to discard the background noise.

The next step of XCMS preprocessing process after peak peaking is the grouping of the peaks representing the same analyte across samples. The peaks initially are grouped according to their mass and then a kernel density estimator calculates the overall distributions of peaks in chromatographic time. Moreover, a simultaneous retention time correction for all peaks in a single step is performed. At this stage, the important parameter to take into account is the band width of peak groups (*bw*) which is dependent on the chromatographic peak width of each study. The average peak width of chromatographic peaks generated after acquisition was around 6-10 s, thus a *bw* of 10 s was selected. Finally,

the filling of missed data is performed by rereading the raw data and integrating them in the regions of the missing peaks. Thus, the peaks that were missed during the previous steps will be filled and will be distinguished from the ones that miss due to absence in the samples.

In order to determine the isotope and adduct peaks CAMERA algorithm was utilized. Briefly, this algorithm use the feature list generated from XCMS procedure and groups the features according to the retention time. Then the presence of isotope and adduct ions are assessed in these groups simplifying the following metabolite identification process.

1.3.4. UPLC-HRMS data evaluation

As mentioned above the reproducibility of LC-MS based metabolomics data is an issue that has to be taken into account. In such studies, variability in the analytical methods can be originated from different sources. However, repeatable data are essential in order to proceed with the statistical analysis and extract reliable results. Therefore, commonly the assessment of metabolomics data quality prior to any analysis of the data is of great importance. In this context, a careful creation of a run sequence ([see Results and Discussion 1.3.2](#) for further details) and examination of the data derived from pooled QC samples acquired regularly throughout the whole run could provide a good overview of the LC-MS variability (De Vos et al. 2007, Want et al. 2010). In our case, a sequence of 197 sample injections was performed including 25 QCs which were used to calculate the LC and MS performance over the 98.5 hour of continuous analysis.

In order to evaluate the performance over time, initially, the extracted ion chromatograms (XICs) of 12 randomly selected ions from the 25 QCs, with a wide range in elution time and molecular weights, were utilized. The first parameter to evaluate was the LC performance and specifically, the retention time variation. This was calculated in second deviation from the mean retention time of each ion from the 25 QCs ([Figure 37a](#)). The maximum deviation in retention time for all selected peaks was 3 s for scarce cases while for the majority of them was 2s. Moreover, the MS performance was evaluated investigating the peak intensities and the mass accuracy of these 12 randomly selected ions. In particular, the peak intensity variation, indicating the drift of sensitivity over time, was calculated as the percentage of deviation from the mean intensity ([Figure 37b](#)). The majority of the peaks presented a deviation smaller than 20% while in some scarce cases this reached the 50%.

Chapter 3: Metabolomics approaches in *Acronychia* genus

Finally, the mass accuracy was assessed during the entire run period by calculating the mass error in parts per million (ppm) of the 12 selected peaks in all 25 QCs. The results are depicted in **Figure 37c** indicating a high accuracy (< 1.5 ppm) throughout the run sequence.

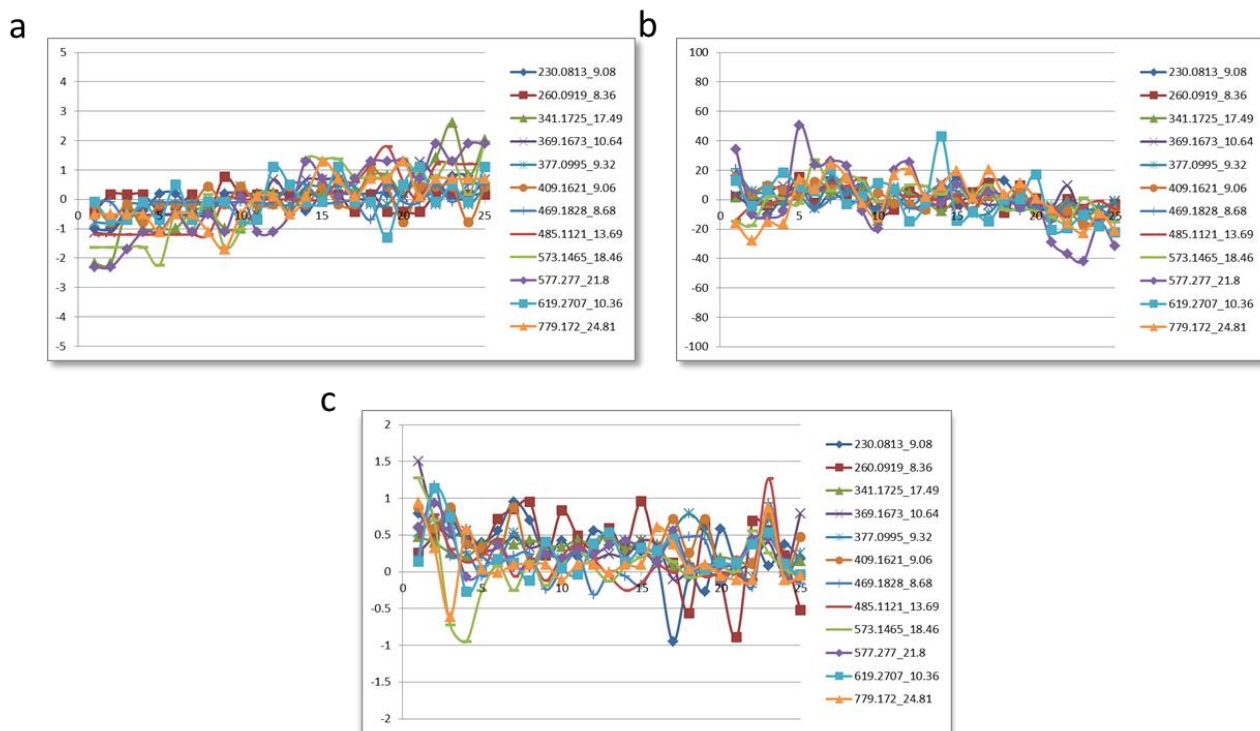


Figure 37: Stability of the UPLC-LTO-Orbitrap system over the 98.5 hours of continuous spectral acquisition; a. retention time variation b. signal intensity variation c. mass accuracy variation of 12 randomly selected peaks from the 25 QCs

The above mentioned findings indicate a satisfactory performance of the UPLC-LTO-Orbitrap instrument during the entire run period. After confirming the stability of the system, the evaluation of the metabolomics data was performed. For this purpose the preprocessed data were used to observe the clustering of all samples using PCA analysis and specifically to detect the clustering pattern of QC samples. Ideally, if the analytical variations were totally absent the QC samples would be identical. In practice, this is quite impossible as there are many parameters that render repeatability a challenging issue for LC-MS analysis. In **Figure 38** the 3D PCA scores plot (including the three first components) of the preprocessed data analysed in ESI positive mode using normalization by mean value and pareto scaling in Metaboanalyst web-based platform is displayed. All QC samples were clustered tightly giving an indication for the reliability of the data. This tight classification

suggests that the differences between the test samples from the different extracts are likely to reveal difference in metabolite profiles rather than analytical variations.

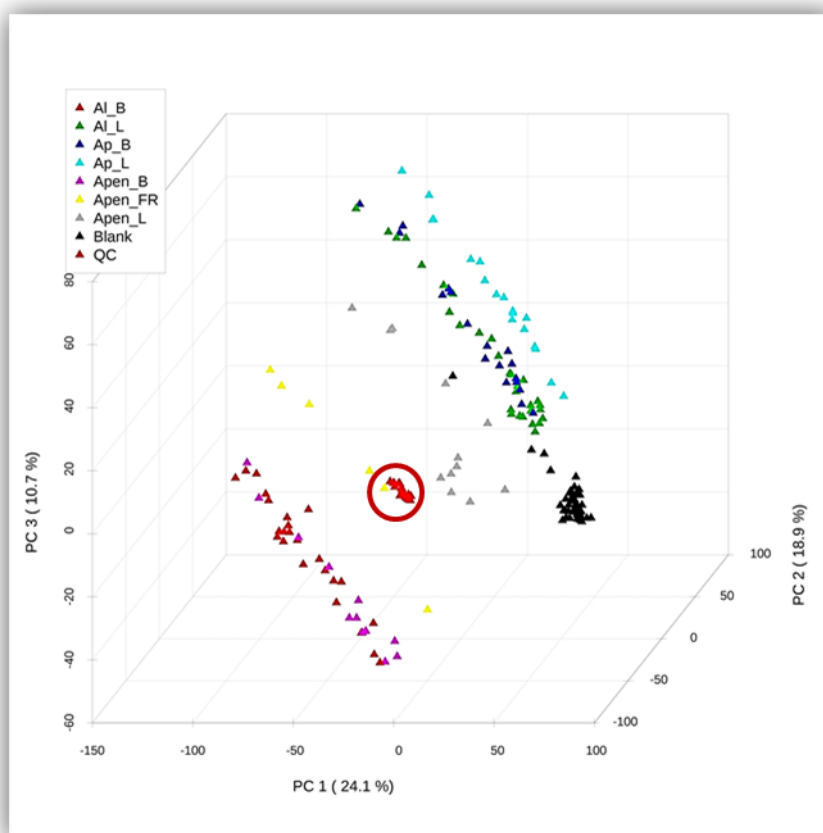


Figure 38: 3D PCA scores plot (PC1 vs. PC2 vs. PC3) of all samples analyzed in positive mode; QC samples indicated in red circle

Although this clustering gives a first idea about the reproducibility of the LC-MS analysis and the reliability of the data, in some cases may be misleading. These are the cases when the first components are representing a large proportion of the variability. Hence, the large biological variability between samples may tend the QC samples to cluster together masking possible variations between them due to analytical issues. Therefore, another way to assess the quality of the data is the investigation of the time dependency of the PCA components. As the PCA scores represent weighted average trajectories of the original variables, the exploration of their time dependency gives an insight into the trends and drifts over time (Gika et al. 2007). **Figure 39** illustrates the time dependency of the PC1 as all the samples are represented in run order. All QC samples are observed in the 2 SD limits and minor variances are visible between the QC samples throughout the 98.5 hours run

giving a further confidence for the quality of the data. In addition, the same observation was performed for blank samples which were more frequent than the QC samples. Blank samples were also presented stable over the run period indicating absence of carry over. Overall, a more significant variability is observed between the extract samples confirming that the difference detected is attributed to biological variation and not to analytical variation.

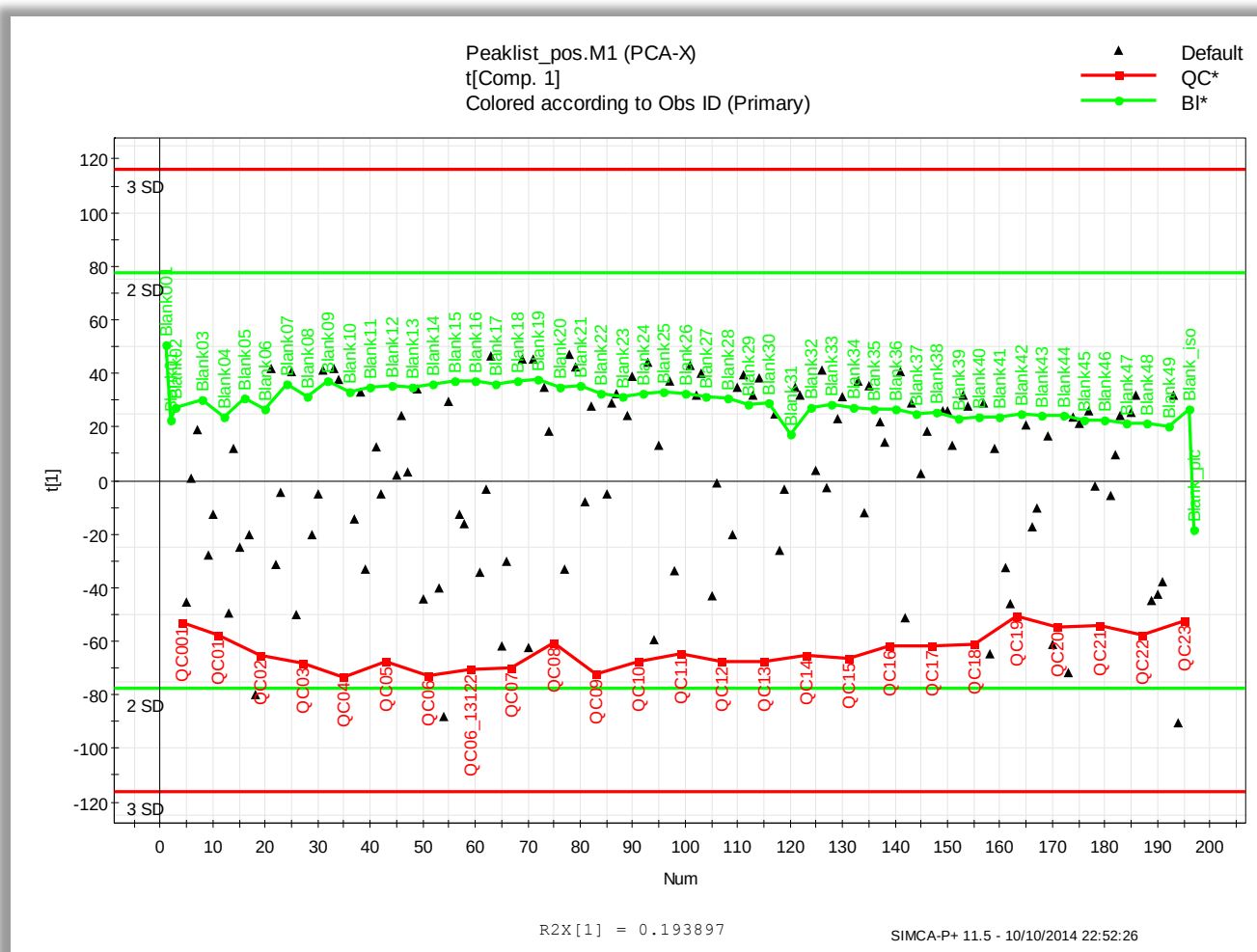


Figure 39: Control chart of PC1 versus samples in run sequence order for analysis in ESI(+); QC samples are represented by red dots, blank samples by green dots and extract samples by black triangles

After the assessment of the robustness of the UPLC-HRMS system, a more detailed investigation of the variability of the MS signal was performed in order to accept the individual ions as potential biomarkers. According to FDA criteria for bioanalytical methods, analytes do not have to exceed a coefficient of variation of 15% and of 20% for compounds close to the limit of quantification with the 33% of the QC values allowed to fall outside the

acceptance criteria (FDA 2001). However, for identification of potential biomarkers in urine samples an acceptance criterion of 30% of variation in the signal value for at least the 70% of the peaks is proposed by Want et al. (Want et al. 2010).

Considering the aforementioned criteria, the relative standard deviation (RSD%) of each peak from the peak list generated after the processing was calculated. In addition, the ions were separated according to the peak intensities in order to have an insight into the percentage of peaks of each subgroup and the number of them that meet the acceptance criteria. As illustrated in the [figure X](#), different acceptance criteria lead to significant percentage of peaks that fell into these limits. For instance, the majority of the peaks (60.1%) presented an intensity of 10^5 order of magnitude. The 63.6% of these peaks presented a RSD% smaller than 20% and the 79.9% of them a RSD% smaller than 30%. Same observations were performed also for the rest group of peaks with different intensities. Overall, the 60% of the total number of peaks met the acceptance criteria of a RSD < 20% while when the acceptance criteria were increased to RSD < 30% of the average value, a significant increase of 77% of the total peaks was calculated to fall into the limits. Hence, according to Want et al. a high level of confidence for the identification of biomarkers from the specific dataset was established.

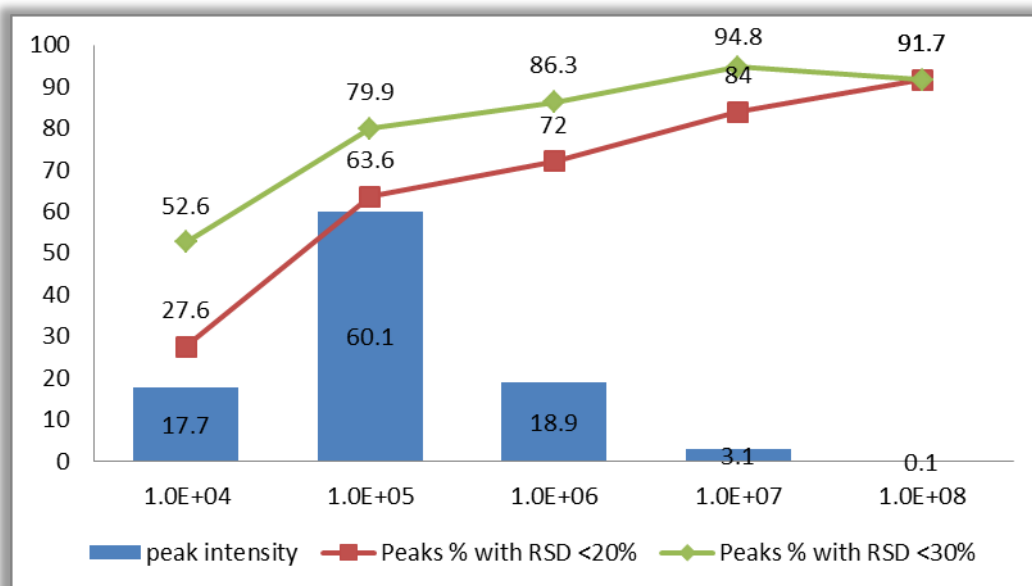


Figure 40: Distribution of peaks according to their intensities in ESI(+) generated after preprocessing and their percentage with RSD < 20% and RSD < 30%

Overall, the abovementioned findings demonstrate the great advantages from the hyphenation of UPLC Acquity chromatography with LTQ-Orbitrap mass spectrometry. Great reproducibility was achieved in terms of retention time, chromatographic peak area and mass accuracy over the 98.5 hours of operation giving a strong confidence for the quality of the dataset.

1.3.5. UPLC-HRMS data pretreatment

Despite the strong evidence obtained concerning the reliability of the dataset, the big number of the features generated after the preprocessing (8093) indicated the presence of 'false' features. Therefore, a filtering step for the reduction of the features was performed. This step is crucial as information, presumably valueless, is eliminating from the dataset. In the current study, initially, the elimination of peaks corresponding to background noise such as impurity and solvent peaks was performed by subtracting the common features present in both extract and blank samples. Moreover, peaks presenting RSD > 25% in the QC samples were removed assuming that they correspond to random features which are not related to any biological information. In [Figure 41](#) is clearly represented the importance of data filtering in *Acronychia* samples dataset using a PCA analysis importing both peak tables, before and after the filtering, in Metaboanalyst platform. PCA scores plot was used to illustrate the variance between the samples in the two first principal components and thus observe the grouping of the samples according to their biological differences. Obviously, in [Figure 41A](#) 'false' peaks complicate significantly the classification of the samples by adding an important level of variance which is not corresponding to any biological information. Interestingly, the elimination of the 'false' peaks ameliorates significantly the clustering of the biologically relevant samples.

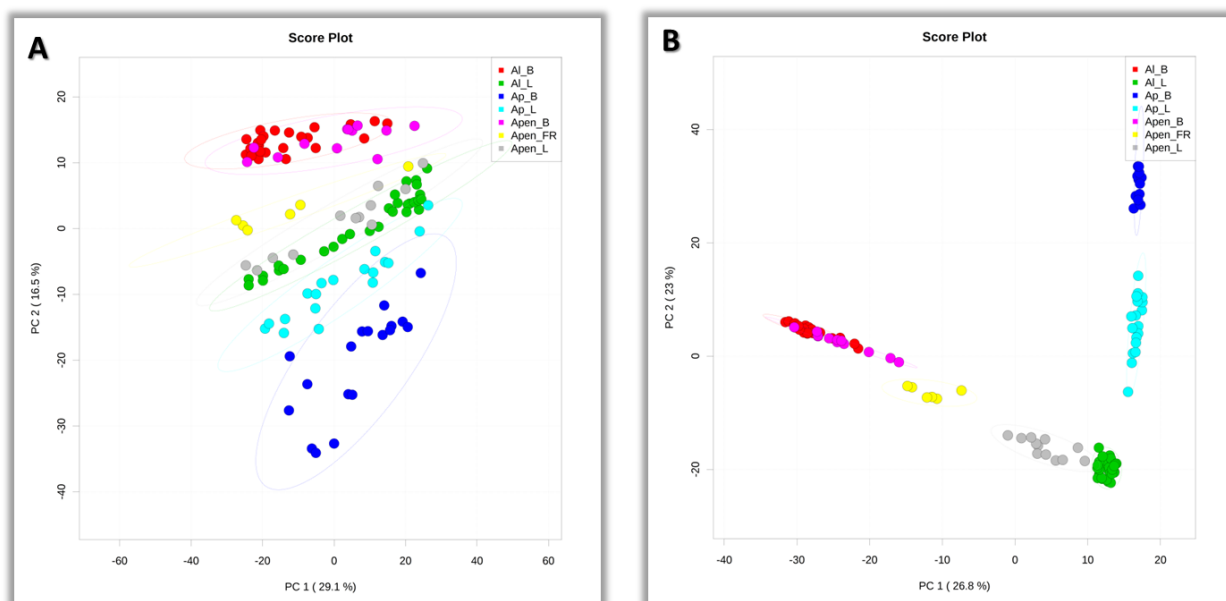


Figure 41: PCA score plots (PC1 vs PC2) of *Acronychia* samples using the peak table A. before data filtering B. after data filtering; *A. laurifolia* barks are coloured in red, *A. laurifolia* leaves in green, *A. porteri* barks in blue, *A. porteri* leaves in turquoise, *A. pedunculata* barks in pink, *A. pedunculata* fruits in yellow and *A. pedunculata* leaves in grey

Following a rationale to reduce further the features, filtering of the features that did not vary significantly in the different sample groups was performed. Therefore, p-values were calculated using ANOVA and features that presented p-values greater than 0.05 were discarded resulting to 4679 features. As expected, the removal of these features did not have a strong effect on the classification of the biologically different samples. However, this elimination was reflected in the PCA analysis as increase of the total variance account which reached the 51% instead of 49.8% that was accounted before this filtering step.

Another important consideration concerning data pretreatment step of LC-MS data is the handling of missing values which was performed using Metaboanalyst. By default Metaboanalyst replace missing values with very small values under the detection limit (the half of the minimum value detected in the data). In addition, Metaboanalyst gives the possibility to the user to exclude variables according to the percentage of the missing values presence or manually. Finally, the possibility of missing value imputation with diverse algorithms is also available (Xia et al. 2011). In our dataset, detailed inspection of missing values uncovered two types of missing values. In some cases, missing values were detected randomly mainly concerning small peaks close to noise level. Presumably these missing values were caused by technical issues such as ion suppression phenomena or

misalignment of the intensity signal during preprocessing. In other cases, missing values were detected consistently in specific biological groups indicating a biological reason of absence. Therefore, the replacement of the missing values in a consistent and automated way is necessary to convert the data into compatible matrices for further statistical analysis without eliminating the important biological information. In this context, it was considered useful to discard the variables with $> 85\%$ taking into account the number of samples defining a different biological group. The rest missing values were imputed using different imputation algorithms available in Metaboanalyst platform [Probabilistic PCA (PPCA), Bayesian PCA (BPCA) and Singular Value Decomposition Imputation (SVD Impute)] and compare the results by the impact of missing value imputation on PCA analysis. However, the PCA scores plots did not reveal any significant difference between the different imputation algorithms indicating an effective preprocessing result and arbitrarily the PPCA algorithm was chosen to continue with the downstream analysis (Figure A 74).

1.3.6. UPLC-HRMS dataset evaluation

Although unsupervised PCA was used extensively during the previous steps to evaluate the impact of the individual actions on the classification outcome, after the selection of the optimal parameters in previous stages, PCA was applied at the end of this procedure to evaluate the dataset and explore the presence of potential outliers. Using Metaboanalyst platform, row-wise and column-wise normalization of the dataset was performed to correct unpredictable dilution mistakes and to adjust the differences in fold differences between the different features, respectively. The effect of normalization is illustrated in Figure 42 representing the significant impact of normalized intensities for the comparison between the different samples.

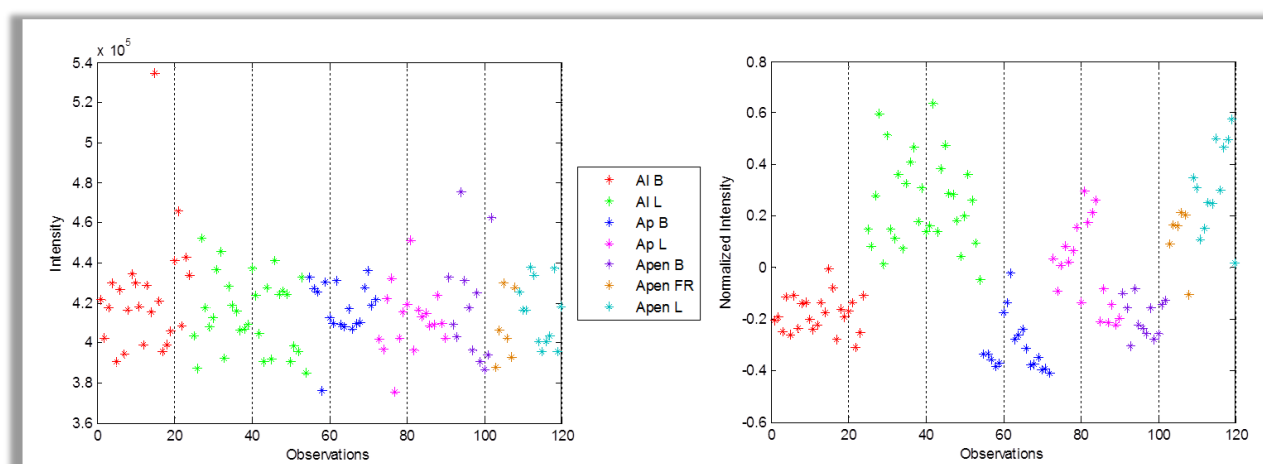


Figure 42: Plots of a randomly selected variable (feature 525.2866_6.47) obtained from the UPLC-ESI(+)-HRMS dataset of *Acronychia* samples as observed in the different samples before (left) and after (right) normalization using Metaboanalyst platform; *A. laurifolia* barks are coloured in red, *A. laurifolia* leaves in green, *A. porteri* barks in blue, *A. porteri* leaves in pink, *A. pedunculata* barks in purple, *A. pedunculata* fruits in yellow and *A. pedunculata* leaves in turquoise

Subsequently, a PCA model was calculated accounting 15 components and explaining the 91.8% of the dataset variance (Figure A 75). In particular, principal component 1 (PC1), 2 (PC2) and 3 (PC3) explained the 27.1%, 23.3% and 9.9% of variance, respectively. The evaluation of the PCA scores plot confirmed the good quality of the dataset since all samples belonging to diverse biological groups were tightly grouped together. Specifically, the reproducibility of the 6 analytical repetitions can be deduced from the overlapping of the scores belonging to the same group. Consequently, discrimination of the samples observed in PC1 and PC2 is highly attributed to inherent biological differences between the annotated groups (Figure 43). Examination of the loading plot may reveal the biomarkers responsible for this classification (see Results and Discussion 3.2 for further details).

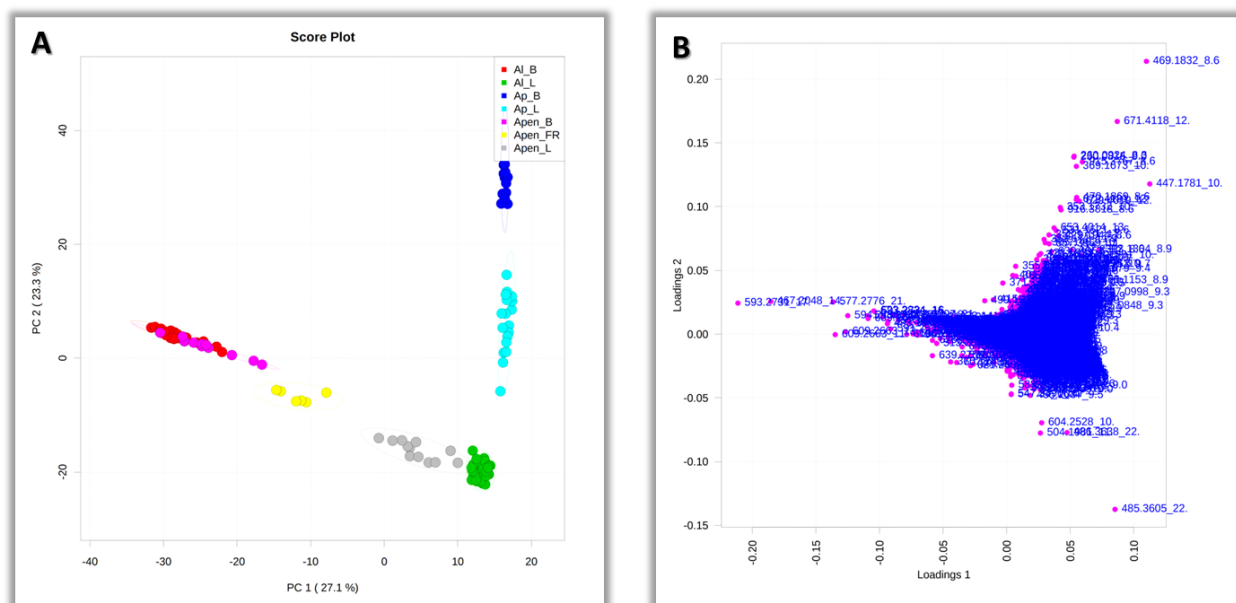


Figure 43: A. PCA score plot (PC1 vs PC2) of *Acronychia* samples obtained by the UPLC-ESI(+)-HRMS dataset; *A. laurifolia* barks are coloured in red, *A. laurifolia* leaves in green, *A. porteri* barks in blue, *A. porteri* leaves in turquoise, *A. pedunculata* barks in pink, *A. pedunculata* fruits in yellow and *A. pedunculata* leaves in grey; B. PCA loading plot (PC1 vs PC2) indicating the features (m/z , rt) responsible for the classification

2. UPLC-HRMS based metabolite identification and dereplication in different *Acronychia* extracts

Metabolomic or metabolic profiling of plant extracts using modern UPLC-HRMS instrumentation is a very automated and high through put process to get an insight at the metabolite composition of different plant extracts (Hall 2006). Nevertheless, one of the main bottlenecks of this procedure is the identification of the known metabolites as well as the chemical assignment of unknown metabolites (Nakabayashi et al. 2013, Wolfender et al. 2013). In plant metabolomics field, this is more challenging comparing to metabolomics studies using mammalian fluids or bacterial material. This is due to the extremely larger number of plant secondary metabolites that occur in nature (approximately 200.000 (Bino et al. 2004)) comparing to the metabolites present in mammalian biological fluids (approximately 20.000) and bacterial material (approximately 60.000). In addition, a lot of plant secondary metabolites are occurring exclusively in specific genus or even species. Beside the fact that a number of freely available databases are existing e.g. METLIN (<http://metlin.scripps.edu/>), PubChem (<http://pubchem.ncbi.nlm.nih.gov/>), ChemSpider (<http://chemspider.com/>), the creation of a universal database containing all plant secondary metabolites has not yet been accomplished rendering the identification of plant secondary metabolites a labour-intensive step.

During the last years, a number of commonly used tools and strategies have been established in order to define the metabolite identity based on LC-MS data (Kind et al. 2010, Kueger et al. 2012). The main spectrometric characteristic that can be obtained from this analysis is the molecular weight. However, this information cannot lead directly to the structure of a metabolite as a large number of compounds possess the same molecular weight indicating the inadequate structure information obtained by MS comparing to NMR. The use of high resolution mass spectrometers is of great importance in order to determine the molecular weight with high accuracy (< 3 ppm). In addition, the combination of accurate mass measurements with relative isotope abundance accuracy is utilized for the prediction of the elemental composition of metabolites with high efficiency (Kind et al. 2006). Determining the molecular formula of a metabolite constitutes the basis for subsequent structure elucidation. Nevertheless, the number of possible metabolites is usually large and further filtering steps are applied to narrow down the candidate numbers (e.g. heuristic filtering (Kind et al. 2007)). Furthermore, mass spectrometers enabling the

generation of multi stage mass spectrometry experiments (ion traps, triple quadrupoles or hybrids) can be exploited to obtain further structural information. Based on the interpretation of MS/MS and MSⁿ spectra important structural information can be extracted. In particular, the implementation of high mass accuracy in the MSⁿ level may provide high quality data for metabolite identification (van der Hooft et al. 2012). Similarly to MS spectra, is highly unlikely that MS/MS and MSⁿ data of plant secondary metabolites are available in databases in order to confirm the identity of the metabolites. However, the investigation of the main losses in combination with the candidate structures often provides important information for the identity of the metabolites. In addition, software such as MassFrontier (HighChem) or free available on-line platforms such as MetFrag (<http://msbi.ipb-halle.de/MetFrag/>) may provide *in silico* calculations of the fragmentations of different metabolites. In order to improve the confidence of the metabolite annotation, other orthogonal parameters are important to be taken into consideration. Commonly, in LC-MS based metabolomics, the retention time is utilized as an essential parameter which gives an estimation of the hydrophobicity of the metabolites (Dunn 2008). This information may reduce further the number of candidate structures in case that they possess significantly different logP values (octanol-water partition coefficient).

In natural products field, due to the lack of a universal plant secondary metabolites MS database, the multiple physicochemical and spectrometric characteristics collected after the acquisition of LC-MS data are combined and compared usually to in-house databases and literature data in order to identify with confidence the known metabolites present in mixtures, also named dereplication process. Particularly, in the current study, predetermination of specific features corresponding to known metabolites was performed by matching these features to multiple databases and then a screening of the datasets for the occurrence of these metabolites was followed, characterizing this procedure as a bottom-up identification and dereplication strategy (van der Hooft et al. 2013).

