

# **The Complexity of Natural Products at the Chemistry–Biology Interface**

BY

FENG QIU

B.S., Chengdu University of Traditional Chinese Medicine, 2005

DISSERTATION

Submitted as partial fulfillment of the requirements  
for the degree of Doctor of Philosophy in Pharmacognosy  
in the Graduate College of the  
University of Illinois at Chicago, 2013

Chicago, Illinois

Defense Committee:

Dr. Guido F. Pauli, Chair and Advisor  
Dr. Jimmy Orjala  
Dr. Shaonong Chen  
Dr. J. Brent Friesen  
Dr. James B. McAlpine  
Dr. Daneel Ferreira, The University of Mississippi  
Dr. David S. Seigler, University of Illinois at Urbana-Champaign

This dissertation is dedicated to my parents for their love and support, without which this work would not have been possible.

## ACKNOWLEDGEMENTS

This dissertation was only made possible by the efforts and support of many individuals. First and foremost, I would like to thank my advisor, Dr. Guido F. Pauli, for his guidance and constant support throughout the course of my Ph.D. study. His extensive knowledge and meticulous attitude taught me how to be an excellent scientist and inspired me to dedicate myself to the field of Pharmacognosy. I am very grateful to the late professor, Dr. Norman R. Farnsworth, for serving on my preliminary examination. He impressed me not only by his profound knowledge of Pharmacognosy, but also by his humor, generosity, and unbounded enthusiasm. I also thank Drs. Steve M. Swanson and Brian T. Murphy for serving as committee members on my preliminary examination. I express my appreciation to Dr. James B. McAlpine for his continued help for editing my scientific writings, without which none of my proposals, publications and dissertation would have been completed. I thank Dr. J. Brent Friesen for sharing his insightful knowledge of countercurrent chromatography and resolving many complex instrumentation problems. I also thank Dr. David C. Lankin for his support in NMR experiments as well as suggestions on my dissertation. I am also thankful to Dr. Shao-Nong Chen for his enormous help and kind advice in every aspect of my research.

I would like to thank Dr. Birgit U. Jaki for her guidance during my lab rotation. My thanks also go to all my lab colleagues. Dr. Tanja Gödecke prepared plant extract for my project. Dr. José G. Napolitano provided me with valuable suggestions on NMR experiments. Dr. Shi-Hui Dong helped patiently to proofread my manuscripts. Drs. María F. Rodríguez-Brasco, Ayano Imai, Chang Hwa Hwang, Kaisu Riihinen, Charlotte Simmler, and Mr. Rene F. Ramos Alvarenga also provided their enthusiastic help in my research work. I feel privileged to work with all of them who never hesitated to share their expertise and experiences, asking nothing in return. In addition, many thanks go to Sanghyun Cho and Geping Cai at ITR/UIC for conducting the TB biological experiments presented in this dissertation.

## **ACKNOWLEDGEMENTS (Continued)**

I kindly acknowledge helpful discussions and NMR instrument support provided by the following individuals at the National Research Council, Halifax, Canada: Mr. Ian Burton as well as Drs. Tobias Karakach and John Walter. I am also grateful for CS hardware support provided by the following individuals at Wrightwood Technologies, Chicago (IL): Samuel Pro, Warren Friedl, Luke Pro, Tom Burdick, and Nick Novak, as well as Dr. Lucas Chadwick.

Finally, I am indebted to the Department of Medicinal Chemistry and Pharmacognosy for providing teaching assistantships and tuition waiver for my first two years here, and the College of Pharmacy for offering me a W.E. van Doren Scholarship. The work presented in this dissertation was made possible by the following grants from the National Center for Complementary and Alternative Medicine (NCCAM) of the National Institutes of Health (NIH): P50 AT000155, RC2 AT005899, and R44 AT004534.

FQ

## TABLE OF CONTENTS

1. A REVIEW OF NATURAL PRODUCT COMPLEXITY.....	1
1.1 Introduction.....	1
1.2 Chemical Complexity of Natural Products.....	2
1.2.1 Structural Diversity and Similarity.....	2
1.2.2 Chemical Variations.....	4
1.2.3 Steep Concentration Gradient.....	5
1.2.4 Residual Complexity.....	5
1.2.5 Methods and Challenges.....	6
1.3 Biological Complexity of Natural Products.....	15
1.3.1 Synergistic and Multitarget Effects.....	15
1.3.2 Methods and Challenges.....	17
1.4 Development of Hypotheses.....	19
2. EXPERIMENTAL.....	23
2.1 Materials.....	23
2.1.1 Plant Materials and Extracts.....	23
2.1.2 Solvents.....	24
2.2 Methods.....	25
2.2.1 Extraction.....	25
2.2.2 Chromatography.....	26
2.2.3 Nuclear Magnetic Resonance Spectroscopy.....	32
2.2.4 Antituberculosis Activity Evaluation.....	33
2.3 Data Analysis.....	35
3. ACCELERATING AND TARGETING OF NATURAL PRODUCT SEPARATION.....	36
3.1 Introduction.....	36
3.2 Sample Cutting.....	40
3.2.1 Enrichment of Terpene Lactones from <i>Ginkgo biloba</i> .....	40
3.2.2 Enrichment of Prenylphenols from <i>Humulus lupulus</i> .....	42
3.2.3 Enrichment of Lipophilic Components from <i>Humulus lupulus</i> .....	44
3.3 Prediction of Countercurrent Separation of Ginkgo Terpene Lactones.....	46
3.4 Resolution of <i>Actaea</i> Triterpenes by Chromatographic Orthogonality.....	51
3.5 Solvent Orthogonality.....	54
3.5.1 Resolution of Ginkgo Terpene Lactones.....	54
3.5.2 Resolution of Prenylphenols from <i>Humulus lupulus</i> .....	58
3.5.3 Resolution of <i>Actaea</i> Triterpenes.....	60
3.6 Conclusion.....	64
4. EXPANDING NMR UTILITY IN NATURAL PRODUCT CHARACTERIZATION.....	65
4.1 Introduction.....	65
4.2 Visualization of Elution Profiles.....	69

## TABLE OF CONTENTS (Continued)

4.2.1	CS of Crude Extract of <i>Camellia sinensis</i> .....	69
4.2.2	CS of Ginkgo Terpene Lactones.....	73
4.2.3	HPLC Purification of Triterpenes from <i>Oplopanax horridus</i> .....	79
4.3	Evaluation of Residual Complexity of Triterpenes from <i>Oplopanax horridus</i> .....	82
4.4	Identification of a New <i>Actaea</i> Triterpene in a Residually Complex Mixture.....	85
4.5	Dereplication of <i>Actaea</i> Triterpenes in Residually Complex Mixtures.....	92
4.6	Conclusion.....	100
5.	RATIONAL NAMING AND VIRTUAL PARTITIONING OF NATURAL PRODUCTS.....	101
5.1	Introduction.....	101
5.2	The New Naming System.....	106
5.3	Construction of Database.....	111
5.4	Canonical Discriminant Analysis.....	115
5.5	Development of Classification Binary Trees.....	118
5.6	Dereplication of <i>Actaea</i> Triterpenes in Residually Complex Mixtures.....	125
5.7	Conclusion.....	133
6.	INTEGRATING CHEMISTRY AND BIOLOGY OF NATURAL PRODUCTS.....	135
6.1	Quantitative Purity–Activity Relationship of Triterpenes from <i>Oplopanax horridus</i> .....	135
6.2	Biochemometric Evaluation of Antituberculosis Principles in <i>Humulus lupulus</i> .....	141
6.3	Conclusion.....	146
7.	SUMMARY AND PERSPECTIVES.....	147
	CITED LITERATURE.....	154
	APPENDICES.....	162
	Appendix A Compound Index.....	162
	Appendix B <sup>1</sup> H NMR Spectra and Data of Isolated Compounds.....	166
	Appendix C VBA Code of ActaMatch.....	180
	Appendix D VBA Code of ActaPredict.....	188
	Appendix E Database of Methyl NMR Data of <i>Actaea</i> Triterpenes.....	192
	VITA.....	217

## LIST OF TABLES

TABLE I. THE DESIGN OF THE PRESENT STUDY.....	22
TABLE II. STANDARDIZED CS SS IN THE ChMWat AND HEMWat FAMILIES.....	30
TABLE III. PARTITION COEFFICIENTS OF FIVE TERPENE LACTONES IN THREE POLARITY-ADJUSTED SS.....	49
TABLE IV. <sup>1</sup> H AND <sup>13</sup> C NMR COMPARISON BETWEEN 27-HYDROXYACTEIN ( <b>29</b> ) AND ACTEIN ( <b>15</b> ).....	90
TABLE V. EXAMPLES OF ACTAEA TRITERPENES AND THEIR CHARACTERISTIC HMBC CORRELATIONS OF Me-26 AND Me-27.....	94
TABLE VI. REPRESENTATIVES OF THE ACTAEA TRITERPENES.....	109
TABLE VII. THE MAJOR TYPES OF ACTAEA TRITERPENES INCLUDED IN THE IN-HOUSE DATABASE, ALONG WITH THEIR COMMON AND NEW SYSTEMATIC NAMES.....	112
TABLE VIII. CLASSIFICATION RESULTS FOR ACTAEA TRITERPENES CONTAINING SEVEN METHYL GROUPS WITH $\delta_H < 1.90$ USING THE CDA ANALYSIS.....	115
TABLE IX. CLASSIFICATION RESULTS FOR ACTAEA TRITERPENES CONTAINING SEVEN METHYL GROUPS WITH $\delta_H < 1.90$ USING THE CBT-7 PARTITIONING.....	121
TABLE X. THE ACCURACY OF THE PREDICTION PERFORMANCE OF THE CBT-7 ESTIMATED BY LEAVE-ONE-OUT CROSS-VALIDATION.....	123
TABLE XI. DEREPLICATION RESULTS OF THE TRITERPENES CONTAINED IN THE RESIDUALLY COMPLEX SAMPLES Ar.18.1.1.7 AND Ar.18.1.3.2.4.....	132
TABLE XII. qHNMR IMPURITY PROFILES AND ANTI-TB ACTIVITY OF THE <i>O. HORRIDUS</i> TRITERPENE FRACTIONS AGAINST STRAIN H <sub>37</sub> Rv.....	137
TABLE XIII. ANTI-TB ACTIVE PRINCIPLES IN HOPS EXTRACT WITH PEARSON'S CORRELATIONS $\geq 0.80$ .....	144

## LIST OF FIGURES

Figure 1. Structural Diversity and Similarity of NPs Exemplified for the Complex Chemical Constituents of <i>Actaea racemosa</i> .....	3
Figure 2. Synergistic and Multitarget Effects of NPs.....	16
Figure 3. Five Plants Selected as Test Cases in the Present Study.....	23
Figure 4. The Conceptual Design of a New Separation Scheme for NPs.....	37
Figure 5. Preparation of Terpene Lactone-Enriched Sample from Ginkgo Leaf Extract by Liquid-Liquid Partitioning.....	40
Figure 6. Identification of Biomarker Compounds in Ginkgo Preparations by <sup>1</sup> H NMR.....	41
Figure 7. Preparation of a Prenylphenol-Enriched Sample from Hops Crude Extract by Liquid-Liquid Partitioning.....	43
Figure 8. Example of a Sample-Cutting Procedure Facilitated by HSCCC Fractionation.....	44
Figure 9. The Protocol for Determining Partition Coefficients of Multiple Components in Mixtures by qHNMR.....	48
Figure 10. Preparative Separation of <i>Actaea</i> Triterpenes by Using Orthogonal Chromatographic and Solvent Systems.....	52
Figure 11. Proposed Procedure for the Separation of Five Ginkgo Terpene Lactones.....	55
Figure 12. Orthogonal CS Process for the Separation of Three Flavonoids and Five Ginkgo Terpene Lactones.....	57
Figure 13. Resolving Three Congeners from Hops by HSCCC with Orthogonal SSs.....	59
Figure 14. Resolving Four Congeners from Black Cohosh by VLC and Orthogonal SSs.....	61
Figure 15. Application of NMR as an Offline Elution Detector.....	66
Figure 16. Offline qHNMR Analysis of HSCCC Fractionation of Green Tea Extract.....	70
Figure 17. The Proposed CCC-NMR Method for Chemical Standardization (Quality Control) of Crude Extracts.....	72
Figure 18. Offline qHNMR Analysis of the 2nd Step of HSCCC Purification of GA and GB.....	74
Figure 19. Matching and Prediction of Compound Purity by Gaussian Fitting of qHNMR-based Elution Profiles of GA and GB.....	77
Figure 20. Offline qHNMR Analysis of the 2nd Step of HSCCC Purification of GC and GJ.....	78

## LIST OF FIGURES (Continued)

Figure 21. Offline qHNMR Analysis of the Preparative HPLC Purification of Triterpenes from <i>O. horridus</i> .....	81
Figure 22. Impurity Profiling of an <i>O. horridus</i> Triterpene Sample (Oh.17.8.21) by COSY Analysis.....	83
Figure 23. Discovery of a “Hidden” Component by Inconsistencies between TLC and <sup>1</sup> H NMR Profiles.....	86
Figure 24. Identification of a New Triterpenoid ( <b>29</b> ) as a Minor Component in a Residually Complex Sample by 2D NMR Spectral Subtraction.....	88
Figure 25. The Residual Complexity of the Fraction Ar.18.1.1.4 Indicated by the <sup>1</sup> H NMR Signals of the Cycloartane Protons (H-19a) of <i>Actaea</i> Triterpenes.....	91
Figure 26. The Dereplication Problem for <i>Actaea</i> Triterpenes Arising from the Identification of Their Highly Variable Side Chains.....	93
Figure 27. Identification of Three Major <i>Actaea</i> Triterpenes in a Residually Complex Sample (Ar.18.1.3.2.4) by Pattern Recognition of Characteristic HMBC Correlations.....	95
Figure 28. Resolving the Residual Complexity of an <i>Actaea</i> Triterpene Sample (Ar.18.3.1) by Pattern Recognition of Characteristic HMBC Correlations.....	97
Figure 29. The Residual Complexity of the Sample Ar.18.1.3.1 Indicated by the <sup>1</sup> H NMR Signals of the Cycloartane Protons (H-19a) of <i>Actaea</i> Triterpenes.....	98
Figure 30. The Concept of Using Methyl Groups of Triterpenes as Partial Structural Indicators (“Me Cams”) to Map the Full Skeleton of Molecules.....	103
Figure 31. A Comparison between the Processes of Liquid and Virtual Partitioning of NPs.....	104
Figure 32. The Aglycone of Tetracyclic <i>Actaea</i> Triterpenes.....	106
Figure 33. The Basic Rules of the New Naming System for <i>Actaea</i> Triterpenes.....	107
Figure 34. Example of the Change of CIP Designation without a Change in Configuration.....	108
Figure 35. Structures of the Representative Subclasses of the <i>Actaea</i> Triterpenes.....	110
Figure 36. A Screenshot and the Composition of the In-house Database with Respect to Triterpene Classes.....	113
Figure 37. A Screenshot and Operational Procedures of ActaMatch Using Compound <b>12</b> as an Example.....	115
Figure 38. Classification of All <i>Actaea</i> Triterpenes Contained in the In-house Database by CDA Analysis.....	117

## LIST OF FIGURES (Continued)

Figure 39. The CBTs Developed for the Classification of <i>Actaea</i> Triterpenes with Five (CBT-5) and Six (CBT-6) Me Groups ( $\delta_H < 1.90$ ).....	118
Figure 40. The CBT Developed for Classification of <i>Actaea</i> Triterpenes with Seven Me Groups ( $\delta_H < 1.90$ ).....	119
Figure 41. A Screenshot and Operational Procedures of ActaPredict Using Compound <b>12</b> as an Example.....	124
Figure 42. The New <i>In Silico</i> Tool ActaFinder.....	126
Figure 43. Dereplication of <i>Actaea</i> Triterpenes in a Residually Complex Sample (E1).....	127
Figure 44. Dereplication of <i>Actaea</i> Triterpenes in a Residually Complex Sample (E2).....	130
Figure 45. The Experimental Design for the Establishment of Quantitative Purity–Activity Relationships of NPs.....	136
Figure 46. Purity–Activity Relationships of the <i>O. horridus</i> Triterpene Fractions.....	139
Figure 47. The Major Steps of the Biochemometric Approach.....	142
Figure 48. Step 4 of the Biochemometric Evaluation by Pearson’s Correlation Analysis.....	143
Figure 49. The Contour Plot of 3D HSCCC-GC-MS Matrix.....	145
Figure 50. The Comparison between the Classic Reductionist Model and the Proposed Gut Model in Bioactive NP Research.....	152
Figure 51. A Flowchart for Future NP Research Deriving from the “Gut Model” .....	153

## LIST OF ABBREVIATIONS

MeCN	acetonitrile
Å	Ångstrom
AQ	data acquisition time (in second) for the NMR spectra
BGF	bioassay-guided fractionation
<i>n</i> -BuOH	<i>n</i> -butanol
CBT	classification binary tree
CCC	countercurrent chromatography
CDA	canonical discriminant analysis
CDCl <sub>3</sub>	deuterated chloroform
CHCl <sub>3</sub>	chloroform
ChMWat	CS solvent system consisting of chloroform, methanol and water
CIP	Cahn-Ingold-Prelog (priority rules for stereoconfiguration)
COSY	<sup>1</sup> H- <sup>1</sup> H correlation spectroscopy
CS	countercurrent separation
δ	chemical shift (ppm)
CH <sub>2</sub> Cl <sub>2</sub> or DCM	dichloromethane
DMSO	dimethyl sulfoxide
ELSD	evaporative light scattering detector
EtOAc	ethyl acetate
EtOH	ethanol
eV	electron Volt
FID	free induction decay (in NMR)
Fr	chromatographic fraction
GC	gas chromatography

## LIST OF ABBREVIATIONS (Continued)

GUESS	general useful estimation of solvent systems
H <sub>2</sub> O	deionized water
H <sub>2</sub> SO <sub>4</sub>	sulfuric acid
HEMWat	CS solvent system consisting of <i>n</i> -hexane, ethyl acetate, methanol and water
HEMSoWat	HEMWat solvent system modified with dimethyl sulfoxide
HterAc	CS solvent system consisting of <i>n</i> -hexane, methyl tert-butyl ether and acetonitrile
HMBC	heteronuclear multiple-bond correlation spectroscopy ( <sup>1</sup> H– <sup>13</sup> C)
HSQC	heteronuclear single quantum coherence spectroscopy ( <sup>1</sup> H– <sup>13</sup> C)
HPLC	high-performance liquid chromatography
HSCCC	high-speed countercurrent chromatography
Hz	Hertz
i.d.	inner diameter
IR	infrared spectroscopy
<i>K</i>	partition coefficient
$\lambda$	wavelength (nm)
LG	Lorentzian–Gaussian window function in NMR processing
LOD	limit of detection
MABA	microplate Alamar blue assay
Me	methyl group
MeOH	methanol
mg	milligram(s)
$\mu$ g	microgram(s)
ng	nanogram(s)

## LIST OF ABBREVIATIONS (Continued)

mL	milliliter(s)
$\mu\text{L}$	microliter(s)
MHz	megaHertz
MIC	minimum inhibitory concentration ( $\mu\text{g/mL}$ )
min	minute
mm	millimeter(s)
mol	mole
MS	mass spectrometry
MTBE	methyl tert-butyl ether
NMR	nuclear magnetic resonance
NP	natural product
NP-	normal phase (chromatography)
NS	number of scans
OAc	acetoxy group
o1p	transmitter offset (in NMR)
PCA	principal component analysis
pH	$-\log_{10}[\text{H}^+]$ (measure of acidity)
ppm	parts per million
qHNMR	quantitative proton ( $^1\text{H}$ ) NMR
QPAR	quantitative purity–activity relationship
<i>r</i>	Pearson's correlation coefficient
$R_f$	retention factor: migration distance of an analyte as a fraction of distance of origin to solvent front in thin-layer chromatography
RG	receiver gain setting (in NMR)
RP	reversed phase (chromatography)

## LIST OF ABBREVIATIONS (Continued)

rpm	rotations per minute
$R_s$	peak resolution in chromatography
SCE	single chemical entity
$S_f$	retention factor: volume of stationary phase as a fraction of total coil volume in countercurrent chromatography
SS	solvent system
SW	spectral width (in NMR)
TB	tuberculosis
TD	number of time domain points (in NMR)
TLC	thin-layer chromatography
UV	ultraviolet light, i.e., 20–400 nm
VBA	Visual Basic Application
Vis	visible light, i.e., 400–700 nm
VLC	vacuum liquid chromatography

## PREFACE

The diverse structures of natural products (NPs) contribute to a broad range of bioactivities as well as a large number of molecular scaffolds for drug discovery and development. However, the chemical and biological complexity associated with NPs also provides enormous difficulties and challenges for mining these molecules (see **Chapter 1**). This dissertation demonstrates new concepts and methodologies for the isolation, structural determination, and biological evaluation of NPs. Two powerful tools, countercurrent separation (CS) and nuclear magnetic resonance (NMR), are highlighted and used extensively in the present study for unraveling the chemical and biological complexity of NPs from exemplified plants, including *Actaea racemosa*, *Camellia sinensis*, *Ginkgo biloba*, *Humulus lupulus*, and *Oplopanax horridus*.

To expedite the separation process for NPs, **Chapter 3** develops several methods which can be implemented to enhance conventional fractionation schemes. Sample-cutting was initially applied as a pre-separation procedure to concentrate the target compounds from the complex mixtures. A combination of “shake-flask” experiments with qHNMR analysis enables simultaneous measurement of partition coefficients ( $K$ ) of multiple components in mixtures. The  $K$  values can be used to predict CS elution profiles and to guide selection and optimization of CS solvent systems. While both polarity and selectivity are key factors influencing the separation performance, the latter is more vital for resolving NP congeners which exhibit similar chemical properties. The concept of chromatographic and solvent orthogonality is, therefore, proposed and explored for the design of efficient fractionation procedures.

In **Chapter 4**, the basic application of NMR is expanded to include several aspects of NP research. First, 1D qHNMR was used as an offline detector for chromatography. The combination of these two methods, chromatography and NMR, enables a new level of assessment of chemical compositions in chromatographic fractions. In particular, qHNMR analysis together with Gaussian curve fitting enables the quantitative representation of elution

## PREFACE (Continued)

profiles of any NMR-detectable analytes, regardless of whether they have been previously identified or not. In addition, the power of 2D NMR techniques was exploited for the analysis of the chemical complexity of NP mixtures. 2D H,H–Correlation via *J*-coupling was used for evaluation of the residual complexity of purified NPs. Differential analysis of HSQC spectra facilitated identification of new compounds in complex mixtures. Furthermore, pattern recognition of HMBC spectra enabled rapid dereplication of multiple NP congeners in residually complex samples. The resolving power of these tools was further enhanced by the use of high-sensitivity cryo-microprobe NMR instruments.

**Chapter 5** details a new concept of NP structural dereplication that takes advantage of easily discernible methyl <sup>1</sup>H NMR signals. Based on a virtual partitioning technique (classification binary trees, CBTs), a predictive computational model was generated that enables rapid dereplication of more than 170 known *Actaea* triterpenes and facilitates elucidation of new compounds. A combination of CBTs, <sup>1</sup>H NMR deconvolution, fingerprinting <sup>1</sup>H NMR signals, and qHNMR led to the unambiguous identification of minor constituents in residually complex triterpene samples. Upon assembling an in-house NMR database, a software application called ActaFinder was developed to assist automatic dereplication of *Actaea* triterpenes. These methodologies have the potential to be applicable to other classes of NPs.

In **Chapter 6**, a preliminary study was carried out for establishing quantitative purity–activity relationship (QPAR) of NPs based on the foregoing NMR-based analysis of residual complexity. It highlighted the importance of characterizing the biological effect of varying impurities which are common occurrences in NPs, even if they have been repeatedly purified. In order to evaluate potential synergistic effects originating from multiple components in the crude extract, a biochemometric approach was employed by using a combination of CS, GC-MS dereplication and statistical analysis. This led to identification of bioactive principles without the need of physical isolation.

# 1. A REVIEW OF NATURAL PRODUCT COMPLEXITY

## 1.1 Introduction

Throughout human history, NPs have served as the foundation of the active ingredients of traditional medicines. In 1804, the German pharmacist Friedrich Sertürner isolated morphine from opium produced by cut seed pods of the poppy, *Papaver somniferum*. This discovery initiated an era wherein the active ingredients from plants could be purified, studied, and administered. Since then, more than 200,000 compounds have been isolated and identified from natural sources (including plants, microbes, and animals), and many have been evaluated as potential drug leads. NPs have been a great source for the discovery and development of therapeutic agents to treat various diseases such as cancer, infections, cardiovascular, and cerebral disorders. Successful examples include salicin (Leroux, 1831), penicillin (Fleming, 1928), reserpine (Müller, 1952), paclitaxel (Wani, 1966), and artemisinin (Tu, 1972). Today, ~40% of therapeutic agents originate from natural sources. These are directly derived by the use of semi-synthetic NP analogs, or indirectly through the use of synthetic compounds based on NP pharmacophores. In addition, the chemopreventive effects of NPs have received much attention, leading to an increased popularity of health-promoting dietary supplements.

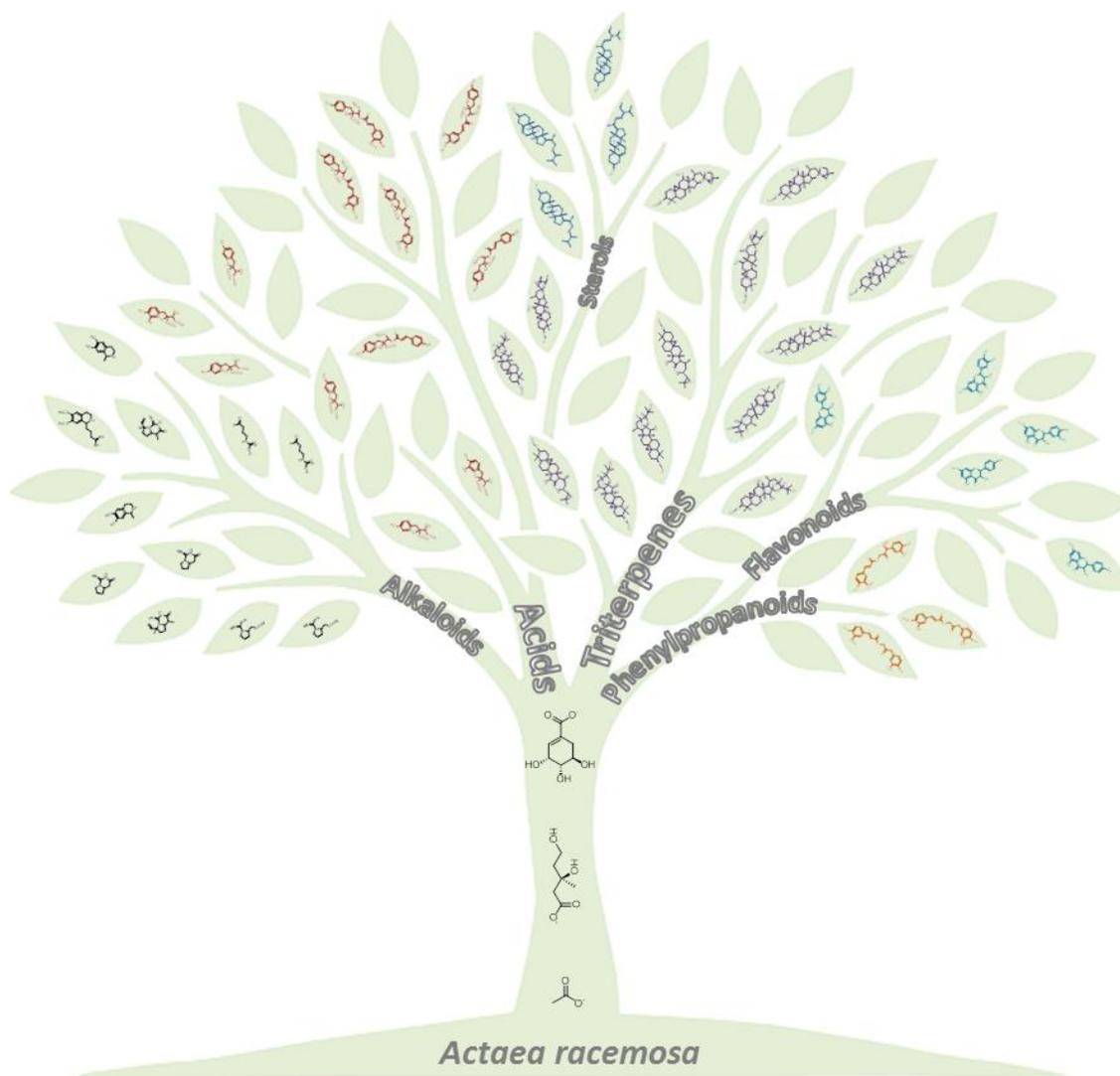
The chemical diversity of NPs have provided many structural scaffolds for drug design and have consistently served as an inspiration for drug discovery and development. While the exploration of NPs is thought to be a fruitful activity for the discovery of novel bioactive compounds, their chemical and biological complexity has limited the efficient separation and identification of these compounds, and has resulted in decreased interest in NP research within the pharmaceutical industry. Conventional research methods for NP discovery are unable to keep pace with the constantly increasing needs of new drugs. Similar challenges also apply to the chemical/biological characterization and standardization of NPs, which is of ultimate necessity for the assurance of their quality, safety, and efficacy when they are used in dietary supplements.

## 1.2 Chemical Complexity of Natural Products

### 1.2.1 Structural Diversity and Similarity

The structural diversity of NPs represents a major characteristic of their chemical complexity. To date, plants are known to produce more than 100,000 NPs. The vast majority of these compounds, commonly referred to as secondary metabolites, do not participate directly in the growth and development of the plant. Although the function of secondary metabolites in the producing organism is still controversial, their chemical diversity has been well recognized. Two hypotheses have been proposed to explain the abundant chemical diversity of plant NPs. According to Feeny *et al.* (1992), secondary metabolites produced by plants alone are involved in physiological responses caused by the interactions with their biotic and abiotic environments. The diversity of compounds produced by plants is the result of the great diversity of plant life and accompanying defense strategies. However, Jones and Firn suggested that organisms that produce and screen many chemicals will likely have an enhanced fitness because the greater chemical diversity leads to more possibilities of producing the rare metabolites with useful and potent biological activities (Jones *et al.*, 1991; Firn *et al.*, 2003, 2006). Surprisingly, the chemical diversity of NPs originates from only a few universal building blocks: acetate, mevalonate, shikimate, methionine, and glucose. These basic elements undergo a variety of biosynthetic transformations and combinations that lead to numerous classes of NPs such as carbohydrates, fatty acids/esters, aromatic polyketides (e.g., phenols and quinones), terpenoids, steroids, phenylpropanoids (e.g., lignans and lignin, coumarins, flavonoids, and isoflavonoids), and alkaloids. It is also found that along the secondary metabolite biosynthetic pathways, plants have the ability to perform *in vivo* combinatorial chemistry that produces a larger number of different but structurally related molecules. These congeners have the same backbones, but differ in side chains, functional groups, or sometimes profoundly in their stereoconfigurations. Thus, not only can all classes of NPs potentially occur within a single plant, but also every class

can consist of numerous but structurally similar compounds, all together creating a highly complex system of NP chemistry in plants (see **Figure 1**).



**Figure 1. Structural Diversity and Similarity of NPs Exemplified for the Complex Chemical Constituents of *Actaea racemosa***

Starting with basic building blocks, diverse secondary metabolites are produced in *A. racemosa*. Based on their structures, each class can be divided into several subclasses which consist of numerous congeners. For example, various types of triterpenes have been identified in *A. racemosa*, in which the cycloartane-type triterpenes are found to have more than 100 members with apparently minor but factually pronounced structural differences.

### 1.2.2 Chemical Variations

Owing to their genetic variability, plants can produce a huge array of NPs, many of which are associated with particular plant genera or species, leading to large chemical variability (Yang *et al.*, 2010; Wolf *et al.*, 2012). Plants in the *Actaea* genus have been widely used as alternative medicines or dietary supplements. A recent study showed a distinction of the phenolic and triterpene constituents of 10 *Actaea* species (He *et al.*, 2006; Ma *et al.*, 2011). It indicated that two chromones, cimifugin and cimifugin glucoside, were contained in *A. dahurica* and *A. foetida*, but not in *A. racemosa*. However, several triterpene glycosides, such as 25-O-acetylcimigenol-3-O-xyloside and cimiracemoside A, were only detected in *A. racemosa*. Furthermore, a preliminary genetic study using a random amplified polymorphic DNA (RAPD) analysis revealed unique DNA profiles of three *Actaea* species: *A. racemosa*, *A. americana*, and *A. rubifolia* (Xu *et al.*, 2002). Supposedly, these data on genetic variations could be further linked to the characteristic chemical profiles associated with different species. In addition to genetic influence, the chemical variation can be caused by environmental factors such as the change in seasons and climate, geographical variations, and soil conditions (e.g., soil pH and composition). As a result of these intrinsic and/or extrinsic factors, the type and level of secondary metabolites can be significantly different, even within a single plant species (Orians *et al.*, 2003). The pattern of secondary metabolites can also exhibit a high variability in different parts of a single plant. For example, ginkgolides are characteristic terpene lactones in *Ginkgo biloba*, of which the structures only differ in the hydroxy substitution at C-1, C-3, and C-7. Among these compounds, ginkgolides A, B, and C, were found in both leaves and root bark. Interestingly, however, ginkgolide J, was found only in leaves, whereas ginkgolide M was isolated only from the root bark (Nakanishi, 2005). Nevertheless, none of these compounds were found in the fruits.

### 1.2.3 Steep Concentration Gradient

The diverse secondary metabolites produced in plants are present at significantly different abundance levels, which also contributes to the complexity of NP chemistry. Taking dried ginkgo leaves as an example, the content of ginkgolic acids can be as high as ~2% (Choi *et al.*, 2004), while the total flavonoids can be ~1% (Deng *et al.*, 2003), and the total terpene lactones are only ~0.2% (Lichtblau *et al.*, 2002). Looking further at the individual compounds, the content of the most abundant terpene lactone, bilobalide, is ~0.1%, whereas the least abundant one, such as ginkgolide J, can be as low as ~0.01% (Lichtblau *et al.*, 2002). These values exemplify the steep concentration gradient of NPs which creates problems when they are chemically and biologically characterized. First, it could be difficult to identify and/or isolate the minor compounds, not only because of their limited quantity, but also due to interferences with major compounds. Second, the commonly used bioassay-guided fractionation (BGF) might lead the scientists to focus on the major compounds but overlook the minor ones which might be significantly bioactive. For instance, prenylphenols are potentially estrogenic compounds in *Humulus lupulus*, among which 8-prenylnaringenin (8-PN) is found to be the most potent phytoestrogen (Milligan *et al.*, 2000; Overk *et al.*, 2005). However, in the crude extract of hop strobiles, the 8-PN content was only about 1/240 of the most abundant but estrogenically inactive xanthohumol (Stevens *et al.*, 1999).

### 1.2.4 Residual Complexity

The previous introduction explains why chemical complexity is an intrinsic characteristic of NPs. Consequently, whenever NPs are purified from natural material, some of this complexity will be retained along the entire fractionation pathway. This relationship is perpetuated in the form of residual complexity, which in principle affects all NPs regardless of how “pure” they are. Residual complexity can be static or dynamic, referring to as impurity patterns (type and level) that are either constant or fluctuating depending on conditions (Chen *et al.*, 2009). As discussed

in Section 1.2.2 (p. 4), minor compounds in a plant extract can exhibit considerable or even most potent bioactivity. Similar considerations apply for the impurities in NP preparations: although they may be present only in minor amounts, their biological functions may have profound impact on the biological evaluation of purified NPs. This problem has been underestimated (see Section 1.3, p. 15).

### **1.2.5 Methods and Challenges**

Owing to their chemical complexity, isolation and structural determination of NPs from their mixtures can be laborious and time-consuming. As a great number of secondary metabolites are produced in plants, the discovery of compounds of interest within the complex mixtures can become similar to that of “finding a needle in a haystack.” Resolution of NP diversity requires a highly selective approach which is dependent of their chemical and physical characteristics. In reality, however, most of this information is unattainable until the single compound is refined from the mixtures. This difficulty is exacerbated when identifying and separating NP congeners, which exhibit highly similar properties. The steep concentration gradient of NPs adds more challenges to the characterization of minor components. Analysis of the residual complexity can be also challenging, not only because the level of impurities might be too low to be effectively detected, but because these minor constituents could be chemically complex as well. The common occurrence of chemical variations even within a single plant species highlights the importance of metabolomic profiling and chemical standardization of NPs, so that the targeted and constant bioactivity can be maintained. A variety of chromatographic and spectroscopic methods have been developed and shown to be effective in resolving the chemical complexity of NPs. Among these techniques, the power of countercurrent separation (CS) and nuclear magnetic resonance (NMR) deserves particular attention.

### 1.2.5.1 CS Resolves Chemical Complexity

Since the first employment of chromatography by the Russian botanist Mikhail Tsvet in 1906, the technology has advanced rapidly and provided various separation tools for resolving the chemical complexity of NPs. Of all modern chromatographic techniques, CS is particularly worthy because it has demonstrated many distinct advantages over other preparative techniques for the separation of NPs but has not received similar attention as other forms of LC in the past decade. First, CS can be applied to virtually any class of NPs because selectivity can be achieved over a full range of polarity through the use of appropriate solvent systems (SSs). Although GC and HPLC exhibit superior resolving power, when they are carried out with large sample loading, resolution is lost due to issues with surface-to-volume ratios and flow dynamics. This is not the case in CS where both phases are liquid. Additionally, since CS does not use a solid support, permanent adsorption onto a column is avoided, and theoretically 100% recovery of the analytes can be achieved. Adding all these characteristics together, CS represents a potentially ideal method for NP isolation, begging the question why it has not been developed further.

So far, CS has contributed largely to key NP discoveries. One of the most famous examples is the isolation of the antitumor agents camptothecin and taxol from the stem bark of *Camptotheca acuminata* and *Taxus brevifolia*, respectively. While present in rather low quantities in the crude extract, these two compounds were successfully purified by sequential steps of Craig countercurrent distribution (CCD). While the isolation was carried out laboriously, neither mass losses nor chemical changes of the eventual products occurred, because of the mild countercurrent distribution methodology (Wall *et al.*, 1996).

Although CS has been proven to be effective and versatile for a diverse range of NPs, older CS equipment such as CCD (Craig *et al.*, 1958) or the later introduced droplet countercurrent chromatography (DCCC) (Tanimura *et al.*, 1970) is limited for wide application due to their disadvantages, such as slow speed (which can be up to several days per run),

significant amount of solvent consumed, and large physical size of the instrument. More recent advancements in CS instruments and methodologies have significantly enhanced its performance and strengthened its power for the separation and analysis of NPs. For example, the invention of high-speed countercurrent chromatography (HSCCC) by Yoichiro Ito *et al.* (1982) dramatically improved the efficiency and resolution of CS, permitting a separation in hours or, more recently, even minutes. In HSCCC, retention of the stationary phase is achieved by using a multi-layer coiled column at a high rotation speed, which enables a higher flow-rate of the mobile phase. By taking advantage of the liquid nature of the stationary phase in CS, the “sweet spot” of CS has been extended to the highly retained analytes by recent developments of elution-extrusion (EECCC) (Berthod *et al.*, 2007) and back-extrusion CCC (BECCC) (Lu *et al.*, 2008). Further new technology relates to the graphic representation of EECCC: Reciprocal symmetry (ReS) and shifted reciprocal symmetry (ReSS) plots and are capable of capturing the high-resolution “sweet spot” of CS in the center of the chromatograms (Friesen *et al.*, 2007). The choice of an appropriate SS is fundamental to the success of a CS and can require a significant time investment, which in practice requires the vast majority of the time devoted to a CS experiment design. Finally, the establishment of the TLC-based GUESS protocol provides a rational and efficient approach for the selection and optimization of CS SSs (Friesen *et al.*, 2005, 2007).

CS has been applied to almost every class of NPs, indicating a high adaptability for the chemical diversity of NPs in terms of polarity, pH, and size of molecules. In particular, the versatile selectivity of CS SSs enables excellent separation of NP congeners. A good example is the preparative isolation of prenylated phenolics from *Humulus lupulus* (hops). Due to their closely related structures, these phytoestrogenic compounds exhibit highly similar chemical properties. Chadwick *et al.* (2005) developed a multidimensional HSCCC approach which led to the isolation of more than 20 prenylated phenolics from spent hops. In particular, the two isomeric flavanones, 6-PN and 8-PN, were shown to be well resolved in a one-step fractionation

using an HEMWat SS (6:4:6:4). Furthermore, this HSCCC approach allowed for a loss-free purification, especially of 8-PN which showed the most potent estrogenic activity, but is present in hops at low concentrations (<10 ppm). A recent HSCCC application by Dahlberg *et al.* (2010) reported a successful purification of tetrahydro-iso- $\alpha$  acid (THIAA) congeners that are found in commercially available modified hop extracts. It was determined that an HEMWat SS (7:3:5:5, pH 5.3) is optimal for separating THIAA *cis* and *trans* diastereomers, whereas the binary SS of hexanes and aqueous buffer (pH 6.8) is optimal for the isolation of individual congeners. Modern CS technology has also been employed for chiral separations of NPs. Imitating the mechanism of a chiral HPLC column, Ma *et al.* (2003) developed an HSCCC method for separating a pair of *N*-(3,5-nitrobenzoyl)-D,L-amino acids using an HEMWat SS with *N*-dodecanoyl-L-proline-3,5-dimethylanilide as a chiral selector in the stationary phase. Surprisingly, the two enantiomers were resolved with less than 5% overlap. Compared with HPLC, CS-based chiral separation is more commercially economical and flexible when choosing optimal chiral conditions.

CS has been shown to be a powerful tool for resolving the chemical complexity of NPs. Similar to other chromatographic techniques, selection of separation conditions, especially appropriate SSs, remains a major challenge for the development of an optimal CS method, which is detailed in the following aspects:

(1) Partition coefficients of analytes are predictors of CS resolution and, therefore, important parameters for optimizing the SS conditions. Traditionally, *K* values are measured by analytical GC or HPLC with authentic standards as identification reference. This approach becomes impractical for compounds for which authentic standards are unavailable.

(2) Inter- and intra-molecular interactions contribute to partition behavior and, thus, can alter partition coefficients. CS is a dynamic process in which the changing molecular interactions affect the partition behavior of analytes. Thus, whenever *K* values are calculated directly from the mixtures in a shake-flask experiment, they may not truly predict the CS behavior of analytes.

In fact, when taking a close look at the CS literature, it can be observed that analytes with high  $K$  values ( $K > 2$ ) eluted much earlier in the actual CS run than what had been predicted from the shake-flask experiment--even when ignoring the additional impact of stationary phase loss which often has not been fully considered due to practical limitations.

(3) Efficient selection and optimization of SSs requires a more rational approach such as correlation of structural characteristics and partition behavior of analytes. Some work has explored the behavior of representative molecules in classical SSs (Koehler *et al.*, 1988; Makovskaya *et al.*, 1995; Ghasemi *et al.*, 2007). However, it is still difficult to establish such relationships for many NPs due to their diverse and complex structures.

(4) While CS chromatographers have developed a variety of SSs, exploitation of selectivity remains a semi-empirical process. In order to best use each SS's selectivity and enhance resolution of multi-step CS, a better understanding of the correlations between the chemical properties of solvents and analytes is required.

#### 1.2.5.2 NMR Resolves Chemical Complexity

NMR spectroscopy plays an important role in NP discovery. In addition to structural elucidation of pure compounds, NMR has been used for both qualitative and quantitative analysis of NP mixtures. Compared to other spectroscopic methods, NMR offers the benefit of providing more detailed structural information that facilitates detection of novel chemotypes, dereplication of known compounds, and characterization of metabolomic profiles. In addition, under quantitative conditions, the  $^1\text{H}$  NMR signals are proportional to molar concentration, enabling a direct comparison of concentrations of all compounds, without the need for calibration curves of individual compounds. However, NMR analysis of NP mixtures remains a major challenge because the spectra are often complicated by severe peak overlap that can significantly hinder the identification and accurate quantification of constituents.

With the development of high-resolution and high-sensitivity NMR instrumentation, together with chemometric methodologies, NMR-based metabolomic analysis has increased resolution/resolving power for chemically complex NPs. To this end, a variety of 2D NMR techniques has been exploited in the analysis of NP mixtures. For example, Xi *et al.* (2006) developed an automatic screening method for amino acids in complex biological samples. Given a database of 2D COSY spectra for the metabolites of interest, this method provides a list sorted by the heuristic likelihood of each metabolite being present in a sample. Owing to the chemical complexity of NP mixtures, their 2D NMR spectra might appear to be highly complex for full and complete interpretation. To address this challenge, Schroeder *et al.* (2007) developed a simple procedure for the differential analysis of arrays of DQF-COSY spectra. It was effectively applied for the detection and characterization of new NPs from a small library of fungal extracts. Recently, HSQC has been shown to be particularly useful in mixture analysis. The presence of certain metabolites within an extract can be clearly distinguished by comparison of the HSQC spectra of mixtures of known reference compounds and those of the extracts. Lewis *et al.* (2007) reported a 2D  $^1\text{H}$ - $^{13}\text{C}$  NMR protocol for the identification and quantification of the most abundant metabolites in plant extracts. Compared to traditional 1D  $^1\text{H}$  NMR analysis, this method was more efficient and just as accurate in the determination of molar concentrations.

One-dimensional/two-dimensional NMR was also used to compare and classify different plant samples based on the chemometric analysis of their NMR spectra. Upon acquisition, a process known as “bucketing” is applied to digitalize the spectra to generate numeric values for further statistical analysis. Owing to the size of the data sets, an appropriate chemometric method is required to reduce the number of variables while maintaining the distinguishing characteristics. Principal component analysis (PCA) is commonly used for such purpose. PCA is an unsupervised pattern recognition method in which all samples are grouped with the maximum separation of all samples based on the discrimination of signals in the spectra. As a

result, the spectroscopic complexity can be simplified into two or three most discriminating components which allow differentiation of the samples. NMR-based metabolomic profiling methodology has been used in chemo-taxonomic analysis of plant species as well as quality standardization of dietary supplements (Kim *et al.*, 2010).

The power of NMR has also been employed to resolve the residual complexity of NPs. Chen *et al.* (2009) examined the dynamic residual complexity of desmethylxanthohumol (DMX) by means of quantitative  $^1\text{H}$  NMR (qHNMR) in a setting that mimics *in vitro* and physiological conditions. The results suggested that measureable estrogenic activity of even high-purity DMX is principally attributed to its degradation products, such as the potent phytoestrogen, 8-PN. Similarly, Schinkovitz *et al.* (2008) used 1D/2D NMR for a stability study of Z-ligustilide, a major bioactive constituent of medicinal plants of the *Apiaceae* family. Identification of its six key degradation products in the residually complex sample was enabled by 2D NMR experiments such as COSY and 1D SelTOCSY. Quantitative proton NMR analysis of these residuals led to the recognition of variations in time- and process-dependent sample purity.

Recently, the screening of NP sources has suffered from low efficiency due to a high probability of duplicate findings. Dereplication, the procedure of rapid identification of known compounds, is, therefore, important to avoid rediscovery of previously characterized compounds. A variety of NMR-based tools has been developed for this purpose. For example, Lambert *et al.* (2005) used a hyphenated technique, HPLC-SPE-NMR, for dereplication of isoflavonoids in the *Smirnowia iranica* extract prior to preparative-scale isolation. The structures of 10 new isoflavonoids and of seven known constituents were elucidated from online NMR analysis upon HPLC separation. This information could be used to direct preparative isolation work. In addition, NMR databases have been used to facilitate NP dereplication. A good example is CSEARCH, a database containing  $^{13}\text{C}$  NMR reference spectra of over 500,000 compounds (Chen *et al.*, 1993). It allows for rapid identification of known compounds and provides insights into the structural elucidation of unknown compounds. Furthermore, AntiMarin

is a  $^1\text{H}$  NMR databases dedicated for dereplication of NPs derived from marine and microorganism sources (Lang *et al.*, 2008). Unlike in a full spectral search, only the number of methine, methylene, and methyl groups is required to find the match compound(s) in the database. Despite these advantages, some limitations are associated with these methods:

(1) For LC-hyphenated techniques, such as LC-UV, LC-MS and LC-NMR, performance of one dimension of highly resolved separation is the prerequisite for structural analysis. Thus, these methods depend on optimization of LC conditions for sufficient resolution of the complex mixtures, which may be a time-consuming and demanding process.

(2) The NMR database search usually requires the query data which are difficult to obtain from the NMR spectra of complex mixtures, and otherwise still needs isolation of pure compounds. The data availability also relates to the successful dereplication. Consequently, these methods may not satisfy the purpose of rapid dereplication of constituents in mixtures.

While all the above NMR applications have been shown to be powerful tools, the relatively low sensitivity of NMR creates a major problem when dealing with the mass-limited samples or minor constituents in mixtures. Recent development involving highly sensitive probe technology, such as cryo-microprobes, has dramatically improved detection sensitivity and, thus, reduced NMR experimental data collection time. In these probes, the electronic components, one of the main electrical sources of noise in NMR spectroscopy measurements, are cooled down to 20 K using liquid helium. This reduces the noise and increases sensitivity by up to a factor of 4. As a result, it is possible to achieve up to 16-fold increase in the signal-to-noise ratio per scan. Furthermore, by using reduced detection volume of NMR probes, such as a 1.7 mm cryoprobe, the increase in mass-sensitivity compared to the 5 mm cryoprobe can be estimated as the product of  $1/9 \times 750/30$ , or a factor of 2.7. As a result of these advancements, current state-of-the-art NMR allows the analysis of NPs at the nanomole-scale (Dalisay *et al.*, 2009). Despite their high sensitivity, it is necessary to be aware of some limitations associated with these techniques. Ideally, a chosen NMR solvent is expected to yield high solubility to

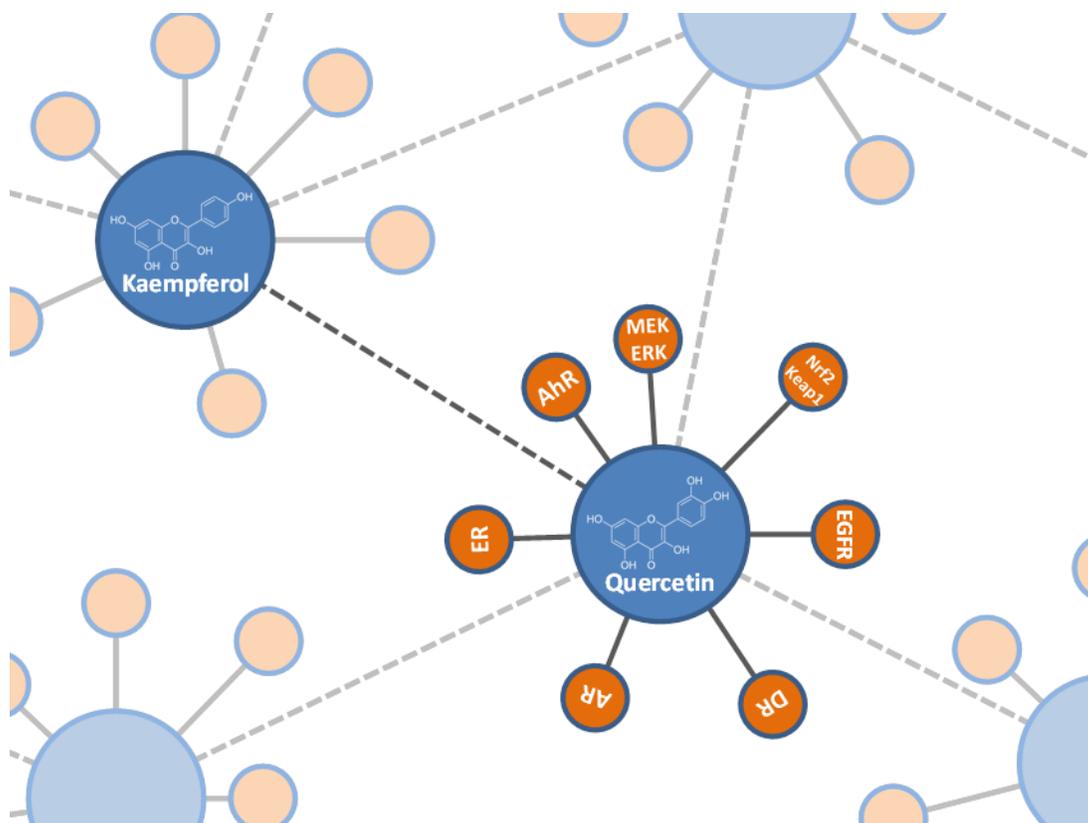
maximize the amount of sample in solution. In practice however, even when the sample is available at the gram level, it may only be analyzed at  $\sim 5 \text{ ng}/40 \mu\text{L}$ . Thus, microprobes may not yield significant advantage for the sample with low solubility because using a smaller sample volume cannot help to concentrate the sample. On the contrary, for the sample with excellent solubility, the significantly high concentration in the reduced volume may negatively impact the resolution and reproducibility of spectra. Therefore, it requires a rational selection of instruments and experimental conditions according to the sample properties for optimal quality of NMR spectra.

## 1.3 Biological Complexity of Natural Products

### 1.3.1 Synergistic and Multitarget Effects

For a long time, the search for bioactive compounds in plants has been directed at the determination of single or a few chemical entities. However, traditional systems of medicines such as Traditional Chinese Medicine (TCM) generally assume that the therapeutic efficacy of plants results from multiple constituents which act in synergy (Gertsch *et al.*, 2011). It is also common in traditional medicines that a combination of different herbs is used as a single remedy in which the synergistic effects of the complex constituents of multiple plant sources are used to achieve an optimal therapeutic efficacy. For this reason, there has been debate on whether NP discovery should focus on the “bioactive principles” rather than single compounds. While these theories are largely based on empirical evidence through a long history of clinical application, scientists are still in the process of finding convincing proof and mechanisms behind these biological complexities of NPs.

In modern pharmacology, drug synergy is defined as drug interaction in ways that enhance or magnify one or more effects, or side effects. For example, the analgesic efficacy of codeine can be enhanced when mixed with acetaminophen or ibuprofen. This phenomenon has also been observed in purified NPs. Stermitz *et al.* (2000) showed that the antimicrobial potential of *Berberis fremontii* is not only caused by antimicrobial agents such as berberine, but also by multidrug-resistance (MDR) inhibitors such as 5'-methoxyhydrocarpin. This plant constituent has no antimicrobial activity alone, but strongly potentiates the action of berberine and other NorA substrates. A mechanism model was suggested, according to which the MDR inhibitors block the NorA pump of the cell membrane in bacteria, preventing the extrusion of accumulated berberine in cells and thus potentiating its antibiotic action.



**Figure 2. Synergistic and Multitarget Effects of NPs**

Taking quercetin as an example, its anticancer activity can be explained by its interactions with multiple targets (solid lines), such as aryl hydrocarbon receptor (AhR), estrogen receptor (ER), androgen receptor (AR), death receptor (DR), and epidermal growth factor receptor (EGFR). Its activity can be enhanced by interaction with the congener kaempferol (dashed line). Extrapolating these findings, the bioactivity of the crude extracts could result from a rather complex mechanism involving synergistic and multitarget effects of multiple components.

The biological complexity of NPs is also attributed to their multitarget effects--a single compound might be able to interact with multiple target molecules. This effect possibly results from the fact that the biosynthesis of the complex structure of NPs involves a variety of enzymes. These enzymes have distinct architectures and molecule-binding cavities, with which the product molecule under synthesis must interact. Therefore, the core structure of each product molecule inherits diverse binding groups as well as a certain level of flexibility which allow interaction with a variety of targets (Ji *et al.*, 2009). As a result of their multitarget effects,

NPs can exhibit a broad spectrum of bioactivities. For example, quercetin is an anti-oxidative flavonoid widely distributed in plants. This compound has been shown to interact with some receptors which are involved in the development of cancers induced by certain chemicals (**Figure 2**). It has also been shown to modulate several signal transduction pathways involving MEK/ERK and Nrf2/keap1, which are associated with the processes of inflammation and carcinogenesis (Murakami *et al.*, 2008). Another study also showed the synergistic effect of quercetin and kaempferol in the reduction of cell proliferation. Their combination was more effective than the additive effect of each compound (Ackland *et al.*, 2005). These two flavonoids, together with other congeners such as galangin, isorhamnetin and catechins, often coexist in plants, suggesting that their anticancer activity in the crude extract results from a complex mechanism involving both synergistic and multitarget effects (**Figure 2**).

### 1.3.2 Methods and Challenges

The “isobole method” has been widely adopted for the evaluation of synergy in NPs (Berenbaum *et al.*, 1989). In this method, different dose combinations of two compounds are investigated for the same type and magnitude of biological effect. The additive interaction means that the effect of two compounds is a pure summation effect. With antagonistic interaction, the overall effect is less than what is expected from the sum of the separate effects. Conversely, synergistic interaction results in an overall effect that is larger than the sum of the separate effects. As shown in the foregoing examples, the current study mainly focuses on the interactions of purified NPs. However, crude plant extracts are diverse and complex systems, both chemically and biologically. Evaluation of the synergistic effects of multiple components in the crude extract is challenging. Inui *et al.* (2007) initiated a CS-based analysis of synergy in the crude extract of an antituberculosis (anti-TB) ethnobotanical, *Oplopanax horridus*. Instead of using pure isolates, the CS fractions were studied using the isobole method to confirm the presence of synergistically active compounds. The results provided guidance for further

separation of bioactive constituents in a bioassay-guided isolation procedure. For synergy research, it is vital to exclude the possibility of activity loss in the course of separation. Currently, CS might be an excellent technique for this purpose, while being capable of providing high throughput and high-resolution separations at the same time. It can be expected that the implementation of parallel spectroscopic analysis in a CS-based synergy study can provide further insights into the structural characterization of the active constituents, and that such an approach might enable correlation with biological effects and generation of structure–activity relationships (SARs).

While residual complexity is likely to be ubiquitous in purified NPs, it is also likely to be overlooked in biological evaluation. Even minor constituents (impurities) can make significant contributions to the overall activity of the sample. For example, a recent study revealed an inverse correlation between purity and anti-TB activity of various reference samples of ursolic acid. This leads to the conclusion that the antimycobacterial activity is not caused by the single pure compound ursolic acid, but can possibly be related to its synergistic effects with various impurities (Jaki *et al.*, 2008). Therefore, the characterization of the residual complexity of NPs is an important endeavor, especially when they are used in biological assessment. The establishment of purity–activity relationships (PARs) is a potentially important tool for the evaluation of the biological impact of molecular interactions when analyzing residually complex samples.

## 1.4 Development of Hypotheses

The diverse structures of NPs contribute to a broad spectrum of bioactivities as well as a great number of molecular scaffolds for drug discovery and development. However, the chemical and biological complexity of NPs also creates great difficulties and challenges in researching these molecules. Traditionally, the isolation of bioactive compounds from natural sources follows the methodology of more or less exhaustive BGF procedure. By repeatedly using a variety of chromatographic techniques, the isolation and characterization of a pure and bioactive compound is the anticipated end result. However, this classical approach is confronted by unprecedented challenges that modern NP research faces as a result of various issues that have arisen when scientists revisit their previous work.

In the first place, one might ask if the results obtained have been proportional to the tremendous efforts expended. The fractionation/purification processes not only require intensive and long-time manpower, but also consume large quantities of expensive and potentially toxic solvents and sorbents. These problems are attributed to the chemical complexity and low concentration of NPs. In order to establish a more efficient and economical methodology for NP separation, **this study hypothesizes that the conventional fractionation process can be accelerated and targeted by appropriate sample preparation methods, and more importantly by a rational selection and optimization of separation conditions.**

### **[Hypothesis A]**

Although a handful of pure compounds may be obtained through BGF, results of subsequent spectroscopic/spectrometric analyses (e.g., UV-Vis, IR, MS, NMR, etc.) may turn out to be a disappointment in that a few, or worse, all of the isolated compounds were either uninteresting or had been previously characterized. This unexpected outcome has encouraged scientists to repeat the same process until more interesting molecules had been isolated. Structural dereplication is, therefore, important to improve the efficiency of NP isolation. **This study is expanding the utility of NMR for the “in-process” characterization of NPs,**

**especially their complex mixtures, to facilitate the structural elucidation and dereplication, both dynamically for fractionation procedures and statically for purified but residually complex samples. Additionally, it is hypothesized that the combination of NMR spectroscopy with chromatography could assist in the design of more efficient fractionation schemes as well as the metabolomic profiling of NPs. [Hypothesis B]**

The major challenge in NMR analysis of NP mixtures is the severe peak overlap which hinders the interpretation of the NMR data. In particular, the structural similarity of NP congeners creates a dereplication problem due to their similar spectroscopic patterns. As a result, it could be rather difficult for visual determination of their minor differences in the complex spectra. Thus, it becomes difficult to employ conventional methodology, such as NMR databases, for dereplication. However, **this study hypothesizes that the chemical shifts of certain readily discernible protons/carbons and structural characteristics of NPs can be statistically correlated, and this correlation can be successfully integrated into a pattern recognition model for automated dereplication. [Hypothesis C]**

While scientists are frequently excited about isolating novel chemical structures or unprecedented structural backbones, they are often disappointed by the subsequent bioassay results indicating that these chemically interesting molecules are not biologically active. Moreover, it can be even more frustrating that the pure compounds isolated through BGF do not exhibit any bioactivity. Upon the successful characterization of synergistic effects of some NPs, the question can be raised that relates to conventional methodology, i.e., whether isolation is a prerequisite for biological evaluation. Although identification of bioactive principles is given much preference over the identification of single pure compounds, it is a more important challenge for the chemical and biological characterization of multiple components in a complex mixture. **This study employs a methodology that statistically integrates the chemical and biological evaluation of NPs, by which the purity–activity relationship can be established,**

**and by which the active principles can be determined in the mixtures without physical separation. [Hypothesis D]**

As discussed in the previous sections, a variety of tools are available today to enhance and/or accelerate drug discovery from natural sources. In particular, CS and NMR spectroscopy are powerful tools to unravel the chemical complexity of NPs. To prove the aforementioned hypotheses, the present study makes extensive use of CS and NMR techniques and especially their combinations in separation, structural elucidation, and biological evaluation of NPs (**Table 1**). Consequently, the resulting methodologies have potential to widen the applicability of the single conventional technique in NP research, and more importantly can enhance their adaptability to the resolution of NP complexity, and, thus, can contribute to a better understanding of their chemical and biological functions.