In this context, a number of different species and organ samples were analysed by UPLC-HRMS platform. Among the species subjected to analysis, *A. laurifolia* and *A. pedunculata* have been extensively investigated by various groups resulting in a large number of literature data available concerning their chemical composition (de Silva et al. 1979, Rahmani et al. 1996, Cui et al. 1999, Pathmasiri et al. 2005, Kozaki et al. 2014). On the other hand, *A. porteri* is reported only once for the presence of three methoxyflavones (Lichius et al. 1994). LC-MS based metabolomics approaches constitute a suitable tool for

Chapter 3: Metabolomics approaches in *Acronychia* genus

the rapid, automated and simultaneous exploration of the chemical profile of plant extract samples. Therefore, based on the known compounds reported in the literature from all *Acronychia* species an in-house database was constructed. In particular, for each compound, structure and origin related data including physicochemical characteristics (logP), and spectrometric characteristics (accurate mass, MS/MS fragments) were collected from the literature and the dereplication of them was successfully performed in the different *Acronychia* samples (Table XIV, Figure 44).

The annotation of the metabolites in different *Acronychia* species and organs was performed exploiting their retention times and MS data and comparing them with the in-house generated database. The MS data, accurate mass ions and MS/MS fragment ions, were collected during the analysis in ESI positive and negative mode. The majority of the identified metabolites detected from their $[M+H]^+$ or $[M+Na]^+$ ions generated in ESI(+). In ESI(-) the majority of the known metabolites were not ionized, however, a part of acetophenones were additionally observed in ESI(-). Using this approach, a total number of 33 metabolites were detected, including AtA, acetophenone monomers, alkaloids, lignans, triterpenes and phenolics. For the identification of the metabolites, the consistency of the extracted ions after preprocessing was verified in terms of average mass and retention time by tracing the compounds in the raw data. In order to define the confidence of the identification metabolites were annotated according to MSI recommendations (Sumner et al. 2007). Thus, metabolites with annotation level 1 were compared to standard compounds available in-house while metabolites with annotation level 2 were putatively annotated comparing their physicochemical properties and spectral characteristics with compounds reported in the literature.

Table XIV: Annotated metabolites in *Acronychia* extracts based on literature data

ID	RT	Formula	Ion	Average Mass <i>m/z</i>	Error (ppm)	RDB	MS/MS				Identification	Elemental Composition	Annot. level
Alkaloids(10 compounds)													
1	7.46	C15H18NO3	[M+H] ⁺	260.1284	0.923	7.5	242.1176	188.0706	C15H16O2N	C11H10O2N	oligophyllicine	C15H17NO3	2
		C15H17O3NNa	[M+Na] ⁺	282.1102	0.480								
2	7.93	C12H11NO4Na	[M+Na] ⁺	256.0582	0.688	7.5					2,3-methelenedioxy 4,7-dimethoxy quinoline	C12H11NO4	2
3	8.36	C15H18NO2	[M+H] ⁺	244.1335	0.715	7.5	188.0707	202.0863	C11H10NO2	C12H12NO2	oligophyline	C15H17NO2	2
4	8.37	C14H14NO4	[M+H] ⁺	260.0918	0.137	8.5	245.0685	230.045	C13H11O4N	C13H9O3N	skimmianine maculosidine kokusaginine	C14H13NO4	2
		C14H13NO4Na	[M+Na] ⁺	282.0737	0.072								
5	8.87	C13H11NO3Na	[M+Na] ⁺	252.0633	0.697	8.5					pteleine γ-fagarine evolitrine	C13H11NO3	2
6	9.09	C13H12NO3	[M+H] ⁺	230.0812	0.218	8.5	215.0579			C12H9O3N	pteleine γ-fagarine evolitrine	C13H11NO3	2
7	9.15	C12H12NO4	[M+H] ⁺	234.0762	0.494	7.5					2,3-methelenedioxy 4,7-dimethoxy quinoline	C12H11NO4	2
8	9.9	C17H22NO4	[M+H] ⁺	304.1545	0.478	7.5	248.0924		C13H14NO4		preskimmianine	C17H21NO4	2
		C17H21NO5Na	[M+Na] ⁺	326.1364	0.309								
9	9.92	C15H17NO2Na	[M+Na] ⁺	266.1154	0.789	7.5					oligophyline	C15H17NO2	2
10	10.13	C17H15NO5Na	[M+Na] ⁺	336.0839	-1.945	10.5					melicopidine	C17H15NO5	2
Acetophenone monomers (11 compounds)													
11	8.69	C14H19O5	[M+H] ⁺	267.1228	0.261	5.5	249.1127		C14H17O4		acronyculatin D	C14H18O5	2

Chapter 3: Metabolomics approaches in *Acronychia* genus

		C14H18O5Na	[M+Na] ⁺	289.1048	0.537								
12	9.35	C13H16O4Na	[M+Na] ⁺	259.0942	0.617	5.5					6-demethylacronylin	C13H16O4	2
		C13H15O4	[M-H] ⁻	235.0977	0.458	6.5	191.1079		C12H15O2				
13	12.13	C19H26O5	[M+H] ⁺	335.1853	-0.091	6.5	279.1234		C15H19O5		acronyculatin B	C19H26O5	2
			[M-H] ⁻	333.1705	-0.742	7.5				acronyculatin C			2
14	14.14	C14H19O4	[M+H] ⁺	251.1280	0.774		195.0652		C10H11O4				
		C14H18O4Na	[M+Na] ⁺	273.1099	0.475	5.5					acronyline	C14H18O4	2
		C14H17O4	[M-H] ⁻	249.1133	0.071	6.5							
15	15.67	C19H25O4	[M+H] ⁺	317.1750	0.612						acronyculatin G/E	C19H24O4	2
		C19H24O4Na	[M+Na] ⁺	339.1567	0.176	7.5							
16	16.77	C15H19O5	[M+H] ⁺	279.1227	0.035	6.5					acronyculatin A	C15H18O5	2
		C15H17O5	[M-H] ⁻	277.1081	-0.097	7.5							
17	17.56	C19H27O4	[M+H] ⁺	319.1905	0.232		263.1279		C15H19O4		1-[2',4'-dihydroxy-3',5'-di-(3''-methylbut-2''-enyl)-6'-methoxy]phenylethanone	C19H26O4	2
		C19H25O4Na	[M+Na] ⁺	341.1725	0.643	6.5	285.1101		C15H18O4Na				
		C19H25O4	[M-H] ⁻	317.1755	-0.922	7.5							
18	19.45	C19H27O4	[M+H] ⁺	319.1906	0.606	6.5	263.1279		C15H19O4		1-[2',4'-dihydroxy-3',5'-di-(3''-methylbut-2''-enyl)-6'-methoxy]phenylethanone	C19H26O4	2
19	19.63	C19H25O4	[M+H] ⁺	317.1751	0.518	7.5					acronyculatin G/E	C19H24O4	2
20	21.8	C19H27O4	[M+H] ⁺	319.1906	0.606	6.5	263.1279		C15H19O4		1-[2',4'-dihydroxy-3',5'-di-(3''-methylbut-2''-enyl)-6'-methoxy]phenylethanone	C19H26O4	2
		C19H25O4	[M-H] ⁻	317.1755	-0.544	7.5							
Acronychia-type acetophenones (9 compounds)													
21	16.76	C32H42O9Na	[M+Na] ⁺	593.2719	-0.428	11.5	357.1676		C19H26O5Na		acrovestenol isomer	C32H42O9	1
		C32H41O9	[M-H] ⁻	569.2754	-0.216	12.5							
22	17.59	C32H43O9	[M+H] ⁺	571.2897	-0.751						acrofolione A	C32H42O9	1
		C32H42O9Na	[M+Na] ⁺	593.2719	-0.428	11.5	341.1729		C19H26O4Na				

Lignans (2 compounds)										
30	9.34	C20H18O6Na	[M+Na] ⁺	377.0996	-0.025	11.5		asarinin	C20H18O6	2
31	8.69	C24H30O8Na	[M+Na] ⁺	469.1832	-0.786	9.5		yangambin	C24H30O8	2
Triterpenes (1 compound)										
32	25.09	C30H50ONa	[M+Na] ⁺	449.3723	-0.929	5.5		b-amyrin	C30H50O	2
Phenolics (1 compound)										

Chapter 3: Metabolomics approaches in *Acronychia* genus

In a next step, in order to get an insight into the presence of these metabolites in the different *Acronychia* samples all the peaks generated after the preprocessing corresponding to the aforementioned metabolites were utilized. The visualization of the data was performed using heatmap analysis combined by hierarchical cluster analysis (HCA). Heatmap visualization allows the representation of the relative concentration of individual metabolites in different samples represented by a colour scale in a matrix. The combination with HCA is commonly used to facilitate comparative observations between the different sample groups and thus, pattern recognition. HCA is an unsupervised clustering method which seeks to build a hierarchy of clusters according to their descriptors. There is a big number of clustering algorithms available (Andreopoulos et al. 2009) however, in metabolomics field, mainly agglomerative techniques are used to produce a series of partitions of the data. The results are illustrated in dendrograms representing the divisions made at each successive stage of analysis while the distances between clusters are defined using selected metrics. Based on the annotated metabolites, and using Heatmap visualization combined with HCA (wards method, euclidian distance), dissimilarities were observed in terms of metabolite occurrence giving a general overview of their relative levels in the various samples (Figure 44).

In particular, alkaloids were mainly accumulated in *A. laurifolia* leaves (metabolites 3, 4, 5, 6, 7, 9) and *A. porteri* barks (metabolites 1, 2, 3, 4, 5, 6, 7, 8) while smaller quantities were observed in *A. pedunculata* leaves (metabolites 1, 4). Interestingly, alkaloids such as oligophylicine (1), oligophyline (3), preskimmianine (8) have been reported only in *A. oligophylebia* (Wen-Hao et al. 1984) and is the first time that they were detected in these species. However, these alkaloids are considered as precursors of furoquinoline alkaloids (Storer et al. 1972) already reported from these species.

Acetophenones (monomers and dimmers) were principally appeared in *A. laurifolia* barks and *A. pedunculata* barks which is in agreement with previous literature data (Su et al. 2003, Kouloura et al. 2012). It is worth noting that some differences were observed between the samples of *A. laurifolia* and *A. pedunculata* barks regarding the occurrence of different acetophenone derivatives. In particular, samples KL4727B (*A. laurifolia* bark) and VN0179L (*A. pedunculata* bark) were dominated by metabolites 14, 15, 18, 22, 24 and 29 while metabolites 10 and 11 were exclusively found in KL4727B. On the other hand, K4652B, K4853B, KL5197B (*A. laurifolia* bark) and VN0874B (*A. pedunculata* bark) were more rich in metabolites 16, 17, 19, 20, 21, 23, 25, 27 and 28. In *A. laurifolia* leaves, low

abundances of acetophenone monomers were detected while leaves and fruits of *A. pedunculata* metabolites 13, 15, 17, 18, 22, 24, 29 were detected in significant amounts. The aforementioned observations regarding acetophenone derivatives suggest that these compounds are only present in *A. laurifolia* and *A. pedunculata* species while total absence is detected in *A. porteri*. This remark is of crucial importance as in the literature AtA are proposed as chemotaxonomic markers of the genus (Adsersen et al. 2007).

Concerning lignan composition of the studied samples, lignans were detected principally in *A. porteri* species and particularly, asarinin (30) was exclusively identified in *A. porteri* leaves samples. Asarinin was isolated from *A. muelleri* leaves (Davenport et al. 1954) and it's the first time to be detected in *A. porteri*. The other annotated lignan, yangambin, was spotted in both *A. porteri* leaves and barks samples. Although yangambin (31) was isolated from *A. laurifolia* roots in our dataset was not detected in *A. laurifolia* samples (Cui et al. 1999). The unique triterpene detected in the sample set was b-amyrin (32). B-amyrin was found in significant amounts only in *A. laurifolia* leaves which is in accordance with the literature data (Govindachari et al. 1969). Finally, the 4-geranyloxy ferulic acid isolated from barks of *A. baueri* (Prager et al. 1966) was detected mainly in *A. laurifolia* leaves and *A. porteri* barks.

Overall, the application of UPLC-ESI(±)-HRMS metabolic profiling approach for the dereplication of known compounds in different *Acronychia* samples resulted in the phytochemical characterization of studied and unexplored *Acronychia* species. A number of different groups of compounds were detected giving a better insight into the chemical composition of the individual samples. Nevertheless, hypothesis that supported the occurrence of *Acronychia*-type acetophenones in all *Acronychia* species rendering them chemotaxonomic markers of the genus come under question due to their absence from *A. porteri* samples.

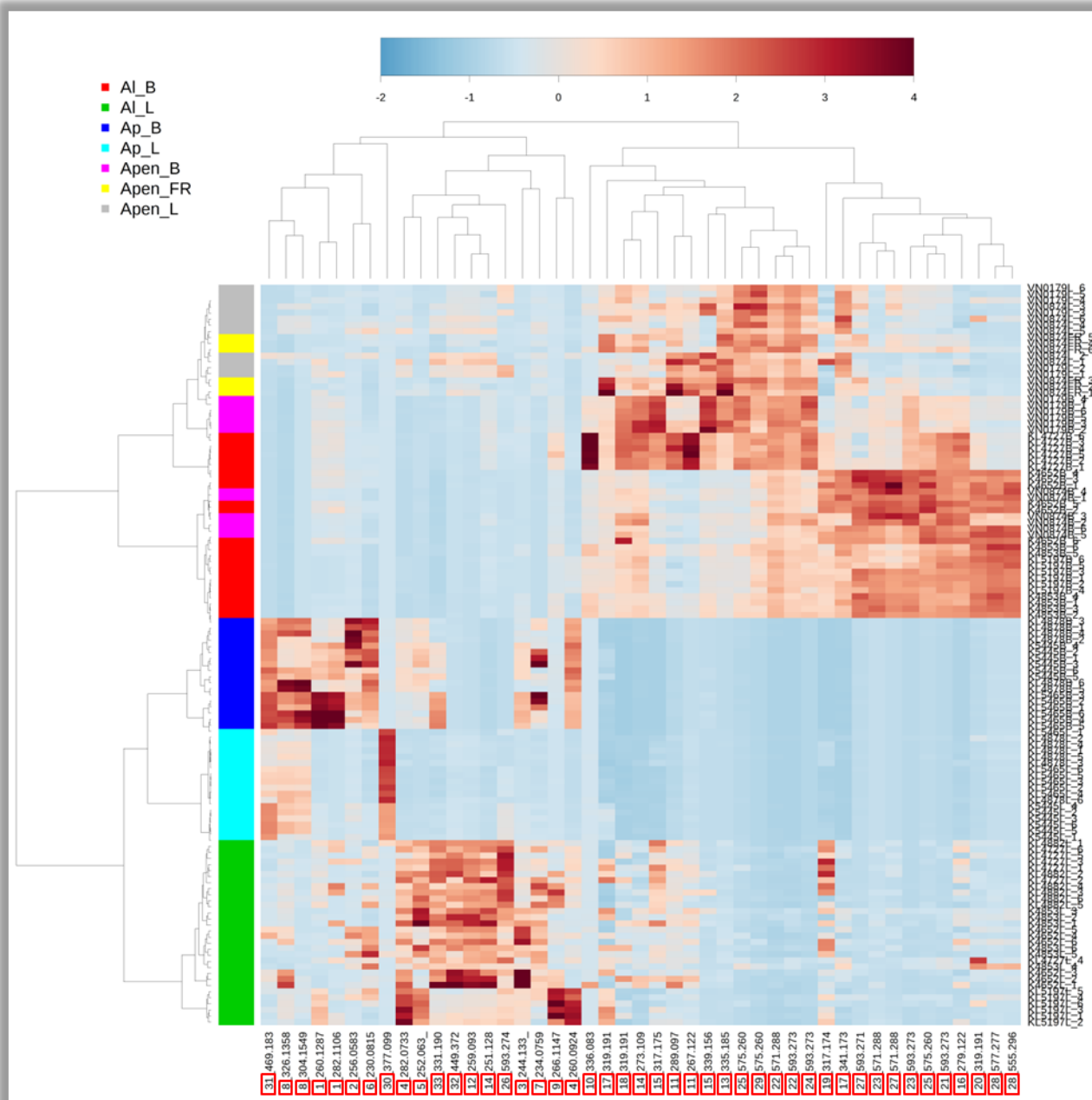


Figure 44: Heatmap visualization based on ANOVA combined with HCA of the annotated metabolites (numbers presented in red boxes are referred to the numbers of Table XIV) in the dataset of different *Acronychia* samples, particularly *A. laurifolia* barks are coloured in red, *A. laurifolia* leaves in green, *A. porteri* barks in blue, *A. porteri* leaves in turquoise, *A. pedunculata* barks in pink, *A. pedunculata* fruits in yellow and *A. pedunculata* leaves in grey

3. ^{13}C NMR based identification of secondary metabolites and dereplication in *Acronychia* extracts

NMR based metabolomics profiling constitutes a rapid, reproducible and non-destructive approach for the investigation of the metabolite composition of plant extracts (Kim et al. 2010). Moreover, NMR is considered as the most adaptable technique for unambiguous structure elucidation of natural products providing highly specific evidence for the identity of a metabolite (Robinette et al. 2011, Halabalaki et al. 2014). However, the identification and structural elucidation of metabolites in complex mixtures is not trivial. Limitations of this technique, such as low sensitivity and extensive signal overlapping, are circumventing the simultaneous robust identification of multiple metabolites in different concentration levels.

^1H NMR is the most popular technique for metabolite profiling of plant extracts (Schripsema 2010) due to the very high natural abundance of ^1H resulting in a higher sensitivity measurements comparing to NMR experiments based on less abundant nuclei (Dunn et al. 2005). Despite this advantage, ^1H NMR technique suffers from extensive signal overlapping hampering significantly the identification procedure. Each metabolite is represented by a number of signals which are spread out in a range of 0-10 ppm leading often in strong overlaps in the majority of spectral regions. Therefore, the application of separation techniques prior to spectral acquisition is usually proposed for dereplication purposes (Halabalaki et al. 2014).

Towards this direction, recently, a dereplication strategy based on ^{13}C NMR data is suggested in order to obtain reliable information concerning the metabolite composition of a mixture (Hubert et al. 2014). According to this strategy, ^{13}C NMR data are acquired, aligned and analysed by HCA in order to detect similarities between the samples and the results are visualized by heatmap toward the objective to identify the metabolites by exploring the individual clusters of chemical shifts generated after the analysis. This is performed by matching the ^{13}C NMR chemical shift clusters to a locally built database using ACD/NMR Workbook Suite 2012 containing the ^{13}C NMR data of 900 natural products, among them all reported compounds from *Acronychia* species. This may considered as a top-down dereplication strategy since a number of signals corresponding to a part of molecule is used as the input for a database search (van der Hooft et al. 2013).

Chapter 3: Metabolomics approaches in *Acronychia* genus

The main advantage of ^{13}C NMR acquisition consists in the more comprehensive structural description of NPs than ^1H NMR by the detection of all ^{13}C resonances, specifically, concerning those compounds possessing a noteworthy number of quaternary carbons. Moreover, the notably higher spectral width of ^{13}C NMR comparing to ^1H NMR and the decoupling of ^{13}C NMR spectra leads to an enhanced resolution of ^{13}C NMR signals reducing significantly signal overlapping generally observed in ^1H NMR spectra.

Nevertheless, the drawback of low sensitivity, much lower than ^1H NMR, probably remains the reason why ^{13}C NMR based metabolomics approaches have not widely developed. This difference in sensitivity concerning ^{13}C NMR in comparison to ^1H NMR is mainly due to the lower abundance of ^{13}C and leads to long acquisition times in order to obtain desirable results. Towards an attempt to increase the sensitivity (or increase S/N ratio) three main possibilities may be accounted namely application of stronger magnetic field, increased concentration and noise reduction. Improvements in probe technologies have contributed significantly in the direction of the two latter possibilities. Specifically, reduced detection volume NMR probes (microprobes) allow the acquisition of several microliters of solvent and thus, a much greater sample concentration is achieved resulting in greater sensitivity measurements. Moreover, the introduction of cryoprobes reduced significantly an important source of noise, the electronic noise, keeping the probe (not the sample) in very low temperature (Kovacs et al. 2005).

Based on the aforementioned strategy, the dereplication of known compounds in different *Acronychia* extracts was attempted using the ^{13}C NMR data obtained from a high magnetic field NMR spectrometer equipped with a cryogenic probe. Up to now, the application was restricted in simple mixtures of natural products, obtained after a fractionation procedure of the extracts (Hubert et al. 2014, Oettl et al. 2014). Therefore, in this case, all 120 samples were acquired using 4096 scans in order to obtain the desirable spectral information. After the preprocessing of the data (binning, alignment, filtering), a data matrix of 120 samples and 855 variables corresponding to ^{13}C intensities was subjected to pattern recognition analysis using HCA (euclidian distance, ward's method). A number of clusters were defined possibly corresponding to different metabolites (Figure 45). Thereafter, the signals of each cluster were used as an input for ^{13}C NMR database search. The results were evaluated by confirming the ^{13}C chemical shifts of the proposed metabolites. Moreover, a cross-check with the results obtained from the UPLC-HRMS dereplication approach was performed to enhance the reliability of the findings (Table XV).

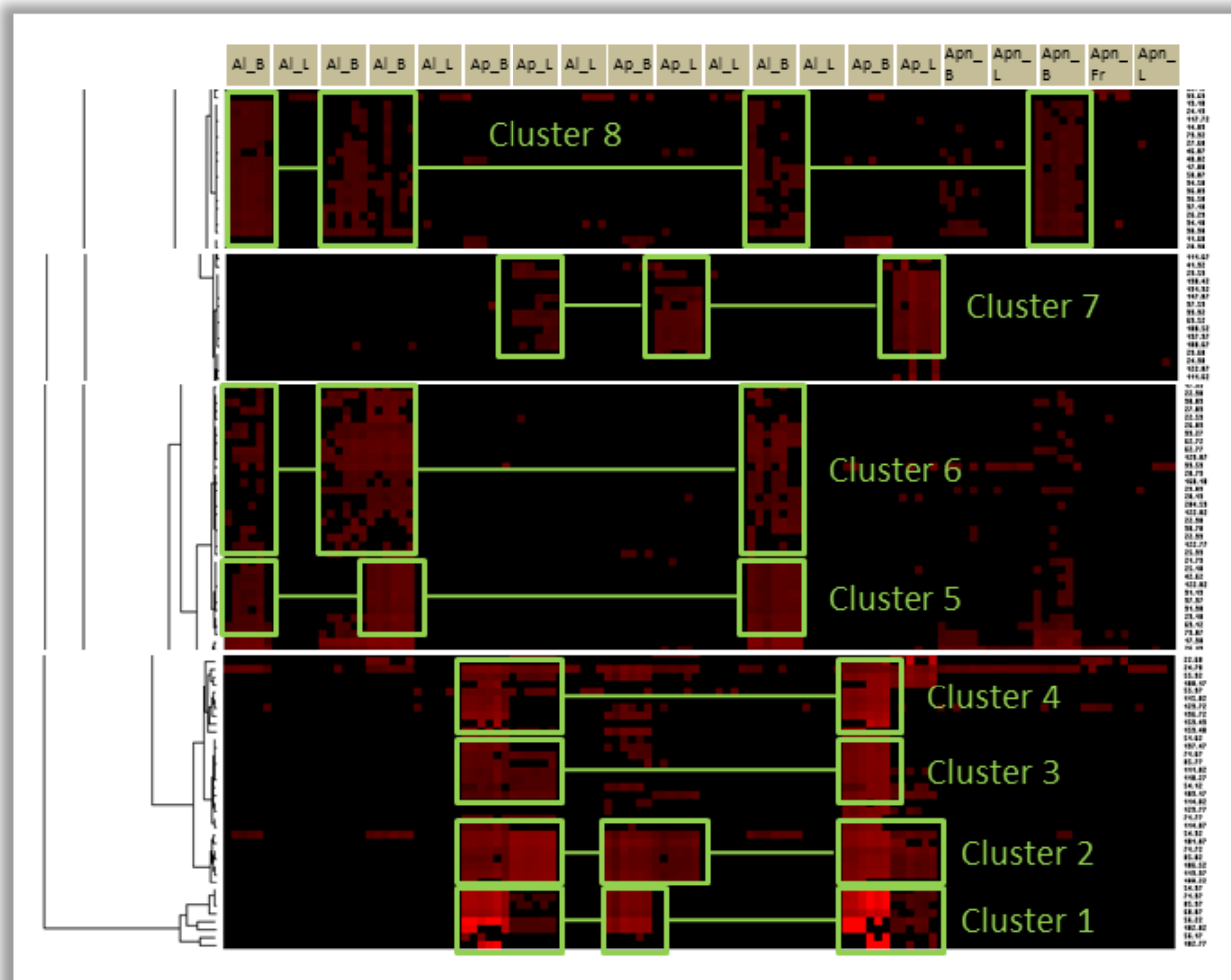


Figure 45: ^{13}C NMR chemical shift clusters revealed after HCA of the ^{13}C NMR data obtained from different *Acronychia* samples, particularly *A. laurifolia* barks are abbreviated as Al_B, *A. laurifolia* leaves as Al_L, *A. porteri* barks as Ap_B, *A. porteri* leaves as Ap_L, *A. pedunculata* barks as Apn_B, *A. pedunculata* fruits as Apn_Fr and *A. pedunculata* leaves as Apn_L.

The first three clusters observed at the heatmap were assigned to lignan compounds exclusively detected in *Acronychia porteri* samples. Specifically, cluster 1 was attributed to yangambin, a lignan which was also detected by UPLC-HRMS dereplication strategy, and interestingly in both cases was traced in *A. porteri* leaves and barks samples. The database search for cluster 2 proposed the occurrence of asarinin. Asarinin was also observed by UPLC-HRMS mainly in *A. porteri* leaves while the current strategy suggested its presence in both *A. porteri* leaves and barks samples. Moreover, another lignan, sesamolin, was observed for the first time in *A. porteri* samples which is described in the literature as constituent of *A. laurifolia* roots (Cui et al. 1999). Despite the similar structures of asarinin and sesamolin and thus, the very close ^{13}C NMR data, the applied strategy revealed

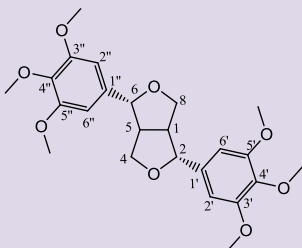
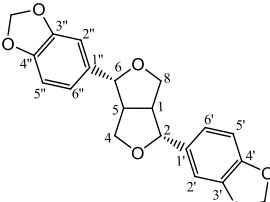
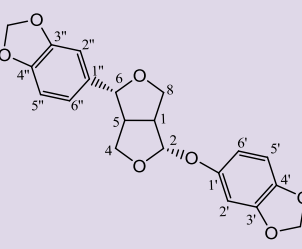
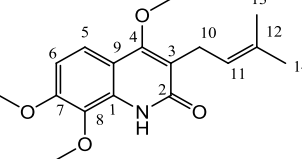
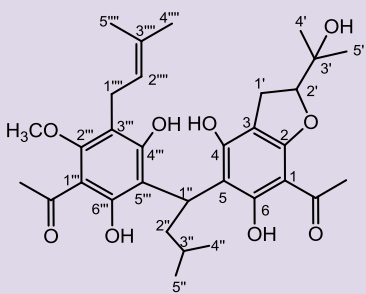
successfully the presence of these two compounds in separate clusters demonstrating a good precision. Cluster 4 consisted of 7 ^{13}C NMR signals was attributed to preskimmianine, a quinolone alkaloid precursor of furoquinoline alkaloids, which was mainly observed in *A. porteri* barks. This finding is in absolute accordance with the UPLC-HRMS dereplication results.

Furthermore, two clusters (5 and 6) unravelled the presence of acetophenone compounds. The database search in these cases proposed a number of different acetophenone derivatives, therefore, verification of the structures by comparison of their structural characteristics was essential to define the metabolites represented by these specific clusters. For instance, for cluster 5 the structures of acronylin, acrofolione B and acropyranol B were suggested. Among them, after comparison of the ^{13}C NMR chemical shifts with the literature data, acrofolione B was proposed due to the presence of features at δ 92.0 and 98.8 attributed to C-2' and C-1 of the structure. Accordingly, cluster 6 was attributed to acrovestone and not to acronyculatin B due to the presence of ^{13}C NMR signals corresponding to carbon atoms of the isopentyl chain. Both acrofolione B and acrovestone were found mainly in *A. laurifolia* bark samples their absence in *A. porteri* samples confirmed the doubt arose from the UPLC-HRMS findings concerning the characterization of AtA as chemotaxonomic markers.

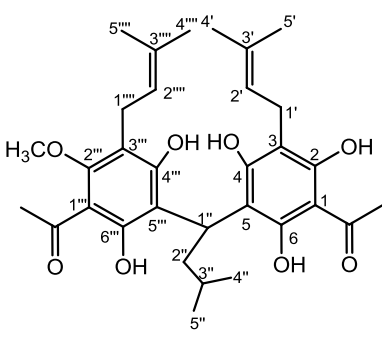
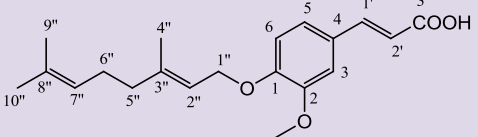
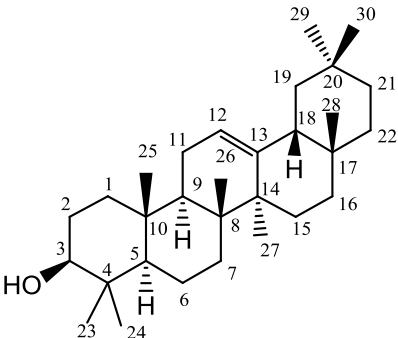
The presence of a cluster located exclusively in *A. porteri* leave samples was deduced as 4'-geranyloxyferulic acid. However, according to UPLC-HRMS dereplication approach this metabolite was mainly observed in *A. laurifolia* leave samples introducing an uncertainty for the assignment of this compound. Finally, b-amyirin was additionally observed as a cluster of 15 ^{13}C NMR chemical shifts located in *A. laurifolia* bark and leave samples. B-amyirin was also found by UPLC-HRMS in *A. laurifolia* leave samples.

The application of a ^{13}C NMR based dereplication strategy has led successfully to the identification of several known compounds reported from *Acronychia* genus. This approach expanded in more complex mixtures, such as plant extracts, was capable to discriminate structures with close ^{13}C NMR data. Moreover, the combination of this strategy with a UPLC-HRMS dereplication approach may provide more reliable evidences for the metabolite composition of plant extracts.

Table XV: Identification of metabolites in *Acronychia* samples using the ^{13}C NMR based dereplication strategy and verification of the results by comparison with the literature data and the results obtained by the previously performed UPLC-HRMS dereplication strategy

Clusters	Database results	Structure verification	Validation by MS	Occurrence
Cluster 1	Yangambin 	102.8 / 102.0 (C-2', 6', C-2'', 6'') 56.2/ 56.4 (3', 3'', 5', 5'' -OCH ₃) 60.8 (4', 4''-OCH ₃) 86.0(C-2, C-6) 75.0 (C-4, C-8) 55.0 (C-1, C-5)	Table XIV: ID 31	Ap_B
Cluster 2	Arsarin 	55.2 (C-1/5) 85.8 (C-2/6) 71.8 (C-4/8) 101.8 (-O-CH ₂ -O) 106.5 (C-2'/2'') 108.2 (C-5'/5'') 120.0 (C-6'/6'')	Table XIV: ID 30	Ap_B, Ap_L
Cluster 3	Sesamolin 	119.8 (C-6') 71.8 / 71.67(C-8) 109.4 (C-5') 54.2 (C-1) 118.2 (C-6') 85.8 (C-6) 51.6 (C-5) 111.0 (C-6'')	NO	Ap_B
Cluster 4	Preskimmianine 	159.4 / 159.6 (C-8, C-4) 129.6 (C-12) 115.8 (C-5) 56.0 (C-4, 7, 8) 108.2 (C-6) 55.8 (C-4, 7, 8) 24.8 (C-14)	Table XIV: ID 8	Ap_B
Cluster 5	Acrofolione B 	73.8 (C-3') 63.4 (-CH ₃ O) 23.2 (C-1''''') 92.0 (C-2') 98.8 (C-1) 31.2 (CH ₃ CO-1) 122.0 (C-2''''') 41.6 (C-2'') 25.1 (C-4''''') 21.8 (C-4'', 5'')	Table XIV: ID 24	Al_B, Apn_B

Chapter 3: Metabolomics approaches in *Acronychia* genus

Cluster 6	Acrovestone 	26.8 (C-3'') 122.8/ 123.0 (C-2') 22.4 (C-1') 38.8 (C-2'') 23.0 (C-1''') 122.0 (C-2''') 204.6 (-CO-1, 1''') 23.4 (C-4'', 5'') 160.2 (C-6, 6'') 28.8 (C-1'') 33.4/ 33.6 (CH ₃ CO- 1) 62.6 (CH ₃ O- 2'')	Table XIV: AI_B ID 28
Cluster 7	4'-geranyloxyferulic acid 	108.6 (C-5) 128.0 (C-4) 59.6 (CH ₃ O) 33.9 (C-5'') 27.6 (C-6'') 147.8 (C-1, 2) 131.2 (C-1') 138.4 (C-2') 23.4 (C-9'', 10'')	Table XIV: Ap_L ID 33
Cluster 8	Beta-amyrin 	31.0 (C-20) 14.6 (C-24, 25) 38.8 (C-1, 4) 34.6 (C-21) 26.4 (C-15) 37.2 (C-22) 36.8 (C-10) 46.8 (C-19) 47.6 (C-18) 48.0 (C-9) 79.8 (C-3) 27.4 (C-2) 119.8 (C-12) 24.6 (C-27) 19.2 (C-6)	Table XIV: AI_B, Apr_B ID 32

4. Investigation of *Acronychia* species taxonomy using NMR and LC-MS based metabolomics approaches

The genus *Acronychia* is consisted of 48 species distributed widely in Australasia and New Caledonia. One widespread species *A. pedunculata* is distributed throughout Malaysia and westward to western India and its northern boundaries are northern India and

southern China (Hartley 1974). Regarding the taxonomy of *Acronychia* genus, there is a continuous discussion concerning its relationship with the genera *Euodia* and *Melicope* which is not unravelled until today (Appelhans et al. 2014). This is supported also from the occurrence of a specific group of compounds, prenylated acetophenones which constitute valuable chemotaxonomic markers of the subfamily Rutoideae, tribe Xanthoxyleae sensu Engler (Adersen et al. 2007). Moreover, the botanical synonymy of several *Acronychia* species is confused in the literature data. For instance, *A. pedunculata* and *A. laurifolia* due to their morphological similarity have been described as synonyms by Hartley (Hartley 1974) while other sources refer to them as different species (Epifano et al. 2013). *Acronychia* species have a long tradition in eastern folk medicine used for multiple purposes among them asthma, ulcers and rheumatism. Furthermore, some species possess an important dietary role as the fruits and aerial parts are used in salads and as condiments (Epifano et al. 2013). The use of herb materials in traditional medicine systems and in diet necessitates the authentication of raw material in terms of safety and efficacy. In particular, closely related species, which do not differ significantly as far as their morphological characteristics are concerned, exhibit different properties. Thus, the clarification of taxonomic issues in *Acronychia* genus is of great importance due to its extensive use in eastern world. Metabolomics profiling approaches have been proven a powerful tool for the investigation of similarities and differences in different biological samples by exploring their metabolite composition simultaneously in an untargeted and unbiased way (Tikunov et al. 2005). Specifically in plant metabolomics field metabolic analysis has been used extensively in the last years for the discrimination of different species using either NMR (Kim et al. 2010) or LC-MS (Kim et al. 2012) or both techniques (Safer et al. 2011, Porzel et al. 2014).