**TABLE I. THE DESIGN OF THE PRESENT STUDY**

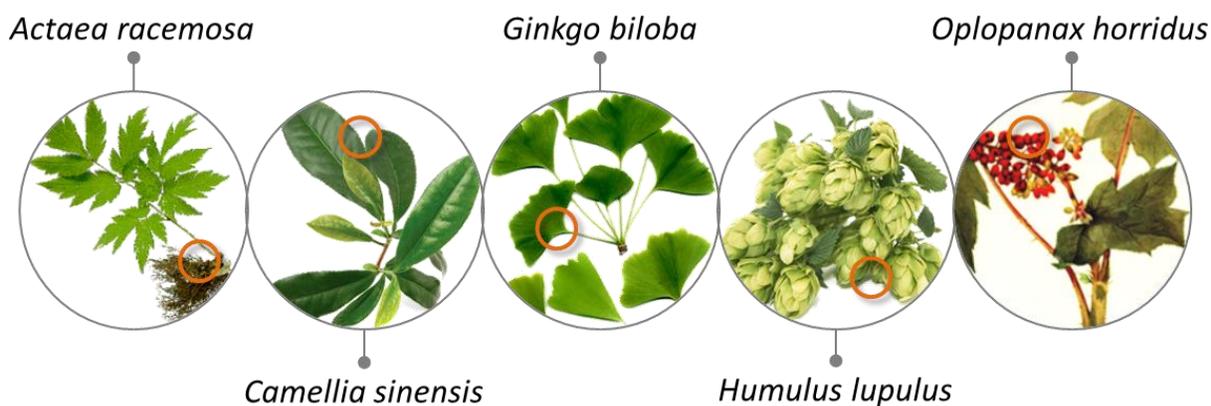
CHAPTER: Topic	Aspect of NP research	Aspect of NP complexity	Hypothesis	SECTION: Methodology
3 Accelerating and Targeting of NP Separation	Separation/ Isolation	<ul style="list-style-type: none"> <li>- Structural Diversity &amp; Similarity</li> <li>- Steep Conc. Gradient</li> <li>- Residual Complexity</li> </ul>	A The conventional fractionation process can be accelerated and targeted by appropriate sample preparation methods, and a rational design of optimal separation procedures.	3.2 Sample Cutting 3.3 Prediction of Elution Profiles 3.4 Chromatographic Orthogonality 3.5 Solvent Orthogonality
4 Expanding NMR Utility in NP Characterization	Structural Elucidation	<ul style="list-style-type: none"> <li>- Structural Diversity &amp; Similarity</li> <li>- Steep Conc. Gradient</li> <li>- Residual Complexity</li> </ul>	B NMR can be employed in analysis of NP mixtures to facilitate the structural elucidation and dereplication. It can also assist the design of efficient fractionation schemes as well as metabolomic profiling of NPs.	4.2 Elution Visualization (by <sup>1</sup> H) 4.3 Residual Complexity Evaluation (by COSY) 4.4 Differential Analysis (by HSQC) 4.5 Pattern Recognition (by HMBC)
5 Rational Naming and Virtual Partitioning of NPs	Structural Dereplication	<ul style="list-style-type: none"> <li>- Structural Similarity</li> <li>- Residual Complexity</li> </ul>	C The chemical shifts of certain protons/ carbons and structural characteristics of NPs can be statistically correlated, and this correlation can be successfully integrated into a pattern recognition model for automated dereplication.	5.2 Rational Naming System 5.5 Classification Binary Trees of <sup>1</sup> H NMR of Methyl Groups 5.6 Automated Dereplication
6 Integrating Chemical and Biological Evaluation of NPs	Chemical and Biological Evaluation	<ul style="list-style-type: none"> <li>- Synergism</li> <li>- Antagonism</li> <li>- Residual Complexity</li> </ul>	D The chemical and biological evaluation of NPs can be statistically integrated, by which the purity–activity relationship can be established, and active principles can be determined in the mixtures without physical separation.	6.1 Quantitative Purity–Activity Relationship 6.2 Biochemometrics

## 2. EXPERIMENTAL

### 2.1 Materials

#### 2.1.1 Plant Materials and Extracts

In order to prove that new concepts and theories are effective in practical applications, four popular plants used as herbal remedies or dietary supplements in the United States and worldwide, including *Actaea racemosa* (Nutt.) L. (syn. *Cimicifuga racemosa*, black cohosh), *Camellia sinensis* (L.) Kuntze (green tea), *Ginkgo biloba* L. (ginkgo), and *Humulus lupulus* L. (hops), as well as an indigenous ethnobotanical *Oplopanax horridus* (Sm.) Miq. (Devil's club), were selected as test cases in the present study (**Figure 3**).



**Figure 3. Five Plants Selected as Test Cases in the Present Study**

The circles in orange indicate the parts of plants used in the present study. *A. racemosa*: Roots and rhizomes; *C. sinensis*: Leaves; *G. biloba*: Leaves; *H. lupulus*: Strobiles; *O. horridus*: Fruits.

**Raw materials.** Authentic *A. racemosa* roots and rhizomes were obtained through Naturex (formerly Pure World Botanicals, South Hackensack, NJ), and voucher specimens are

deposited in the College of Pharmacy, UIC (BC# 066). Dried fruits of *O. horridus* were collected from wild specimens of the plant in the vicinity of Anchorage, Alaska and authenticated by David C. Smith at Alaska Green Gold (Anchorage, AK) in 2007. Voucher specimens are deposited in the College of Pharmacy, UIC (BC# 390).

**Commercial extracts.** The supercritical fluid extract of *H. lupulus* strobiles (Lot# 1994) was acquired from Northern Brewer (Hallertau, Germany). The crude extracts of *G. biloba* leaves (Lot# G96-63-A9) and *C. sinensis* leaves (Lot# G57-19-A9) were acquired from Naturex.

### 2.1.2 Solvents

All organic solvents used throughout the extraction and fractionation process were analytical grade (Pharmco-AAPER, Brookfield, CT) and redistilled on a rotary evaporator prior to use. All organic solvents used for analytical instrumentation were HPLC or GC grade (Sigma-Aldrich, St. Louis, MO and Fischer Scientific, Hampton, NH). The deuterated solvents used for NMR analysis were obtained from Sigma-Aldrich, St. Louis (MO) and Cambridge Isotope Laboratories, Andover (MA). Water was prepared by deionization to 18.2 MΩ/cm at 25 °C on a Milli-Q Synthesis A10 Water Purification System (Millipore, Bedford, MA). All waste solvents were properly disposed of in accordance of regulations accepted by the Occupational Safety and Health Administration (OSHA) and the U.S. Environmental Protection Agency (EPA).

## 2.2 Methods

### 2.2.1 Extraction

Percolation was employed for the preparation of the crude plant extracts. Dried plant materials were initially pulverized into a fine powder, moistened with an appropriate amount of the specified solvent and allowed to stand for approximately 4 h, after which the moist mass was packed into a separatory funnel as a percolator. Additional solvent was added to form a shallow layer above the mass, and the mixture was allowed to steep for 24 h. The liquid was allowed to drip slowly into an Erlenmeyer flask. Additional solvent was added to produce the required volume. Finally, the percolate was concentrated to a syrupy residue on a rotary evaporator. The details for the preparation of crude extracts of *O. horridus* and *A. racemosa* are described as follows:

***O. horridus***. The dried fruits (7 kg) of *O. horridus* were pulverized and percolated sequentially with PE and DCM at room temperature. The crude organic extracts were concentrated *in vacuo* (<40 °C) to yield 2515 and 194 g of PE and DCM syrup residues, respectively. A combination of 200 and 15 g of PE and DCM residues was exhaustively fractionated on VLC for further separation.

***A. racemosa***. The dried roots/rhizomes (1 kg) of *A. racemosa* were pulverized, homogenized, and percolated with 11 L of fresh MeOH at room temperature. The crude organic extract was concentrated *in vacuo* to yield 164 g of a syrupy residue. The residue was reconstituted in deionized water (250 mL) and partitioned with EtOAc (20 × 250 mL, 52 g) and *n*-BuOH (19 × 250 mL, 32 g). The EtOAc partition was subjected to column chromatography for the isolation of cycloartane triterpenoids.

Liquid-liquid extraction (partitioning) was used both for the preparation of enriched samples and in the measurement of partition coefficients. A sample was dissolved in a fresh biphasic solvent system using equal volumes of upper and lower phase in a separatory funnel. The funnel was closed and shaken vigorously until the sample was well dissolved and

distributed at equilibrium between the two phases, during which process the stopcock was opened periodically to vent the vapor pressure buildup. Centrifugation was applied if needed to break the emulsion and/or accelerate phase separation. After the mixture completely settled, the two phases were separated into two flasks and evaporated to dryness *in vacuo*.

## 2.2.2 Chromatography

### 2.2.2.1 Thin-Layer Chromatography (TLC)

Analytical TLC was used for the selection of solvent conditions to be used for preparative chromatography, e.g., VLC, MPLC, and CCC, as well as for composition analysis of the resulting fractions. NP-TLC was performed on pre-coated Alugram SIL G/UV 0.20 mm thick silica gel 60 aluminum plates with fluorescent indicator UV<sub>254</sub> (10 × 20 cm; Macherey-Nagel, Düren, Germany). RP-TLC was performed on pre-coated Alugram RP-18W/UV 0.15 mm thick silica gel C<sub>18</sub> plates with fluorescent indicator UV<sub>254</sub>. The sample solutions were carefully spotted at ~0.5 cm from the bottom edge of the plates using Wiretrol 10  $\mu$ L disposable pipets (Drummond Scientific, Broomall, PA). Plates were developed in closed TLC chambers using the following SSs:

*n*-Hexane–EtOAc–MeOH (NP)

*n*-Hexane–DCM–MeOH (NP)

CHCl<sub>3</sub>–MeOH (NP)

MeOH–H<sub>2</sub>O (RP)

For NP-TLC, EtOAc and CHCl<sub>3</sub>/CH<sub>2</sub>Cl<sub>2</sub>-based SSs were selected based on their different selectivity for the specific compounds. MeOH and *n*-hexane were used for adjustment of the overall polarity. The ratio of the solvent compositions was optimized as appropriate for the sample polarity.

The TLC plates were removed from the chamber after they developed to within ~1.0 cm from the top edge. After the solvent was allowed to evaporate, the plates were examined under

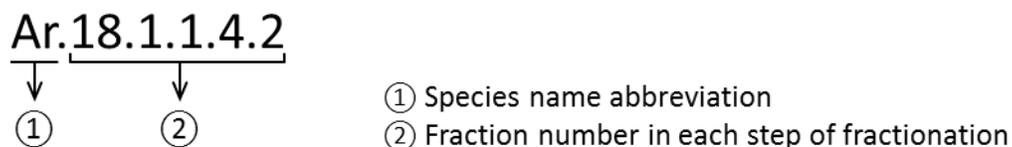
UV light at 254 and 365 nm, and pictures were taken for permanent records using a digital camera. The plates were sprayed evenly with a fresh mixture of general purpose vanillin staining reagent (6 g vanillin, 2.5 mL conc. H<sub>2</sub>SO<sub>4</sub>, 250 mL EtOH) and heated with a heat gun until vivid colorful spots were visible. All stained TLC plates were immediately scanned at 300 dpi using an office scanner, and the images were appropriately cropped and saved in TIFF format. Finally, the original images were processed within Adobe Photoshop by applying Autolevels to increase the contrast of the colored spots over the background, which improved visualization of minor compounds.

#### 2.2.2.2 Vacuum Liquid Chromatography (VLC)

As a quick and convenient method, VLC was employed for the initial large-scale fractionation of crude extracts as well as purification of enriched samples. Compared to other open column methods, VLC reduces the retention time of constituents on the column, and, thus, minimizes the irreversible loss of isolates. The operational procedures comprise the following steps: An appropriate amount (usually 10–20 times sample weight) of NP silica gel 60 (40–63 μm; Merck, Darmstadt, Germany) was initially poured into a VLC column, and a reduced pressure was applied using a diaphragm vacuum pump (Vacuubrand, Wertheim, Germany) until the silica gel was firmly and evenly packed. A disk of filter paper was placed above the silica gel. The sample was dissolved in solvent and mixed with silica gel at a ratio of 1:1 or 1:2 in a mortar. Once the solvent had evaporated, the sample was evenly applied to the column and covered with another disk of filter paper. When the mass of the sample was lower (e.g., <50 mg), the sample was dissolved in a small volume of solvent and directly applied to the top of silica gel in the column using a glass pipette. Sand was added and the column was topped with some cotton to reduce streaking. Afterwards, solvent in aliquots (2.5–3 times of column dead volume) was poured into the column in an isocratic or step-wise gradient fashion, and fractions were collected in round bottom flasks under reduced pressure. Finally, the fractions collected were concentrated on a

rotary evaporator and transferred to small weighed vials. The major SSs used in NP-VLC were *n*-hexane–EtOAc–MeOH or *n*-hexane–DCM–MeOH. Their ratios were optimized using NP-TLC in which the target compounds were best separated with  $R_f$  values of 0.2–0.3.

The fractions (combined if necessary) produced in any types of preparative chromatography were coded as the following format:



The above example indicates the 2<sup>nd</sup> fraction produced in the 5<sup>th</sup> step of fractionation of *Actaea racemosa*.

### 2.2.2.3 Medium-Pressure Liquid Chromatography (MPLC)

The advantages of using MPLC in the separation of NPs include higher loading capacities and higher throughput when compared to other preparative chromatographic methods. Four types of MPLC columns were used in the current study: (1) A glass MPLC column (30 × 450 mm) packed with Polygoprep RP C<sub>18</sub> silica gel (25–40 μm; Macherey-Nagel, Düren, Germany); (2) An MPLC column (20 × 300 mm) packed with Polygoprep RP C<sub>18</sub> silica gel; (3) A Lobar column packed with LiChroprep RP C18 silica gel (10 μm, 11 × 200 mm; Merck); and (4) An MPLC column (13 × 1000 mm) packed with NP silica gel 60. All columns were used with a six-port valve connected to a sample loop, a Waters Delta 501 HPLC pump (Milford, MA) and an LKB Bromma 2111 Multitrac (Bromma, Sweden) or Foxy Jr. (Teledyne Isco, Lincoln, NE) fraction collectors. All RP-MPLC runs were conducted using an isocratic or step-wise gradient elution using SSs MeOH–H<sub>2</sub>O or MeCN–H<sub>2</sub>O. For NP-MPLC, isocratic elution used a SS consisting of *n*-hexane, EtOAc, and water. TLC was used for the optimization of the solvent conditions, which were similar to those used in VLC. The columns were equilibrated with the mobile phase for 30

min at the beginning. The sample solution was loaded into the loop and injected onto the column. Based on the system being used, flow rates were set to 0.5–5 mL/min, and fractions were collected at 1–4 min intervals.

#### 2.2.2.4 High-Speed Countercurrent Chromatography (HSCCC)

Owing to its high capacity and efficiency, HSCCC has become a suitable preparative separation method for NPs. In this study, two HSCCC instruments were used: (1) A CCC-1000 high-speed countercurrent chromatograph (Pharma-Tech Research, Baltimore, MD), equipped with a 120 mL column consisting of three serially connected coils of 1.6 mm i.d. PTFE tubing; and (2) A TBE-20A high-speed countercurrent chromatograph (Tauto Biotech, Shanghai, China), equipped with a 20 mL coil of 0.8 mm i.d. PTFE tubing. Both HSCCC systems were equipped with a six-port valve with a loop (3 and 0.5 mL, respectively) for sample injection and a PTR HPLC pump (Pharma-Tech Research, Baltimore, MD) for solvent delivery. Fraction collection was accomplished with a Foxy Jr. HPLC fraction collector.

The CS SSs used in the current study mainly included various members of ChMWat and HEMWat families (Friesen *et al.*, 2005). As shown in **Table 2**, the ChMWat family was standardized to eight SSs with different compositions, which were coded by a series of integer values from -3 to +4 indicating increasing polarity. Similarly, the HEMWat family consists of 16 standardized SSs with a polarity range from -7 to +8. Selection of these SSs was based on the GUESS method developed by Friesen *et al.* (2007). Preparation was performed by thoroughly mixing the individual components in a separatory funnel. After the solvent mixture settled into two phases, the upper and lower phases were separated and stored in Pyrex bottles.

**TABLE II. STANDARDIZED CS SS IN THE ChMWat AND HEMWat FAMILIES<sup>a</sup>**

ChMWat SSs (v/v)				HEMWat SSs (v/v)				
Code	<u>CHCl<sub>3</sub></u>	<u>MeOH</u>	<u>Water</u>	Code	<u>hexane</u>	<u>EtOAc</u>	<u>MeOH</u>	<u>Water</u>
-3	10	0	10	-7	9	1	9	1
-2	10	1	9	-6	8	2	8	2
-1	10	2	8	-5	7	3	7	3
0	10	3	7	-4	7	3	6	4
+1	10	4	6	-3	6	4	6	4
+2	10	5	5	-2	7	3	5	5
+3	10	6	4	-1	6	4	5	5
+4	10	7	3	0	5	5	5	5
				+1	4	6	5	5
				+2	3	7	5	5
				+3	4	6	4	6
				+4	3	7	4	6
				+5	3	7	3	7
				+6	2	8	2	8
				+7	1	9	1	9
				+8	0	10	0	10

<sup>a</sup> Friesen *et al.*, 2005

HSCCC fractionation was performed in elution–extrusion mode (EECCC) using the following procedures. When using the 120 mL HSCCC, the stationary phase (upper phase of ChMWat or lower phase of HEMWat) was initially pumped into the coil at a flow rate of 5 mL/min. After the coil was completely filled with stationary phase, the mobile phase was pumped into the coil at a flow rate of 1 mL/min in tail-to-head (ChMWat) or head-to-tail (HEMWat) mode, with the centrifuge spinning at 800–1000 rpm. The stationary phase displaced by the mobile phase was collected in a graduated cylinder. The sample was prepared for injection by dissolving it in 1.5 mL each of upper and lower phase of the SS. Injection of the sample using the 3 mL sample loop occurred after the hydrodynamic equilibrium was

established in the column, i.e., only mobile phase eluted from the column. The volume of stationary phase extruded during the equilibration process was measured, and the stationary phase retention factor ( $S_f$ ) calculated. Fraction collection was started at the time of injection and performed in 3 or 4 min intervals. After elution of the appropriate volume of mobile phase (usually  $K = 2.5\text{--}3.0$ ), the eluent was switched to the stationary phase until all of the mobile phase plus one column volume of stationary phase was extruded. The same operational procedure was applied to the 20 mL HSCCC, except that the rotation speed was increased to 1600 rpm, the flow rate was 0.50 mL/min, and the fractions were collected in 2 min intervals. In preparation for further NMR analysis, all fractions were evaporated to dryness in a SpeedVac and stored in a desiccator *in vacuo* for 24 h to eliminate variations from residual water. The fractions were combined as appropriate according to their compositions as determined by TLC and/or NMR.

#### 2.2.2.5 High-Performance Liquid Chromatography (HPLC)

Semi-preparative HPLC was carried out using a YMC ODS-AQ column (120 Å, 5  $\mu\text{m}$ , 20  $\times$  250 mm), a Waters Delta 600 solvent pump, an online solvent degasser (helium sparging), a Waters 996 photodiode array (PDA) detector, a Waters 717 plus autosampler, and an HPLC fraction collector. The system was controlled by a Windows-based desktop computer running Waters Empower 2 chromatography software. The separation was conducted by isocratic elution using 90% aq. MeOH at a flow rate of 6 mL/min. The sample solution was injected after the column was equilibrated for 10 min. Fraction collection was started immediately after sample injection and performed in 1 min intervals.

#### 2.2.2.6 Gas Chromatography (GC)

GC was performed on an Agilent 7890A GC system with an Agilent 7000A GC/MS Triple Quadrupole mass spectrometer and fitted with an HP-5ms capillary column (30 m  $\times$  0.25 mm  $\times$

0.25  $\mu\text{m}$ ). Helium was used as carrier gas at a flow rate of 1 mL/min. The sample was dissolved in an appropriate amount of *n*-hexane. Split injection (ratio 1:10) of 1  $\mu\text{L}$  after 5 min of solvent acquisition delay was applied for all samples. The oven temperature was increased from 40 to 310  $^{\circ}\text{C}$  with a linear gradient of 15  $^{\circ}\text{C}/\text{min}$ , and was held at 310  $^{\circ}\text{C}$  for 10 min. MS was detected in electron impact (EI) mode at -70 eV with a centroid scan from 50–650  $m/z$ . The separation was monitored on a desktop computer using Agilent Mass Hunter software. The GC peaks were tentatively identified by searching the NIST 08 Mass Spectra Library.

### 2.2.3 Magnetic Nuclear Resonance Spectroscopy (NMR)

NMR spectra were recorded on a Bruker Avance 400 MHz, Avance 600 MHz, and Avance 700 MHz spectrometers (Karlsruhe, Germany). The 400 MHz instrument (5 mm probe) was maintained by Dr. Aleksej Kronic in the College of Pharmacy, UIC, the 600 MHz instrument (5 mm probe) by Dr. Benjamin Ramirez at the Center for Structural Biology, UIC. Access to the 700 MHz NMR spectrometer (1.7 mm micro-cryoprobe) was provided by the Institute for Marine Biosciences, National Research Council, Halifax, Canada. All NMR spectrometers were operated and regulated at 25  $^{\circ}\text{C}$  (298 K). The test sample was dissolved in the appropriate NMR solvent and transferred to an NMR tube. The NMR tube was securely sealed, placed in a spinner turbine, and inserted in the NMR probe. The signal was manually locked and shimmed, after which the pulse programs ( $^1\text{H}$ : zg30;  $^{13}\text{C}$ : zgpg30; COSY: cosygpqf; HSQC: hsqcedetgp; HMBC: hmbcgp1pndqf) were selected for the specific types of NMR spectra to be recorded. The acquisition parameters such as solvent type, data size (TD), spectral width (SW), transmitter offset (o1p/o2p), and acquisition time (AQ) were set as needed. Based on the solution concentration, the acquisition parameters including receiver gain (RG) and number of scans (NS) were properly selected so that sufficient signal-to-noise ratio (S/N) could be achieved.

The acquired data were processed offline using one or more of the following software packages: NUTS (AcornNMR, Livermore, CA), ACD/Labs NMR Suite (Advanced Chemistry

Development, Toronto, Canada), and MestReNova (Mestrelab Research, Santiago de Compostela, Spain). For  $^1\text{H}$  NMR spectra, the S/N ratio and digital resolution was enhanced by applying Lorentzian-Gaussian window functions (the LB and GF values were optimized for each sample) and two or three times zero-filling of the FID data. After Fourier transformation, the spectra were manually phase corrected, and baseline flattening was performed by applying  $n$ th-order polynomial correction ( $n \leq 10$ ). Finally, the chemical shifts ( $\delta_{\text{H}}$ , ppm) were referenced to the solvent residual signals (chloroform- $d$   $\delta_{\text{H}}$  7.240; methanol- $d_4$   $\delta_{\text{H}}$  3.310; pyridine- $d_5$   $\delta_{\text{H}}$  8.740; DMSO- $d_6$   $\delta_{\text{H}}$  2.500). Likewise, for 2D NMR spectra, the resolution was increased by applying zero-filling and forward linear-prediction along  $t_2$  dimension prior to Fourier transformation. Phase correction was performed when processing phase-sensitive spectra such as HSQC. The 100% integral method of qHNMR was applied for quantifications. This method is based on the sum and relations of all integrals obtained for both the analytes and the impurities (Pauli *et al.*, 2001, 2005).

#### 2.2.4 Antituberculosis Activity Evaluation

The Microplate Alamar Blue Assay (MABA) was used for evaluation of the anti-TB activity of fractions and isolates. *Mycobacterium tuberculosis* H<sub>37</sub>Rv was obtained from the American Type Culture Collection (Rockville, MD). H<sub>37</sub>Rv was cultured in 100 mL of Middlebrook 7H9 broth supplemented with 0.2% (v/v) glycerol, 10% (v/v) OADC (oleic acid, albumin, dextrose, catalase), and 0.05% (v/v) Tween 80, a culture medium referred to as 7H9GC-T80. The cultures were incubated in 300 mL nephelometer flasks on a rotary shaker (New Brunswick Scientific, Edison, NJ) at 150 rpm and 37 °C until they reached an optical density of 0.4–0.5 at 550 nm. The bacteria were washed and suspended in 20 mL of phosphate-buffered saline and passed through an 8  $\mu\text{m}$  pore size filter to eliminate bacterial clumps. The filtrates were aliquoted and stored at –80 °C. Rifampin was solubilized according to the manufacture's recommendation. Stock solutions were filter sterilized (0.22  $\mu\text{m}$  pore size) and stored at –80 °C. The 7H12 media

consisted of Middlebrook 7H9 broth supplemented with 0.1% casitone, 0.1% palmitic acid (5.6 mg/mL free acid in EtOH, Sigma), 10% albumin (50 mg/mL in water, Sigma), and 0.1% catalase (4 mg/mL in water, Sigma).

The minimum inhibitory concentrations (MICs) of each sample was determined using the MABA assay. Testing was performed in black, clear-bottomed, 96-well microplates (Packard Instrument, Meriden, CT) in order to minimize background fluorescence. Initial sample dilutions were prepared in DMSO, and subsequent 2-fold dilutions were performed in 0.1 mL of 7H12 media on the microplate. The inocula were initially diluted in 7H12 media to achieve approximately  $2 \times 10^5$  cfu/mL, and 0.1 mL was added to individual wells. Wells containing samples only were used to detect autofluorescence of samples. Additional control wells consisted of bacteria only (B) and medium only (M). Plates were incubated at 37 °C. At day 7 of incubation of plates inoculated with H<sub>37</sub>Rv, 20  $\mu$ L of Alamar blue solution (Trek Diagnostic Systems, Cleveland, OH) and 12.5 mL of 20% Tween 80 were added to all the wells, and plates were reincubated at 37 °C for 24 h. Fluorescence was measured in a Victor II multilabel fluorometer (Perkin-Elmer Life Sciences, Boston, MA) in bottom reading mode with excitation at 530 nm and emission at 590 nm. A background subtraction was performed on all wells using the mean of triplicate M wells. Percent inhibition was defined as  $1 - (\text{test well FU} / \text{mean FU of triplicate B wells}) \times 100$ . The lowest drug concentration effecting an inhibition of 90% was considered the MIC.

## 2.3 Data Analysis

Numerical data analysis was aided by various computer applications. Basic calculations and 2D chart graphing were accomplished in Microsoft Excel 2010. Graphical 3D plots were generated in SigmaPlot 11.0 (Systat Software Inc., Chicago, IL). Gaussian fitting was performed using an interactive model (developed by Professor Thomas O'Haver, the University of Maryland at College Park) in OpenOffice.org Calc 3.3 (Oracle, Redwood, CA). Statistical analysis including Pearson's correlation and linear regression was performed by SPSS 13.0 (SPSS Inc., Chicago, IL). The canonical discriminant analysis (CDA) was performed in Microsoft Excel 2010 with the XLSTAT-Pro 7.5 add-on (Addinsoft, Paris, France). The classification binary trees (CBTs) were generated by the classification and regression tree (CART) analysis within the Salford Predictive Miner v6.6 (Salford Systems, San Diego, CA). Chemical structures were drawn in ChemSketch 12.0 (Advanced Chemistry Development, Toronto, Canada) and ChemBioDraw Ultra 12.0 (CambridgeSoft, Cambridge, MA).

### 3. ACCELERATING AND TARGETING OF NATURAL PRODUCT SEPARATION<sup>1</sup>

#### 3.1 Introduction

As discussed in Chapter 1, the conventional separation process for bioactive NPs is far less efficient than expected. Investigation of the disadvantages has revealed that the process often starts with the crude extracts in which the target compounds usually have low content. Considering an ~20% permanent loss when using absorption chromatography, any given compound will be only recovered at less than approximate half of its original content after three purification steps. Frequently, the content of bioactive compounds is much lower than 1% and several rounds of purification are required to separate the structurally similar compounds. As a result, the yield of bioactives is frequently much below 0.1% or even 0.01% of the extract.

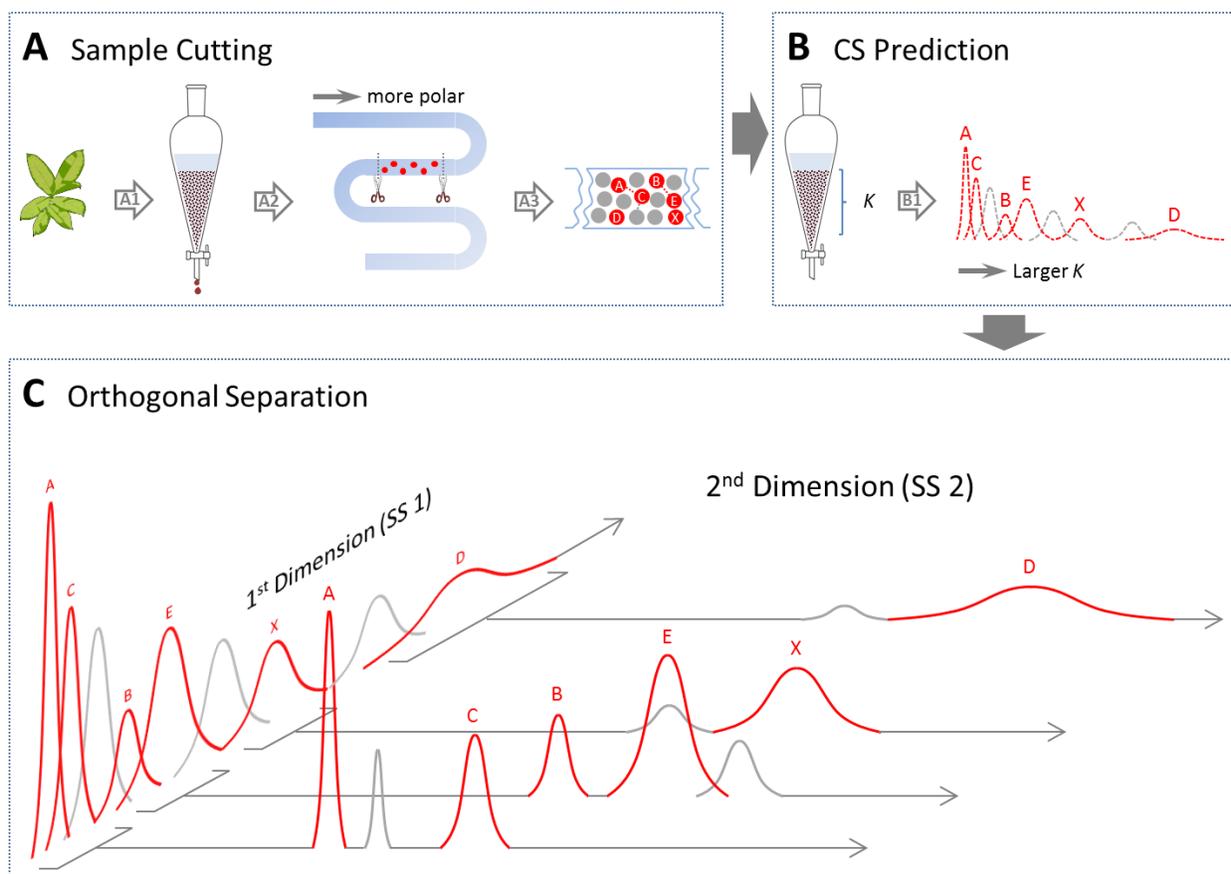
Additionally, the traditional process is almost blinded to scientists due to the lack of structural information of the bioactive constituents. While planar chromatography like TLC is frequently used as a fractionation monitoring tool for the optimization of the chromatographic conditions, it does not provide us with any structural information about the sample constituents but only predicts the relative polarity of the constituents and separation performance. In reality, however, these predictions appear to be relatively unreliable as the actual separation is usually performed on a different separation medium, e.g., column chromatography, which shows different selectivity of constituents in the sample.

As a result of these problems, the separation is often inefficient and frequently leads to loss of target compounds or may even lose track of them completely. In order to overcome these difficulties, simplification and visualization are suggested to unravel the complexity of the separation process and improve the efficiency and effectiveness in isolation of pure compounds.

**Figure 4** shows the experimental design of a new separation scheme for NPs. The details are described as follows:

---

<sup>1</sup> Contents presented in Sections 3.2.1, 3.3, and 3.5.1 has been partially published in: Qiu, F.; Friesen, J. B.; McAlpine, J. B.; Pauli, G. F. Design of Countercurrent Separation of Ginkgo biloba Terpene Lactones by Nuclear Magnetic Resonance, *Journal of Chromatography A* 2012, 1242, 26–34



**Figure 4. The Conceptual Design of a New Separation Scheme for NPs**

Panel A shows the flowchart for an efficient sample preparation approach. The crude plant extracts are processed by a liquid-liquid partitioning technique (shake-flask or CS), affording enriched samples of target compounds. In Panel B, the CS profiles of the target compounds are predicted and simulated by measuring their  $K$  values in the candidate SSs in a shake-flask. Thus, the CS conditions can be optimized prior to actual separation. Panel C shows that an orthogonal separation is employed to enhance the resolution of NPs, using the different selectivity of various chromatographic methods and/or SSs in sequential purification schemes.

**Steps A1 to A3:** Instead of directly using crude extracts at the start, the sample is initially divided into a few fractions according to the range of polarity. The purpose of this **sample-cutting** approach is to prepare an enriched sample by removing the unwanted compounds by use of a liquid-liquid partitioning technique.

**Step B: CS** is the initial choice for separation of NPs. As a prime consideration, the process of CS is highly and accurately predictable by measuring the partition coefficients ( $K$  values) of the constituents. Secondly, CS is more economical and environmentally friendly. It also has, theoretically, 100% recovery of the sample. Combined with NMR techniques, the process of chromatographic separation can be structurally visualized and thus becomes more readily optimized and targeted to the compounds of interests. In particular, the measurement of  $K$  values by NMR ( **$K$ -by-NMR**) allows for simultaneous prediction of separation performance of multiple components in mixtures, which facilitates selection and optimization of SSs. Using these  $K$  values, the isolates can be readily located in the fractions, making CS a highly efficient targeted separation.

**Step C:** When using column or countercurrent chromatography, the optimized SSs are the keys to unlock the complexity of NPs and relate to efficient isolation of pure compounds. Compared to repeatedly using the same chromatographic methods and SSs, the concept of **chromatographic and solvent orthogonality** may be more effective and efficient in separation of NPs, especially those congeners with similar chemical properties. By using the different selectivity of orthogonal chromatographic methods (e.g., NP vs. RP) and/or SSs (e.g., EtOAc-based vs.  $\text{CHCl}_3$ -based), the complex NPs can be well resolved in a simplified fractionation scheme. Additionally, **solvent modification** is an enhancement to applied orthogonality in which a minute amount of a modifier acts as a resolution enhancer by changing the physical and/or chemical properties of analyte molecules.

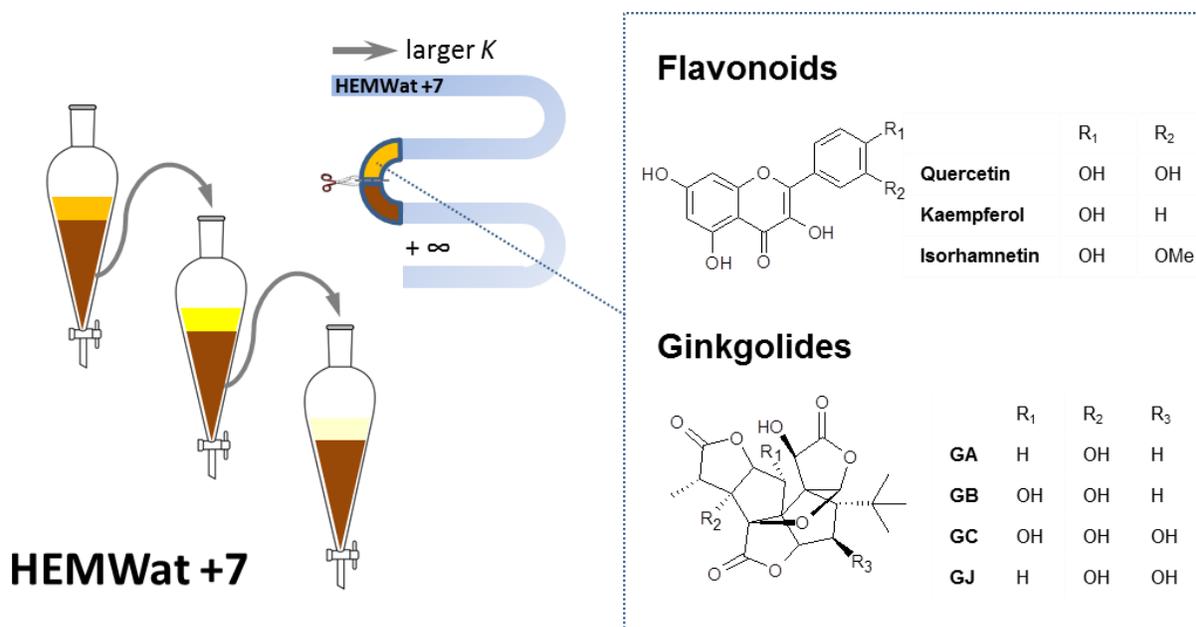
Chromatography is usually coupled with spectroscopy as an online or offline detector for recording elution profiles. Some limitations are associated with commonly used spectroscopic methods, such as UV and ELSD. For example, UV is unsuitable for detecting UV-inactive molecules. Both UV and ELSD provide limited structural information of elutes. The current study uses qHNMR as a fractionation control tool or an offline detector in chromatographic separation. Compared with other detection methods, NMR is universal for all molecules with NMR-active

nuclei (e.g.,  $^1\text{H}$  and  $^{13}\text{C}$ ) and the  $^1\text{H}$  signals have a linear relationship with analyte concentration. Most importantly, NMR visualizes the separation process by providing valuable structural information of the eluents. Therefore, it is particularly powerful in target separation when combined with *K*-by-NMR methodology.

## 3.2 Sample Cutting

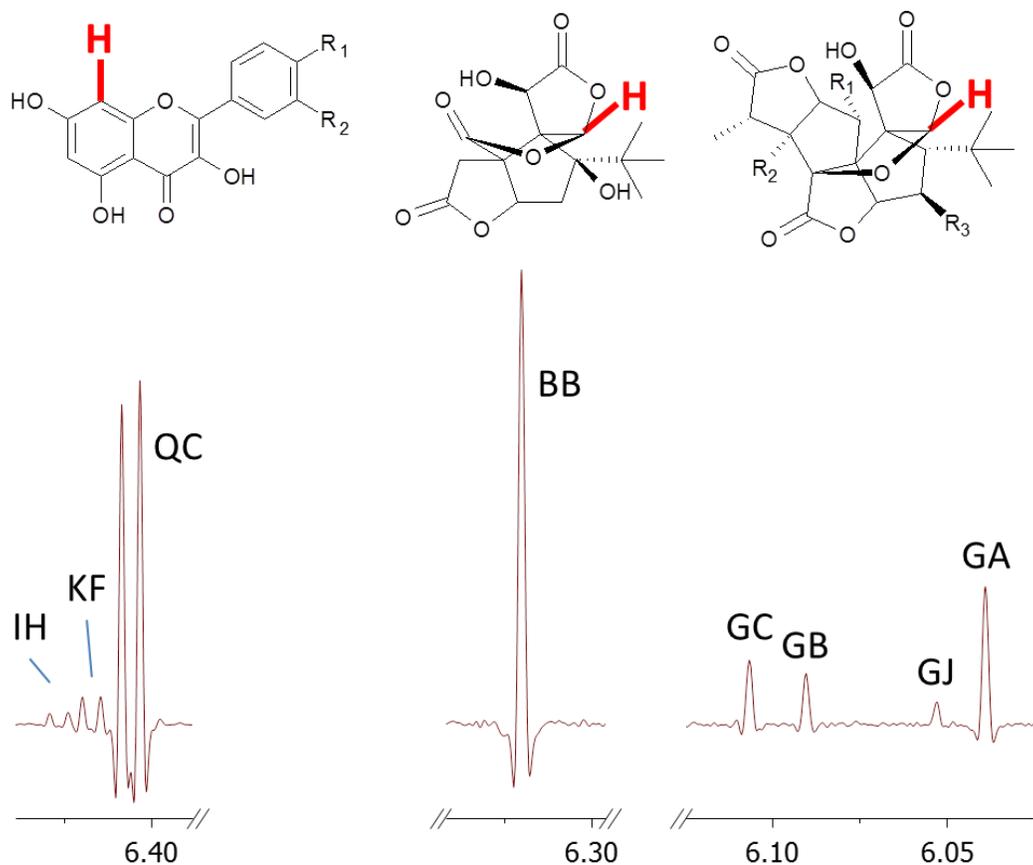
### 3.2.1 Enrichment of Flavonoids and Terpene Lactones from *Ginkgo biloba*

Ginkgo is known as a living fossil in the plant kingdom and has been valued for its pharmaceutical properties for centuries. Ginkgo leaf extract is one of the most popular dietary supplements in the United States and worldwide (Blumenthal *et al.*, 2006). It has demonstrated beneficial pharmacological effects against neurodegenerative diseases, cardiovascular diseases, oxidative stress, cancer, tinnitus, geriatric, and psychiatric disorders (Chan *et al.*, 2007; Mahadevan *et al.*, 2008). Pharmacognostic studies have established flavonoids and terpene lactones as the main bioactive compounds. The major flavonoids found in ginkgo leaf include isorhamnetin (IR, **1**), kaempferol (KF, **2**), and quercetin (QC, **3**), and the major terpene lactones: bilobalide (BB, **4**), ginkgolides A (GA, **5**), B (GB, **6**), C (GC, **7**), and J (GJ, **8**).



**Figure 5. Preparation of a Terpene Lactone-Enriched Sample from Ginkgo Leaf Extract by Liquid-Liquid Partitioning**

Terpene lactones content increased 3-fold by removing hydrophilic substances from the crude extract using HEMWat +7 in a shake-flask.



**Figure 6. Identification of Biomarker Compounds in Ginkgo Preparations by  $^1\text{H}$  NMR**

The presence of five terpene lactones and three flavonoids in the terpene lactone-enriched sample (TLES) can be readily identified by their  $^1\text{H}$  NMR characteristic signals with distinct chemical shifts (methanol- $d_4$ , 400 MHz):  $\delta_{\text{H}}$  6.311 (BB), 6.029 (GA), 6.081 (GB), 6.097 (GC), 6.044 (GJ), 6.476 (IR), 6.433 (KF), and 6.401 (QC).

The content of total terpene lactones in ginkgo leaf extract is only ~6%. Clearly, enrichment of these compounds prior to separation is necessary for large-scale preparative isolation. The sample cutting methodology for the preparation of an enriched sample uses liquid-liquid partitioning and requires selection of the optimal SSs with polarity matching that of the target compounds. The biphasic HEMWat system is considered a first choice due to its coverage of a wide polarity range. Ginkgo leaf extract is mainly comprised of ginkgolic acids,

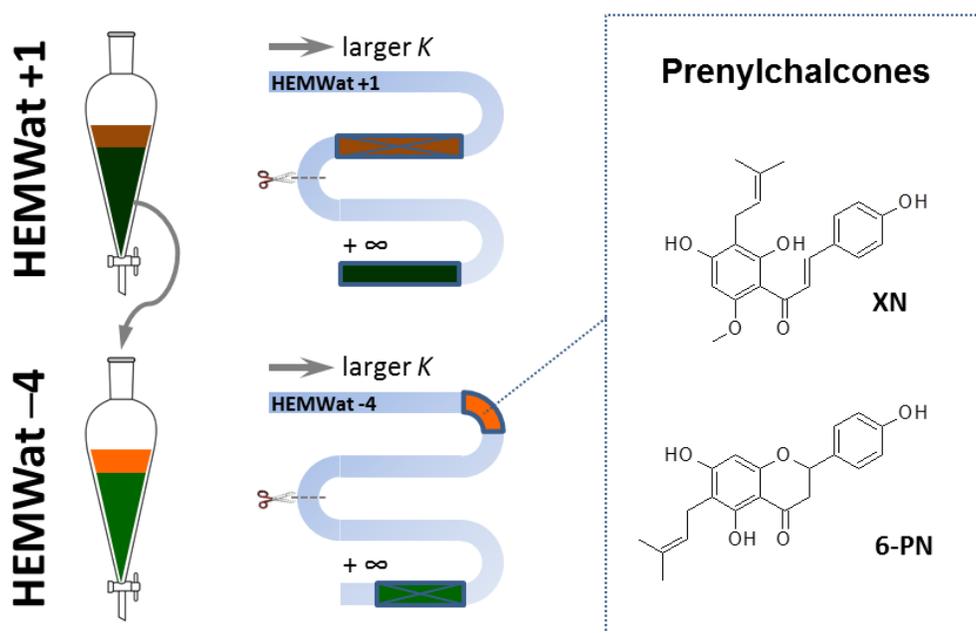
flavonoids, and terpene lactones. Therefore, HEMWat +7 is an appropriate SS with which all ginkgolic acids can be removed in the aqueous phase (lower phase). While minor terpene lactones (especially more polar GC/GJ) may be distributed in the lower phase, three repetitive partitionings is thought to be sufficient to achieve concentration with minimal loss of these compounds.

As shown in **Figure 5**, ginkgo leaf extract (600 mg) was added to HEMWat +7 with 5 mL each of upper and lower phase. The solution was shaken vigorously and centrifuged to break up any emulsion and to accelerate phase separation. The upper phase was separated from the solution. Another 5 mL of fresh upper phase was added to the remaining lower phase. The abovementioned partitioning procedure was repeated twice. The upper-phase solutions were combined and evaporated in a SpeedVac to dryness and yielded ~200 mg of terpene lactone-enriched sample (TLES). As a result, by removing the hydrophilic constituents in the crude extract, the terpene lactones were enriched 3-fold (to ~18%) in the TLES. Further qHNMR analysis of the TLES (**Figure 6**) revealed that ~80% of the contents were flavonoids including quercetin, kaempferol and isorhamnetin. While flavonoids and terpene lactones exhibit similar polarities, these compounds can be well resolved in CS (see Section 3.5.1, p. 54; Section 4.2.2, p. 73).

### **3.2.2 Enrichment of Prenylphenols from *Humulus lupulus***

Hops are well known for its use in the brewing industry and as a mild sedative in phytomedicines. Recent studies have been initiated to explore their phytoestrogenic properties and other potential benefits to women's health. Prenylphenols have been found in hops with bioactivities to alleviate symptoms related to menopause (Kitaoka *et al.*, 1998; Milligan *et al.*, 1999). Previous separation of prenylphenols from hops began with crude extracts by initial fractionation using column chromatography. In addition to prenylphenols, crude hops extracts contain lipophilic substances and hydrophilic  $\alpha/\beta$ -acids (Neve *et al.*, 1991; Verzele *et al.*, 1991).

Based on the polarity distribution of these constituents, **Figure 7** shows a design of the sample-cutting method for preparation of prenylphenol-enriched sample (PPES) by two steps of liquid-liquid partitioning. The procedure was carried out in a shake-flask in a manner similar to that described for the preparation of TLES. Initially, 520 mg of crude hops extract was partitioned in HEMWat +1, in which acids and prenylphenols were separated in the aqueous lower phase, leaving less polar substances in the organic upper phase. The lower phase was dried in a SpeedVac and further partitioned in HEMWat -4. As a result, the acids were removed in the lower phase, and 260 mg of PPES was obtained from the upper phase and used for targeted isolation of xanthohumol (XN, **9**) and 6-prenylnaringenin (6-PN, **10**) (see Section 3.5.2, p. 58).



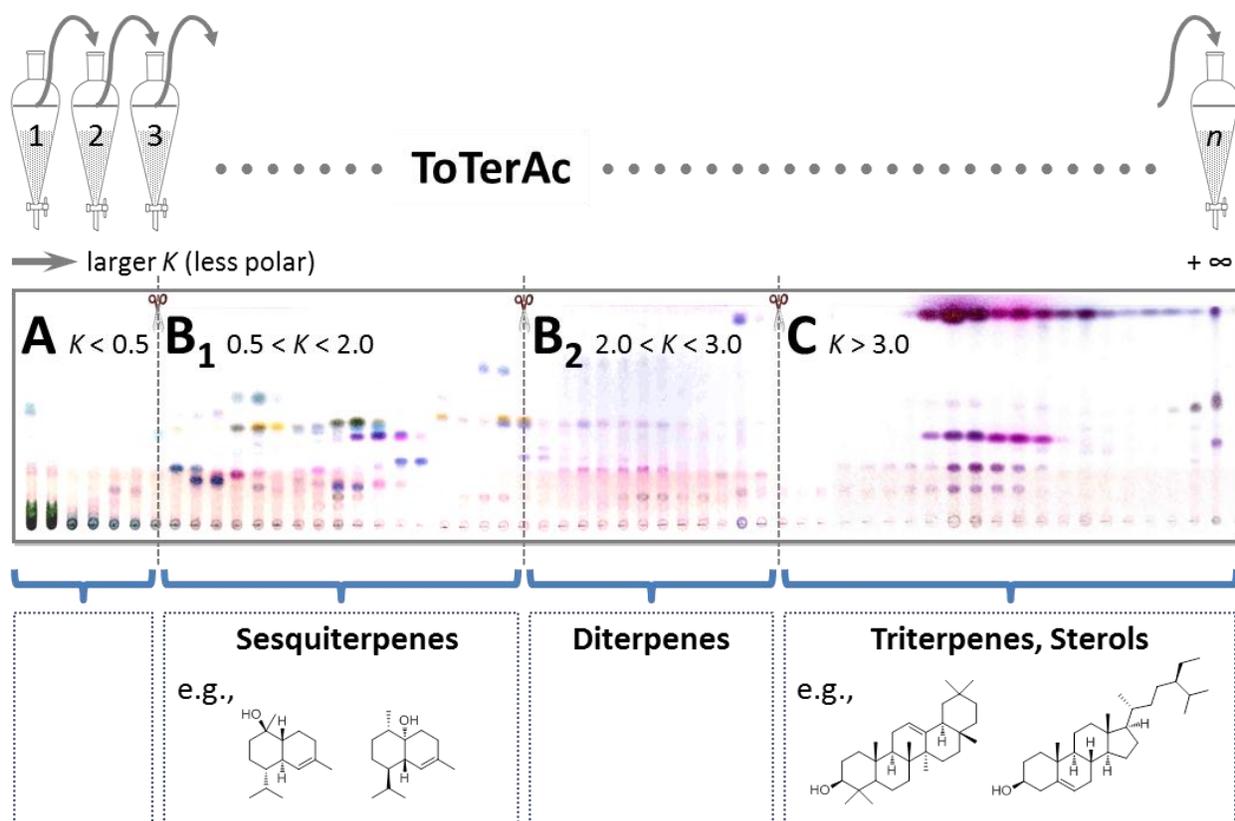
**Figure 7. Preparation of a Prenylphenol-Enriched Sample from Hops Crude Extract by Liquid-Liquid Partitioning**

By removing lipophilic and acidic substances in two steps of shake-flask partitioning using HEMWat +1 and -4 SSs, the content of XN and 6-PN was enriched by 2-fold.

### 3.2.3 Enrichment of Lipophilic Components from *Humulus lupulus*

In addition to liquid-liquid partitioning in shake-flasks, CS is also an efficient and effective approach for sample cutting in the preparation of enriched sample. By multiple partitions during CS process, the components within a narrow range of polarity can be well separated based on the compound classes.

The lipophilic hops extract serves as an excellent example. By using the ToTerAc SS (toluene–methyl tert-butyl ether–acetonitrile, 1:0.1:1, v/v), 100 mg of lipophilic hops extract was fractionated on a 120 mL HSCCC instrument in normal-phase mode with stationary phase retention of 56% at a flow rate of 1.5 mL/min and rotation speed of 1000 rpm. According to the  $K$  values, the fractionation by HSCCC resulted in three regions: A ( $K < 0.5$ ), B ( $0.5 < K < 3.0$ ) and C ( $K > 3.0$ ), of which B is considered as the “sweet spot” where the best separation is usually achieved. In **Figure 8**, TLC analysis of HSCCC fractions indicated the chemical diversity and complexity in these three regions and showed that various chemically distinct components exhibit similar TLC behavior in terms of  $R_f$  value. Further GC/MS analysis (see Section 6.2, p. 141) as well as preparative isolation revealed that Region C contained the less polar triterpenes and sterols. However, Region B consisted of two subregions  $B_1$  ( $0.5 < K < 2.0$ ) and  $B_2$  ( $2.0 < K < 3.0$ ) which mainly contained sesquiterpenes and diterpenes, respectively. As a consequence, the lipophilic hops extract was separated into three fractions enriched with different types of compounds, making further separation and/or characterization more focused on the compounds of interest. In addition, the CS-based sample cutting enhances the reproducibility of the process due to its focus on the partition coefficients.



**Figure 8. Example of a Sample-Cutting Procedure Facilitated by HSCCC Fractionation**

Lipophilic hops extract was separated into three enriched fractions, individually comprising specific types of compounds, which are well-suited for chemical and biological characterization.

Compared to conventional column chromatography, liquid-liquid partitioning in both shake-flask and CCC is a more efficient approach for sample cutting in the initial step of separation. The process of sample cutting not only enriches target compounds, but also removes unwanted components which may negatively affect the further separation and characterization. This removal is particularly useful in the standardization or quality control of the raw materials for dietary supplements. The complex composition of the crude extracts could cause significant interference with MS or NMR analysis. However, the enriched sample is much less compositionally complex and thus is more suitable for use, both qualitatively and quantitatively, for evaluation of biomarker compounds in the extracts.

### 3.3 Prediction of CS of Ginkgo Terpene Lactones

The partition coefficient is not only an important theoretical parameter in CS, but also of much practical value. It determines when a particular compound elutes from the CS column, i.e., the retention volume. In addition,  $K$  values allow prediction of whether any two particular compounds can be separated, i.e., their resolution ( $R_s$ ) from one another. Thus, the measurement of  $K$  values of target compounds is an important aid in the selection of appropriate CS SSs. The classical approach is to partition the pure analyte in the candidate biphasic SS and calculate the concentration ratio of the compound distributed into the two phases. However, for NPs which are unavailable commercially or otherwise in pure form or for unknown bioactive compounds, this requires prior purification from crude mixtures or synthesis. Both often preclude the practicality of the CS experiment from the start. While HPLC or GC is frequently used to work with mixtures, these methods are often associated with detection problems. Owing to the absence of carbon-carbon unsaturation, terpene lactones show only a weak maximum UV absorption around 220 nm. Therefore, UV is unsuitable for detection and quantitation purposes, leaving refractive index (RI), evaporative light scattering (ELSD), NMR or MS as likely potential techniques. RI and ELSD lack selectivity for distinction of the individual terpene lactones. Moreover, both RI and ELSD, as well as MS require calibration with authentic standards for quantitative analysis. However, qHNMR is not limited by these factors due to the almost ubiquitous occurrence of protons in organic compounds. Furthermore, the absolute integral of the  $^1\text{H}$  NMR signal of an analyte has a linear relationship with analyte concentration in the sample (Pauli *et al.*, 2005). Taking these advantages of NMR, the present study used a combination of classical shake-flask experiments and qHNMR analysis to simultaneously determine multiple  $K$  values of target analytes in mixtures. As shown in **Figure 6**, the signals of H-12 of the five terpene lactones are well-separated singlets with chemical shifts  $\delta_{\text{H}}$  6.311 (BB), 6.029 (GA), 6.081 (GB), 6.097 (GC) and 6.044 (GJ) which could be used for determination and quantitation purposes (van Beek *et al.*, 1993; Choi *et al.*, 2003; Li *et al.*, 2004). Because the

upper and lower-phase samples were analyzed under identical experimental conditions, the  $K$  values of each terpene lactone could be calculated as the ratio of absolute integral of corresponding individual H-12 signals in the two phases.

An experimental protocol for determining  $K$  values in mixtures by qHNMR ( $K$ -by-NMR) was developed as shown in **Figure 9**. The procedure consists of five steps:

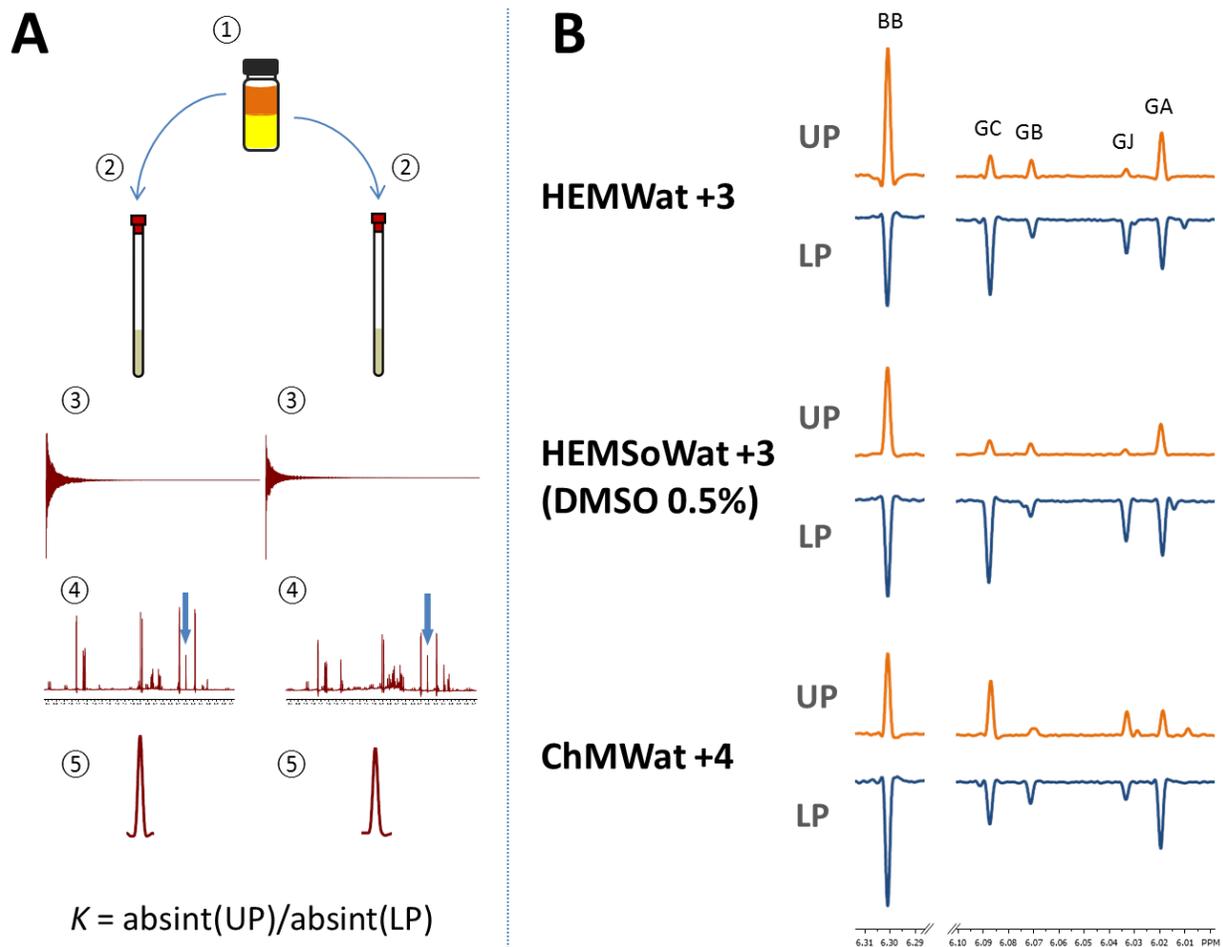
**Step 1:** A certain amount of sample is dissolved in a biphasic SS with an equal volume of upper and lower phase. The solution is shaken vigorously until the solutes are thoroughly dissolved. If necessary, the solution is centrifuged to break up any emulsion until the surface between upper and lower phase is completely clear.

**Step 2:** Equal aliquots of upper and lower-phase solution are accurately transferred in two vials and evaporated *in vacuo* to dryness. Before NMR analysis, the two samples are stored in a desiccator *in vacuo* overnight to eliminate any moisture.

**Step 3:** The two samples are dissolved in equal volumes of an appropriate NMR solvent, accurately dosed with an analytical syringe, and transferred into the same type of NMR tubes. After the NMR spectrometer is properly locked and shimmed,  $^1\text{H}$  NMR spectra of the two samples are recorded, respectively, under the same acquisition parameters which include pulse program, number of data points (TD), acquisition time (AQ), receiver gain (RG), and number of scans (NS).

**Step 4:** The FIDs of the two spectra are processed using identical conditions (window function, zero-filling, and solvent-peak calibration). The two spectra are properly phased and baseline corrected.

**Step 5:** The characteristic signals of target compounds in the mixture are identified in the spectra according to the literature data. The signals used for calculations of  $K$  values are ideally, but not necessarily, singlets without an overlap with other signals. The  $K$  value of each target compound is calculated as the ratio of absolute integral of corresponding characteristic signals in the two samples.



**Figure 9. The Protocol for Determining Partition Coefficients of Multiple Components in Mixtures by qHNMR**

Panel A: The experimental procedure of *K*-by-NMR. Panel B: Using this protocol, the partition coefficients of five ginkgo terpene lactones in a series of biphasic SSs were calculated as the ratio of absolute integral (absint) of corresponding characteristic signals in the UP and LP samples.

By using this *K*-by-NMR protocol, the *K* values of five ginkgo terpene lactones in various SSs were measured in TLES as shown in **Figure 9B** and **Table 3**. These *K* values were used as guidance for the selection of appropriate CS SSs. As determined by their *K* values, BB, GA, and GB could be separated in ChMWat +4 (10:7:3, v/v), while the GC/GJ pair was poorly resolved. In HEMSoWat +3/0.5% (4:6:4:6, 0.5% DMSO, v/v), however, all five terpene lactones could be well resolved.

**TABLE III. PARTITION COEFFICIENTS OF FIVE TERPENE LACTONES IN THREE POLARITY-ADJUSTED SS**

	ChMWat +4	HEMWat +3	HEMSoWat +3/0.5%
BB	0.68	0.78	0.84
GA	0.35	1.16	1.46
GB	0.53	1.16	1.18
GC	1.24	3.74	4.63
GJ	1.21	4.47	5.34

For CS, when the operational parameters including flow rate (*Q*) and stationary phase retention (*S<sub>f</sub>*) remain constant, the analytes' retention is solely dependent on their partition coefficient in the SS, namely, *K* values. Therefore, the retention volume of analytes can be calculated using the following equation:

$$V_R = V_m + K \times V_s \quad (1)$$

where *V<sub>R</sub>* is the retention volume of an analyte; *V<sub>m</sub>* and *V<sub>s</sub>* denotes the volume of mobile and stationary phase, respectively.

Thus, the retention time (*t*) can be calculated as:

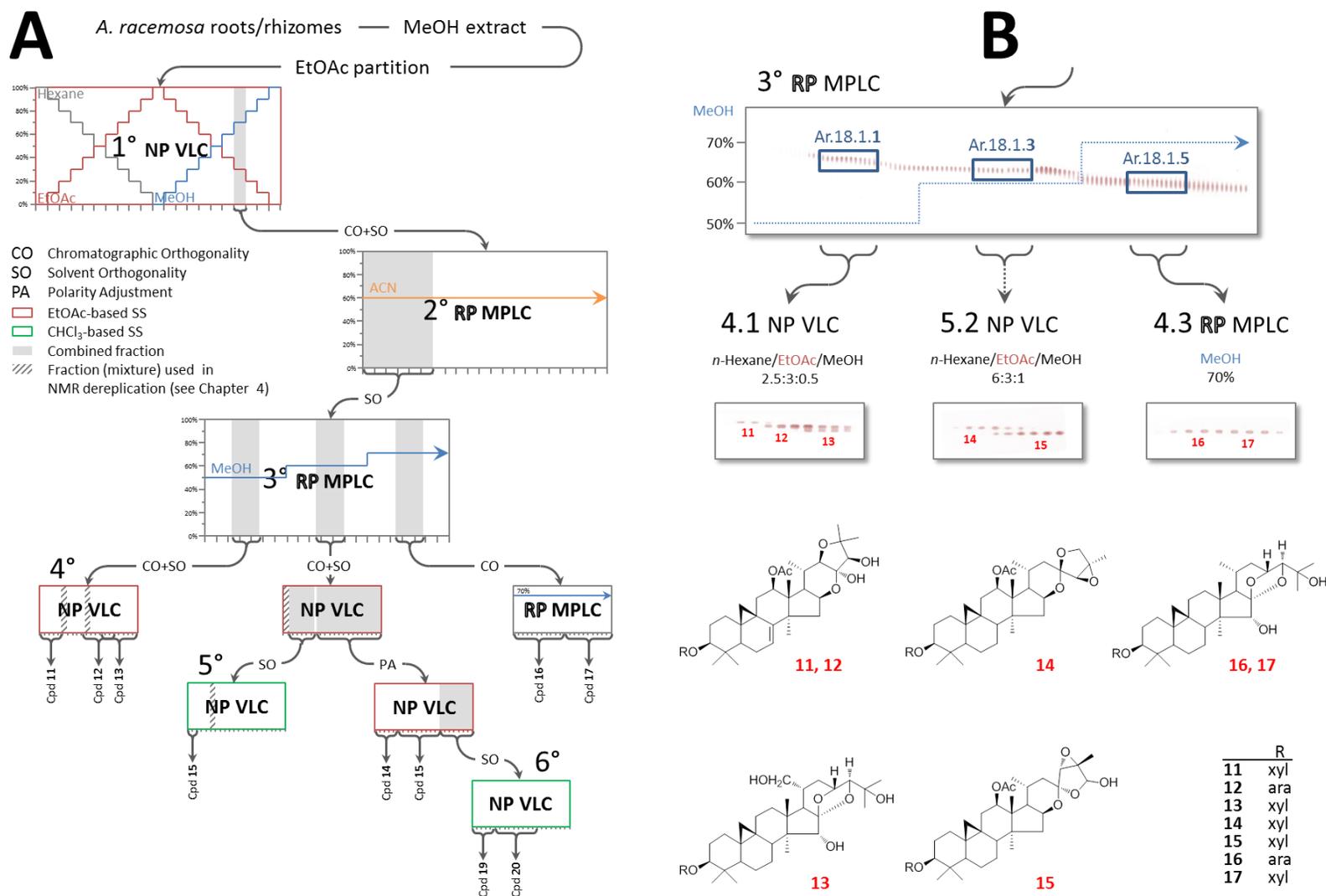
$$t = \frac{V_C - V_0}{Q} (1 - S_f + K \times S_f) \quad (2)$$

where  $V_C$  and  $V_0$  denotes the column volume and dead volume, respectively. Therefore, the time when compounds elute from the column can be predicted prior to actual separation. The further CS results (see Section 4.2.2, p. 73) have shown a high accuracy of this mathematical relationship which demonstrates its reliable applicability in targeted isolation. By adding the power of NMR to the  $K$ -based separation, CS becomes a structurally visualized process which is highly efficient in isolation of target compounds from complex mixtures.

### 3.4 Resolution of *Actaea* Triterpenes by Chromatographic Orthogonality

Owing to their closely related chemical properties, structurally similar NP congeners are usually difficult to separate by traditional chromatographic methods. Similar compounds are best separated using orthogonal chromatographic systems with different selectivities for the individual structural characteristics. In the present work, 2D solid–liquid chromatographic systems combining reversed and normal phase modes were used in the separation of cycloartane triterpenoids from black cohosh.

As shown in **Figure 10**, in the 3<sup>rd</sup> step of fractionation, a triterpene-enriched sample was fractionated on a user-assembled RP MPLC system (MPLC column #2, see p. 30) with a stepwise gradient elution using aq. MeOH (50–70%) as mobile phase. According to RP-TLC analysis, all the fractions gave rise to single broadened spots with gradually decreasing  $R_f$  value along the fractionation pathway. Using NP-TLC instead, these fractions were shown to be mixtures of more than two or three major compounds. Thus, fractions were combined according to the NP-TLC profiles and further resolved in a second preparative dimension using NP-VLC. The combined fraction (Ar.18.1.1) was fractionated on an NP-VLC (silica gel 60) using *n*-hexane–EtOAc–MeOH (2.5:3:0.5, v/v) as the SS, which led to separation of three triterpenoids: cimracemoside F (**11**), cimracemoside G (**12**), and 21-hydroxycimigenol xyloside (**13**). Similarly, Ar.18.1.3 was further separated by NP-VLC using the less polar SS of *n*-hexane–EtOAc–MeOH (6:3:0.5, v/v), which yielded two acteols: 23-*epi*-26-deoxyactein (**14**) and actein (**15**). While Ar.18.1.5 was unresolved in the 4<sup>th</sup> fractionation, an additional round of separation on a different RP column (Lobar, C<sub>18</sub> silica gel) led to resolution of a critical pair of cimigenol-type triterpenoids: cimigenol xyloside (**16**) and cimigenol arabinoside (**17**). It is noteworthy that the purity of these isolates was more than 90% determined by qHNMR analysis, which indicates that the performance of VLC using optimized SSs can be comparable or even better than HPLC separation. Besides, VLC is a more efficient and economical approach in large-scale purification, when compared to preparative HPLC.



**Figure 10. Preparative Separation of *Actaea* Triterpenes by Orthogonal Chromatographic and Solvent Systems**

Panel A shows the complete fractionation flowchart. Panel B details the 3<sup>rd</sup>–5<sup>th</sup> steps of fractionation enhanced by chromatographic orthogonality (RP vs. NP) and monitored by TLC analysis.

As shown in this example, both polarity and selectivity affect the separation on silica gel-based LC. The latter is a more important factor for resolution of NP congeners, especially when employing NP-LC. This also shows that structurally related compounds can have either very close or vastly different chromatographic behavior. This serves as an example of the chemical complexity and chromatographic overlap of plant metabolomes, and explains why the establishment of optimal chromatographic conditions largely remains an empirical exercise to date. For the best separation results, both RP- and NP-LC need to be optimized based on polarity and selectivity for the specific sample. In practical use, RP-LC might be better used as a procedure for sample cutting or “polarity-based fractionation” in the initial step of separation. However, as many more candidate SSs are available for NP-LC, it is more suitable for “selectivity-based fractionation” which can better resolve structurally similar compounds. The details of SSs optimization and the use of orthogonal SSs are discussed in the next section.

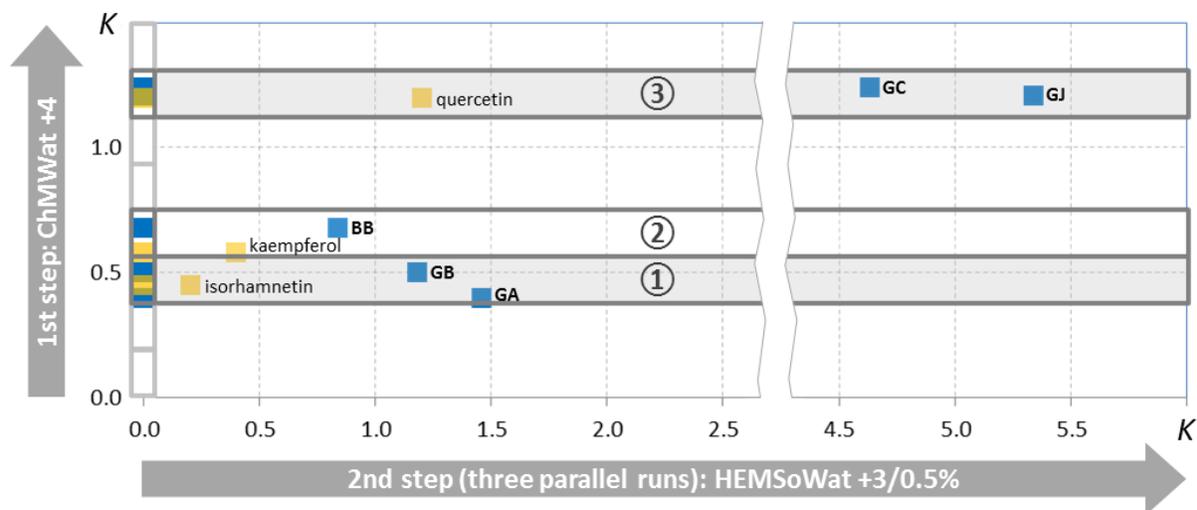
### 3.5 Solvent Orthogonality

Optimization of SSs is a key step toward the success of chromatographic separation. Traditionally, separation of NPs has been performed by multi-step fractionation using the same generic SSs at various volume ratios in order to resolve difficult-to-separate compounds. While SS polarity is often given more consideration than its selectivity, the latter can significantly affect the separation, especially for compounds with similar structural characteristics.  $\text{CHCl}_3/\text{CH}_2\text{Cl}_2$  and EtOAc, which are commonly used major constituents of SSs in liquid chromatography, exhibit different relative dipole moments and acid/base properties. Accordingly, their chemical selectivity can be mapped in a triangular fashion (Johnson *et al.*, 2007). However, these differences in physicochemical properties can also translate into profound selectivity differences and chromatographic orthogonality. For the purpose of developing a more efficient approach for the separation of structurally similar NPs, the current study has explored the use of EtOAc- and  $\text{CHCl}_3/\text{CH}_2\text{Cl}_2$ -based orthogonal SSs in preparative chromatography.

#### 3.5.1 Resolution of Ginkgo Terpene Lactones

According to the  $K$  values measured by qHNMR-guided shake-flask experiments, it was initially determined that all five terpene lactones could be separated in a single HSCCC run by using HEMSoWat +3/0.5%. In practice, however, the TLES sample contained other compounds such as flavonoids and benzoic acid derivatives which can exhibit chromatographic overlap with terpene lactones. Trial fractionation of the enriched sample was performed on a 20 mL HSCCC by using ChMWat +4 and HEMSoWat +3/0.5%, respectively.  $^1\text{H-NMR}$  analysis of the ChMWat +4 fractions indicated that kaempferol, isorhamnetin, and quercetin overlapped with BB, GA/GB and GC/GJ, respectively. In HEMSoWat +3/0.5%, however, kaempferol and isorhamnetin eluted at the front, while quercetin overlapped with GA/GB. Based on these results, the most practical approach for purifying the five terpene lactones from the enriched sample was a two-step HSCCC fractionation procedure, using ChMWat +4 and HEMSoWat +3/0.5% as a pair of

orthogonal SSs (**Figure 11**). Using this orthogonality-enhanced HSCCC, the three originally interfering flavonoids, which are also considered bioactive compounds in *G. biloba*, can be purified as well.

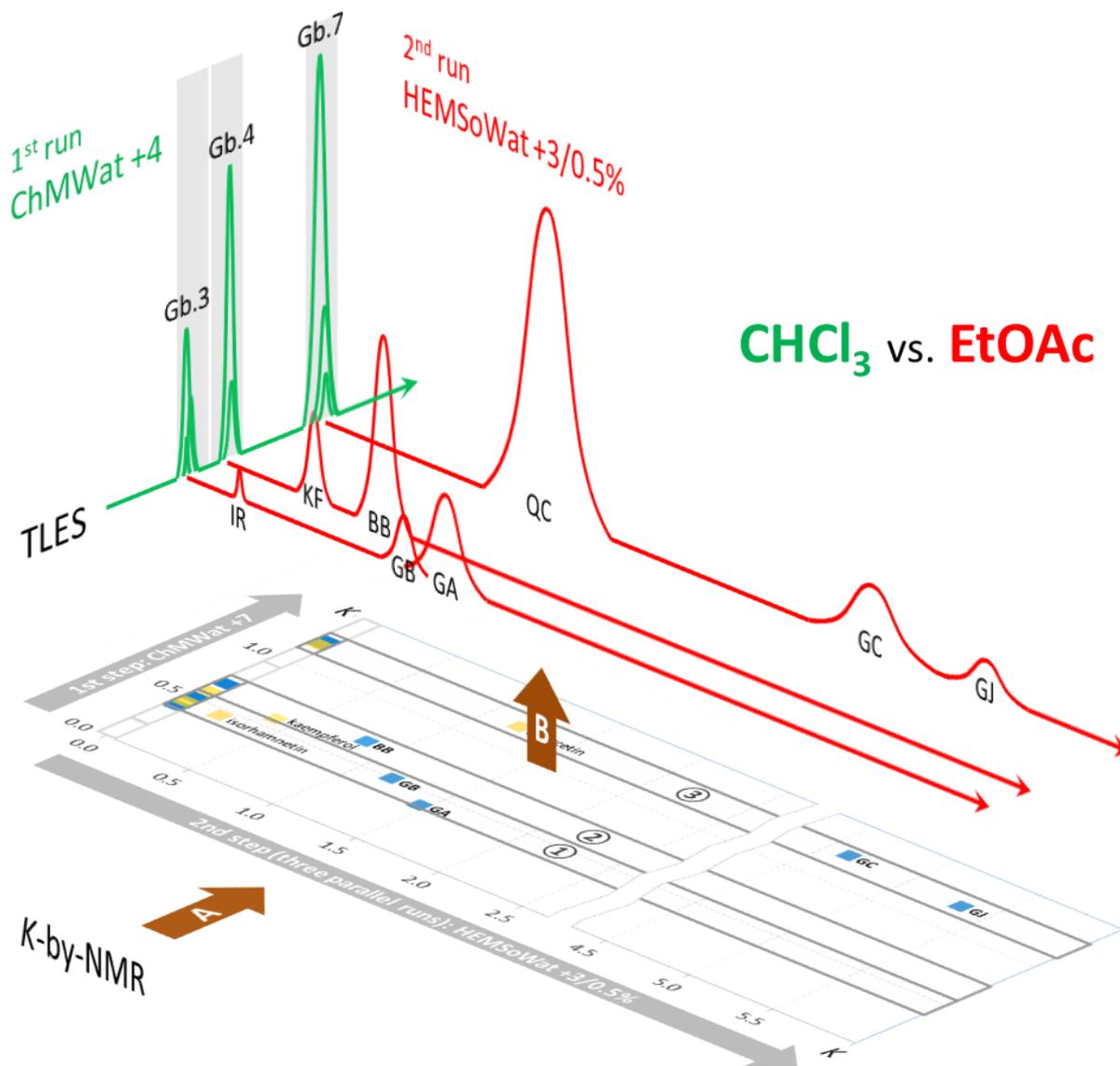


**Figure 11. Proposed Procedure for the Separation of Five Ginkgo Terpene Lactones**

An enriched sample (TLES) is initially separated by ChMWat +4 into three combined fractions, zones ①–③: ① a mixture of isorhamnetin, GA and GB; ② a mixture of kaempferol and BB; and ③ a mixture of quercetin, GC and GJ. These three samples are subsequently separated by an orthogonal SS of HEMSoWat +3/0.5%.

Based on the proposed fractionation scheme, 100 mg of TLES was initially fractionated on a 120 mL HSCCC instrument in normal-phase mode by using ChMWat +4. With a flow rate of 1 mL/min and rotation speed of 800 rpm, stationary phase retention was 50% when the hydrodynamic equilibrium was established within the column. The fractions were collected in 12 mL aliquots until a total volume of 180 mL ( $K = 2$ ) eluted from the column. According to  $^1\text{H}$  NMR analysis, the fraction Gb.3 ( $K = 0.35\text{--}0.55$ ) contained GA, GB, kaempferol and isorhamnetin;

Gb.4 ( $K = 0.55\text{--}0.75$ ) contained BB and kaempferol; Gb.7 ( $K = 1.15\text{--}1.35$ ) contained GC, GJ, and quercetin. These three combined fractions were each subjected to the second step of HSCCC (20 mL coil) in normal-phase mode using HEMSoWat +3/0.5%. The fractionation process was monitored offline and evaluated by qHNMR (see Section 4.2.2, p. 73). As suggested by the qHNMR analysis, all five terpene lactones were well separated from the interfering flavonoids. Meanwhile, the two critical pairs of GA/GB and GC/GJ were also resolved (**Figure 12**).

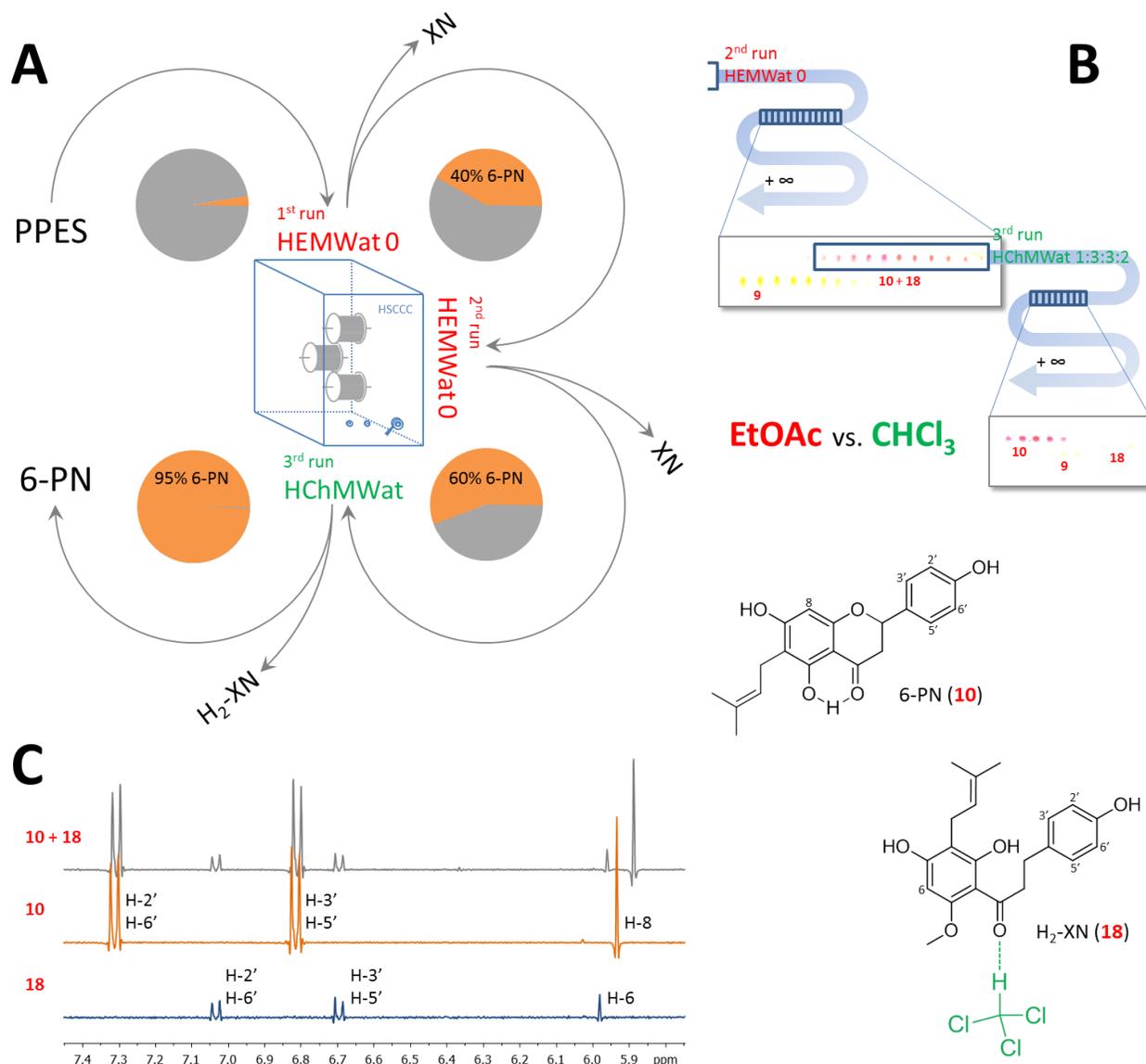


**Figure 12. Orthogonal CS Process for the Separation of Three Flavonoids and Five Ginkgo Terpene Lactones**

Step A: The CS process was predicted and mapped from the *K*-by-NMR measurements. Step B: Guided by the “*K*-map”, the targeted separation was performed in a 2D CS procedure using a pair of orthogonal SSs, ChMWat +4 and HEMSoWat +3/0.5%. As a result, five ginkgo terpene lactones together with three flavonoids were efficiently separated with excellent resolution.

### 3.5.2 Resolution of Prenylphenols from *Humulus lupulus*

Orthogonal SSs were also used in targeted isolation of 6-prenylnaringenin (6-PN) from the prenylphenol enriched sample (PPES) of hops. PPES comprised of xanthohumol (XN) as a major component, ~2% 6-PN, and other minor prenylphenols. The sample cutting approach using CCC was implemented into the separation process. Based on the previous study (Chadwick *et al.*, 2005), HEMWat 0 was an optimized SS to resolve XN from other prenylphenols. Therefore, 90 mg of PPES was initially fractionated on a 120 mL HSCCC in reversed-phase mode using HEMWat 0. NP-TLC was used to analyze the fractions. According to the literature data, XN and 6-PN appear as yellowish and pinkish spots, respectively, on the TLC plate stained by vanillin/EtOH/H<sub>2</sub>SO<sub>4</sub>. Therefore, all the fractions showing the pinkish spot on TLC were combined, leading to 4.2 mg of further enriched sample (6PNES) with ~40% 6-PN. The same HSCCC process was repeated once for the enriched sample (6PNES) and afforded 2.8 mg of crude 6-PN (~60%). TLC results suggested that 6-PN was well separated in the second step of HSCCC only with a minor overlap with XN. However, further NMR analysis revealed that these previously considered pure fractions were in fact a mixture of 6-PN and minor 1,2-dihydroxanthohumol (H<sub>2</sub>-XN, **18**) (**Figure 13**).



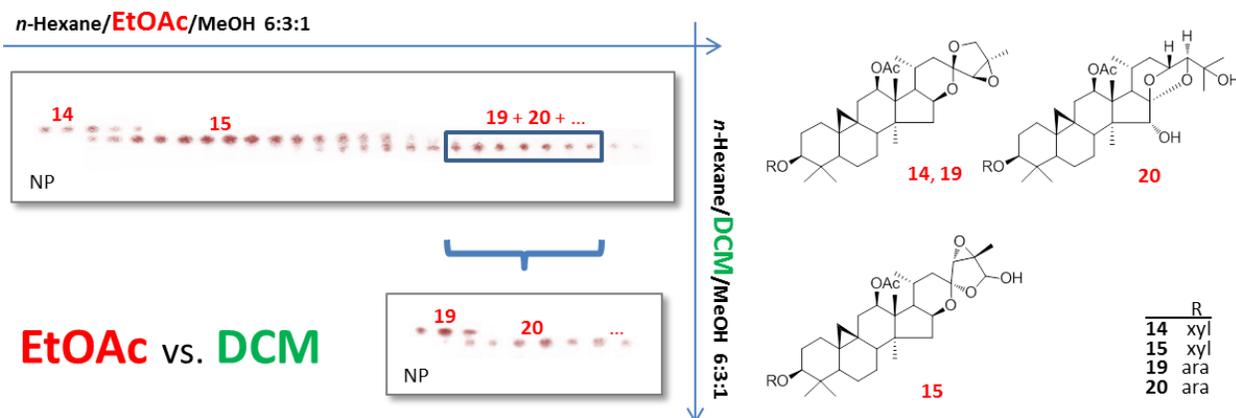
**Figure 13. Resolving Three Congeners from Hops by HSCCC with Orthogonal Ss**

Panel A: By implementing the sample-cutting concept, 6-PN was efficiently concentrated by removing the major interfering component, XN, in the first two steps of HSCCC using HEMWat 0. Panel B: TLC analysis of the 2<sup>o</sup> fractions suggested 6-XN was well separated only with a minor overlap with XN. However, in Panel C, NMR analysis revealed the previously considered pure fractions were a mixture of 6-PN and H<sub>2</sub>-XN. In the 3<sup>rd</sup> step of HSCCC, this critical pair was completely resolved using an orthogonal SS, HChMWat 1:3:3:2. This possibly resulted from the selective formation of an intermolecular hydrogen bond between CHCl<sub>3</sub> and the carbonyl group of H<sub>2</sub>-XN.

Apparently, these two compounds could not be resolved but co-eluted by the EtOAc-based SS.  $\text{CHCl}_3$  was considered as an alternative solution but none of SSs in the well-known family of ChMWat matched the polarity of these two compounds. Thus, a new SS HChMWat was developed by replacing EtOAc in HEMWat with  $\text{CHCl}_3$ . The ratio of four components was empirically determined as 1:3:3:2, of which the polarity was considered appropriate for separation of prenylphenols. As expected, the critical pair of 6-PN and  $\text{H}_2$ -XN were fully resolved in a second dimension of RP-HSCCC using HChMWat as SS. 6-PN was successfully isolated with qHNMR purity of 95%. The surprising selectivity of  $\text{CHCl}_3$  for  $\text{H}_2$ -XN likely originated from the selective formation of an intermolecular hydrogen bond between  $\text{CHCl}_3$  and the carbonyl group of  $\text{H}_2$ -XN, an interaction which cannot occur in 6-PN, resulting in  $\text{H}_2$ -XN being more retained than 6-PN in RP mode.

### 3.5.3 Resolution of *Actaea* Triterpenes

In the example of the separation of triterpenoids from black cohosh using orthogonal chromatography systems (see Section 3.4, p. 51), the combined fraction Ar.18.1.3 was resolved by NP-VLC with *n*-hexane–EtOAc–MeOH (12:6:1, v/v), leading to the isolation of actein (**14**) and 23-*epi*-26-deoxyactein (**15**). As indicated by TLC analysis using an EtOAc-based SS (**Figure 14**), an additional compound was separated and eluted after compound **15**. However, when using a DCM-based SS for TLC development, these fractions were shown to be a mixture of more than two major compounds. Thus, they were combined and subjected to an orthogonal separation on NP-VLC eluting with *n*-hexane–DCM–MeOH (12:6:1, v/v). As a result, two compounds were isolated and identified as cimracemoside N (**19**) and 12-acetoxycimigenol arabinoside (**20**).



**Figure 14. Resolving Four Congeners from Black Cohosh by VLC and Orthogonal SSs**

While actein (**14**) and 23-*epi*-26-deoxyactein (**15**) were separated by the EtOAc-based SS, cimracemoside N (**19**) and 12-acetoxycimigenol arabinoside (**20**) could be resolved in the second dimension using a DCM-based SS as indicated by the DCM-based TLC monitoring. This also shows that the spatially very different triterpenes **14/19**, **20**, and **15** can have surprisingly similar chromatographic properties, i.e., that structural differences are not predictive of ease of separation and can still lead to co-elution.

In summary, the examples shown above represent two distinct scenarios of chromatographic separation: (1) The structurally related compounds can be well resolved; (2) However, the structurally different compounds can still co-elute. Polarity is an important contributing factor for the separation performance of many LC techniques, such as those used in the present study. Excluding other effects, any two analytes with different polarity can theoretically be separated. In practice, however, their structural differences can lead to their different selective interactions with stationary and/or mobile phases. As a result, their separation can also be affected by chromatographic conditions which may exhibit different selectivity on the partial structures or particular functional groups or even stereocenters of the analytes. An excellent example is the chiral separation of racemic compounds into their enantiomers. While being only different in stereochemistry, two enantiomers of the same analyte can exhibit different affinity to the single-enantiomer stationary phase, provided that an appropriate chiral

column is used. Therefore, a critical pair of analytes with minor structural differences and almost equal polarity can be well separated by their highly selective interactions with the chromatographic media. However, despite having distinct structures, the analytes can still show similar polarity and, thus, can be poorly resolved using nonselective chromatographic conditions.

Both polarity and selectivity need to be taken into account when selecting optimal separation conditions. The latter is more important for NP congeners with similar chemical properties. Appropriate use of orthogonal SSs can be efficient and effective in highly selective separations of complex NPs. The pair of EtOAc and  $\text{CHCl}_3/\text{CH}_2\text{Cl}_2$  is an excellent example for the development of orthogonal SSs, as has been shown in above three cases. For future reference, a rational approach for using EtOAc and  $\text{CHCl}_3/\text{CH}_2\text{Cl}_2$ -based SSs in chromatographic separation is suggested as follows: EtOAc is the first choice for developing the SS families. In CS, EtOAc-based SSs like HEMWat have a wider coverage of polarity for NPs, and are more stable with higher  $S_f$  values and less loss of stationary phase during the fractionation process. Furthermore, in line with the concept of green chemistry, EtOAc is greener (less toxic) than chlorinated solvents. Subsequently,  $\text{CHCl}_3/\text{CH}_2\text{Cl}_2$  can be used as an alternative when the EtOAc-based SS fails to separate the target compounds. While a variety of organic solvents are used in contemporary research for the preparative separation of NPs, their orthogonal potential in chromatography has not been systematically studied. Owing to the chemical complexity of biological samples, the selection of SSs for their chromatographic separation is mostly empirical and based on trial-and-error. However, by implementing the concept of orthogonality in solvent selections, SS optimization can be more efficient and effective. Additionally, the use of orthogonal SSs has several beneficial consequences. Exhaustive fractionation procedures can be reduced to fewer steps, and, thus, use less materials (e.g., solvents and sorbents) and require less effort. Compounds can also be separated and purified with reduced permanent loss in a simplified fractionation scheme. The

selectivity of commonly used organic solvents has been thoroughly studied and systematically summarized, such as the well-known Snyder's solvent selectivity triangle (Snyder *et al.*, 1974). This provides a solid theoretical foundation for developing orthogonal SSs.

Solvent modification is another useful approach in SSs optimization for improving the resolution of difficult-to-separate compounds. There are several mechanisms of resolution enhancement by mobile phase additives or modifiers. For example, acid or base in SSs can prevent peak tailing by reducing the interaction of ionized compounds with free silanols on silica. DMSO is used in injection to increase the sample solubility in the mobile phase. At the onset of CS of ginkgolides in the current study, it was observed that when trace quantities of DMSO were used to increase sample loading with an HEMWat +3 CS, better resolution of GA and GB was obtained. Upon further investigation it was found that addition of 0.5% DMSO in the SS increased the resolution of GA and GB, thus, allowing separation of these two otherwise almost equipolar lactones. Although the ability of an additional solvent to change partitioning behavior is well known, the effect of such a trace amount of a solvent is without parallel. It is proposed to be a consequence of the effect of DMSO on the intermolecular and/or intramolecular hydrogen bonding of the terpene lactones in solution. Clearly, DMSO has potential as a modifier in CS SSs by acting as a "resolution enhancer" for compounds with similar properties. Other high-boiling solvents such as *N,N*-dimethylformamide (DMF) and *N*-methyl-2-pyrrolidone (NMP) may also have similar capability and these are worth investigation in future studies.

Finally, combining the orthogonal use of both chromatographic methods (e.g., RP vs. NP, VLC vs. CCC) and SSs (e.g., MeOH vs. MeCN, EtOAc-based vs. CHCl<sub>3</sub>/CH<sub>2</sub>Cl<sub>2</sub>-based), termed as "multidimensional orthogonality", can potentially provide much higher resolution by multiplying the selectivity in sequential purifications.

### 3.6 Conclusion

This chapter introduced several innovative concepts and applications for accelerating and targeting separation/isolation of NPs. It has been demonstrated that shake-flask or CS sample cutting is efficient and effective for preparation of enriched samples. The process of sample cutting not only increases the content of target compounds but also removes contamination which may interfere with further chromatographic fractionation and/or spectroscopic analysis. Compared to column chromatography, liquid-liquid partitioning is a convenient high throughput technique with minimum sample loss. More importantly, a large pool of SS candidates provides wide polarity coverage for the separation of NPs, making it a universal method for sample enrichment. The preparation of semi-purified samples enriched with biomarker compounds is of particular use in the chemical and biological standardization of crude plant extracts. One aspect of the chemical complexity of NPs is represented by a large number of congeners which exhibit similar chromatographic and spectroscopic behavior due to their closely related structures. However, the use of orthogonal conditions in chromatography enables an efficient separation of these congeners in a shortened fractionation procedure. The term, orthogonality, refers to different selectivity of both solid and mobile phases in chromatography. Although NP congeners exhibit similar polarity, their minor structural differences may lead to their selective interaction with the solid and/or mobile phase. As a result, their resolution may be significantly enhanced by alternative use of the chromatographic methods and/or SSs with different selectivity. While chromatography often needs empirical optimization, some important factors can be considered for choosing appropriate orthogonal conditions: (1) the acidic or basic properties of target compounds as well as SSs, and (2) the occurrence of intra- or inter-molecular hydrogen bonding.

## 4. EXPANDING NMR UTILITY IN NATURAL PRODUCT CHARACTERIZATION<sup>1</sup>

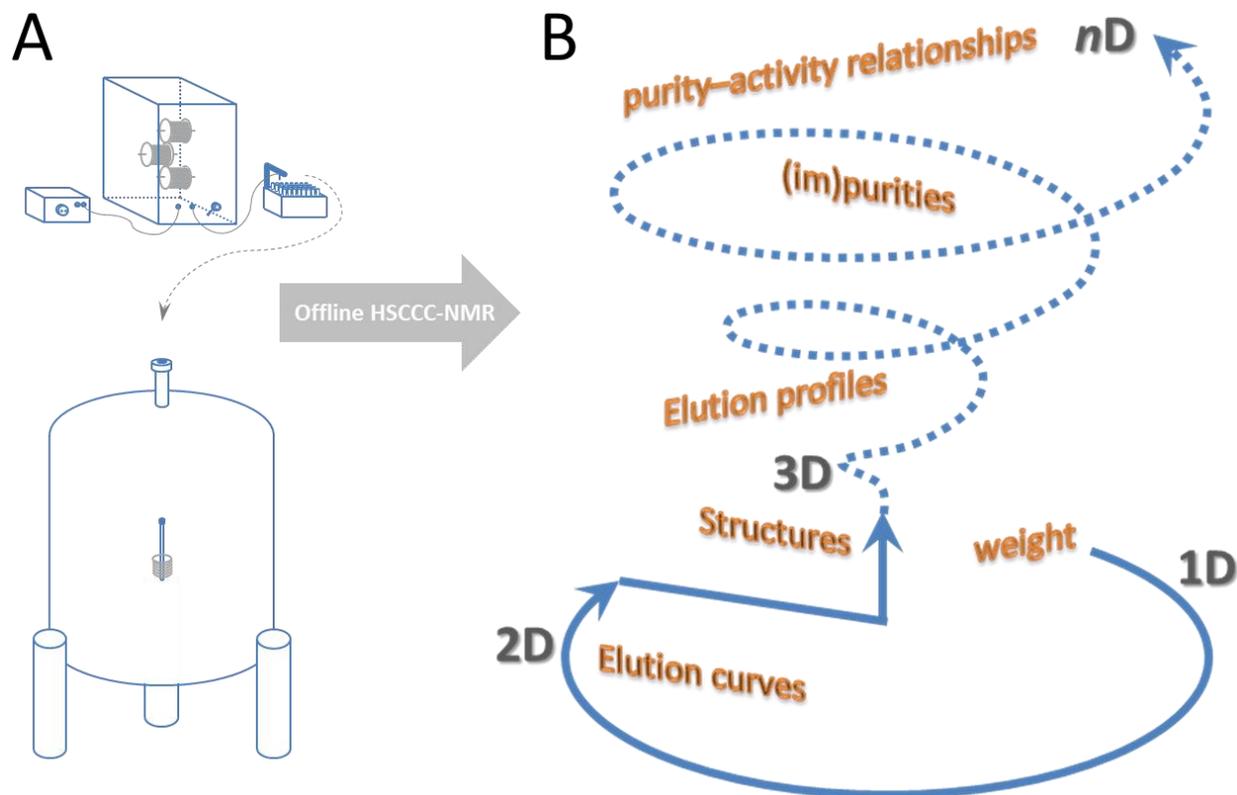
### 4.1 Introduction

In NP research, NMR spectroscopy is frequently used at the final stage of fractionation for structural elucidation of pure compounds. However, the capacity of NMR techniques reaches far beyond this function in the analysis of NPs. This chapter seeks to show how to expand the utility of NMR for resolution of the NP complexity and expedite NP discovery. The following aspects of extensive applications of NMR are introduced and exemplified.

**Elution Detector.** Much like UV, ELSD or MS, NMR may also be used as an elution detector for monitoring the fractionation process. However, the tandem of chromatography and NMR provides information well beyond elution curves generated by UV, RI, or ELSD detectors. This combination of two methods enables the assessment of chemical compositions in chromatographic fractions by providing valuable structural information. Currently, the hyphenation of LC to NMR includes three online modes of techniques: (1) on-flow (continuous flow); (2) stopped-flow; and (3) loop-storage (Exarchou *et al.*, 2005). The on-flow mode allows a rapid screening with <sup>1</sup>H NMR of a mixture, but only for the more intense signals of the major constituents. In both stopped-flow and loop storage modes, the analytes can be examined with more time-intensive 1D and 2D NMR experiments. However, frequent stops of flow may have negative impact on the quality of separation. This is apparently not suitable for the analysis of complex NP mixtures. The loop-storage mode may still be associated with the limitation arising from the low concentration of analytes. In addition, all these online techniques require the use of costly deuterated solvents in chromatographic separation. For CS, this becomes a problem as the SS usually consists of multiple organic solvents. Thus, the use of NMR as an offline detector may be more advantageous when analyzing NPs by chromatography.

---

<sup>1</sup> Content presented in Section 4.2.2 has been partially published in: Qiu, F.; Friesen, J. B.; McAlpine, J. B.; Pauli, G. F. Design of Countercurrent Separation of Ginkgo biloba Terpene Lactones by Nuclear Magnetic Resonance, *Journal of Chromatography A* 2012, 1242, 26–34



**Figure 15. Application of NMR as an Offline Elution Detector**

When coupled offline with chromatography (Panel A), NMR expands the traditional elution detection to multiple dimensions which altogether can qualitatively and quantitatively visualize the elution profiles of target compounds and even minor impurities (Panel B). Unveiling the full map of separation procedures facilitates the evaluation of residual complexity as well as the establishment of purity–activity relationships (PARs).

The offline use of NMR could provide the researchers with more ease and freedom of running and optimizing the NMR experiments (**Figure 15**). Without the limitations of mechanical interfaces, the NMR can be offline coupled to any type of chromatographic systems. After the separation, the diluted sample can be concentrated and analyzed in an appropriate type and volume of NMR solvent. Thus, the fractions can be readily performed under quantitative conditions. Combining the qHNMR variant of NMR analysis with Gaussian curve fitting enables

the full quantitative representation of elution profiles of any NMR-detectable analytes, regardless of whether it has been identified or not. This provides unique information about the composition of eluates/fractions, which can be used to study the presence of minor constituents and analytes with overlapping elution/extrusion behavior in CS, i.e., providing information about the residual complexity of the resulting fractions. This information can be further correlated with bioactivity, leading to the establishment of purity–activity relationships (PARs) (Jaki *et al.*, 2008).

**Quality Controller.** Quality control of raw materials is an important aspect in standardization of dietary supplements. LC-MS and LC-UV are widely accepted methods for qualitative and quantitative evaluation of crude plant extracts. Both methods require authentic standards, and are unsuitable for compounds that exhibit poor ionization or lack a UV chromophore. However, NMR is not limited by these factors. The integral of  $^1\text{H}$  NMR signals exhibit a linear relationship with the analyte concentration under quantitative conditions (Pauli *et al.*, 2005). Therefore, qHNMR is a viable alternative method for the standardization of crude plant extracts and preparations. It is also important to note that chemical complexity of samples typically results in significant peak overlap in the  $^1\text{H}$  NMR spectra which may hinder the identification of particular compounds. However, with the aid of *K*-based CS, crude extracts can be transformed into highly enriched fractions in which the concentration of target analytes is higher and less interference with other constituents is observed, rendering these enriched fractions more suitable for qHNMR analysis.

**Resolution Enhancer.** The resolution of the complexity of NPs not only refers to the identification of multiple components in mixtures but also has to consider various minor impurities present in repeatedly purified materials. Often, due to signal overlap, chemical shift dispersion of 1D NMR is insufficient to resolve these complexities of NPs. Instead, more powerful 2D NMR techniques aid mixture analysis and identification of molecules through their spin–spin couplings, whereas the NMR signals are dispersed into an additional frequency dimension and give rise to unique cross-peak patterns for each molecule present (see below).

In particular, by taking advantage of the enhanced limit of detection of today's cryo-microprobe NMR technologies, acquisition of 2D NMR spectra of microgram samples can be accomplished in a few minutes. Characterization of minor constituents can be achieved at the nanomole level, thus making it a highly efficient tool in the analysis of mass-limited samples and for the evaluation of residual complexity of purified materials.

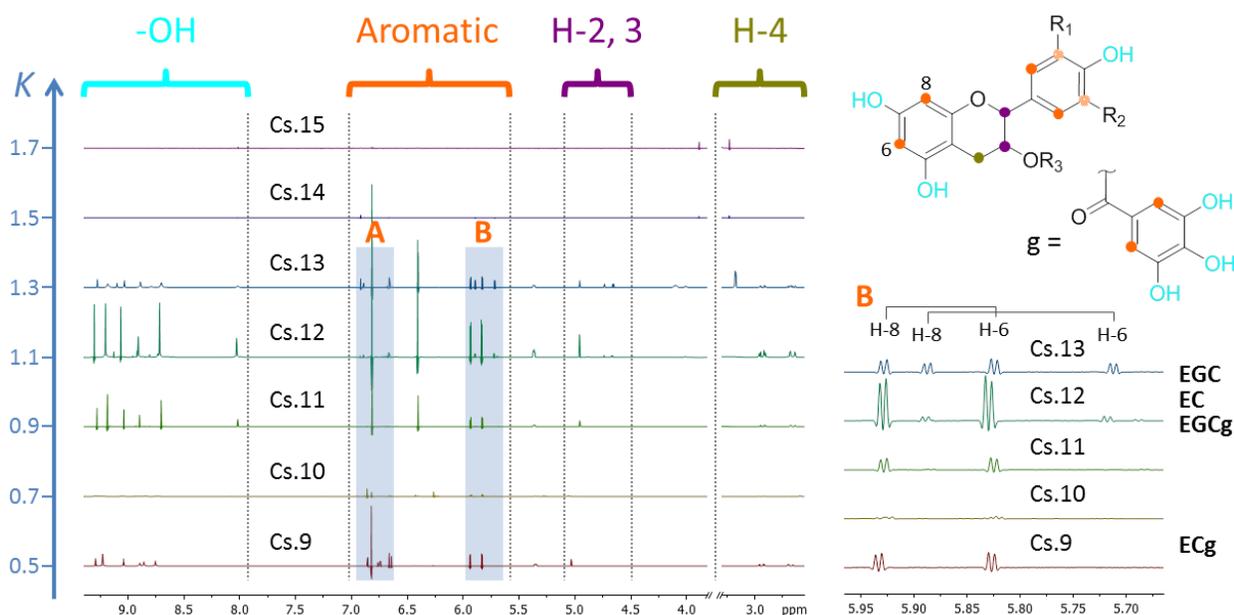
**Target Identifier.** While NMR spectroscopy has been extensively used for structural determination of purified compounds, the use of NMR analysis of NP mixtures has received less attention, likely, due to the challenges associated with severe peak overlap. However, the complex spectra can still be interpreted by implementing the following concepts. **(A) Sequential NMR elution profiles:** For any fractionation process, the elution profiles are represented by subsequent chemical information of fractions. Therefore, NMR can be used as a virtual separation tool for the comparative analysis of spectra of subsequent fractions. This enables extraction of signals of target compounds from complex spectra. **(B) NMR pattern recognition:** The structural characteristics of any single chemical entity (SCE) are represented by a unique pattern of signals in the 1D/2D NMR spectra. Much like biometric recognition, a sub-portion of these complex spectroscopic patterns ("information overflow") might be sufficient to distinguish the different chemical entities. Therefore, NMR pattern recognition is of particular use in structure dereplication of NPs in mixtures, and can facilitate the targeted isolation of compounds of interest, combining both aspects of (A) and (B). This methodology can also be expanded to structural predictions as well as metabolomic profiling of NPs.

## 4.2 Visualization of Elution Profiles

In CS, using  $K$  values measured in a shake-flask experiment, analytes can be readily located in CS fractions. However, the complete individual elution profiles including peak shape (peak width and height) and resolution (peak overlap) cannot be predicted only from the partition behavior of compounds, but also depends on the instrumental conditions such as the number of theoretical plates of the CS column, operating temperature, and pressure. While UV and ELSD are traditionally used in recording elution profiles, the chromatograms provide limited or no structural information. Instead, using NMR detection, it is not only possible to construct the deconvoluted peaks with the aid of Gaussian fitting, but also the full structures of each analyte can be established. More importantly, NMR can visualize UV-inactive molecules and even minor impurities that co-elute with the main isolates. The development of online CS-NMR requires special interfaces and multiple solvent suppression techniques, and has only been demonstrated once with rather limited utility (Spraul *et al.*, 1997). Although further work is needed to overcome the difficulties of CS-NMR hyphenation, the present study has made extensive use of offline qHNMR to perform post-column analysis of the chromatographic fractions.

### 4.2.1 CS of Crude Extract of *Camellia sinensis*

The leaves of *C. sinensis* are widely used as a source of beverages and dietary supplements such as green tea. Green tea concentrated extract contains mainly polyphenols and caffeine, and has been shown to exhibit anti-oxidative activity which is linked to a reduction of cancer risk and heart disease (Graham *et al.*, 1992; Yen *et al.*, 1995; Kahkonen *et al.*, 1999). The chemical characterization and standardization of biomarkers in the crude extracts are important for the clinical evaluation of green tea preparations. In the present study, HSCCC-NMR was used for the selective analysis of catechins in crude green tea extract.



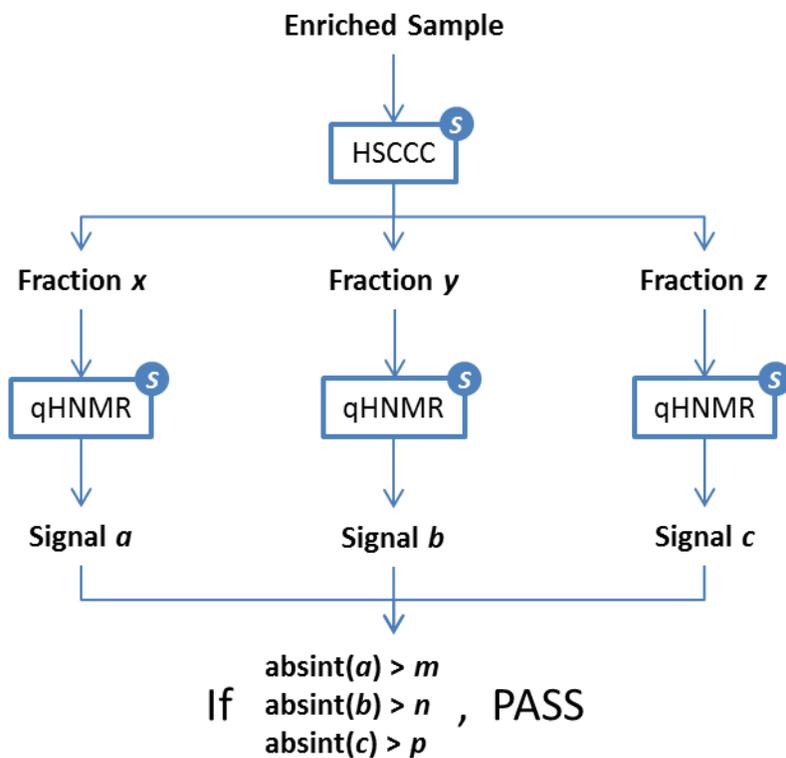
**Figure 16. Offline qHNMR Analysis of HSCCC Fractionation of Green Tea Extract**

Commercial extract of green tea was separated on a 120 mL HSCCC in RP mode using an HEMWat SS (1:7:1:7, v/v), leading to a series of fractions enriched with caffeine and four major catechins. All these biomarkers can be directly analyzed by qHNMR (DMSO- $d_6$ , 400 MHz), making the offline hyphenation of qHNMR and HSCCC suitable for the standardization of these biomarkers in crude extracts.

Commercial green tea extract (50 mg) was initially fractionated on a 20 mL HSCCC in RP mode using an HEMWat SS (1:7:1:7, v/v). Operating at a flow rate of 0.5 mL/min and a rotation speed of 1200 rpm, the stationary-phase retention was 55% when the hydrodynamic equilibrium was established within the column. The eluates were collected at 4 min intervals (2 mL/fraction). Fractions Cs.9 to Cs.15 were dried, dissolved in 500  $\mu$ L DMSO- $d_6$  and subjected to NMR analysis (400 MHz) under identical acquisition and processing parameters. **Figure 16** shows stacked spectra of these seven fractions which can be divided into four fingerprinting regions of catechins: ① <3.50 (H-4); ② 4.50–5.25 (H-2 and H-3); ③ 5.50–7.00 (aromatic protons); and ④ >8.00 (hydroxy groups). Region ③ can be further divided into two subregions: ③-A 5.50–6.00 (H-6 and H-8) and ③-B 6.00–7.00 (H-2', H-3', H-5' and H-6'). Associated with

the elution order, region ③-A is the most differentiating region for the identification of catechins in the fractions. Thus, the four major catechins were readily recognized as epicatechin gallate (ECg, **21**) in Cs.9 to Cs.10, epigallocatechin gallate (EGCg, **22**) in Cs.11 to Cs.13, EC (**23**) in Cs.12, and EGC (**24**) in Cs.13. In addition, caffeine could be identified in Cs.14 and Cs.15.

It is interesting to note that signals of H-6 and H-8 of ECg and EGCg were overlapped at ~5.83 and ~5.93 ppm in the crude extract. However, upon their separation, while still not highly purified yet by the one-step of HSCCC, these signals were also “separated” and thus easily identified and quantified. The same results were observed for the pair of EC and EGC. For a general application, HSCCC can separate a complex biological matrix into fractions with relatively simple and enriched compositions which are more suitable for NMR analysis. Adding to the advantage of *K*-based separation, HSCCC fractionation combined with qHNMR analysis is a potential technique for use in standardization or quality control of plant extracts. For example, **Figure 17** shows a proposed CCC-NMR method for the standardization of plant extracts. A required amount of sample is initially subjected to HSCCC fractionation using standardized conditions (optimized SS and operational parameters including running mode, constant flow rate, and controlled temperature). Based on the retention volume (or *K* value) of individual target compounds, a series of fractions are collected, combined and analyzed by qHNMR under standardized conditions (appropriate pulse program, receiver gain, and number of scans). The quantity of each target compound is evaluated by the integrals of characteristic <sup>1</sup>H NMR signals. It is noteworthy that the HSCCC process is not necessarily optimized to isolate pure compounds but sufficient to separate the complex sample into fractions that yield less crowded NMR spectra, and thus, the target compounds can be more readily identified and quantified such as the abovementioned example of catechins in green tea.



S Standardized process

**Figure 17. The Proposed CCC-NMR Method for Chemical Standardization (Quality Control) of Crude Extracts**

The crude extract or enriched sample is initially fractionated by a standardized HSCCC process. Based on the retention volume of the biomarker compounds, only fractions containing these compounds are collected and subjected to qHNMR analysis under standardized conditions. Using absolute integrals of characteristic signals for the biomarker compounds, the quality of the sample can be evaluated.

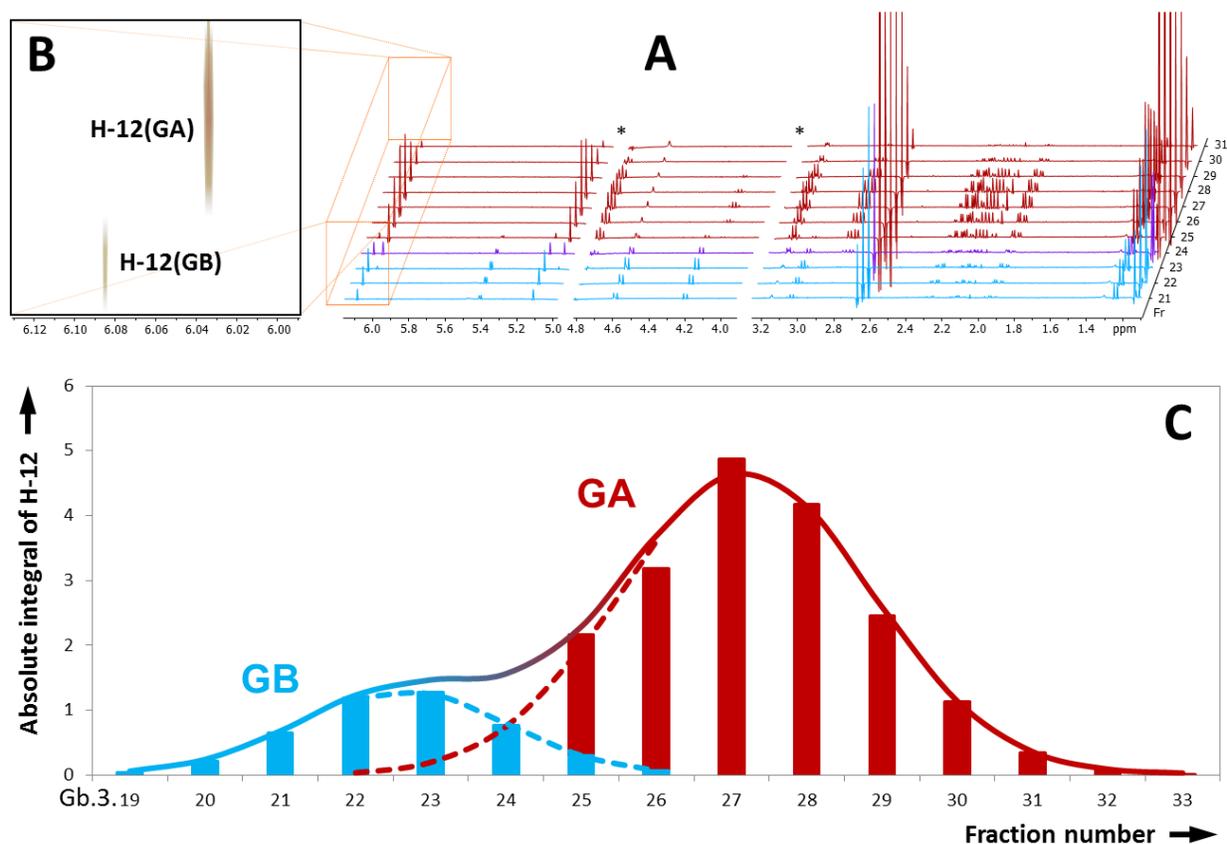
#### 4.2.2 CS of Ginkgo Terpene Lactones

In the above discussion on the separation of ginkgolides using orthogonal SSs (see Section 3.5.1, p. 54), three combined fractions, Gb.3, Gb.4, and Gb.7, were re-fractionated in a second dimension of HSCCC ( $V_c = 20$  mL) in RP mode using HEMSoWat +3/0.5% SSs. As all fractions were analyzed offline by qHNMR under the same experimental conditions, the absolute integral of the  $^1\text{H}$  NMR signal of H-12 had a linear relationship with the concentration of each terpene lactone in the sample. By using this absolute integral as a measure of the relative molar mass of each compound, the elution pattern of GA and GB could be reconstructed as shown in **Figure 18**. Theoretically, elution in HSCCC follows a Gaussian distribution as long as it does not exhibit sample solubility or overloading problems. Accordingly, the CCC chromatographic peaks of GA and GB could be simulated by Gaussian curve fitting following established mathematical relationships (Walsh *et al.*, 1995; Di Marco *et al.*, 2001) as follows:

$$y_{\text{GA}} = 4.6533e^{-\left(\frac{n-27.19}{2.3604}\right)^2} \quad (R^2 = 0.995) \quad (3)$$

$$y_{\text{GB}} = 1.3014e^{-\left(\frac{n-22.62}{1.9700}\right)^2} \quad (R^2 = 0.997) \quad (4)$$

where  $n$  is the fraction number and  $y$  is the absolute integral of H-12. According to these simulated chromatographic peaks, the achieved experimental resolution ( $R_s$ ) of GA and GB was approximated as follows.



**Figure 18. Offline qHNMR Analysis of the 2nd Step of HSCCC Purification of GA and GB**

Panel A shows the stacked plots of the  $^1\text{H}$  NMR spectra of 11 fractions Gb.3.21 to Gb.3.31 (500  $\mu\text{L}$  methanol- $d_4$ , 400 MHz). Panel B shows a contour plot of the 5.990–6.130 ppm range of these spectra. Using the absolute integral of H-12 as a measure of the relative molar mass of the terpene lactones in each fraction, the individual terpene lactone content of the 11 fractions are shown in Panel C. In addition, the corresponding elution curves were reconstructed by Gaussian fitting. Asterisks denote the omitted residuals solvent signals (3.320 and 4.860 ppm).

For retention time-based chromatography like GC or HPLC, resolution ( $R_s$ ) is defined as:

$$R_s = \frac{2(t_{R2} - t_{R1})}{w_1 + w_2} \quad (5)$$

In analogy, for retention volume-based HSCCC,  $R_s$  can be calculated using the equation:

$$R_s = \frac{2(V_{R2} - V_{R1})}{w_1 + w_2} \quad (6)$$

In **Figure 18C**, it is shown that  $V_R(\text{GA}) \approx 27.2$  mL,  $w(\text{GA}) \approx 6.7$  mL,  $V_R(\text{GB}) \approx 22.6$  mL, and  $w(\text{GB}) \approx 5.6$  mL. Thus, the resolution between GA and GB was  $\sim 0.74$ . In analytical chromatography, the emphasis is on the theoretical plates, while the critical parameters for preparative chromatography are the amount of compound produced per unit time, product purity, recovery, and separation cost (Cazes *et al.*, 2005). Although  $R_s$  of GA and GB (0.74) was less than 1.0, which is considered insufficient resolution in analytical chromatography, recovery of GB was remarkably high at  $>75\%$  with purity of  $>95\%$  if only the three fractions Gb.3.21, Gb.3.22 and Gb.3.23 had been combined. This exemplifies the great preparative efficiency of CS.

By using Equation (1), the  $K$  values of GA and GB could also be calculated more precisely as 1.45 and 1.16, respectively. In the shake-flask experiment (see Section 3.3, p. 46), the  $K$  values of these two ginkgolides were measured as 1.46 and 1.18. The remarkable consistency between the shake-flask and actual HSCCC experiments (1.45 vs. 1.46,  $\Delta = -0.7\%$ ; 1.16 vs. 1.18,  $\Delta = -1.7\%$ ) demonstrates that the measurement of  $K$  values by NMR ( $K$ -by-NMR) can efficiently guide targeted isolation by rapid tracing of the compounds of interest in a series of CS fractions.

Owing to the limit of detection (LOD) in qHNMR, a compound at the beginning or ending of elution which is present at a very low concentration levels may not be detected. As a result, it can be difficult to determine if two compounds exhibit overlap and which fractions should be combined for the best balance of recovery and purity. However, this problem can be solved when considering that the chromatographic peaks generally follow a Gaussian distribution. Accordingly, the Gaussian distribution acts as a constraint for a fitting procedure, in which the breadth of chromatographic peaks can be predicted using the values of more concentrated fractions in the peak center. To establish this method, the present study used the elution of GA and GB as an example. The relative quantity of GB in fractions before Gb.3.21 and after Gb.3.24 was predicted to be below the LOD under the chosen conditions of the NMR

experiment. In order to facilitate a prediction, the chromatographic peak of GB was reconstructed by Gaussian curve fitting using the relative quantities of the four major fractions Gb.3.21 to Gb.3.24:

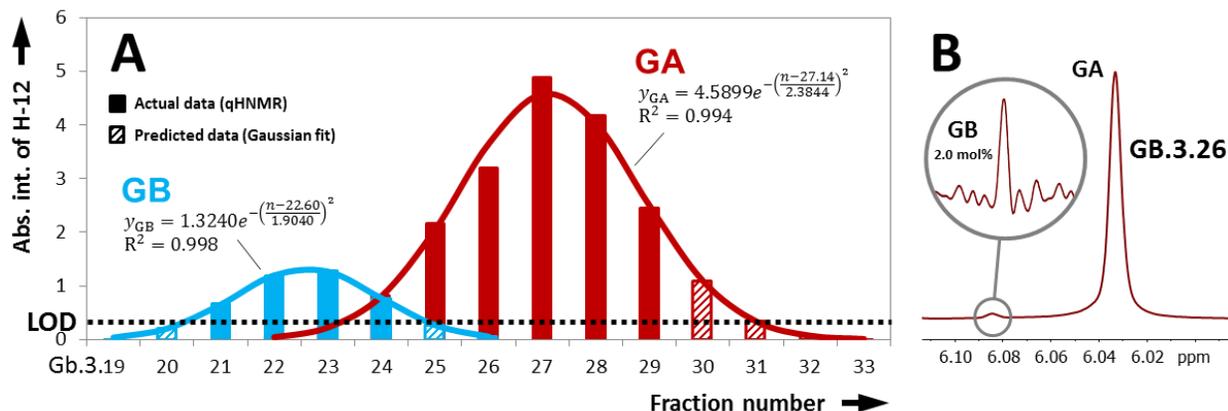
$$y_{GB} = 1.3240e^{-\left(\frac{n-22.60}{1.9040}\right)^2} \quad (R^2 = 0.998) \quad (7)$$

Thus, by using Eq. 7, the relative molar masses of GB in Gb.3.20 and Gb.3.25 were predicted as 0.2061 and 0.2782, respectively. It was also expected that Gb.3.26 contained ca. 1.7 mol% of GB, which was not actually detected by NMR using the standard acquisition parameters (NS = 256). To evaluate the accuracy of this predicted value, Gb.3.26 was re-analyzed by qHNMR with a much increased NS value of 8192 (8k) for 4-fold improved LOD. As expected, GB was confirmed to be present at 2.0 mol% (**Figure 19**). Similarly, the chromatographic peak of GA was best fitted to five fractions Gb.3.25 to Gb.3.29 as a Gaussian equation:

$$y_{GA} = 4.5899e^{-\left(\frac{n-27.14}{2.3844}\right)^2} \quad (R^2 = 0.994) \quad (8)$$

The calculated relative molar masses of GA in Gb.3.30 and Gb.3.31 were 1.0888 and 0.3339, respectively, and thus are close to the values from a high-sensitivity qHNMR experiment 1.0826 ( $\Delta = 0.57\%$ ) and 0.3530 ( $\Delta = -5.4\%$ ), respectively.

This demonstrates that Gaussian curve fitting makes it possible to predict the quantities of compounds of interest even in fractions which contain concentrations below the LOD of the qHNMR experiment. In addition, the breadth of each chromatographic peak as well as the overlapping peak can be visually illustrated. It is therefore possible to determine which fractions may be combined for the best yield with a required level of purity without performing chemical analysis of all fractions.



**Figure 19. Matching and Prediction of Compound Purity by Gaussian Fitting of qHNMR-based Elution Profiles of GA and GB**

Panel A: The elution of GA and GB was each mathematically described as a Gaussian equation. The relative molar mass of these two ginkgolides in the fractions shown in the shaded area was predicted from the equations. Panel B: GB was determined to have a molar content of 2.0 mol% in Gb.3.26 by qHNMR analysis, employing 8k scans (NS = 8192), which matched the predicted value of 1.6 mol% obtained from Gaussian fitting.

Similarly, qHNMR analyses of fractions Gb.4.53 to Gb.4.59 and fractions Gb.4.64 to Gb.4.69 followed by Gaussian curve fitting indicated that GC and GJ eluted following a Gaussian distribution with the following equations:

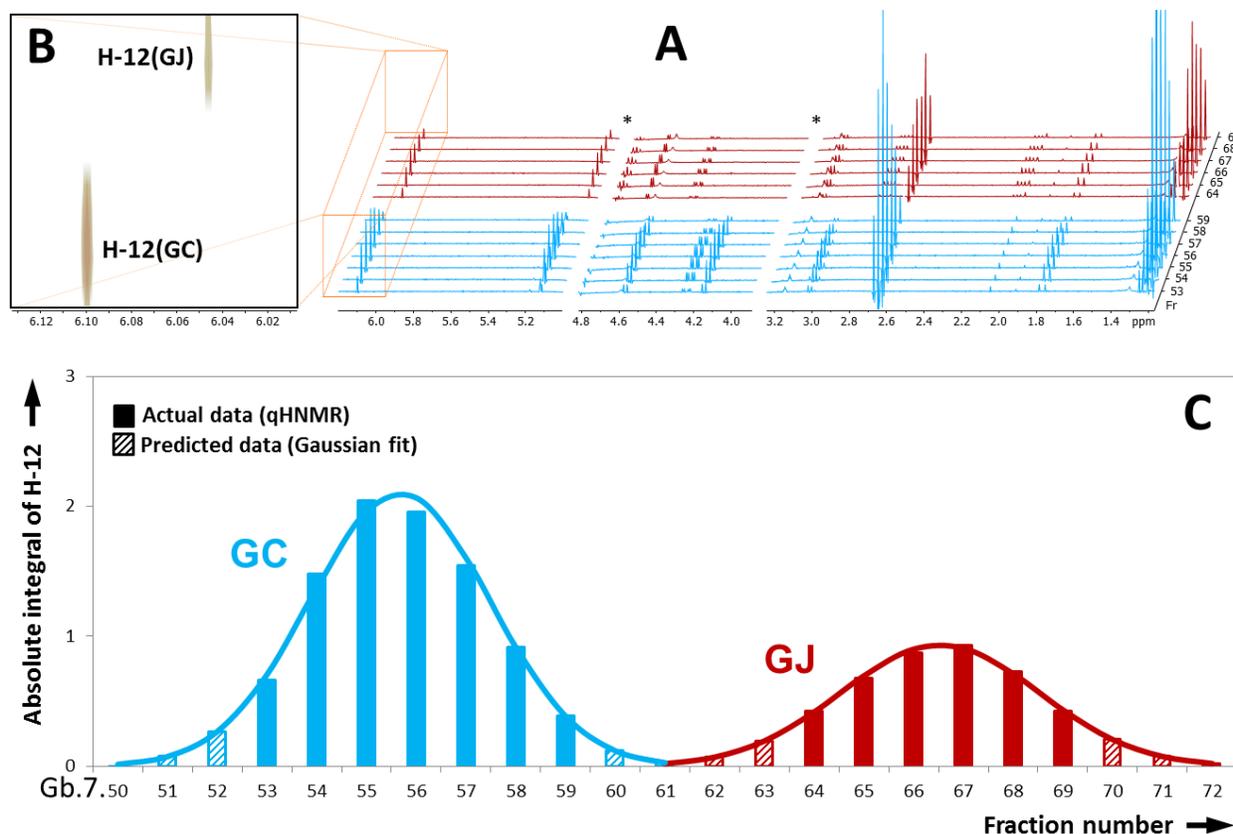
$$y_{GC} = 2.1008e^{-\left(\frac{n-55.67}{2.5586}\right)^2} \quad (R^2 = 0.985) \quad (9)$$

$$y_{GJ} = 0.9341e^{-\left(\frac{n-66.53}{2.8228}\right)^2} \quad (R^2 = 0.991) \quad (10)$$

In **Figure 20C**, it can be observed that GC and GJ were completely separated. Their  $R_s$  was calculated as 1.50 (GC:  $V_R \approx 222.7$  mL and  $w \approx 28.5$  mL; GJ:  $V_R \approx 266.3$  mL and  $w \approx 29.4$  mL) which is considered baseline resolution.