In the current study, the analysis of the *Acronychia* extracts was performed by NMR and LC-ESI(±)-MS techniques collecting multiple snapshots of the metabolite composition of each of the different samples (see Results and Discussion 1 for further details). The parallel analysis with these two techniques is used to provide a better insight at the chemical composition of the different species and organs of *Acronychia* samples. Due to the complexity of the acquired data, chemometrics analysis techniques were applied for the handling of both large metabolomics datasets because of their ability to provide interpretable and rigorous models for complex correlated datasets (Trygg et al. 2006).

4.1. Classification of *Acronychia* samples using NMR based metabolomics

NMR spectroscopy is mainly used in plant metabolomics for the classification of different samples and identification of biomarkers. NMR is a suitable method for the simultaneous detection of diverse groups of secondary metabolites. Moreover, NMR constitutes a valuable technique for the structure elucidation of natural products (Robinette et al. 2011, Halabalaki et al. 2014). Actually, in the majority of applications in plant metabolomics field, ^1H NMR spectra of multiple plant extracts are used for classification purposes among the other applications (Wang et al. 2004, Kim et al. 2005, Safer et al. 2011, Zhi et al. 2012). However, in cases that extensive signal overlapping is occurring projection of 2D JRES data (pJRES) are utilized to reduce the complexity of the data and thus enhance the resolution of the method.

During this study, as aforementioned (see Results and Discussion 1.2) ^1H NMR data presented extensive signal congestion in the aliphatic region while pJRES data were characterized by more comprehensive signals. In order to take advantage of this enhanced resolution obtained from the pJRES spectra, the classification of the different *Acronychia* samples based on their metabolite profiling by NMR was performed using the pJRES spectra dataset.

For the analysis of this large dataset an unsupervised principal component analysis (PCA) method was applied. The JRES data were reduced by PCA and a model comprising of 12 principal components accounting for the 92.3% of the total variance of the dataset was constructed using Metaboanalyst platform (Figure A 76). Nevertheless, the first three components were utilized in order to get an insight into the main metabolic differences of the various *Acronychia* samples explaining the 71.6% of the overall variance. Looking at the score plot of the first and second component (PC1/PC2) a clear separation of *A. laurifolia* and *A. pedunculata* species versus *A. porteri* species was observed along the PC1 axis indicating the close relationship of *A. laurifolia* and *A. pedunculata* species. The discrimination of the different organs of *A. laurifolia* and *A. pedunculata* species was mainly characterized by PC2. A distinct separation of the leaves and barks was observed while the fruit samples were placed in between leave and bark samples. It is worth to note that also VN0179B an *A. pedunculata* sample was not clustered together with the other bark samples mainly across PC2 designating a slightly different metabolic composition from the

rest *A. laurifolia* and *A. pedunculata* bark samples. Finally, good discrimination of *A. porteri* barks and leaves was detected by PC3 (Figure 46).

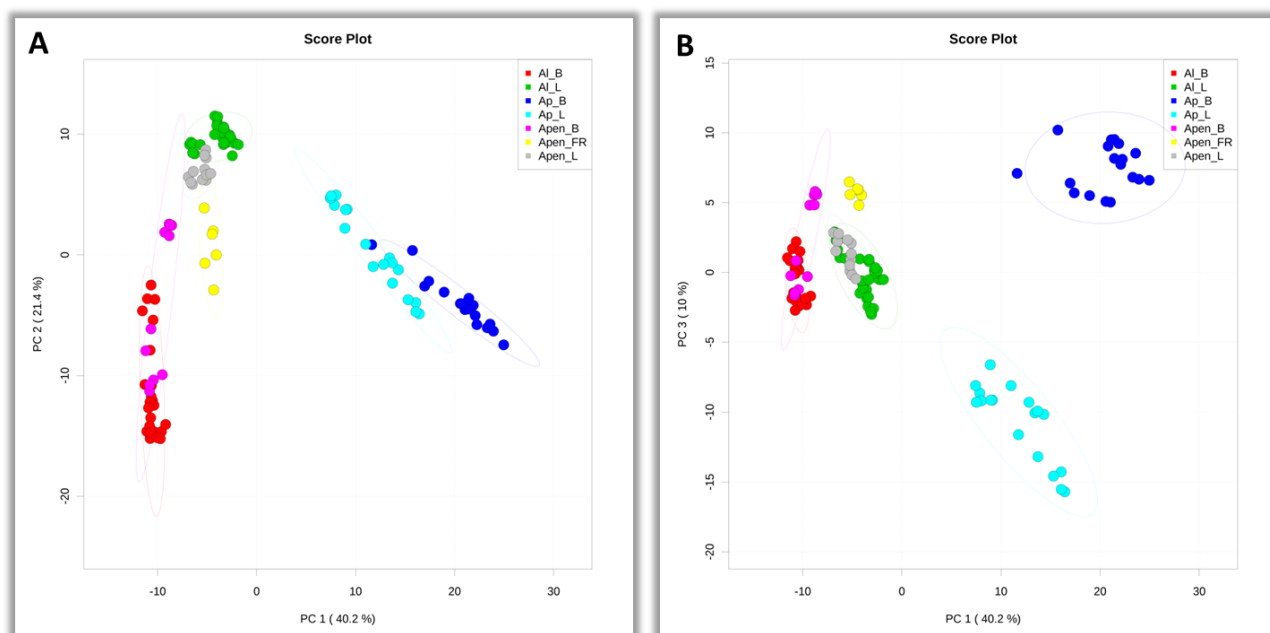


Figure 46: PCA scores plots of A. PC1/PC2 and B. PC1/PC3 obtained from the pJRES data of *Acronychia* samples; *A. laurifolia* barks are coloured in red, *A. laurifolia* leaves in green, *A. porteri* barks in blue, *A. porteri* leaves in turquoise, *A. pedunculata* barks in pink, *A. pedunculata* fruits in yellow and *A. pedunculata* leaves in grey

Similarly to PCA, hierarchical cluster analysis (HCA), another unsupervised method, was used to assess the differences and similarities between *Acronychia* samples based on the multivariate distance between each sample and clustering them according to the relative proximity of their metabolite profiles (Fukusaki et al. 2005). Moreover, outlier detection can also be performed by HCA as they usually form a distant branch that joins the main cluster at a very high level (Xia et al. 2011). In the current study, Metaboanalyst platform was used to perform HCA. Based on ward's method for clustering HCA of the multiple *Acronychia* samples revealed similar clusters with PCA, confirming the robustness of the analysis (Figure 47). Two strong branches were formed representing the differences between *A. laurifolia* and *A. pedunculata* species versus *A. porteri* species. In addition, the barks and the leaves from each of the two main groups were separated in distinct branches of smaller distance. By means of HCA, the relationship of the *A. pedunculata* fruit samples and VN0179B *A. pedunculata* bark samples was defined closer to *A. laurifolia* and *A. pedunculata* leave samples.

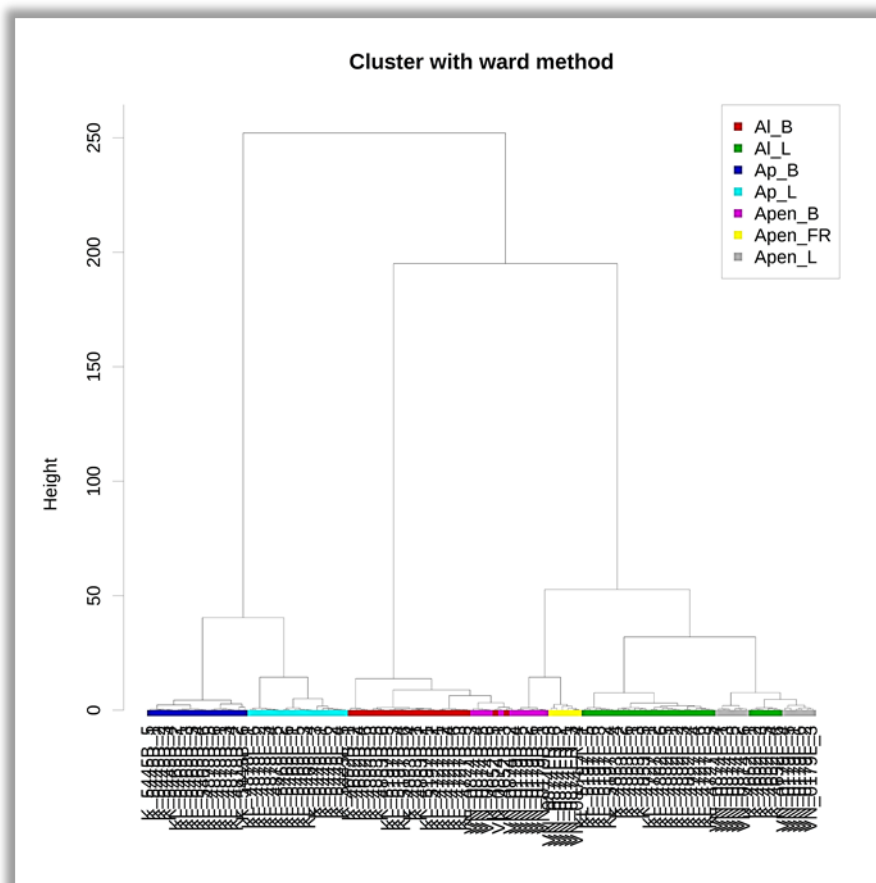


Figure 47: Dendrogram of different *Acronychia* samples based on hierarchical cluster analysis obtained from pJRES data using ward's method for clustering and pearson distance; *A. laurifolia* barks are coloured in red, *A. laurifolia* leaves in green, *A. porteri* barks in blue, *A. porteri* leaves in turquoise, *A. pedunculata* barks in pink, *A. pedunculata* fruits in yellow and *A. pedunculata* leaves in grey

Despite the fact that NMR is a versatile method for the identification of metabolites, in plant metabolomics field due to the lack of NMR databases of plant secondary metabolites, studies are mainly focused in the identification of sugar, aminoacids and organic acid compounds based on ^1H NMR chemical shift values (Verpoorte et al. 2007, Kim et al. 2010, Robinette et al. 2011). However, prior knowledge of the metabolite composition of the investigated species may give some clues about the categories or the metabolites that are responsible for the classification (biomarkers). This is effectuated comparing representative chemical shifts with reference data from the literature or available reference compounds. In the current study, the followed extraction protocol involving the utilization of EtOAc was not expected to afford the above mentioned commonly reported metabolites. The array of metabolites that were likely to occur in the extracts was mainly focused on the secondary medium polarity metabolites. Therefore, the exploration of the chemical nature of the

biomarkers was performed investigating the loading plots (Figure 48) together with the literature data. Moreover, in order to enhance the resolution of the features and get more information for structural determination, 2D NMR experiments (homonuclear COSY and heteronuclear HSQC, HMBC) were also acquired for all the 20 different biological samples (Figure A 77).

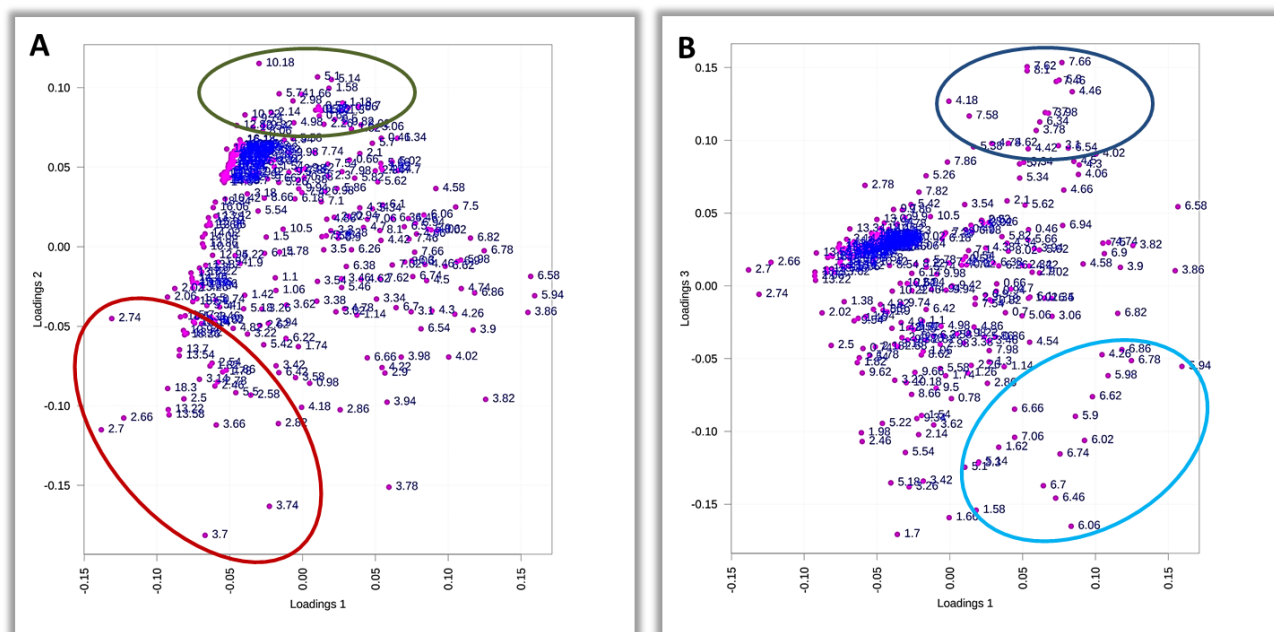


Figure 48: PCA loadings plots of A. PC1/PC2 and B. PC1/PC3 based on JRES dataset of *Acronychia* samples; indicated discriminating features in red for *A. laurifolia* / *A. pedunculata* bark samples; in green *A. laurifolia* / *A. pedunculata* leaf samples; in blue for *A. porteri* bark samples; in turquoise for *A. porteri* leaf samples

In particular, discriminating signals for *A. laurifolia* / *A. pedunculata* bark samples were observed in the PC1/PC2 loadings plot among them chemical shifts corresponding to AtA compounds (features in ranges δ 1.38, 2.02- 2.06, 2.5- 2.84, 3.7- 3.74, 4.82, 5.22, 10.10- 10.42, 13.22- 13.58 and 15.26). To confirm the findings, the ^1H NMR, 2D JRES and HSQC spectra were used and all the aforementioned signals were corresponded to the literature data (Kouloura et al. 2012). For instance, characteristic ^1H and ^{13}C resonances of the methoxy groups of AtA were detected at δ 3.7 and 3.72 with the respective carbons at δ 62.3 and 63.0. Moreover, signals detected at δ 10.10-10.42, 13.22- 13.58 and 15.26 revealed the presence of characteristic hydroxy groups of AtA. Likewise, signals responsible for the discrimination of *A. porteri* bark samples were identified at δ 8.1, 7.66, 6.3 and 4.46 in the PC2/PC3 loadings plot. The respective resonances for the ^{13}C were extracted from the HSQC spectra at δ 123.6, 144.1, 115.2 and 59.0, respectively. According to the

forementioned signals and literature data the presence of furoquinoline alkaloids is suggested (de Silva et al. 1979). A number of other features were found responsible for the discrimination of *Acronychia* sample groups, however, their reliable correlation with metabolites was hard to be supported.

4.2. Classification of *Acronychia* samples using UPLC-ESI(+)-HRMS based metabolomics

Before analysing the data using multivariate data analysis (MVDA) the similarities and differences of the samples were evaluated by visual inspection of their metabolic profiles (Figure 49). Figure 49 illustrates more or less diverse patterns for each sample group that was analysed. In addition, major metabolites present in each group could be detected and attributed to known metabolites. Nevertheless, this approach could not provide any information regarding the significance of difference between these samples as well as the contribution of the major or minor metabolites in the classification of the various samples. Therefore, the LC-ESI(+)-MS dataset was subjected initially to unsupervised MVDA methods to explore the relative variability within the various sample groups and the investigation of the metabolites associated with this classification.

RT: 0.00 - 30.00

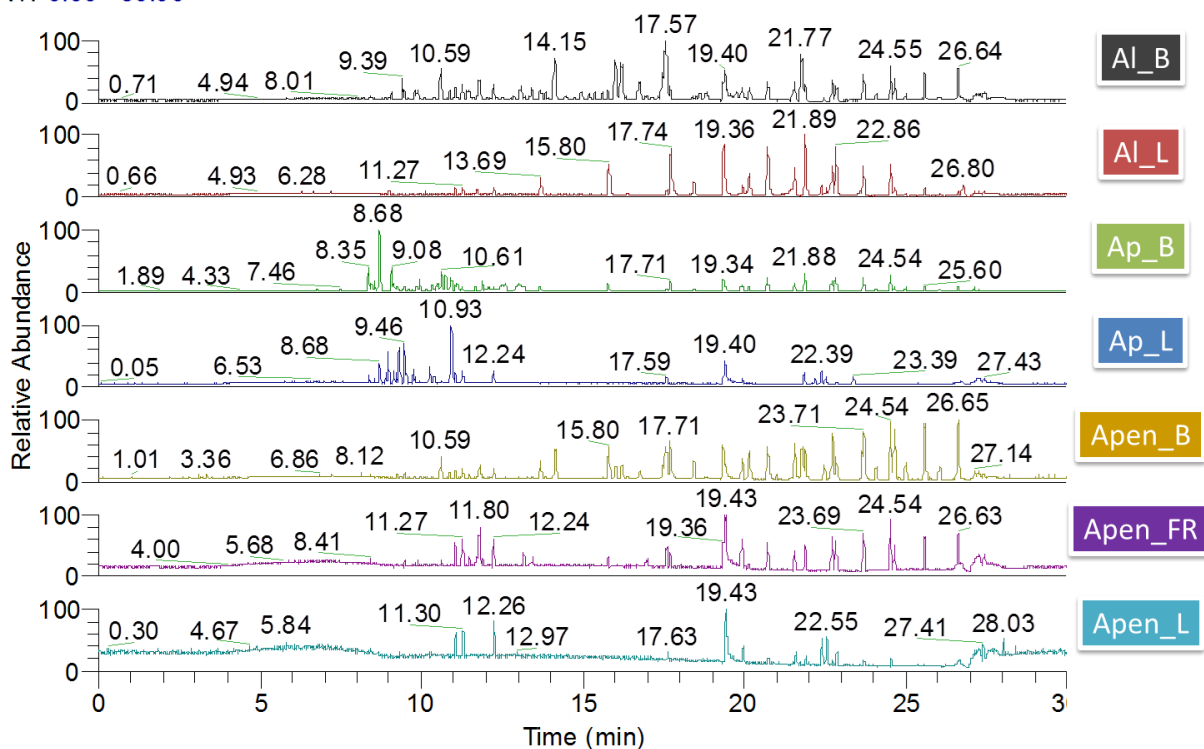


Figure 49: Base peak chromatograms of the seven different *Acronychia* samples. Al_B: *A. laurifolia* barks, Al_L: *A. laurifolia* leaves, Ap_B: *A. porteri* barks, Ap_L: *A. porteri* leaves, Apen_B: *A. pedunculata* barks, Apen_FR: *A. pedunculata* fruits, Apen_L: *A. pedunculata* leaves

Initially, PCA was used to observe the discrimination of the different *Acronychia* sample groups with an unsupervised manner. Hence, the data matrix imported to the Metaboanalyst platform (after filtering, missing value imputation, normalization and scaling) was analysed by PCA. Using Metaboanalyst web platform, a PCA model of 15 components was calculated accounting for 91.8% of the dataset variance. In the scores plot of PC1/PC2 the clustering of the different *Acronychia* sample groups was observed (Figure 50). In particular, distinct clustering of four groups was observed in Figure 50. The first group, separated from the rest on the PC1 as observed in the scores plot, is composed from samples of *A. laurifolia* and *A. pedunculata* barks as well as *A. pedunculata* fruits. The overlapping of *A. laurifolia* and *A. pedunculata* bark samples denoted similar metabolite composition of these two biological samples whereas *A. pedunculata* fruit samples seems to be related with the latter samples as far as their composition is concerned due to their proximity according to PC1, PC2 and PC3. The second group is consisted of *A. laurifolia* and *A. pedunculata* leaves which showed a distinct but close metabolite profile. Discrimination of the two organs of *A. porteri* was also visible along PC2. In the PC2/PC3 scores plot similar observations to PC1/PC2 scores plot were made concerning the clustering of the different sample groups. However, better discrimination of *A. porteri* bark and leave samples was observed due to their clear separation on PC3. Overall, very good discrimination of the different organs was performed using PCA analysis. Moreover, discrimination of *A. porteri* from the other species was observed while as far as *A. laurifolia* and *A. pedunculata* are concerned, only the different parts were clustered separately supporting the chemotaxonomic confusion of the aforementioned species. One main outlier was defined by PCA analysis of *Acronychia* species corresponding to one sample of *A. porteri* leaves (KL4878_2) which was discarded for further downstream analysis.

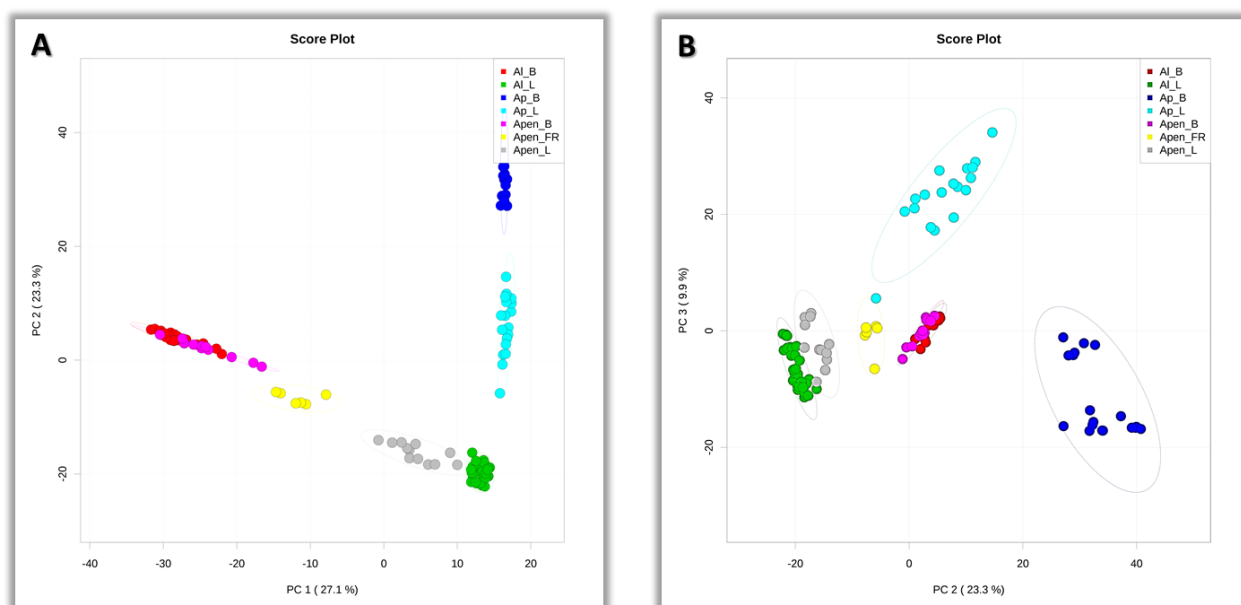


Figure 50: PCA scores plots of A. PC1/PC2 and B. PC2/PC3 of *Acronychia* samples using UPLC-ESI(+)-HRMS data; *A. laurifolia* barks are coloured in red, *A. laurifolia* leaves in green, *A. porteri* barks in blue, *A. porteri* leaves in turquoise, *A. pedunculata* barks in pink, *A. pedunculata* fruits in yellow and *A. pedunculata* leaves in grey

In order to confirm the robustness of our method, HCA was also used as an alternative unsupervised method to assess the differences between *Acronychia* samples and for the detection of biomarkers based on ward's method for clustering. As illustrated in the dendrogram (Figure 51), *A. pedunculata* fruit samples are presenting a closest connection to *A. laurifolia* and *A. pedunculata* bark samples, as grouped in one separate cluster, and all these differ significantly from the rest of the samples. *A. porteri* barks and leaves samples are more closely related between them comparing with the different organs of *A. laurifolia* and *A. pedunculata*. In addition, the discrimination of *A. laurifolia* and *A. pedunculata* species in different clusters was not obtained indicating the close relation or synonymy of these species as observed also by PCA. The sample KL4878_2 was also identified as outlier by HCA indicating the robustness of the analysis.

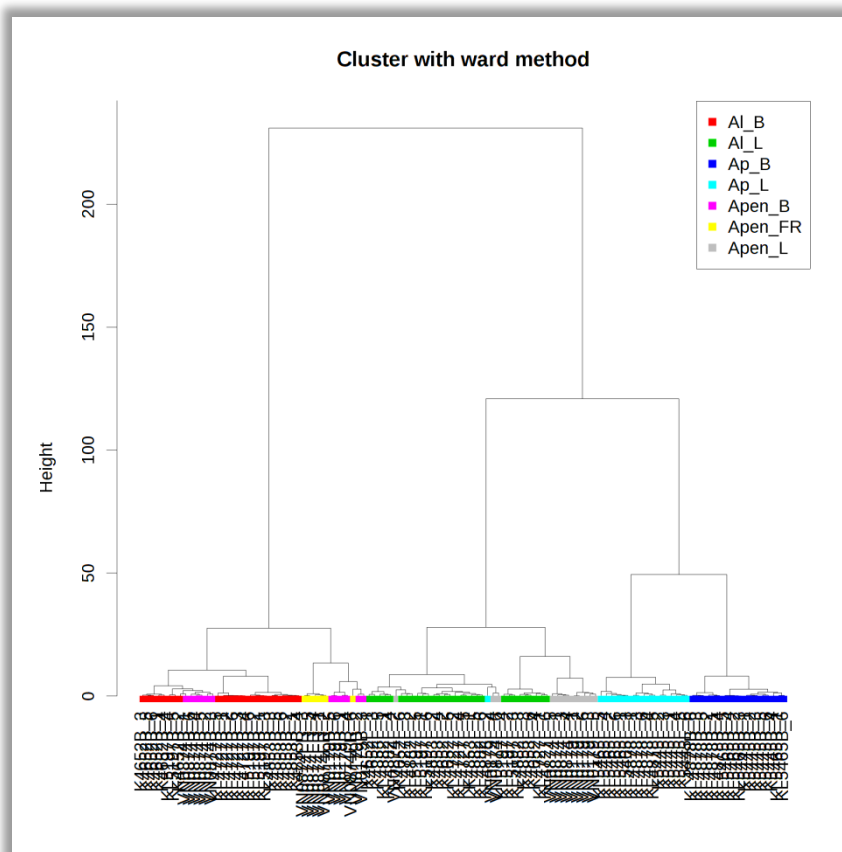


Figure 51: Dendrogram of different *Acronychia* samples based on hierarchical cluster analysis of UPLC-ESI(+)-HRMS data using ward's method for clustering and pearson distance; *A. laurifolia* barks are coloured in red, *A. laurifolia* leaves in green, *A. porteri* barks in blue, *A. porteri* leaves in turquoise, *A. pedunculata* barks in pink, *A. pedunculata* fruits in yellow and *A. pedunculata* leaves in grey

4.2.1. Biomarkers identification

After the establishment of the PCA model for the classification of the different *Acronychia* samples, the identification of biomarkers responsible for the classification of different species and organs of *Acronychia* samples was performed exploiting the most discriminatory signals observed in PCA loading plots of the respective score plots that were used to explain the variation between the different samples (Figure 52). In particular, as mentioned above four main groups were identified using PCA, which were confirmed by HCA. Thus, in order to identify the biomarkers for each of the four groups, their loadings on PC1, PC2 and PC3 were defined (Figure 52). Precisely, the variables with the smallest values on PC1 were considered the biomarkers for *A. laurifolia* and *A. pedunculata* bark samples. The discriminant variables for *A. laurifolia* and *A. pedunculata* bark samples were determined the variables presenting the largest values on PC1 and the smallest on PC2.

Accordingly, the biomarkers for the discrimination of *A. porteri* bark and leaf samples were determined by the smallest and largest values on PC3, respectively (Figure 52).

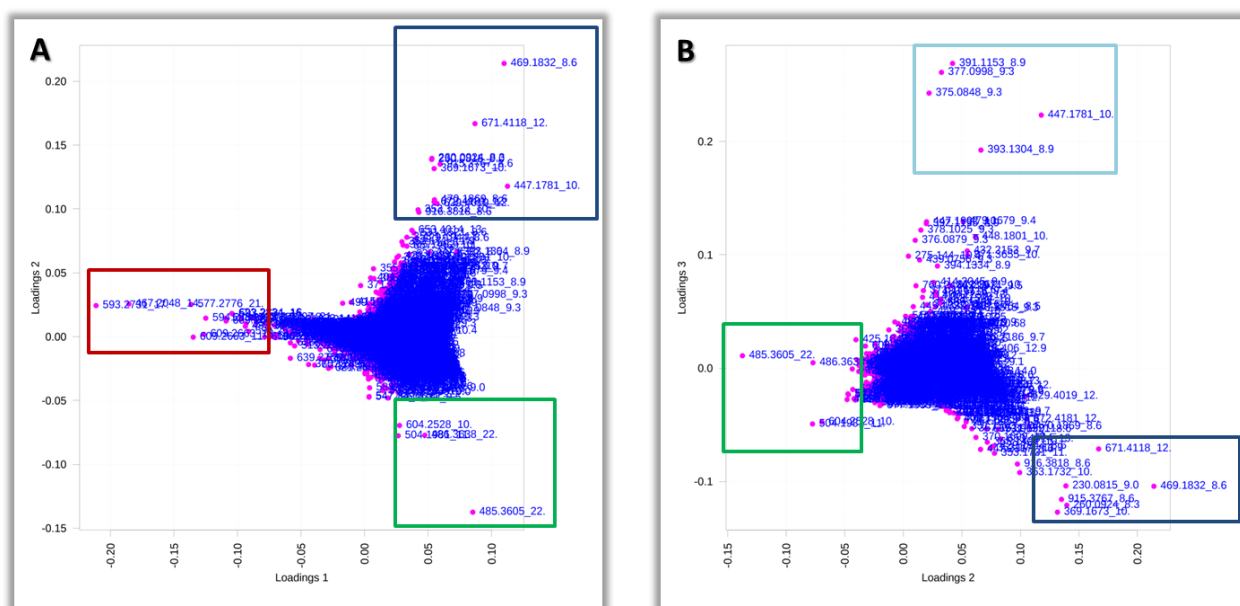


Figure 52: PCA loadings plots of A. PC1/PC2 and B. PC2/PC3 using UPLC-ESI(+)-HRMS dataset of *Acronychia* samples; indicated discriminating features in red for *A. laurifolia* /*A. pedunculata* bark samples; in green *A. laurifolia* /*A. pedunculata* leaf samples; in blue for *A. porteri* bark samples; in turquoise for *A. porteri* leaf samples

In this context, the 20 more significant features accountable for this classification were extracted from the loadings table for each group and are summarized in Table XVI. The application of both UPLC and HRMS for the analysis and afterwards the processing with XCMS and CAMERA algorithms resulted in the generation of features characterized by great robustness as is illustrated in Table XVI. As a result, multiple features corresponding to the same metabolites were characterized as discriminatory variables. In order to confirm the significance of the biomarkers, additionally analysis of variance (ANOVA) was performed in Metaboanalyst platform to compare quantitatively discriminatory variables across the groups. Because of the multiple-testing issue, FDR or Bonferonni corrected P-values computed from Metaboanalyst platform used to assess the significance (Broadhurst et al. 2006). The significance of the discriminating features of interest was also assessed by box-whisker plots (Figure A 78, Figure A 79). As a matter of fact, the concentration of all discriminant features was found to differ significantly in between the diverse biological groups giving the evidence that these could be used as biomarkers.