The examples of the two ginkgolide pairs demonstrate that qHNMR cannot only define the chemical compositions in HSCCC fractions but can also be used to generate elution curves such as are otherwise created by UV, RI, or ELSD detection. Using a combination of qHNMR analysis and nonlinear curve fitting, the elution profiles of identified compounds can be

represented quantitatively. Additionally, NMR fraction analysis in conjunction with Gaussian fitting allows for the prediction of the composition of fractions which are outside the LOD/LOQ window of the NMR detection experiment.



**Figure 20. Offline qHNMR Analysis of the 2nd Step of HSCCC Purification of GC and GJ**

Panel A shows the stacked plots of the  $^1\text{H}$  NMR spectra of 13 fractions Gb.7.53 to Gb.7.59 and Gb.7.64 to Gb.7.69 (500  $\mu\text{L}$  methanol- $d_4$ , 400 MHz). Panel B shows a contour plot of the 6.099–6.130 ppm range of these spectra. Using the absolute integral of H-12 as a measure of the relative molar mass of the terpene lactones in each fraction, the individual terpene lactone content of the 13 fractions are shown in Panel C. The relative molar masses of the ginkgolides contained in the remaining seven fractions are shown as shaded bars and was predicted from Gaussian curve fitting.

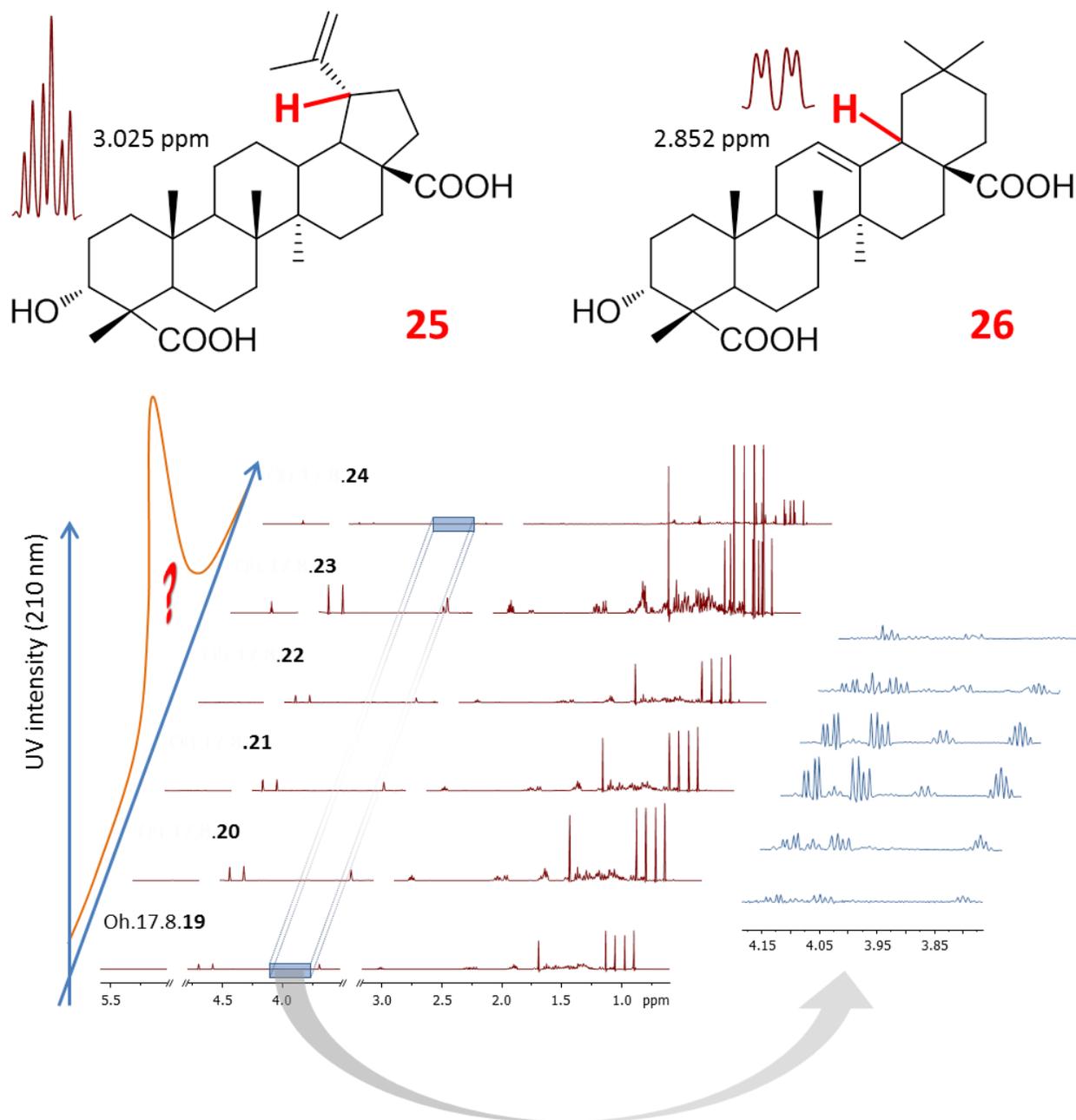
### 4.2.3 HPLC Purification of Triterpenes from *Oplopanax horridus*

While the majority of NP purification efforts focus on the identification of purified compounds, co-eluting impurities (<10%) are frequently overlooked, particularly as long as they do not interfere with the spectroscopic analysis of the main constituent. However, considering that the impurities may exert significant influence on the accurate biological assessment (Jaki *et al.*, 2008), it is critically necessary to evaluate the residual complexity of purified NPs. As presented in the previous two examples of critical ginkgolide pairs, offline qHNMR was also used to identify the minor impurities in subsequent chromatographic fractions. The visualization of co-eluting impurities not only reveals their elution profiles, but also provides qualitative and quantitative information for the development of the purity–activity relationships (see Section 6.1, p. 135).

Devil's club is an ethnobotanical which has been used by Native Americans to treat diabetes and a variety of tumors (Tai *et al.*, 2006). Recent studies showed that its crude extracts inhibit *M. tuberculosis* (Inui *et al.*, 2007). BGF led to the isolation of active polyynes and sesquiterpenes (Kobaisy *et al.*, 1997; Inui *et al.*, 2010). In a continuation of efforts to identify anti-TB active constituents from Devil's club, an exhaustive fractionation procedure using NP-VLC was carried out for the *n*-hexane and DCM partitions of the crude extract. Eluting with a stepwise gradient of *n*-hexane–EtOAc and EtOAc–MeOH, a triterpene-enriched fraction (TEF) was obtained after two steps of VLC fractionation. A 100 mg sample of TEF was subjected to HPLC purification using an isocratic elution with 90% aq. MeOH. TLC analysis of the fractions, Oh.17.8.19 to Oh.17.8.24, showed that they all gave a single spot with the same  $R_f$  value. However, their  $^1\text{H}$  NMR analyses indicated that Oh.17.8.19 to Oh.17.8.22 contained a known triterpene, 3 $\alpha$ -hydroxylup-20(29)-ene-23,28-dioic acid (**25**), while Oh.17.8.23 and Oh.17.8.24 were a mixture of **25** and its isomer, 3 $\alpha$ -hydroxyolean-12-ene-23,28-dioic acid (**26**). The ratio of these two isomeric triterpenes in Oh.17.8.23 and Oh.17.8.24 was measured as 7:3 and 3:7, respectively, by using their compound-specific marker signals: a doublet of triplets (dt) at  $\delta_{\text{H}}$

3.025 with  $J = 4.8$  and  $10.8$  Hz (H-18) for **25**, and a doublet of doublets (dd) at  $\delta_{\text{H}}$  2.852 with  $J = 4.2$  and  $13.8$  Hz (H-17) for **26**.

An expanded view of the range 3.50–4.50 ppm in the spectra revealed an additional group of minor signals which were assumed to belong to glyceride-type impurities based on their chemical shifts and splitting patterns. A stack of six spectra (**Figure 21**) shows that these impurities were present in all six fractions with a molar content of ca. 0.5–8.0 mol%. Additional weak signals were observed in the 5.30–5.70 ppm region which indicated further residual complexity of the purified triterpenes. In this example, offline qHNMR analysis expanded the visualization of elution profiles to multiple dimensions in which both major compounds and minor impurities can be qualitatively and quantitatively described. While 1D  $^1\text{H}$  NMR is often insufficient to fully characterize all the components, 2D NMR techniques are available to advance the analysis of complex mixtures as well as to evaluate the residual complexity, as will be shown in the following.



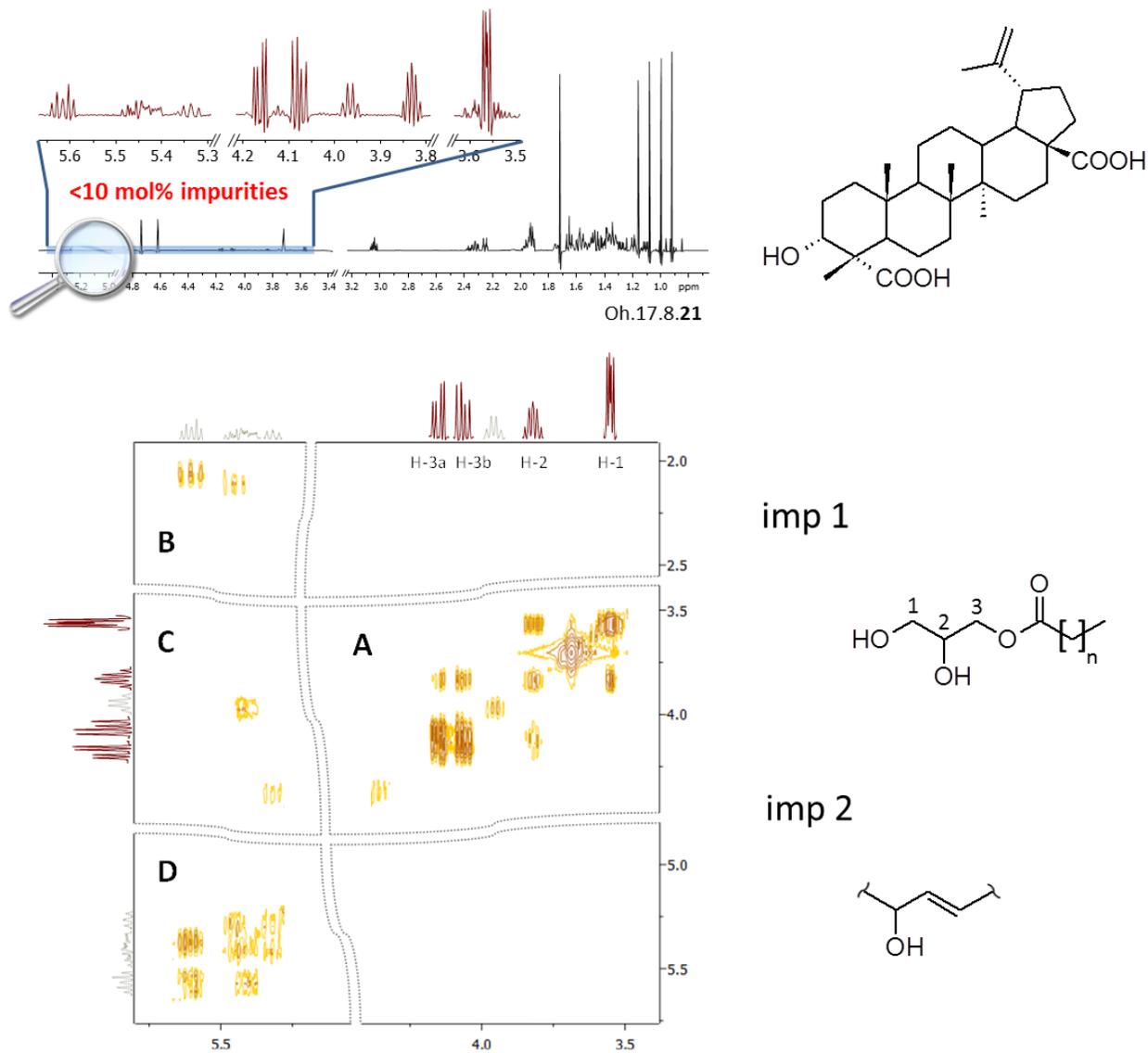
**Figure 21. Offline qHNMR Analysis of the Preparative HPLC Purification of Triterpenes from *O. horridus***

A known triterpene (**25**) eluted in fractions Oh.17.8.19 to Oh.17.8.24 while co-eluting with its isomer (**26**) in Oh.17.8.23 and Oh.17.8.24. These two compounds were distinguished by their compound-specific marker signals: dt  $\delta_{\text{H}}$  3.052 (H-18, **25**) and dd  $\delta_{\text{H}}$  2.852 (H-17, **26**) (500  $\mu\text{L}$  methanol- $d_4$ , 600 MHz). At the same time, all six fractions contained ca. 0.5–7.5 mol% of glycerides that were identified by the characteristic signals in  $\delta_{\text{H}}$  3.50–4.50 and further increased the residual complexity of the fractions.

### 4.3 Evaluation of Residual Complexity of Triterpenes from *Oplopanax horridus*

The homonuclear correlation spectroscopy (COSY) is the mother of all 2D NMR methods and is widely used to identify spins of protons which are coupled to each other. In a COSY spectrum, the conventional 1D  $^1\text{H}$  NMR spectrum of chemical shift appears along a diagonal ridge running from the lower left to upper right corner of the plot. The important information acquired from a COSY spectrum is the off-diagonal cross peaks, which are produced by spin coupling of protons separated by three chemical bonds in the topology  $\text{H}-\text{C}-\text{C}-\text{H}$  ( $^3J_{\text{HH}}$ ), i.e., the protons are “adjacent” and their spin coupling is transmitted through the intervening carbon–carbon bond. To date, COSY is still among the most popular forms of 2D NMR spectroscopy and is routinely performed as a highly useful method for identifying different molecules through their unique cross-peak patterns (Xi *et al.*, 2006).

In order to verify the presence of minor glycerides in purified *Oplopanax* triterpenes (see Section 4.2.3, p. 79), COSY was used to determine whether there were the typical  $^1\text{H}-^1\text{H}$  correlations that can be expected for such signals observed in the range 3.50–4.50 ppm. Fraction Oh.17.8.21 which contained 6.1% glyceride impurities ( $\sim 130 \mu\text{g}$ ) was selected for a COSY experiment. Using a 400 MHz NMR spectrometer, the time-domain data size (TD) was set to 2048K/256K (F2/F1) and the number of scans (NS) increased to 64, for an improved LOD for such minor quantity of impurities. As expected, the acquired COSY spectrum clearly showed four  $^1\text{H}-^1\text{H}$  correlations which reveal the spin system of H-1 to H-3 in a monoglyceride (imp 1, **27**) (**Figure 22A**). As the methylene signals of the fatty acid chain (except for the triplet of 2H at 2.372 ppm) are severely overlapped with those of the methylenes of the triterpenes in the crowded region of 0.80–2.00 ppm, it was difficult to determine the length of fatty acid chain in both the 1D and the 2D spectra. However, it at least became clear that the aliphatic chain is free of double bonds as no olefinic signals were observed in the region 4.50–6.00 ppm.



**Figure 22. Impurity Profiling of an *O. horridus* Triterpene Sample (Oh.17.8.21) by COSY Analysis**

A monoglyceride was identified by its characteristic  $^1\text{H}$ - $^1\text{H}$  correlation pattern shown in region A of the COSY spectrum (methanol- $d_4$ , 400 MHz). The observed correlations in regions B, C, and D indicate two minor impurities both of which possess an allylic hydroxyl group.

In addition, the COSY spectrum also revealed two minor impurities (imp 2, **28a/b**) both of which have an allylic hydroxyl group in the molecule. These results were deduced from a correlation of two protons in the 5.30–5.70 ppm region which indicates a double bond (**Figure 22B**). Furthermore, one of these two protons further correlates with a proton at 3.50–4.50 ppm that is consistent with structure –O–C–H (**Figure 22C**), while the other one correlates with two protons at ~2.00 ppm which are consistent with an allylic group (**Figure 22D**). The total content of these two impurities was determined as 5.9 mol%.

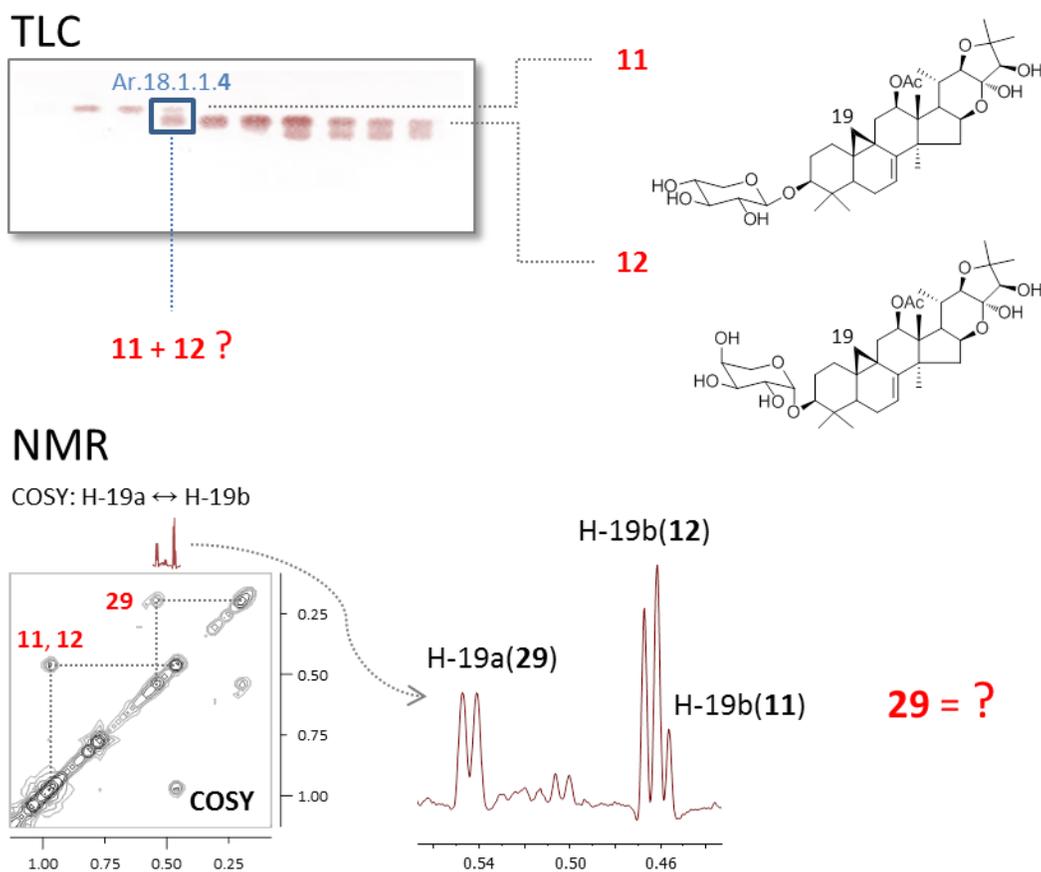
With a qHNMR purity of 85.8%, the purified triterpene **25** contained one structurally related impurity (2.2%) and at least three structurally unrelated impurities (12.0%), which exemplifies the importance of residual complexity of purified NPs. In our experience, even repeatedly purified reference materials of biosynthetically diverse NPs such as triterpenes often exhibit surprisingly high degrees of residual complexity. In addition, the observed co-occurrence of considerably different chemical species is of importance regarding bioactivity. Therefore, the evaluation of the degree and pattern of residual complexity of purified NPs should be considered a prerequisite for their biological assessment. The example of the *Oplopanax* triterpene also shows that 1D and 2D NMR are useful tools for qualitative and quantitative evaluation of the residual complexity. In addition to classic COSY experiments, the modern heteronuclear correlation experiments such as HSQC and HMBC are also powerful techniques in identifying the complex mixtures of NPs, as shown in the following sections 4.4 and 4.5.

#### 4.4 Identification of a New *Actaea* Triterpene in a Residually Complex Mixture

Traditionally, NMR analysis of complex mixtures is considered a complicated process as the severe peak overlap can significantly hinder the identification of the components. Therefore, structure elucidation of compounds in NP discovery usually occurs at the final stage of separation, i.e., the isolation of “pure” compounds. This classical approach for the identification of new compounds can be inefficient as specific structural information remains unconsidered throughout most of the fractionation pathway, and this feature contributes significantly to the common isolation of previously discovered or “uninteresting” compounds. The present section develops a new concept of 2D NMR subtraction which extracts useful information from the complex NMR spectra of sequential chromatographic fractions. Beyond the capability of identifying the compounds in the mixtures, NMR can be used as a virtual separation tool for resolving those compounds that are present in minor quantity or are difficult to separate by chromatography.

As mentioned in the discussion of chromatographic orthogonality in Section 3.4 (p. 51), a combined fraction (Ar.18.1.1) was re-fractionated by NP-VLC using an EtOAc-based SS, which yielded 10 subfractions Ar.18.1.1.1 to Ar.18.1.1.10. In **Figure 23**, TLC analysis of these fractions clearly shows the elution order of three major compounds. According to NMR analysis of selected fractions, Ar.18.1.1.2 and Ar.18.1.1.6 were determined to contain cimracemoside F (**11**) and cimracemoside G (**12**), respectively. Thus, Ar.18.1.1.4 was supposedly a mixture of **11** and **12**. However, the result of  $^1\text{H}$  NMR analysis was inconsistent with this assumption based on TLC analysis. When excluding the signals of **11** and **12**, a residual pair of H-19 signals was observed at  $\delta_{\text{H}}$  0.196 and 0.544 which indicated the presence of another cycloartane triterpene (**29**). Owing to the limited quantity of the sample (0.4 mg), further separation was rendered impractical. However, considering that the nearby fractions Ar.18.1.1.2 and Ar.18.1.1.6 represent the known triterpenoids **11** and **12**, respectively, compound **29** could be separated virtually by subtracting the spectra of these two fractions from the spectrum of Ar.18.1.1.4. Thus the

structural elucidation of **29** became possible without the need for further chromatographic separation.

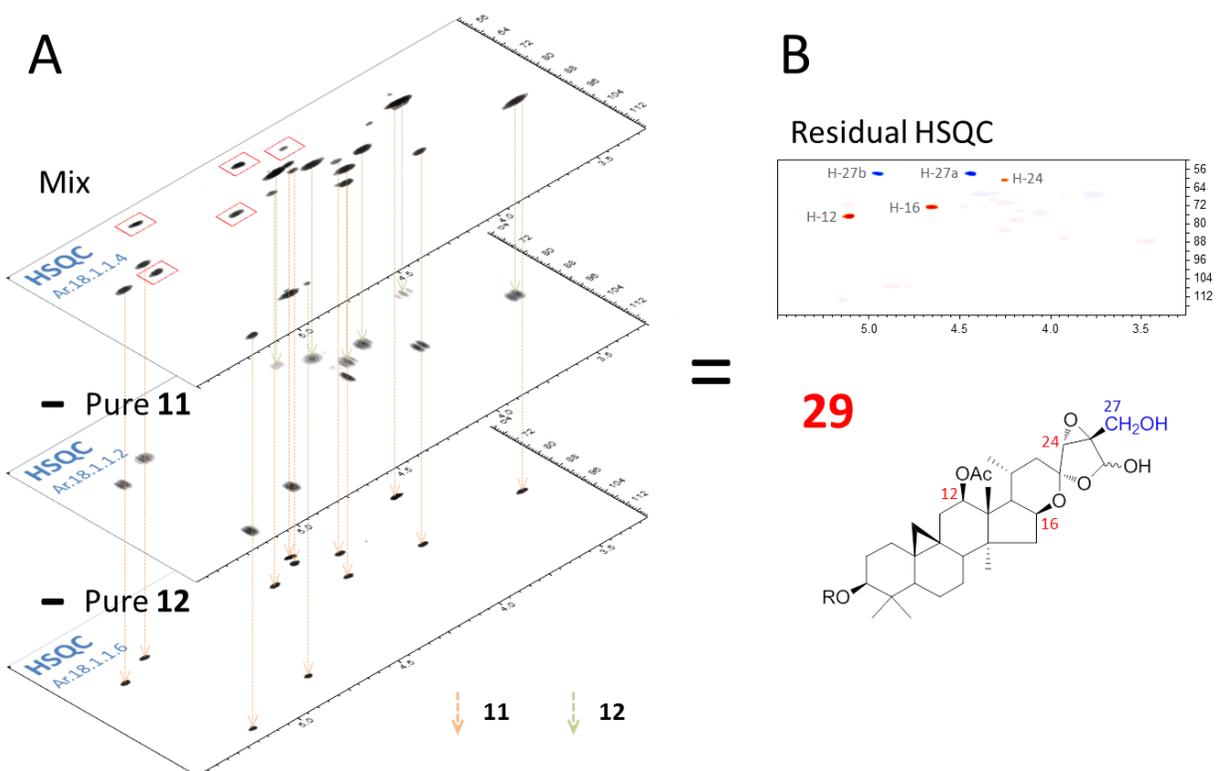


**Figure 23. Discovery of a “Hidden” Component by Inconsistencies between TLC and  $^1\text{H}$  NMR Profiles**

While TLC shows the fraction Ar.18.1.1.4 as a mixture of two compounds **11** and **12**, the  $^1\text{H}$  NMR signals of H-19 reveal an additional cycloartane triterpene (**29**). The unusual structural characteristics of **29** in the NMR spectrum triggered interest in the further elucidation of its structure by 2D NMR.

Apparently, selection of appropriate NMR methods is important for turning this concept into real application. While  $^1\text{H}$  NMR spectra can be acquired rapidly, severe peak overlap causes insufficient resolution of the spectra and, thus, makes accurate spectral subtraction difficult. Similar problems exist for COSY. However, 2D heteronuclear spectroscopy can provide increased resolution and enhanced shift dispersion and, therefore, provides improved metabolite specificity by utilizing the greater  $^{13}\text{C}$  chemical shift dispersion on one axis of the 2D spectrum. Comparing commonly used HMBC and HSQC spectra, the latter is much less complex by providing only  $^1J_{\text{C,H}}$  correlations. In addition, it is a more sensitive experiment compared to  $^1\text{H}$ - $^{13}\text{C}$  experiments and, thus, is less demanding in terms of data acquisition time. The cross peaks in 2D spectra are generally more visually distinguishable than the 1D peaks, making the 2D spectra easier to compare. In summary, HSQC was considered the most appropriate method for performing 2D NMR spectral subtraction.

The HSQC spectra of these three fractions Ar.18.1.1.2, Ar.18.1.1.4 and Ar.18.1.1.6 were acquired on a 700 MHz cryo-microprobe NMR spectrometer under identical experimental conditions: Solvent 35  $\mu\text{L}$  pyridine- $d_5$  (99.96 %D), TD 1024K/256K (F2/F1), and NS 2. The residual solvent peak of pyridine- $d_5$  was calibrated to  $\delta_{\text{H}}$  7.22 ( $^1\text{H}$ ) and  $\delta_{\text{C}}$  123.9 ( $^{13}\text{C}$ ) as reference signals to align the spectra for accurate spectra subtraction. It is well known that the major structural differences of the cycloartane-type triterpenoids in black cohosh arise from modifications of their side chains at C-17 which are often biosynthetically cyclized onto C-16. The observed NMR signals of the side-chain protons are highly indicative of the structures. As the side chains usually contain ether, epoxide, and/or acetoxy groups, their fingerprinting region (e.g.,  $\delta_{\text{H}}$  3.00–5.50,  $\delta_{\text{C}}$  50.0–100.0) in the HSQC spectra are appropriate for spectral comparison.



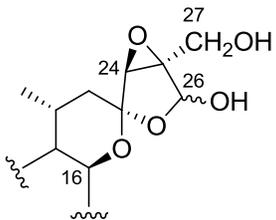
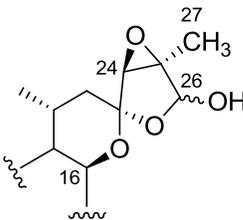
**Figure 24. Identification of a New Triterpenoid (29) as a Minor Component in a Residually Complex Sample by 2D NMR Spectral Subtraction**

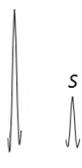
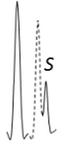
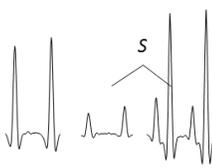
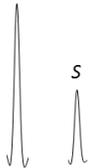
Panel A: Led by a comparative analysis, the characteristic signals of compound **29** were “isolated” from the HSQC spectra of Ar.18.1.1.4 by subtracting the signals of compounds **11** and **12**, which were readily available in the HSQC spectra of the nearby fractions Ar.18.1.1.2 and Ar.18.1.1.6, respectively. Panel B: Further analysis of residual HSQC map unambiguously determined **29** as 27-hydroxyactein, which was likely biosynthesized by oxidation of Me-27 in actein.

**Figure 24** shows the aligned spectra in stack mode with an expansion of the fingerprint region of side chains and sugar moieties. As a result, the cross peaks in Ar.18.1.1.2 and Ar.18.1.1.6 were subtracted from Ar.18.1.1.4, leaving four cross peaks which belong to compound **29**. The characteristic signal at  $\delta_{\text{H}}$  4.660 and  $\delta_{\text{C}}$  72.75 (H-16) indicated that the structure of **29** is similar to actein. However, comparative analysis of  $^1\text{H}$  NMR spectra of **29** and actein showed that one methyl signal ( $\text{CH}_3$ -27) was missing in **29**. The  $^1\text{H}$  NMR signals of H-24 at  $\delta_{\text{H}}$  3.948/3.803 (26*R*/26*S*) in actein are shifted downfield to  $\delta_{\text{H}}$  4.253/4.153 in **29** (**Table 4**).

Moreover, H-26 in actein is shifted downfield from  $\delta_{\text{H}}$  5.771/5.748 to  $\delta_{\text{H}}$  6.144/6.189 in **29**. In contrast, in the HSQC spectrum of **29**, one pair of geminal protons was observed at  $\delta_{\text{H}}$  4.437/4.949 and  $\delta_{\text{C}}$  57.78, which resonated as two doublets ( $J$  12.6 Hz) in the  $^1\text{H}$  NMR spectrum. These data suggest that the 27- $\text{CH}_3$  in **29** had been oxidized to a  $-\text{CH}_2\text{OH}$ . The HMBC correlation between H-27a and C-26 ( $\delta_{\text{C}}$  96.57) further confirmed that the structure of **29** should be 27-hydroxyactein. As **11** and **12** are known glycosides, it was readily determined that **29** is a xyloside based on the intensities and locations of the xylose sugar signals in the HSQC spectrum. It is noteworthy that the content of **29** was determined to be only ~20 mol%, i.e., ~80  $\mu\text{g}$  in the 0.41 mg sample. This demonstrates the power of the micro-cryo NMR probe in the identification of compounds in residually complex and mass-limited samples. In addition to **11**, **12**, and **29** as major components, Ar.18.1.1.4 exhibited large residual complexity, containing more than 10 minor (<10 mol%) cycloartane triterpenoids according to the observed H-19 signals (**Figure 25**).

TABLE IV.  $^1\text{H}$  AND  $^{13}\text{C}$  NMR COMPARISON BETWEEN 27-HYDROXYACTEIN (**29**) AND ACTEIN (**15**)<sup>a</sup>

Position	27-Hydroxyacteïn ( <b>29</b> )			Actein ( <b>15</b> )		
	$^1\text{H}$ NMR signal	$\delta_{\text{H}}$	$\delta_{\text{C}}^{\text{b}}$	$^1\text{H}$ NMR signal <sup>c</sup>	$\delta_{\text{H}}$	$\delta_{\text{C}}^{\text{e}}$
16		4.660 4.721 <sup>d</sup> m	72.75 72.23 <sup>d</sup>		4.630 4.630 <sup>d</sup> m	73.10 73.00 <sup>d</sup>
24		4.253 4.153 <sup>d</sup> s	60.85 60.87 <sup>d</sup>		3.948 3.803 <sup>d</sup> s	62.93 63.46 <sup>d</sup>
26		6.144 6.189 <sup>d</sup> s	96.57 95.81 <sup>d</sup>		5.771 5.748 <sup>d</sup> s	98.20 98.45 <sup>d</sup>
27a		4.437 4.444 <sup>d</sup> d (12.6)	57.78 57.38 <sup>d</sup>		1.799 1.645 <sup>d</sup> s	13.15 13.06 <sup>d</sup>
27b		4.949 4.606 <sup>d</sup> d (12.6)		-	-	-

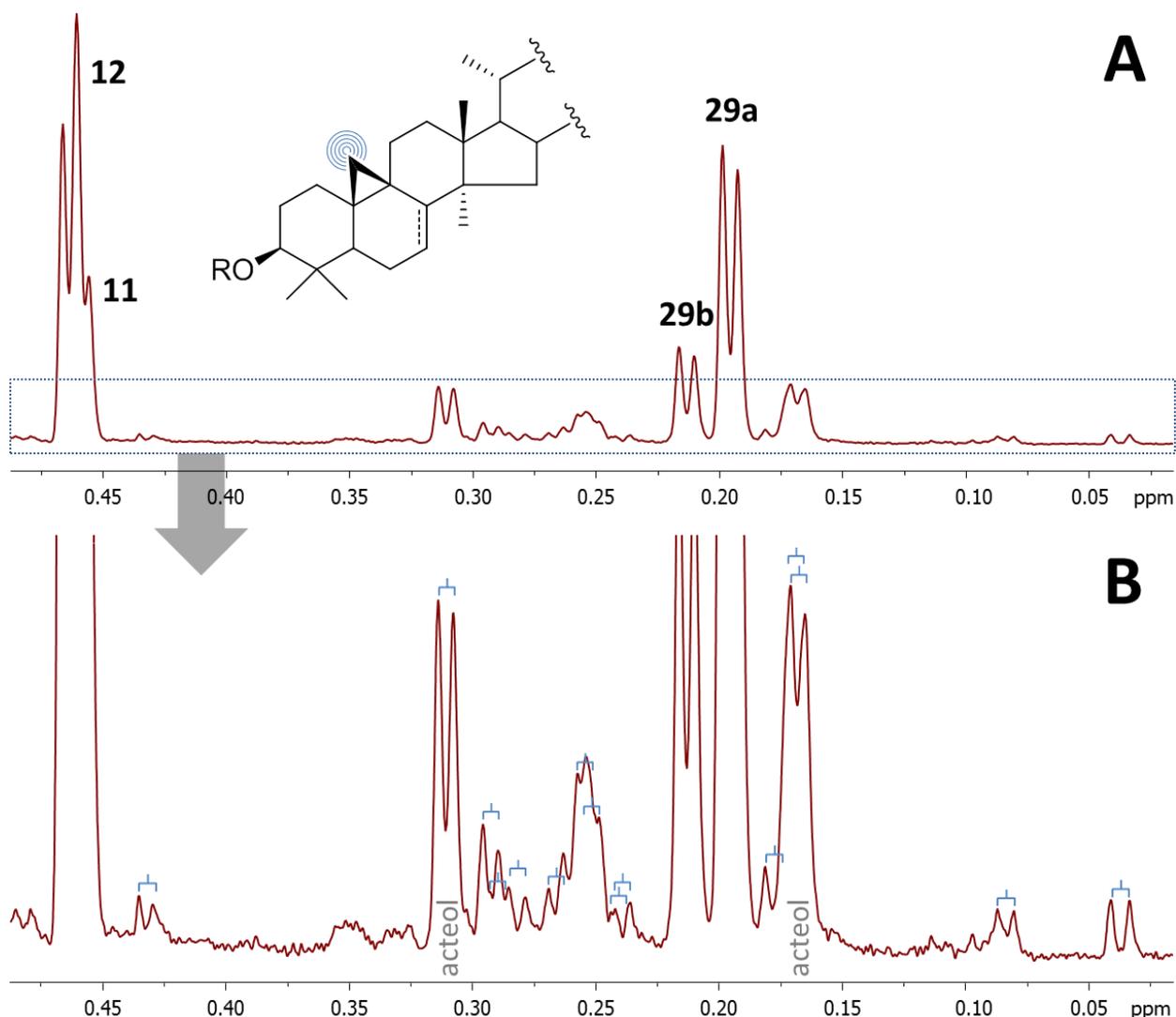
<sup>a</sup> The NMR spectra were measured in pyridine- $d_5$  (99.96 %D) at 400 MHz

<sup>b</sup> The  $^{13}\text{C}$  chemical shifts of **29** were determined from the HSQC spectra

<sup>c</sup> The intensity of methyl signal ( $\text{CH}_3$ -27) of **15** is not shown in proportion to other signals

<sup>d</sup> The NMR data of the major  $26\alpha$ -OH (26S) anomer of **29**

<sup>e</sup> The  $^{13}\text{C}$  data of **15** was taken from the literature (Kusano *et al.*, 1998)



**Figure 25. The Residual Complexity of the Fraction Ar.18.1.1.4 Indicated by the  $^1\text{H}$  NMR Signals of the Cycloartane Protons (H-19a) of *Actaea* Triterpenes**

In addition to **11**, **12**, and **29** as major components (Panel A), Ar.18.1.1.4 exhibited large residual complexity, containing more than 10 minor cycloartane triterpenoids based on the H-19 signals as observed in a blow-up of the  $^1\text{H}$  NMR spectrum (Panel B). It remains challenging to elucidate all these minor components in such highly residually complex sample of triterpenes solely based on the  $^1\text{H}$  NMR spectrum. However, using the HMBC pattern recognition (see next section), two compounds with H-19a protons resonating at 0.168 and 0.311 ppm were tentatively identified as two acteol glycosides.

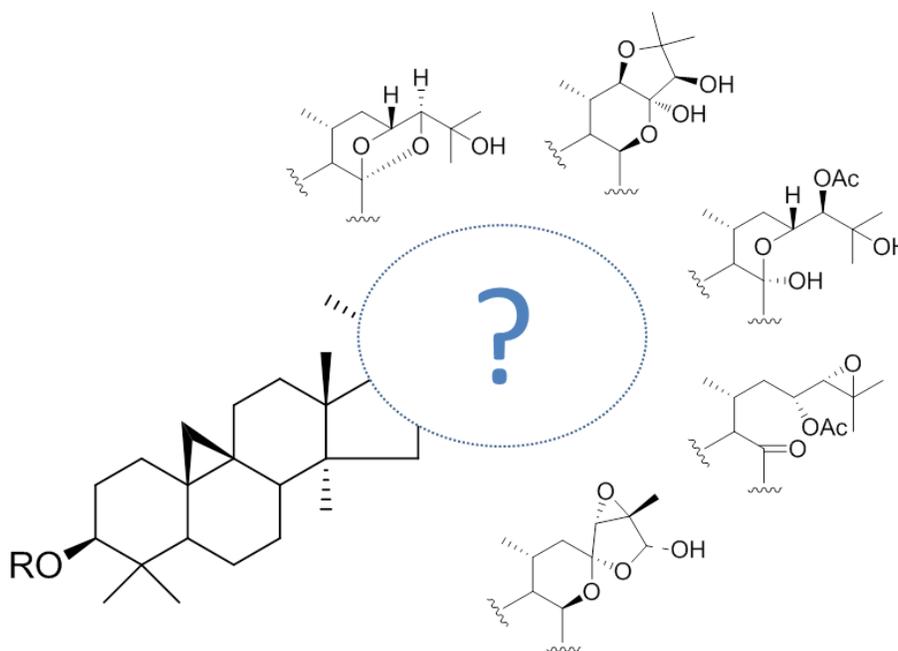
#### 4.5 Dereplication of *Actaea* Triterpenes in Residually Complex Mixtures

The rapid identification of known NPs, a process known as dereplication, is strategically important for scientists involved in screening for novel bioactive compounds from natural sources. The most common procedures used to identify compounds prior to purification are based on LC-UV, LC-MS, or a combination of the two. Both these two procedures require authentic standards for qualitative and quantitative analysis. However, NMR is not limited by these factors due to the almost ubiquitous occurrence of protons in organic compounds. Recently, NMR analysis of complex mixtures for dereplication of known compounds has received much attention. As NP mixtures exhibit large chemical complexity, the full interpretation of their NMR spectra is still challenging. The unique structure of an SCE results in a unique pattern of signals in the 1D/2D NMR spectra which are usually referred to as fingerprints. Much like biometric recognition, a small portion of these unique spectroscopic patterns is thought to be sufficient to distinguish different chemical entities. Thus, the identification of individual components in a mixture could be focused on the recognizable and characteristic subregions of the complex NMR spectra. Starting from this hypothesis, an HMBC-based approach was developed for dereplication of NPs using their characteristic  $^1\text{H}$ - $^{13}\text{C}$  correlations, which is similar to the process of pattern recognition.

As mentioned in Section 4.4 (p. 85), all of the known cycloartane triterpenes from *Actaea* fall into a few basic structural skeletons with the major difference in their C-17 side chains. Therefore, identification of the structures of side chains is key to the success of dereplication of these triterpenes. Comparing the structural characteristics of these side chains, they are mainly differentiated by the partial structures at C-24 and C-25 (**Figure 26**). Additionally, it is known that one or two terminal methyl groups ( $\text{CH}_3$ -26 and  $\text{CH}_3$ -27) are attached to C-25. As a result, the HMBC correlations between these methyl protons and proximal C-25 and C-24 produce uniquely recognizable patterns for each type of the *Actaea* triterpene, and, thus, can be used as indicators of the skeletal structures. Another advantage of using these  $^1\text{H}$ - $^{13}\text{C}$  correlations for

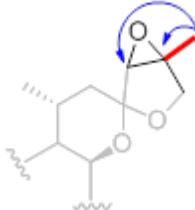
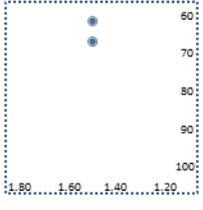
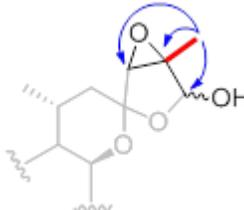
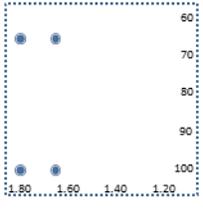
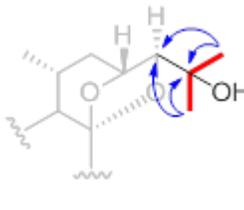
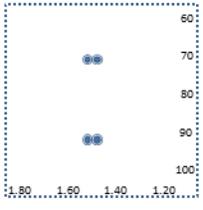
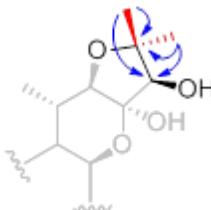
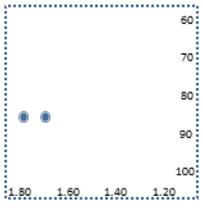
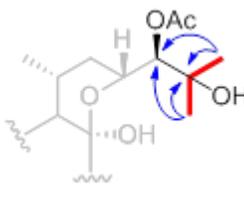
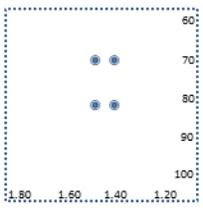
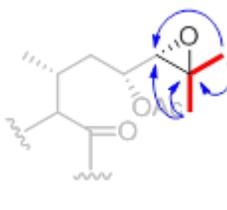
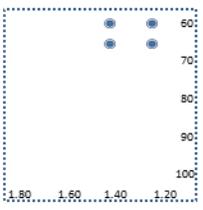
dereplication is that the methyl signals are more visually recognizable due to their greater intensities in the NMR spectra, making it possible to identify minor compounds in the residually complex sample.

**Table 5** lists the examples of *Actaea* triterpenes and their characteristic HMBC correlations regarding the terminal methyl groups. The  $^{2/3}J_{C,H}$  correlations between H-26/H-27 and the two proximal carbons C-25 and C-24, which are shown as cross peaks in the HMBC spectrum, form a uniquely recognizable pattern for each type of the *Actaea* triterpene. These specific patterns as simulated in **Table 5** are observed within a small region of  $\delta_H$  1.10–1.90 and  $\delta_C$  50.0–110.0 of the HMBC spectra. Two examples are given below for the dereplication of *Actaea* triterpenes in residually complex samples using this new HMBC pattern recognition approach.



**Figure 26. The Dereplication Problem for *Actaea* Triterpenes Arises from the Identification of Their Highly Variable Side Chains**

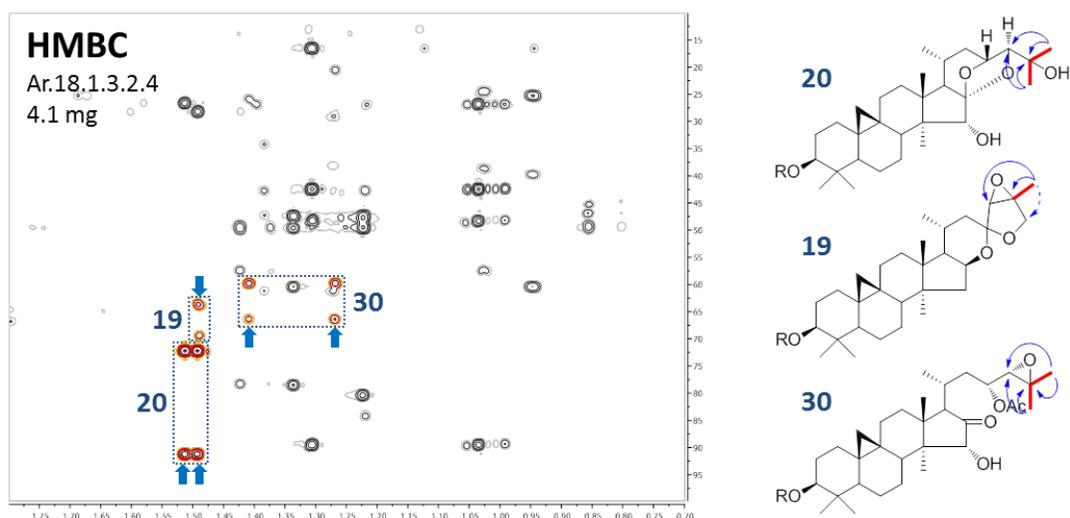
**TABLE V. EXAMPLES OF ACTAEA TRITERPENES AND THEIR CHARACTERISTIC HMBC CORRELATIONS OF Me-26 AND Me-27**

Compound Class	HMBC Correlation		Side Chain	Recognition Pattern
	$\delta_H$	$\delta_C$		
Acteol (26-H)	<i>one pair</i> $1.50 \pm 0.05$	$68 \pm 2$ $63 \pm 2$		
Acteol (26-OH)	<i>one pair</i> $1.65 \pm 0.05^a$ $1.80 \pm 0.05^b$	$65 \pm 2$ $100 \pm 2$		
Cimigenol	<i>two pairs</i> $1.45-1.55$	$72 \pm 2$ $90 \pm 2$		
Cimiracemoside	<i>two pairs</i> $1.68-1.82$	$84 \pm 2$ $84 \pm 2$		
Hydroshengmanol	<i>two pairs</i> $1.40-1.50$	$70 \pm 2$ $82 \pm 2$		
23-O-Acetylshengmanol	<i>two pairs</i> $1.25-1.45$	$60 \pm 2$ $65 \pm 2$		

<sup>a</sup> The  $^1\text{H}$  NMR data of 26 $\alpha$ -OH (26S) epimer

<sup>b</sup> The  $^1\text{H}$  NMR data of 26 $\beta$ -OH (26R) epimer

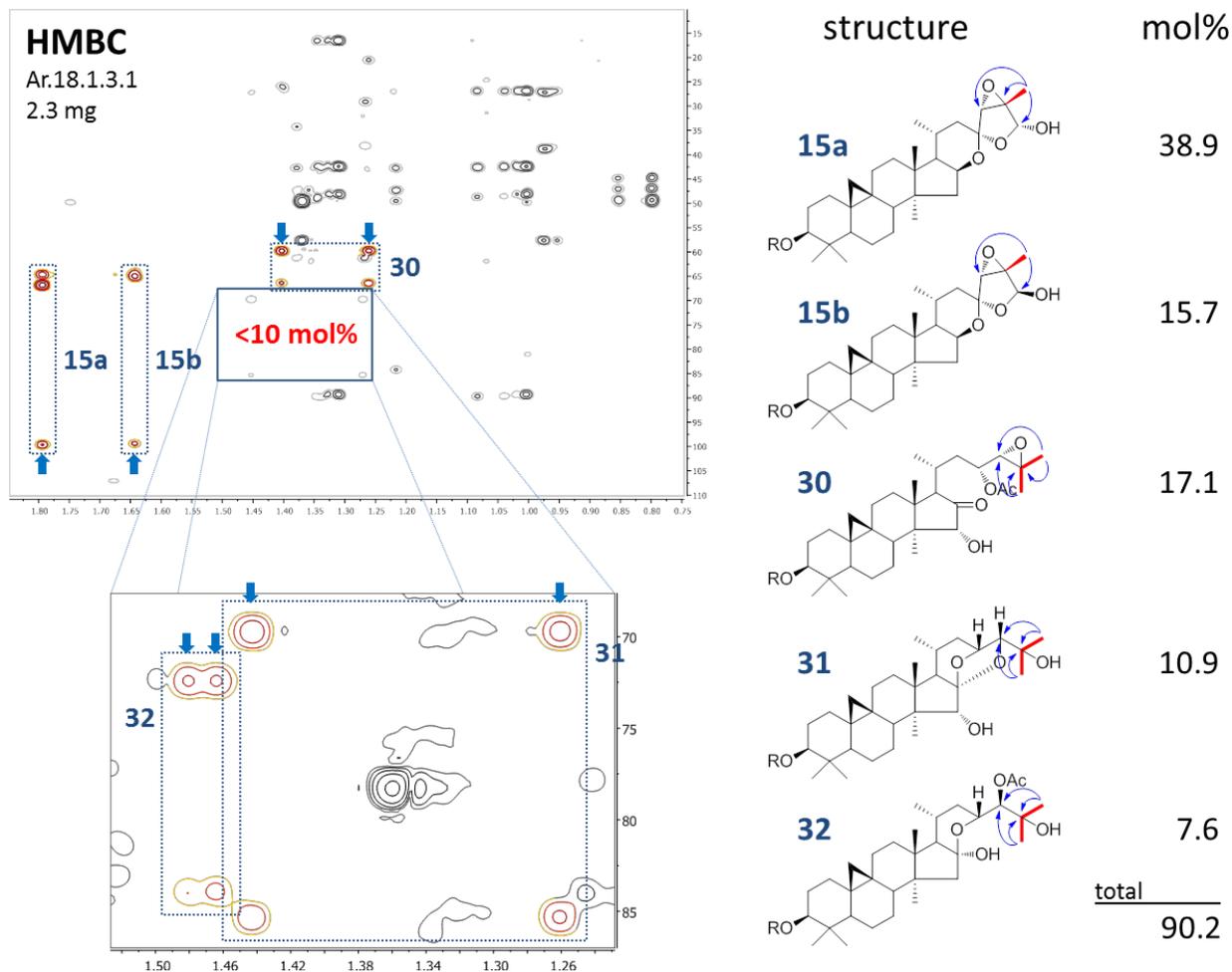
**Case 1:** The sample for this case (Ar.18.1.3.2.4, 4.1 mg) was generated in the 5<sup>th</sup> step of fractionation of black cohosh extracts (see **Figure 10**). The sample was initially subjected to <sup>1</sup>H NMR analysis. Each *Actaea* triterpene gives rise to a pair of doublets in the range of  $\delta_{\text{H}}$  0.20–1.00, due to their cyclopropane methylene protons, H-19a/b. On the basis of these characteristic signals, it is known that this sample contained three major triterpenes. The fact that none of the H-19a signals are shifted downfield to ~1.00 ppm indicates that all three triterpenes are saturated at C-7 and C-8. The HMBC spectrum was then acquired on a 700 MHz cryo-microprobe NMR under the following conditions: Solvent 35  $\mu\text{L}$  pyridine-*d*<sub>5</sub> (99.96 %D), TD 2048K/256K (F2/F1), and NS 4. As shown in **Figure 27**, these three triterpenes are readily recognized as the cimigenol glycoside (**20**), the acteol glycoside (26-H, **19**), and the 23-O-acetylshengmanol glycoside (**30**), based on pattern recognition of their specific HMBC correlations. Their molar ratio was measured by qHNMR as 62:22:16 by using the integral of their H-19b signals.



**Figure 27. Identification of Three Major *Actaea* Triterpenes in a Residually Complex Sample (Ar.18.1.3.2.4) by Pattern Recognition of Characteristic HMBC Correlations**

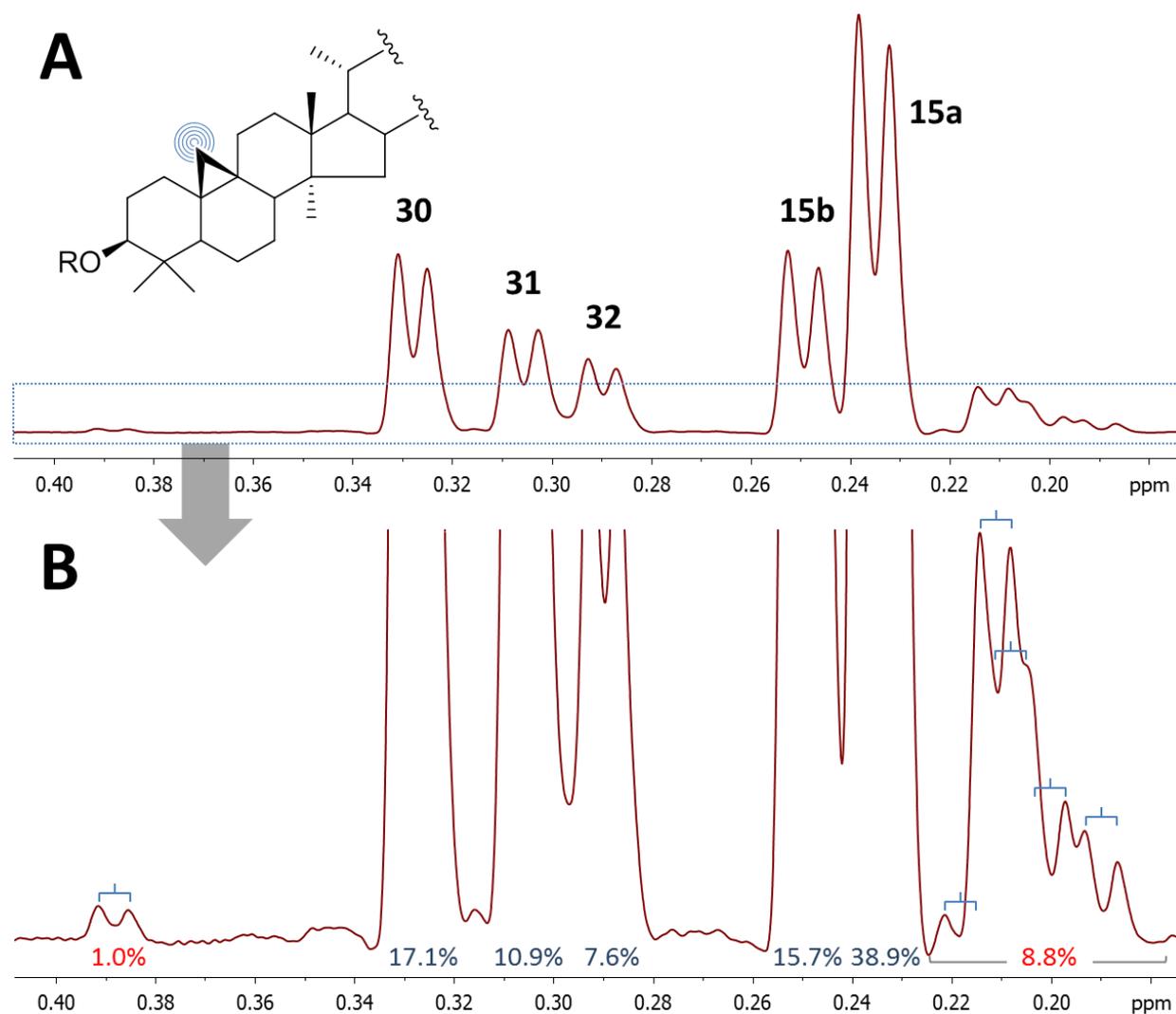
NMR experimental conditions: 700 MHz, micro-cryoprobe, 35  $\mu\text{L}$  pyridine-*d*<sub>5</sub>, TD 2048K/256K (F2/F1), and NS 4.

**Case 2:** The development of cryo-microprobe NMR has significantly extended the scope of qNMR, pushing the LOD down into the nanomole scale and making it possible to efficiently dereplicate compounds in materials with limited or minor quantities. For example, Sample 2 (Ar.18.3.1, 2.3 mg) was a fraction obtained in the 4<sup>th</sup> step of fractionation of black cohosh extracts (see **Figure 10**). Its <sup>1</sup>H NMR spectrum revealed the presence of three major triterpenes (each > 10 mol%) according to the presence of H-19 signals in the range of  $\delta_{\text{H}}$  0.20–1.00. Further HMBC analysis identified them as the acteol glycosides (26-OH, **15a** and **15b**) and the 23-O-acetylshengmanol glycoside (**30**) (**Figure 28**). As acteols with a hydroxy substituent at C-26 represent mixtures of the C-26 epimers in solution with a molar ratio of ca. 7:3, a pair of HMBC correlations between H-27 and C-24/C-26 can be observed for each epimer. In addition to these three triterpenes, the upper field of the <sup>1</sup>H NMR spectrum of Sample 2 showed more than two additional pairs of minor signals for H-19a, indicating a high residual complexity of this sample. By increasing the intensities of the HMBC spectrum, two additional HMBC correlation patterns were revealed, which apparently represented the 24-*epi*-cimigenol glycoside (**31**) and the hydroshengmanol glycoside (**32**), respectively. Because the <sup>1</sup>H NMR spectra were acquired under quantitative conditions (qHNMR), their molar content could be calculated by the “100% method” of qHNMR using the integrals of their H-19a signals (**Figure 29**). As a result, the total content of the overall five identified triterpenes account for 90.2% of the sample.



**Figure 28. Resolving the Residual Complexity of an *Actaea* Triterpene Sample (Ar.18.3.1) by Pattern Recognition of Characteristic HMBC Correlations**

Using the same NMR conditions in Case 1, the minor constituents down to the 0.20 mg (0.30  $\mu\text{mol}$ ) level were identified in the 2.3 mg sample. The identification of four co-occurring minor constituents, belonging to four different triterpene skeleton types, in a repeatedly purified NP emphasizes the critical need for the evaluation of residual complexity of reference materials, especially when used for biological assessment.



**Figure 29. The Residual Complexity of the Sample Ar.18.1.3.1 Indicated by the  $^1\text{H}$  NMR Signals of the Cycloartane Protons (H-19a) of *Actaea* Triterpenes**

Panel A shows the specific region for H-19a signals in the  $^1\text{H}$  NMR spectrum indicating five major cycloartane triterpenes. Based on their relative intensities, these signals can be assigned to the corresponding compounds identified in the HMBC spectrum. A blow-up of the  $^1\text{H}$  NMR spectrum further unveils additional minor components which consist of at least six cycloartane triterpenes, as shown in Panel B. The molar content of these compounds was calculated by the “100% method” of qHNMR using the integrals of their H-19a signals.

It is particularly noteworthy that the remaining 9.8% of content shows a higher order of residual complexity for this sample. As the H-19a signals of cycloartane triterpenes are all doublets with  $J$  values of ca. 4.20 Hz, it is known that these minor components consist of at least six cycloartane triterpenes. Surprisingly, in this case, the very minor constituents down to the 20  $\mu\text{g}$  (ca. 30 nmol) level in the 2.3 mg sample were detected in the  $^1\text{H}$  NMR spectrum using a 700 MHz 1.7 mm cryo-microprobe. Based on the overlapping of the H-19a signals, it is safe to conclude that the H-19a signals of additional minor components can be buried under those of major ones, such as **15** and **30**. The structures of these minor compounds remained unidentified because their limited quantity was possibly beyond the LOD of HMBC spectrum under the current data acquisition conditions. It is also possible that their characteristic HMBC correlations were overlapped with those of same-type identified components. Future studies may enhance the NMR experimental conditions for an improved LOD in order to resolve this high residual complexity arising from the very minor and diverse impurities. In addition, the COSY spectra may be employed for the identification and pattern recognition of cycloartane protons (H-19a/b).

These two cases demonstrate that NMR pattern recognition represents a powerful approach for rapid dereplication of NPs of both pure compounds and complex mixtures. While limited examples are illustrated here, additional studies employed statistical analysis for developing computational models of pattern recognition which will improve the reliability of dereplication results (see Chapter 5, p. 93). Finally, the approach presented here, combined with contemporary qHNMR methodology using cryo-microprobes, has potential for the standard characterization of residual complexity of NP reference materials.

## 4.6 Conclusion

Chapter 4 introduced an extension of conventional NMR methodologies in NP research and exemplified practical applications for raw plant materials used in popular dietary supplements. The experimental results demonstrated that NMR is a versatile and universal analytical technique and can be employed for various aspects of NP research in terms of their qualitative and quantitative characterization. When offline coupled with chromatography, NMR enables expansion of traditional elution detection to multiple dimensions, which qualitatively and quantitatively visualizes the elution profiles of target compounds and even of minor impurities. Unveiling the full map of separation procedures facilitates the evaluation of residual complexity as well as the establishment of purity–activity relationships (see Section 6.1, p. 135). Despite the chemical complexity of NP mixtures, the individual constituents can still be characterized by a combination of 2D NMR techniques, comparative analysis, and pattern recognition. These tools not only expedite identification of new compounds at the early stage of separation, but also assist in the dereplication of known compounds in mixtures without their physical isolation, making the separation a highly efficient and targeted procedure. Looking forward, expansion of the NMR utilities for mixture analysis together with implementation of chemometric techniques should have a wide application in NP research. As the abovementioned NMR methodologies enable simultaneous identification and quantification of chemical constituents of complex mixtures, they can be applied to metabolomic profiling of NPs, revealing the full chemical image of crude NPs such as botanicals for the comprehensive interpretation of their metabolomic profiles, and, thus, providing a chemical foundation for their biological evaluation as well as standardization of their quality.

## 5. RATIONAL NAMING AND VIRTUAL PARTITIONING OF NATURAL PRODUCTS<sup>1</sup>

### 5.1 Introduction

The genera *Cimicifuga* and *Souliea*, now considered to be part of the genus *Actaea* (Campton *et al.*, 1998), have been the source of almost 200 triterpenes possessing the cycloartane skeleton (Li *et al.*, 2006). Almost all these triterpenes have been accorded trivial names, to a large extent derived from the Latin binomial or common names associated with the source plant. These names, at best, provide clues as to the origin of the compound, but seldom have any indication of the actual structure to the non-cognoscenti, and certainly do not help the scientist in the search for novel triterpenes. A non-comprehensive list of these names includes: acerinol, acerionol, acteol, bugbanoside, cimiaceroside, cimicidanol, cimicidol, cimicifoetiside, cimicifol, cimicifugenol, cimicifugoside, cimifoetiside, cimifoside, cimigenol, cimigol, cimilactone, cimiracemoside, cimiside, dahurinol, foetidinol, heracleiforinol, and shengmanol (Li *et al.*, 2006). The names are also not practical, even for the specialist, because the similarity of names gives no indication of similarity of structure. One example is the cimiracemosides A, M, and P, which have completely different ring systems and differ in the sites of oxygenation at C-12, C-15, C-16, C-21, C-23, and C-26. Another illustration is reflected by the fact that most of the triterpenes from this genus occur as glycosides, usually at the C-3-oxygen, and for the most part these are named with the suffix “-oside”, whereas the aglycones have the suffix “-ol”. However, even this simple convention is not universally followed as cimicidol, cimicifol, and acteol are all glycosides, whereas acerinol and heracleiforinol, although being alcohols, no longer have that functional group at C-3.

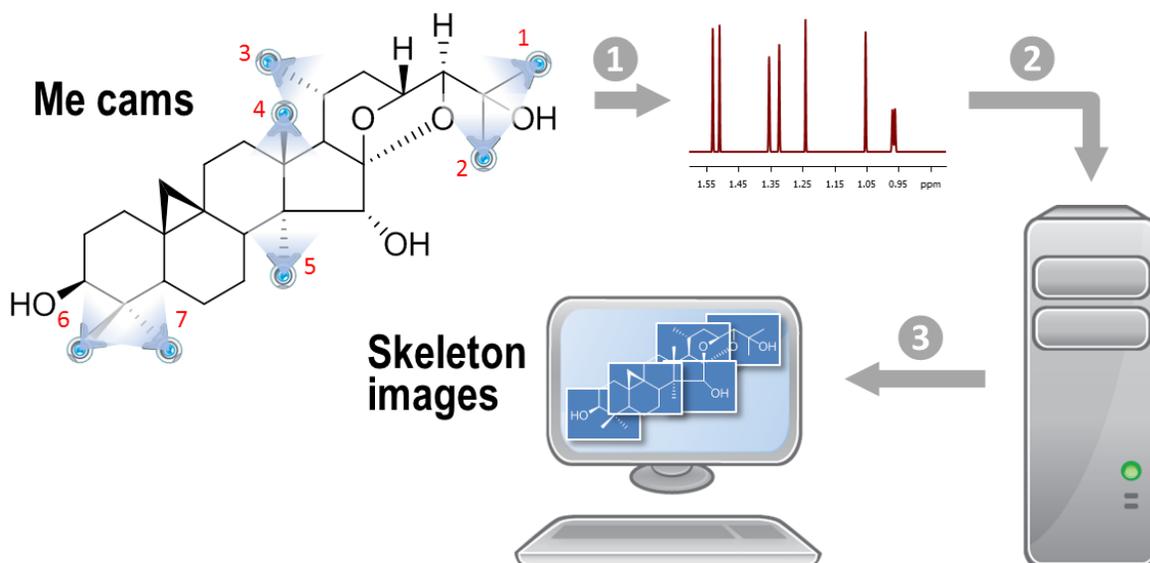
Here a new rational naming system is proposed, which will simplify the deduction of all known *Actaea* triterpene structures as well as congeners yet to be discovered, given the

---

<sup>1</sup> Content presented in this chapter has been partially published in: Qiu, F.; Imai, A.; McAlpine, J. B.; Lankin, D. C.; Burton, I.; Karakach, T.; Farnsworth, N. R.; Chen, S.-N.; Pauli, G. F. Dereplication, Residual Complexity and Rational Naming: The Case of *Actaea* Triterpenes, *Journal of Natural Products* 2012, 75, 432–443

knowledge of only the cycloartane skeleton. In reclassifying *Cimicifuga* within the genus *Actaea*, botanists have given chemists an opportunity to systematize this nomenclature, and add support to the aforementioned reasoning for a new naming scheme. The use of the generic name *Actaea* as the basis for the new naming system is further justified by the recent discovery of several very closely related triterpenes from *Actaea vaginata* (previously *Souliea vaginata*), a species never considered to be a *Cimicifuga* (Zhou *et al.*, 2004, 2005, 2006).

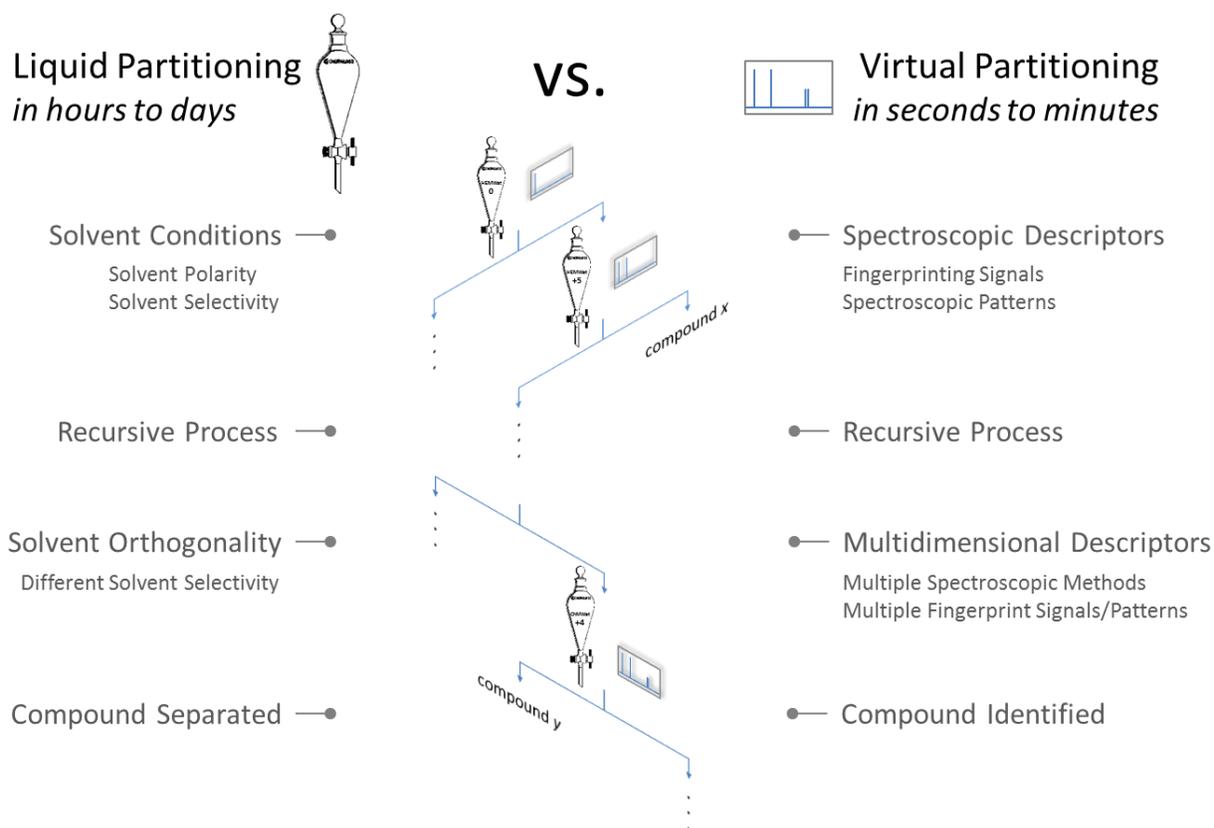
An HMBC-based pattern recognition method for rapid dereplication of *Actaea* triterpenes was introduced in Chapter 4. Although 2D NMR spectra provide more structural information, they require much longer data acquisition time compared to 1D  $^1\text{H}$  NMR spectra. The quality of spectra is also dependent of sample mass and NMR instruments. This chapter developed an alternative dereplication method using more readily obtainable  $^1\text{H}$  spectra. It relies on the fact that most of these compounds have five to seven skeletal methyl (Me) groups serving as the “surveillance units” (“Me cams”) for the neighboring segments of their molecules. Therefore, their full structures can be mapped by combining all “surveillance images” provided by each of the Me groups as “surveillance units” (**Figure 30**). Furthermore, despite the complexity of  $^1\text{H}$  NMR spectra, most of the Me signals of *Actaea* triterpenes are singlets with relatively high intensity, making them more distinguishable than methines and methylenes. The history of using only Me groups in the determination of structures initially dates back to the late 1950s and into the 1960s for steroids (Shoolery *et al.*, 1958; Bhacca *et al.*, 1964) and triterpenes (Lavie *et al.*, 1964; Tursch *et al.*, 1967; Cheung *et al.*, 1969). These studies analyzed the additive intramolecular shielding or deshielding effects of proximate substituents on the chemical shifts of the Me groups, and as a result, the substitution pattern of the substituents could be deduced and the structures of triterpenes could be elucidated using this approach. Two more recent studies used this approach for the structural elucidation of cardiac glycosides (Pauli, 1993, 1995) and unsaturated  $\text{C}_{27}$  sterols (Wilson *et al.*, 1996).



**Figure 30. The Concept of Using Methyl Groups of Triterpenes as Partial Structural Indicators (“Me Cams”) to Map the Full Skeleton of Molecules**

Step ①: Imitating the mechanism of a biometric system, the raw data (e.g., Me shifts and multiplicities) are collected by each of the “Me cams”; Step ②: The collected data are processed *in silico* for pattern recognition. Step ③: The patterns of methyl NMR data are converted to visible images, representing partial structures from which the full structure can be assembled.

The present study hypothesizes that the relationship between the Me shifts and structural characteristics is statistically correlated, and that correlation can be further integrated into a pattern recognition model for structural determination. Starting from this hypothesis, we aimed to establish a novel methodology that uses classification binary trees (CBTs) (Kokkinofta *et al.*, 2003; Petrakis *et al.*, 2005, 2008) for a rapid and automatic structural dereplication. The CBTs function as a virtual partitioning process that classifies members of the population based on several dichotomous dependent variables (**Figure 31**). In the present study, an in-house database currently containing the Me shifts of ~180 *Actaea* triterpenes was assembled. Using the Me shifts in the database as the training set, the CBTs for the classification of *Actaea* triterpenes were generated by classification and regression tree (CART) (Brown *et al.*, 2009) analysis.



**Figure 31. A Comparison between the Processes of Liquid and Virtual Partitioning of NPs**

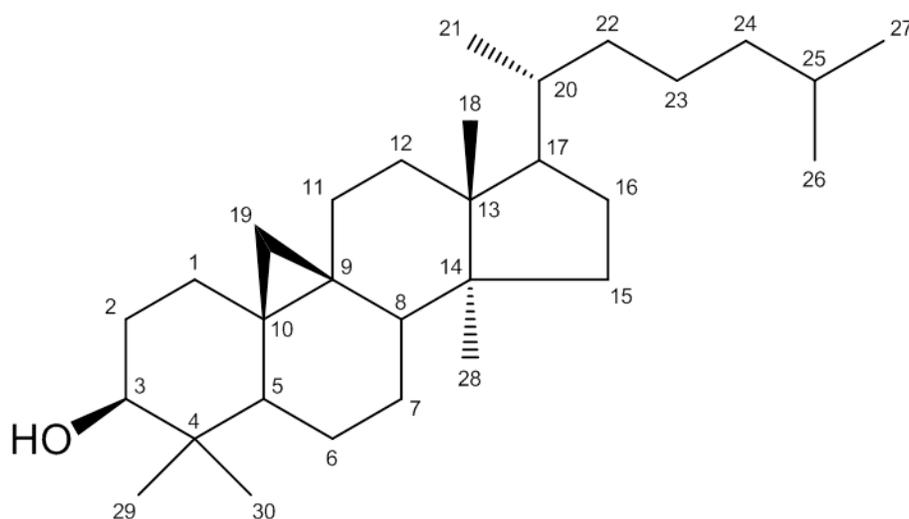
The basic mechanism of the CBTs is that a group of chemical entities is virtually partitioned using several spectroscopic descriptors in a recursive process until all SCEs with the same spectroscopic characteristics are separated/classified, and thus can be identified.

Furthermore, triterpenes are a good example to demonstrate an important signature of NPs, i.e., that certain levels of characteristic impurity patterns, referred to as residual complexity, remain visible along the entire (bio-)analytical pathway (Chen *et al.*, 2009). The term, residual complexity, refers to the easily overlooked impurity profile of isolated NPs, which may exert a significant influence on their accurate biological assessment (Schinkovitz *et al.*, 2008; Jaki *et al.*, 2008; Chen *et al.*, 2009). It is therefore important to characterize both qualitatively and quantitatively the impurity profile of isolated NPs. In the present study, both

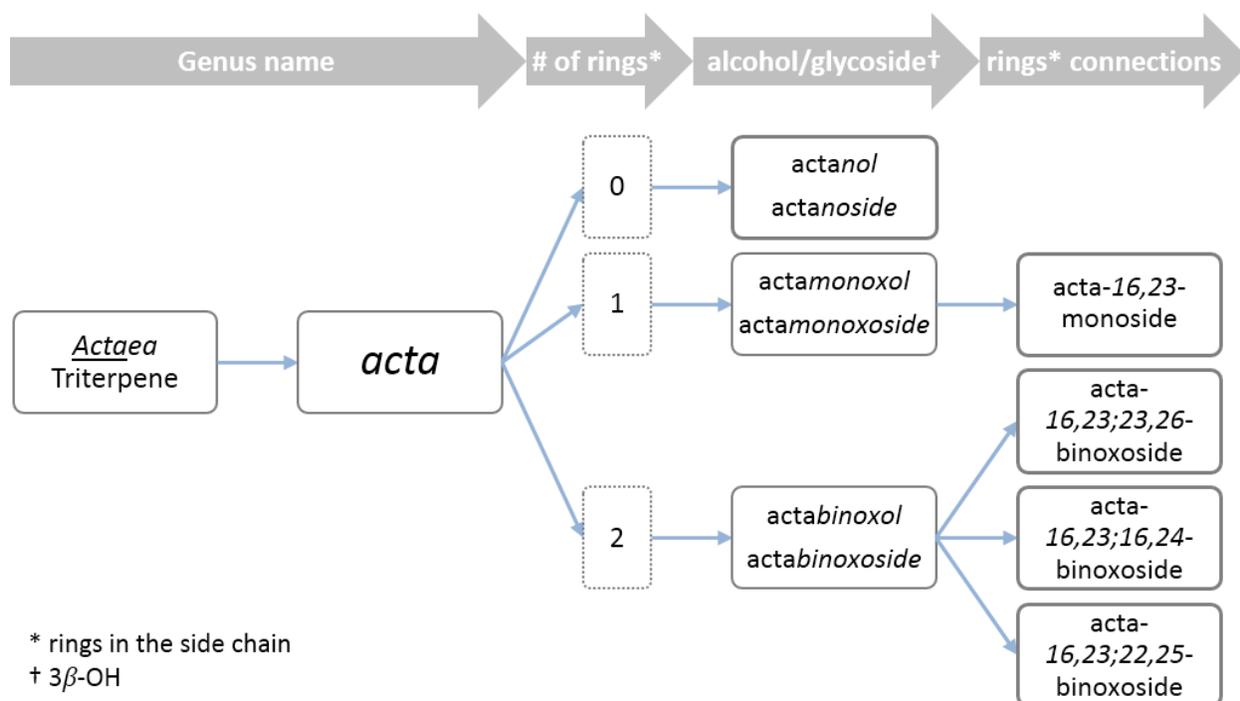
classification models and the database search were utilized as *in silico* tools to dereplicate *Actaea* triterpenes in residually complex samples of purified reference materials.

## 5.2 The New Naming System

All of the known cycloartane triterpenes from *Actaea* fall into only a few basic structural skeletons. As far as C-20 to C-27 are concerned, there are acyclic compounds in which these carbons have no connections between themselves other than the basic carbon chain, as shown in **Figure 32**. Then, there are other compounds in which some of these carbons are involved in one or two rings, usually formed by ether or acetal oxygens, often onto C-16. Accordingly, the new system (**Figure 33**) would name the acyclic aglycone compounds as *actanols*, those with a single oxygen bridge forming a further ring as *actamonoxols*, and those with two oxygen-containing rings *actabinoxols*. Oxiranes are not included in these root names but are accounted for as substituents. These names all include the  $3\beta$ -hydroxy group. Where this group is part of a glycosidic linkage, the suffix would be “-oside”, e.g., actabinoxoside. All of the substituents and other structural modifications need to be affixed using standard chemical nomenclature, with prefixes arranged in alphabetical order.



**Figure 32. The Aglycone of Tetracyclic *Actaea* Triterpenes**

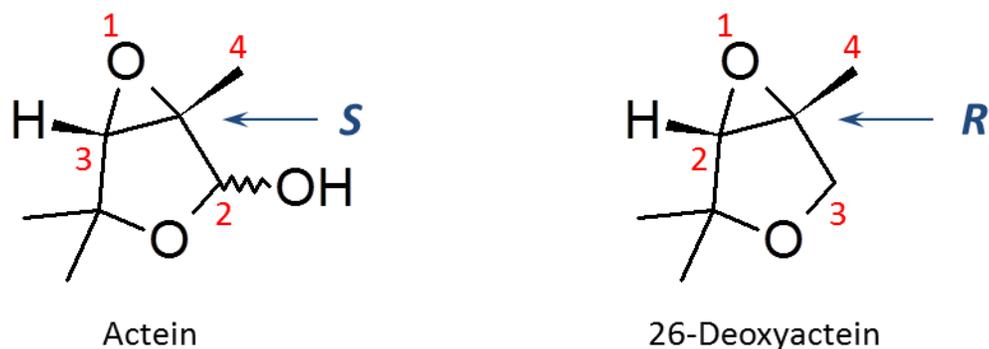


**Figure 33. The Basic Rules of the New Naming System for *Actaea* Triterpenes**

Using the genus name as the basis, the *Actaea* triterpene aglycones are initially named as actanols, actamonols or actabinoxols based on the number of rings in their side chains at C-17 which reflects their major structural differences. Then they are further differentiated by the positions of the rings' connections, such as acta-16,23;23,26-binoxoside. This concept of combining the botanical and chemical information in the nomenclature can be applied to other classes of NPs.

The absolute configuration of these triterpenes would be designated via the Cahn-Ingold-Prelog (CIP) system rather than the simpler  $\alpha/\beta$  system commonly used in steroid nomenclature, because the  $\alpha/\beta$  nomenclature fails in bicyclic caged rings that occur in many of the actabinoxols. This problem has been addressed until now by partial use of both systems, however the CIP system works universally, and we are advocating its use in all cases except for glycosidic linkages, where the  $\alpha/\beta$  and D/L system for sugars is well accepted and fully understood. The CIP system have a disadvantage in that the stereodesignation of a specific stereocenter can change without a requisite change in the configuration, but rather by changes

in CIP-preferences of nearby substituents, and hence a change in the precedence number of substituents to that carbon (**Figure 34**).



**Figure 34. Example of the Change of CIP Designation without a Change in Configuration**

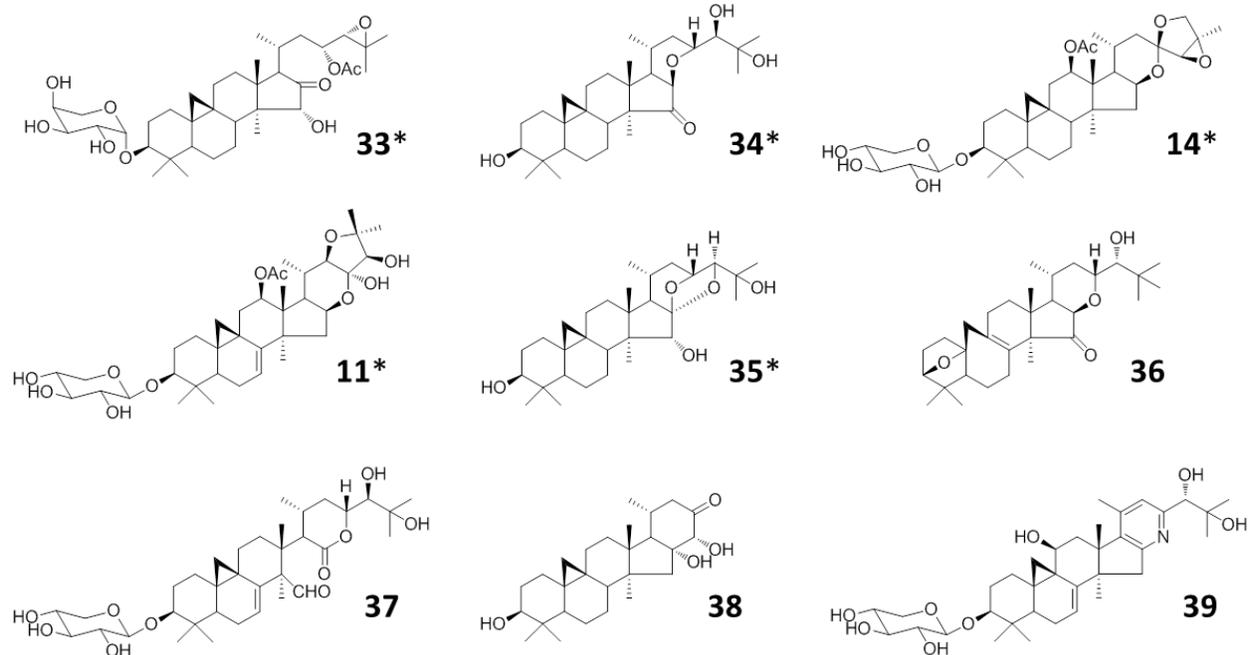
The C-25 of actein and 26-deoxyactein have the same configuration but different CIP designations by virtue of an interchange in the precedence numbers of C-24 and C-26.

Although such a naming scheme will occasionally result in reasonably long names, they will be readily understood for all *Actaea* triterpenes by organic chemists with nothing more than the basic knowledge of **Figure 32**. Some representatives of *Actaea* triterpenes are listed in **Table 6** and their structures are given in **Figure 35**. The first five skeletons include more than 90% of the known triterpenes from *Actaea*. There are, however, a limited number of compounds that do not share these basic structures. Some compounds, which are missing carbons at the end of the chain, are readily accommodated by the “nor” prefix, and there are a couple of types where carbon-carbon bonds are cleaved and use of the “seco” prefix is required.

**TABLE VI. REPRESENTATIVES OF THE ACTAEA TRITERPENES. STRUCTURAL TYPES 1–5 (\*) COVER MORE THAN 90% OF KNOWN STRUCTURES**

Common Name	New Systematic Name
23- <i>O</i> -Acetylshengmanol arabinoside ( <b>33</b> )*	(23 <i>R</i> )-23-Acetoxy-(24 <i>S</i> )-24,25-epoxy-(15 <i>R</i> )-15-hydroxy-16-oxo-3- <i>O</i> - $\alpha$ -L-arabinopyranosylactanoside
Dahurinol ( <b>34</b> )*	(24 <i>R</i> )-24,25-Dihydroxy-15-oxoacta-(16 <i>R</i> ,23 <i>R</i> )-16,23-monoxol
23- <i>epi</i> -26-Deoxyactein ( <b>14</b> )*	(12 <i>R</i> )-12-Acetoxy-(24 <i>R</i> ,25 <i>R</i> )-24,25-epoxy-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-(16 <i>S</i> ,23 <i>R</i> )-16,23;23,26-binoxoside
Cimiracemoside F ( <b>11</b> )*	(12 <i>R</i> )-12-Acetoxy-7,8-didehydro-(23 <i>R</i> ,24 <i>S</i> )-23,24-dihydroxy-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-(16 <i>S</i> ,22 <i>R</i> )-16,23;22,25-binoxoside
Cimigenol ( <b>35</b> )*	(15 <i>R</i> )-15,25-Dihydroxyacta-(16 <i>S</i> ,23 <i>R</i> ,24 <i>S</i> )-16,23;16,24-binoxol
Acerionol ( <b>36</b> )	3-Deoxy-8,9-didehydro-(24 <i>R</i> )-24,25-dihydroxy-(3 <i>S</i> ,10 <i>S</i> )-3,10-epoxy-15-oxo-9,10-secoacta-(16 <i>R</i> ,23 <i>R</i> )-16,23-monoxol
Compound <b>37</b> <sup>a</sup>	7,8-Didehydro-(24 <i>R</i> )-24,25-dihydroxy-15-formyl-16-oxo-15,16-seco-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-(23 <i>R</i> )-16,23-monoxoside
Foetidinol ( <b>38</b> )	(16 <i>R</i> ,24 <i>R</i> )-16,24-Dihydroxy-23-oxo-25,26,27-trinoracta-16,24-carbamanol
Cimicifugadine ( <b>39</b> ) [an alkaloid]	(11 <i>S</i> ,24 <i>S</i> )-11,24,25-Trihydroxy-7,8,16,17,20,22,23, <i>N</i> -octadehydro-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-16,23-monozoside

<sup>a</sup> No common name has been assigned



**Figure 35. Structures of the Representative Subclasses of the Actaea Triterpenes**

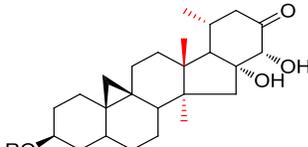
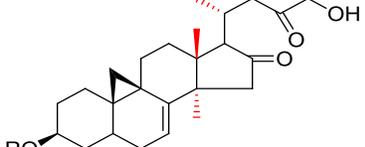
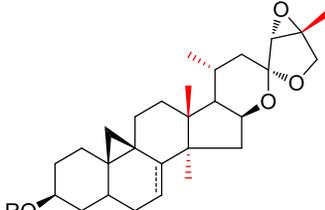
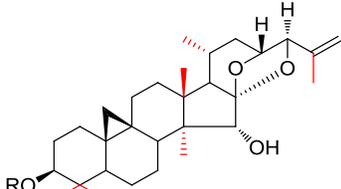
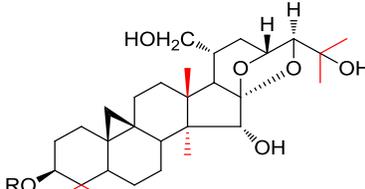
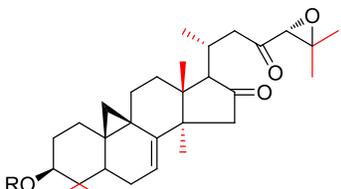
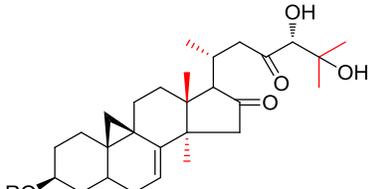
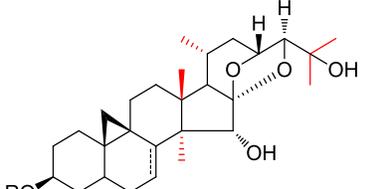
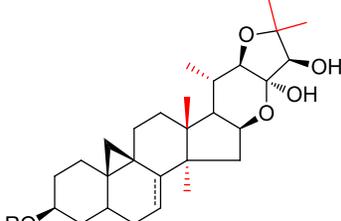
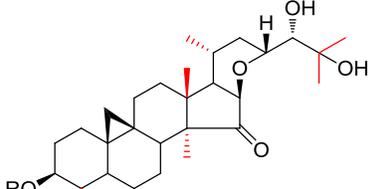
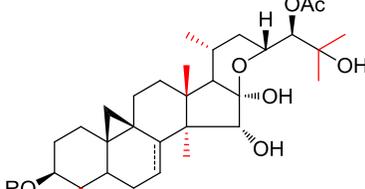
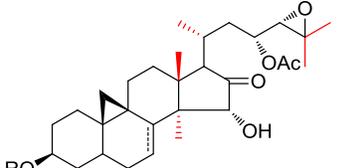
The first five skeletons (\*) include >90% of currently known structures.

### 5.3 Construction of Database

In order to obtain sufficient data to develop effective classification models, an extensive literature search was carried out through SciFinder (American Chemical Society, Washington D.C.) to locate reports of *Actaea* triterpenes with spectroscopic data. For each reported triterpene, the collected data include the references, the chemical structures, the compound names and types, the species names, the NMR field strength and solvents, and the  $^1\text{H}$  NMR data of the cyclopropane methylene (H-19a/b) and all Me groups including chemical shift and assignment. All chemical shift values were recorded with two decimal places for a homogenous dataset. While pyridine- $d_5$  was used for most triterpene glycosides, the less polar solvent  $\text{CDCl}_3$  was used in a few cases, especially for triterpene aglycones. Owing to the effect of various solvents on the chemical shifts, the NMR solvent was also noted for each compound. Mining of other NMR acquisition parameters including temperature was omitted due to the frequent lack of reporting in the literature.

The data sets of  $^1\text{H}$  NMR spectra, predominantly measured in pyridine- $d_5$ , for ~180 *Actaea* triterpenes, representing all *Actaea* triterpenes found in SciFinder, were collected (**Figure 36**; Appendix E, p. 192). The major types of compounds are (both trivial and new names given; see also **Table 7**): cimigenols (50, 33%; acta-16,23;16,24-binoxols), acteols (16, 11%; acta-16,23;23,26-binoxols), hydroshengmanols (16, 11%; acta-16,23-monoxols), cimiracemosides (12, 8%; acta-16,23;22,25-binoxols), 23-O-acetylshengmanols (11, 7%; 16-oxo-actanols), cimicidanol (7, 5%; 16,23-dioxo-actanols), dahurinols (5, 3%; acta-16,23-monoxols), foetidonols (5, 3%; acta-16,24-carbamonols), and cimicidols (4, 3%; 16,23-dioxo-actanols). Some rare subtypes, such as 15,16-secocimicidols, alkaloids, cimicifugenols, cimilactones, and heracleiforinols comprising one or two known compounds, were also included.

**TABLE VII. THE MAJOR TYPES OF ACTAEA TRITERPENES INCLUDED IN THE IN-HOUSE DATABASE, ALONG WITH THEIR COMMON (⊗) AND NEW SYSTEMATIC NAMES (⊙)<sup>a</sup>**

 <p>⊗ Foetidinol ⊙ (16<i>R</i>,24<i>R</i>)-16,24-Dihydroxy-23-oxo-25,26,27-trinoracta-16,24-carbamanol</p>	 <p>⊗ Trinorcimicidol ⊙ 7,8-Didehydro-24-hydroxy-16,23-dioxo-25,26,27-trinoractanol</p>	 <p>⊗ Acteol ⊙ (24<i>R</i>,25<i>R</i>)-24,25-Epoxyacta-(16<i>S</i>,23<i>R</i>)-16,23;23,26-binoxol</p>
 <p>⊗ 25-Dehydrocimigenol ⊙ 25,26-Didehydro-(15<i>R</i>)-15-hydroxyacta-(16<i>S</i>,23<i>R</i>,24<i>S</i>)-16,23;16,24-binoxol</p>	 <p>⊗ 21-Hydroxycimigenol ⊙ (15<i>R</i>)-15,21,25-Trihydroxyacta-(16<i>S</i>,23<i>R</i>,24<i>S</i>)-16,23;16,24-binoxol</p>	 <p>⊗ Cimicidanol ⊙ 7,8-Didehydro-(24<i>S</i>)-24,25-epoxy-16,23-dioxoactanol</p>
 <p>⊗ Cimicidol ⊙ 7,8-Didehydro-(24<i>R</i>)-24,25-dihydroxy-16,23-dioxoactanol</p>	 <p>⊗ Cimigenol ⊙ (15<i>R</i>)-15,25-Dihydroxyacta-(16<i>S</i>,23<i>R</i>,24<i>S</i>)-16,23;16,24-binoxol</p>	 <p>⊗ Cimiracemoside ⊙ (23<i>R</i>,24<i>S</i>)-23,24-Dihydroxyacta-(16<i>S</i>,22<i>R</i>)-16,23;22,25-binoxol</p>
 <p>⊗ Dahurinol ⊙ (24<i>R</i>)-24,25-Dihydroxy-15-oxoacta-(16<i>R</i>,23<i>R</i>)-16,23-monoxol</p>	 <p>⊗ Hydroshengmanol ⊙ (24<i>S</i>)-24-Acetoxy-(15<i>R</i>,16<i>R</i>)-15,16,25-trihydroxyacta-(23<i>R</i>)-16,23-monoxol</p>	 <p>⊗ 23-O-Acetylshengmanol ⊙ (23<i>R</i>)-23-Acetoxy-(24<i>S</i>)-24,25-epoxy-(15<i>R</i>)-15-hydroxy-16-oxoactanol</p>

<sup>a</sup> The methyl groups used for the dereplication models are indicated in red.



Each compound was given a unique number (“Link #”, the 3<sup>rd</sup> column) which hyperlinks to their structural images and systematic names in a separate file created in ChemBioDraw. The methyl shifts were entered in ascending order, i.e., Me1 < Me2 ... Me9, each followed by the corresponding assignment in an associated separate column. Using Excel’s built-in database tools, the data entries can be readily sorted and filtered according to any required criteria, thus facilitating data search and analysis. The database is being updated whenever new data becomes available.

Moreover, an *in silico* “search-and-match” function (ActaMatch, Appendix C, p. 180) was developed using Microsoft Visual Basic for Applications 7.0 (VBA) code in a separate worksheet (**Figure 37**). Pearson’s correlation is a statistical method which generates a coefficient ( $r$ ) as a measure of the linear dependence between two variables  $x$  and  $y$ , giving a value between +1 and –1 inclusive. It is widely used as a measure of the strength of linear dependence between variables. The ActaMatch search function is based on the Pearson’s Coefficient ( $r$ ) as a measure of the similarity of the pattern of Me shifts between the investigated compound and the compounds in the database (Eq. 11).

$$r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \quad (11)$$

where  $x$  and  $y$  denote the Me shifts of the investigated compound and any compound in the database, respectively. By entering the Me shifts of the investigated compound and an appropriate  $r$  value ( $r_0$ ), the VB-coded program determines which compound(s) in the database fulfill  $r \geq r_0$  and lists the hit(s) on the output page sorted by  $r$  values.

**ActaMatch**

Go Reset Export Print Help

**INPUT**

Methyl groups	Doublet
0.99	<input type="checkbox"/>
1.08	<input type="checkbox"/>
1.32	<input type="checkbox"/>
1.33	<input checked="" type="checkbox"/>
1.41	<input type="checkbox"/>
1.71	<input type="checkbox"/>
1.79	<input type="checkbox"/>
2.14	<input type="checkbox"/>
	<input type="checkbox"/>
	<input type="checkbox"/>
Cyclopropane	Doublet

**PARAMETERS**

Field strength: 700MHz Formula: C H O N  
 NMR solvent: C5D5N Sugar: arabinopyranose  
 r (similarity): 0.9980 \* must fill in

**RESULTS**

4 compounds were found!

Comparison chart showing Input (red) and Hit #3 (blue) peaks.

**MATCH LIST**

Name (point to see)	Field	Solvent	Formula	H-19	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	Me9	Me10	Sugar	Similarity
Cimiracemoside	500MHz	C5D5N	C37H56O11	0.50, 1.01	1.02 30	1.08 28	1.33 21	1.34 29	1.41 18	1.70 27	1.78 26	2.13 12-OAc	-	-	xy(pyr) 3	0.9997
Cimiracemoside	500MHz	C5D5N	C37H56O11	0.48, 1.00	1.02 30	1.06 28	1.31 21	1.33 29	1.40 18	1.69 27	1.77 26	2.10 12-OAc	-	-	xy(pyr) 3	0.9994
Cimiracemoside	500MHz	C5D5N	C37H56O11	0.47, 1.00	0.97 30	1.05 28	1.28 29	1.31 21	1.39 18	1.68 27	1.76 26	2.11 12-OAc	-	-	ara(pyr) 3	0.9998
Actaeaeopoxide	600MHz	C5D5N	C37H56O11	0.50, 0.98	1.03	1.07	1.31	1.33	1.41	1.69	1.77	2.15 12-OAc	-	-	xy(pyr) 3	0.9988

**Figure 37. A Screenshot and Operational Procedures of ActaMatch Using Compound 12 (Cimiracemoside G) as an Example**

Step ①: The methyl shifts are entered in an ascending order; Step ②: Pearson's coefficient ( $r$ ) is set as a threshold for the search; Step ③: The **Go** button is clicked to run the search program; Step ④: The number of hits are shown in the results box; Step ⑤: The details of the hits are listed. The unmatched parameters entered in step ② are indicated in red. The comparison chart is displayed in step ④ by clicking each row.

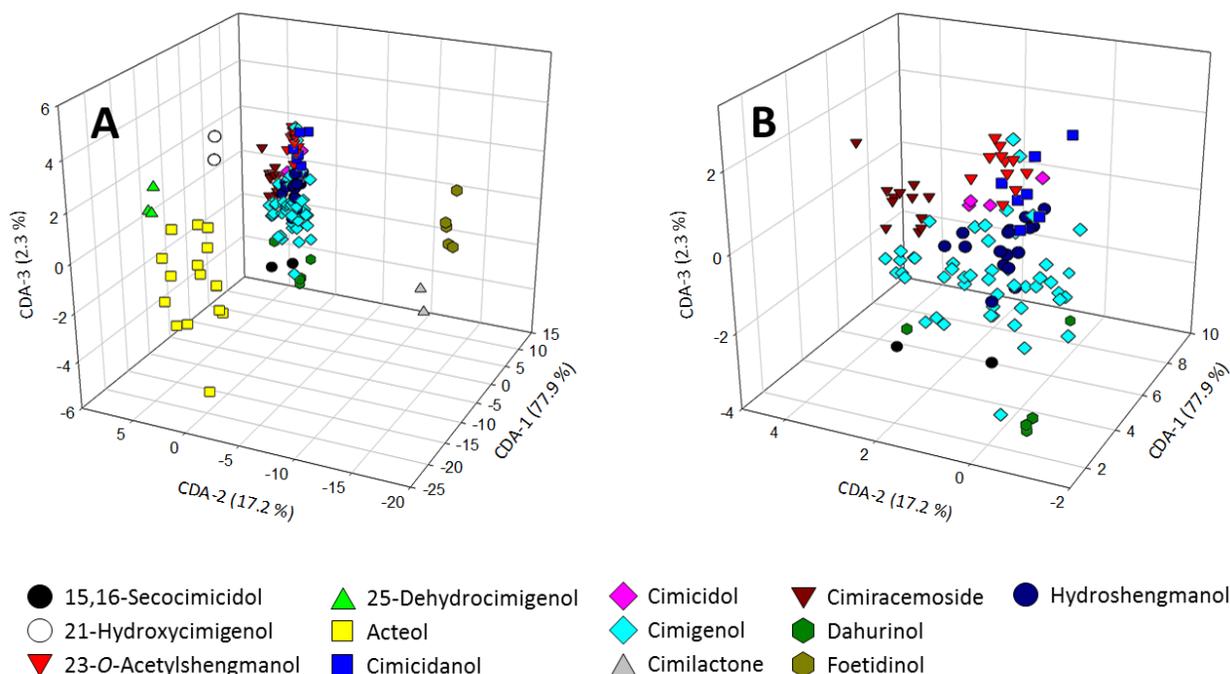
## 5.4 Canonical Discriminant Analysis

A canonical discriminant analysis (CDA) was initially performed for all compounds in the database using their Me  $\delta_H$  values (Me1, Me2, etc.) in ascending order. All the Me groups with  $\delta_H < 1.90$  were used in this analysis as Me signals with  $\delta_H > 1.90$  are either acetoxy (OAc) or methoxy groups (OMe) which are not essential base structural components for *Actaea* triterpenes. In order to create a dimensionally homogenous data set for CDA analysis, compounds with less than seven analyzed Me groups ( $\delta_H$  0.70–1.90) are given extra variable(s) with value 0. The result for all compounds is visualized in **Figure 38**, representing a 3D CDA plot. The first factor (CDA-1) represents 77.9% variation in the original data, whereas CDA-2 and CDA-3 account for 17.2 and 2.3% variation, respectively. All three factors explain a total of 97.4% of cumulative variance in a highly significant analysis (Wilks'  $\lambda = 0.00$ ,  $F_{\text{approx}} = 35.16$ ,  $df_1 = 84$ ,  $df_2 = 718$ ,  $P < 0.0001$ ).

**TABLE VIII. CLASSIFICATION RESULTS FOR ACTAEA TRITERPENES CONTAINING SEVEN METHYL GROUPS WITH  $\delta_H < 1.90$  USING THE CDA ANALYSIS**

Type <sup>a</sup>	Total	Correct (%)	Classified Type							
			SE	CR	CO	CG	AS	CA	HS	DA
SE	2	100	<b>2</b>	0	0	0	0	0	0	0
CR	12	91.7	0	<b>11</b>	0	1	0	0	0	0
CO	4	75.0	0	0	<b>3</b>	0	0	1	0	0
CG	50	84.0	0	1	0	<b>42</b>	2	0	4	1
AS	11	100	0	0	0	0	<b>11</b>	0	0	0
CA	7	28.6	0	0	0	0	4	<b>2</b>	1	0
HS	16	81.2	0	0	0	3	0	0	<b>13</b>	0
DA	5	100	0	0	0	0	0	0	0	<b>5</b>

<sup>a</sup> 15,16-Secocimicidol (SE); Cimircemoside (CR); Cimicidol (CO); Cimigenol (CG); 23-O-Acetylshengmanol (AS); Cimicidanol (CA); Hydroshengmanol (HS); Dahurinol (DA).



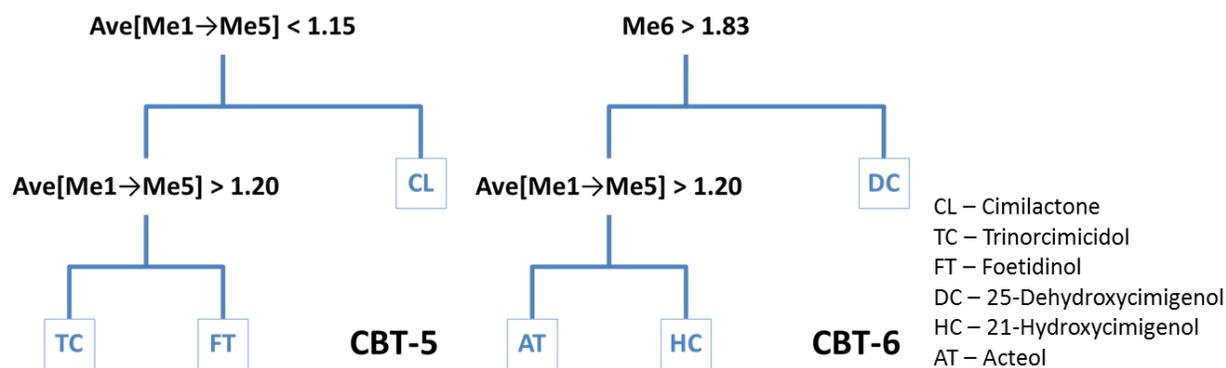
**Figure 38. Classification of All *Actaea* Triterpenes Contained in the In-house Database by CDA Analysis**

The 3D plot (panel A; axes CDA-1 = 77.9%, CDA-2 = 17.2%, CDA-3 = 2.3%) shows that the first 3 factors account for 97.4% of the total variance in the Me shifts of the compounds. Panel B shows the sub-cluster of all triterpenes with seven Me groups ( $\delta_{\text{H}} < 1.90$ ) having CDA-1 scores between 0 and 10, which form further sub-clusters depending on the specific skeleton types.

The majority of triterpenes (~80%) that have seven Me groups ( $\delta_{\text{H}} < 1.90$ ) are clustered in a space shown in **Figure 38B** and their classification results are listed in **Table 8**. Relying only on the variances of the Me shifts, all the *Actaea* triterpenes in the database can be classified with an overall correct rate of 86.9% by the model derived from CDA analysis. Considering inescapable variations of reported  $^1\text{H}$  chemical shift information due to inconsistencies in, e.g., temperature and calibration (TMS vs. residual solvent), the discriminative power of the model could be further improved in the future by using a standardized NMR acquisition protocol.

## 5.5 Development of Classification Binary Trees

CART is a machine learning technique ideal for large and unbalanced data sets with many descriptors (Steinberg *et al.*, 2006). It generates a tree-like graph or model as a binary-decision support tool to identify the origin or class of the samples. In order to build a more accurate classification model for *Actaea* triterpenes, the classification binary trees (CBTs) were developed by CART analysis and used to partition the compounds into structurally similar clusters of aglycones. In order to avoid overfitting the data, the trees were appropriately pruned while the prediction accuracy was maintained at the optimum.

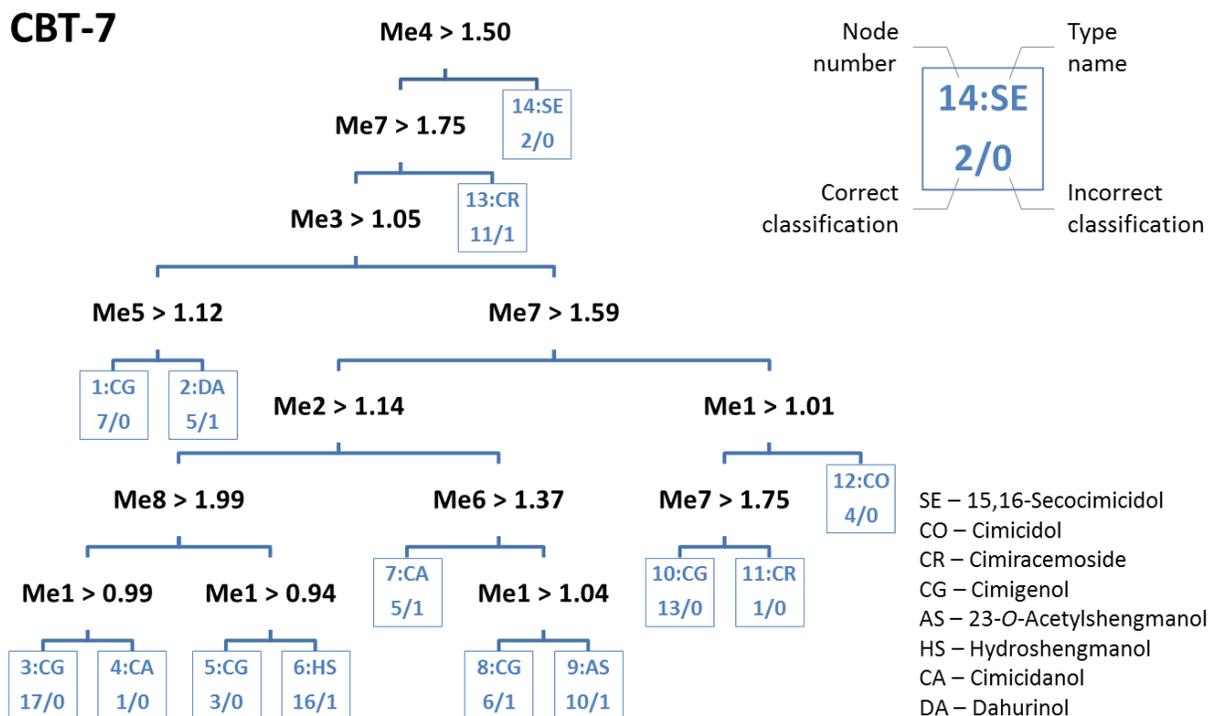


**Figure 39. The CBTs Developed for the Classification of *Actaea* Triterpenes with Five (CBT-5) and Six (CBT-6) Me Groups ( $\delta_{\text{H}} < 1.90$ )**

Ave[Me1→Me5] denotes the average of all five Me shifts (Me1 to Me5). In case the answer to a given descriptor/splitter is yes, it branches to the right child node.

The compounds in the database were initially divided into three subgroups according to the number of skeletal Me groups (five, six, or seven;  $\delta_{\text{H}} < 1.90$ ) within the molecules. **Figure 39** shows the resulting CBT from CART analysis to classify *Actaea* triterpenes with five (CBT-5) or six Me groups (CBT-6) by using their Me shifts. Both CBTs consist of three terminal nodes

(leaves) and two non-terminal nodes. From the top (root) of the tree, the compounds were split into groups according to the Me shifts used as descriptors at each node. Using these two CBTs, all of the triterpenes with five or six Me groups in the database were correctly classified. Similarly, the CBT for the classification of *Actaea* triterpenes with seven Me groups (CBT-7) is depicted in **Figure 40**, which is characterized by 14 terminal nodes and 13 non-terminal nodes, with an overall success rate of 94.4%. The percentage of correct classification for each structural subtype is shown in **Table 9**.



**Figure 40. The CBTs Developed for the Classification of *Actaea* Triterpenes with Seven Me Groups ( $\delta_H < 1.90$ )**

In CART analysis, the importance of a variable is usually determined by looking at every node in which a variable appears and taking into account its suitability as a splitter. The

importance score of the variables used in the generation of CBT-7 was calculated using the Salford Predictive Miner as follows: Me7 100.00, Me6 98.44, Me4 68.19, Me1 64.78, Me5 62.80, Me3 57.30, Me2 50.71, and Me8 29.20. These scores reflect the contribution of each Me signal to the classification of *Actaea* triterpenes, with the contribution stemming from all the variables' roles as primary splitters and as surrogates to any of the primary splitters. Here, Me7 and Me6 are ranked as the two most important. More than 75% of the Me7 and Me6 protons are assigned to either H-26 or H-27. Both of these Me groups are located in the aglycone side chain, which often cyclizes with C-16, and, thus, are highly indicative of the major structural differences of *Actaea* triterpenes. This explains why Me7 and Me6 are the most important indicators of the aglycone type. By further looking at the resulting CBT, Me7 is found to be the major classifier for cimracemosides (nodes #11 and #13) and cimigenols (node #10) with an OAc group at C-25. This is highly consistent with their structural characteristics. Under the neighboring effect of a 25-epoxy function, both the H-26 and H-27 signals of cimracemosides shift downfield to  $\delta_{\text{H}} > 1.75$ . Acetylation of the OH group at C-25 has been seen only in cimigenols, which results in H-26 shifting to the range of  $\delta_{\text{H}} 1.59\text{--}1.75$ . The variable Me4 can be used to classify 15,16-secocimcidols: their Me4 protons are either H-18 or H-27 with apparent  $\delta_{\text{H}} > 1.50$ . Me1 is ranked as the fourth most important classifier, covering ~50% of the investigated compounds which are cimigenols (nodes #3 and #5), cimicidols (node #12), cimicidanols (node #8), 23-O-acetylshengmanols (node #16), and hydroshengmanols (node #6). Consistently, the protons of their Me1s are all H-21, which is also a Me group in the side chain of the aglycone. While any of the first four important variables cannot distinguish cimigenols (node #1) and dahurinols (node #2), Me5 works well to differentiate these two types of compounds. The two Me groups, Me3 and Me2, show much less importance because the underlying protons are the geminal Me groups at C-4 (H-29 and H-30), which are located in the least structurally diverse region of the aglycones. However, the differences in their chemical

shifts, regardless of their assignment, are also useful to split the compounds into subgroups which can be further classified using other discriminating Me groups.

**TABLE IX. CLASSIFICATION RESULTS FOR ACTAEA TRITERPENES CONTAINING SEVEN METHYL GROUPS WITH  $\delta_H < 1.90$  USING THE CBT-7 PARTITIONING**

Type	Total	Correct (%)	Classified Type							
			SE	CR	CO	CG	AS	CA	HS	DA
SE	2	100	<b>2</b>	0	0	0	0	0	0	0
CR	12	100	0	<b>12</b>	0	0	0	0	0	0
CO	4	100	0	0	<b>4</b>	0	0	0	0	0
CG	50	92.0	0	1	0	<b>46</b>	1	0	1	1
AS	11	90.9	0	0	0	0	<b>10</b>	1	0	0
CA	7	85.7	0	0	0	1	0	<b>6</b>	0	0
HS	16	100	0	0	0	0	0	0	<b>16</b>	0
DA	5	100	0	0	0	0	0	0	0	<b>5</b>

Interestingly, by using CBT-7, in the terminal nodes #1, 3, 5, 8, and 10, the 40 known cimigenols are partitioned into five subgroups, and each group has its own structural characteristics. All 13 cimigenols in node #10 have an OAc group at C-25. Four of the seven cimigenols in node #1 have an OAc at C-25 and an extra OAc within the sugar moiety. Cimigenols in node #5 either have an OAc at C-12 or an OMe at C-25. As a matter of fact, it is easy to distinguish OAc and OMe according to the Me chemical shift: The signals for OAc are usually observed at  $2.0 \pm 0.2$  ppm, while OMe groups resonate at  $3.2 \pm 0.2$  ppm. In addition, six cimigenols classified in node #8 have an OH group at C-12. However, all 17 cimigenols in node #3 are free of any OAc or OMe within the aglycone. These results indicate that the presence

and position of OH, OAc, and OMe in cimigenols may also be determined based on the Me shifts.

In addition to this dereplication capability, the CBT models have potential to predict the aglycone type of unknown *Actaea* triterpenes yet to be discovered. Owing to limited data available, leave-one-out cross-validation (LOOCV) was used to estimate the accuracy of the predictions. Overall, the predictions are 80.4% correct for CBT-7. As summarized in **Table 10**, cimigenols, cimiracemosides, and hydroshengmanols, which comprise the majority of compounds in the database, have excellent prediction rates of 80.0, 91.7, and 100%, respectively. For 23-O-acetylshengmanols, three of 11 (i.e., 27.3%) are incorrectly predicted as cimicidanols. This is understandable because both compound types have the same epoxide group at C-24 and C-25, leading to difficulties in differentiating them by the terminal Me groups in the side chain. The minor classes of *Actaea* triterpenes, including cimicidanols, cimicidols, and dahurinols, are 50–60% correctly predicted. Despite their structural similarity with other types of *Actaea* triterpenes, the use of additional descriptors, such as the chemical shifts of cyclopropane methylene protons (H-19a/b) and/or other protons, will likely improve predictive accuracy.

**TABLE X. THE ACCURACY OF THE PREDICTION PERFORMANCE OF THE CBT-7 ESTIMATED BY LEAVE-ONE-OUT CROSS-VALIDATION**

Type	Total	Correct (%)	Predicted Type							
			SE	CR	CO	CG	AS	CA	HS	DA
SE	2	100	<b>2</b>	0	0	0	0	0	0	0
CR	12	91.7	0	<b>11</b>	0	1	0	0	0	0
CO	4	50.0	0	0	<b>2</b>	0	1	1	0	0
CG	50	80.0	0	1	1	<b>40</b>	2	4	1	1
AS	11	72.7	0	0	0	0	<b>8</b>	3	0	0
CA	7	57.1	0	0	0	1	2	<b>4</b>	0	0
HS	16	100	0	0	0	0	0	0	<b>16</b>	0
DA	5	60.0	0	0	0	2	0	0	0	<b>3</b>

The CBT models were automated by a VBA program named ActaPredict (Appendix D, p. 188) using a separate worksheet within the Excel database file (**Figure 41**). In both ActaMatch and ActaPredict, additional functions were implemented in order to create a user-friendly operating environment. For example, an input check can be initially performed. When errors are found in the data queries, further program execution is aborted with display of a warning message. Furthermore, two command buttons were added, one of which was assigned with the corresponding VB program while the other was coded for data reset. Finally, the ActaMatch and ActaPredict modules together with the spreadsheet database were integrated into an application suite within a macro-enabled Excel file named ActaFinder.

ActaPredict

Go Reset Export Print Help

**INPUT**

CH3	$\delta$ (H, ppm)
Me1	0.99
Me2	1.08
Me3	1.32
Me4	1.33
Me5	1.41
Me6	1.71
Me7	1.79
Me8	2.14

Note: Me1 < Me2 < ... < Me8

CH2	$\delta$ (H, ppm)
H-19a	
H-19b	

**RESULTS**

New systematic name  
**Acta-16,23;22,25-binoxol**

Structure

Note : Your H-19 data is incomplete or incorrect!

**RESULTS**

Aglycone type	7,8-ene	Hydroxyl groups					Acetyl/methoxyl groups				Sugar moiety
		1-	3-	7-	12-	15-	12-	23-	24-	25-	
Cimiracemoside	?						-OAc				

**Figure 41. A Screenshot and Operational Procedures of ActaPredict Using Compound 12 (Cimiracemoside G) as an Example**

Step ①: The methyl shifts are entered in an ascending order; Step ②: The **Go** button is clicked to run the program; Step ③: The structure can be predicted and shown in the results box; Step ④: The substituents and their positions can be also identified.

## 5.6 Dereplication of *Actaea* Triterpenes in Residually Complex Mixtures

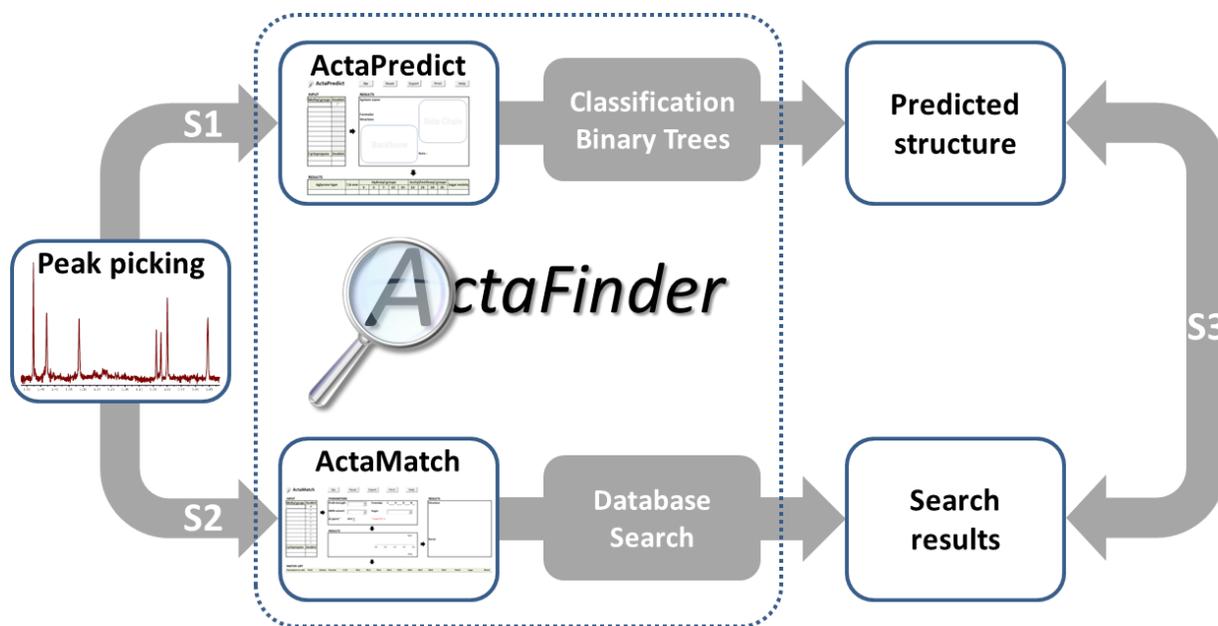
Traditionally, the complexity of NP mixtures makes it a challenge to identify the components by full interpretation of NMR spectra. In the present study, the concept of using only Me shifts for the dereplication of multicomponent mixtures, such as residually complex (impure) mixtures of triterpenes, has the particular advantage that Me resonances are usually singlets of relatively high intensity. While the Me groups resonate in the same range of 0.8–2.0 ppm as several aliphatic methines and methylenes, the signals of the latter are much more complex and their intensities distribute over a much broader range due to  $J$  coupling. In approximation, comparing a *ddd* methylene (1H) with a singlet Me (3H) signal, the individual spectral lines of the former are ~25 fold lower in intensity. Accordingly, Me groups associated with minor triterpenes of more than 4% become visible even in overlapped regions of the spectra. While chemical shift dispersion limits the number of components for which all Me signals can be identified, 2D NMR methods prove useful in the further unraveling of this “hidden” spectroscopic information.

This study establishes an *in silico* dereplication approach to identify *Actaea* triterpenes in both pure forms and residually complex mixtures by using a combination of the CBTs and database search, using the following three steps (**Figure 42**).

**Step 1 [ActaPredict]:** The Me shifts are analyzed by the CBTs, and the triterpene skeleton is determined; the substituents on the skeleton as well as the sugar moieties are identified by the presence of characteristic  $^1\text{H}$  NMR signals.

**Step 2 [ActaMatch]:** The  $^1\text{H}$  NMR data of Me groups are also used to search the hits with  $r \geq r_0$ , where  $r_0$  is the threshold of similarity defined by the user; based on our experience, exact hits have  $r$  values  $> 0.998$ .

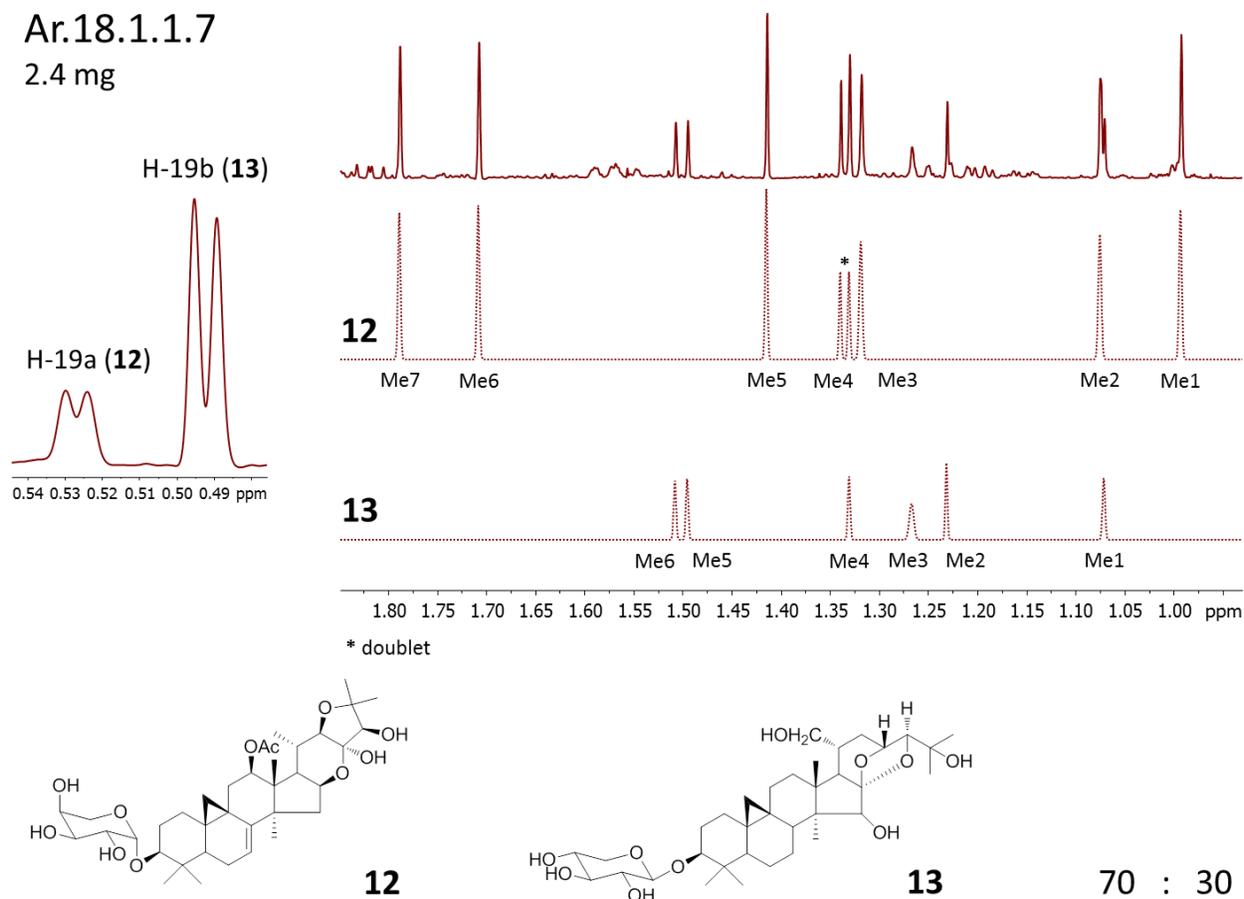
**Step 3:** The results from steps 1 and 2 are compared for consistency.