Table XVI: Biomarkers responsible for the classification of different *Acronychia* species and organ samples

No	Rt	m/z	MS/MS	Loss	Adduct/Isotope	EC	RDB	Name	Category	Annotation level	FDR	Class
Biomarkers for <i>A. porteri</i> leaves												
1	8.93	391.1153	373.1616	H ₂ O	[M+Na] ⁺	C ₂₁ H ₂₀ O ₆	11.5		flavanone	3	7.23E-27	A.p_L
1	8.94	392.1195			M+1	C ₂₁ H ₂₀ O ₆					1.26E-26	A.p_L
2	8.91	393.1304	no		[M+Na] ⁺	C ₂₁ H ₂₂ O ₆	10.5		flavonoid	3	3.86E-39	A.p_L
2	8.91	394.1334			M+1	C ₂₁ H ₂₀ O ₆					4.14E-38	A.p_L
3	9.32	439.0756	no		[M+2Na-H] ⁺	C ₂₂ H ₁₈ O ₇			flavonoid	3	3.22E-55	A.p_L
4	9.34	375.0848	no		[M+Na] ⁺	C ₂₀ H ₁₆ O ₆	12.5		flavonoid	3	8.25E-45	A.p_L
4	9.34	376.0879			M+1	C ₂₀ H ₁₆ O ₆					2.18E-44	A.p_L
5	9.33	377.0998	no		[M+Na] ⁺	C ₂₀ H ₁₈ O ₆	11.5	asarinin	lignan	2	7.73E-50	A.p_L
5	9.33	378.1025			M+1	C ₂₀ H ₁₈ O ₆					2.46E-49	A.p_L
6	9.49	479.1679	no		[M+Na] ⁺	C ₂₅ H ₂₈ O ₈	11.5		flavanone	3	2.81E-26	A.p_L
7	9.54	463.1914	no		[M+Na] ⁺	C ₂₅ H ₂₈ O ₇	11.5			4	5.62E-19	A.p_L
8	9.78	432.2153	400.1887	CH ₄ O	[M+Na] ⁺	C ₂₅ H ₃₁ NO ₄	10.5	Acidissiminol epoxide	tyramine derivatives	3	3.11E-37	A.p_L
9	9.92	414.2045	no		[M+K] ⁺	C ₂₂ H ₃₃ NO ₄	6.5		alkaloid	3	3.22E-55	A.p_L
10	10.41	275.1440	257.1523	H ₂ O	[M+H] ⁺	C ₂₀ H ₁₉ O	11.5			4	2.01E-07	A.p_L
			219.0654	C ₄ H ₈								A.p_L
11	10.94	871.3655	no		[2M+Na] ⁺	C ₂₅ H ₂₈ O ₆	11.5		flavonoid	3	1.39E-27	A.p_L
11	10.94	872.3671			M+1	C ₂₅ H ₂₈ O ₆					2.33E-27	A.p_L
Biomarkers for <i>A. porteri</i> barks												

Chapter 3: Metabolomics approaches in *Acronychia* genus

12	8.37	260.0924	245.0687	CH3	[M+H] ⁺	C14H13NO4	8.5	Skimmianine Maculosidine Kokusaginine	furoquinoline alkaloid	2	1.31E-47	A.p_B
		227.0579		CH3+H2O								A.p_B
13	8.69	429.1944	411.1808	H2O	[M+H-H2O] ⁺	C24H30O8		yangambin	lignan	2	8.08E-45	A.p_B
13	8.69	447.2317	429.1908	H2O	[M+H] ⁺	C24H30O8					2.5E-42	A.p_B
13	8.69	469.1832	no		[M+Na] ⁺	C24H30O8	9.5				2.63E-68	A.p_B
13	8.69	470.1869			M+1	C24H30O8					6.09E-70	A.p_B
13	8.69	915.3767	no		[2M+Na] ⁺	C24H30O8					4.45E-42	A.p_B
13	8.69	916.3818			[2M+Na] ⁺ M+1	C24H30O8					2.5E-42	A.p_B
14	8.69	531.1521	no		[M+H] ⁺	C26H26O12	13.5		phenolic glycoside	3	2.07E-71	A.p_B
15	9.09	230.0815	215.0581	CH3	[M+H] ⁺	C13H11NO3	8.5	Pteleine Evolitrine g-fagarine	furoquinoline alkaloid	2	1.52E-41	A.p_B
16	10.57	353.1732	201.0524	C10H16O	[M+Na] ⁺	C20H26O4	7.5			4	3.83E-44	A.p_B
17	10.68	369.1673	337.1414	CH4O	[M+Na] ⁺	C20H26O5	7.5			4	1.27E-26	A.p_B
18	10.78	365.1402	309.1338	C4H9	[M+Na] ⁺	C20H22O5	9.5			4	1.28E-21	A.p_B
19	11.13	353.1726	335.186	H2O	[M+Na] ⁺	C20H26O4	7.5			4	1.72E-28	A.p_B
20	11.83	657.3969	no		[M+Na] ⁺	C36H58O9	7.5		triterpenoid glycoside	3	3.8E-119	A.p_B
21	11.98	353.171	201.0524	C10H16O	[M+Na] ⁺	C20H26O4	7.5			4	1.39E-30	A.p_B
			335.186	H2O								A.p_B
22	12.03	629.4019	527.3339	C5H10O2	[M+Na] ⁺	C35H58O8	6.5		triterpenoid	3	5.18E-41	A.p_B

glycoside											
23	12.52	671.4118	569.3452	C5H10O2	[M+Na] ⁺	C37H60O9	7.5	Goyaglycoside a/ Momordicoside K	triterpenoid glycoside	3	1.13E-72 A.p_B
23	12.53	672.4181			M+1	C37H60O9				3	4.79E-72 A.p_B
24	13.07	653.4014	551.3339	C5H10O2	[M+Na] ⁺	C37H58O8	8.5		triterpenoid	3	1.59E-12 A.p_B
Biomarkers for <i>A. laurifolia</i> / <i>A. pedunculata</i> leaves											
25	8.69	465.2449	no		[M+Na] ⁺	C23H38O8	4.5			4	2.02E-27 A.I_L
26	9.01	425.1846	407.205	H2O	[2M+Na+K-H] ⁺		9.5				4.69E-09 A.I_L
27	9.02	449.2499	no		[M+Na] ⁺	C23H38O7	4.5			4	6.78E-33 A.I_L
28	9.14	433.2561	no		[M+Na] ⁺	C23H38O6	4.5			4	5.58E-36 A.I_L
29	9.14	549.3034	no		[M+Na] ⁺	C28H46O9	5.5			4	1.64E-31 A.I_L
30	9.53	455.2037	323.1256		[M+Na] ⁺	C24H32O7	8.5			4	8.74E-26 A.I_L
31	10.13	529.2103	no		[M+Na] ⁺	C19H38I5	0.5			4	4.22E-05 A.I_L
32	10.3	588.2562	531.2228		[M+Na] ⁺	C32H39NO8	13.5			4	1.83E-44 A.I_L
33	10.32	604.2528	572.226	CH4O	[M+Na] ⁺	C32H39NO9	13.5			4	4.73E-37 A.I_L
33	10.32	605.2531			M+1						7.39E-36 A.I_L
34	10.72	284.1263	no		[M+Na] ⁺	C15H19NO3	6.5		alkaloid	3	5.31E-21 A.I_L
35	11.72	504.1981	275.0897	C14H15NO2	[M+Na] ⁺	C27H31NO7	12.5			4	2.93E-22 A.I_L
35	11.73	505.203			M+1	C27H31NO7				4	1.99E-22 A.I_L
36	16.47	547.34	501.0783		[M+Na] ⁺	C33H48O5	9.5			4	1.95E-30 A.I_L
37	17.35	556.341	no		[M+Na] ⁺	C34H47NO4	11.5			4	1.06E-31 A.I_L
38	17.88	563.3358	no		[M+Na] ⁺	C33H48O6	9.5			4	5.19E-33 A.I_L
39	21	547.3396	no		[M+Na] ⁺	C33H48O5	9.5			4	1.91E-23 A.I_L

Chapter 3: Metabolomics approaches in *Acronychia* genus

40	22.45	485.3605	no		[M+Na]+	C29H50O4	4.5		4	4.74E-32	A.I_L
40	22.45	486.3638			M+1	C29H50O4				3.69E-32	A.I_L
41	22.46	501.3358	no		[M+Na]+	C32H46O3	9.5		4	1.57E-26	A.I_L
Biomarkers for <i>A. laurifolia</i> / <i>A. pedunculata</i> barks											
42	10.03	483.1983	425.1577	C3H6O	[M+Na]+	C25H32O8	9.5	phloroglucinol derivative	3	1.1E-48	A.I_B
			357.1678								A.I_B
			289.1051								A.I_B
43	10.61	609.266	291.0842		[M+Na]+	C32H42O10	11.5		4	3.76E-70	A.I_B
			359.147								A.I_B
44	10.86	373.1659	no		[M+Na]+	C19H26O6	6.5	sesquiterpenoid	3	7.22E-52	A.I_B
45	11.6	357.1681	301.105	C4H8	[M+Na]+	C19H26O5	6.5		4	1.51E-45	A.I_B
46	11.79	609.2663	357.1674		[M+Na]+	C32H42O10	11.5	AtA OH type-L	3	7.27E-60	A.I_B
46	11.79	610.2704			M+1	C32H42O10				3.1E-60	A.I_B
47	13.4	609.2663	357.1678		[M+H-H2O]+				4	7.27E-60	A.I_B
48	14.16	467.2048	273.1102		[M+Na]+	C25H32O7	9.5		4	2.17E-63	A.I_B
48	14.16	468.2061			M+1	C25H32O7				1.5E-62	A.I_B
49	16.13	623.2834	387.1781		[M+Na]+	C33H44O10	11.5		4	5.72E-17	A.I_B
50	16.76	593.2731	357.1675		[M+Na]+	C32H42O9	11.5	AtA	3	2.36E-50	A.I_B
51	17.45	591.2576	no		[M+Na]+	C32H40O9	12.5	AtA	3	1.38E-43	A.I_B
52	17.49	341.1731	285.1103	C4H8	[M+Na]+	C19H26O4	6.5	Cohulupone	3	3.51E-32	A.I_B
53	17.49	373.1989	341.1782	CH4O	[M+Na]+	C20H30O5	5.5	hemiterpene	4	4.39E-31	A.I_B
54	17.59	593.2731	341.1725		[M+Na]+	C32H42O9	11.5	Acrofolione A	1	2.36E-50	A.I_B

Results and Discussion

54	17.58	594.2736			M+1	C32H42O9					6.49E-76	A.I_B
55	18.77	593.2731	341.1729		[M+Na]+	C32H42O9	11.5	Acropyranol A	AtA	1	2.36E-50	A.I_B
56	19.47	593.2733	533.2547		[M+Na]+	C32H42O9	11.5	Acrofolone B	AtA	1	1.32E-47	A.I_B
57	21.8	577.2776	341.1726	C13H16O4	[M+Na]+	C32H42O8	11.5	Acrovestone	AtA	1	9.39E-24	A.I_B
57	21.8	578.2797			M+1	C32H42O8					9.09E-24	A.I_B

Chapter 3: Metabolomics approaches in *Acronychia* genus

For the identification of the metabolites represented from each feature or group of features a strategy, comprising multiple steps was applied (Figure 53). First of all, metabolites presenting the same retention time were grouped together in order to assess whether they attributed to adduct, fragment or dimer ions of the same metabolite or coeluting metabolites. For this step, the valuable information from CAMERA algorithm was exploited. Thus, features corresponding to $[M+H]^+$, $[M+Na]^+$, $[2M+Na]^+$, $[M+H-H_2O]^+$, $[M+H-HCOOH]^+$ were identified and ascribed to the same metabolites along with the isotope ions M , $M+1$, $M+2$, $M+3$. After the grouping step, all biomarkers were matched with the metabolites of the in-house database of *Acronychia* compounds in terms of accurate mass, retention time and MS/MS spectra when available. Thus, the biomarkers corresponding to reported compounds from the literature were annotated after verification of the isotopic pattern and RDB values. For the determination of the possible structures of the rest discriminating features, accurate mass search was performed in publicly available databases (METLIN, MassBank and HMDB) and MS/MS search (MetFrag). In order to reduce the number of candidate metabolites, isotopic pattern and RDB value match were performed and possible correlations with the reported metabolites from the genus were considered.

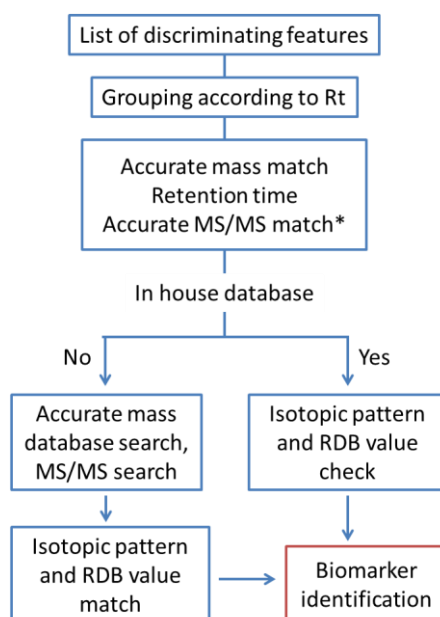


Figure 53: Flowchart summarizing the main steps for HRMS identification

Summarizing the information of Table XVI, PCA of *Acronychia* samples lead to the determination of biomarkers for each of the four distinct groups that were revealed. In particular, *A. porteri* leaves presented significant accumulation of flavonoid compounds

which is in absolute coherence with the unique available literature concerning the phytochemical investigation of *A. porteri* leaves (Lichius et al. 1994). Particularly, the biomarkers (metabolites 1-4, 6, 9, 11) attributed to flavonoid compounds have not been reported from *Acronychia* species. However, their common biosynthetic pathways with the ones reported in the literature imply a higher level of confidence concerning their annotation. Interestingly, a wide array of compounds, chemically diverse, is defined as the discriminant variables for *A. porteri* barks among them furoquinoline alkaloids (metabolites 12, 15), a lignan (metabolite 13) and a number of triterpenoid glycosides. Despite the fact that *A. porteri* barks have not been investigated previously, a number of furoquinoline alkaloids (Cui et al. 1999, Kouloura et al. 2012), several lignans (Davenport et al. 1954, Cui et al. 1999) and triterpenes (Govindachari et al. 1969, Rahmani et al. 1996) reported in the literature from other species assisted the annotation of these biomarkers. As far as the *A. laurifolia* / *A. pedunculata* leaves group is concerned, a number of biomarkers were found but could not be identified. Finally, the majority of biomarkers identified for *A. laurifolia* / *A. pedunculata* barks are belonging to *Acronychia*-type acetophenones (AtA). Apart from the other metabolites that were identified, based on the previous work (Chapter 2, Results and Discussion 2) metabolite 46 was assigned as AtA with additional OH group and additional ring type L. Following this strategy for the identification of biomarkers in different *Acronychia* sample groups a better insight of the chemical composition of these groups was obtained.

4.3. Discrimination of *A. laurifolia* and *A. pedunculata* species

As mentioned above, the synonymy of *A. laurifolia* and *A. pedunculata* species is an obscure issue (see Chapter 1, Introduction 2.1 for further details). Previously, UPLC-ESI(+)-HRMS dataset obtained from *Acronychia* samples was reduced by PCA in order to obtain the maximum variation between the samples. Four distinct groups were formed; however, the discrimination of *A. laurifolia* and *A. pedunculata* species was not possible. As a consequence, the use of another tool for understanding the difference between the two groups and reveal possible discriminating metabolites was necessary. A supervised model, such as PLS-DA or OPLS-DA analysis, can be used as they provide a correlation between the variables and the groups so the detection of chemotaxonomic markers between these two species may be expected. PLS-DA and OPLS-DA are prediction and regression methods that reveal information from the X dataset related to known information Y dataset. OPLS-DA can

Chapter 3: Metabolomics approaches in *Acronychia* genus

be considered as a modification of the traditional PLS-DA, with integral orthogonal signal correction filter (Bylesjö et al. 2006). The separation of the Y-predictive and Y-orthogonal variation, the discriminating and the one that do not contribute to the classification, respectively, facilitates the interpretation and visualization of the model specifically when the orthogonal variation is significant. However, the predictions that are made with both PLS and OPLS methods are identical (Tapp et al. 2009). Therefore, in the current study, OPLS-DA in Simca 13.0 software was used to get a better understanding of the relevant metabolite variations of different metabolic profiles and thus reveal statistically and potentially significant biomarkers responsible for the discrimination of *A. laurifolia* and *A. pedunculata* species. The two different organs were examined separately. In both cases, a clear separation of the different species was obtained (Figure 54A, Figure 55A). In order to define the variables that presented the strongest correlation with this classification, the S-plot, the OPLS-DA loadings plot was utilized (Figure 54B, Figure 55B). The S-plot visualizes the covariance (p) and correlation (pcorr) between the metabolites and the modelled class designation. Therefore, S-plot contributes at the identification of statistically significant and potentially biologically significant metabolites, based both on contributions to the model and their reliability, respectively.

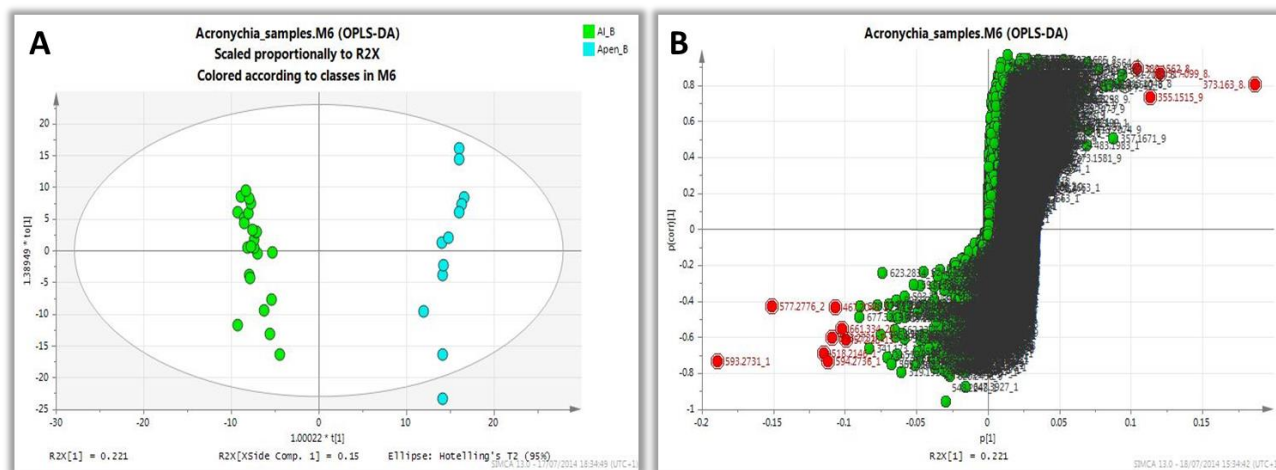


Figure 54: A. OPLS-DA score plot and B. S-plot obtained by UPLC-ESI(+)-HRMS dataset of *A. laurifolia* (in green) and *A. pedunculata* (in turquoise) bark samples; possibly discriminating features are annotated with red colour in the S-plot

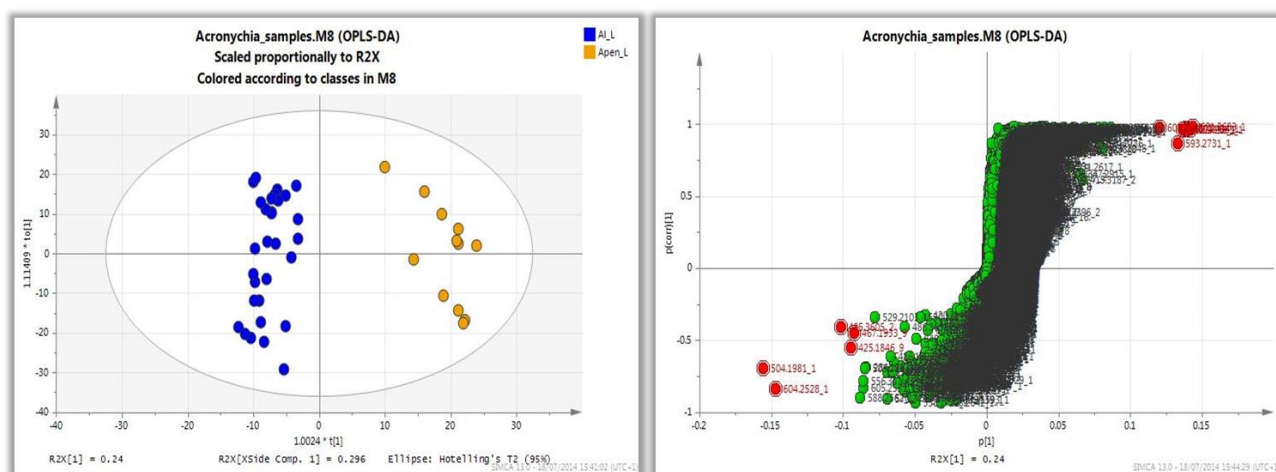


Figure 55: A. OPLS-DA scores plot and B. S-plot obtained by UPLC-ESI(+)-HRMS dataset of *A. laurifolia* (in blue) and *A. pedunculata* (in yellow) leave samples; possibly discriminating features are annotated with red colour in the S-plot

OPLS-DA despite the fact that is a very good method to argue classification between two groups, as a supervised method, is very prone to over-fit. Thus, before the interpretation of the potentially discriminant variables, the validation of the method is compulsory. Therefore, extensive model validation by means of cross validation technique was performed. Following this principal, each time a number of samples is removed from the model and the model is constructed based on the remaining data. The removed samples are predicted from the constructed model and this is performed until all the samples will be predicted once. In the current study, using SIMCA 13.0 software the two models were evaluated by assessment of the cross validated scores based on 7-fold cross validation. In particular, the 120 samples were divided into 7 groups randomly and each time the omitted samples were calculated based on the models constructed from the rest samples. These results were evaluated using the cross validated score plots (Figure 56). A good separation of the different species was also observed in the scores plot using the cross validated scores instead of the normal scores indicating the robustness of the model.

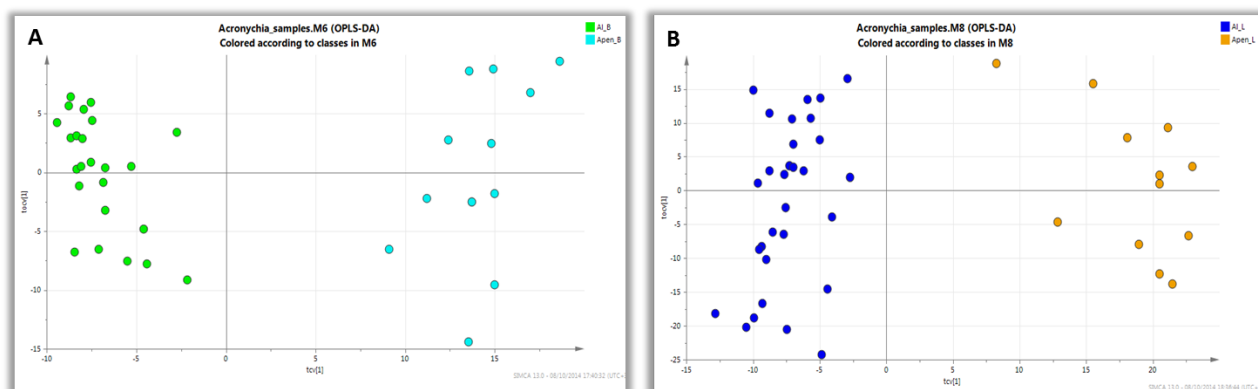


Figure 56: Score plots obtained using the cross validated score values concerning the UPLC-ESI(+)-HRMS dataset of *A. A. laurifolia* (in green) and *A. pedunculata* (in turquoise) bark samples and B. *A. laurifolia* (in blue) and *A. pedunculata* (in yellow) leaf samples

For the identification of the possible key metabolites contributed to the discrimination of the two species the variable importance in the projection (VIP) values derived from the OPLS-DA model were considered. VIPs are ranked according to the importance of each variable for the classification, and VIP values >1 are considered statistically significant. In addition, the correlation coefficients (pcorr) of the variables relative to the first model score value in the OPLS-DA model were also extracted from S-plot calculated by Pearson correlation. Cut-off values for $p(\text{corr}) < |0.7|$ were used to select metabolites that most strongly contributed to differences between two species. Thus, the integrals of metabolites which were meeting the VIP and correlation coefficients criteria may account for the discrimination. For the identification of the metabolites, the same strategy mentioned above was followed (Figure 53) resulting in the characterization of several biomarkers which can be used for the separation of the two species for each organ. This information is summarized in Table XVII.

In order to confirm the significance of the discriminating variables univariate t-test analysis were performed for both datasets using Metaboanalyst platform and the significant features extracted from the OPLS-DA analysis were further evaluated based on the FDR values (Bonferonni corrected P-values). The results were illustrated by box-whisker plots (Figure A 80, Figure A 81).

Table XVII: Specific biomarkers responsible for the discrimination of *A. laurifolia* and *A. pedunculata* bark samples

No	Rt	m/z	Adduct/Isotope	MS/MS	Loss	EC	RDB	Name	Category	Annotation level	FDR	Class
1	13.92	518.2146	[M+Na] ⁺			C28H33NO7	12.5		Alkaloid	4	1.04E-04	Al_B
2	13.84	572.2600	[M+Na] ⁺			C32H39NO7	13.5		Alkaloid	4	6.90E-04	Al_B
3	17.59	593.2731	[M+Na] ⁺	341.1725	C13H16O5	C32H42O9	11.5	Acrofolione A	AtA	1	9.04E-04	Al_B
	17.58	594.2736	M+1								8.87E-06	Al_B
4	8.43	317.0990	[M+Na] ⁺			C15H18O6	6.5			4	1.75E-09	Apen_B
5	8.20	373.1630	[M+Na] ⁺	301.105	C4H8O	C19H26O6	6.5			4	1.93E-07	Apen_B
6	8.89	389.1562	[M+Na] ⁺			C19H26O7	6.5			4	1.39E-10	Apen_B
7	9.38	355.1515	[M+Na] ⁺			C19H24O5	7.5			4	7.67E-06	Apen_B

Table XVIII: Specific biomarkers responsible for the discrimination of *A. laurifolia* and *A. pedunculata* leave samples

No	Rt	m/z	Adduct/Isotope	MS/MS	Loss	EC	RDB	Name	Category	Annotation level	FDR	Class
1	11.72	504.1981	[M+Na] ⁺	275.0897	C14H15NO2	C27H31NO7	12.5			4	9.92E-05	Al_L
2	10.32	604.2528	[M+Na] ⁺	572.226	CH4O	C32H39NO9	13.5			4	5.27E-09	Al_L
3	11.44	639.2756	[M+Na] ⁺			C33H44O11	11.5			4	3.06E-19	Apen_L
4	11.79	609.2663	[M+Na] ⁺	357.1674	C13H16O5	C32H42O10	11.5	AtA OH type-L	AtA	3	1.59E-21	Apen_L
5	13.25	607.2494	[M+Na] ⁺			C32H40O10	12.5		AtA	4	1.59E-21	Apen_L
6	17.00	621.2683	[M+Na] ⁺			C33H42O10	12.5			4	4.35E-23	Apen_L
7	17.59	593.2731	[M+Na] ⁺	341.1725	C13H16O5	C32H42O9	11.5	Acrofolione A	AtA	1	3.31E-11	Apen_L

Despite the close relation of *A. pedunculata* and *A. laurifolia*, according to OPLS-DA analysis based on the UPLC-ESI(+)-HRMS data these two species are differing significantly as far as the metabolite composition of bark and leave samples are concerned. This finding supports that *A. pedunculata* and *A. laurifolia* are different species presenting important differences in key metabolites concentrations. Particularly a number of secondary metabolites found to contribute significantly to this classification such as AtA and alkaloid compounds. However, it is important to take into consideration the different origin of the two species. *A. pedunculata* samples were collected in Vietnam while *A. laurifolia* samples were harvested in Malaysia. Thus, the variation observed in those two *Acronychia* species may be arise from the growing conditions in different locations and reflected to their different metabolite composition.

5. Statistical integration of different metabolomics techniques for the identification of metabolites

Currently, in plant metabolomics a number of different analytical techniques are applied for the metabolite profiling of plant extracts resulting in the reliable characterization of the metabolome (Hall 2006). Among them, the most popular are NMR and MS methodologies providing complementary information for the identification of biomarkers (Dunn et al. 2005). Nevertheless, in plant metabolomics field, the identification of biomarkers constitutes the main bottleneck of this high throughput procedure. The complex nature of plant secondary metabolites renders the identification procedure of biomarkers a difficult task. In addition, the great number of plant metabolites including the potentially new natural products in plant extracts has hindered the establishment of a universal database containing all the spectroscopic and spectrometric data (Wolfender et al. 2013). Thus, the structure elucidation of biomarkers in complex mixtures is still a time consuming and user dependent step (Moco et al. 2007). In traditional natural products chemistry studies, the structure elucidation of pure isolated metabolites is performed using mainly NMR and MS data. In these cases, the combination of these two techniques is implemented “manually” as the data are referring to the same compound. However, in metabolomics studies this is not applicable. The plant extracts are analysed as complex mixtures, hence the direct integration of both techniques with the respect of characterization of single molecules is not feasible. On the top of that, metabolomics

generate a large amount of data which impose the use of multivariate analysis methods for their handling.

The integration of two datasets by statistical methods is usually referred to the literature using projection based methods such as PLS, OPLS and O2PLS regression methods (Richards et al. 2010). The two datasets are denoted as X and Y and after the fitting the relation of the respective variables is assumed based on their correlation and covariance. In these cases, the obscure part of the analysis is the interpretation of the results since the variable selection of the biological related information is non-trivially accurate due to multicollinearity phenomena (Næs et al. 2001). A novel computational methodology called “sparse PLS” (sPLS) is introduced for a predictive purpose analysis and is used to match the information from two datasets and unravel correlated features (Le Cao et al. 2008). In metabolomics studies the datasets are characterized by high dimensionality containing noise and multicollinearities, while the sample replication is always restricted especially comparing with the number of generated features. The application of sPLS utilizes the properties of PLS regression analysis maximising the covariance between each linear combination (components) associated to each data set. Additionally, its unique characteristics include the variable selection from both data sets based on soft-thresholding penalization (Lasso) on the loading vectors. This penalization approach of sPLS loadings contributes to overcome multicollinearity issues and leads to the extraction of the more relative biologically related correlations removing noise interferences by shrinking their coefficients towards zero (Le Cao et al. 2011). Therefore, this method permits the integration of two datasets and the selection of variables in one step procedure following an unsupervised approach.

In this context, Dejean et al have built mixOmics package implemented in R environment which includes sPLS algorithm and a number of graphical outputs to visualize the results (Dejean et al. 2014). In metabolomics studies, adaptable visualisation techniques constitute a crucial prerequisite to unravel the biological information by the high dimensionality of the generated data. In this package a number of graphical outputs are available to resume the results and obtain a better insight into the relationship between two datasets.

5.1. Integration of NMR and MS datasets of *Acrorychia* extracts samples using sPLS

In plant metabolomics studies, as previously mentioned the identification of biomarkers is a crucial step to unravel the biological information from the complex datasets. Therefore, the integration of different information from multiple techniques is really valuable. In this context, the exploration of the correlation between NMR and MS data of *Acrorychia* extracts was performed in order to get a better insight into the structural characteristics of biomarkers. The integration of the measurements from both techniques was executed using sPLS algorithm incorporated in mixOmics package. In particular, for the X matrix the dataset obtained from the pJRES data of *Acrorychia* samples was selected. Concerning the Y matrix, the features generated by UPLC-ESI(+)-HRMS corresponding to the biomarkers were selected since the handling of the entire dataset was too much complicated. For the optimal implementation of sPLS analysis, two parameters had to be tuned, the number of dimensions (sparsity degrees) and the number of variables to be selected. Concerning the number of dimensions, the selection was based on the predictive ability of the model by performing cross validation calculations, evaluating the Q^2 values for each dimension (Le Cao et al. 2008). In the current study, initially a number of 10 dimensions was chosen and after a 10 fold cross validation approach and taking into account the proposed cut-off of 0.0975, the application of 4 sparsity degrees for the analysis was decided (Figure A 82). The number of variables to be selected is more challenging issue given the complexity of the two datasets. According to the literature, variable selection in sPLS is performed based on the biological information that needs to be answered and the experience of the scientist on the dataset (Le Cao et al. 2008). Only in cases of sPLS discriminant analysis specific criteria have been proposed for the selection of the number of variables (Le Cao et al. 2011). In our case, concerning *Acrorychia* extract samples, an arbitrary selection of 50 variables for each dimension for the X matrix was decided while all variables of the Y matrix were accounted for the analysis, considering it appropriate for the next interpretation step.

The evaluation of the model was performed using a 10 fold cross validation procedure. The predictive power of the model was estimated by the cumulative Q^2 (predicted variation) for the 4 components which was calculated for 0.977 indicating the reliability of the model to predict the UPLC-HRMS features from the NMR data.

5.2. Identification of metabolites based on the sPLS model of NMR and MS datasets

In order to explore the associations between the NMR and MS datasets the sPLS loadings were exploited. The visualization of the correlated loadings from the X and Y datasets was performed initially by the examination of the correlation circle plot. In this plot, the sPLS variables are represented as vectors projected into a space spanned by two (2D plot) or three (3D plot) latent variables, inside a circle of radius 1. The relationship between the variables is denoted by their relative position on the space. The angle between two variables is representing the type of correlation; for instance a very sharp angle denotes a positive correlation, an obtuse angle a negative correlation and when the vectors of two loadings are placed perpendicularly it means that no correlation is presented between them. In addition, the length of the vectors, their relative distance between the centre and the circumference of radius 1, is characteristic for the strength of correlation. Usually in the 2D correlation circle plots two circles are designed to assist the interpretation, the external one of radius 1 and another one of radius 0.5 in order to visualise better the more important variables (Gonzalez et al. 2012). Specifically, in the correlation circle plot of *Acronychia* samples in the first dimension a group of correlated X (red elements) and Y (blue elements) variables is clearly observed (blue dashed rectangle) (Figure 57A). The strong correlation between the variables is also denoted by their relative position close to the circumference of the external cycle. In the second dimension, a group of strongly correlated X and Y variables was detected close to 1 (green dashed rectangle) together with three other groups of variables that were presented less correlated (Figure 57A). Interestingly, in the third dimension a group of variables was revealed to have strong positive correlation (yellow rectangle) despite the fact that was not obvious from the first two dimensions. For this reason, the 3D correlation circle plot was visualized enhancing the interpretability of the model (Figure A 83).