**Figure 42. The New *In Silico* Tool ActaFinder**

The ActaFinder is comprised of two modules ActaPredict (step S1) and ActaMatch (step S2). In a third step (S3), the results of S1 and S2 are compared for consistency. This approach can potentially be adopted for other NPs using characteristic and readily accessible  $^1\text{H}$  chemical shift information, such as that of Me groups.

Two examples (Ar.18.1.1.7, Ar.18.1.3.2.4) of residually complex triterpene reference materials were chosen to illustrate this approach. Sample Ar.18.1.1.7 (2.4 mg) was produced in the 4<sup>th</sup> step of fractionation of the black cohosh extracts (see **Figure 10**). It was initially subjected to  $^1\text{H}$  NMR analysis using the conditions stated in the Experimental Section 2.2.3 (p. 32). Each *Actaea* triterpene gives rise to a pair of doublets in the range of  $\delta_{\text{H}}$  0.2–1.0, due to its cyclopropane methylene protons, H-19a/b. Based on these characteristic signals, it was determined that this sample contained two major triterpenes (**12**, **13**) (**Figure 43**). The Me signals of each triterpene were readily recognized based on their integral values relative to the individual H-19 signals. The overlapped Me signals (Me2 of **12** and Me1 of **13** at 1.08 ppm) were deconvoluted by using the Line Fitting function in MestReNova software, and the individual spectra of the two triterpenes were extracted from the  $^1\text{H}$  NMR spectrum of the mixture.



**Figure 43. Dereplication of *Actaea* Triterpenes in a Residually Complex Sample (Ar.18.1.1.7)**

The sample was dissolved in 35  $\mu\text{L}$  of pyridine- $d_5$  in a 1.7 mm NMR tube. The  $^1\text{H}$  NMR spectra were recorded on a 700 MHz NMR spectrometer equipped with a 1.7 mm cryo-microprobe using the following acquisition parameters: TD 32 K, SW 14423 Hz, AT 2.3 s, NS 128, and RG 57. The exact composition of Ar.18.1.1.7 was analyzed by a combination of  $^1\text{H}$  NMR spectral deconvolution, CBT partitioning of Me  $^1\text{H}$  chemical shifts, and characteristic  $^1\text{H}$  NMR sugar signals. The results showed that Ar.18.1.1.7 exhibited moderate residual complexity, which is frequently found with *Actaea* triterpene reference materials, and can be considered a “clean” 70:30 mixture of the two triterpenes **12** and **13**.

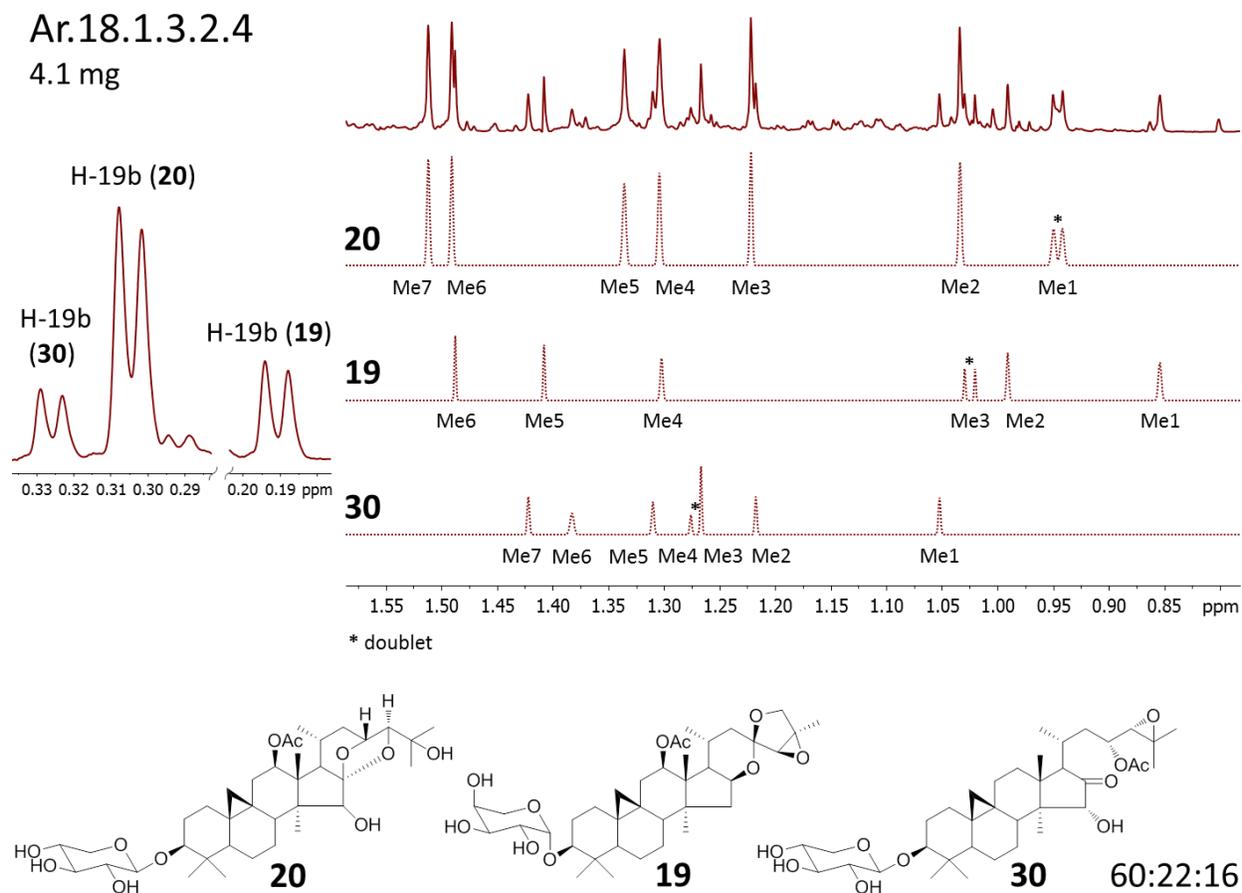
Compounds **12** and **13** have seven and six Me groups with  $\delta_{\text{H}} < 1.90$ , respectively. By using the CBT-7 partitioning in ActaPredict, **12** was dereplicated as an acta-16,23;22,25-binoxol, formerly often designated as cimracemoside. An additional Me signal at 2.14 ppm indicates that an OAc may be present at C-12, a position which is commonly acetoxyated in actabinoxols. Its

H-19a signal was observed at 0.98 ppm, indicating the presence of a  $\Delta^{7,8}$ -double bond. Close inspection of the region for the sugar moieties (3.7–5.0 ppm) identified a characteristic *dd* signal (11.9, 1.4 Hz, H-5'b) of arabinopyranose (*arap*) at 3.790 ppm, bearing the same integral as H-19b of **12**. Compound **13** was dereplicated as the xylopyranoside of a 21-hydroxylated acta-16,23;16,24-binoxol, formerly classified as 21-hydroxycimigenol, by CBT-6 partitioning. This was substantiated by the lack of a doublet among the Me signals due to hydroxylation of the Me at C-21. A characteristic *dd* signal (11.2, 9.8 Hz, H-5'b) of xylopyranose (*xylp*) was observed at 3.755 ppm, exhibiting the same integral as H-19a of **13**. Summarizing all the dereplication results and further observations, the structures of **12** and **13** were identified as shown in **Figure 43**. In addition, because the  $^1\text{H}$  NMR spectra were acquired under quantitative conditions (qHNMR), their molar ratio was determined to be 70:30 from the integrals of their H-19a/b signals.

The residually complex sample Ar.18.1.3.2.4 (4.1 mg) was the same as used for Case 1 in Section 4.5 (p. 92). In the  $^1\text{H}$  NMR spectrum of this fraction, the H-19a/b signals indicated that its composition is more complicated, with at least six minor triterpenes being present along with the main component, **20**. Initial identification targeted the major constituents **20**, **19**, and **30**, which had contents of more than 10 mol% and allowed full Me deconvolution: Based on the integral of their H-19b signals, individual Me signals were identified and extracted from the  $^1\text{H}$  NMR spectrum by deconvolution of the overlapped peaks. Compound **20** was dereplicated as an acta-16,23;16,24-binoxol (formerly: cimigenol) in node #5 of the CBT-7. A Me signal at 2.14 ppm further indicated that **20** is acetylated at C-12. Compound **19** was dereplicated as a 24,25-epoxy derivative of an acta-16,23;23,26-binoxol (formerly: acteol) with an OAc (2.15 ppm) at C-12. Compound **30** was dereplicated as the 23-acetate of a 16-oxo-actanol (formerly: 23-O-acetylshengmanol). The fact that none of the H-19a signals is shifted downfield to  $\sim 1.00$  ppm indicated that all three triterpenes are saturated at C-7 and C-8. A characteristic *dd* signal (11.9, 1.4 Hz, H-5'b) of arabinopyranose (*arap*) was observed at 3.832 ppm, exhibiting the same

integral as H-19a of **19**. Two overlapped *dd* signals were observed at 3.730 and 3.744 ppm, and both are characteristic for H-5'b of *xylp*. Their integrals were identical with those of H-19b of **20** and **30**, respectively. Therefore, the structures of **20**, **19**, and **30** were identified as shown in **Figure 44**. Their molar ratio was measured by qHNMR as 62:22:16 based on the integral of their H-19b signals.

While the aforementioned general considerations put the threshold of identifiable Me signals (vs. overlapping CH<sub>2</sub>/CH protons) around the 5% level, we were still able to tentatively identify *R*- and *S*-actein [(12*R*)-12-acetoxy-(24*R*,25*S*)-24,25-epoxy-(26*R*&*S*)-26-hydroxy-3-*O*-β-D-xylopyranosylacta-(16*S*,23*R*)-16,23;23,26-binoxoside] as two minor constituents of Ar.18.1.3.2.4 (~3 and ~5% impurities, respectively). Evidence for this assignment came from both the CBT analysis and the characteristic <sup>23</sup>J HMBC cross peaks between the small Me-28 signals at 0.87 and 0.80 ppm and the bridgehead carbons C-8/13/14, which all resonate in the narrow range ~44–46 ppm. The dereplication results are consistent with those obtained by the HMBC pattern recognition approach (see Section 4.5, p. 92). The successful dereplication of five triterpenes including three minor triterpenes in a 4.1 mg sample also demonstrates the power of cryo-microprobe NMR analysis of residually complex NPs.



**Figure 44. Dereplication of *Actaea* Triterpenes in a Residually Complex Sample (Ar.18.1.3.2.4)**

Using the analogous approach as for sample Ar.18.1.1.7, analysis of Ar.18.1.3.2.4 led to the identification of three major triterpenes, **20**, **19** and **30**, in this residually rather complex mixture. Because these compounds were present in a 62:22:16 ratio, their Me signals were readily distinguished and amenable to CBT dereplication. Interestingly, the minor impurities giving rise to Me singlets at 0.87 and 0.80 ppm could be assigned to *R*- and *S*-actein (**15**), at ~3 and ~5 mol% abundance, respectively, using their known Me-28 chemical shifts and characteristic HMBC coupling patterns.

In order to verify the dereplication results by the CBTs, the  $^1\text{H}$  NMR data of the Me groups of individual triterpenes identified in the mixture samples Ar.18.1.1.7 and Ar.18.1.3.2.4 were searched by ActaMatch. The results are shown in **Table 11** and indicate that all the investigated compounds are highly correlated with their best hits ( $r > 0.998$ ). It is noteworthy that the triterpenes with the same aglycone but different sugar moieties may exhibit a high correlation with  $r > 0.999$  in terms of the  $^1\text{H}$  NMR properties of Me groups. For example, adding to cimracemoside G, two hits, both of which are cimracemoside F data from two different sources, matched compound **12** with high  $r$  values of 0.9994 and 0.9997, respectively. For cimracemosides F and G, the different sugar moieties *xylp* vs. *arap* have only a negligible chemical shift effect on the H-29 and H-30 Me groups, which results in a minor difference in the  $r$  value. However, inconsistencies in the NMR experimental conditions of reported data may also contribute to this minor difference. As a result, rather than identifying matches solely on the basis of correlation ranking, glycosides often require verification on the basis of characteristic sugar signals which are readily available. Compound **30** is another good example to illustrate this concept: whereas the best match to **30** was an arabinopyranoside with  $r = 0.9997$ , the sugar was identified as xylopyranose based on the characteristic  $^1\text{H}$  NMR signals.

**TABLE XI. DEREPLICATION RESULTS OF THE TRITERPENES CONTAINED IN THE RESIDUALLY COMPLEX SAMPLES Ar.18.1.1.7 AND Ar.18.1.3.2.4**

Compound	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	<i>r</i>
Sample Ar.18.1.1.7									
<b>12<sup>a</sup></b>	0.99	1.08	1.32	1.33 <sup>c</sup>	1.41	1.71	1.79	2.14	0.9998
cimiracemoside G <sup>b</sup> (ara)	0.97	1.05	1.28	1.31 <sup>c</sup>	1.39	1.68	1.76	2.11	
<b>13</b>	1.08	1.23	1.27	1.34	1.50	1.51	–	–	0.9975
21-dehydrocimigenol xyl <sup>b</sup>	1.04	1.20	1.24	1.28	1.46	1.48	–	–	
Sample Ar.18.1.3.2.4									
<b>20<sup>a</sup></b>	0.94 <sup>c</sup>	1.03	1.22	1.30	1.34	1.49	1.51	2.14	0.9992
12-acetoxycimigenol xyl <sup>b</sup>	0.92 <sup>c</sup>	0.98	1.21	1.25	1.31	1.47	1.49	2.10	
<b>19</b>	0.85	0.99	1.03 <sup>c</sup>	1.30	1.41	1.49	2.15	–	0.9994
cimiracemoside N <sup>b</sup> (ara)	0.85	0.96	1.02 <sup>c</sup>	1.27	1.42	1.48	2.14	–	
<b>30</b>	1.05	1.22	1.27	1.28 <sup>c</sup>	1.31	1.38	1.42	2.07	0.9997
23-O-acetylshengmanol <sup>b</sup>	1.05	1.21	1.25	1.26 <sup>c</sup>	1.30	1.37	1.40	2.06	

<sup>a</sup> Main component.

<sup>b</sup> Triterpene with the best match (*r*).

<sup>c</sup> Indicates Me-21 doublets.

## 5.7 Conclusion

This study introduces two new tools for the efficient study of triterpenes present in *Actaea*, a genus extensively used in complementary and alternative medicine and a major source of these NPs. A new semi-systematic naming scheme links the compound name to the molecular structure, and a rapid dereplication tool utilizes the readily available information from the  $^1\text{H}$  NMR chemical shifts of Me groups as well as an in-house database. A rational naming scheme plays an important role in dereplication, because unambiguous compound names provide crucial links between the literature and the molecular structures. *Actaea* triterpenes served as examples to demonstrate how these tools can be developed and utilized in practice. By using the Me shifts as indicators of structural characteristics, two classification models based on CDA and CBTs were generated for the *in silico* classification of *Actaea* triterpenes according to their aglycone type. This concept has potential to be adopted for any other class of NPs with characteristic and readily accessible chemical shift information, such as that of Me groups.

Both CDA and CBTs exhibit high accuracy when classifying the *Actaea* triterpenes in our in-house database using only Me chemical shifts. Comparing these two methods in practical use, CBTs are more straightforward, simple to understand, and to interpret. The CBT model can be easily implemented in a procedural computer algorithm such as the VBA code. Therefore, CBTs are not only efficient in the dereplication of triterpenes, as shown, but are also a promising dereplication tool for other NPs, such as steroids, other terpenoids, and peptides.

Using a combination of characteristic  $^1\text{H}$  Me shifts (Me-28) and  $^{2/3}J_{\text{C,H}}$  HMBC coupling patterns, we were able to tentatively assign *R*- and *S*-actein as two minor constituents in the sample Ar.18.1.3.2.4 (**Figure 43**), present only at the 3–5% level. While further results will be reported in due course, it is safe to conclude that the approach presented, combined with contemporary (q)NMR methodology using 700 MHz 1.7 mm cryo-microprobe equipment, has potential for the standard characterization of residual complexity of NP reference materials,

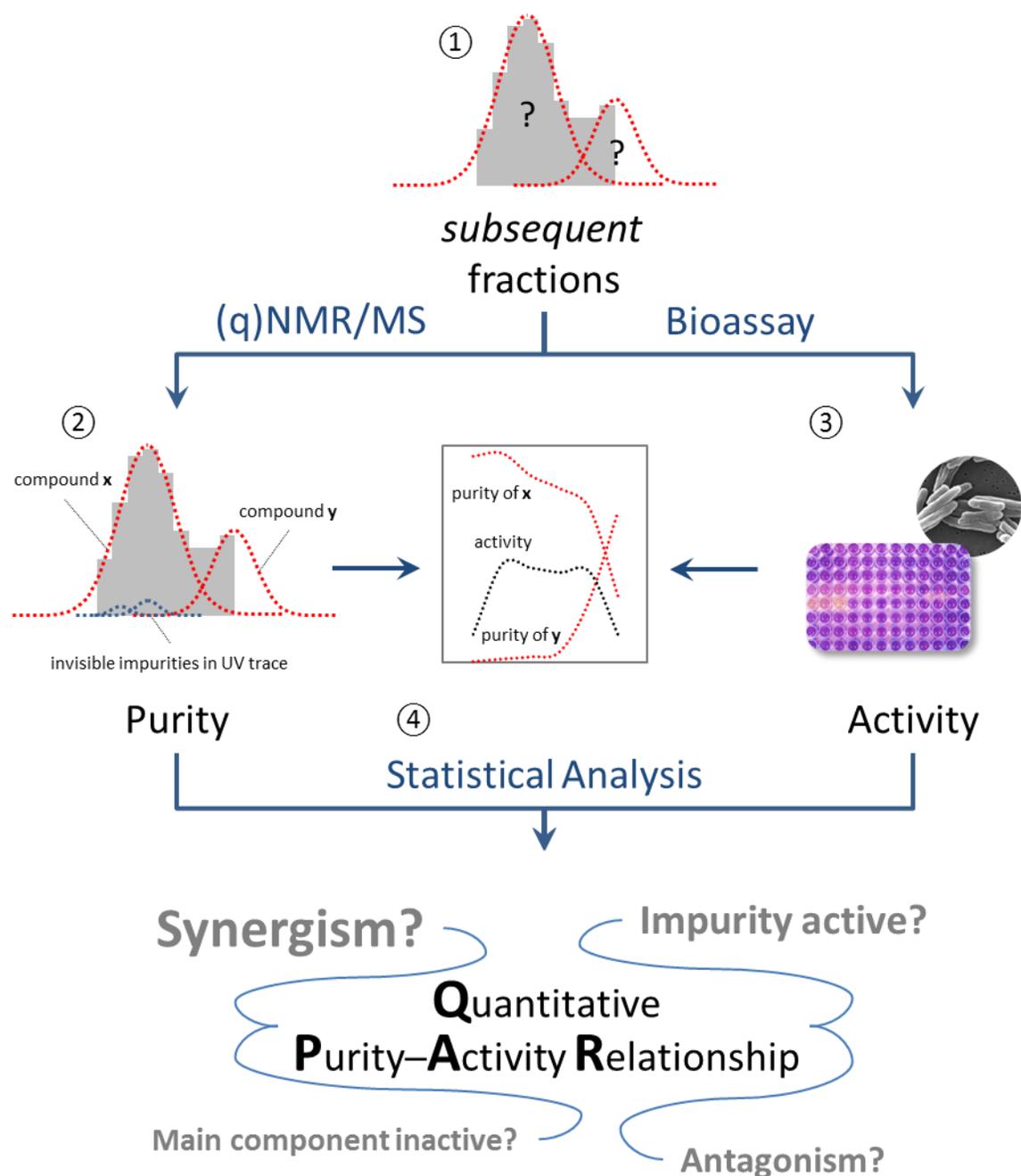
allowing analysis of several minor constituents down to the 10–20  $\mu\text{g}$  level in a 1–5 mg sample. Recently, the power of HSQC in the analysis of complex mixtures has been shown (Lewis, 2007; Xi, 2008). Future studies will also adopt HSQC-DEPT which is not only more sensitive than other 2D  $^1\text{H}$ – $^{13}\text{C}$  experiments, but also provides an extra dimension by tying the  $^{13}\text{C}$  chemical shifts to the appropriate  $^1\text{H}$  chemical shifts for the methyl groups. HSQC-DEPT is particularly useful in determining and differentiating the methyl groups of individual triterpenes in complex mixtures and, thus, improves the dereplication process.

In our experience, even repeatedly purified reference materials of biosynthetically diverse NPs such as triterpenes often exhibit surprisingly high degrees of residual complexity. In this regard, the case of Ar.18.1.3.2.4 is particularly noteworthy, because it shows that constitutionally and spatially distinct NPs can exhibit similar or even identical chromatographic behavior, even in multi-step purification procedures. This study identified Ar.18.1.3.2.4 as a mixture of more than five compounds which belong to at least four different skeleton types: one acta-16,23;16,24-binoxol, one actanol, and three acta-16,23;24,26-binoxols belonging to two different subtypes. While the different abundance levels are important parameters of residual complexity, the observed co-occurrence of considerably different chemical species is of broader importance regarding bioactivity. First, the evaluation of the degree and pattern of residual complexity of purified NPs should be considered a prerequisite for their biological assessment. Second, assumptions about 3D structural similarities can potentially be misleading and have to be verified for each particular sample used in a bioassay. The dereplication tools introduced here, in combination with qualitative and quantitative  $^1\text{H}$  NMR analysis, might inspire future applications for a wider range of NPs.

## 6. INTEGRATING CHEMISTRY AND BIOLOGY OF NATURAL PRODUCTS

### 6.1 Quantitative Purity–Activity Relationship of Triterpenes from *O. horridus*

Typically, when a substance is pharmacologically evaluated, the assumption is made that the sample represents an SCE or a defined mixture of known chemical entities with known composition. However, when bioactive materials require isolation from complex matrices, they are likely to retain residual complexity even in a refined stage. The term, residual complexity, refers to the easily overlooked impurity profile of isolated NPs, which may exert a significant influence on their accurate biological assessment. Residual complexity can be static or dynamic, referring to impurity patterns that are either constant or fluctuating depending on conditions, respectively. A previous study has introduced the concept of purity–activity relationships (PARs) in NP research (Jaki *et al.*, 2008). The PAR profiles of NPs were generated by a correlation of qHNMR purity description and anti-TB biological data of different batches of the same natural product. However, as these samples originated from various research and commercial sources, the different impurity profiles of each sample resulted in unique PARs patterns which are incomparable between the samples for a quantitative analysis. The present study explores an extension of the PAR concept by enabling the quantitative evaluation of the effects of varying minor components (impurities) on the bioactivity of a purified NP. **Figure 45** shows an experimental design for the establishment of quantitative purity–activity relationships (QPARs) of NPs. Instead of using samples of different origin, the proposed approach uses subsequent chromatographic fractions for impurity profiling and biological assessment. In these fractions, the various major and minor components are present at different concentration levels following the individual elution profiles. Consequently, it is possible to correlate the concentrations of each component with the measured bioactivity of each fraction. This leads to a quantitative model which allows a comprehensive analysis of bioactivity in residually complex samples as it relates to individual components and their interactions.



**Figure 45. The Experimental Design for the Establishment of Quantitative Purity–Activity Relationships of NPs**

Step ①: The NPs are purified chromatographically, yielding a series of fractions (gray bars); Step ②: All fractions are subject to qNMR and/or MS analysis for qualitative and quantitative identification of major constituents as well as minor impurities; Step ③: In parallel, the fractions are evaluated by a high-throughput bioassay for the targeted bioactivity; Step ④: The chemical and biological data obtained in steps ② and ③ by statistical analysis generates a quantitative model that correlates the purity and activity of NPs.

The following example utilizes the antimycobacterial active triterpenes from Devil's club to illustrate the QPAR methodology. In Section 4.2.3 (p. 79), a series of HPLC fractions were acquired from a triterpene enriched sample.  $^1\text{H}$  NMR analysis and search of the literature revealed that the major components in the fractions Oh.17.8.19 to Oh.17.8.22 were the known triterpene,  $3\alpha$ -hydroxylup-20(29)-ene-23,28-dioic acid (tt 1, **25**), while Oh.17.8.23 and Oh.17.8.24 were 7:3 and 3:7 mixtures of **25** and its isomer,  $3\alpha$ -hydroxyolean-12-ene-23,28-dioic acid (tt 2, **26**). Minor quantities of **26** were also present in Oh.17.8.20 to Oh.17.8.22. The molar ratio of **25** and **26** in each fraction was determined from their characteristic signals at  $\delta_{\text{H}}$  3.052 (H-18, **25**) and  $\delta_{\text{H}}$  2.852 (H-17, **26**). On the basis of impurity signals observed in the  $^1\text{H}$  NMR spectra of these fractions, COSY analysis was employed and confirmed the presence of monoglyceride (imp 1, **27**) and polyene impurities (imp 2, **28**). Thus, using the 100% qHNMR method, the molar content of each component in all fractions was calculated (**Table 12**).

**TABLE XII. qHNMR IMPURITY PROFILES AND ANTI-TB ACTIVITY (IN MIC) OF THE *O. HORRIDUS* TRITERPENE FRACTIONS AGAINST STRAIN H<sub>37</sub>Rv**

Fr (Oh.17.8.X)	.19	.20	.21	.22	.23	.24
tt 1 (mol%)	94.9	96.3	85.8	79.9	70.8	29.3
tt 2	0.0	1.2	2.2	4.7	28.0	69.4
imp 1	5.1	2.5	6.1	7.7	0.6	1.3
imp 2	0.0	0.0	5.9	7.7	0.6	0.0
MIC ( $\mu\text{g/mL}$ ) <sup>a</sup>	220	118	121	125	124	226

<sup>a</sup> The highest test concentration was 256  $\mu\text{g/mL}$ .

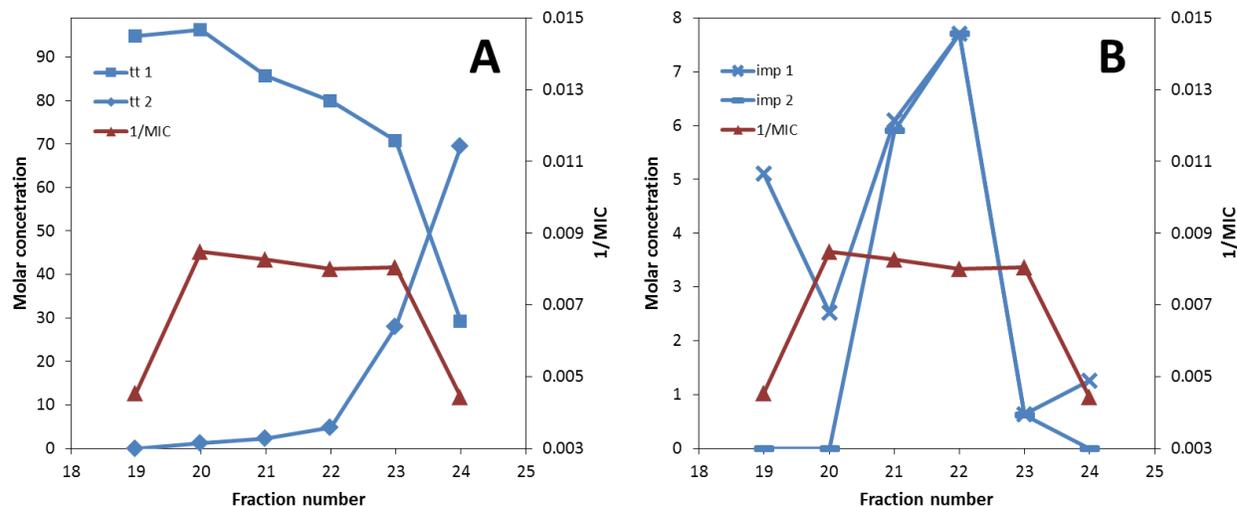
In addition to providing the results of purity evaluation, **Table 12** also summaries the anti-TB activity of the six samples determined by the microplate Alamar blue assay (MABA). In the present study, the MIC values ( $\mu\text{g/mL}$ ) are the lowest drug concentrations effecting an inhibition of 90% of *M. tuberculosis*. As a lower MIC is an indication of a better activity, the inverse of MIC ( $1/\text{MIC}$ ) was used here to represent a quantitative measure of anti-TB activity. These data led to the establishment of QPARs for the anti-TB active fractions as visually illustrated in **Figure 46**. Apparently, curve shape and progression demonstrate that the activity and purity of the triterpene **25** were not proportional, nor were they correlated in a linear or logarithmic fashion. Therefore, the activity must be attributed to both the major components and the impurities or to the impurities only. It is further observed that the anti-TB activity of **25** appeared to be inhibited by the presence of monoglycerides. However, this inhibition was counteracted by the co-occurrence of polyene analogues representing minor impurities in the triterpene fractions.

In order to quantitatively evaluate the effects of two impurities on the observed activity of **25**, the relationship between sample activity and purity can be further correlated in a mathematical model. For this study, two conditions were initially hypothesized: (1) the concentration and activity of any component as an SCE follows a linear relationship; (2) the combination of activities of all components in the sample is only additive. Based on these two hypotheses, linear regression analysis becomes an appropriate method for the establishment of QPARs. As a result, the anti-TB activity of **25** ( $1/\text{MIC}$ ) can be expressed as a linear function of molar contents of three impurities **26**, **27**, and **28**:

$$1/\text{MIC} = 0.011 - 0.007 \times \text{tt2} - 0.114 \times \text{imp1} + 0.081 \times \text{imp2} \quad (R^2 = 0.951) \quad (12)$$

When extrapolating this function toward "100% purity" (i.e.,  $\text{tt2} = \text{imp1} = \text{imp2} = 0$ ), **25** exhibits an MIC of 91  $\mu\text{g/mL}$ . This means that the impurities had a negative impact on the observed activity of **25**, leading to higher than expected MIC values for the six samples. The effect of each

impurity on the overall activity can be explained by the coefficients for the corresponding independent variable in the model (Eq. 12). Clearly, both **26** and **27** show adverse effects, while **28** exhibits a promoting function.



**Figure 46. Purity–Activity Relationships of the *O. horridus* Triterpene Fractions**

Panel A shows that the activity and purity of **25** were not proportional, nor were they correlated in a linear or logarithmic fashion. Panel B illustrates the correlation between the impurities (imp 1 = monoglyceride, imp 2 = polyene) and anti-TB activity. This demonstrates that the impurities have a major impact on the observed anti-TB activity of triterpene fractions dominated by triterpene **25**.

Glycerol is the main carbon source usually employed in the culture of mycobacteria. A recent study has identified a monoglyceride lipase Rv0183 from *M. tuberculosis*, suggesting that this lipolytic enzyme may be involved in the degradation of host cell lipids. Thus, the monoglycerides can be hydrolyzed by Rv0183, releasing free glycerol and absorbable fatty acid which are essential elements for the growth of *M. tuberculosis* (Cotes *et al.*, 2007; Singh *et al.*, 2010). This explains the antagonist effect of monoglyceride impurities on the anti-TB activity of

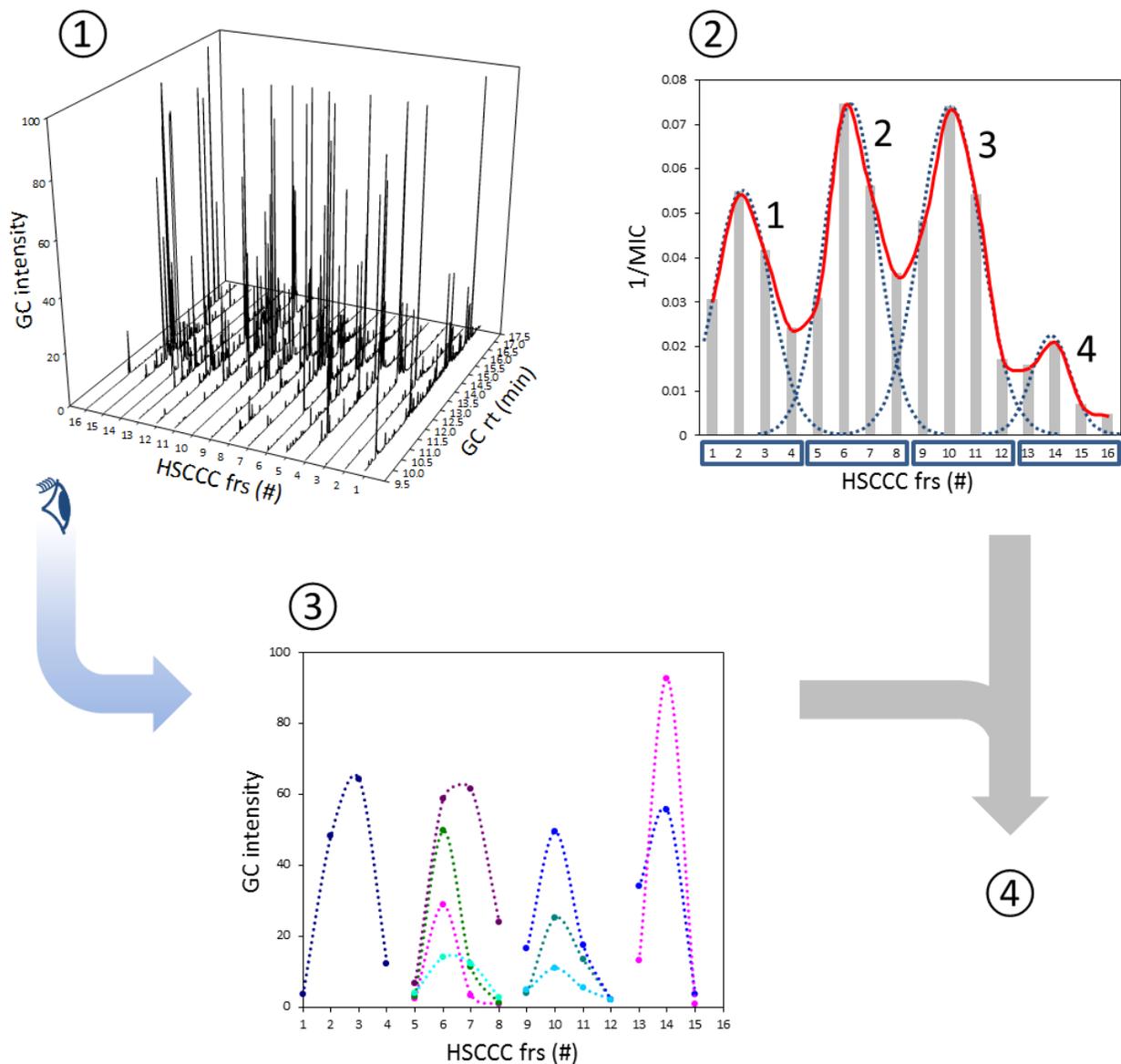
triterpenes. However, several reports have confirmed polyynes as relatively potent anti-TB compounds with MIC < 60  $\mu\text{g}/\text{mL}$  (Deng *et al.*, 2008; Inui *et al.*, 2010). Therefore, even minor polyynes may exert significant enhancement of the anti-TB activity of the triterpene in the sample.

The triterpenes **25** and **26** exist as an isomeric pair derived from betulinic acid and oleanolic acid, respectively, by epimerization at C-3 and oxidation of 23-Me to a carboxylic acid group. These two precursors have previously been reported to be active against *M. tuberculosis* with MIC values of 62 and 29  $\mu\text{g}/\text{mL}$ , respectively (Copp *et al.*, 2007). However, in contrast to the literature, the QPAR model developed here implies that the anti-TB activity of **26** is actually weaker than that of **25**, resulting in a higher MIC value for the combination of the two relative to that of pure **25**.

## 6.2 Biochemometric Evaluation of Antituberculosis Principles in *H. lupulus*

For a long time, the widely accepted methodology for identification of bioactive principles from herbal medicines has been BGF with the ultimate goal of isolating active compounds. However, the BGF concept has overlooked an important characteristic of botanical drugs which consist of a multitude of chemical components possibly exerting synergistic therapeutic effects. Therefore, the BGF approach may or may not be suitable for materials that exhibit synergistic effects between multiple agents such as herbal medicines. To overcome this limitation, a biochemometric approach has recently been developed to evaluate the bioactive principles of herbal medicines (Inui, 2008). This approach integrates the chemical and biological data of subsequent fractions and identifies multiple bioactive compounds in complex mixtures without the need for their individual isolation. With the aid of *K*-based CS which can be used as a chromatographic method, the bioactive principles can be readily located in the fractions, and, thus, can be directly used for further chemical and biological characterization. The following case exemplifies an application of biochemometrics in the determination of the anti-TB active principles in an herbal medicine.

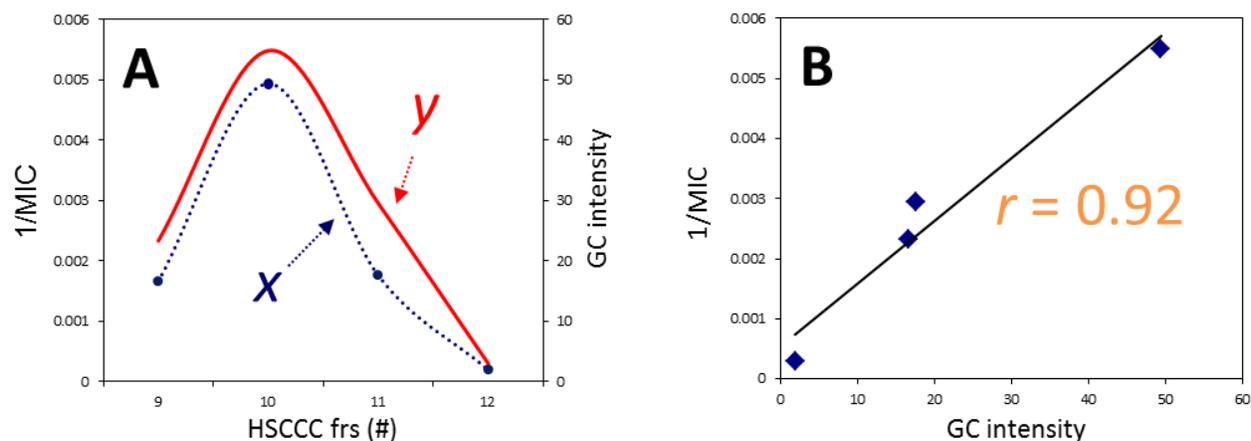
The essential oil of hops has been reported and confirmed to have antibacterial activity against *M. tuberculosis*. As shown in **Figure 8**, the lipophilic hops extract containing essential oil components can be well separated by HSCCC in elution-extrusion mode using the HterAc SS which is optimized for the lipophilicity of essential oils. Subsequent GC-MS analysis suggested that the fractions in the A- ( $K < 0.5$ ) and B- ( $0.5 < K < 3.0$ ) regions contained sesquiterpenes and fatty ketones, whereas the fractions in the C-region ( $K > 3.0$ ) consisted mainly of fatty acids and alcohols. Further preparative purification of the combined C-fractions using VLC afforded two triterpenes ( $\alpha$ -amyrin **40** and  $\beta$ -amyrin **41**), a sterol ( $\beta$ -sitosterol **42**), and a fatty alcohol (1-tetracosanol **43**).



**Figure 47. The Major Steps of the Biochemometric Approach**

Step ①: All preparative HSCCC fractions of lipophilic hops extract were subjected to GC-MS analysis, leading to the formation of a 3D HSCCC-GC-MS matrix; Step ②: A biochromatogram was generated by anti-TB activity evaluation of all fractions using the MABA assay. The biochromatogram was further deconvoluted to individual biopeaks; Step ③: The constituents in the subsequent fractions related to each biopeak were extracted from the front-view dimension (HSCCC-MS) of the 3D matrix; Step ④: Pearson's correlation analysis was performed to determine similarities between each peak in the HSCCC-MS dimension and potentially correlated biopeaks (Figure 48).

Anti-TB bioassays of the fractions in all three regions showed that the strongest bioactivity was concentrated in the B-fractions. Combination of the analytical data resulting from GC-MS analysis of all HSCCC fractions resulted in the creation of a 3D HSCCC-GC-MS matrix (Figure 47-①). Incorporating the quantitative biological evaluation of all HSCCC fractions generated a 4D data set. This data was subject to biochemometric evaluation, which consisted of a Pearson's correlation analysis of the deconvoluted biochromatogram and the 3D matrix (Figure 48).



**Figure 48. Step 4 of the Biochemometric Evaluation by Pearson's Correlation Analysis**

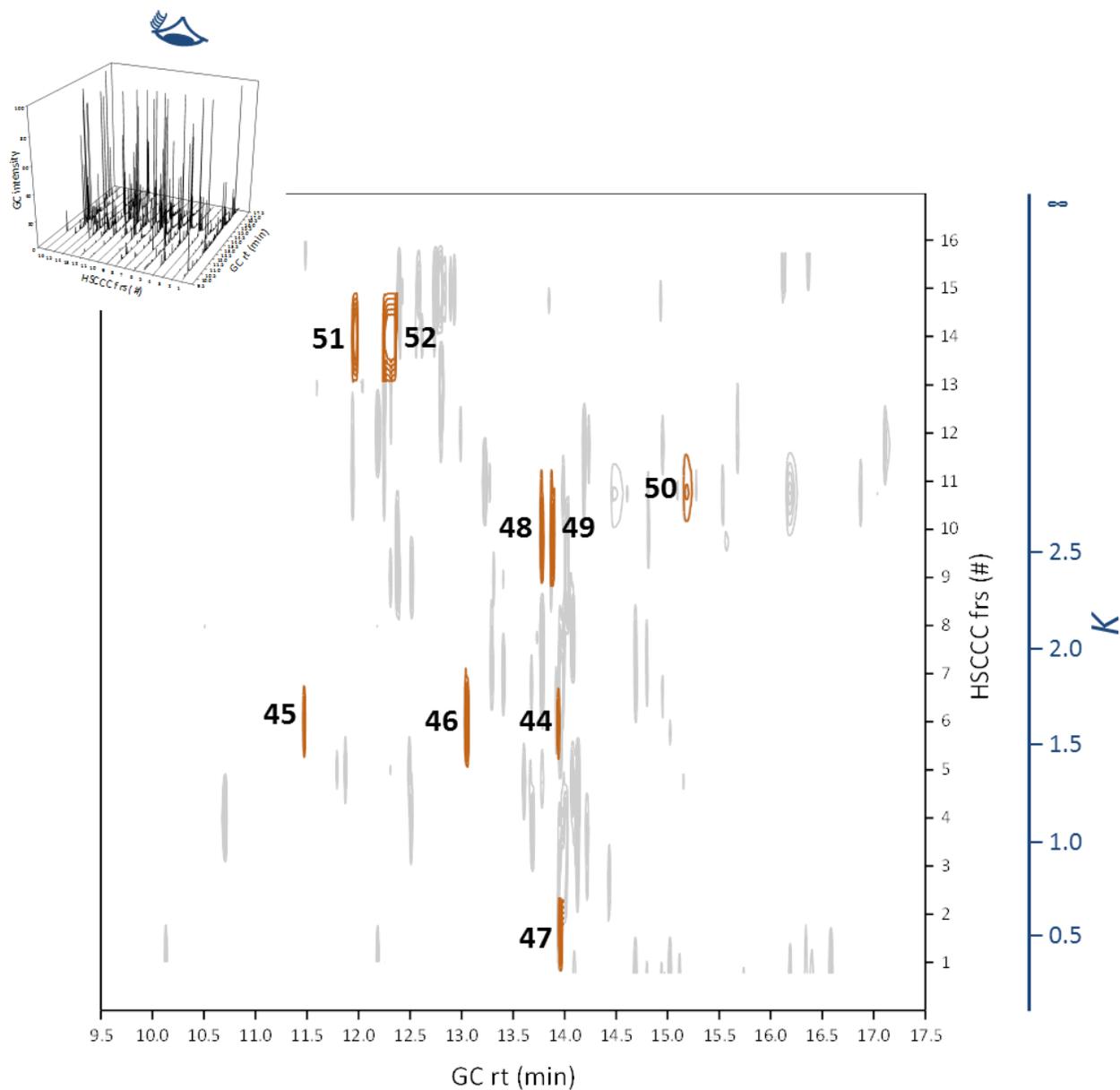
The two panels A and B illustrate that a linear relationship exists between the GC intensity (blue line in panel A) of compound **49** in four subsequent fractions and the anti-TB activity (1/MIC values, red line) of these fractions with Pearson's correlation ( $r$ ) of 0.92.

**TABLE XIII. ANTI-TB ACTIVE PRINCIPLES IN HOPS EXTRACT WITH PEARSON'S CORRELATIONS  $\geq 0.80$** 

#	Proposed Structure	GC rt (min)	<i>r</i>	MS Match (%)
44	Unknown	13.65	0.95	
45	2-Dodecanone	11.41	0.85	73.9
46	Unknown	12.94	0.84	
47	Cadinol or related	13.66	0.87	40.4
48	Cubenol or related	13.48	0.80	42.7
49	Unknown	13.57	0.92	
50	Unknown	15.22	0.93	
51	Caryophyllene or related	11.88	0.81	31.4
52	Unknown	12.17	0.92	

In the present study, the biochemometric analysis probed the 4D data for linear relationships between the concentration of chemical constituents and observed biological activities (**Figure 49**). As a result, 10 GC peaks with Pearson's correlations of  $r \geq 0.80$  were identified (**Table 13**), indicating that the underlying phytoconstituents are associated with the observed bioactive CCC peaks and, thus, represent the anti-TB principles in the lipophilic hops extract. The NIST Mass Spectral Library search suggested that these compounds are mainly sesquiterpenes, such as cadinol and cubenol (or related compounds).

The results show that biochemometric analysis is capable of determining bioactive constituents in complex NPs without the need for the isolation of single compounds. In addition, biochemometrics can also guide preparative, CS-based isolation efforts aimed at the further characterization of the active principles by isolation and full structure elucidation as well as for synergy/antagonism studies. Future studies will be required to further characterize the anti-TB active compounds by *K*-targeted isolation combined with spectroscopic analysis.



**Figure 49. The Contour Plot of the Pearson's Correlation Values in the 3D HSCCC-GC-MS Matrix**

The active components with  $r \geq 0.80$  (indicated in orange) mainly distributed in the B-region of the HSCCC separation, indicating that the SS was suitable for the lipophilic hops extract. This chart can be used as a "guide map" for  $K$ -targeted isolation of these proposed active compounds for further chemical and biological characterization.

### 6.3 Conclusion

This chapter has emphasized the importance of analyzing interactions of multiple chemical components in the biological evaluation of NPs. The results of the QPARs of *O. horridus* triterpenes study suggest that purity should be routinely investigated for all isolates, especially when they are used for biological assessment. A novel approach was introduced for the establishment of QPARs using the chemical and biological profiles of subsequent chromatographic fractions. Additionally, the results highlight the importance of integrating advanced NMR techniques for structural elucidation or dereplication in the characterization of minor impurities. The complexity of biologically active NPs extends beyond their structural diversity and encompasses interactions between the main component and the residually complex matrix. The establishment of QPARs provides a powerful tool for analyzing and quantifying these interactions as well as their resulting effects on the bioactivity. Furthermore, the explanatory model of these interactions even enables prediction of the bioactivity of isolates from the same batch of purification, which provides a potential technique for biological standardization of botanicals as crude extracts or refined phytochemicals. In addition to NPs, these methodologies should equally apply to the products of parallel and combinatorial synthesis. The concept of integration of chemical and biological complexities was also adopted in identification of bioactive principles in NP mixtures. As botanical drugs consist of multiple chemical components, their interactions are an integral part of their overall bioactivity, which is often considered to be a result of the synergistic and/or antagonistic effects. The case of hops demonstrates that the biochemometric approach is effective for connecting the biological assays with the chemical profiles, and determining the bioactive principles in mixtures without requiring physical isolation of SCEs. This methodology can be implemented in any workflow of fractionation aimed at the discovery of bioactive principles in NPs.

## 7. SUMMARY AND PERSPECTIVES

NPs have been proven to be a good source of drug candidates, whether in the form of herbal remedies or refined (purified) compounds. For decades, tremendous efforts have been devoted to the development of methods and tools for NP research, and have achieved significant success. However, due to their chemical and biological complexities, the characterization of NPs still remains a challenging and time-consuming process. In order to resolve these complexities and improve the efficiency of NP discovery, especially for the separation and identification of bioactive compounds, the present study has developed several new concepts, methods, and applications. Using four popular herbal supplements and one ethnobotanical as case studies, these methods have been shown to be effective and potentially applicable in a broader variety of NPs. These results are promising as they will direct future applications as well as improvements of the methods themselves.

**Sample Preparation.** Enrichment is a crucial step for an efficient separation of target compounds from NP mixtures (Section 3.2, p. 40). Liquid-liquid partitioning is an appropriate method for preparation of enriched samples, which can minimize sample loss and reduce the complexity of the fractionation tree. The concept of sample-cutting can be also applied to sample preparation for biological evaluation and quality standardization of NPs, because target compounds can be efficiently concentrated while interfering components are removed, using liquid-liquid partitioning with appropriate SSs.

**Optimization of Separation Conditions.** The prediction of separation performance, such as CS methods led by  $K$  values, is important for the rational optimization of chromatographic conditions (Section 3.3, p. 46). Recent developments of CS modeling and simulation software as exemplified by ProMISE (Sutherland *et al.*, 2009, 2011) and prEEdiCCCT (Gallagher *et al.*, 2010) have expanded such prediction from calculation of the resolution factor to visualization of the behavior of the analytes in the CS process. Future studies could apply the intelligent programs to facilitate the computer-aided design of CS processes.

Chromatographic and solvent orthogonality are useful tools for the rational design of efficient fractionation schemes with enhanced resolution (Section 3.4, p. 51; Section 3.5, p. 54). While both polarity and selectivity are key factors influencing the separation performance, the latter is more important for resolution of structurally similar congeners which are common occurrences in NPs. Therefore, orthogonal conditions are highly effective in the separation of target compounds by offering multidimensional selectivity that acts as a resolution enhancer. Until recently, the selection of optimal separation conditions has virtually remained an empirical operation. Systematic exploration of the relationship between the structural characteristics of NPs and the selectivity of the stationary and mobile phases is required to establish a protocol for developing optimized orthogonal conditions for the purification of the target compounds.

**Mixture Analysis by 1D/2D NMR.** Structural visualization of the separation process can not only guide isolation, but also can provide chemical information for the optimization of the separation conditions such as the selection of appropriate orthogonal SSs (Section 4.2, p. 69; Section 4.3, p. 82). NMR techniques play a vital role in developing such applications. The introduction of highly sensitive cryo-microprobe NMR spectrometers makes it possible to monitor and track trace amounts of constituents, even in complex mixtures (Section 4.4, p. 85). Evolving from the current work, future studies could make more extensive use of 1D/2D NMR techniques for the structural elucidation and dereplication of NP mixtures. These applications, if combined with *K*-by-NMR methodology, could potentially further increase the efficiency of CS.

Currently, LC-MS and database searches are the most popular approaches for structural dereplication, which is an integral part of NP discovery. However, they are still dependent of the additional dimensions of highly resolved separation as well as reproducible ionization. Free of such limitations, the use of pattern recognition in 1D/2D NMR spectroscopy has shown to be a powerful tool in analysis of NP mixtures (Section 4.5, p. 92; Chapter 5, p. 101). While NMR analysis associated with pattern recognition has been widely used for metabolomic profiling and quality evaluation of NPs, their applications in structural elucidation and dereplication have been

initiated, but so far have received little attention (Kowalski *et al.*, 1972; Goux *et al.*, 1989; Dzeroski *et al.*, 1998). In addition to binary trees, future studies could investigate other statistical methods and computer techniques in building more accurate and intelligent pattern recognition models, and employ such methodology in a wider range of NPs.

**Biological Evaluation of NPs.** Residual complexity is an important, but often overlooked, aspect in purified NPs. Impurities, even in trace amounts, may have profound impact on the bioactivity of the main components, such as was shown in the present study (Section 6.1, p. 135). The overall bioactivity is a result of constituent interactions that are represented by synergistic and/or antagonistic effects. It is critical, therefore, that the nature of these interactions be properly analyzed and quantified. Commonly used methods for analysis of drug interactions, such as isobolographic analysis, fractional analysis, and logistic regression analysis (Bovill, 1998), could be adopted in future studies for exploration of NP interactions and assist to identify the bioactive principles.

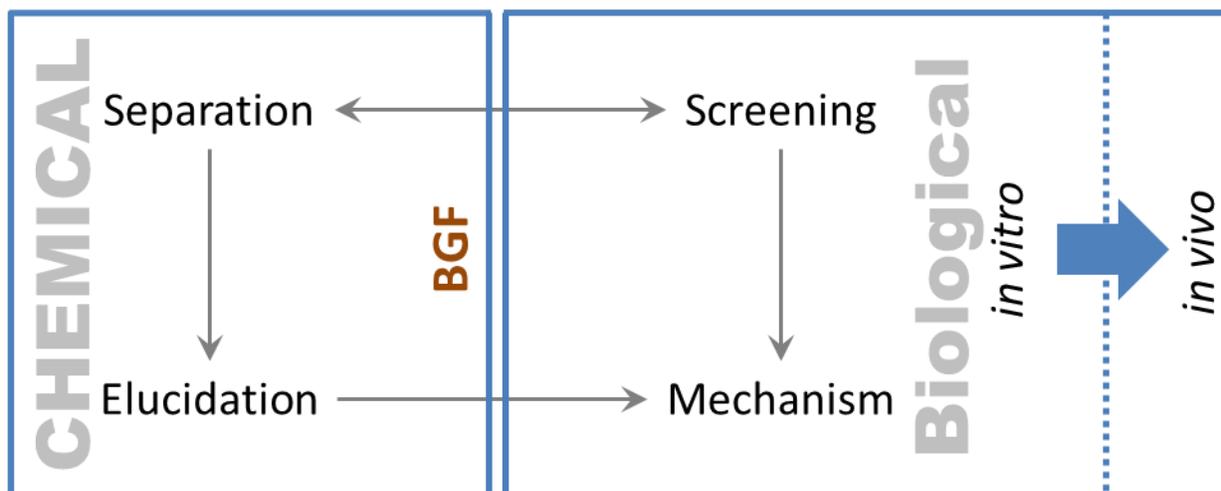
Currently, the synergy and antagonism studies of NPs mainly focus on combinations of purified compounds. However, this methodology is not able to unveil the mechanism of action of traditional medicines, which commonly use crude extracts as remedies and represent highly complex systems in both chemistry and biology. The synergistic/antagonistic interactions may not only result from a multitude of chemical components, but can also be affected by the concentration gradient of the components in the crude extract. Many other factors should be also taken into account, such as the effects of the extraction method and physiological conditions. None of these can be replicated by simply using refined phytochemicals. Moreover, the actual situation is that a combination of different plants is often used to formulate a complex herbal remedy, the chemical and biological properties of which could be extremely complex. This has created an unprecedented challenge for characterizing the traditional medicines in the process of NP discovery.

**Intertwining Chemistry and Biology of NPs.** Traditionally, BGF has been applied as a reductionist approach which simplifies NP discovery by creating a discrete transit between chemical and biological studies. Led by biological screening at each step of fractionation, isolation and characterization of pure and bioactive molecules is the anticipated end result. However, this model overlooks the complex relationships between the metabolomic chemistry and biology of NPs. Consequently, the results might not truly reveal the properties of these NPs or might even be misleading. Based on the present study, a new model for bioactive NP research is proposed in which the interface between chemical characterization and biological evaluation could be significantly enhanced, thus, termed the “gut model.” In this model, many concepts and methodologies developed in the present study are implemented to create additional layers and contact points between chemical and biological studies that help resolve the complexity of NPs (**Figure 50**).

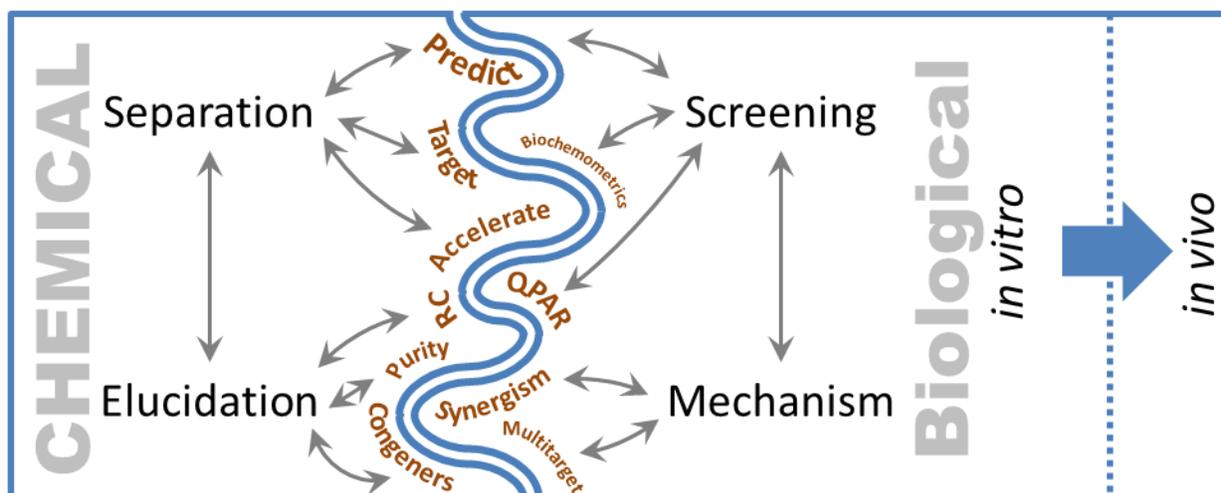
The chemistry–biology interface of this model works in analogy to a gap junction between cells where multiple ion channels are present for cellular communication. Thus, by linking the added points at the two sides of the interface in the “gut model,” multiple channels can be built up to enhance the interactions between the chemical and biological evaluation of NPs. These channels can together form a complex network for a comprehensive resolution of NP complexity. Furthermore, by defining more points of contact between chemical and biological characterization, future studies could enrich the chemistry–biology interface and improve the understanding of NP complexity. Derived from the “gut model,” **Figure 51** suggests a flowchart for future NP research which combines the innovative methodologies developed in the present study. Upon biochemometric evaluation, NP mixtures can be directly subjected to spectroscopic analysis for structural characterization of the potentially bioactive components. This can be facilitated by computer-aided techniques, such as database searches and pattern recognition of spectroscopic fingerprint signals. Whenever separation of pure compounds is required, more efficient fractionation schemes can be designed by predicting elution profiles and more rational

selection of chromatographic conditions. By monitoring NP separation procedure through detailed spectroscopic analysis of the fractions, the isolation and identification of target compounds from the mixtures is accelerated by structural dereplication. Finally, the chemical profiling of impurities in purified NPs leads to the establishment of quantitative purity–activity relationships which can be valuable for biological assessment of the isolates and the characterizing of their mechanisms of action.

## A Reductionist Model

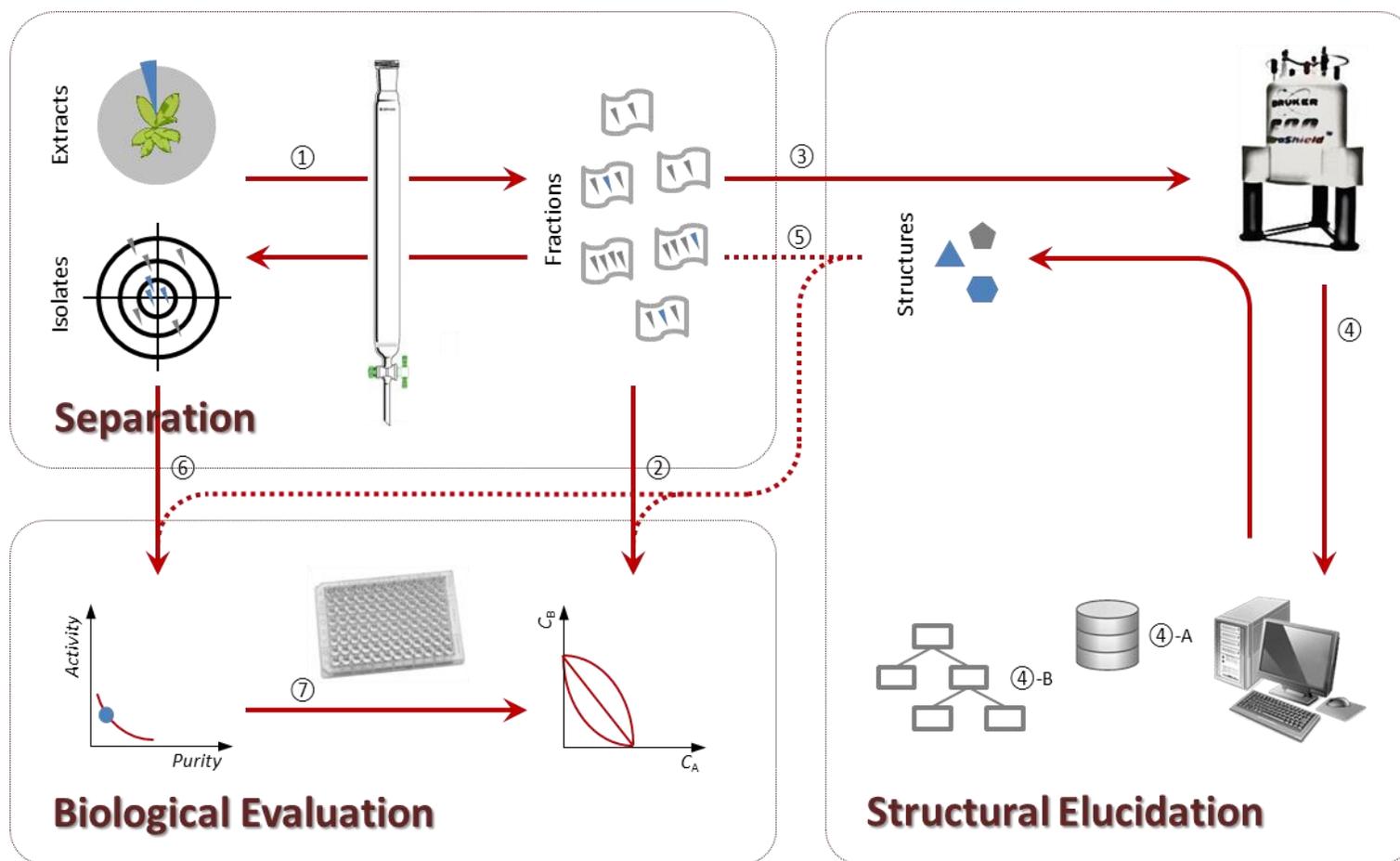


## B Gut Model



**Figure 50. The Comparison between the Classic Reductionist Model and the Proposed Gut Model in Bioactive NP Research**

Panel A: The reductionist model is characterized by rather discrete transits between chemical and biological studies. Led by biological screening at each step of fractionation, isolation and characterization of pure and bioactive molecules is the end result. Panel B: In the gut model, however, the interface between chemical characterization and biological evaluation is significantly enhanced by providing additional contact points between the chemical and the biological studies. Through these points, multiple “communication channels” can be built up to link the chemical and biological properties of NPs. The added dimension of information from the chemistry–biology interface are likely to add new insights towards the resolution of NP complexity.



**Figure 51. A Proposed Flowchart for Future NP Research Deriving from the “Gut Model”**

Upon biochemometric evaluation (Steps ① and ②), NP mixtures can be directly subjected to spectroscopic analysis for structural characterization of the potentially bioactive components (Step ③). This can be facilitated by computer-aided techniques, such as database searches (④-A) and pattern recognition of spectroscopic fingerprint signals (④-B). The NMR analysis can guide target isolation of the bioactive compounds (Steps ⑤). Finally, the chemical profiling of impurities in purified NPs leads to the establishment of quantitative purity–activity relationships (Step ⑥) which can be valuable for biological assessment of the isolates and the characterizing of their mechanisms of actions (Step ⑦).

## CITED LITERATURE

- Ackland, M. L.; Van de Waarsenburg, S.; Jones, R. *In Vivo* **2005**, *19*, 69–76.
- Akihisa, T.; Hideshima, R.; Koike, K.; Kimura, Y.; Nikaido, T. *Chem. Pharm. Bull.* **1999**, *47*, 1157–1160.
- Ali, Z.; Khan, S. I.; Fronczek, F. R.; Khan, I. A. *Phytochemistry* **2007**, *68*, 373–382.
- Bedir, E.; Kahn, I. *Chem. Pharm. Bull.* **2000**, *48*, 425–427.
- Berenbaum, M. C. *Pharmacol. Rev.* **1989**, *41*, 93–141.
- Berthod, A.; Deroux Jean, M.; Bully, M.; Conway, W. D.; Petroski, R. J. *Modern Countercurrent Chromatography*; American Chemical Society, Washington D.C., 1995; p. 16.
- Berthod, A.; Friesen, J. B.; Inui, T.; Pauli, G. F. *Anal. Chem.* **2007**, *79*, 3371–3382.
- Bhacca, N. S.; Williams, D. H., Applications of NMR Spectroscopy in Organic Chemistry; Illustrations From the Steroid Field; Holden-Day: San Francisco, CA, 1964.
- Blumenthal, M.; Ferrie, G. K. L.; Cavaliere, C.; *HerbalGram* **2006**, *71*, 64.
- Boroczky, K.; Laatsch, H.; Wagner-Dobler, I.; Stritzke, K.; Schulz, S. *Chem. Biodivers.* **2006**, *3*, 622–634.
- Bradshaw, J.; Butina, D.; Dunn, A. J.; Green, R. H.; Hajek, M.; Jones, M. M.; Lindon, J. C.; Sidebottom, P. J. *J. Nat. Prod.* **2001**, *64*, 1541–1544.
- Brown, S. D.; Tauler, R.; Walczak, B. *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*; Elsevier: Amsterdam, 2009; Vol. 3, pp. 542–567.
- Carlson, E. E. *Acs. Chem. Biol.* **2010**, *5*, 639–653.
- Cazes, J. *Encyclopedia of Chromatography*; CRC Press, 2005; Vol. 2
- Chadwick, L. R.; Fong, H. H. S.; Farnsworth, N. R.; Pauli, G. F. *J. Liq. Chromatogr. Relat. Technol.* **2005**, *28*, 1959–1969.
- Chan, P.-C.; Xia, Q.; Fu, P.-P. *J. Environ. Sci. Health. C Environ. Carcinog. Ecotoxicol. Rev.* **2007**, *25*, 211–244.
- Chen, L.-R.; Robien, W. *Chemometr. Intell. Lab.* **1993**, *19*, 217–223.
- Chen, S.-N.; Fabricant, D. S.; Lu, Z. Z.; Fong, H.H.; Farnsworth, N. R. *J. Nat. Prod.* **2002**, *65*, 1391–1397.
- Chen, S.-N.; Lankin, D. C.; Chadwick, L. R.; Jaki, B. U.; Pauli, G. F. *Planta Med.* **2009**, *75*, 757–762.
- Chen, S.-N.; Lankin, D. C.; Chadwick, L. R.; Jaki, B. U.; Pauli, G. F. *Planta Med.* **2009**, *75*, 757–762.

- Chen, S.-N.; Lankin, D. C.; Nikolic, D.; Fabricant, D. S.; Lu, Z.-Z.; Ramirez, B.; van Breeman, R. B.; Fong, H. H. S.; Farnsworth, N. R.; Pauli, G. F. *J. Nat. Prod.* **2007**, *70*, 1016–1023.
- Chen, S.-N.; Li, W.; Fabricant, D. S.; Santarsiero, B. D.; Mesecar, A.; Fitzloff, J. F.; Fong, H. H. S.; Farnsworth, N. R. *J. Nat. Prod.* **2002**, *65*, 601–605.
- Cheung, H. T.; Williamson, D. G. *Tetrahedron* **1969**, *25*, 119–128.
- Cheung, H. T.; Wong, C.-S.; Yan, T.-C. *Tetrahedron Lett.* **1969**, *10*, 5077–5080.
- Choi, Y. H.; Choi, H. K.; Hazekamp, A.; Bermejo, P.; Schilder, Y.; Erkelens, C.; Verpoorte, R. *Chem. Pharm. Bull.* **2003**, *51*, 158–161.
- Choi, Y. H.; Choi, H.-K.; Peltenburg-Looman, A. M. G.; Lefeber, A. W. M.; Verpoorte, R. *Phytochem. Anal.* **2004**, *15*, 325–330.
- Compton, J. A.; Culham, A.; Jury, S. L. *Taxon* **1998**, *47*, 593–634.
- Conway, W. D. *J. Chromatogr. A* **2011**, *1218*, 6015–6023.
- Copp, B. R.; Pearce, A. N. *Nat. Prod. Rep.* **2007**, *24*, 278–297.
- Corley, D. G.; Durley, R. C. *J. Nat. Prod.* **1994**, *57*, 1484–1490.
- Cotes, K.; Dhouib, R.; Douchet, I.; Chahinian, H.; De Caro, A.; Carriere, F.; Canaan, S. *Biochem. J.* **2007**, *408*, 417–427.
- Craig, L. C.; King, T. P. *Fed. Proc.* **1958**, *17*, 1126–1134.
- Dahlberg, C. J.; Traub, J. S.; Tripp, M. L.; Bland, J. S.; Carroll, B. J. *J. Sep. Sci.* **2010**, *33*, 2828–2832.
- Dalisay, D. S.; Molinski, T. F. *J. Nat. Prod.* **2009**, *72*, 739–744.
- Dan, C.; Zhou, Y.; Ye, D.; Peng, S.; Ding, L.; Gross, M.L.; Qiu, S.-X. *Org. Lett.* **2007**, *9*, 1813–1816.
- de Folter, J.; Sutherland, I. A. *J. Chromatogr. A* **2009**, *1216*, 4218–4224.
- de Folter, J.; Sutherland, I. A. *J. Chromatogr. A* **2011**, *1218*, 6009–6014.
- Deng, F.; Zito, S. W. *J. Chromatogr. A* **2003**, *986*, 121–127.
- Deng, S.-X.; Wang, Y.-H.; Inui, T.; Chen, S.-N.; Farnsworth, N. R.; Cho, S.; Franzblau, S. G.; Pauli, G. F. *Phytother. Res.* **2008**, *22*, 878–882.
- Di Marco, V. B.; Bombi, G. G. *J. Chromatogr. A* **2001**, *931*, 1–30.
- Dzeroski, S.; Schulze-Kremer, S.; Heidtke, K. R.; Siems, K.; Wettschereck, D.; Blockeel, H. *Appl. Artif. Intell.* **1998**, *12*, 363–383.

- Exarchou, V.; Krucker, M.; van Beek, T. A.; Vervoort, J.; Gerothanassis, I. P.; Albert, K. *Magn. Reson. Chem.* **2005**, *43*, 681–687.
- Fan, Y.-S.; Yao, Z.; Zhang, Y.-W.; Duan, H.-Q. *J. Asian Nat. Prod. Res.* **2009**, *11*, 588–596.
- Firn, R. D.; Jones, C. G. *Nat. Prod. Rep.* **2003**, *20*, 382–391.
- Firn, R. D.; Jones, C. G. *Trends. Plant Sci.* **2006**, *11*, 112–113.
- Friesen, J. B.; Pauli, G. F. *Anal. Chem.* **2007**, *79*, 2320–2324.
- Friesen, J. B.; Pauli, G. F. *J. Chromatogr. A* **2007**, *1151*, 51–59.
- Friesen, J. B.; Pauli, G. F. *J. Liq. Chromatogr. Relat. Technol.* **2005**, *28*, 2777–2806.
- Friesen, J. B.; Pauli, G. F.; *Anal. Chem.* **2007**, *79*, 2320–2324.
- Gertsch, J. *Planta Med.* **2011**, *77*, 1086–1098.
- Ghasemi, J.; Saaidpour, S. *Anal. Chim. Acta* **2007**, *604*, 99–106.
- Gödecke, T.; Lankin, D. C.; Nikolic, D.; Chen, S. N.; van Breemen, R. B.; Farnsworth, N. R.; Pauli, G. F. *J. Nat. Prod.* **2009**, *72*, 433–437.
- Gödecke, T.; Nikolic, D.; Lankin, D. C.; Chen, S. N.; Powell, S. L.; Dietz, B.; Bolton, J. L.; van Breemen, R. B.; Farnsworth, N. R.; Pauli, G. F. *Phytochem. Anal.* **2009**, *20*, 120–133.
- Goux, W. J. *J. Magn. Reson.* **1989**, *85*, 457–469.
- Graham, H. N. *Prev. Med.* **1992**, *21*, 334–350.
- Hänsel, R. V.; Schulz, J. *Arch. Pharm.* **1998**, *321*, 37–40.
- Harvey, A. L. *Drug Discov. Today* **2008**, *13*, 894–901.
- He, K.; Pauli, G. F.; Zheng, B.; Wang, H.; Bai, N.; Peng, T.; Roller, M.; Zheng, Q. *J. Chromatogr. A* **2006**, *1112*, 241–254.
- Hostettmann, K.; Potterat, O.; Wolfender, J.-L. *Pharm. Ind.* **1997**, *59*, 339.
- Ignatova, S.; Sumner, N.; Colclough, N.; Sutherland, I. *J. Chromatogr. A* **2011**, *1218*, 6053–6060.
- Inui, T. Phytochemical and Biochemometric Evaluation of the Alaskan Anti-TB Ethnobotanical: *Oplopanax horridus*. PhD Dissertation, University of Illinois at Chicago, Chicago, 2008.
- Inui, T.; Wang, Y.-H.; Deng, S.-X.; Smith, D. C.; Franzblau, S. G.; Pauli, G. F. *J. Chromatogr. A* **2007**, *1151*, 211–215.
- Inui, T.; Wang, Y.-H.; Nikolic, D.; Smith, D. C.; Franzblau, S. G.; Pauli, G. F. *J. Nat. Prod.* **2010**, *73*, 563–567.

- Ito, Y.; Sandlin, J.; Bowers, W. G. *J. Chromatogr.* **1982**, *244*, 247–258.
- Jaki, B. U.; Franzblau, S. G.; Chadwick, L. R.; Lankin, D. C.; Zhang, F.; Wang, Y.; Pauli, G. F. *J. Nat. Prod.* **2008**, *71*, 1742–1748.
- Jaracz, S.; Malik, S.; Nakanishi, K. *Phytochemistry* **2004**, *65*, 2897–2902.
- Ji, H. F.; Li, X. J.; Zhang, H. Y. *EMBO Rep.* **2009**, *10*, 194–200.
- Johnson, A. R.; Vitha, M. F. *J. Chromatogr. A* **2011**, *1218*, 556–586.
- Jones, C. G.; Firn, R. D. *Philos. T. Roy. Soc. B* **1991**, *333*, 273–280.
- Kadota, S.; Li, J.-X.; Tanaka, K.; Namba, T. *Tetrahedron* **1995**, *51*, 1143–1166.
- Kahkonen, M. P.; Hopia, A. I.; Vuorela, H. J.; Rauha, J. P.; Pihlaja, K.; Kujala, T. S.; Heinonen, M. *J. Agr. Food Chem.* **1999**, *47*, 3954–3962.
- Kalchhauser, H.; Robien, W. *J. Chem. Inform. Comput. Sci.* **1985**, *25*, 103–108.
- Kim, H. K.; Choi, Y. H.; Verpoorte, R. *Nat. Protoc.* **2010**, *5*, 536–549.
- Kitaoka, M.; Kadokawa, H.; Sugano, M.; Ichikawa, K.; Taki, M.; Takaishi, S.; Iijima, Y.; Tsutsumi, S.; Boriboon, M.; Akiyama, T. *Planta Med.* **1998**, *64*, 511–515.
- Kobaisy, M.; Abramowski, Z.; Lermer, L.; Saxena, G.; Hancock, R. E. W.; Towers, G. H. N.; Doxsee, D.; Stokes, R. W. *J. Nat. Prod.* **1997**, *60*, 1210–1213.
- Koeda, M.; Aoki, Y.; Sakurai, N.; Nagai, M. *Chem. Pharm. Bull.* **1995**, *43*, 771–776.
- Koehler, M. G.; Grigoras, S.; Dunn, W. J. *Mol. Inform.* **1988**, *7*, 150–159.
- Kokkinofa, R.; Petrakis, P. V.; Mavromoustakos, T.; Theocharis, C. R. *J. Agr. Food Chem.* **2003**, *51*, 6233–6239.
- Kowalski, B. R.; Bender, C. F. *Anal. Chem.* **1972**, *44*, 1405–1411.
- Kusano, A.; Shibano, M.; Kitagawa, S.; Kusano, G.; Nozoe, S.; Fushiya, S. *Chem. Pharm. Bull.* **1994**, *42*, 1940–1943.
- Kusano, A.; Shibano, M.; Kusano, G. *Chem. Pharm. Bull.* **1995**, *43*, 1167–1170.
- Kusano, A.; Shibano, M.; Kusano, G. *Chem. Pharm. Bull.* **1996**, *44*, 167–172.
- Kusano, A.; Shibano, M.; Kusano, G. *Chem. Pharm. Bull.* **1999**, *47*, 1175–1179.
- Kusano, A.; Shibano, M.; Kusano, G.; Miyase, T. *Chem. Pharm. Bull.* **1996**, *44*, 2078–2085.
- Kusano, A.; Shibano, M.; Tsukamoto, D.; Kusano, G. *Chem. Pharm. Bull.* **2001**, *49*, 437–441.

- Kusano, A.; Shimizu, K.; Idoji, M.; Shibano, M.; Minoura, K.; Kusano, G. *Chem. Pharm. Bull.* **1995**, *43*, 279–283.
- Kusano, A.; Takahira, M.; Shibano, M.; Miyase, T.; Kusano, G. *Chem. Pharm. Bull.* **1999**, *47*, 511–516.
- Kusano, A.; Takahira, M.; Shibano, M.; Miyase, T.; Okuyama, T.; Kusano, G. *Heterocycles* **1998**, *48*, 1003–1013.
- Kusano, A.; Takahira, M.; Shibano, M.; In, Y.; Ishida, T.; Miyase, T.; Kusano, G. *Chem. Pharm. Bull.* **1998**, *46*, 467–472.
- Kusano, A.; Takahira, M.; Shibano, M.; Miyase, T.; Kusano, G. *Chem. Pharm. Bull.* **1998**, *46*, 1001–1007.
- Kusano, G.; Hojo, S.; Kondo, Y.; Takemoto, T. *Chem. Pharm. Bull.* **1977**, *25*, 3182–3189.
- Kusano, G.; Idoji, M.; Sogoh, Y.; Shibano, M.; Kusano, A.; Iwashita, T. *Chem. Pharm. Bull.* **1994**, *42*, 1106–1110.
- Kusano, G.; Murakami, Y.; Sakurai, N.; Takemoto, T. *Yakugaku Zasshi* **1976**, *96*, 82–85.
- Kusano, G.; Takemoto, T. *Yakugaku Zasshi* **1975**, *95*, 1133–1137.
- Kusano, G.; Uchida, H.; Murakami, Y.; Sakurai, N.; Takemoto, T. *Yakugaku Zasshi* **1976**, *96*, 321–325.
- Lambert, M.; Staerk, D.; Hansen, S. H.; Sairafianpour, M.; Jaroszewski, J. W. *J. Nat. Prod.* **2005**, *68*, 1500–1509.
- Lang, G.; Mayhudin, N. A.; Mitova, M. I.; Sun, L.; van der Sar, S.; Blunt, J. W.; Cole, A. L. J.; Ellis, G.; Laatsch, H.; Munro, M. H. G. *J. Nat. Prod.* **2008**, *71*, 1595–1599.
- Lavie, D.; Benjaminov, B. S.; Shvo, Y. *Tetrahedron* **1964**, *20*, 2585–2592.
- Lewis, I. A.; Schommer, S. C.; Hodis, B.; Robb, K. A.; Tonelli, M.; Westler, W. M.; Sussman, M. R.; Markley, J. L. *Anal. Chem.* **2007**, *79*, 9385–9390.
- Li, C.-Y.; Lin, C.-H.; Wu, C.-C.; Lee, K.-H.; Wu, T.-S. *J. Agric. Food Chem.* **2004**, *52*, 3721–3725.
- Li, J. W.-H.; Vederas, J. C. *Science* **2009**, *325*, 161–165.
- Li, J.-X.; Kadota, S.; Hattori, M.; Yoshimachi, S.; Shiro, M.; Oogami, N.; Mizuno, H.; Namba, T. *Chem. Pharm. Bull.* **1993**, *41*, 832–841.
- Li, J.-X.; Yu, Z.-Y. *Curr. Med. Chem.* **2006**, *13*, 2927–2951.
- Lichtblau, D.; Berger, J. M.; Nakanishi, K. *J. Nat. Prod.* **2002**, *65*, 1501–1504.
- Liu, Y.; Chen, D.; Si, J.; Tu, G.; An, D. *J. Nat. Prod.* **2002**, *65*, 1486–1488.
- Lobstein-Guth, A.; Briancon-Scheid, F.; Anton, R. *J. Chromatogr.* **1983**, *267*, 431–438.

- Lu, Y.; Pan, Y.; Berthod, A. *J. Chromatogr. A* **2008**, *1189*, 10–18.
- Lu, Y.; Sun, C.; Liu, R.; Pan, Y. *J. Chromatogr. A* **2007**, *1146*, 125–130.
- Lu, Y.; Sun, C.; Wang, Y.; Pan, Y. *J. Chromatogr. A* **2007**, *1151*, 31–36.
- Mahadevan, S.; Park, Y. *J. Food. Sci.* **2008**, *73*, 14–19.
- Makovskaya, V.; Dean, J. R.; Tomlinson, W. R.; Comber, M. *Anal. Chim. Acta* **1995**, *315*, 193–200.
- Matthias, H.; Christian, W.; Bjorn, B. *Pharm. Pharmacol. Lett.* **2001**, *1*, 15–16.
- Milligan, S. R.; Kalita, J. C.; Heyerick, A.; Rong, H.; De Cooman, L.; De Keukeleire, D. *J. Clin. Endocr. Metab.* **1999**, *84*, 2249–2252.
- Milligan, S. R.; Kalita, J. C.; Pocock, V.; Van De Kauter, V.; Stevens, J. F.; Deinzer, M. L.; Rong, H.; De Keukeleire, D. *J. Clin. Endocr. Metab.* **2000**, *85*, 4912–4915.
- Murakami, A.; Ashida, H.; Terao, J. *Cancer Lett.* **2008**, *269*, 315–325.
- Nakanishi, K. *Bioorgan. Med. Chem.* **2005**, *13*, 4987–5000.
- Nishida, M.; Yoshimitsu, H. *Chem. Pharm. Bull.* **2011**, *59*, 1243–1249.
- Nishida, M.; Yoshimoto, H.; Nohara, T. *Chem. Pharm. Bull.* **2003**, *51*, 354–356.
- Orians, C. M.; Lower, S.; Fritz, R. S.; Roche, B. M. *Biochem. Syst. Ecol.* **2003**, *31*, 233–247.
- Pauli, G. F. *Attacking Cardenolides in the 1 ppm Range*; Annual Meeting of the Gesellschaft für Arzneipflanzenforschung (GA), Halle, Germany, 1995.
- Pauli, G. F. *Cardenolide aus Adonis aleppica Boiss. – Isolierung und Strukturaufklärung*. Ph.D. Dissertation, Heinrich-Heine-University, Düsseldorf, Germany, 1993.
- Pauli, G. F.; Jaki, B. U.; Lankin, D. C. *J. Nat. Prod.* **2005**, *68*, 133–149.
- Pauli, G. F.; Pro, S. M.; Friesen, J. B. *J. Nat. Prod.* **2008**, *71*, 1489–1508.
- Pellett, J.; Lukulay, P.; Mao, Y.; Bowen, W.; Reed, R.; Ma, M.; Munger, R. C.; Dolan, J. W.; Wrisley, L.; Medwid, K.; Toltl, N. P.; Chan, C. C.; Skibic, M.; Biswas, K.; Wells, K. A.; Snyder, L. R. *J. Chromatogr. A* **2006**, *1101*, 122–135.
- Petrakis, P. V.; Agiomyrgianaki, A.; Christophoridou, S.; Spyros, A.; Dais, P. *J. Agr. Food Chem.* **2008**, *56*, 3200–3207.
- Petrakis, P. V.; Touris, I.; Liouni, M.; Zervou, M.; Kyrikou, I.; Kokkinofa, R.; Theocharis, C. R.; Mavromoustakos, T. M. *J. Agr. Food Chem.* **2005**, *53*, 5293–5303.
- Pietta, P. G.; Mauri, P. L.; Rava, A.; *Chromatographia* **1990**, *29*, 251–253.
- Sakurai, N.; Koeda, M.; Aoki, Y.; Nagai, M. *Chem. Pharm. Bull.* **1995**, *43*, 1475–1482.

- Sakurai, N.; Koeda, M.; Inoue, T.; Nagai, M. *Chem. Pharm. Bull.* **1994**, *42*, 48–51.
- Schinkovitz, A.; Pro, S. M.; Main, M.; Chen, S. N.; Jaki, B. U.; Lankin, D. C.; Pauli, G. F. *J. Nat. Prod.* **2008**, *71*, 1604–1611.
- Schröder, F. C.; Gibson, D. M.; Churchill, A. C. L.; Sojikul, P.; Wursthorn, E. J.; Krasnoff, S. B.; Clardy, J. *Angew. Chem. Int. Edit.* **2007**, *46*, 901–904.
- Shao, Y.; Harris, A.; Wang, M.; Zhang, H.; Cordell, G. A.; Bowman, M.; Lemmo, E. *J. Nat. Prod.* **2000**, *63*, 905–910.
- Shoolery, J. N.; Rogers, M. T. *J. Am. Chem. Soc.* **1958**, *80*, 5121–5135.
- Singh, G.; Singh, G.; Jadeja, D.; Kaur, J. *Crit. Rev. Microbiol.* **2010**, *36*, 259–269.
- Snyder, L. R. *J. Chromatogr.* **1974**, *92*, 223–230.
- Spraul, M.; Braumann, U.; Renault, J.-H.; Thépenier, P.; Nuzillard, J.-M. *J. Chromatogr. A* **1997**, *766*, 255–260.
- Steinbeck, C. *Nat. Prod. Rep.* **2004**, *21*, 512–518.
- Steinberg, D.; Golovnya, M. *CART 6.0 User's Guide*; Salford Systems: San Diego, CA, 2006; pp. 61–63.
- Stermitz, F. R.; Lorenz, P.; Tawara, J. N.; Zenewicz, L. A.; Lewis, K. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1433–1437.
- Stevens, J. F.; Taylor, A. W.; Deinzer, M. L. *J. Chromatogr. A* **1999**, *832*, 97–107.
- Sun, L.; Yan, J.; Nan, Y.; Zhang, H.; Qiu, M. *Molecules* **2008**, *13*, 1712–1721.
- Sun, L.-R.; Qing, C.; Zhang, Y.-L.; Jia, S.-Y.; Li, Z.-R.; Pei, S.-J.; Qiu, M.-H.; Gross, M. L.; Qiu, S.-X. *J. Beilstein J. Org. Chem.* **2007**, *3*, No. 3.
- Sun, L.-R.; Yan, J.; Zhou, L.; Li, Z.-R.; Qiu, M.-H. *Molecules* **2011**, *16*, 5701–5708.
- Tanimura, T.; Pisano, J. J.; Ito, Y.; Bowman, R. L. *Science* **1970**, *169*, 54–56.
- Tursch, B.; Savoir, R.; Ottinger, R.; Chiurdoglu, G. *Tetrahedron Lett.* **1967**, *8*, 539–543.
- Ulrich-Merzenich, G.; Panek, D.; Zeitler, H.; Vetter, H.; Wagner, H. *Indian J. Exp. Biol.* **2010**, *48*, 208–219.
- van Beek, T. A. *Bioorg. Med. Chem.* **2005**, *13*, 5001–5012.
- van Beek, T. A.; van Veldhuizen, A.; Lelyveld, G. P. *J. Nat. Prod.* **1997**, *60*, 735–738.
- van Beek, T. A.; van Veldhuizen, A.; Lelyveld, G. P.; Piron, I.; Lankhorst, P. P. *Phytochem. Anal.* **1993**, *4*, 261–268.

- Wagenaar, F. L.; Hochlowski, J. E.; Pan, J. Y.; Tu, N. P.; Searle, P. A. *J. Chromatogr. A* **2009**, *1216*, 4154–4160.
- Wall, M. E.; Wani, M. C. *J. Ethnopharmacol.* **1996**, *51*, 239–253.
- Walsh, S.; Diamond, D. *Talanta* **1995**, *42*, 561–572.
- Watanabe, K.; Mimaki, Y.; Sakagami, H.; Sashida, Y. *Chem. Pharm. Bull.* **2002**, *50*, 121–125.
- Wei, Y.; Ito, Y. *J. Chromatogr. A* **2006**, *1115*, 112–117.
- Wende, K.; Mugge, C.; Thurow, K.; Schöpke, T.; Lindequist, U. *J. Nat. Prod.* **2001**, *64*, 986–989.
- Wilson, W. K.; Sumpter, R. M.; Warren, J. J.; Rogers, P. S.; Ruan, B.; Schröpfer, Jr. G. J. *J. Lipid Res.* **1996**, *37*, 1529–1555.
- Wolfender, J.-L.; Ndjoko, K.; Hostettmann, K. *Phytochem. Anal.* **2001**, *12*, 2–22.
- Xi, Y.; de Ropp, J. S.; Viant, M. R.; Woodruff, D. L.; Yu, P. *Anal. Chim. Acta* **2008**, *614*, 127–133.
- Xi, Y.; de Ropp, J.; Viant, M.; Woodruff, D.; Yu, P. *Metabolomics* **2006**, *2*, 221–233.
- Xu, H.; Fabricant, D. S.; Piersen, C. E.; Bolton, J. L.; Pezzuto, J. A.; Fong, H.; Totura, S.; Farnsworth, N. R.; Constantinou, A. I. *Phytomedicine* **2002**, *9*, 757–762.
- Ye, W.-C.; Zhang, J.-W.; Che, C.-T.; Ye, T.; Zhao, S.-X. *Planta Med.* **1999**, *65*, 770–772.
- Yen, G.-C.; Chen, H.-Y. *J. Agr. Food Chem.* **1995**, *43*, 27–32.
- Yoshimitsu, H.; Nishida, M.; Sakaguchi, M.; Nohara, T. *Chem. Pharm. Bull.* **2006**, *54*, 1322–1325.
- Zhang, Q.-W.; Ye, W.-C.; Che, C.-T.; Zhao, S.-X. *J. Asian Nat. Prod. Res.* **1999**, *2*, 45–49.
- Zhang, Q.-W.; Ye, W.-C.; Hsiao, W.-L.; Zhao, S.-X.; Che, C.-T. *Chem. Pharm. Bull.* **2001**, *49*, 1468–1470.
- Zhou, L.; Yang, J.-S.; Zou, J.-H.; Tu, G.-Z. *Chem. Pharm. Bull.* **2004**, *52*, 622–624.
- Zhou, L.; Yang, J.-S.; Tu, G.-Z.; Zou, J.-H. *Chem. Pharm. Bull.* **2006**, *54*, 823–826.
- Zhou, L.; Yang, J.-S.; Tu, G.-Z.; Zu, J.-H. *Chem. Pharm. Bull.* **2006**, *54*, 823–826.
- Zhou, L.; Yang, J.-S.; Wu, X.; Zou, J.-H.; Xu, X.-D.; Tu, G.-Z. *Heterocycles*, **2005**, *64*, 1409–1414.
- Zhou, L.; Yang, J.-S.; Zou, J.-H.; Tu, G.-Z. *Chem. Pharm. Bull.* **2004**, *52*, 622–624.
- Zhu, N.; Jiang, Y.; Wang, M.; Ho, C.-T. *J. Nat. Prod.* **2001**, *64*, 627–629.

## APPENDICES

### Appendix A: Compound Index

#	Compound Name	L <sup>a</sup>	I <sup>b</sup>	M <sup>c</sup>	Page(s)
<i>Actaea</i> sp.					
11	(12 <i>R</i> )-12-Acetoxy-7,8-didehydro-(23 <i>R</i> ,24 <i>S</i> )-23,24-dihydroxy-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-(16 <i>S</i> ,22 <i>R</i> )-16,23;22,25-binoxoside		●	●	51, 52, 85–91, 109, 110
12	(12 <i>R</i> )-12-Acetoxy-7,8-didehydro-(23 <i>R</i> ,24 <i>S</i> )-23,24-dihydroxy-3- <i>O</i> - $\alpha$ -L-arabinopyranosylacta-(16 <i>S</i> ,22 <i>R</i> )-16,23;22,25-binoxoside		●	●	51, 52, 85–91, 127, 128, 131, 132
13	(15 <i>R</i> )-15,21,25-Trihydroxy-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-(16 <i>S</i> ,23 <i>R</i> ,24 <i>S</i> )-16,23;16,24-binoxoside			●	51, 52, 127, 128, 131, 132
14	(12 <i>R</i> )-12-Acetoxy-(24 <i>R</i> ,25 <i>R</i> )-24,25-epoxy-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-(16 <i>S</i> ,23 <i>R</i> )-16,23;23,26-binoxoside		●	●	51, 52, 60, 61, 109, 110
15	(12 <i>R</i> )-12-Acetoxy-(24 <i>R</i> ,25 <i>S</i> )-24,25-epoxy-(26 <i>R</i> & <i>S</i> )-26-hydroxy-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-(16 <i>S</i> ,23 <i>R</i> )-16,23;23,26-binoxoside		●	●	51, 52, 60, 61, 90, 96–98, 108
16	(15 <i>R</i> )-15,25-Dihydroxy-3- <i>O</i> - $\alpha$ -L-arabinopyranosylacta-(16 <i>S</i> ,23 <i>R</i> ,24 <i>S</i> )-16,23;16,24-binoxoside		●		51, 52
17	(15 <i>R</i> )-15,25-Dihydroxy-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-(16 <i>S</i> ,23 <i>R</i> ,24 <i>S</i> )-16,23;16,24-binoxoside		●		51, 52
19	(12 <i>R</i> )-12-Acetoxy-(24 <i>R</i> ,25 <i>R</i> )-24,25-epoxy-3- <i>O</i> - $\alpha$ -L-arabinopyranosylacta-(16 <i>S</i> ,23 <i>R</i> )-16,23;23,26-binoxoside		●	●	60, 61, 90, 95, 128–132
20	(12 <i>R</i> )-12-Acetoxy-(15 <i>R</i> )-15,25-dihydroxy-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-(16 <i>S</i> ,23 <i>R</i> ,24 <i>S</i> )-16,23;16,24-binoxoside		●	●	60, 61, 95, 128–132
29	(12 <i>R</i> )-12-Acetoxy-(24 <i>R</i> ,25 <i>S</i> )-24,25-epoxy-(26 <i>R</i> & <i>S</i> )-26,27-dihydroxy-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-(16 <i>S</i> ,23 <i>R</i> )-16,23;23,26-			●	85–91
30	(23 <i>R</i> )-23-Acetoxy-(24 <i>S</i> )-24,25-epoxy-(15 <i>R</i> )-15-hydroxy-16-oxo-3- <i>O</i> - $\beta$ -D-xylopyranosylactanoside			●	95–98, 128–132
31	(15 <i>R</i> )-15,25-Dihydroxy-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-(16 <i>S</i> ,23 <i>R</i> ,24 <i>R</i> )-16,23;16,24-binoxoside			●	96–98
32	(24 <i>S</i> )-24-Acetoxy-(15 <i>R</i> ,16 <i>R</i> )-15,16,25-trihydroxy-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-(23 <i>S</i> )-16,23-monoxoside		●		96–98
33	(23 <i>R</i> )-23-Acetoxy-(24 <i>S</i> )-24,25-epoxy-(15 <i>R</i> )-15-hydroxy-16-oxo-3- <i>O</i> - $\alpha$ -L-arabinopyranosylactanoside	●			109, 110

## Appendix A (Continued)

#	Compound Name	L <sup>a</sup>	I <sup>b</sup>	M <sup>c</sup>	Page(s)
<i>Actaea sp.</i>					
34	(24 <i>R</i> )-24,25-Dihydroxy-15-oxoacta-(16 <i>R</i> ,23 <i>R</i> )-16,23-monoxol	●			109, 110
35	(15 <i>R</i> )-15,25-Dihydroxyacta-(16 <i>S</i> ,23 <i>R</i> ,24 <i>S</i> )-16,23;16,24-binoxol	●			109, 110
36	3-Deoxy-8,9-didehydro-(24 <i>R</i> )-24,25-dihydroxy-(3 <i>S</i> ,10 <i>S</i> )-3,10-epoxy-15-oxo-9,10-secoacta-(16 <i>R</i> ,23 <i>R</i> )-16,23-monoxol	●			109, 110
37	7,8-Didehydro-(24 <i>R</i> )-24,25-dihydroxy-15-formyl-16-oxo-15,16-seco-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-(23 <i>R</i> )-16,23-monoxoside	●			109, 110
38	(16 <i>R</i> ,24 <i>R</i> )-16,24-Dihydroxy-23-oxo-25,26,27-trinoracta-16,24-carbamonal	●			109, 110
39	(11 <i>S</i> ,24 <i>S</i> )-11,24,25-Trihydroxy-7,8,16,17,20,22,23, <i>N</i> -octadehydro-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-16,23-monazoside	●			109, 110
<i>Camellia sinensis</i>					
21	Epicatechin gallate (ECg)			●	70, 71
22	Epigallocatechin gallate (EGCg)			●	70, 71
23	Epicatechin (EC)			●	70, 71
24	Epigallocatechin (EGC)			●	70, 71
<i>Ginkgo biloba</i>					
1	Isorhamnetin (IR)	●	●		40, 41
2	Kaempferol (KF)	●	●		40, 41
3	Quercetin (QC)	●	●		40, 41

**Appendix A (Continued)**

#	Compound Name	L <sup>a</sup>	I <sup>b</sup>	M <sup>c</sup>	Page(s)
<i>Ginkgo biloba</i>					
4	Bilobalide (BB)		●	●	40, 41, 54–57
5	Ginkgolide A (GA)		●	●	40, 41, 54–57, 73–76
6	Ginkgolide B (GB)		●	●	40, 41, 54–57, 73–76
7	Ginkgolide C (GC)		●	●	40, 41, 54–57, 77, 78
8	Ginkgolide J (GJ)		●	●	40, 41, 54–57, 77, 78
<i>Humulus lupulus</i>					
9	Xanthohumol (XN)		●	●	43, 58–60
10	6-Prenylnaringenin (6-PN)		●	●	43, 58–60
18	1,2-Dihydroxanthohumol (H <sub>2</sub> -XN)		●	●	58-60
40	$\alpha$ -Amyrin			●	141
41	$\beta$ -Amyrin			●	141
42	$\beta$ -Sitosterol		●		141
43	1-Tetracosanol		●		141
44	Unknown			●	144, 145
45	Unknown			●	144, 145

**Appendix A (Continued)**

#	Compound Name	L <sup>a</sup>	I <sup>b</sup>	M <sup>c</sup>	Page(s)
<i>Humulus lupulus</i>					
46	Unknown			●	144, 145
47	Unknown			●	144, 145
48	Unknown			●	144, 145
49	Unknown			●	144, 145
50	Unknown			●	144, 145
51	Unknown			●	144, 145
52	Unknown			●	144, 145
<i>Oplopanax horridus</i>					
25	3 $\alpha$ -Hydroxylup-20(29)-ene-23,28-dioic acid		●	●	79–84, 137–140
26	3 $\alpha$ -Hydroxyolean-12-ene-23,28-dioic acid			●	81, 82, 137–140
27	Monoglyceride			●	79–84, 137–140
28	Polyne analogue			●	79–84, 137–140

<sup>a</sup> L = data taken from the literature

<sup>b</sup> I = isolated in the present study

<sup>c</sup> M = identified in mixtures

## Appendix B: <sup>1</sup>H NMR Spectra and Data of Isolated Compounds

#	Compound Name	Page(s)
Compound 1	Isorhamnetin	167, 168
Compound 2	Kaempferol	167, 168
Compound 3	Quercetin	167, 168
Compound 4	Bilobalide	169, 171
Compound 5	Ginkgolide A	170, 171
Compound 6	Ginkgolide B	170, 171
Compound 7	Ginkgolide C	170, 171
Compound 8	Ginkgolide J	170, 171
Compound 9	Xanthohumol	172, 173
Compound 10	6-Prenylnaringenin	172, 173
Compound 11	(12 <i>R</i> )-12-Acetoxy-7,8-didehydro-(23 <i>R</i> ,24 <i>S</i> )-23,24-dihydroxy-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-(16 <i>S</i> ,22 <i>R</i> )-16,23;22,25-binoxoside	174
Compound 14	(12 <i>R</i> )-12-Acetoxy-(24 <i>R</i> ,25 <i>R</i> )-24,25-epoxy-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-(16 <i>S</i> ,23 <i>R</i> )-16,23;23,26-binoxoside	175
Compound 15	(12 <i>R</i> )-12-Acetoxy-(24 <i>R</i> ,25 <i>S</i> )-24,25-epoxy-(26 <i>R</i> & <i>S</i> )-26-hydroxy-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-(16 <i>S</i> ,23 <i>R</i> )-16,23;23,26-binoxoside	176
Compound 16	(15 <i>R</i> )-15,25-Dihydroxy-3- <i>O</i> - $\alpha$ -L-arabinopyranosylacta-(16 <i>S</i> ,23 <i>R</i> ,24 <i>S</i> )-16,23;16,24-binoxoside	177
Compound 25	3 $\alpha$ -Hydroxylup-20(29)-ene-23,28-dioic acid	179
Compound 32	(24 <i>S</i> )-24-Acetoxy-(15 <i>R</i> ,16 <i>R</i> )-15,16,25-trihydroxy-3- <i>O</i> - $\beta$ -D-xylopyranosylacta-(23 <i>S</i> )-16,23-monoxoside	178

Appendix B (Continued)

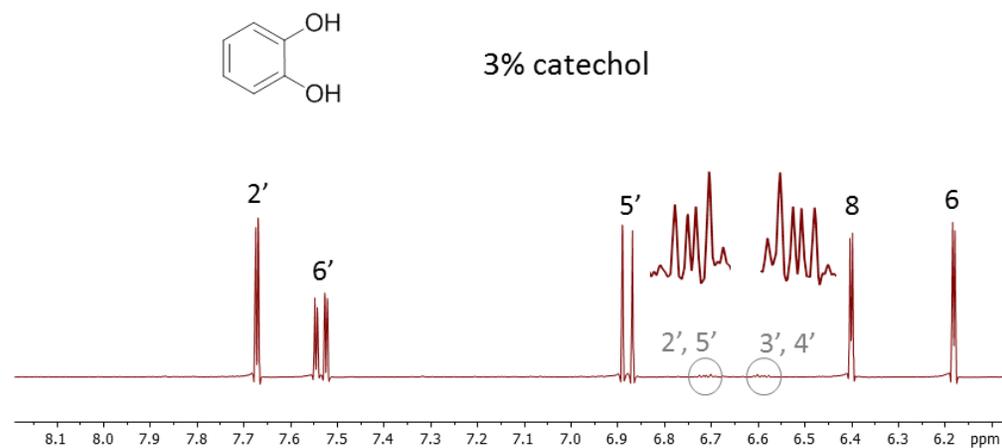
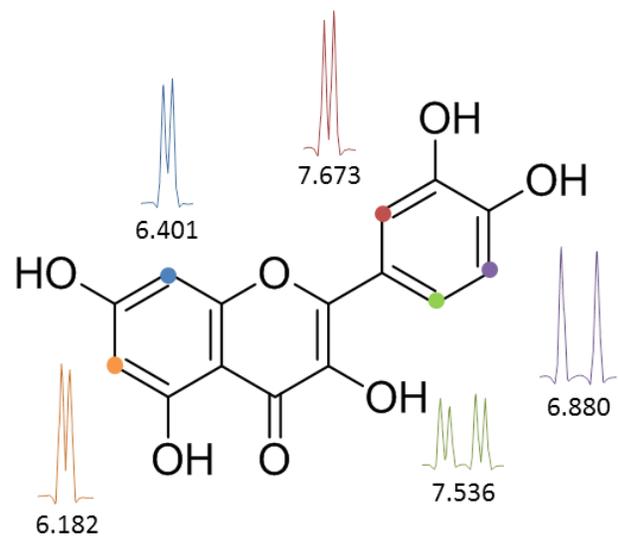
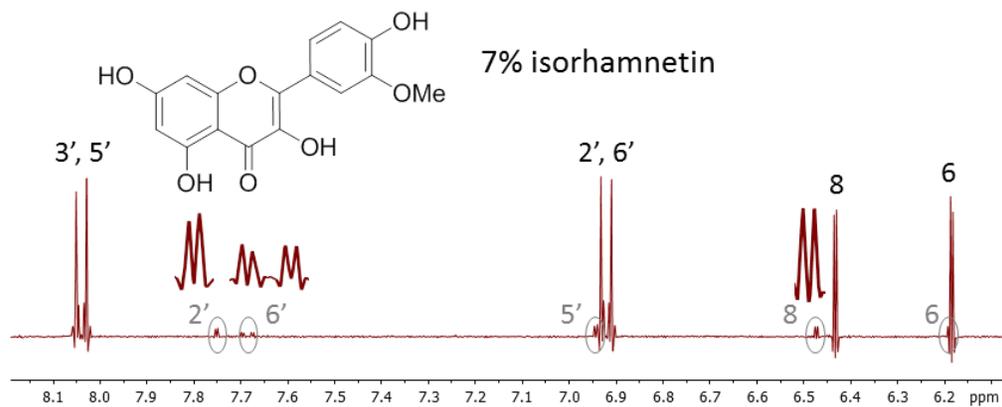
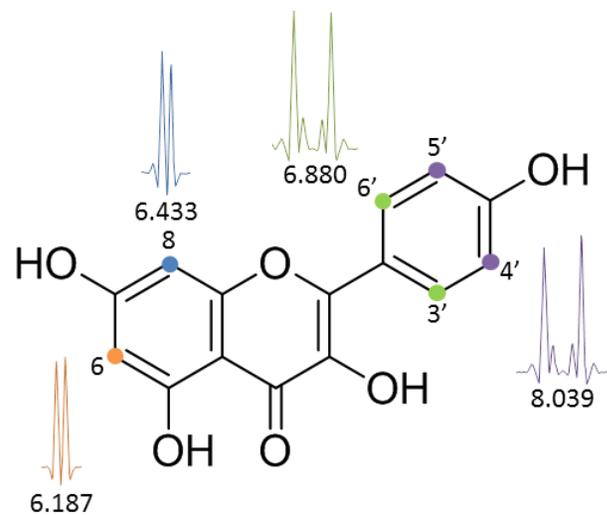


Figure B1. <sup>1</sup>H NMR Spectra of Compounds 1–3 (DMSO-*d*<sub>6</sub>, 400 MHz)

## Appendix B (Continued)

TABLE B1. <sup>1</sup>H NMR ASSIGNMENTS FOR COMPOUNDS 1–3<sup>a</sup>

	Compound 1		Compound 2		Compound 3	
	$\delta_{\text{H}}$ (ppm)	$m^{\text{b}}$ ( $J$ in Hz)	$\delta_{\text{H}}$ (ppm)	$m^{\text{b}}$ ( $J$ in Hz)	$\delta_{\text{H}}$ (ppm)	$m^{\text{b}}$ ( $J$ in Hz)
H-6	6.193	d (2.0)	6.187	d (2.1)	6.182	d (2.0)
H-8	6.476	d (2.0)	6.433	d (2.1)	6.401	d (2.0)
H-2'	7.753	d (2.0)	6.880	$m^{\text{c}}$	7.673	d (2.0)
H-3'	-	-	8.039	$m^{\text{c}}$	-	-
H-5'	6.945	d (8.4)	8.039	$m^{\text{c}}$	6.880	d (8.4)
H-6'	7.688	dd (8.4, 2.0)	6.880	$m^{\text{c}}$	7.536	dd (8.4, 2.0)
OMe	3.842	s	-	-	-	-

<sup>a</sup> The spectra were recorded in DMSO- $d_6$  at 400 MHz

<sup>b</sup> Multiplicity of the signal: s = singlet; d = doublet; dd = doublet of doublets; m = multiplet

<sup>c</sup> AA'XX' spin system

Appendix B (Continued)

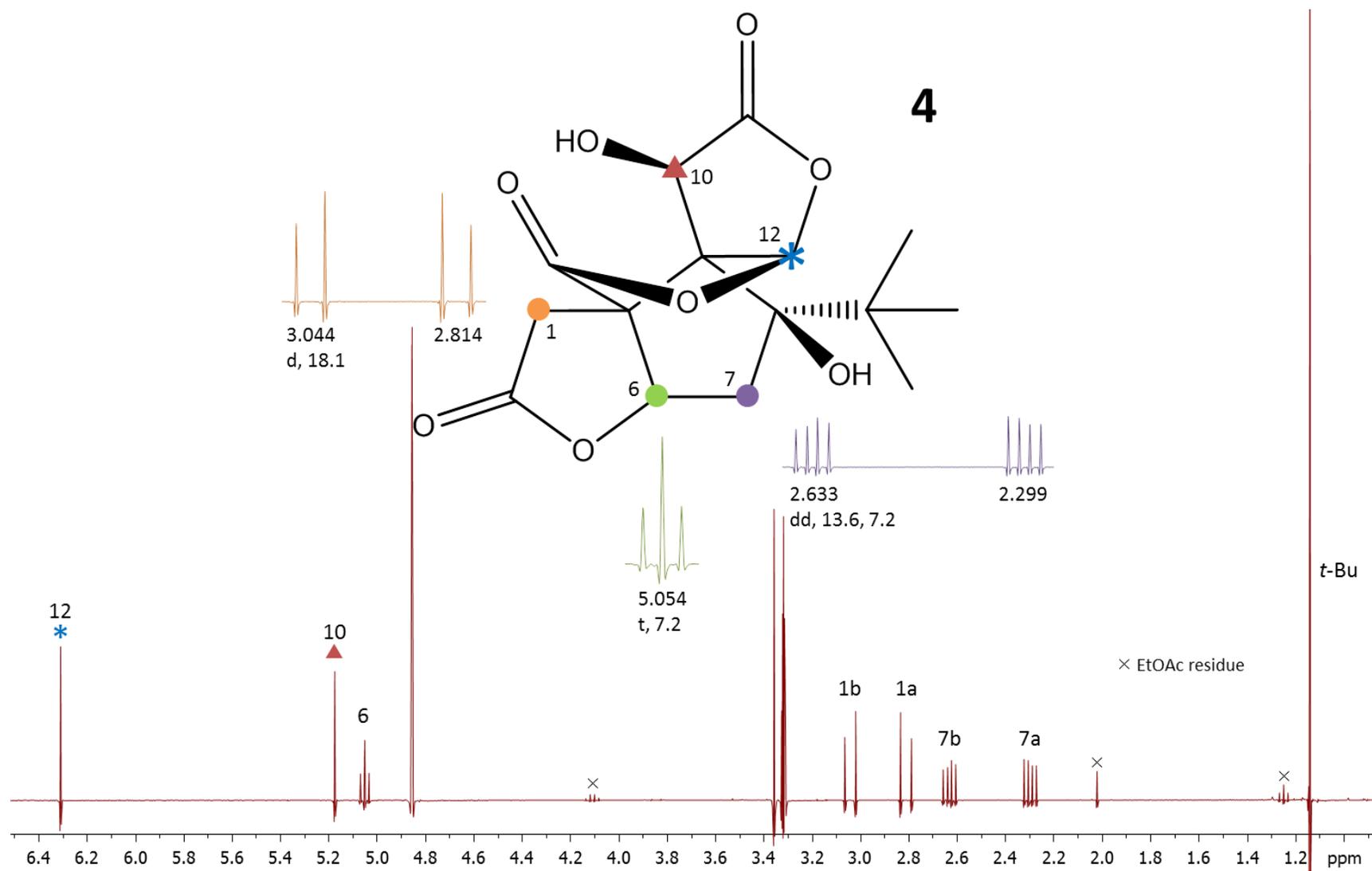


Figure B2. <sup>1</sup>H NMR Spectra of Compounds 4 (methanol-*d*<sub>4</sub>, 400 MHz)

Appendix B (Continued)

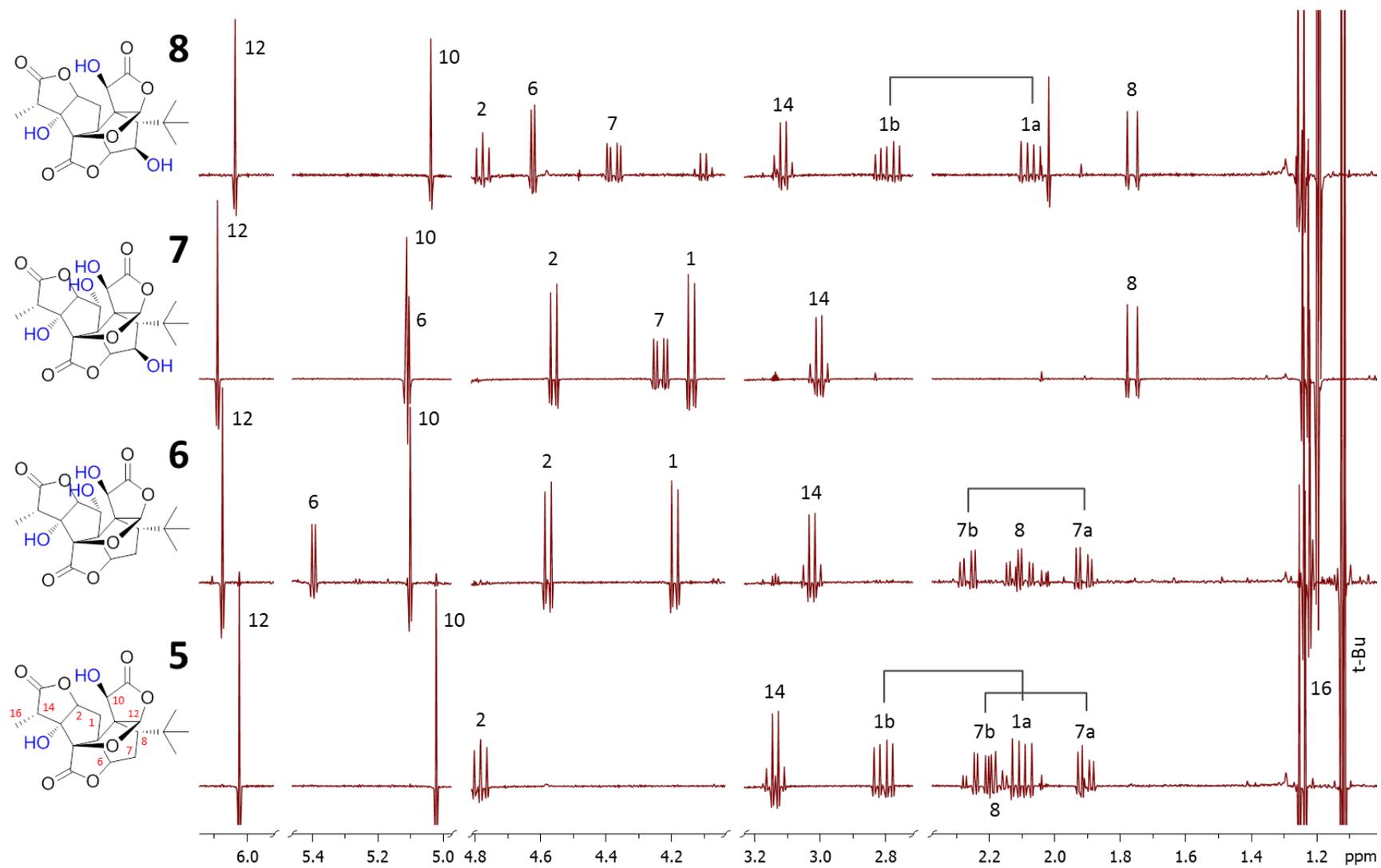


Figure B3.  $^1\text{H}$  NMR Spectra of Compounds 5–8 (methanol- $d_4$ , 400 MHz)

## Appendix B (Continued)

**TABLE B2. <sup>1</sup>H NMR ASSIGNMENTS FOR COMPOUNDS 4–8<sup>a</sup>**

	Compound 4		Compound 5		Compound 6		Compound 7		Compound 8	
	$\delta_{\text{H}}$	$m^{\text{b}}$ (J in Hz)	$\delta_{\text{H}}$	$m^{\text{b}}$ (J in Hz)	$\delta_{\text{H}}$	$m^{\text{b}}$ (J in Hz)	$\delta_{\text{H}}$	$m^{\text{b}}$ (J in Hz)	$\delta_{\text{H}}$	$m^{\text{b}}$ (J in Hz)
H-1a	2.814	d (18.1)	2.094	dd (15.5, 8.2)	4.188	d (7.8)	4.137	d (7.6)	2.068	dd (15.5, 8.0)
H-1b	3.044	d (18.1)	2.805	dd (15.5, 7.2)	-	-	-	-	2.802	dd (14.7, 7.5)
H-2	-	-	4.791	t (7.6)	4.575	d (7.8)	4.558	d (7.7)	4.774	t (7.6)
H-6	5.054	t (7.0)	~5.00	(overlapped <sup>c</sup> )	5.394	d (4.1)	5.109	d (4.5)	4.621	d (4.2)
H-7a	2.299	dd (13.6, 7.2)	1.900	dd (13.8, 5.2)	1.905	dd (14.2, 4.9)	4.232	dd (12.4, 4.4)	4.374	dd (12.3, 4.3)
H-7b	2.633	dd (13.6, 7.2)	2.218	dd (13.8, 4.0)	2.260	dd (13.7, 4.9)	-	-	-	-
H-8	-	-	2.164	dd (13.5, 5.2)	2.084	dd (13.8, 4.3)	1.758	d (12.4)	1.758	d (12.3)
H-10	5.177	s	5.021	s	5.099	s	5.110	s	5.037	s
H-12	6.312	s	6.023	s	6.074	s	6.089	s	6.036	s
H-14	-	-	3.133	q (7.2)	3.022	q (7.1)	3.001	q (7.1)	3.109	q (7.3)
H-16	-	-	1.240	d (7.2)	1.226	d (7.1)	1.230	d (7.1)	1.243	d (7.2)
<i>t</i> -Bu	1.142	s	1.110	s	1.120	s	1.194	s	1.186	s

<sup>a</sup> The spectra were recorded in methanol-*d*<sub>4</sub> at 400 MHz. The assignments were made based on the literature (van Beek *et al.*, 2005)

<sup>b</sup> Multiplicity of the signal: s = singlet; d = doublet; dd = doublet of doublets; t = triplet; q = quartet

<sup>c</sup> Signal was overlapped by HOD

Appendix B (Continued)

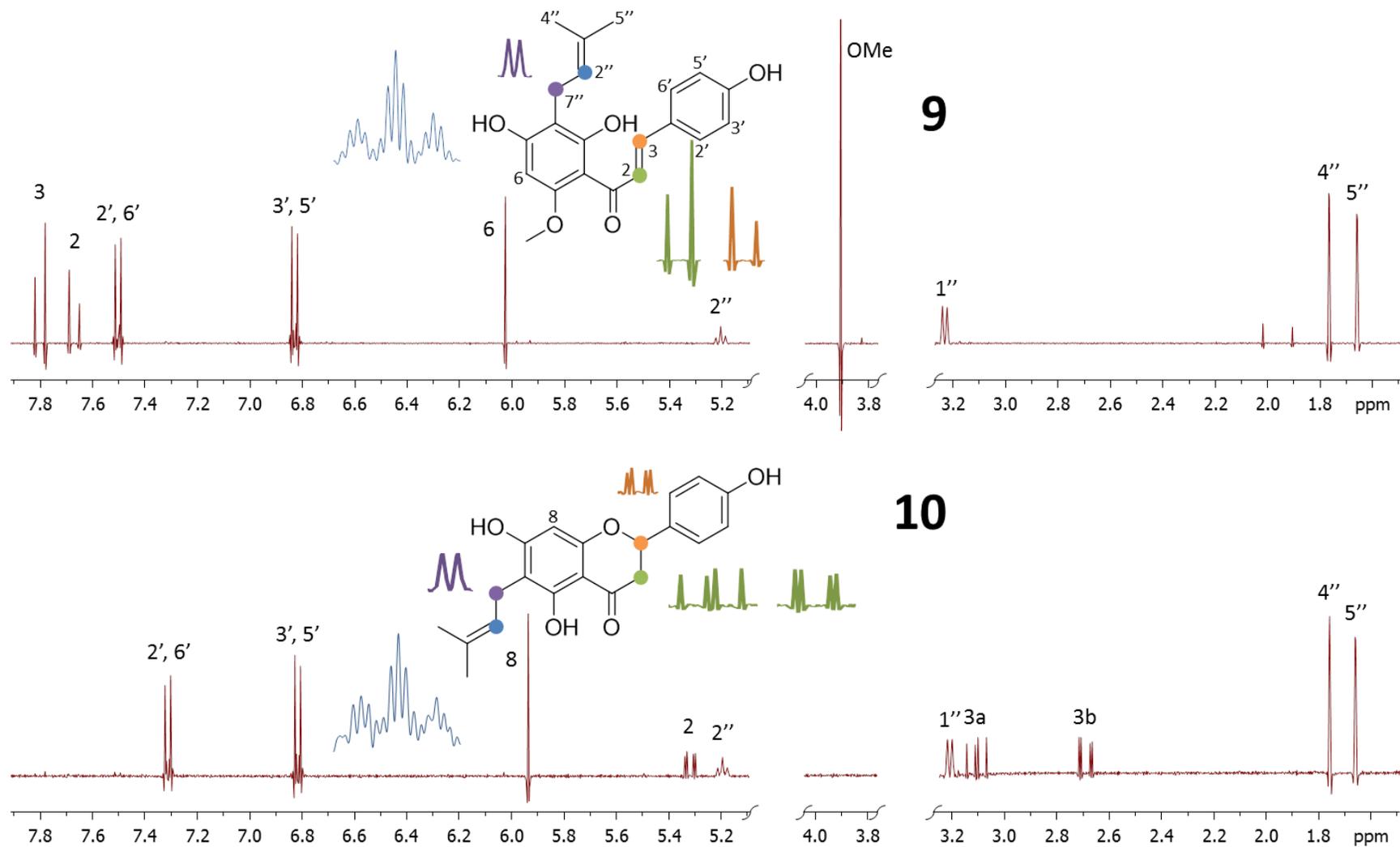


Figure B4. <sup>1</sup>H NMR Spectra of Compounds **9** and **10** (methanol-*d*<sub>4</sub>, 400 MHz)

## Appendix B (Continued)

TABLE B3. <sup>1</sup>H NMR ASSIGNMENTS FOR COMPOUNDS 9 AND 10<sup>a</sup>

	Compound 9		Compound 10	
	$\delta_{\text{H}}$	$m^{\text{b}}$ (J in Hz)	$\delta_{\text{H}}$	$m^{\text{b}}$ (J in Hz)
H-2	7.670	d (15.6)	5.313	dd (12.8, 3.2)
H-3a	7.801	d (15.5)	2.684	dd (16.8, 3.2)
H-3b	-	-	3.101	dd (16.8, 12.8)
H-6	6.023	s	-	-
H-8	-	-	5.933	s
H-2'	7.503	$m^{\text{c}}$	7.311	$m^{\text{c}}$
H-3'	6.828	$m^{\text{c}}$	6.814	$m^{\text{c}}$
H-5'	6.828	$m^{\text{c}}$	6.814	$m^{\text{c}}$
H-6'	7.503	$m^{\text{c}}$	7.311	$m^{\text{c}}$
H-1''	3.229	br d (7.2)	3.204	br d (7.2)
H-2''	5.198	ts	5.189	ts
H-4''	1.760	d (0.8)	1.751	d (0.8)
H-5''	1.653	d (0.8)	1.653	d (0.8)

<sup>a</sup> The spectra were recorded in methanol-*d*<sub>4</sub> at 400 MHz. The assignments were made based on the literature (Hänsel *et al.*, 1988)

<sup>b</sup> Multiplicity of the signal: s = singlet; d = doublet; br d = broad doublet; dd = doublet of doublets; tq = triplet of septets; m = multiplet

<sup>c</sup> AA'XX' spin system

## Appendix B (Continued)

Pages 174–178:

- Methyl signals (used for dereplication by CBTs)
- ▲ Fingerprinting signals of the cyclopropyl protons
- \* Fingerprinting signals of the side chain
- ◆ Fingerprinting signals of the sugar moiety

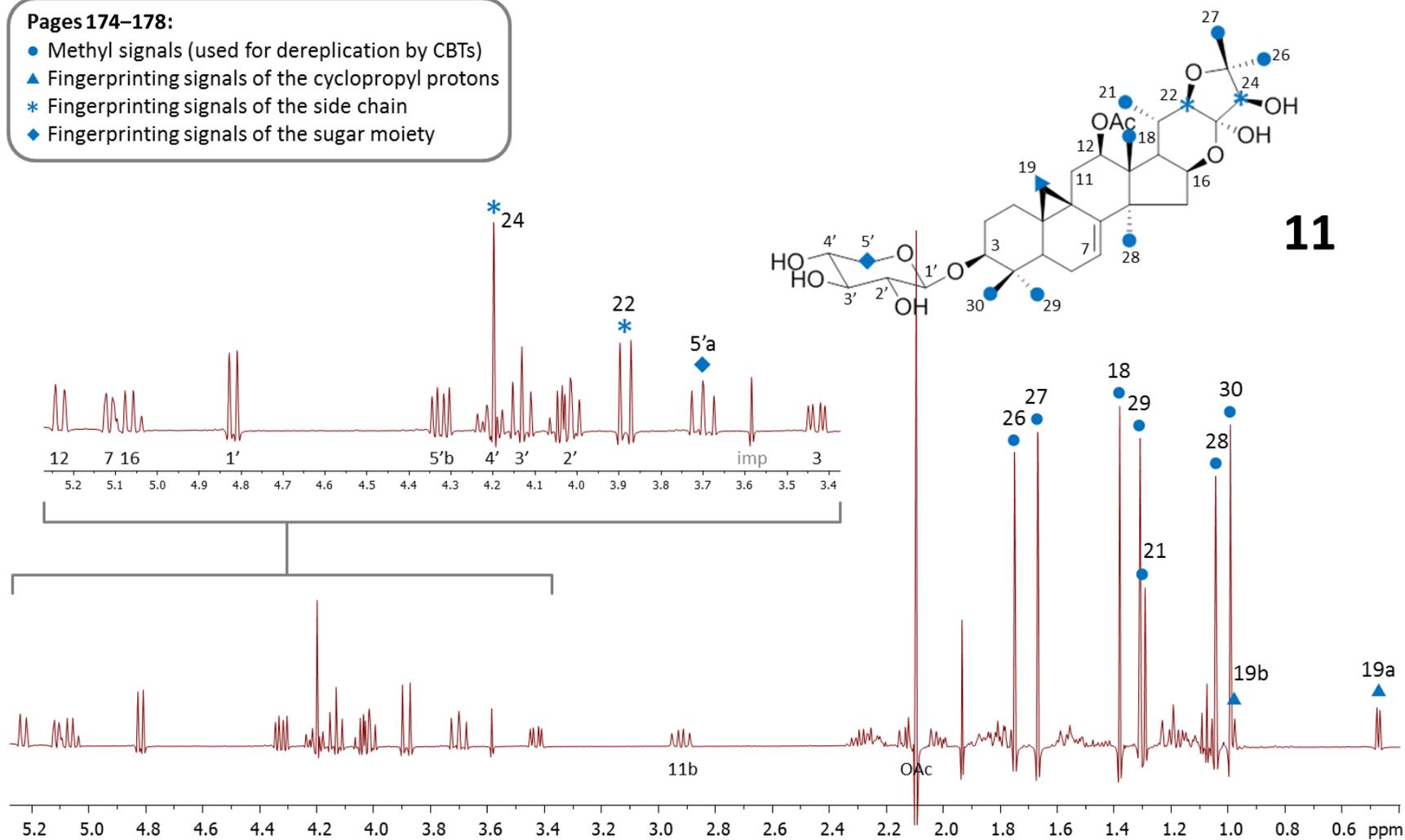


Figure B5. <sup>1</sup>H NMR Spectra of Compound 11 (pyridine-*d*<sub>5</sub>, 400 MHz)