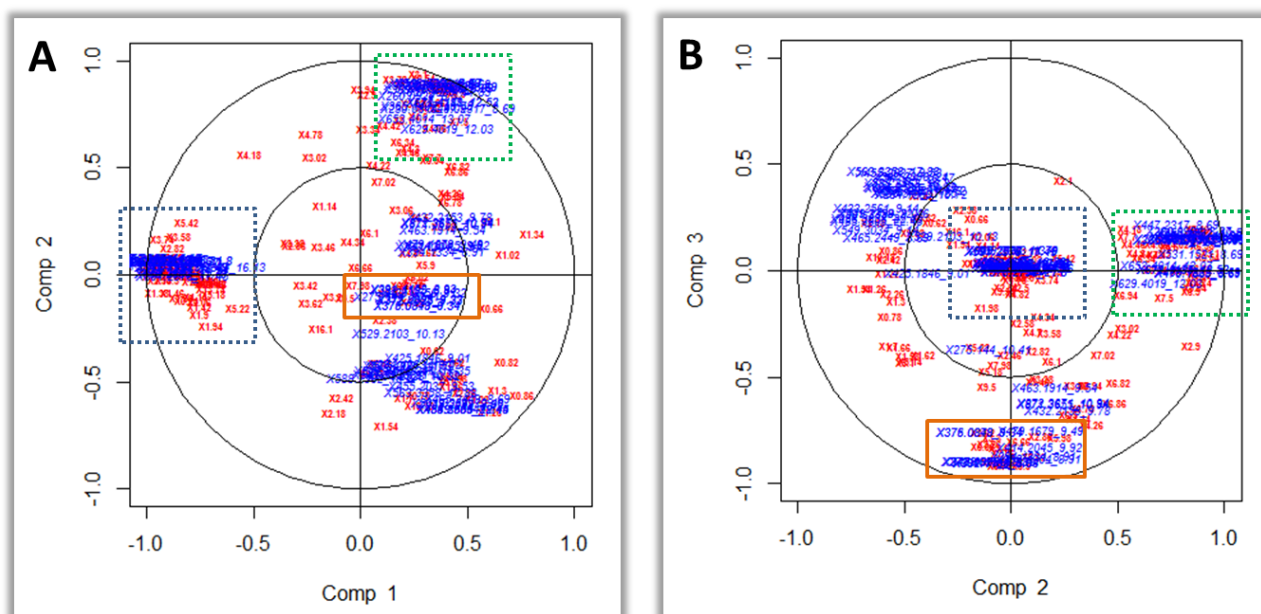


Figure 57: Correlation Circle plots for dimensions A. 1 and 2 and B. 2 and 3 obtained from the sPLS analysis of pJRES (red elements) and UPLC-ESI(+)-HRMS (blue elements) data of *Acronychia* samples

The correlation circle plots contributed significantly to the evaluation of the sPLS results by presenting the general picture of the correlated variables. However, the extraction of detailed information from this kind of representations is not straight forward for the majority of cases. A more efficient representation of the correlations of the specific features from both datasets can be achieved using clustered image maps (CIM). CIM is based on the hierarchical clustering of the discrete datasets combined with a coloured heatmap indicating the correlation between subsets of variables (Gonzalez et al. 2012). Therefore, information regarding the proximity between correlated variables can be extracted and utilized for the identification of biomarkers.

In order to evaluate the results obtained from the sPLS analysis of *Acronychia* samples, in the first step the correlation of the MS features with the respective NMR signals of the identified biomarkers was investigated. In particular, a principal cluster at the left top of the CIM including identified AtA compounds (e.g. metabolites 47, 50, 51, 54-57 of Table XVI) presented strong correlation with NMR signals that correspond to these chemical structures (Figure 58). Specifically, downfield shifted signals at δ 15.26, 13.26-13.62 were revealed corresponding to the hydroxyl groups of these molecules that forms hydrogen bonds. Moreover, signals at δ 5.18, 5.22, 3.22, 3.68, 1.35, 1.38 and 1.46 were found to be correlated with the AtA MS features corresponding to the protons of the two isoprenyl

groups. Two signals at δ 2.5 and 2.7 were detected matching with the protons of the acetyl groups. And finally characteristic signals at δ 2.03 and 4.82 corresponding to the isopentyl chain were also observed to be correlated with the MS features of AtA. Another compound that was successfully identified was yangambin. Strong correlations were revealed between the MS features of yangambin and the aromatic protons at δ 6.54 as well as with the protons of the hydrofuran rings at δ 4.73, 4.5, 3.9 and 3.1. The protons of the methoxyl groups were also traced at δ 3.83 and 3.78. Interestingly, a cluster corresponding to flavonoid compounds as deduced from HRMS features presented strong correlation with NMR signals at a range of δ 5.9-7.2 and some of them presented also correlation with signals at 3.6 and 4.26 probably corresponding to methoxy groups on the structures.

The above mentioned findings support that sPLS analysis may successfully be applied for the integration of MS and NMR datasets obtained from different analytical metabolomics platforms acquiring the same samples. This integration is of valuable importance for the annotation or structural elucidation of biomarkers which constitutes a bottleneck in the overall plant metabolomics pipeline. The interpretation of this information necessitates specific visualization tools such as clustered image maps (CIM). The CIM revealed clusters attributed to specific categories of compounds according to both MS and NMR signals providing an increased level of confidence for their identification.

Chapter 3: Metabolomics approaches in *Acronychia* genus

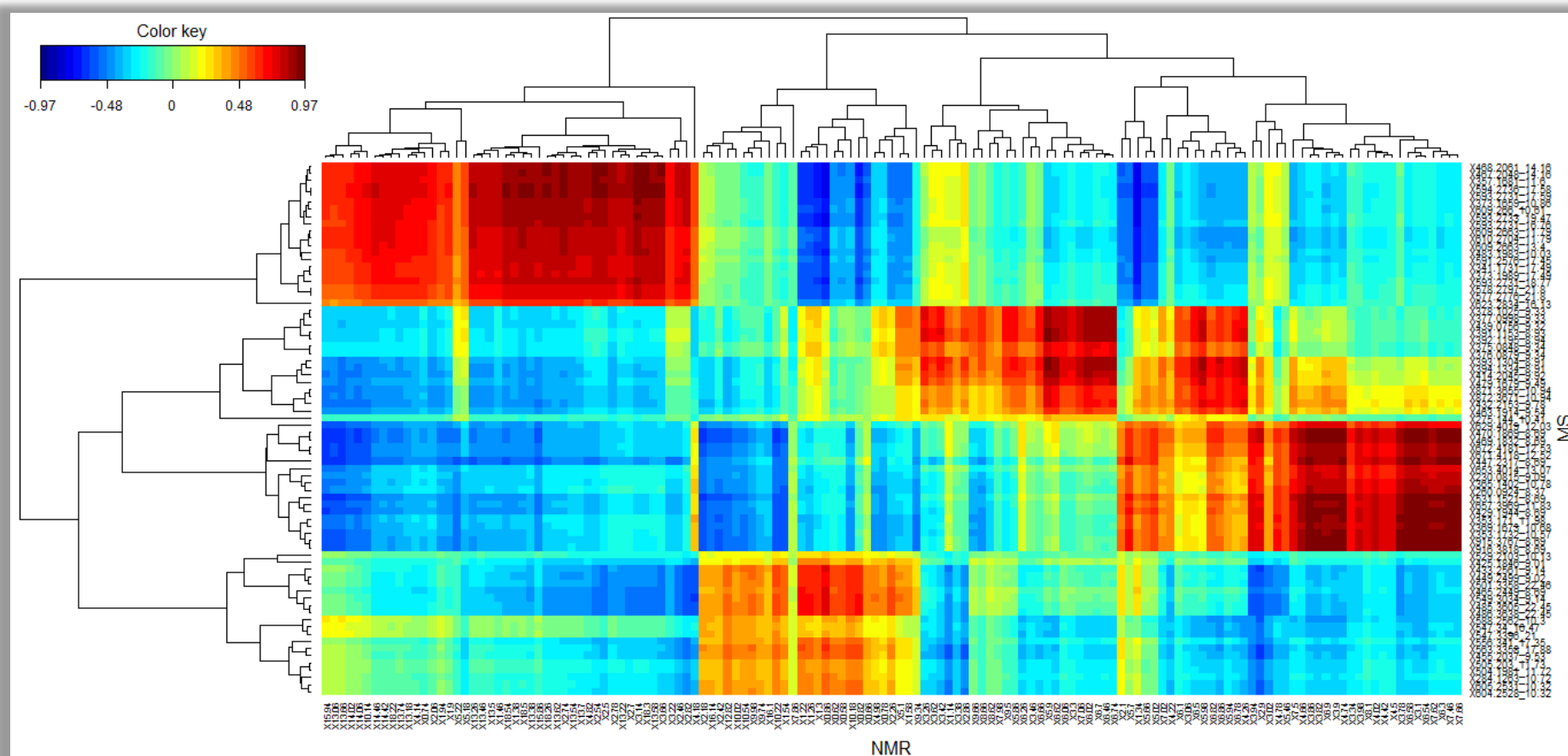


Figure 58: CIM on the *Acronychia* pJRES and UPLC-ESI(+)-HRMS datasets analysed with the sPLS; the red and blue colours indicate positive and negative correlations respectively, whereas yellow or light blue indicate small correlation values

6. Discovery of 5-LO inhibitors from *Acronychia* samples using UPLC-HRMS based metabolomics and PLS regression

Medicinal plants constitute a rich source of bioactive compounds which has been exploited during the years resulting in the launch of important drugs in the market (Balunas et al. 2005, Mishra et al. 2011, Dias et al. 2012). However, the discovery of pharmacologically active metabolites in plant extracts is a complicated task. The main characteristic of plant extracts is their complex nature. A wide array of secondary compounds present in these complex mixtures co-exist contributing to significant activity. Despite the clear advantage of the natural products as drugs due to their unique structures (Newman et al. 2012), the tracing of bioactivity in complex mixture remains a challenging issue and the improvement of the drug discovery methods is compulsory (Butler 2004). Toward the direction to replace methodologies including laborious and time consuming processes multiple successful drug discovery strategies have been proposed for the tracking of bioactivity in plant extracts focusing on individual active compounds using mainly reductionist approaches (Michel et al. 2013, Potterat et al. 2013). However, it is increasingly supported that multiple active compounds are hidden behind any given activity of medicinal herbal preparations and often synergistic or antagonism phenomena are taking place (Gilbert et al. 2003). In addition, another characteristic that complicates the drug discovery process is the presence of small amounts of highly active compounds in herbal preparations that are not detected due to moderated activity that is reflected to the whole mixture. Thus, the necessity of holistic approach to comprehend the contribution of each compound or the interaction of the compounds in the overall bioactivity of the complex mixture prior to any isolation step is of great impact for the drug discovery effort. Recently, the introduction of metabolomics techniques in drug discovery process has been proven a great tool towards this direction (Yuliana et al. 2011, Wolfender et al. 2012). Metabolomics profiling approaches utilizing state of art analytical and spectroscopic techniques provide a broad insight into the metabolite composition of a given complex sample. In parallel, high throughput bioassays deliver a pharmacological activity profile for these complex mixtures. By applying chemometrics data analysis the combination of the metabolic profile with the pharmacological activity profile of the different samples is performed unravelling the relation of certain marker compounds with this activity.

Chapter 3: Metabolomics approaches in *Acronychia* genus

To address this issue an appropriate MVDA method has to be employed. PLS-regression (PLS-r) is a supervised method specifically established to make good predictions in multivariate problems. Its goal is to analyse or predict a set of dependent variables from a set of independent variables or predictors. This method attempts to find an optimal decomposition of the predictor dataset given a matrix of responses (Bartel et al. 2013). Applying this technique, the dimensionality of the data acquired with various analytical techniques is reduced by combining correlated variables to form latent variables. By means of the new latent variables, the response in the dataset can be explained in relation to one or more variables (Abdi 2010). PLS-r has proven to be a very versatile method for multivariate data analysis and the number of applications is steadily increasing in research fields like bioinformatics, machine learning and chemometrics (Mehmood et al. 2012). However, lately the modification of the PLS-r analysis method led to the introduction of the orthogonal projections to latent structures (OPLS). The main idea of the OPLS method is to separate the systematic variation in the X variable into two parts that which is linearly related to Y (Y-predictive) and that which is orthogonal to Y (Y-orthogonal). This gives rise to a much better interpretability as the orthogonal variation is not accounted for the prediction (Trygg et al. 2002). Nevertheless, it is well known that the predictions of the single response OPLS and the single response PLS-r result in identical regressions (Verron et al. 2004, Indahl 2014).

The application of metabolomics approaches for the discovery of bioactive compounds is an emerging field growing the last years. Up today only 6 publications have been found to address this issue. In all cases, NMR based metabolomics data are used as independent variables and are correlation with a pharmacological activity (response). In particular, only 3 publications among them are applying regression models to predict the activity and identify the bioactive metabolites by relating the response with more correlated metabolites (Roos et al. 2004, Cho et al. 2009, Yuliana et al. 2011). The rest 3 literature data are dealing with this problematic using discriminant analysis and revealing the bioactive compounds by matching them with the most important discriminating variables. In the current study, for the first time the correlation of UPLC-HRMS data with a pharmacological activity is reported constructing both PLS and OPLS regression models to predict the activity of new samples and to reveal bioactive compounds.

In this context, different organs and different species of the genus *Acronychia* have been selected in order to trace pharmacologically active metabolites. *Acronychia* species

have been traditionally used in folk medicine for their anti-inflammatory and antipyretic effects to treat asthma, ulcers and rheumatism (Epifano et al. 2013). Based on the traditional use of *Acronychia* species, 5-lipoxygenase (5-LO) enzyme was selected to evaluate the anti-inflammatory potency of the obtained extracts. 5-LO is a key enzyme involved in the inflammation and allergy process through catalysis of the first step in the biosynthesis of leukotrienes (LTs) from arachidonic acid. Based on the multiple potent pathophysiological actions of LTs in respiratory and cardiovascular diseases, the pharmacological intervention with 5-LO is a challenge in the development of new therapeutics (Pergola et al. 2010) (see Chapter 1 Results and Discussion 8 for further details).

6.1. PLS and OPLS regression model for prediction of 5-LO inhibition

In a first step, in order to get an insight into the correlation of the chemical profile of multiple *Acronychia* samples and their capacity to inhibit 5-LO enzyme (Table A 17) a PLS multivariate calibration method was applied which utilizes two blocks of data sets the independent X, in this case the UPLC-ESI(+)-HRMS data, and the dependent Y, % inhibition of 5-LO data at a concentration of 1 µg/mL, and relates them using regression (Wold et al. 2001). Using Simca 13.0 software, an optimum number of 6 latent variables was defined for the construction of a significant and robust model explaining the 69% of the total variance of X dataset and the 97.5% of Y dataset (Figure A 84). The PLS-r scores plot (Figure 59) illustrates the correlation between X (t scores) and Y (u scores) datasets as explained from the first component accounting for the 22.7% and 44.3% of the total variance, respectively. U scores, correspondingly to t scores, are representations of the observations in the Y space as situated on the projection plane. The interpretability of model using the regression line is not easy in PLS-r models as the scores represent the global variation of the dataset, including the variation that is not related to the Y dataset. Therefore, as observed in the scores plot the overall correlation of the X and Y dataset is not very strong ($R^2 = 0.44$) indicating the presence of orthogonal variation on X dataset which is not related to Y dataset. This is totally comprehensive for a complex mixture of compounds such as extracts containing metabolites possessing a potential 5-LO inhibition capacity and others which are not interacting with 5-LO enzyme.

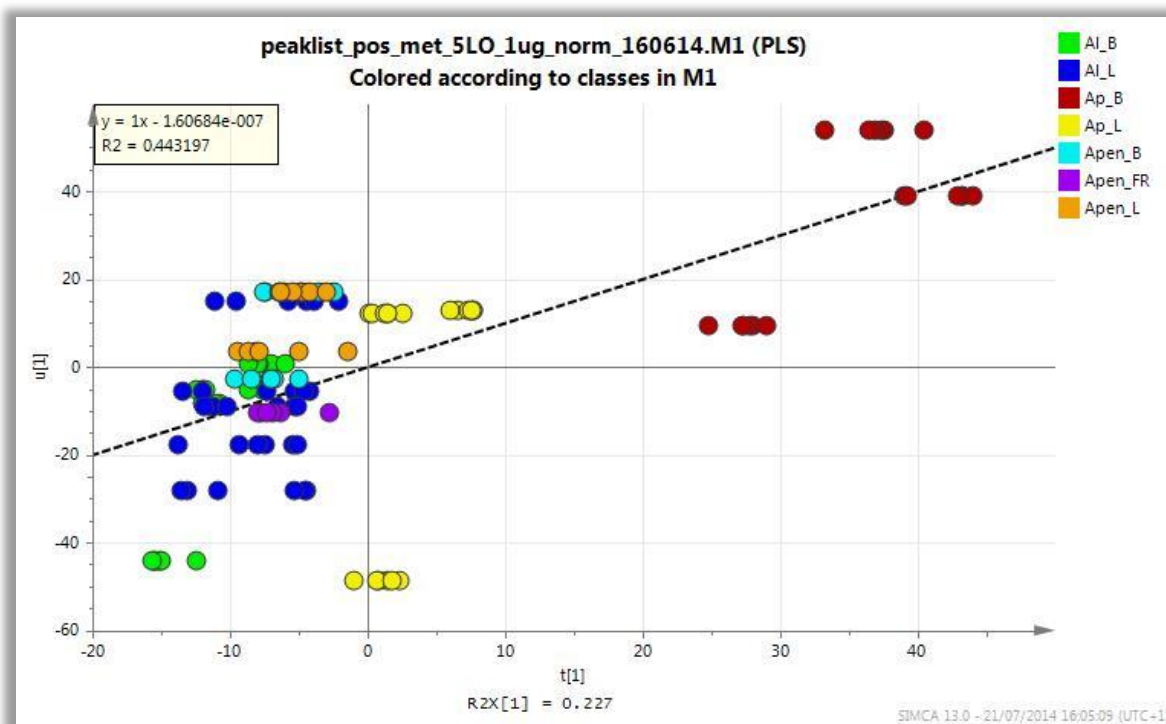


Figure 59: PLS-r scores plot (T1/ U1) obtained from the UPLC-ESI(+)-HRMS data and the response dataset expressed as the % inhibition of 5-LO; *A. laurifolia* barks are coloured in green, *A. laurifolia* leaves in blue, *A. porteri* barks in red, *A. porteri* leaves in yellow, *A. pedunculata* barks in turquoise, *A. pedunculata* fruits in purple and *A. pedunculata* leaves in orange

Since PLS-r is a supervised multivariate method, an extensive validation of the model has to be performed. A quantitative measure of the goodness of fit of the model is given by R^2X and R^2Y (explained variation), indicating the amount of X and Y variables explained by the model, respectively. More important than fit, however, is the predictive ability of the model. The predictive power of the model is represented by Q^2 (predicted variation) which is estimated using cross-validation values to calculated how accurately the dependent variable is predicted by a given set of independent variables. Therefore, Q^2 value has to be arbitrarily close to one (indication prediction ability) (Eriksson et al. 2006). In the current model, the cumulative Q^2 for the 6 components were calculated for 0.959 demonstrating that using this model the pharmacological response can be predicated with accuracy from the UPLC-ESI(+)-HRMS data.

In addition, as PLS models tend to over-fit the assessment of the model significance was performed using the results obtained from the response on a 100 permutation test. The order of the Y variables is randomly permuted and each of the randomized group generates

a new set of R^2 and Q^2 values, which are plotted against the correlation coefficient between the original Y values and the permuted Y values in the SIMCA software (Figure 60) (Eriksson et al. 2006). The intercepts for R^2 and Q^2 lines in this plot are a measure of the over-fit. A model is considered valid when $R^2 \text{ int} < 0.4$ and $Q^2 \text{ int} < 0.05$ (Eriksson et al. 2002). In this case the R^2 and Q^2 intercepts were calculated for 0.324 and -0.675 indicating that the model was not over-fitting.

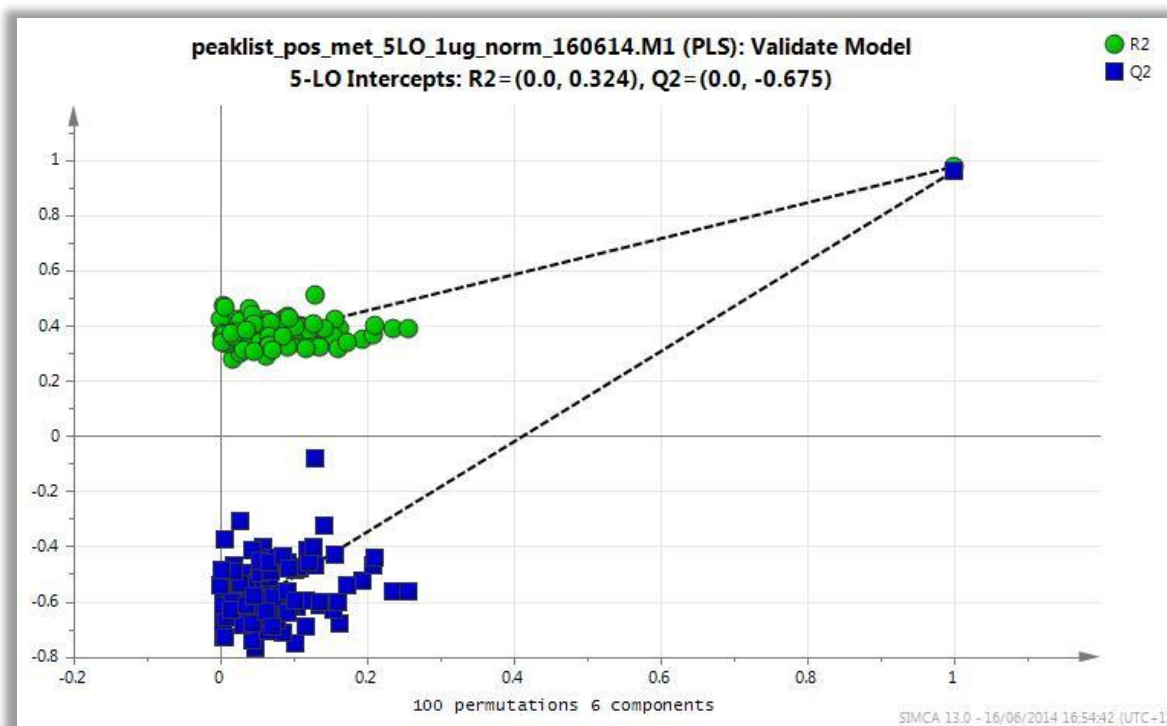


Figure 60: Results from the response permutation test; the vertical axis gives the R^2 and Q^2 values of each model; the horizontal axis represents the correlation coefficient between the original Y and the permuted Y; the PLS model is significant since the R^2 and Q^2 of the original model are always larger than the corresponding values of the models fitted to the permuted responses.

A representative plot for the PLS-r model is the regression plot of the original Y-variables and the predicted Y-variables which gives an estimation of the ability of the model to predict the capacity of new samples to inhibit 5-LO (Figure 61). The strong correlation of the original and the predicted Y-variables as illustrated in the PLS regression plot of this model indicate that this model can be used for the evaluation of the 5-LO inhibition. In addition, the RMSEE (Root Mean Squared Error of Estimation) represents one standard deviation in the metric of the Y variable. With a RMSEE of 1.77 at a range of 13- 60 % 5-LO product formation of the control a reliable prediction is possible.

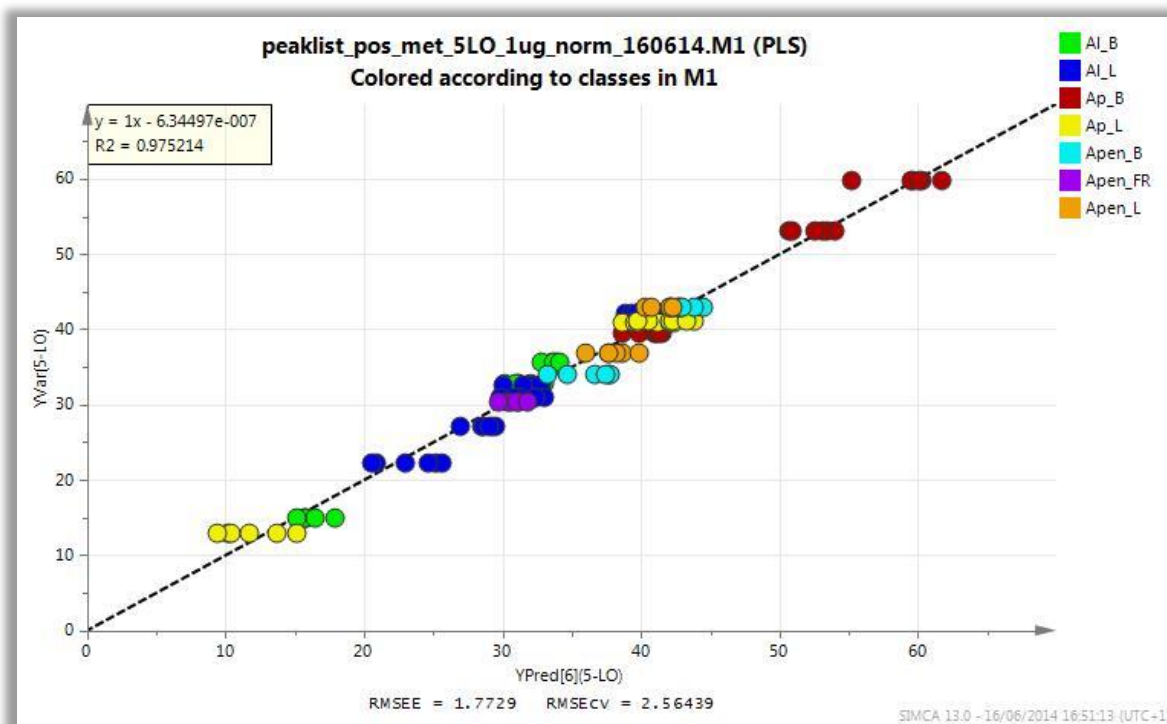


Figure 61: Regression plot of the original Y-variables (YVar) and the predicted Y-variables (Ypred) based on the PLS-r model; *A. laurifolia* barks are coloured in green, *A. laurifolia* leaves in blue, *A. porteri* barks in red, *A. porteri* leaves in yellow, *A. pedunculata* barks in turquoise, *A. pedunculata* fruits in purple and *A. pedunculata* leaves in orange

For the sake of interpretation reasons, the same dataset was analysed by OPLS. Since single Y OPLS method is an extended version of single Y PLS in which the PLS model is rotated, placing the Y-predictive part of the model in the first component. Therefore, the first component is much more interpretable comparing to PLS. In addition, PLS and OPLS with the same number of components give identical prediction (Eriksson et al. 2008). As a matter of fact, in the OPLS score plot of the first component, the correlation of the spectrometric data and the response values was well described by a linear relationship ($R^2 = 0.98$) since only the correlated X-variables were accounted for this model (Figure 62), in contrast to the PLS-r model where the total variation of the X dataset was accounted (Figure 59). Specifically, the analysis overview of the OPLS model demonstrated that a large amount (62.8%) of the X dataset (spectrometric data) was orthogonal to the Y dataset (pharmacological response). In this case that the orthogonal variation was not taken into account, the Y variance found strongly correlated to the X-predictive variance ($R^2_{\text{cum}} = 0.983$) (Figure A 85). The extensive validation of the PLS-r model indicated a good predictive ability of the Y dataset from the X dataset, however, validity of the OPLS model was also

demonstrated plotting the cross validated scores of the X-predictive variables of the UPLC-ESI(+)-HRMS dataset and the Y scores of the 5-LO inhibition response (Figure A 86).

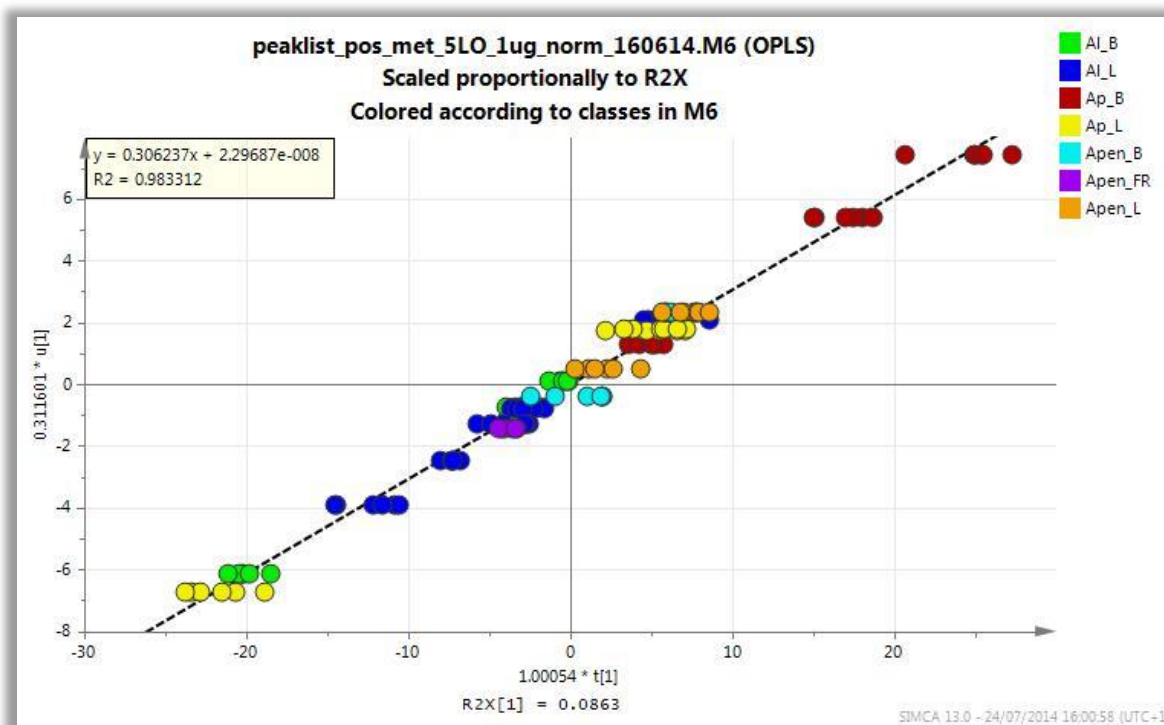


Figure 62: OPLS scores plot (T1/ U1) obtained from the UPLC-ESI(+)-HRMS data and the response dataset expressed as the % inhibition of 5-LO; *A. laurifolia* barks are coloured in green, *A. laurifolia* leaves in blue, *A. porteri* barks in red, *A. porteri* leaves in yellow, *A. pedunculata* barks in turquoise, *A. pedunculata* fruits in purple and *A. pedunculata* leaves in orange

6.2. Identification of bioactive compounds by PLS and OPLS analysis

In order to reveal the bioactive metabolites the variable importance in the projection (VIP) and the correlation coefficient values derived from the PLS-r and OPLS model were exploited. VIPs are ranked according to the importance of each variable for the regression, and generally those variables with $VIP > 1$ are considered statistically significant. Consequently, VIPs calculated from PLS-r and OPLS analysis may vary significantly since in the first case the entire X dataset is accounted for the regression while in the second case the orthogonal X variables are not taken into account. Moreover, the correlation coefficients obtained from both analyses should not differ, since the predictions of the single response OPLS and the single response PLS-r are identical. In contrast to VIPs which are positive values, correlation coefficients give additionally the information of positive or negative

Chapter 3: Metabolomics approaches in *Acronychia* genus

correlation. Actually, differences concerning VIPs and correlation coefficient between PLS and OPLS were illustrated for the current study generating the correlation coefficient line plots obtained from both analyses and variables with VIP values >1 were annotated in red colour (Figure 63).

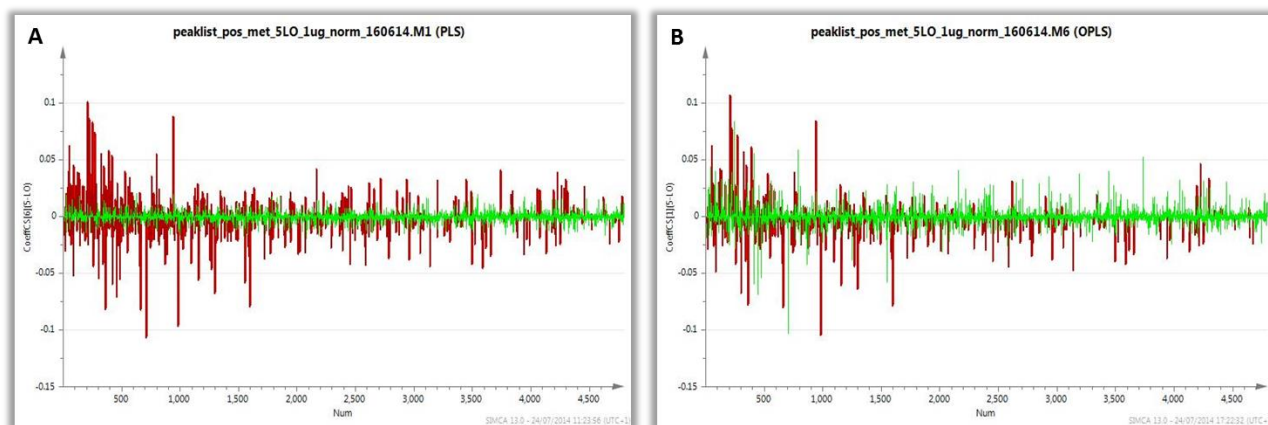


Figure 63: Correlation coefficients line plot of all variables obtained from the A. PLS and B. OPLS model; the red lines indicate the features with VIP >1

In this context, the correlation coefficients from each analysis were extracted in order to evaluate the strength and the direction of the relationship between the each metabolite and the 5-LO inhibition capacity. Correlation coefficient cut-offs were calculated for a significance level of 0.005, based on the FDR values (Bonferonni corrected P-values), in order to identify the features presenting significant correlation. Overall, the features which met this criterion along with VIPs > 1 were identified as markers of potential bioactive compounds (Figure A 87, Figure A 88). The results were resumed in Table XIX presenting the compounds which corresponded to the major contributing features and their main occurrence in the dataset.

Table XIX: Major contributing features attributed to possibly bioactive compounds; compounds 5 and 7 are referred to Table XVII and compounds 4, 16 and 17 to Table XVI

Feature	Compound	VIP	R	Occurrence
260.0924@8.37	Skimmianine	12.8668	0.10742	Ap_B
	Maculosidine			
	Kokusaginine			
355.1515@9.38	7	4.35349	0.084992	Apen_B
447.1992@10.42		2.27167	0.078178	
373.163@8.2	5	1.16067	0.072091	Apen_B
230.0815@9.09	Pteleine	11.5305	0.06307	Ap_B
	Evolitrine			
	g-fagarine			
369.1673@10.68	17	13.5231	0.06182	Ap_B
375.0848@9.34	4 (flavonoid)	1.20603	0.057702	Ap_L
353.1732@10.57	16	9.95771	0.057572	Ap_B
577.2776@21.8	Acrovestone	2.4245	0.046955	Al_B
915.3767@8.69	Yangambin	12.8117	0.045634	Ap_B
469.1832@8.69	Yangambin	17.0701	0.04269	Ap_B
261.0954@8.37	Skimmianine	5.02794	0.040975	Ap_B
	Maculosidine			
	Kokusaginine			

After the evaluation of the VIPs and the correlation coefficients obtained from both PLS and OPLS analysis 12 features among the thousands features describing the metabolic profile of *Acronychia* samples were found to be strongly positively correlated with the activity dataset. In the same time, other features meeting the abovementioned criteria were found to be negatively related with the pharmacological activity and attributed to metabolites exhibiting antagonism effects. Interestingly, different chemical structural patterns were recognised among the positively correlated metabolites. Feature 260.0924@8.37, representing a furoquinoline alkaloid possibly skimmianine, maculosidine or kokusaginine, was identified as the most correlated one. In addition 261.0954@8.37 feature, corresponding to the M+1 ion of skimmianine, maculosidine or kokusaginine was extracted among the significant features indicating the robustness of the method. It is worth to note that also 230.0815@9.09 is attributed to a furoquinoline alkaloid which suggests that this group of compounds may present significant inhibition of 5-LO.

Chapter 3: Metabolomics approaches in *Acronychia* genus

Unfortunately, there is no literature data confirming this finding. Moreover, features 469.1832@8.69 and 915.3767@8.69 corresponding to yangambin, a symmetric lignan, have demonstrated significant positive contribution to the regression. This observation is also supported by literature data demonstrating significant inhibition of 5-LO by yangambin (Lim et al. 2009). Finally, acrovestone, an AtA derivative, was characterized by significant correlation. Acrovestone, was evaluated previously, and demonstrated potent inhibitory activity (IC_{50} = 2.7 μ M) against 5-LO (see Chapter1 Results and Discussion 8 for further details).

The identification of compounds with reported pharmacological activity consist a proof of concept that the integration of high accuracy, efficient and robust analytical techniques with the appropriate multivariate data analysis method can be applied to reveal bioactive compounds from complex mixtures such as plant extracts. The statistical and chemical validation of this method seems promising for the application of metabolomics approaches with the attempt to discover new target compounds. Metabolomics, correlating the chemical profile with the bioactivity prior to any isolation step, can contribute significantly to overcome problems such as discovery of known bioactive compounds or loss of bioactivity after fractionation steps occurring using time-consuming and ineffective traditional approaches. Overall, these techniques can enhance the efficacy of drug discovery in terms of time and quality of the results.

Conclusion

In this chapter, novel dereplication strategies and integrated metabolomics approaches were developed optimised and applied aiming to the identification of novel active compounds in complex mixtures such as plant extracts. As a proof of concept, different *Acronychia* samples were utilised based on literature data and knowledge obtained during previous studies. Different *Acronychia* samples belonging to diverse species, organs and different geographical locations were selected. In particular, 20 different biological samples (3 different species, 3 different organs and 2 locations) were collected and comprised the working material towards the development of a comprehensive metabolomics workflow incorporating NMR and LC-MS approaches.

The initial step of this effort was the development and optimisation of sample preparation procedures which are one of the most critical processes for any metabolomic workflow with strong impact to the validity and reliability of the data obtained. A unified extraction scheme was applied for the simultaneous preparation of the samples designated for NMR and LC-MS acquisitions along with their pharmacological assessment. Moreover, method validation steps for the entire procedure were performed for both platforms using analytical and statistical means giving new insight in plant metabolomics pipeline. Based on the developed methodology reliable datasets were obtained from this downstream analysis allowing their exploitation for the investigation of essential biological questions.

Much attention was also given to the utilisation of the different potentials offered by the analytical platforms used, aiming to the development of an integrated and complete metabolomic workflow. Thus, data from ^1H NMR, pJRES NMR and ^{13}C NMR experiments as well as data from UPLC-ESI(+)-HRMS/MS and UPLC-ESI(-)-HRMS/MS analysis were obtained and analysed, both in parallel and in combination. pJERS was proven a useful alternative for complex mixtures such as plant extracts offering higher spectra resolution while ^{13}C data can significantly contribute to the classification of samples but also in identification of metabolites of interest especially in combination with other techniques.

For the first time the combination of the UPLC-ESI($\pm$)-HRMS as well as the ^{13}C NMR datasets was utilized for the dereplication of compounds in all different plant samples giving

a better insight into their chemical composition and therefore, qualitative information related to the classification patterns. Furthermore, the hyphenation of data derived from both positive and negative ionisation mode in LC-MS analysis strengthened considerably the dereplication procedure. This strategy resulted to useful information regarding their value as metabolomic tools but also enabled the in depth characterization of *Acronychia* samples metabolite composition without any isolation step. Briefly, a total number of 33 known compounds belonging to acetophenones, alkaloids, lignans, terpenes and other phenolic compounds were determined in *Acronychia* samples by UPLC-ESI($\pm$)-HRMS with different contribution in the clustering patterns while new structures were proposed. It is important to highlight, that 8 *Acronychia* metabolites were also revealed by the ^{13}C NMR approach, verifying the UPLC-ESI($\pm$)-HRMS results.

Furthermore, since the identification of statistically significant metabolites responsible for the classification and/or discrimination of different samples (biomarkers) comprise a challenging task in any metabolomic study, a systematic effort towards this direction was also performed. Thus, apart the in-house databases (LC-MS) and the combination of data, investigation of public databases was also carried out. In this context, a total number of 57 metabolites were identified belonging to flavonoids, acetophenones, alkaloids, lignans and terpenes. Thus, the identification of the metabolites exploiting the maximum possible information obtained from both platforms based on in house or freely available databases is proposed. Moreover, in order to enhance identification reliability, the integration of UPLC-HRMS and NMR using bioinformatics tools is also suggested for the handling and interpretation of the large amount of data. Therefore, the application of sPLS analysis led to the confirmation of the structural identity of species and organ specific biomarkers. This approach constitutes a promising tool for the structural characterization of metabolites in plant metabolomics studies.

The multiple metabolomics platform was also utilised for the thorough investigation of chemotaxonomic issues associated with *Acronychia* species. The main goal was to investigate the metabolome of the different organs, species and plant origins but also to evaluate whether *A. pendunculata* and *A. laurifolia* are identical or comprise different species based on their metabolome analysis. In particular, NMR and LC-MS based metabolomics profiling approaches were used for the determination of variances between the analysed samples. The application of unsupervised PCA analysis allowed the exploration of discriminant features which revealed important species and organ specific biomarkers,

however, the discrimination of *A. laurifolia* and *A. pedunculata* samples wasn't achieved. Alternative statistical tools were thus employed, such as supervised OPLS-DA analysis, in order to further explore the relevant taxonomic issue. Indeed, OPLS-DA analysis of these species enabled the discrimination of the two species and the detection of 14 metabolites significantly different indicating that they probably comprise different species. It is worth noting that *Acronychia*-type acetophenones weren't detected in *A. porteri* samples leading to a serious doubt regarding their future utilisation as chemotaxonomic markers of the genus.

Furthermore, given the fact that identification of an active entity in a plant extract prior to any isolation and purification step encompasses a challenging and still unresolved issue in natural products chemistry, an attempt to identify anti-inflammatory compounds in *Acronychia* samples was also performed. Thus, a novel approach involving the correlation of the UPLC-HRMS data with the pharmacological responses of all extracts against 5-lipoxygenase (5-LO) was performed. Sophisticated statistical tools such as regression analysis (PLS and OPLS) were employed in order to disclose new 5-LO inhibitors while still being in mixture. Statistical and chemical validation of this approach was achieved identifying compounds with known anti-inflammatory activity such as acetophenones, together with new candidates e.g. lignans. Based on our results, metabolomics was proved a valuable tool to predict pharmacological activity in complex mixtures, track bioactive components and direct bioguided isolation under a new perspective.

Overall, this study constitutes a proof of concept of the successful application of dereplication and metabolomics techniques for taxonomic investigation and identification of bioactive compounds in total extracts. That aspires to introduce an alternative and holistic concept in natural products chemistry, in general. Probably, the main strength of this work is that several discriminant features characterising the different samples were revealed, many of them were identified, new ones were proposed while possible new 5-LO inhibitors were unravelled without the need of any fractionation, isolation and purification step.

APPENDIX

3

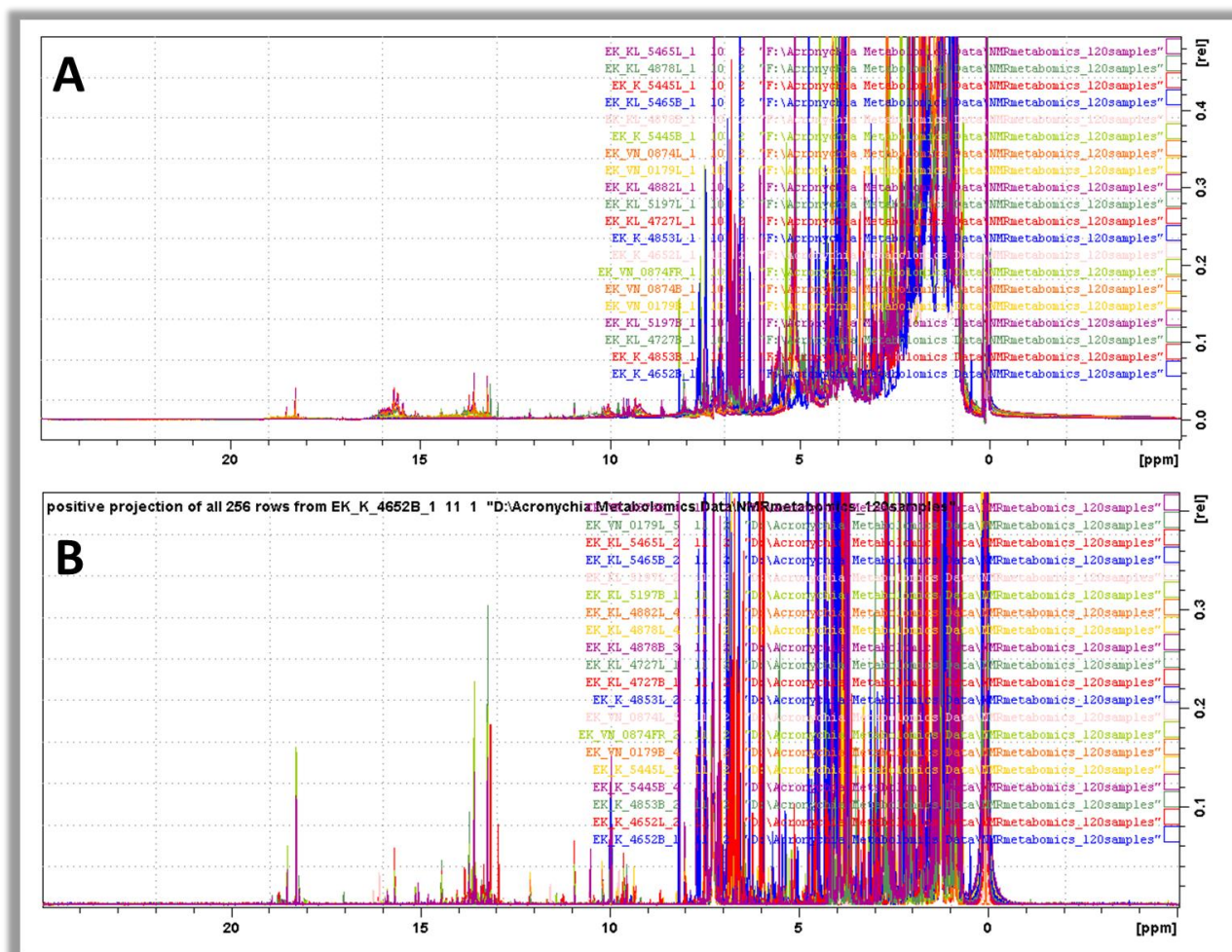


Figure A 73: Comparative overlay of A. ^1H NMR B. pJRES spectra of the 20 different *Acronychia* samples; enhanced resolution and less complexity are observed in pJRES spectra

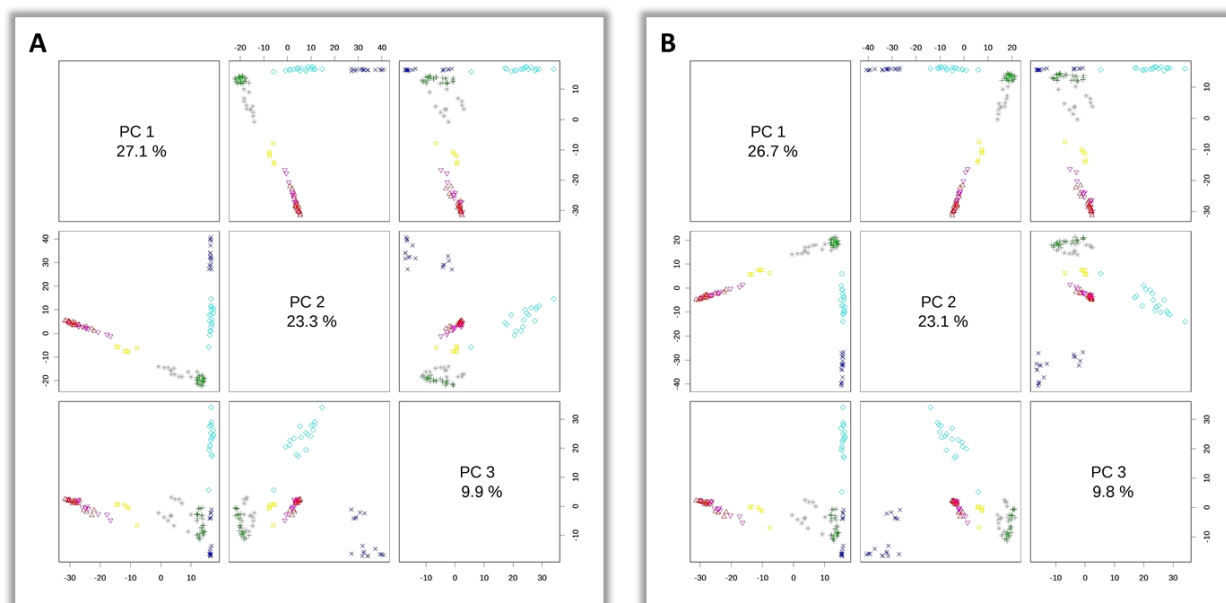


Figure A 74: PCA score plots using A. PPCA, BPCA or SCV Impute algorithms and B. KNN algorithm for the imputation of missing values obtained after the preprocessing of the UPLC-ESI(+)-HRMS dataset of *Acronychia* samples

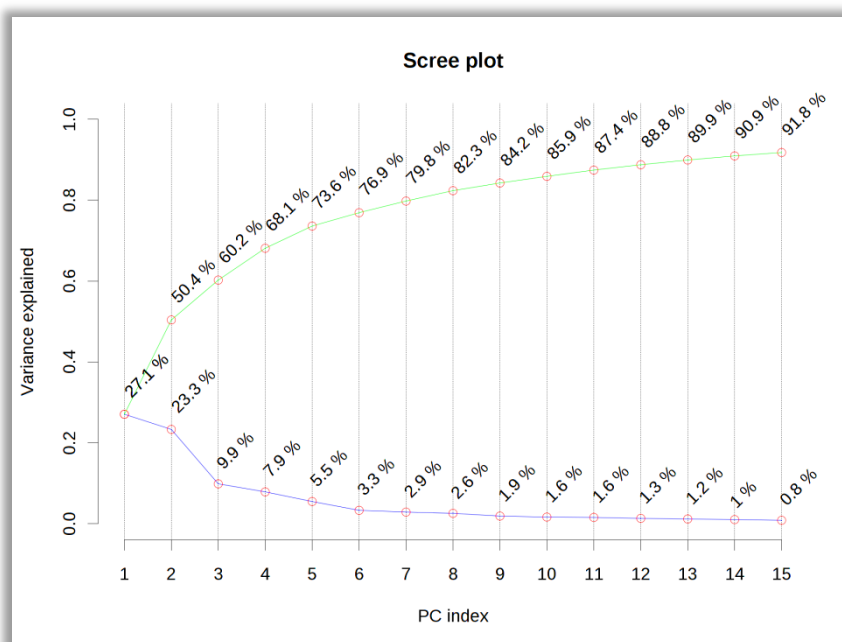


Figure A 75: PCA scree plot obtained using the UPLC-ESI(+)-HMS dataset of *Acronychia* samples

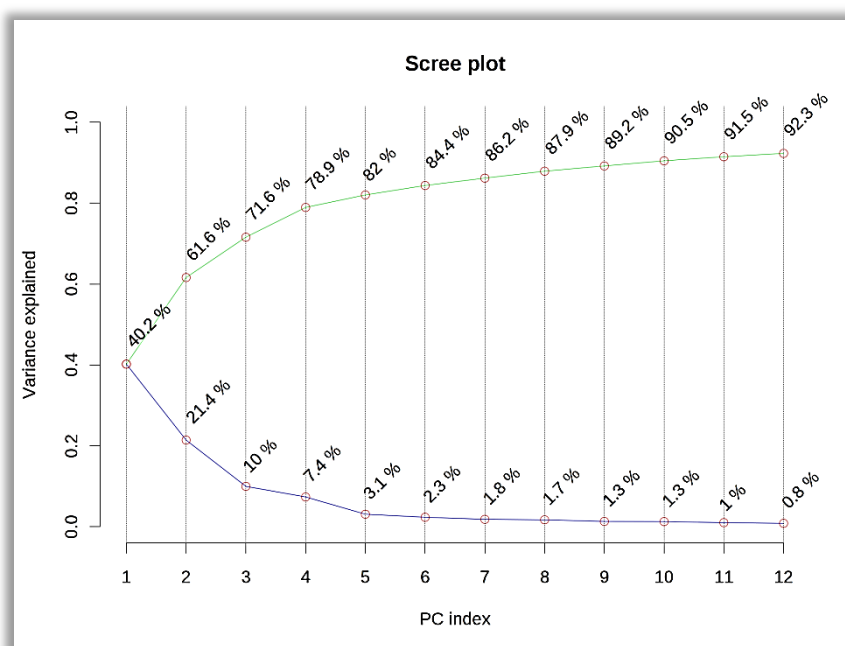


Figure A 76: PCA scree plot obtain from the PCA analysis of the pJRES spectra

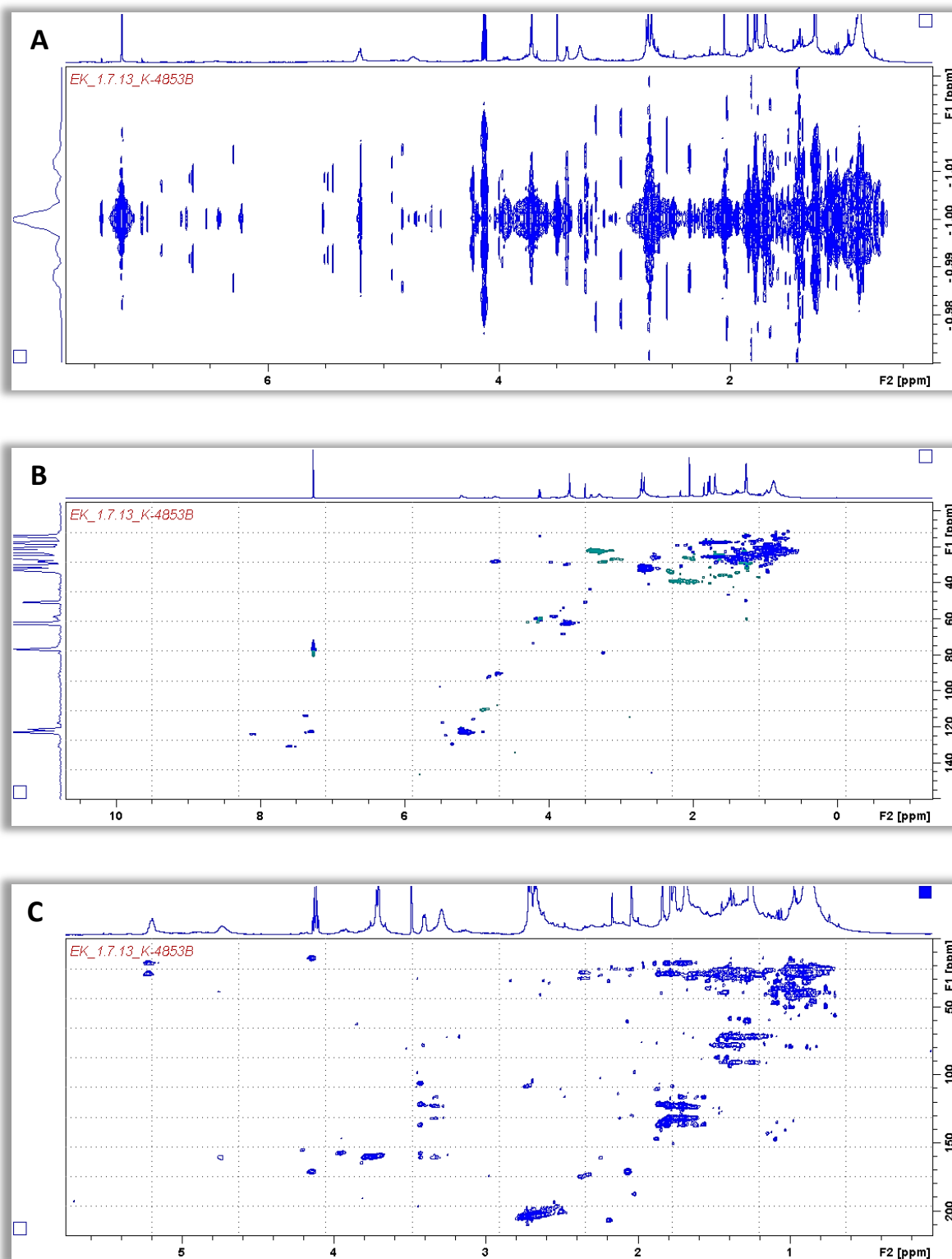


Figure A 77: Representative 2D spectra A. 2DJRES B. HSQC C. HMBC of *A. laurifolia* bark sample (K4853B) used for the structure determination of biomarkers

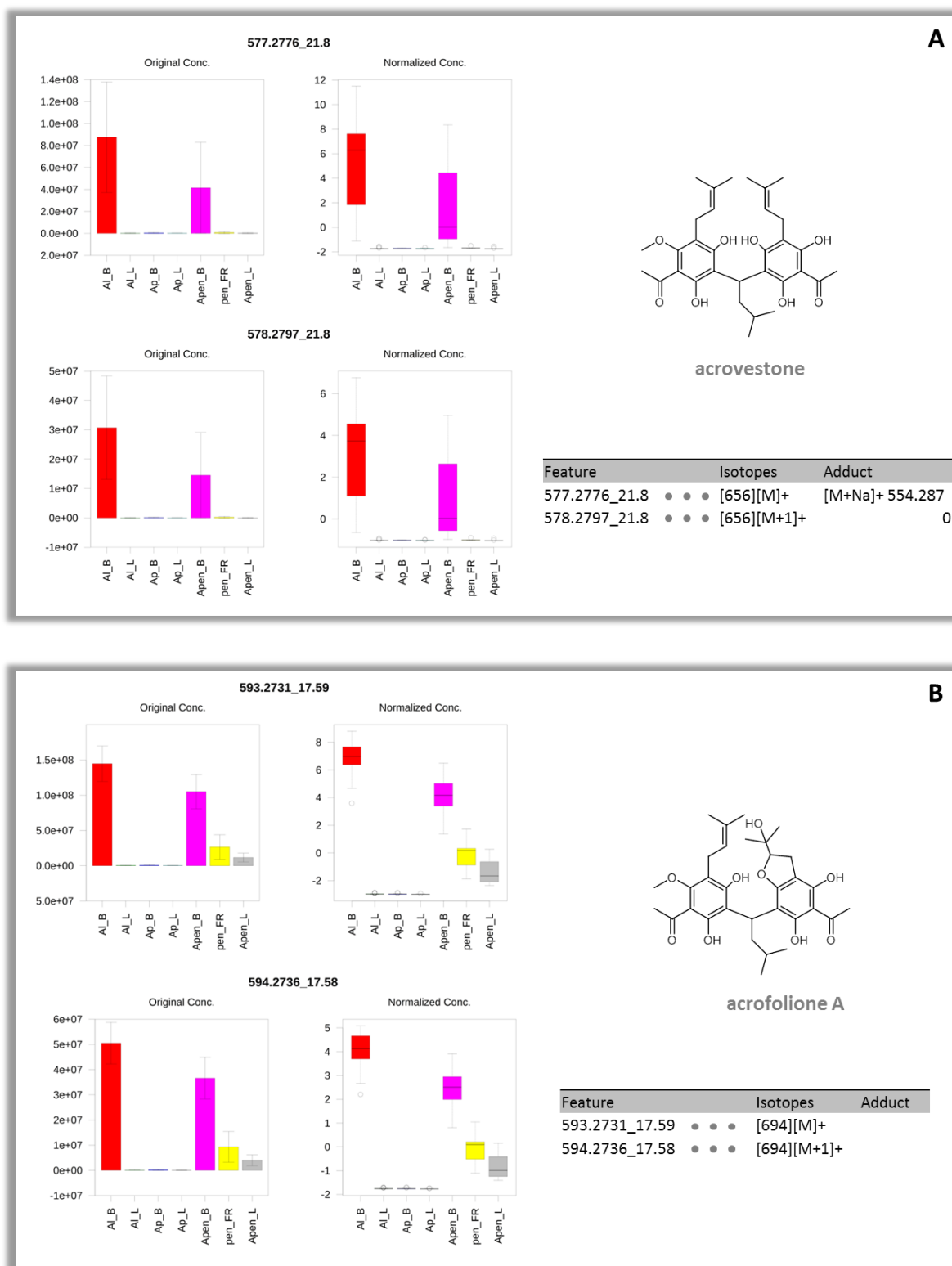


Figure A 78: Bar plots of the original concentrations (mean $\pm$ SD) on the left and box-whisker plots of the normalized concentrations on the right of two examples of biomarkers of *A. laurifolia*/ *pedunculata* bark samples namely A. acrovestone and B. acrofolione A; the identification was based on two discriminant features corresponding to the same metabolite as deduced from CAMERA output thus a segment of CAMERA table is provided

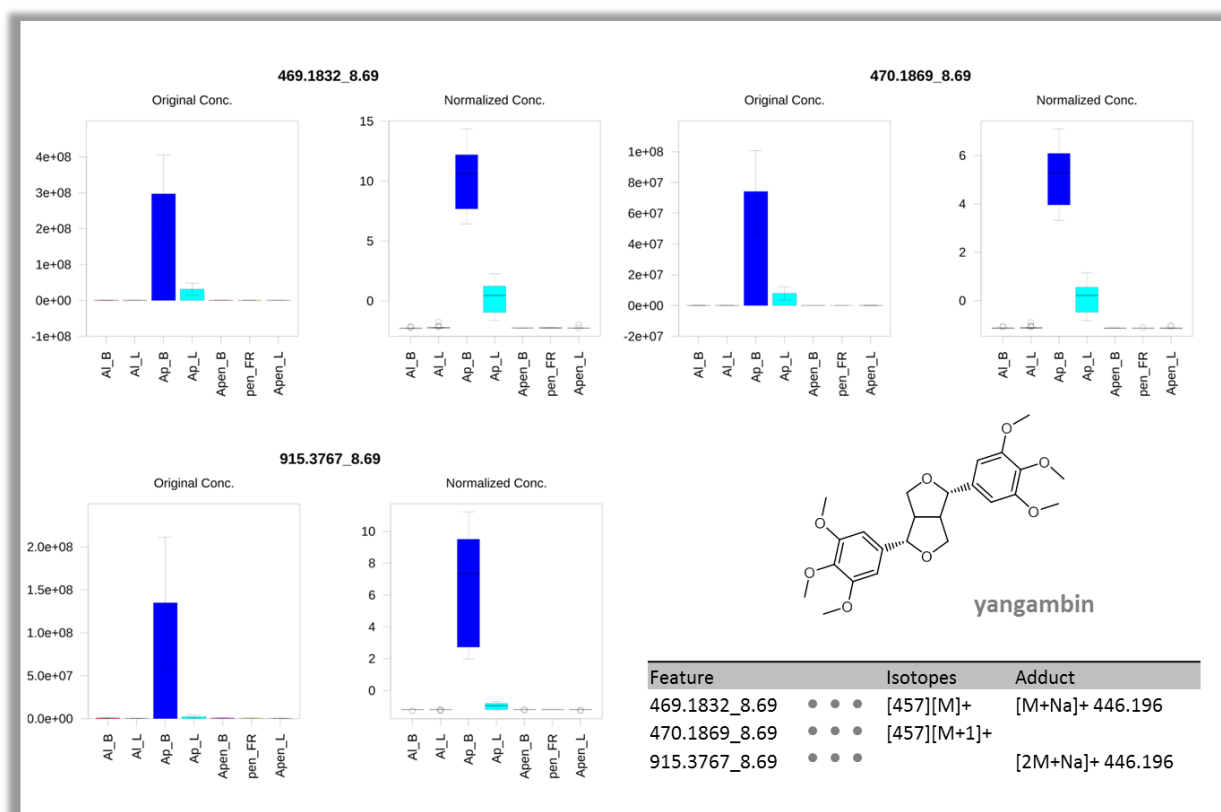


Figure A 79: Bar plots of the original concentrations (mean +/- SD) on the left and box-whisker plots of the normalized concentrations on the right of the biomarker of *A. porteri* bark samples yangambin; the identification was based on six (three of them are presented) discriminant features corresponding to the same metabolite as deduced from CAMERA output thus a segment of CAMERA table is provided

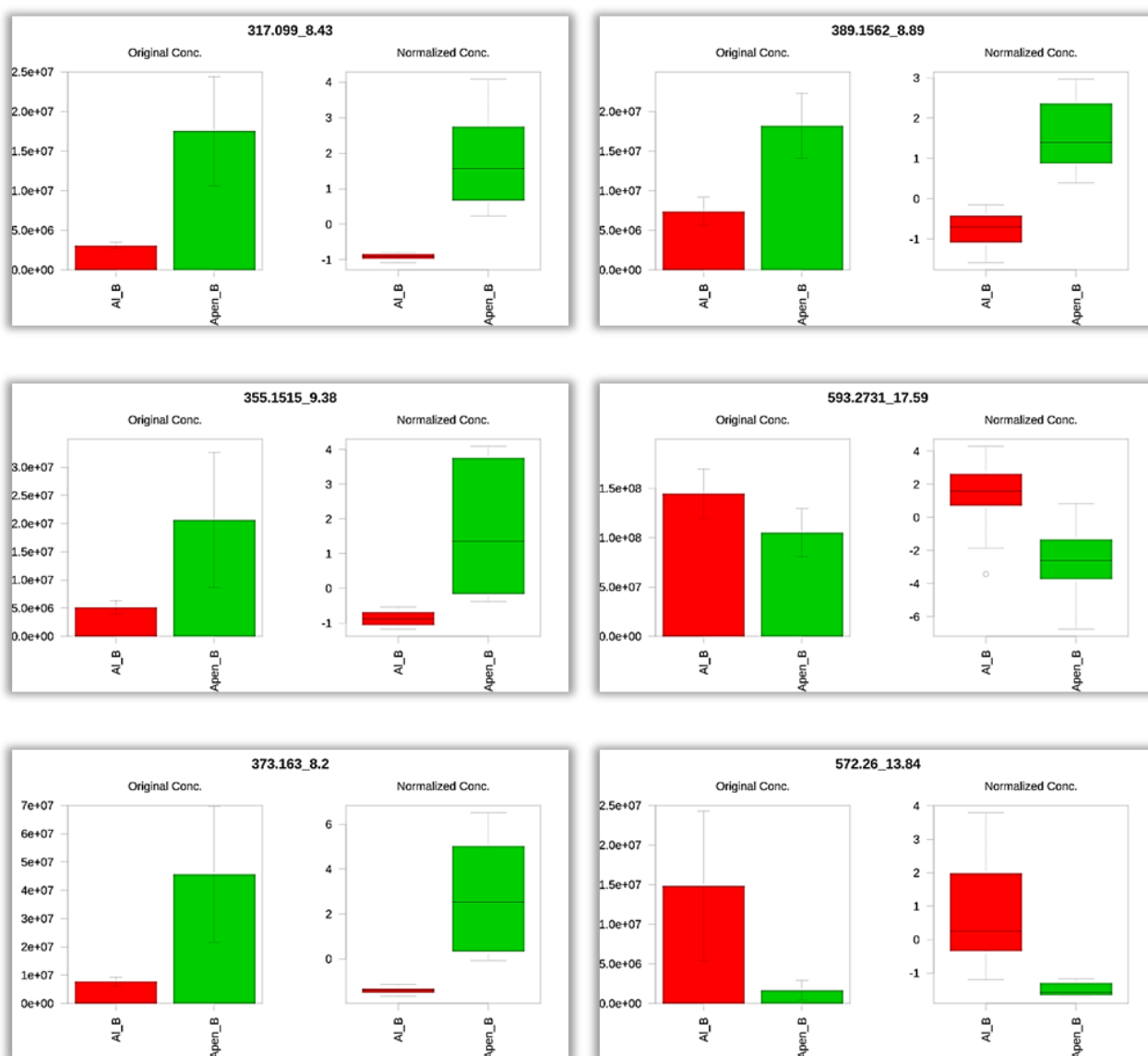


Figure A 80: Bar plots of the original concentrations (mean $\pm$ SD) on the left and box-whisker plots of the normalized concentrations on the right of the discriminant metabolites between *A. laurifolia* and *A. pedunculata* bark samples

Chapter 3: Metabolomics approaches in *Acronychia* genus

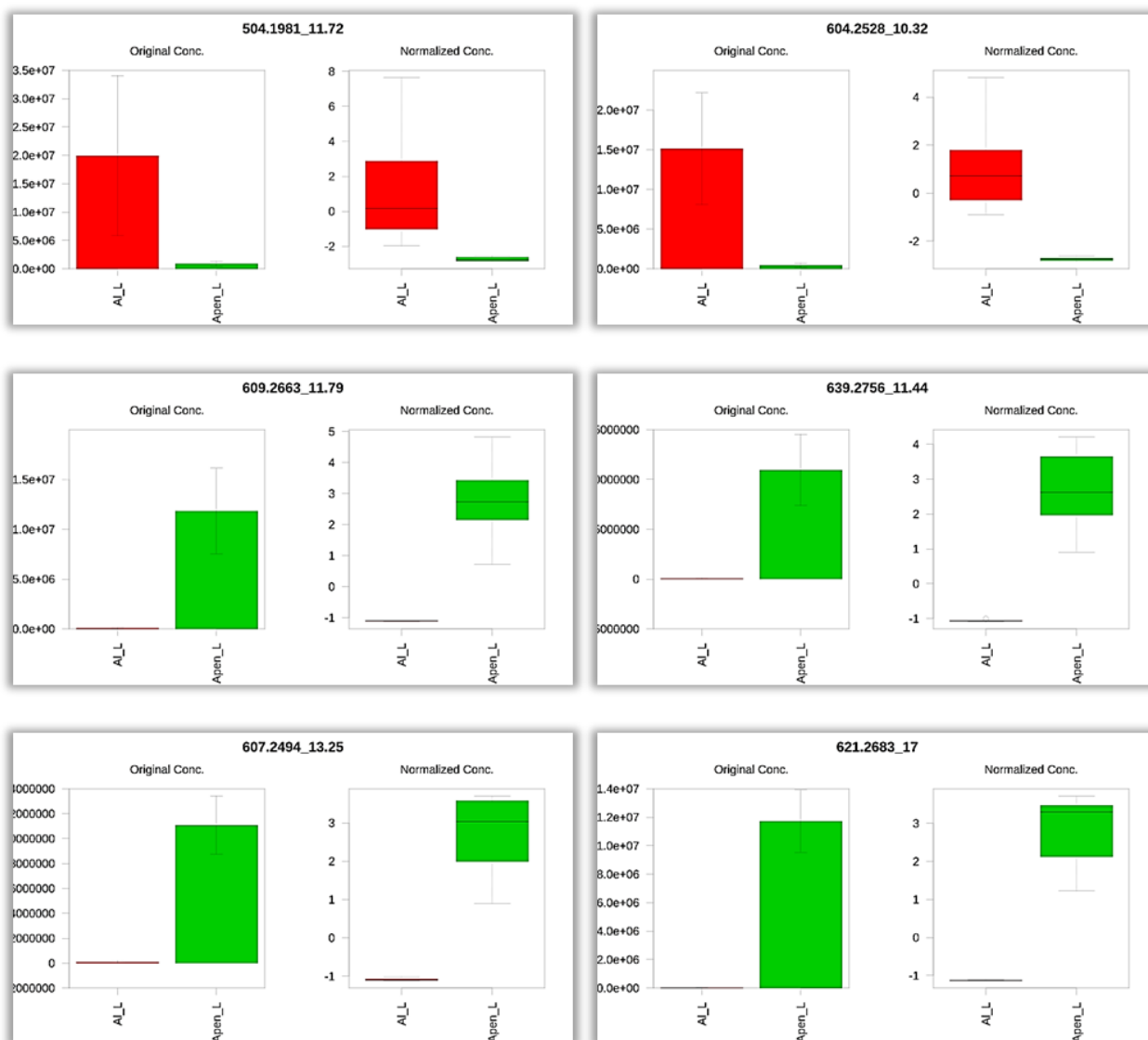


Figure A 81: Bar plots of the original concentrations (mean $\pm$ SD) on the left and box-whisker plots of the normalized concentrations on the right of the discriminant metabolites between *A. laurifolia* and *A. pedunculata* leave samples

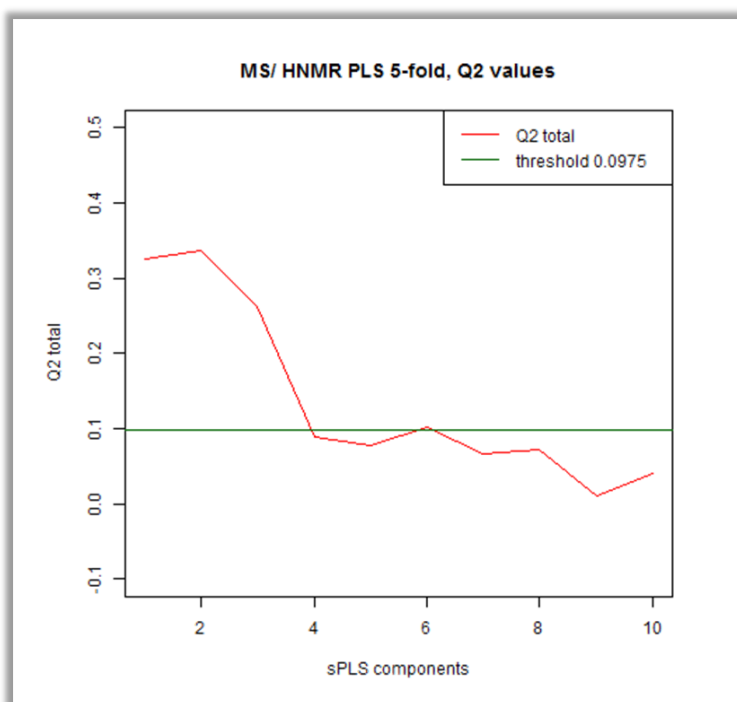


Figure A 82: Marginal contribution of the respective latent variable for different sparsity degrees in sPLS of *Acronychia* extract samples; the horizontal green line indicates the threshold value in Q^2

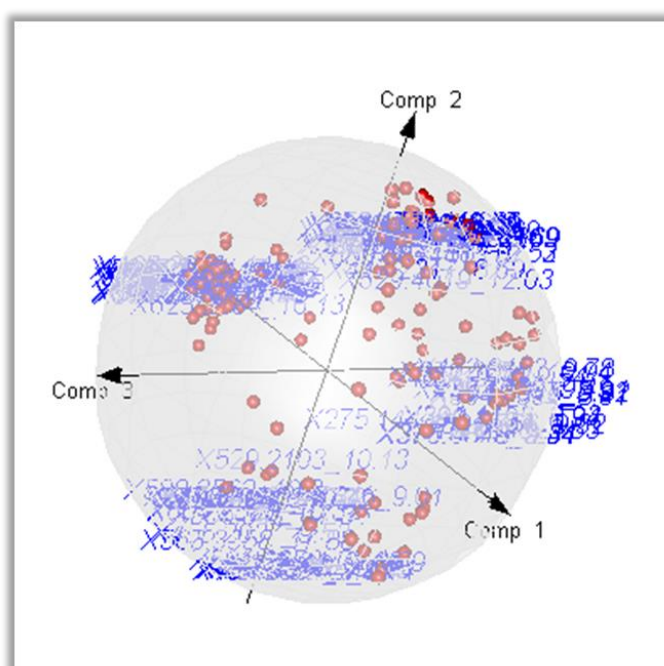


Figure A 83: 3D Correlation Circle plot for dimensions 1, 2 and 3 obtained from the sPLS analysis of pJRES and UPLC-ESI(+)-HRMS of *Acronychia* samples

Chapter 3: Metabolomics approaches in *Acronychia* genus

Table A 17: 5-LO product formation (mean % of control) calculated for all extracts assayed in five different concentrations

concentration	EK-4652 (B)		EK-4652 (L)		EK-4853 (B)		EK-4853 (L)		EK-5445 (B)	
	mean		mean		mean		mean		mean	
	% of control	SEM	% of control	SEM	% of control	SEM	% of control	SEM	% of control	SEM
0.1 µg/ml	103,09	5,92	94,14	12,93	96,54	4,84	80,22	6,52	85,39	5,65
0.3 µg/ml	83,68	4,84	66,31	4,17	71,46	3,48	58,95	2,08	98,37	3,36
1 µg/ml	35,54	4,38	42,09	6,49	31,33	8,87	32,62	3,48	59,82	17,53
3 µg/ml	13,83	1,74	15,43	1,22	16,32	2,14	10,29	1,15	41,52	3,27
10 µg/ml	5,33	1,43	7,29	0,94	4,31	0,98	4,85	0,68	17,02	4,48

concentration	EK-5445 (L)		EKL-4727 (B)		EKL-4727 (L)		EKL-4878 (B)		EKL-4878 (L)	
	mean		mean		mean		mean		mean	
	% of control	SEM	% of control	SEM	% of control	SEM	% of control	SEM	% of control	SEM
0.1 µg/ml	44,79	5,34	70,61	10,24	66,74	5,05	105,95	12,63	93,30	7,31
0.3 µg/ml	34,23	3,47	93,58	2,15	47,08	5,71	83,32	9,20	22,62	18,41
1 µg/ml	12,96	4,80	14,94	10,26	22,25	8,91	39,51	13,82	40,86	3,91
3 µg/ml	7,88	1,36	19,11	1,48	9,95	1,72	28,02	5,37	9,77	1,55
10 µg/ml	5,09	2,77	3,94	0,75	7,53	1,91	8,22	0,21	4,56	1,34

concentration	EKL-4882 (L)		EKL-5197 (B)		EKL-5197 (L)		EKL-5465 (B)		EKL-5465 (L)	
	mean		mean		mean		mean		mean	
	% of control	SEM	% of control	SEM	% of control	SEM	% of control	SEM	% of control	SEM
0.1 µg/ml	78,75	7,91	116,39	6,94	87,23	10,73	95,23	8,68	81,24	6,93
0.3 µg/ml	57,01	7,67	60,28	14,12	40,13	12,21	48,48	11,94	55,71	8,80
1 µg/ml	27,06	11,60	32,76	9,71	31,00	10,54	53,00	15,72	41,13	2,69
3 µg/ml	7,52	1,07	8,57	2,27	11,70	1,74	30,04	2,97	9,53	2,26
10 µg/ml	5,14	1,55	2,54	0,55	6,81	2,61	16,44	1,12	8,61	3,46

concentration	EVN-0179 (B)		EVN-0179 (L)		EVN-0874 (B)		EVN-0874 (FR)		EVN-0874 (L)	
	mean		mean		mean		mean		mean	
	% of control	SEM	% of control	SEM	% of control	SEM	% of control	SEM	% of control	SEM
0.1 µg/ml	111,21	7,04	96,75	5,80	96,27	18,45	93,16	9,67	67,91	8,94
0.3 µg/ml	87,09	4,40	60,56	3,49	56,74	6,65	36,26	16,49	47,76	4,18
1 µg/ml	42,97	6,07	42,94	5,78	33,94	5,20	30,37	6,86	36,77	4,21
3 µg/ml	16,07	3,20	10,66	2,48	26,49	1,57	22,97	3,21	11,50	1,85
10 µg/ml	5,18	1,45	4,41	0,08	5,00	1,97	6,06	1,03	7,24	0,47

Component	R2X	R2X(cum)	Eigenvalue	R2Y	R2Y(cum)	Q2	Limit	Q2(cum)	Significance	Iterations
0	Cent.									
1	0.227	0.227	27.2	0.443	0.443	0.426	0	0.426	R1	1
2	0.112	0.339	13.5	0.226	0.669	0.355	0	0.63	R1	1
3	0.186	0.524	22.3	0.102	0.771	0.27	0	0.73	R1	1
4	0.094	0.618	11.3	0.121	0.892	0.5	0	0.865	R1	1
5	0.0427	0.661	5.12	0.063	0.955	0.536	0	0.937	R1	1
6	0.0288	0.69	3.45	0.0203	0.975	0.347	0	0.959	R1	1

Figure A 84: Overview analysis of the PLS-r model; six component were considered explaining the 69% of the total variance of X dataset and the 97.5% of Y dataset; the predictive ability is represented by Q^2 accounting for 95.9%

Component	R2X	R2X(cum)	Eigenvalue	R2	R2(cum)	Q2	Limit	Q2(cum)	R2Y	R2Y(cum)	EigenvalueY	Significance
Model		0.714			0.983			0.963		1		
Predictive		0.0863			0.983			0.963		1		
P1	0.0863	0.0863	10.4	0.983	0.983	0.963	0.01	0.963	1	1	1	R1
Orthogonal in X(OPLS)		0.628			0							
O1	0.198	0.198	23.7	0	0							R1
O2	0.205	0.402	24.6	0	0							R1
O3	0.118	0.521	14.2	0	0							R1
O4	0.0509	0.572	6.11	0	0							R1
O5	0.031	0.603	3.72	0	0							R1
O6	0.0249	0.628	2.99	0	0							R1

Figure A 85: Overview analysis of the OPLS model; one predictive component was calculated accounting for the 98.3% of the Y variance and six orthogonal components were considered; the predictive ability is represented by Q^2 accounting for 96.3%

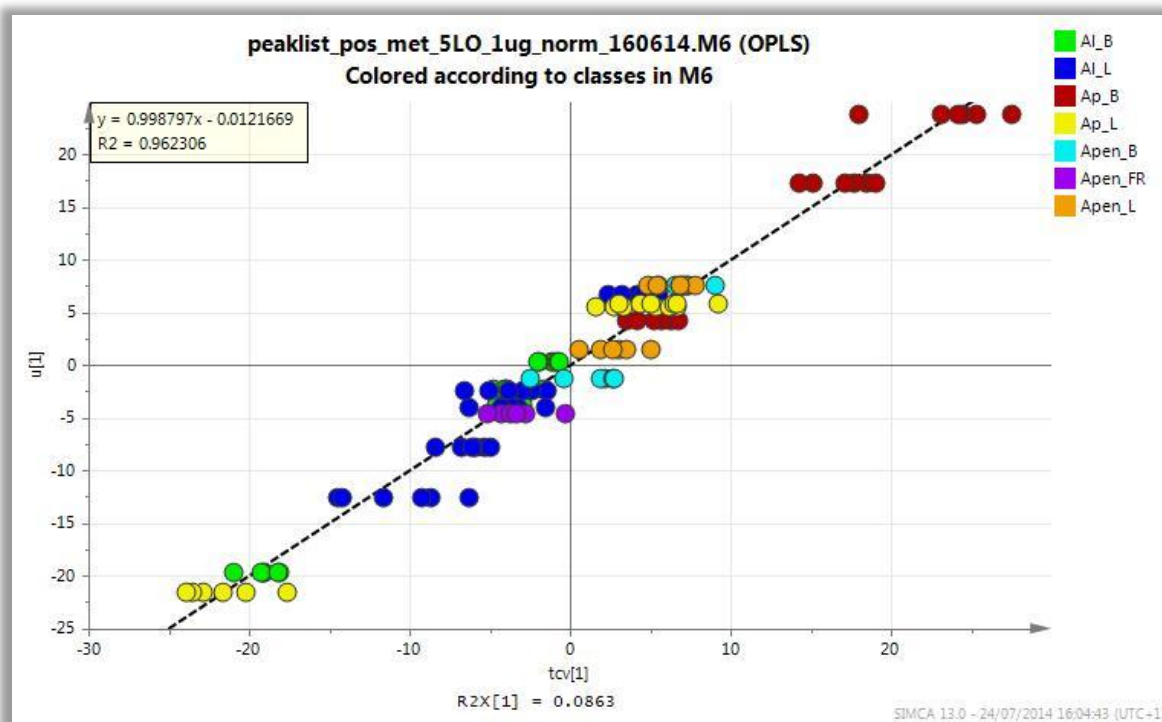


Figure A 86: OPLS scores plot (TCV/U) on the first component for the assessment of the predictability of the model; TCV represents the cross validated scores calculated for the X-predictive variable and U the scores of the Y variable; *A. laurifolia* barks are coloured in green, *A. laurifolia* leaves in blue, *A. porteri* barks in red, *A. porteri* leaves in yellow, *A. pedunculata* barks in turquoise, *A. pedunculata* fruits in purple and *A. pedunculata* leaves in orange

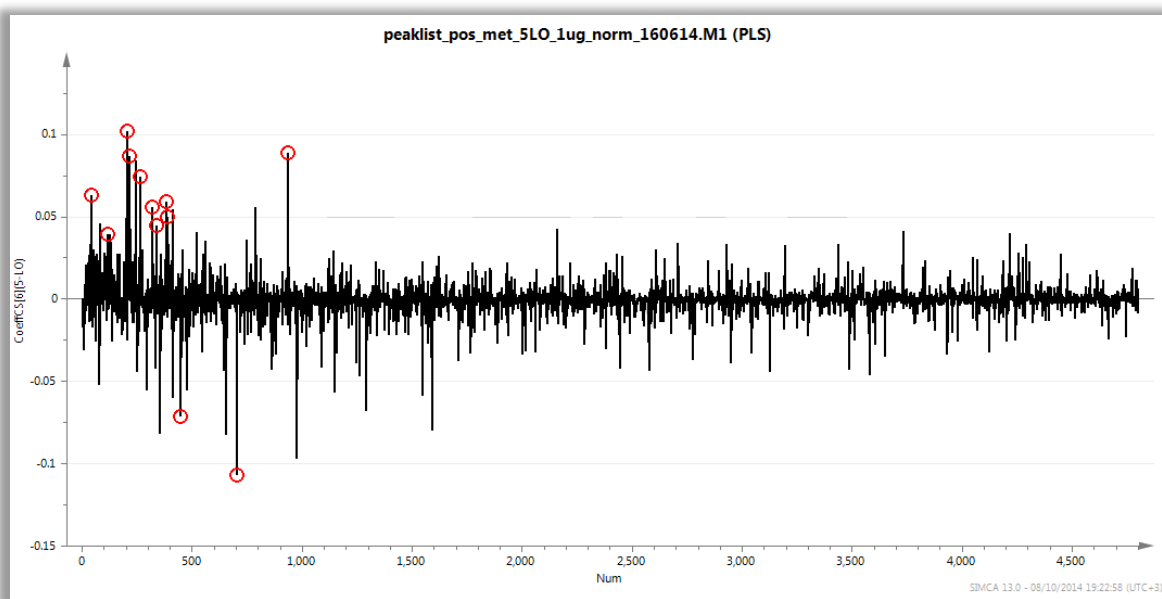


Figure A 87: Correlation coefficients line plot of all variables obtained from the PLS model; major contributing features are assigned with red circles

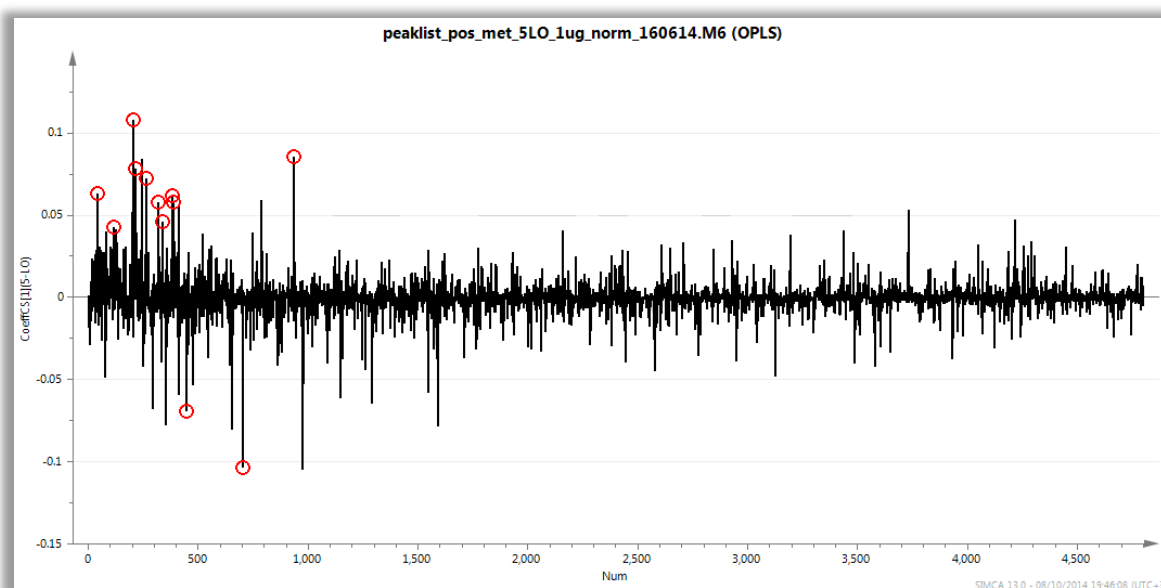


Figure A 88: Correlation coefficients line plot of all variables obtained from the OPLS model; major contributing features are assigned with red circles

References

- Abdi H. 2010. Partial least squares regression and projection on latent structure regression (PLS Regression). *Comput. Stat.* **2**: 97-106.
- Adersen A, Smitt UW, Simonsen HT, Christensen SB, Jaroszewski JW. 2007. Prenylated acetophenones from *Melicope obscura* and *Melicope obtusifolia* ssp. *obtusifolia* var. *arborea* and their distribution in Rutaceae. *Biochem. Syst. Ecol.* **35**: 447-453.
- Adersen A, Smitt UW, Simonsen HT, Christensen SB, Jaroszewski JW. 2007. Prenylated acetophenones from *Melicope obscura* and *Melicope obtusifolia* ssp. *obtusifolia* var. *arborea* and their distribution in Rutaceae. *Biochem. Syst. Ecol.* **35**: 447-453.
- Akhmedov NG, Myshakin EM, Dennis Hall C. 2004. Dynamic NMR and ab initio studies of exchange between rotamers of derivatives of octahydrofuro[3,4-f]isoquinoline-7(1H)-carboxylate and tetrahydro-2,5,6(1H)-isoquinolinetricarboxylate. *Magn. Reson. Chem.* **42**: 39-48.
- Allwood JW, Ellis DI, Goodacre R. 2008. Metabolomic technologies and their application to the study of plants and plant–host interactions. *Physiol. Plantarum* **132**: 117-135.
- Allwood JW, Goodacre R. 2010. An introduction to liquid chromatography–mass spectrometry instrumentation applied in plant metabolomic analyses. *Phytochem. Anal.* **21**: 33-47.
- Andreopoulos B, An A, Wang X, Schroeder M. 2009. A roadmap of clustering algorithms: finding a match for a biomedical application. *Brief. Bioinform.* **10**: 297-314.
- Andrews G, Dean R, Hawkrigde A, Muddiman D. 2011. Improving Proteome Coverage on a LTQ-Orbitrap Using Design of Experiments. *J. Am. Soc. Mass Spectrom.* **22**: 773-783.
- Appelhans MS, Wen J, Wagner WL. 2014. A molecular phylogeny of *Acrorychia*, *Euodia*, *Melicope* and relatives (Rutaceae) reveals polyphyletic genera and key innovations for species richness. *Mol. Phylogenet. Evol.* **79**: 54-68.
- Bagozzi D. 2003. WHO Fact Sheet No.134. *WHO (World Health Organization)*.
- Balunas MJ, Kinghorn AD. 2005. Drug discovery from medicinal plants. *Life Sci.* **78**: 431-441.
- Bankefors J, Broberg S, Nord LI, Kenne L. 2011. Electrospray ionization ion-trap multiple-stage mass spectrometry of *Quillaja* saponins. *J. Mass Spectrom.* **46**: 658-665.
- Bartel J, Krumsiek J, Theis FJ. 2013. Statistical methods for the analysis of high-throughput metabolomics data. *Comput. Struct. Biotechnol. J.* **4**.
- Bayly MJ, Holmes GD, Forster PI, Cantrill DJ, Ladiges PY. 2013. Major Clades of Australasian Rutoideae (Rutaceae) Based on *rbcl* and *atpB* Sequences. *PLoS ONE* **8**: e72493.
- Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH, Trethewey RN, Lange BM, Wurtele ES, Sumner LW. 2004. Potential of metabolomics as a functional genomics tool. *Trends Plant Sci.* **9**: 418-425.
- Blaas N, Humpf H-U. 2013. Structural profiling and quantitation of glycosyl inositol phosphoceramides in plants with fourier transform mass spectrometry. *J. Agric. Food Chem.* **61**: 4257-4269.
- Briskin DP. 2000. Medicinal Plants and Phytomedicines. Linking Plant Biochemistry and Physiology to Human Health. *Plant Physiol.* **124**: 507-514.
- Bro R, Smilde AK. 2003. Centering and scaling in component analysis. *J. Chemometr.* **17**: 16-33.
- Broadhurst D, Kell D. 2006. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* **2**: 171-196.

References

- Burkill IH. (1966) A dictionary of the economic products of the Malay Peninsula. Kuala Lumpur, Malaysia, Ministry of Agriculture & Co-operatives.
- Butler MS. 2004. The Role of Natural Product Chemistry in Drug Discovery. *J. Nat. Prod.* **67**: 2141-2153.
- Bylesjö M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J. 2006. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemometr.* **20**: 341-351.
- Castillo S, Gopalacharyulu P, Yetukuri L, Orešič M. 2011. Algorithms and tools for the preprocessing of LC–MS metabolomics data. *Chemom. Intell. Lab. Syst.* **108**: 23-32.
- Cho SK, Yang S-O, Kim S-H, Kim H, Ko JS, Riu KZ, Lee H-Y, Choi H-K. 2009. Classification and prediction of free-radical scavenging activities of dangyuja (*Citrus grandis* Osbeck) fruit extracts using ¹H NMR spectroscopy and multivariate statistical analysis. *J. Pharm. Biomed. Anal.* **49**: 567-571.
- Choi YH, Kim HK, Linthorst HJ, Hollander JG, Lefeber AW, Erkelens C, Nuzillard JM, Verpoorte R. 2006. NMR metabolomics to revisit the tobacco mosaic virus infection in *Nicotiana tabacum* leaves. *J. Nat. Prod.* **69**: 742-748.
- Chun WY, Chang CC. (1965) Flora Hainanica. Bei Jing, China, Science Press.
- Craig A, Cloarec O, Holmes E, Nicholson JK, Lindon JC. 2006. Scaling and Normalization Effects in NMR Spectroscopic Metabonomic Data Sets. *Anal. Chem.* **78**: 2262-2267.
- Cui B, Chai H, Dong Y, Horgen FD, Hansen B, Madulid DA, Soejarto DD, Farnsworth NR, Cordell GA, Pezzuto JM, Kinghorn AD. 1999. Quinoline alkaloids from *Acronychia laurifolia*. *Phytochemistry* **52**: 95-98.
- Cuyckens F, Claeys M. 2004. Mass spectrometry in the structural analysis of flavonoids. *J. Mass Spectrom.* **39**: 1-15.
- Danielsson R, Bylund D, Markides KE. 2002. Matched filtering with background suppression for improved quality of base peak chromatograms and mass spectra in liquid chromatography–mass spectrometry. *Anal. Chim. Acta* **454**: 167-184.
- Davenport JB, Sutherland MD. 1954. Asarinin in *Acronychia muelleri* W.D. Francis. *Aust. J. Chem.* **7**: 384-386.
- de Silva LB, de Silva ULL, Mahendran M, Jennings R. 1979. Kokusaginine and evolitrine from *Acronychia pedunculata*. *Phytochemistry* **18**: 1255-1256.
- De Silva LB, Herath WM, Liyanage C, Kumar V, Uddin Ahmad V, Sultana A. 1991. Demethylacrovostone from *Achronychia pedunculata* fruits. *Phytochemistry* **30**: 1709-1710.
- De Vos RC, Moco S, Lommen A, Keurentjes JJ, Bino RJ, Hall RD. 2007. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2**: 778-791.
- Dear GJ, James AD, Sarda S. 2006. Ultra-performance liquid chromatography coupled to linear ion trap mass spectrometry for the identification of drug metabolites in biological samples. *Rapid Commun. Mass Spectrom.* **20**: 1351-1360.
- Dejean S, Gonzalez I, Le Cao K-A (2014). mixOmics: Omics Data Integration Project.
- Dettmer K, Aronov PA, Hammock BD. 2007. Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **26**: 51-78.
- Dias DA, Urban S, Roessner U. 2012. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2**: 303-336.
- Dobson CM. 2004. Chemical space and biology. *Nature* **432**: 824-828.
- Domon B, Costello CE. 1988. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **5**: 397-409.
- Dunn WB. 2008. Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Phys. Biol.* **5**: 011001.

- Dunn WB, Broadhurst D, Brown M, Baker PN, Redman CWG, Kenny LC, Kell DB. 2008. Metabolic profiling of serum using ultra performance liquid chromatography and the LTO-Orbitrap mass spectrometry system. *J. Chromatogr. B* **871**: 288-298.
- Dunn WB, Ellis DI. 2005. Metabolomics: Current analytical platforms and methodologies. *TrAC Trend. Anal. Chem.* **24**: 285-294.
- Ebinger K, Weller HN. 2013. Comparison of chromatographic techniques for diastereomer separation of a diverse set of drug-like compounds. *J. Chromatogr. A* **1272**: 150-154.
- Epifano F, Fiorito S, Genovese S. 2013. Phytochemistry and pharmacognosy of the genus *Acronychia*. *Phytochemistry* **95**: 12-18.
- Eriksson L, Johansson E, Kettaneh-Wold N, Trygg J, Wold S. (2006) Multi- and Megavariate Data Analysis. Basic Principles and Applications. Umeå, Umetrics Academy.
- Eriksson L, Johansson E, Lindgren F, Sjöström M, Wold S. 2002. Megavariate analysis of hierarchical QSAR data. *J. Comput. Aid. Mol. Des.* **16**: 711-726.
- Eriksson L, Trygg J, Wold S. 2008. CV-ANOVA for significance testing of PLS and OPLS® models. *J. Chemometr.* **22**: 594-600.
- Fabricant DS, Farnsworth NR. 2001. The value of plants used in traditional medicine for drug discovery. *Environ. Health Perspect.* **109**: 69-75.
- FDA (2001). Guidance for Industry, Bioanalytical method Validation, Food and Drug Administration, Centre for Drug Evaluation and Research (CDER). FDA.
- Fiehn O. (2002) Metabolomics — the link between genotypes and phenotypes. *Functional Genomics*. Town C., Springer Netherlands, 155-171.
- Fiehn O, Sumner L, Rhee S, Ward J, Dickerson J, Lange B, Lane G, Roessner U, Last R, Nikolau B. 2007. Minimum reporting standards for plant biology context information in metabolomic studies. *Metabolomics* **3**: 195-201.
- Fischer L, Szellas D, Rådmark O, Steinhilber D, Werz O. 2003. Phosphorylation- and stimulus-dependent inhibition of cellular 5-lipoxygenase activity by nonredox-type inhibitors. *FASEB J.* **17**: 949-951.
- Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. 2004. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47**: 1739-1749.
- Fukusaki E, Kobayashi A. 2005. Plant metabolomics: potential for practical operation. *J. Biosci. Bioeng.* **100**: 347-354.
- Furtado NAJC, Vessicchi R, Tomaz JC, Galembeck SE, Bastos JK, Lopes NP, Crotti AEM. 2007. Fragmentation of diketopiperazines from *Aspergillus fumigatus* by electrospray ionization tandem mass spectrometry (ESI-MS/MS). *J. Mass Spectrom.* **42**: 1279-1286.
- Gabelica V, Pauw ED. 2005. Internal energy and fragmentation of ions produced in electrospray sources. *Mass Spectrom. Rev.* **24**: 566-587.
- Gad HA, El-Ahmady SH, Abou-Shoer MI, Al-Azizi MM. 2013. Application of chemometrics in authentication of herbal medicines: A review. *Phytochem. Anal.* **24**: 1-24.
- Gibbons S, Udo EE. 2000. The effect of reserpine, a modulator of multidrug efflux pumps, on the in vitro activity of tetracycline against clinical isolates of methicillin resistant *Staphylococcus aureus* (MRSA) possessing the tet(K) determinant. *Phytother. Res.* **14**: 139-140.
- Gika HG, Theodoridis GA, Wingate JE, Wilson ID. 2007. Within-Day Reproducibility of an HPLC-MS-Based Method for Metabonomic Analysis: Application to Human Urine. *J. Proteome Res.* **6**: 3291-3303.
- Gilbert B, Alves LF. 2003. Synergy in plant medicines. *Curr. Med. Chem.* **10**: 13-20.
- Gonzalez I, Cao K-AL, Davis M, Dejean S. 2012. Visualising associations between paired 'omics' data sets. *BioData Min.* **5**: 19.

References

- Goodacre R, Broadhurst D, Smilde A, Kristal B, Baker JD, Beger R, Bessant C, Connor S, Capuani G, Craig A, Ebbels T, Kell D, Manetti C, Newton J, Paternostro G, Somorjai R, Sjöström M, Trygg J, Wulfert F. 2007. Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics* **3**: 231-241.
- Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. 2004. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol.* **22**: 245-252.
- Govindachari TR, Jadhav SJ, Joshi BS, Kamat VN, Mohammed A, Parthasarathy PC, Patankar SJ, Prakash D, Ran DF, Viswanathan NV. 1969. Chemical Investigation of Some Indian Plants: Part IV. *Indian J. Chem.* **7**: 308-310.
- Gross J. (2011) Practical Aspects of Electron Ionization. *Mass Spectrometry*. Gross J., Springer Berlin Heidelberg, 223-248.
- Gustavsson SÅ, Samskog J, Markides KE, Långström B. 2001. Studies of signal suppression in liquid chromatography–electrospray ionization mass spectrometry using volatile ion-pairing reagents. *J. Chromatogr. A* **937**: 41-47.
- Gutowsky HS, Holm CH. 1956. Rate Processes and Nuclear Magnetic Resonance Spectra. II. Hindered Internal Rotation of Amides. *J. Chem. Phys.* **25**: 1228-1234.
- Hackstadt A, Hess A. 2009. Filtering for increased power for microarray data analysis. *BMC Bioinformatics* **10**: 11.
- Hagel J, Facchini P. 2008. Plant metabolomics: analytical platforms and integration with functional genomics. *Phytochem. Rev.* **7**: 479-497.
- Halabalaki M, Bertrand S, Stefanou A, Gindro K, Kostidis S, Mikros E, Skaltsounis LA, Wolfender JL. 2014. Sample preparation issues in NMR-based plant metabolomics: optimisation for *Vitis* wood samples. *Phytochem. Anal.* **25**: 350-356.
- Halabalaki M, Vougiotiannopoulou K, Mikros E, Skaltsounis AL. 2014. Recent advances and new strategies in the NMR-based identification of natural products. *Curr. Opin. Chem. Biol.* **25**: 1-7.
- Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL. 2004. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **47**: 1750-1759.
- Hall RD. 2006. Plant metabolomics: from holistic hope, to hype, to hot topic. *New Phytol.* **169**: 453-468.
- Hartley TG. 1974. A revision of the genus *Acronychia*. (Rutaceae). *J. Arnold Arbor.* **55**: 469-567.
- Holčapek M, Jirásko R, Líska M. 2010. Basic rules for the interpretation of atmospheric pressure ionization mass spectra of small molecules. *J. Chromatogr. A* **1217**: 3908-3921.
- Hooft JJ, Vervoort J, Bino R, Vos RH. 2012. Spectral trees as a robust annotation tool in LC–MS based metabolomics. *Metabolomics* **8**: 691-703.
- Horgen FD, Edrada RA, de los Reyes G, Agcaoili F, Madulid DA, Wongpanich V, Angerhofer CK, Pezzuto JM, Soejarto DD, Farnsworth NR. 2001. Biological screening of rain forest plot trees from Palawan Island (Philippines). *Phytomedicine* **8**: 71-81.
- Hrydziusko O, Viant M. 2012. Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline. *Metabolomics* **8**: 161-174.
- Hubert J, Nuzillard J-M, Purson S, Hamzaoui M, Borie N, Reynaud R, Renault J-H. 2014. Identification of Natural Metabolites in Mixture: A Pattern Recognition Strategy Based on ¹³C NMR. *Anal. Chem.* **86**: 2955-2962.
- Indahl UG. 2014. Towards a complete identification of orthogonal variation in multiple regression from a PLS1 modeling point of view: including OPLS by a change of orthogonal basis. *J. Chemometr.* **28**: 508-517.

- Inui T, Wang Y, Pro SM, Franzblau SG, Pauli GF. 2012. Unbiased evaluation of bioactive secondary metabolites in complex matrices. *Fitoterapia* **83**: 1218-1225.
- Issaq HJ, Van ON, Waybright TJ, Muschik GM, Veenstra TD. 2009. Analytical and statistical approaches to metabolomics research. *J. Sep. Sci.* **32**: 2183-2199.
- Ito Y, Bowman RL. 1970. Countercurrent Chromatography: Liquid-Liquid Partition Chromatography without Solid Support. *Science* **167**: 281-283.
- Jayasinghe P, Bandara B, Ekanayaka E, Thevanesam V. 2006. Screening for antimicrobial activity of *Acrorychia pedunculata* (Ankenda) and *Adenanthera Pavonina* (Madatiya) against bacteria causing skin and wound infections in humans. *Proceedings of the Peradeniya University Research Sessions* **11**: 105.
- Justesen U. 2001. Collision-induced fragmentation of deprotonated methoxylated flavonoids, obtained by electrospray ionization mass spectrometry. *J. Mass Spectrom.* **36**: 169-178.
- Kemsley EK. 1996. Discriminant analysis of high-dimensional data: a comparison of principal components analysis and partial least squares data reduction methods. *Chemom. Intell. Lab. Syst.* **33**: 47-61.
- Khairudin K, Sukiran N, Goh H-H, Baharum S, Noor N. 2014. Direct discrimination of different plant populations and study on temperature effects by Fourier transform infrared spectroscopy. *Metabolomics* **10**: 203-211.
- Kim HK, Choi YH, Erkelens C, Lefeber AWM, Verpoorte R. 2005. Metabolic Fingerprinting of *Ephedra* Species Using ¹H-NMR Spectroscopy and Principal Component Analysis. *Chem. Pharm. Bull.* **53**: 105-109.
- Kim HK, Choi YH, Verpoorte R. 2010. NMR-based metabolomic analysis of plants. *Nat. Protoc.* **5**: 536-549.
- Kim HK, Choi YH, Verpoorte R. 2011. NMR-based plant metabolomics: where do we stand, where do we go? *Trends Biotechnol.* **29**: 267-275.
- Kim HK, Saifullah, Khan S, Wilson EG, Kricun SDP, Meissner A, Goraler S, Deelder AM, Choi YH, Verpoorte R. 2010. Metabolic classification of South American *Ilex* species by NMR-based metabolomics. *Phytochemistry* **71**: 773-784.
- Kim HK, Verpoorte R. 2010. Sample preparation for plant metabolomics. *Phytochem. Anal.* **21**: 4-13.
- Kim Y, Lee J, Park S-H, Lee C, Lee J, Lee D, Kim N, Lee D, Kim H, Lee C. 2012. LC-MS-based chemotaxonomic classification of wild-type *Lespedeza* sp. and its correlation with genotype. *Plant Cell Rep.* **31**: 2085-2097.
- Kind T, Fiehn O. 2006. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics* **7**: 234.
- Kind T, Fiehn O. 2007. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* **8**: 105.
- Kind T, Fiehn O. 2010. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal. Rev.* **2**: 23-60.
- Klagkou K, Pullen F, Harrison M, Organ A, Firth A, Langley GJ. 2003. Approaches towards the automated interpretation and prediction of electrospray tandem mass spectra of non-peptidic combinatorial compounds. *Rapid Commun. Mass Spectrom.* **17**: 1163-1168.
- Koeberle A, Zettl H, Greiner C, Wurglics M, Schubert-Zsilavecz M, Werz O. 2008. Pirinixic Acid Derivatives as Novel Dual Inhibitors of Microsomal Prostaglandin E2 Synthase-1 and 5-Lipoxygenase. *J. Med. Chem.* **51**: 8068-8076.

References

- Kouloura E, Halabalaki M, Lallemand M-C, Nam S, Jove R, Litaudon M, Awang K, Hadi HA, Skaltsounis A-L. 2012. Cytotoxic prenylated acetophenone dimers from *Acronychia pedunculata*. *J. Nat. Prod.* **75**: 1270-1276.
- Kovacs H, Moskau D, Spraul M. 2005. Cryogenically cooled probes—a leap in NMR technology. *Prog. Nucl. Magn. Reson. Spectrosc.* **46**: 131-155.
- Kozaki S, Takenaka Y, Mizushina Y, Yamaura T, Tanahashi T. 2014. Three acetophenones from *Acronychia pedunculata*. *J. Nat. Med.* **68**: 421-426.
- Kruve A, Kaupmees K, Liigand J, Oss M, Leito I. 2013. Sodium adduct formation efficiency in ESI source. *J. Mass Spectrom.* **48**: 695-702.
- Kueger S, Steinhauser D, Willmitzer L, Giavalisco P. 2012. High-resolution plant metabolomics: from mass spectral features to metabolites and from whole-cell analysis to subcellular metabolite distributions. *Plant J.* **70**: 39-50.
- Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S. 2011. CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets. *Anal. Chem.* **84**: 283-289.
- Lamberton JA, Price JR. 1953. Alkaloids of the Australian Rutaceae: *Acronychia baueri* Schott. IV. Alkaloids present in the leaves. *Aust. J. Chem.* **6**: 66-77.
- Larsson J, Gottfries J, Muresan S, Backlund A. 2007. ChemGPS-NP: Tuned for Navigation in Biologically Relevant Chemical Space. *J. Nat. Prod.* **70**: 789-794.
- Laufer SA, Augustin J, Dannhardt G, Kiefer W. 1994. (6,7-Diaryldihydropyrrolizin-5-yl)acetic Acids, a Novel Class of Potent Dual Inhibitors of Both Cyclooxygenase and 5-Lipoxygenase. *J. Med. Chem.* **37**: 1894-1897.
- Le Cao K-A, Boitard S, Besse P. 2011. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* **12**: 253.
- Le Cao KA, Rossouw D, Robert-Granie C, Besse P. 2008. A sparse PLS for variable selection when integrating omics data. *Stat. Appl. Genet. Mol. Biol.* **7**: 35.
- Lecomte MH. (1907-1951) Flore générale de l'Indo-Chine. Paris, France, Masson.
- Lenz EM, Wilson ID. 2006. Analytical Strategies in Metabonomics. *J. Proteome Res.* **6**: 443-458.
- Levandi T, Leon C, Kaljurand M, Garcia-Canas V, Cifuentes A. 2008. Capillary electrophoresis time-of-flight mass spectrometry for comparative metabolomics of transgenic versus conventional maize. *Anal. Chem.* **80**: 6329-6335.
- Lichius JJ, Thoison O, Montagnac A, Pais M, Guéritte-Voegelein F, Sévenet T, Cosson J-P, Hadi AHA. 1994. Antimitotic and Cytotoxic Flavonols from *Zieridium pseudobtusifolium* and *Acronychia porteri*. *J. Nat. Prod.* **57**: 1012-1016.
- Liland KH. 2011. Multivariate methods in metabolomics – from pre-processing to dimension reduction and statistical analysis. *TrAC Trend. Anal. Chem.* **30**: 827-841.
- Lim H-K, Chen J, Sensenhauser C, Cook K, Subrahmanyam V. 2007. Metabolite identification by data-dependent accurate mass spectrometric analysis at resolving power of 60 000 in external calibration mode using an LTQ/Orbitrap. *Rapid Commun. Mass Spectrom.* **21**: 1821-1832.
- Lim H, Son KH, Bae KH, Hung TM, Kim YS, Kim HP. 2009. 5-Lipoxygenase-inhibitory constituents from *Schizandra fructus* and *Magnolia flos*. *Phytother. Res.* **23**: 1489-1492.
- Lindon J, Holmes E, Nicholson J. 2006. Metabonomics Techniques and Applications to Pharmaceutical Research & Development. *Pharmaceut. Res.* **23**: 1075-1088.
- Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR. 2006. Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.* **1**: 387-396.

- Liu J, Yang X, He J, Xia M, Xu L, Yang S. 2007. Structure analysis of triterpene saponins in *Polygala tenuifolia* by electrospray ionization ion trap multiple-stage mass spectrometry. *J. Mass Spectrom.* **42**: 861-873.
- Liu L, Nam S, Tian Y, Yang F, Wu J, Wang Y, Scuto A, Polychronopoulos P, Magiatis P, Skaltsounis L, Jove R. 2011. 6-Bromoindirubin-3'-oxime inhibits JAK/STAT3 signaling and induces apoptosis of human melanoma cells. *Cancer Res.* **71**: 3972-3979.
- Lommen A. 2009. MetAlign: Interface-Driven, Versatile Metabolomics Tool for Hyphenated Full-Scan Mass Spectrometry Data Preprocessing. *Anal. Chem.* **81**: 3079-3086.
- Lubbe A, Ali K, Verpoorte R, Hae Choi Y. (2013) NMR-Based Metabolomics Analysis. *Metabolomics in Practice*. Lämmerhofer M., Weckwerth W. Weinheim, Germany, Wiley-VCH Verlag GmbH & Co. KGaA, 209-238.
- Ludwig C, Gunther U. 2011. MetaboLab - advanced NMR data processing and analysis for metabolomics. *BMC Bioinformatics* **12**: 366.
- Ma YL, Li QM, Van Den Heuvel H, Claeys M. 1997. Characterization of flavone and flavonol aglycones by collision-induced dissociation tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **11**: 1357-1364.
- Marston A, Hostettmann K. 2006. Developments in the application of counter-current chromatography to plant analysis. *J Chromatogr A* **1112**: 181-194.
- Mehmood T, Liland KH, Snipen L, Sæbø S. 2012. A review of variable selection methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst.* **118**: 62-69.
- Melamud E, Vastag L, Rabinowitz JD. 2010. Metabolomic Analysis and Visualization Engine for LC-MS Data. *Anal. Chem.* **82**: 9818-9826.
- Michel T, Halabalaki M, Skaltsounis A-L. 2013. New Concepts, Experimental Approaches, and Dereplication Strategies for the Discovery of Novel Phytoestrogens from Natural Sources. *Planta Med.* **230**: 514-532.
- Mishra BB, Tiwari VK. 2011. Natural products: An evolving role in future drug discovery. *Eur. J. Med. Chem.* **46**: 4769-4807.
- Moco S, Vervoort J, Moco S, Bino RJ, De Vos RCH, Bino R. 2007. Metabolomics technologies and metabolite identification. *TrAC Trend. Anal. Chem.* **26**: 855-866.
- Mohamadi F, Richards NGJ, Guida WC, Liskamp R, Lipton M, Caufield C, Chang G, Hendrickson T, Still WC. 1990. Macromodel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* **11**: 440-467.
- Morris KF, Erickson LE. 1996. NMR Determination of Internal Rotation Rates and Rotational Energy Barriers: A Physical Chemistry Lab Project. *J. Chem. Educ.* **73**: 471.
- Mushtaq MY, Choi YH, Verpoorte R, Wilson EG. 2014. Extraction for Metabolomics: Access to The Metabolome. *Phytochem. Anal.* **25**: 291-306.
- Næs T, Mevik B-H. 2001. Understanding the collinearity problem in regression and discriminant analysis. *J. Chemometr.* **15**: 413-426.
- Nakabayashi R, Saito K. 2013. Metabolomics for unknown plant metabolites. *Anal. Bioanal. Chem.* **405**: 5005-5011.
- Newman DJ, Cragg GM. 2007. Natural Products as Sources of New Drugs over the Last 25 Years. *J. Nat. Prod.* **70**: 461-477.
- Newman DJ, Cragg GM. 2009. Natural product scaffolds as leads to drugs. *Future Med. Chem.* **1**: 1415-1427.
- Newman DJ, Cragg GM. 2012. Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *J. Nat. Prod.* **75**: 311-335.
- Niessen WMA. 2000. Structure elucidation by LC-MS. Foreword. *Analusis* **28**: 885-887.
- Nisbet LJ, Moore M. 1997. Will natural products remain an important source of drug research for the future? *Curr. Opin. Chem. Biol.* **8**: 708-712.

References

- Oettl SK, Hubert J, Nuzillard J-M, Stuppner H, Renault J-H, Rollinger JM. 2014. Dereplication of depsides from the lichen *Pseudevernia furfuracea* by centrifugal partition chromatography combined to ^{13}C nuclear magnetic resonance pattern recognition. *Anal. Chim. Acta* **846**: 60-67.
- Oyama M, Bastow KF, Tachibana Y, Shirataki Y, Yamaguchi S, Cragg GM, Wu T-S, Lee K-H. 2003. Antitumor agents 225. Acrofoliones A and B, two novel cytotoxic acetophenone dimers from *Acronychia trifoliolata*. *Chin. Pharm. J.* **55**: 239-245.
- Oyama M, Bastow KF, Tachibana Y, Shirataki Y, Yamaguchi S, Cragg GM, Wu TS, Lee KH. 2003. Antitumor agents 225. Acrofoliones A and B, two novel cytotoxic acetophenone dimers, from *Acronychia trifoliolata*. *Chin. Pharm. J.* **55**: 239-245.
- Pathmasiri W, El-Seedi HR, Han X, Janson J-C, Huss U, Bohlin L. 2005. Aryl ketones from *Acronychia pedunculata* with cyclooxygenase-2 inhibitory effects. *Chem. Biodivers.* **2**: 463-469.
- Pathmasiri W, el-Seedi HR, Han X, Janson JC, Huss U, Bohlin L. 2005. Aryl ketones from *Acronychia pedunculata* with cyclooxygenase-2 inhibitory effects. *Chem. Biodivers.* **2**: 463-469.
- Pauli GF, Pro SM, Friesen JB. 2008. Countercurrent Separation of Natural Products. *J. Nat. Prod.* **71**: 1489-1508.
- Pergola C, Werz O. 2010. 5-Lipoxygenase inhibitors: a review of recent developments and patents. *Expert Opin. Ther. Pat.* **20**: 355-375.
- Pinkston JD, Wen D, Morand KL, Tirey DA, Stanton DT. 2006. Comparison of LC/MS and SFC/MS for Screening of a Large and Diverse Library of Pharmaceutically Relevant Compounds. *Anal. Chem.* **78**: 7467-7472.
- Pluskal T, Castillo S, Villar-Briones A, Oresic M. 2010. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**: 395.
- Porzel A, Farag M, Mülbradt J, Wessjohann L. 2014. Metabolite profiling and fingerprinting of *Hypericum* species: a comparison of MS and NMR metabolomics. *Metabolomics* **10**: 574-588.
- Potterat O, Hamburger M. 2013. Concepts and technologies for tracking bioactive compounds in natural product extracts: generation of libraries, and hyphenation of analytical processes with bioassays. *Nat. Prod. Rep.* **30**: 546-564.
- Prager RH, Thredgold HM. 1966. Some neutral constituents of *Acronychia baueri*. *Aust. J. Chem.* **19**: 451-454.
- Qiu X, Zhang J, Huang Z, Zhu D, Xu W. 2013. Profiling of phenolic constituents in *Polygonum multiflorum* Thunb. by combination of ultra-high-pressure liquid chromatography with linear ion trap-Orbitrap mass spectrometry. *J. Chromatogr. A* **1292**: 121-131.
- Rådmark O, Werz O, Steinhilber D, Samuelsson B. 2007. 5-Lipoxygenase: regulation of expression and enzyme activity. *Trends Biochem. Sci.* **32**: 332-341.
- Rahmani M, Taufiq-Yap YH, Sukari MA, Ismail HBM. 1996. Constituents of *Acronychia laurifolia*. *Fitoterapia* **67**: 180.
- Rasmussen B, Cloarec O, Tang H, Staerk D, Jaroszewski JW. 2006. Multivariate analysis of integrated and full-resolution ^1H -NMR spectral data from complex pharmaceutical preparations: St. John's wort. *Planta Med.* **72**: 556-563.
- Richards SE, Dumas M-E, Fonville JM, Ebbels TMD, Holmes E, Nicholson JK. 2010. Intra- and inter-omic fusion of metabolic profiling data in a systems biology framework. *Chemom. Intell. Lab. Syst.* **104**: 121-131.
- Robinette SL, Brüscheiler R, Schroeder FC, Edison AS. 2011. NMR in Metabolomics and Natural Products Research: Two Sides of the Same Coin. *Acc. Chem. Res.* **45**: 288-297.

- Rodrigo SK, Jayasinghe ULB, Bandara BMR. 2007. Antifungal, Antioxidant and Cytotoxic Activity of *Acronychia pedunculata* and *Adenanthera pavonina*. *Proceedings of the Peradeniya University Research Sessions* **12**: 94-95.
- Roldan M, Engel B, de Vos RH, Vereijken P, Astola L, Groenenboom M, van de Geest H, Bovy A, Molenaar J, van Eeuwijk F, Hall R. 2014. Metabolomics reveals organ-specific metabolic rearrangements during early tomato seedling development. *Metabolomics* **10**: 958-974.
- Roos G, Röseler C, Büter KB, Simmen U. 2004. Classification and Correlation of St. John's Wort Extracts by Nuclear Magnetic Resonance Spectroscopy, Multivariate Data Analysis and Pharmacological Activity. *Planta Med.* **70**: 771-777.
- Rouzer CA, Ford-Hutchinson AW, Morton HE, Gillard JW. 1990. MK886, a potent and specific leukotriene biosynthesis inhibitor blocks and reverses the membrane association of 5-lipoxygenase in ionophore-challenged leukocytes. *J. Biol. Chem.* **265**: 1436-1442.
- Ruey-Shiuan T, Pierre-Alain C, Bernard T. (1995) Measurement of Partition Coefficient Using Centrifugal Partition Chromatography. *Modern Countercurrent Chromatography*, American Chemical Society. **593**, 143-154.
- Safer S, Cicek SS, Pieri V, Schwaiger S, Schneider P, Wissemann V, Stuppner H. 2011. Metabolic fingerprinting of *Leontopodium* species (Asteraceae) by means of ¹H NMR and HPLC-ESI-MS. *Phytochemistry* **72**: 1379-1389.
- Sangster TP, Wingate JE, Burton L, Teichert F, Wilson ID. 2007. Investigation of analytical variation in metabolomic analysis using liquid chromatography/mass spectrometry. *Rapid Commun. Mass Spectrom.* **21**: 2965-2970.
- Sarker S, Nahar L. (2012) Hyphenated Techniques and Their Applications in Natural Products Analysis. *Natural Products Isolation*. Sarker S. D., Nahar L., Humana Press. **864**, 301-340.
- Sawada M, Shizuma M, Takai Y, Yamada H, Kaneda T, Hanafusa T. 1992. Enantioselectivity in fast-atom bombardment (FAB) mass spectrometry. *J. Am. Chem. Soc.* **114**: 4405-4406.
- Schripsema J. 2010. Application of NMR in plant metabolomics: techniques, problems and prospects. *Phytochem. Anal.* **21**: 14-21.
- Seeger C, Sturm S, Stuppner H. 2013. Mass spectrometry and NMR spectroscopy: modern high-end detectors for high resolution separation techniques-state of the art in natural product HPLC-MS, HPLC-NMR, and CE-MS hyphenations. *Nat. Prod. Rep.* **30**: 970-987.
- Shulaev V, Cortes D, Miller G, Mittler R. 2008. Metabolomics for plant stress response. *Physiol. Plantarum* **132**: 199-208.
- Simmler C, Napolitano JG, McAlpine JB, Chen S-N, Pauli GF. 2014. Universal quantitative NMR analysis of complex natural samples. *Curr. Opin. Chem. Biol.* **25**: 51-59.
- Smith C, Want E, O'Maille G, Abagyan R, Siuzdak G. 2006. XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching and identification. *Anal. Chem.* **78**: 779-787.
- Smith CA. 2014. LC/MS Preprocessing and Analysis with xcms. <http://www.bioconductor.org/>.
- Smolinska A, Blanchet L, Buydens LMC, Wijmenga SS. 2012. NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review. *Anal. Chim. Acta* **750**: 82-97.
- Steinfath M, Groth D, Lisec J, Selbig J. 2008. Metabolite profile analysis: from raw data to regression and classification. *Physiol. Plantarum* **132**: 150-161.

- Storer R, Young DW. 1972. Preskimmianine: The Biogenetic Precursor of Skimmianine from *Dictamnus albus* L. *Tetrahedron Lett.* **13**: 2199-2202.
- Su CR, Kuo PC, Wang ML, Liou MJ, Damu AG, Wu TS. 2003. Acetophenone derivatives from *Acronychia pedunculata*. *J. Nat. Prod.* **66**: 990-993.
- Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reilly MD, Thaden JJ, Viant MR. 2007. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**: 211-221.
- Sumner LW, Mendes P, Dixon RA. 2003. Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* **62**: 817-836.
- Tapp HS, Kemsley EK. 2009. Notes on the practical utility of OPLS. *TrAC Trends Anal. Chem.* **28**: 1322-1327.
- Tautenhahn R, Bottcher C, Neumann S. 2008. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* **9**: 504.
- Taylor LT. 2009. Supercritical fluid chromatography for the 21st century. *J Supercrit Fluids* **47**: 566-573.
- Taylor LT. 2010. Supercritical Fluid Chromatography. *Anal Chem* **82**: 4925-4935.
- Tian O, Schwartz SJ. 2003. Mass spectrometry and tandem mass spectrometry of *Citrus* limonoids. *Anal. Chem.* **75**: 5451-5460.
- Tikunov Y, Lommen A, de Vos CHR, Verhoeven HA, Bino RJ, Hall RD, Bovy AG. 2005. A Novel Approach for Nontargeted Data Analysis for Metabolomics. Large-Scale Profiling of Tomato Fruit Volatiles. *Plant Physiol.* **139**: 1125-1137.
- Trygg J, Holmes E, Lundstedt T. 2006. Chemometrics in Metabonomics. *J. Proteome Res.* **6**: 469-479.
- Trygg J, Wold S. 2002. Orthogonal projections to latent structures (O-PLS). *J. Chemometr.* **16**: 119-128.
- Tyler VE. 1999. Phytochemicals: Back to the Future. *J. Nat. Prod.* **62**: 1589-1592.
- van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **7**: 1-15.
- van der Greef J, Smilde AK. 2005. Symbiosis of chemometrics and metabolomics: past, present, and future. *J. Chemometr.* **19**: 376-386.
- van der Hooft JJ, de Vos RH, Ridder L, Vervoort J, Bino R. 2013. Structural elucidation of low abundant metabolites in complex sample matrices. *Metabolomics* **9**: 1009-1018.
- van der Hooft JJ, Vervoort J, Bino R, de Vos RH. 2012. Spectral trees as a robust annotation tool in LC-MS based metabolomics. *Metabolomics* **8**: 691-703.
- van der Hooft JJJ, Vervoort J, Bino RJ, Beekwilder J, de Vos RCH. 2010. Polyphenol identification based on systematic and robust high-resolution accurate mass spectrometry fragmentation. *Anal. Chem.* **83**: 409-416.
- Verpoorte R, Choi YH, Kim HK. 2007. NMR-based metabolomics at work in phytochemistry. *Phytochem. Rev.* **6**: 3-14.
- Verron T, Sabatier R, Joffre R. 2004. Some theoretical properties of the O-PLS method. *J. Chemometr.* **18**: 62-68.
- Vervoort N, Daemen D, Torok G. 2008. Performance evaluation of evaporative light scattering detection and charged aerosol detection in reversed phase liquid chromatography. *Journal of Chromatography A* **1189**: 92-100.
- Viant MR. 2003. Improved methods for the acquisition and interpretation of NMR metabolomic data. *Biochem. Biophys. Res. Commun.* **310**: 943-948.

- Wang C. (1961) The Forests of China. Massachusetts: Harvard University Maria Moors Cabot Foundation
- Wang D, Liu Z, Guo M, Liu S. 2004. Structural elucidation and identification of alkaloids in *Rhizoma Coptidis* by electrospray ionization tandem mass spectrometry. *J. Mass Spectrom.* **39**: 1356-1365.
- Wang T, Shao K, Chu Q, Ren Y, Mu Y, Qu L, He J, Jin C, Xia B. 2009. Automics: an integrated platform for NMR-based metabolomics spectral processing and data analysis. *BMC Bioinformatics* **10**: 83.
- Wang Y, Tang H, Nicholson JK, Hylands PJ, Sampson J, Whitcombe I, Stewart CG, Caiger S, Oru I, Holmes E. 2004. Metabolomic Strategy for the Classification and Quality Control of Phytomedicine: A Case Study of Chamomile Flower (*Matricaria recutita* L.). *Planta Med.* **70**: 250-255.
- Want EJ, Wilson ID, Gika H, Theodoridis G, Plumb RS, Shockcor J, Holmes E, Nicholson JK. 2010. Global metabolic profiling procedures for urine using UPLC-MS. *Nat. Protoc.* **5**: 1005-1018.
- Weckwerth W, Morgenthal K. 2005. Metabolomics: from pattern recognition to biological interpretation. *Drug Discov. Today* **10**: 1551-1558.
- Weissberg A, Dagan S. 2011. Interpretation of ESI(+)-MS-MS spectra—Towards the identification of “unknowns”. *Int. J. Mass Spectrom.* **299**: 158-168.
- Wen-Hao X, Zhi X. 1984. Chemical studies on alkaloids in the root of *Acronychia oligophylebia* Merr. *Acta Chim. Sinica* **2**: 66-73.
- Wenzel SE, Kamada AK. 1996. Zileuton: the first 5-lipoxygenase inhibitor for the treatment of asthma. *Ann. Pharmacother.* **30**: 858-864.
- Werz O, Szellas D, Steinhilber D, Rådmark O. 2002. Arachidonic Acid Promotes Phosphorylation of 5-Lipoxygenase at Ser-271 by MAPK-activated Protein Kinase 2 (MK2). *J. Biol. Chem.* **277**: 14793-14800.
- Werz OBESBROSD. 2002. Activation of 5-lipoxygenase by cell stress is calcium independent in human polymorphonuclear leukocytes. *Blood* **99**: 1044-1052.
- White C. 2005. Integration of supercritical fluid chromatography into drug discovery as a routine support tool: Part I. Fast chiral screening and purification. *Journal of Chromatography A* **1074**: 163-173.
- White C, Burnett J. 2005. Integration of supercritical fluid chromatography into drug discovery as a routine support tool: II. Investigation and evaluation of supercritical fluid chromatography for achiral batch purification. *Journal of Chromatography A* **1074**: 175-185.
- Wink M. 2003. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* **64**: 3-19.
- Wold S, Esbensen K, Geladi P. 1987. Principal component analysis. *Chemom. Intell. Lab. Syst.* **2**: 37-52.
- Wold S, Sjöström M, Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **58**: 109-130.
- Wolfender J-L, Marti G, Ferreira Queiroz E. 2010. Advances in Techniques for Profiling Crude Extracts and for the Rapid Identification of Natural Products: Dereplication, Quality Control and Metabolomics. *Curr. Org. Chem.* **14**: 1808-1832.
- Wolfender JL, Queiroz EF. 2012. New approaches for studying the chemical diversity of natural resources and the bioactivity of their constituents. *Chimia (Aarau)* **66**: 324-329.
- Wolfender JL, Queiroz EF, Hostettmann K. 2006. The importance of hyphenated techniques in the discovery of new lead compounds from nature. *Expert Opin. Drug Discov.* **1**: 237-260.

References

- Wolfender JL, Rudaz S, Choi YH, Kim HK. 2013. Plant metabolomics: from holistic data to relevant biomarkers. *Curr. Med. Chem.* **20**: 1056-1090.
- Wolfender JL, Waridel P, Ndjoko K, Hobby KR, Major HJ, Hostettmann K. 2000. Evaluation of Q-TOF-MS/MS and multiple stage IT-MSn for the dereplication of flavonoids and related compounds in crude plant extracts. *Analysis* **28**: 895-906.
- Worley B, Powers R. 2014. MVAPACK: A Complete Data Handling Package for NMR Metabolomics. *ACS Chem. Biol.* **9**: 1138-1144.
- Wu T-S, Wang M-L, Jong T-T, McPhail AT, McPhail DR, Lee K-H. 1989. X-Ray crystal structure of acrovestone, a cytotoxic principle from *Acronychia pedunculata*. *J. Nat. Prod.* **52**: 1284-1289.
- Wu TS, Wang ML, Jong TT, McPhail AT, McPhail DR, Lee KH. 1989. X-ray crystal structure of acrovestone, a cytotoxic principle from *Acronychia pedunculata*. *J. Nat. Prod.* **52**: 1284-1289.
- Xia J, Psychogios N, Young N, Wishart DS. 2009. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* **37**: W652-660.
- Xia J, Wishart DS. 2011. Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nat. Protoc.* **6**: 743-760.
- Yuliana N, Jahangir M, Verpoorte R, Choi Y. 2013. Metabolomics for the rapid dereplication of bioactive compounds from natural sources. *Phytochem. Rev.* **12**: 293-304.
- Yuliana ND, Khatib A, Choi YH, Verpoorte R. 2011. Metabolomics for bioactivity assessment of natural products. *Phytother. Res.* **25**: 157-169.
- Yuliana ND, Khatib A, Verpoorte R, Choi YH. 2011. Comprehensive Extraction Method Integrated with NMR Metabolomics: A New Bioactivity Screening Method for Plants, Adenosine A1 Receptor Binding Compounds in *Orthosiphon stamineus* Benth. *Anal. Chem.* **83**: 6902-6906.
- Zhang J Fau - Huang ZH, Huang Zh Fau - Qiu XH, Qiu Xh Fau - Yang YM, Yang Ym Fau - Zhu DY, Zhu da Y Fau - Xu W, Xu W. 2012. Neutral fragment filtering for rapid identification of new diester-diterpenoid alkaloids in roots of *Aconitum carmichaeli* by ultra-high-pressure liquid chromatography coupled with linear ion trap-orbitrap mass spectrometry. *PLoS One* **7**: e52352.
- Zhang T, Nguyen D, Franco P. 2008. Enantiomer resolution screening strategy using multiple immobilised polysaccharide-based chiral stationary phases. *J. Chromatogr. A* **1191**: 214-222.
- Zhang Y, Wen Z, Washburn MP, Florens L. 2009. Effect of Dynamic Exclusion Duration on Spectral Count Based Quantitative Proteomics. *Anal. Chem.* **81**: 6317-6326.
- Zhi H-J, Qin X-M, Sun H-F, Zhang L-Z, Guo X-Q, Li Z-Y. 2012. Metabolic Fingerprinting of *Tussilago farfara* L. Using ¹H-NMR Spectroscopy and Multivariate Data Analysis. *Phytochem. Anal.* **23**: 492-501.
- Zubarev RA, Makarov A. 2013. Orbitrap mass spectrometry. *Anal. Chem.* **85**: 5288-5296.