Appendix B (Continued)

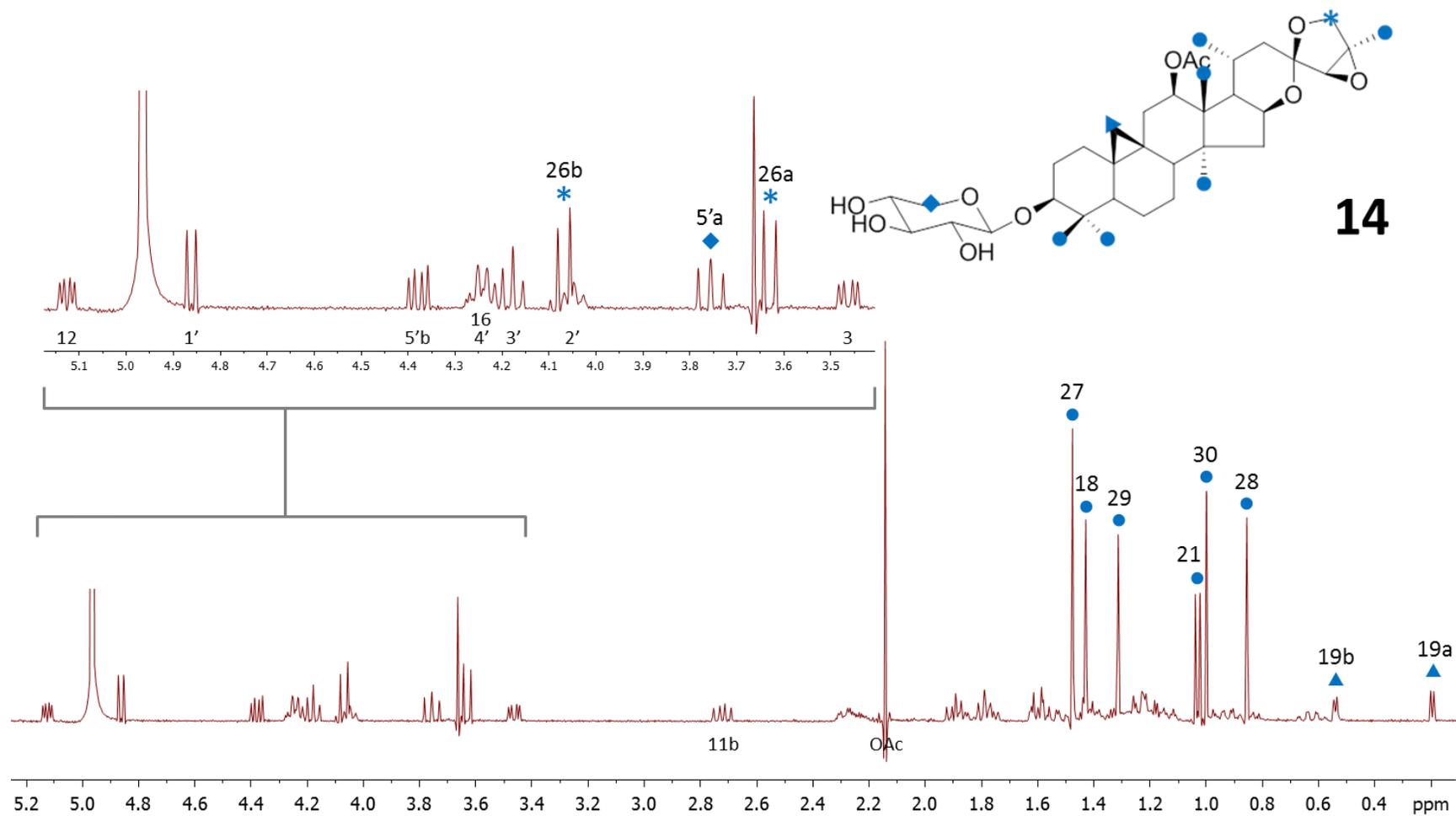


Figure B6. <sup>1</sup>H NMR Spectra of Compound 14 (pyridine-*d*<sub>5</sub>, 400 MHz)

## Appendix B (Continued)

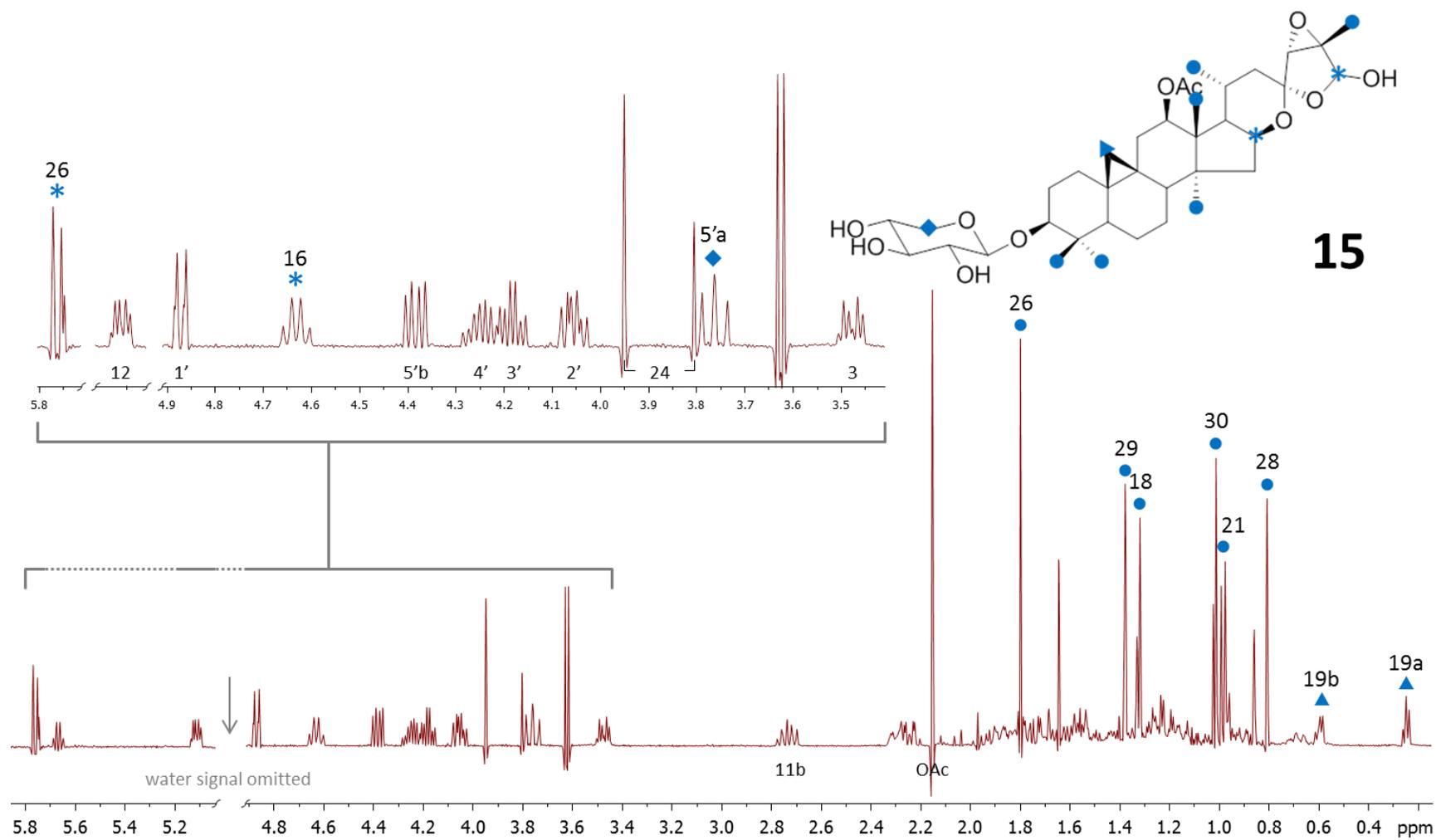


Figure B7.  $^1\text{H}$  NMR Spectra of Compound 15 ( $\text{pyridine-}d_5$ , 400 MHz)

Appendix B (Continued)

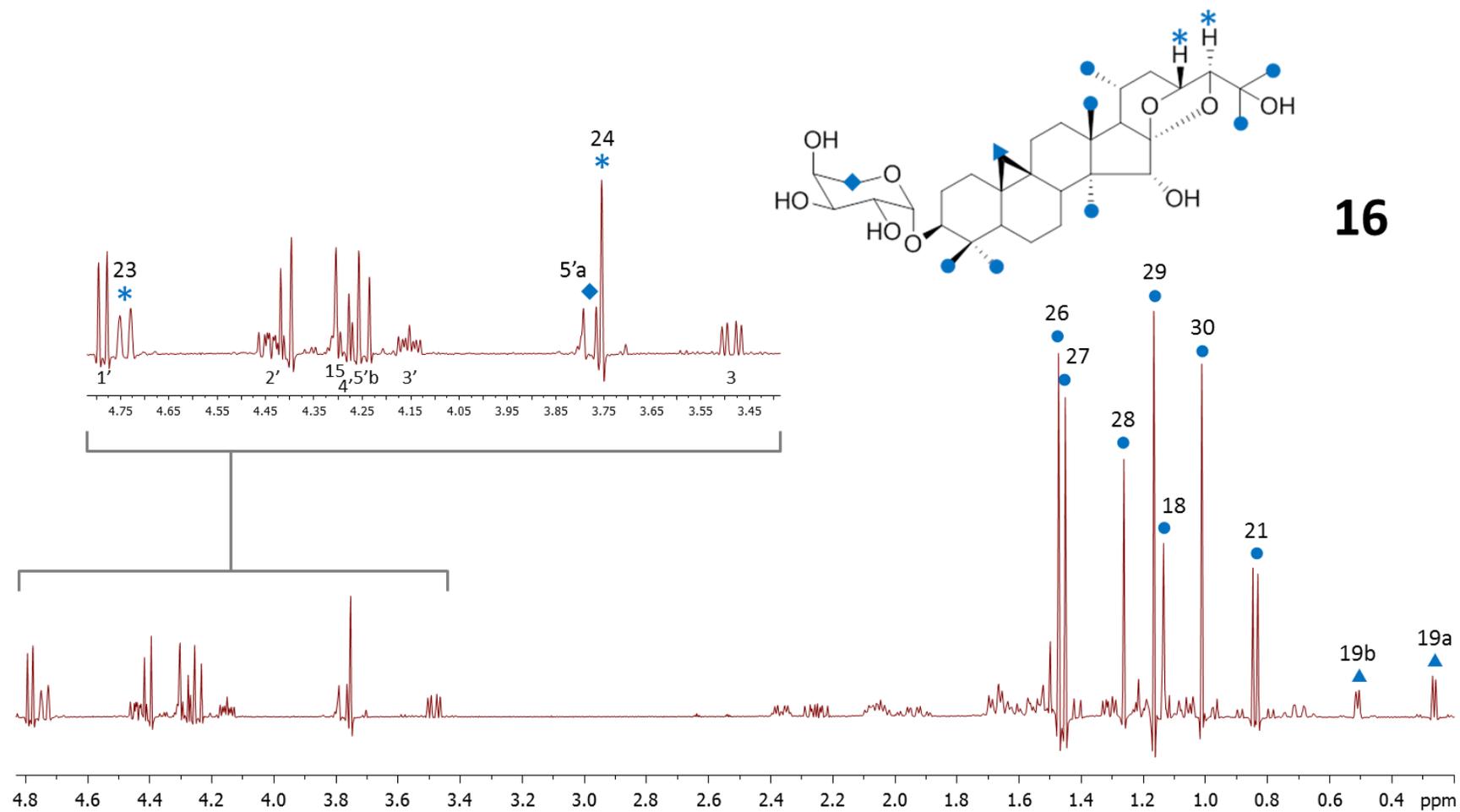


Figure B8. <sup>1</sup>H NMR Spectra of Compound 16 (pyridine-d<sub>5</sub>, 400 MHz)

Appendix B (Continued)

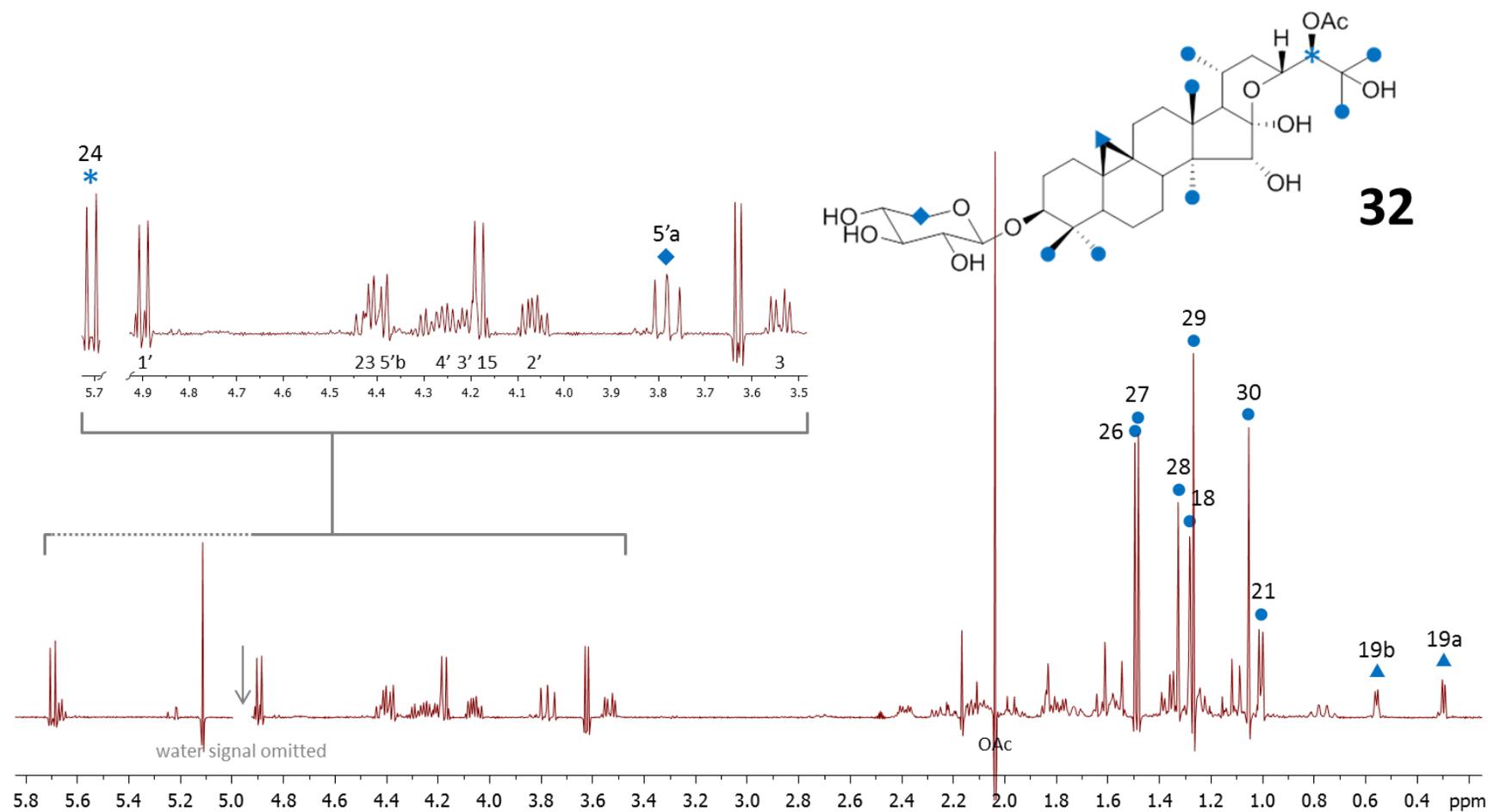


Figure B9. <sup>1</sup>H NMR Spectra of Compound 32 (pyridine-*d*<sub>5</sub>, 400 MHz)

Appendix B (Continued)

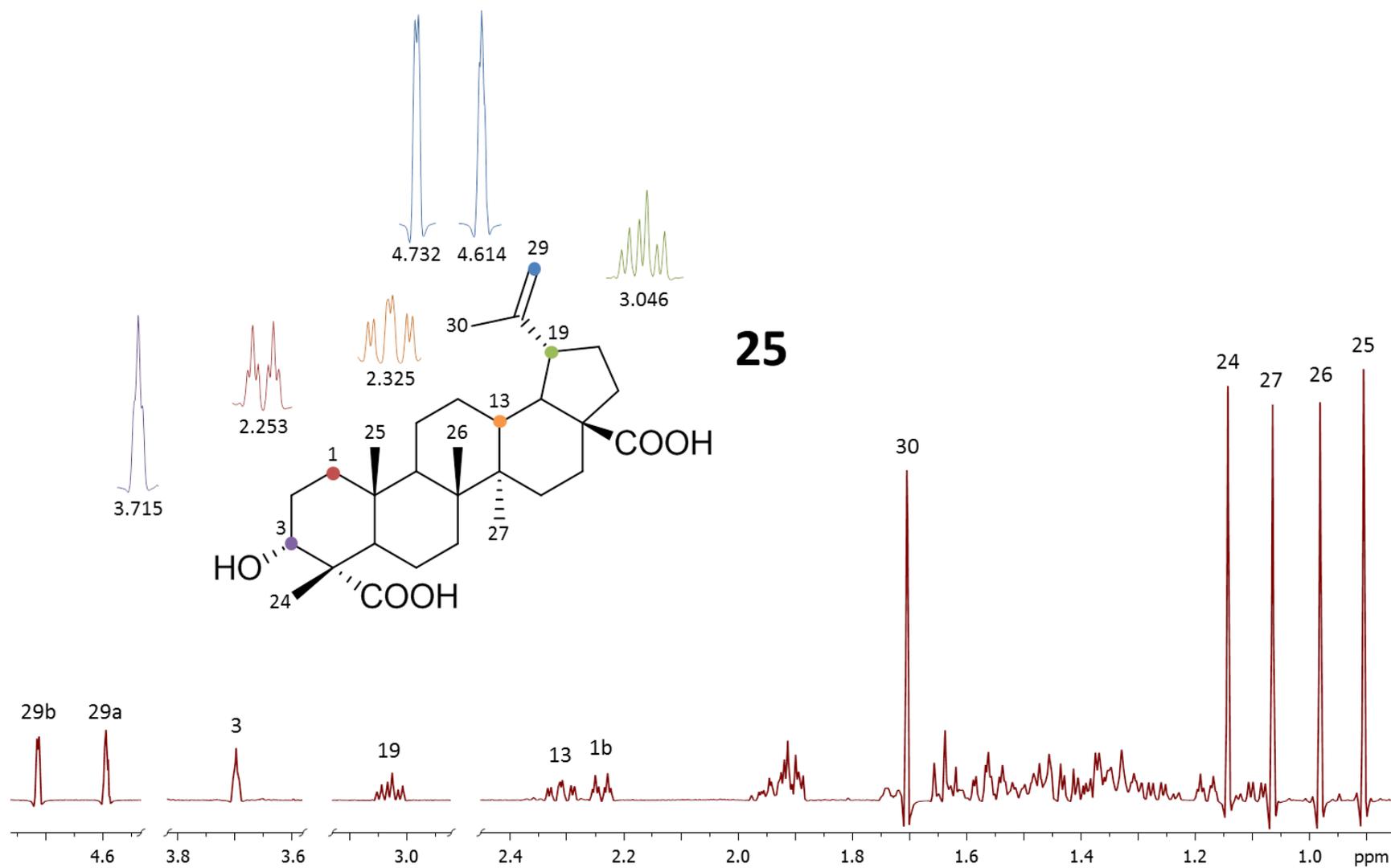


Figure B10. <sup>1</sup>H NMR Spectra of Compound 25 (pyridine-*d*<sub>5</sub>, 400 MHz)

## Appendix C: VBA Code of ActaPredict

```
Sub Predict()  
  
    'Initialization  
    Sheet1.Unprotect  
    Sheet1.Range("B22:N22") = ""  
    Sheet1.Range("E5") = ""  
    Sheet1.Range("L15") = ""  
    Sheet1.Range("J21") = "12-"  
    Sheet1.Image1.Picture = LoadPicture("")  
    Sheet1.Image2.Picture = LoadPicture("")  
    Sheet1.Image3.Picture = LoadPicture("")  
    Sheet1.Image4.Picture = LoadPicture("")  
    Sheet1.Image5.Picture = LoadPicture("")  
    Sheet1.Image6.Picture = LoadPicture("")  
    Sheet1.Image7.Picture = LoadPicture("")  
    Sheet1.Image8.Picture = LoadPicture("")  
    Sheet1.Image9.Visible = False  
  
    'Input Check  
    s = 0  
    For i = 5 To 11  
        m = i - 3  
        If Sheet1.Range("C" & i) = "" Then  
            a = 0  
            Exit For  
        Else  
            If Sheet1.Range("C" & i) < s Or IsNumeric(Sheet1.Range("C" &  
                i)) = False Then  
                a = 1  
                c = 0  
                Exit For  
            Else  
                s = Sheet1.Range("C" & i)  
                c = 1  
            End If  
        End If  
    Next  
  
    If m < 5 Then  
        c = 0  
    End If  
  
    If a = 0 And m >= 5 And m < 8 Then  
        For j = (i + 1) To 11  
            If Sheet1.Range("C" & j) <> "" Then  
                c = 0  
                Exit For  
            Else
```

**Appendix C (Continued)**

```

        c = 1
    End If
Next
End If

Sheet1.Range("P5") = i
Sheet1.Range("P6") = m

If c = 0 Then
    MsgBox "Error! Your data entry is incomplete or incorrect!",
        vbCritical, "ActaFinder v1.0"
Else

Sheet1.Range("C5") = me1
Sheet1.Range("C6") = me2
Sheet1.Range("C7") = me3
Sheet1.Range("C8") = me4
Sheet1.Range("C9") = me5
Sheet1.Range("C10") = me6
Sheet1.Range("C11") = me7
Sheet1.Range("C12") = me8

'n of Me (w/ delta < 2) = 5
If me6 = 0 Or me6 > 1.9 Then
    If WorksheetFunction.Average(Sheet1.Range("C5:C9")) < 1.15 Then
        Sheet1.Range("B22") = "Cimilactone"
        Sheet1.Range("E5") = ""
        Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Cimilactone.emf")
    Else
        If WorksheetFunction.Average(Sheet1.Range("C5:C9")) > 1.2 Then
            Sheet1.Range("B22") = "Foetidonol"
            Sheet1.Range("E5") = ""
            Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Foetidonol.emf")
        Else
            Sheet1.Range("B22") = "Trinorcimicidol"
            Sheet1.Range("E5") = "Actanol"
            Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Trinorcimicidol.emf")
        End If
    End If
End If

'n of Me (w/ delta < 2) = 6
If me6 <> 0 And me7 = 0 Or me7 > 1.9 Then
    If me6 > 1.83 Then
        Sheet1.Range("B22") = "25-dehydrocimigenol"
    End If
End If

```

## Appendix C (Continued)

```

        Sheet1.Range("E5") = "Acta-16,23;16,24-binoxol"
        Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\25-
dehydrocimigenol.emf")
    Else
        If WorksheetFunction.Average(Sheet1.Range("C5:C9")) > 1.2 Then
            Sheet1.Range("B22") = "21-hydroxycimigenol"
            Sheet1.Range("E5") = "Acta-16,23;16,24-binoxol"
            Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\21-
hydroxycimigenol.emf")
        Else
            Sheet1.Range("B22") = "Acteol"
            Sheet1.Range("E5") = "Acta-16,23;23,26-binoxol"
            Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Acteol.emf")
        End If
    End If
End If

'n of Me (w/ delta < 2) = 7
If me7 <> 0 And me8 = 0 Or me8 > 1.9 Then
    If me4 > 1.5 Then
        'node #14
        Sheet1.Range("B22") = "15,16-secocimicidol"
        Sheet1.Range("E5") = ""
        Sheet1.Image1.Picture = LoadPicture("")
    Else
        If me7 > 1.75 Then
            'node #13
            Sheet1.Range("B22") = "Cimiracemoside"
            Sheet1.Range("E5") = "Acta-16,23;22,25-binoxol"
            Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Cimiracemoside.emf")
        Else
            If me3 > 1.05 Then
                If me7 > 1.59 Then
                    If me1 > 1.01 Then
                        'node #12
                        Sheet1.Range("B22") = "Cimicidol"
                        Sheet1.Range("E5") = "16,23-dioxo-actanol"
                        Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Cimicidol.emf")
                    Else
                        If me7 > 1.75 Then
                            'node #11
                            Sheet1.Range("B22") = "Cimiracemoside"
                            Sheet1.Range("E5") = "Acta-16,23;22,25-binoxol"

```

## Appendix C (Continued)

```

Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Cimiracemoside.emf")
    Else
        'node #10
        Sheet1.Range("B22") = "Cimigenol"
        Sheet1.Range("E5") = "Acta-16,23;16,24-binoxol"
        Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Cimigenol.emf")
    End If
End If
Else
    If me2 > 1.14 Then
        If me6 > 1.37 Then
            If me1 > 1.04 Then
                'node #9
                Sheet1.Range("B22") = "23-O-acetylshengmanol"
                Sheet1.Range("E5") = "Actanol"
                Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\23-O-
acetylshengmanol.emf")
            Else
                'node #8
                Sheet1.Range("B22") = "Cimigenol"
                Sheet1.Range("E5") = "Acta-16,23;16,24-
                    binoxol"
                Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Cimigenol.emf")
            End If
        Else
            'node #7
            Sheet1.Range("B22") = "Cimicidanol"
            Sheet1.Range("E5") = "Actanol"
            Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Cimicidanol.emf")
        End If
    Else
        If me8 > 1.99 Then
            If me1 > 0.94 Then
                'node #6
                Sheet1.Range("B22") = "Hydroxyshengmanol"
                Sheet1.Range("E5") = "Acta-16,23-monoxol"
                Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Hydroxyshengmanol.emf")
            Else
                'node #5
                Sheet1.Range("B22") = "Cimigenol"
                Sheet1.Range("E5") = "Acta-16,23;16,24-
                    binoxol"
            End If
        End If
    End If
End If

```

## Appendix C (Continued)

```

        Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Cimigenol.emf")
        End If
    Else
        If me1 > 0.99 Then
            'node #4
            Sheet1.Range("B22") = "Cimicidanol"
            Sheet1.Range("E5") = "Actanol"
            Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Cimicidanol.emf")
        Else
            'node #3
            Sheet1.Range("B22") = "Cimigenol"
            Sheet1.Range("E5") = "Acta-16,23;16,24-
                binoxol"
            Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Cimigenol.emf")
        End If
    End If
End If
End If
Else
    If me5 > 1.12 Then
        'node #2
        Sheet1.Range("B22") = "Dahurinol"
        Sheet1.Range("E5") = "Acta-16,23-monoxol"
        Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Dahurinol.emf")
    Else
        'node #1
        Sheet1.Range("B22") = "Cimigenol"
        Sheet1.Range("E5") = "Acta-16,23;16,24-binoxol"
        Sheet1.Image1.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Cimigenol.emf")
    End If
End If
End If
End If

Sheet1.Image2.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Backbone.emf")

'7,8-ene
x = Sheet1.Range("C16")
y = Sheet1.Range("C17")
If x = "" Or y = "" Or IsNumeric(x) = False Or IsNumeric(y) = False
Then

```

**Appendix C (Continued)**

```

Sheet1.Range("L15") = "Your H-19 data is incomplete or incorrect!"
Sheet1.Range("D22") = "?"
Else
  If WorksheetFunction.Average(x, y) > 0.7 Then
    Sheet1.Range("D22") = "Y"
    Sheet1.Image3.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Bond.emf")
  Else
    Sheet1.Range("D22") = "N"
    Sheet1.Image3.Picture = LoadPicture("")
  End If
End If

'?-O-Ac
If Sheet1.Range("B22") = "Acteol" Then
  If Sheet1.Range("B11") > 1.9 Then
    Sheet1.Range("J22") = "-OAc"
    Sheet1.Image5.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\OAc.emf")
  End If
End If

If Sheet1.Range("B22") = "Cimiracemoside" Then
  If Sheet1.Range("B12") > 1.9 Then
    Sheet1.Range("J22") = "-OAc"
    Sheet1.Image5.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\OAc.emf")
  End If
End If

If Sheet1.Range("B22") = "Hydroxyshengmanol" Then
  Sheet1.Range("L22") = "-OAc"
  If Sheet1.Range("B13") > 3 Then
    Sheet1.Range("M22") = "-OMe"
    Sheet1.Image6.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Me.emf")
  Else
    Sheet1.Image6.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\H.emf")
  End If
End If

If Sheet1.Range("B22") = "Hydroxyshengmanol" Then
  Sheet1.Range("L22") = "-OAc"
End If

If Sheet1.Range("B22") = "23-O-acetylshengmanol" Then
  Sheet1.Range("K22") = "-OAc"

```

## Appendix C (Continued)

```

End If

If Sheet1.Range("B22") = "Cimicidanol" Then
  If Sheet1.Range("B12") > 1.9 Then
    Sheet1.Range("J21") = "15-"
    Sheet1.Range("J22") = "-OAc"
    Sheet1.Image8.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\OAc2.emf")
  Else
    End If
  End If

If Sheet1.Range("B22") = "Dahurinol" Then
  If Sheet1.Range("B12") > 3 Then
    Sheet1.Range("M22") = "-OMe"
    Sheet1.Image7.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Me.emf")
  Else
    Sheet1.Image4.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\H.emf")
  End If
End If

If Sheet1.Range("B22") = "Cimigenol" Then
  If Sheet1.Range("B10") > 1.6 And Sheet1.Range("B11") > 1.6 Then
    If Sheet1.Range("B12") <> "" Then
      If Sheet1.Range("B12") > 1.9 And Sheet1.Range("B12") < 3
Then
        Sheet1.Range("M22") = "-OAc"
        Sheet1.Image4.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Ac.emf")
      Else
        Sheet1.Range("M22") = "-OMe"
        Sheet1.Image4.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\Me.emf")
      End If
    Else
      Sheet1.Range("M22") = "?"
      Sheet1.Image4.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\H.emf")
      Sheet1.Image9.Visible = True
      Sheet1.Range("L15") = Sheet1.Range("L15") + "25-OH may be
halogenated."
    End If
  Else
    If Sheet1.Range("B12") > 1.9 Then
      Sheet1.Range("J22") = "-OAc"
      Sheet1.Image5.Picture =

```

**Appendix C (Continued)**

```
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\OAc.emf")
    Sheet1.Image4.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\H.emf")
    Else
        Sheet1.Image4.Picture =
LoadPicture("C:\Users\fqiu\Desktop\TT_Aglycones\H.emf")
    End If
End If

End If

End If

'Finalization
Sheet1.Range("C5").Select
Sheet1.Protect

End Sub
```

## Appendix D: VBA Code of ActaMatch

```
Sub Match()  
  
    'Initialization  
    'Sheet2.Unprotect  
    Sheet2.Range("E12") = ""  
    Sheet2.Range("AE17") = "Hit # "  
    Sheet2.Range("B21:AS100") = ""  
    Sheet2.Range("A21:A100").Locked = True  
    Sheet2.Range("AU5:AU14") = ""  
    For Each Comment In Sheet2.Comments  
        Comment.Delete  
    Next  
  
    'Input Check  
    s = 0  
    For i = 5 To 13  
        m = i - 4  
        If Sheet2.Range("B" & i) = "" Then  
            a = 0  
            Exit For  
        Else  
            If Sheet2.Range("B" & i) < s Or IsNumeric(Sheet2.Range("B" &  
                i)) = False Then  
                a = 1  
                c = 0  
                Exit For  
            Else  
                s = Sheet2.Range("B" & i)  
                c = 1  
            End If  
        End If  
    Next  
  
    If (m - 1) < 5 Then  
        c = 0  
    End If  
  
    If a = 0 And (m - 1) >= 5 Then  
        For j = (i + 1) To 13  
            If Sheet2.Range("B" & j) <> "" Then  
                c = 0  
                Exit For  
            Else  
                c = 1  
            End If  
        Next  
    End If  
End Sub
```

## Appendix D (Continued)

```

If c = 0 Then
    MsgBox "Error! Your data entry is incomplete or incorrect!",
        vbCritical, "ActaFinder v1.0"
Else

'Database search
r = Sheet2.Range("H8")
n = 5
m = Sheet3.Range("B" & n)
H = 21
Do While m <> ""
    m = Sheet3.Range("B" & n)
    Sheet3.Range("AJ5") = Sheet3.Range("L" & n)
    Sheet3.Range("AJ6") = Sheet3.Range("N" & n)
    Sheet3.Range("AJ7") = Sheet3.Range("P" & n)
    Sheet3.Range("AJ8") = Sheet3.Range("R" & n)
    Sheet3.Range("AJ9") = Sheet3.Range("T" & n)
    Sheet3.Range("AJ10") = Sheet3.Range("V" & n)
    Sheet3.Range("AJ11") = Sheet3.Range("X" & n)
    Sheet3.Range("AJ12") = Sheet3.Range("Z" & n)
    Sheet3.Range("AJ13") = Sheet3.Range("AB" & n)
    Sheet3.Range("AJ14") = Sheet3.Range("AD" & n)
    For i = 5 To 14
        If Sheet3.Range("AJ" & i) = "-" Then
            Sheet3.Range("AJ" & i) = 0
        End If
    Next
    If WorksheetFunction.Pearson(Sheet2.Range("B5:B14"),
        Sheet3.Range("AJ5:AJ14")) >= r Then
        For i = 5 To 14
            If Sheet3.Range("AJ" & i) = 0 Then
                Sheet3.Range("AJ" & i) = "-"
            End If
        Next
        Sheet2.Range("B" & H) = Sheet3.Range("D" & n)
        Sheet2.Range("B" & H).AddComment (Sheet3.Range("C" & n))
        Sheet2.Range("B" & H).Comment.Shape.TextFrame.AutoSize = True
        Sheet2.Range("A" & H).Locked = False
        Sheet2.Range("K" & H) = Sheet3.Range("AJ5")
        Sheet2.Range("O" & H) = Sheet3.Range("AJ6")
        Sheet2.Range("S" & H) = Sheet3.Range("AJ7")
        Sheet2.Range("W" & H) = Sheet3.Range("AJ8")
        Sheet2.Range("AA" & H) = Sheet3.Range("AJ9")
        Sheet2.Range("AE" & H) = Sheet3.Range("AJ10")
        Sheet2.Range("AI" & H) = Sheet3.Range("AJ11")
        Sheet2.Range("AK" & H) = Sheet3.Range("AJ12")
        Sheet2.Range("AM" & H) = Sheet3.Range("AJ13")
        Sheet2.Range("AO" & H) = Sheet3.Range("AJ14")
    End If
End If

```

## Appendix D (Continued)

```

Sheet2.Range("AS" & H) =
WorksheetFunction.Pearson(Sheet2.Range("B5:B14"),
Sheet3.Range("AJ5:AJ14"))

Sheet2.Range("C" & H) = LTrim(Str(Sheet3.Range("J" & n))) +
"MHz"
If Left(Sheet2.ComboBox1.Value, 3) <>
    LTrim(Str(Sheet3.Range("J" & n))) Then
    Sheet2.Range("C" & H).Font.ColorIndex = 3
Else
    Sheet2.Range("C" & H).Font.ColorIndex = 1
End If

Sheet2.Range("D" & H) = Sheet3.Range("I" & n)
If Sheet2.ComboBox2.Value <> Sheet3.Range("I" & n) Then
    Sheet2.Range("D" & H).Font.ColorIndex = 3
Else
    Sheet2.Range("D" & H).Font.ColorIndex = 1
End If

Formula = "C" + LTrim(Sheet2.Range("X4")) + "H" +
LTrim(Sheet2.Range("AA4")) + "O" + LTrim(Sheet2.Range("AD4"))
If Sheet2.Range("AG4") <> 0 Then
    Formula = Formula + "N" + LTrim(Sheet2.Range("AG4"))
End If
Sheet2.Range("E" & H) = Sheet3.Range("F" & n)
If Formula <> Sheet2.Range("E" & H) Then
    Sheet2.Range("E" & H).Font.ColorIndex = 3
Else
    Sheet2.Range("E" & H).Font.ColorIndex = 1
End If

Sheet2.Range("F" & H) = Sheet3.Range("K" & n)

Sheet2.Range("M" & H) = Sheet3.Range("M" & n)
Sheet2.Range("Q" & H) = Sheet3.Range("O" & n)
Sheet2.Range("U" & H) = Sheet3.Range("Q" & n)
Sheet2.Range("Y" & H) = Sheet3.Range("S" & n)
Sheet2.Range("AC" & H) = Sheet3.Range("U" & n)
Sheet2.Range("AG" & H) = Sheet3.Range("W" & n)
Sheet2.Range("AJ" & H) = Sheet3.Range("Y" & n)
Sheet2.Range("AL" & H) = Sheet3.Range("AA" & n)
Sheet2.Range("AN" & H) = Sheet3.Range("AC" & n)
Sheet2.Range("AP" & H) = Sheet3.Range("AE" & n)

Sheet2.Range("AQ" & H) = Sheet3.Range("H" & n)
If Left(Sheet2.Range("AQ" & H), 3) <>
    Left(Sheet2.ComboBox3.Value, 3) Then

```

**Appendix D (Continued)**

```
        Sheet2.Range("AQ" & H).Font.ColorIndex = 3
    Else
        Sheet2.Range("AQ" & H).Font.ColorIndex = 1
    End If

    H = H + 1
End If
n = n + 1
Loop

H = H - 21
If H = 0 Then
    MsgBox "No compound was found!", vbInformation, "ActaFinder v1.0"
    Sheet2.Range("E12") = "No compound was found!"
Else
    If H = 1 Then
        MsgBox "1 compound was found!", vbInformation, "ActaFinder
        v1.0"
        Sheet2.Range("E12") = "1 compound was found!"
    Else
        MsgBox LTrim(Str(H)) + " compounds were found!", vbInformation,
        "ActaFinder v1.0"
        Sheet2.Range("E12") = LTrim(Str(H)) + " compounds were found!"

    End If
End If

End If

End If

'Finalization
Sheet2.Range("B5").Select
Sheet2.Protect

End Sub
```

## Appendix E: Database of Methyl NMR Data of *Actaea* Triterpenes

Literature Name	Previous Type	Sp <sup>a</sup>	Sv <sup>b</sup>	FS <sup>c</sup>	Me1		Me2		Me3		Me4		Me5		Me6		Me7		Me8		
					Assgn		Assgn		Assgn		Assgn		Assgn		Assgn		Assgn		Assgn		Assgn
foetidinol	Foetidonol	<i>h</i>	P	400	0.93	21	1.21	30	1.29	18	1.30	29	1.62	28							
anhydrodahuriny diacetate	Dahurinol	<i>ac</i>	C	60 or 100	0.84		0.86		0.86		0.95	21	1.14		1.28		1.75	26	2.04	2.08	
cimiacerol B	Cimiracemoside	<i>s</i>	P	400	0.89	28	1.08	30	1.22	29	1.23	21	1.24	18	1.68	27	1.76	26			
cimiacerol A	Cimiracemoside	<i>s</i>	P	400	1.09	28	1.11	30	1.23	29	1.24	18	1.25	21	1.68	27	1.77	26			
cimiacerinol	Cimiacerinol	<i>s</i>	P	400	0.89	28	0.98	29	0.99	30	1.07	18	1.25	21	1.67	27	1.76	26			
cimigenol	Cimigenol	<i>r</i>	P	900	0.88	21	1.12	30	1.20	18	1.23	28	1.32	29	1.50	27	1.52	26			
25-O-acetylcimigenol	Cimigenol	<i>h</i>	C	400	0.81	30	0.88	21	0.95	28	1.00	29	1.09	18	1.41	27	1.47	26	1.99	25-OAc	

<sup>a</sup> Species (Sp): *acerina* (*ac*), *asiatica* (*as*), *dahurica* (*d*), *foetida* (*f*), *heracleifolia* (*h*), *podocarpa* (*p*), *racemosa* (*r*), *simplex* (*s*), *vaginata* (*v*)

<sup>b</sup> Solvents (Sv): pyridine-*d*<sub>5</sub> (P), CDCl<sub>3</sub> (C), DMSO-*d*<sub>6</sub> (D)

<sup>c</sup> Field Strength (FS)

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	...	Assgn				
					Assgn													
cimigenyl diacetate	Cimigenol	ac	C	60 or 100										1.95 1.95				
			B	60 or 100	0.86	0.93	0.95	1.15	1.15	1.25	1.25	1.82 1.94						
cimigenyl triacetate	Cimigenol	ac	C	60 or 100	0.85	0.88	1.12	1.18	1.28	1.44	1.96	2.04 2.04						
7 $\beta$ -hydroxycimigenol	Cimigenol	P	r	900	0.89	21	1.13	30	1.22	18	1.25	28	1.32	29	1.51	27	1.53	26
			s	300	0.88	21	1.10	30	1.19	18	1.23	29	1.33	28	1.52	27	1.56	26
25-chlorodeoxycimigenol	Cimigenol	r	P	900	0.87	21	1.08	30	1.14	18	1.19	28	1.33	29	1.70	26	1.71	27
cimigenol-15-O- $\beta$ -D-glucopyranoside	Cimigenol	d	P	400	0.86	21	1.05	30	1.14	18	1.23	26	1.23	29	1.28	28	1.40	27
7,8-didehydrocimigenol	Cimigenol	h	C	400	0.86	30	0.91	21	1.01	29	1.03	18	1.09	28	1.20	27	1.20	26

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	...	Assgn						
					Assgn															
25- <i>O</i> -acetyl-7,8-didehydrocimigenol	Cimigenol	<i>h</i>	C	400	0.85	30	0.90	21	1.01	29	1.03	18	1.06	28	1.42	27	1.48	26	1.99	25-OAc
25- <i>O</i> -methyl-7,8-dehydrocimigenol	Cimigenol	<i>s</i>	P	500	0.90	21	1.11	30	1.17	18	1.19	29	1.27	27	1.28	26	1.44	28	3.22	25-OMe
24- <i>epi</i> -7,8-didehydrocimigenol	Cimigenol	<i>h</i>	C	400	0.85	30	0.90	21	1.01	29	1.03	18	1.06	28	1.22	27	1.33	26		
3-keto-24- <i>epi</i> -7,8-didehydrocimigenol	Cimigenol	<i>h</i>	P	400	0.98	21	1.07	29	1.11	30	1.19	18	1.26	28	1.29	27	1.43	26		
24- <i>epi</i> -acerinol	Acerinol	<i>h</i>	C	400	0.88	18	0.91	21	0.92	30	0.95	28	1.01							
24- <i>O</i> -acetyl-25- <i>O</i> -methyl-7,8-dehydroshengmanol	Dahurinol	<i>s</i>	P	500	1.06	21	1.09	30	1.18	29	1.22	27	1.27	26	1.29	18	1.47	28	2.09	24-OAc
heracleiforinol	Heracleiforinol	<i>h</i>	C	400	0.81	18	0.88	28	0.88	30	1.00	21	1.02	29	1.13	27	1.29	26	2.05	24-OAc
24- <i>O</i> -acetylacerinol	Acerinol	<i>spp.</i>	C		0.87		0.92		0.99		0.99		1.06		1.12		1.23		2.04	24-OAc

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	...	Assgn				
					Assgn													
dahuriny diacetate	Dahurinol	ac	C	60 or 100	0.88	0.88	0.98	1.17	1.17	1.20	1.26	2.04 2.08						
isodahuriny diacetate	Dahurinol	ac	C	60 or 100	0.86	0.86	0.97	1.15	1.15	1.19	1.23	2.05 2.21						
25-O-methylisodahuriny diacetate	Dahurinol	ac	C	60 or 100	0.88	0.88	0.97	21	1.17	1.20	1.23	1.27	2.04 2.13					
cimicidanol	Cimicidanol	h	P	400	1.04	21	1.18	30	1.21	28	1.23	18	1.28	29	1.35	27	1.37	26
cimicifugenol	Cimicifugenol	s	C	500	0.82	29	0.93	30	0.98	28	1.01	21	1.06	18	1.59	27	1.68	26
cimilactone A	Cimilactone	d	P	500	0.83	28	0.95	21	1.00	30	1.23	18	1.31	29	2.11	12-OAc		
3 $\beta$ ,11 $\beta$ -dihydroxy-24,25,26,27-tetranorcychoart-7-en-23,16- $\beta$ -olide-3-O- $\beta$ -D-xylopyranoside	Cimilactone	h	P	500	0.93	21	1.04	28	1.11	30	1.19	18	1.38	29				
cimilactone B	Cimilactone	d	P	500	0.96	21	1.03	28	1.01	30	1.26	18	1.31	29	2.16	12-OAc		

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8						
					Assgn													
cycloartane-24-hydroxy-12 $\beta$ -acetoxy-25,26,27-nor-16,23-dione-3-O- $\alpha$ -L-arabinose	Trinorcimicidol	<i>d</i>	P	400	0.97	28	0.97	30	1.26	21	1.29	29	1.35	18	2.23	12-OAc		
cycloartane-16,24-dihydroxy-12-acetoxy-25,26,27-nor-23-one-3-O- $\alpha$ -L-arabinose	Trinorcimicidol	<i>d</i>	P	400	0.99	21	1.00	30	1.28	29	1.33	18	1.39	28	2.12	12-OAc		
cimicifugoside H-3	Trinorcimicidol		P		1.02	21	1.14	30	1.17	28	1.20	18	1.40	29				
3 $\beta$ ,15 $\alpha$ ,16 $\alpha$ ,24 $\alpha$ -tetrahydroxy-25,26,27-trinor-16,24-cyclocycloart-23-one-3-O- $\beta$ -D-xylopyranoside	Foetidonol	<i>h</i>	P	500	0.90	21	1.06	30	1.15	18	1.31	28	1.32	29				
foetidinol-3-O- $\beta$ -xyloside	Foetidonol	<i>h</i>	P	400	0.92	21	1.16	30	1.26	18	1.41	29	1.60	28				
cimicifugoside H-4	Foetidonol	<i>h</i>	P		0.92	21	1.13	30	1.24	18	1.37	29	1.56	28				
15 $\alpha$ -hydroxyfoetidinol-3-O- $\beta$ -xyloside = cimicifugoside H6	Foetidonol	<i>h</i>	P	400	0.92	21	1.16	30	1.30	18	1.41	29	1.51	28				
cimicifugoside H-6	Foetidonol	<i>h</i>	P		0.92	21	1.16	30	1.30	18	1.42	29	1.52	28				

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8						
					Assgn													
3 $\beta$ ,15 $\alpha$ ,16 $\alpha$ ,24 $\alpha$ -tetrahydroxy-25,26,17-trinor-16,24-cyclocycloart-7-en-23-one-3-O- $\beta$ -D-xylopyranoside	Foetidonol	<i>h</i>	P	500	0.93	21	1.06	30	1.15	18	1.32	29	1.53	28				
12 $\beta$ -acetoxy-3 $\beta$ ,15 $\alpha$ ,16 $\alpha$ ,24 $\alpha$ -tetrahydroxy-25,26,27-trinor-16,24-cyclocycloart-7-en-23-one-3-O- $\beta$ -D-xylopyranoside	Foetidonol	<i>h</i>	P	500	0.99	21	1.05	30	1.34	29	1.40	18	1.55	28	2.16	12-OAc		
cimiracemoside N	Acteol	<i>r</i>	P	300	0.85	28	0.96	30	1.02	21	1.27	29	1.42	18	1.48	27	2.14	12-OAc
acetylacteol-3-O-arabinoside	Acteol	<i>h</i>	D	400	0.77	30	0.84	28	0.87	21	0.96	29	1.13	18	1.43	26	1.96	12-OAc
cimiracemoside P	Acteol	<i>r</i>	P	300	0.85	28	0.93	21	1.02	30	1.33	18	1.33	29	1.65	27	2.16	12-OAc
(26S)-bugbanoside B	Acteol	<i>s</i>	P	400	0.91	28	0.92	21	1.08	30	1.25	18	1.39	29	1.75	27		
(26R)-bugbanoside B	Acteol	<i>s</i>	P	400	0.90	21	0.96	28	1.10	30	1.25	18	1.41	29	1.59	27		
acteol	Acteol	<i>n/a</i>	P	270 or 400	0.79	28	0.96	30	0.97	21	1.27	29	1.34	18	1.75	26	2.11	12-OAc

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8								
					Assgn															
26-deoxyactein	Acteol	<i>n/a</i>	P	270 or 400	0.84	28	0.98	30	1.02	21	1.29	29	1.41	26	1.46	18	2.12	12-OAc		
2'-O-acetylactein	Acteol	<i>f</i>	P	400	0.88	28	1.00	30	1.05	21	1.20	29	1.35	18	1.88	26	2.26	12-OAc	2.28	2'-OAc
2'-O-acetyl-26-deoxyactein	Acteol	<i>f</i>	P	400	0.93	28	0.99	30	1.12	21	1.16	29	1.50	26	1.55	18	2.22	12-OAc	2.25	2'-OAc
cimifoside D	Acteol	<i>f</i>	P	500	0.82	28	0.96	30	1.00	21	1.27	29	1.39	18	1.54	27	2.11	12-OAc		
cimifoside E	Acteol	<i>f</i>	P	400 or 500	0.75	28	0.91	30	0.94	21	1.08	29	1.12	18	1.74	27	2.12	12-OAc		
cimiracemoside O	Acteol	<i>r</i>	P	300	0.80	28	0.98	21	0.98	30	1.37	18	1.79	27	1.79	29	1.99	12-OAc	2.16	4'-OAc
23- <i>epi</i> -26-deoxyactein	Acteol	<i>r</i>	P	500	0.83	28	0.99	30	1.00	21	1.29	29	1.40	18	1.45	27				
12-deacetoxy-15 $\alpha$ -hydroxy-23- <i>epi</i> -26-deoxyactein	Acteol	<i>v</i>	P	500	0.95	21	1.02	30	1.24	28	1.29	18	1.30	29	1.41	27				

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	Assgn							
					Assgn															
12-deacetoxy-23- <i>epi</i> -26-deoxyactein	Acteol	<i>v</i>	P	500	0.83	28	0.95	21	1.01	30	1.22	18	1.31	29	1.43	27				
cimiracemoside I	Acteol	<i>r</i>	P	900	1.01	21	1.04	30	1.07	28	1.26	18	1.35	29	1.47	27				
(26 <i>S</i> )-bugbanoside A	Acteol	<i>s</i>	P	400	0.95	21	1.05	28	1.11	30	1.23	18	1.41	29	1.75	27				
(26 <i>R</i> )-bugbanoside A	Acteol	<i>s</i>	P	400	0.92	21	1.10	28	1.12	30	1.24	18	1.42	29	1.60	27				
cimifugoside (26 <i>S</i> )	Cimicifugoside	<i>s</i>	P	500	0.98	21	1.01	28	1.02	30	1.31	29	1.41	18	1.78	27	2.18	12-OAc		
26-deoxycimicifugoside	Acteol	<i>s</i>	P	400	0.99	30	1.02	21	1.05	28	1.30	29	1.47	27	1.49	18	2.17	12-OAc		
2'- <i>O</i> -malonylcimicifugoside	Acteol	<i>s</i>	P	500	0.96	21	0.99	30	1.00	28	1.16	29	1.38	18	1.79	27	2.21	12-OAc		
3'- <i>O</i> -acetylcimicifugoside	Cimicifugoside	<i>as</i>	P	300	0.98	21	0.98	30	1.02	28	1.31	29	1.42	18	1.81	27	2.00	3'-OAc	2.18	12-OAc

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8								
					Assgn															
4'-O-acetyl-23- <i>epi</i> -26-deoxycimifugoside	Cimicifugoside	<i>as</i>	P	300	0.98	30	1.03	21	1.06	28	1.31	29	1.48	27	1.50	18	1.99	4'-OAc	2.20	12-OAc
cimiracemoside A	Cimigenol	<i>r</i>	P	500	1.04	30	1.24	28	1.30	18	1.30	29	1.50	26	1.50	27				
12,β,21-dihydroxycimigenol-3-O-α-L-arabinopyranoside	Cimigenol	<i>r</i>	P	500	1.00	30	1.23	28	1.28	29	1.49	26	1.49	27	1.52	18				
cimiracemoside B	Cimigenol	<i>r</i>	P	500	1.04	30	1.20	28	1.24	18	1.28	29	1.46	26	1.48	27				
cimiracemoside J	25-Dehydrocimigenol	<i>r</i>	P	300	0.94	21	1.01	30	1.20	28	1.28	29	1.37	18	1.85	27	2.13	12-OAc		
cimiracemoside K	25-Dehydrocimigenol	<i>r</i>	P	300	0.95	21	1.04	30	1.20	28	1.31	29	1.32	18	1.85	27	2.12	12-OAc		
25-anhydrocimigenol-3-O-xyloside	25-Dehydrocimigenol	<i>h</i>	P	400	0.86	21	1.08	30	1.16	18	1.18	28	1.33	29	1.84	27				
25,26-anhydrocimigenol-3-O-β-D-(2'-O-acetyl)xylopyranoside	Cimigenol	<i>v</i>	P	500	0.84	21	0.96	30	1.09	18	1.16	28	1.27	29	1.83	27				

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	...	Assgn						
					Assgn															
2'-O-malonylcimiaceroside B	Cimiracemoside	s	P	500	0.85	28	1.00	30	1.18	29	1.20	18	1.22	21	1.67	27	1.76	26		
2'-O-β-D-glucopyranosyl-(1→2)-β-D-glucopyranosyl-cimiaceroside B	Cimiracemoside	n/a	P	500	0.85	28	1.11	30	1.20	18	1.22	21	1.26	29	1.70	27	1.76	26		
2'-O-(6'''-O-trans-isoferuloyl)-β-D-glucopyranosyl-(1→2)-β-D-glucopyranosylcimiaceroside B	Cimiracemoside	n/a	P	500	0.85	28	1.15	30	1.21	21	1.23	18	1.31	29	1.68	27	1.77	26		
cimiracemoside H	Cimiracemoside	r	P	500	0.87	28	1.02	30	1.35	29	1.38	18	1.38	21	1.72	27	1.79	26	2.11	12-OAc
cimiaceroside B	Cimiracemoside	s	P	400	0.87	28	1.04	30	1.21	18	1.22	21	1.33	29	1.67	27	1.75	26		
cimiaceroside C	Cimiracemoside	f	P	500	0.83	28	1.02	30	1.19	18	1.21	21	1.31	29	1.67	27	1.76	26		
cimiaceroside E	Cimiracemoside	f	P	400 or 500	0.83	28	0.97	30	1.14	18	1.15	21	1.30	29	1.51	26	1.61	27		
(20S,22R,23S,24R)-16β:23;22:25-diepoxy-3β,23,24-trihydroxy-9,19-cycloartane-3-O-β-D-(4'-O-acetyl)xylopyranoside	Cimiracemoside	v	P	500	0.84	28	1.01	30	1.20	18	1.22	21	1.32	29	1.68	27	1.76	26	1.95	4'-OAc

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	12-OAc							
					Assgn															
cimiracemoside F	Cimiracemoside	<i>r</i>	P	500	1.02	30	1.08	28	1.33	21	1.34	29	1.41	18	1.70	27	1.78	26	2.13	12-OAc
				500	1.02	30	1.06	28	1.31	21	1.33	29	1.40	18	1.69	27	1.77	26	2.10	12-OAc
cimiracemoside G	Cimiracemoside	<i>r</i>	P	500	0.97	30	1.05	28	1.28	29	1.31	21	1.39	18	1.68	27	1.76	26	2.11	12-OAc
cimiaceroside A	Cimiracemoside	<i>s</i>	P	400	1.05	30	1.08	28	1.22	18	1.25	21	1.35	29	1.67	27	1.76	26		
		<i>r</i>		900	1.06	30	1.08	28	1.22	18	1.26	21	1.37	29	1.70	27	1.79	26		
1 $\alpha$ -hydroxycimigenol-3-O- $\alpha$ -L-arabinopyranoside	Cimigenol	<i>r</i>	P	900	0.85	21	1.10	30	1.21	18	1.28	28	1.38	29	1.48	26	1.49	27		
		<i>s</i>		400	0.86	21	1.08	30	1.20	18	1.28	28	1.34	28	1.45	26	1.48	27		
cimiracemoside C	Cimigenol	<i>r</i>	P	500	0.86	21	1.03	30	1.15	18	1.19	18	1.28	29	1.47	27	1.49	26		

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	...	Assgn						
					Assgn															
cimiracemoside D	Cimigenol	<i>r</i>	P	500	0.92	21	0.98	30	1.21	29	1.25	28	1.31	18	1.47	27	1.49	26	2.10	12'-OAc
25-O-acetyl-12 $\beta$ -hydroxycimigenol-3-O- $\alpha$ -L-arabinoside	Cimigenol	<i>r</i>	P	500	1.00	30	1.21	28	1.27	29	1.39	21	1.42	18	1.69	26	1.71	27	1.98	25'-OAc
12 $\beta$ -hydroxycimigenol-3-O- $\beta$ -D-arabinopyranoside	Cimigenol	<i>s</i>	P	300	0.99	30	1.23	29	1.26	28	1.37	21	1.43	18	1.51	27	1.55	26		
25-O-acetylcimigenol-3-O- $\alpha$ -L-arabinopyranoside	Cimigenol	<i>d</i>	P	400 or 500	0.89	21	1.00	30	1.12	18	1.16	28	1.26	29	1.64	27	1.66	26	1.94	25'-OAc
cimicifoetoside B	Cimigenol	<i>f</i>	P	500	0.84	21	0.95	30	1.07	29	1.13	18	1.17	28	1.65	27	1.67	26	1.95 2.09	25-, 2'-OAc
cimicifoetoside A	Cimigenol	<i>f</i>	P	500	0.85	21	0.93	30	1.05	29	1.12	18	1.18	28	1.44	26	1.47	27	2.11	2'-OAc
cimigenol 3-O- $\beta$ -D-galactopyranoside	Cimigenol	<i>s</i>	P	400	0.87	21	1.04	30	1.16	18	1.20	28	1.32	29	1.46	27	1.48	26		
25-O-methylcimigenol-3-O- $\beta$ -D-galactopyranoside	Cimigenol	<i>s</i>	P	400	0.87	21	1.03	30	1.13	18	1.20	28	1.27	29	1.29	27	1.39	26		

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	...	Assgn						
					Assgn															
25-O-acetylcimigenol-3-O-β-D-galactopyranoside	Cimigenol	s	P	400	0.87	21	1.04	30	1.15	18	1.21	28	1.30	29	1.67	27	1.70	26	2.02	25-OAc
25-O-acetylcimigenol-3-O-β-D-glucopyranoside	Cimigenol	s	P	400	0.87	21	1.07	30	1.15	18	1.21	28	1.33	29	1.67	27	1.70	26	2.00	25-OAc
1α-hydroxycimigenol-3-O-β-D-galactopyranoside	Cimigenol	s	P	500	0.85	21	1.08	30	1.19	18	1.29	28	1.38	29	1.47	26	1.50	27		
(23R,24S)-16β,23;16α,24-diepoxyoctartane-3β,12β,25-triol-3-O-β-D-xylopyranoside	Foetidonol	n/a	P	500	1.05	30	1.24	28	1.36	29	1.39	21	1.41	18	1.47	27	1.54	26		
1α-hydroxycimigenol-3-O-β-D-xylopyranoside	Cimigenol	r	P	900	0.85	21	1.14	30	1.21	18	1.27	28	1.43	29	1.48	26	1.50	27		
		r		300	0.84	21	1.11	30	1.20	18	1.30	28	1.39	29	1.47	26	1.50	27		
25-O-acetylcimigenol-3-O-xylopyranoside	Cimigenol	r	P	900	0.85	21	1.07	30	1.15	18	1.20	28	1.33	29	1.67	27	1.69	26	1.97	25-OAc
25-O-acetyl-7-β-hydroxycimigenol-3-O-β-D-xylopyranoside	Cimigenol	r	P	900	0.87	21	1.09	30	1.19	18	1.30	28	1.36	29	1.73	27	1.74	26	1.96	25-OAc

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	...	Assgn						
					Assgn															
25-O-acetyl-7- $\beta$ -hydroxycimigenol-3-O- $\beta$ -D-xylopyranoside	Cimigenol	s	P	300	0.88	21	1.08	30	1.19	18	1.30	29	1.35	28	1.73	27	1.75	26	1.98	25-OAc
cimigenol-3-O- $\beta$ -D-xylopyranoside	Cimigenol	r	P	600	0.87	21	1.07	30	1.16	18	1.23	28	1.32	29	1.50	27	1.55	26		
(23 <i>R</i> ,24 <i>S</i> )-16 $\beta$ ,23;16 $\alpha$ ,24-diepoxy-12 $\beta$ -acetoxy-cycloart-3 $\beta$ ,15 $\alpha$ ,25-triol-3-O- $\beta$ -D-xylopyranoside	Cimigenol	n/a	P	500	0.96	21	1.08	28	1.08	30	1.19	18	1.27	27	1.34	29	1.44	26		
cimigenol xylose	Cimigenol	r	P	900	0.86	21	1.07	30	1.16	18	1.20	28	1.33	29	1.48	27	1.51	26		
cimigenol-3-O- $\beta$ -D-xyloside	Cimigenol	r	P	900	0.87	21	1.07	30	1.16	18	1.20	28	1.33	29	1.48	26	1.50	27		
12 $\beta$ -hydroxycimigenol-3-O- $\beta$ -D-xylopyranoside	Cimigenol	s	P	300	1.01	30	1.21	29	1.28	28	1.35	21	1.41	18	1.49	27	1.52	26		
7 $\beta$ -hydroxycimigenol-3-O- $\beta$ -D-xylopyranoside	Cimigenol	s	P	300	0.87	21	1.07	30	1.18	18	1.30	29	1.34	28	1.49	27	1.52	26		
25-O-acetyl-1 $\alpha$ -hydroxycimigenol-3-O- $\beta$ -D-xylopyranoside	Cimigenol	s	P	400	0.85	21	1.11	30	1.19	18	1.29	28	1.38	29	1.66	26	1.68	27	1.99	25-OAc

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	...	Assgn						
					Assgn															
25- <i>O</i> -acetylcimigenol-3- <i>O</i> - $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 3)- $\beta$ -D-xylopyranoside	Cimigenol	s	P		0.81	21	1.01	30	1.10	18	1.15	28	1.26	29	1.67	26	1.71	27	1.95	25-OAc
3'- <i>O</i> -acetyl-cimigenol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Cimigenol	ac	P	400	0.83	21	1.00	30	1.13	18	1.17	28	1.24	29	1.46	27	1.49	26	1.98	3'-OAc
(3',12 $\beta$ )- <i>O</i> -diacetylcimigenol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Cimigenol	as	P	300	0.95	21	0.99	30	1.22	28	1.26	18	1.33	29	1.50	27	1.51	26	1.98 2.13	3', 12-OAc
(4',25)- <i>O</i> -diacetylcimigenol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Cimigenol	as	P	300	0.87	21	1.06	30	1.16	18	1.21	28	1.33	29	1.70	26	1.73	27	1.98 1.98	25-, 4'-OAc
2'- <i>O</i> -acetyl-25- <i>O</i> -methylcimigenol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Cimigenol	as	C	300	0.80	30	0.88	21	0.94	28	0.94	29	1.07	18	1.07	26	1.16	27	2.13	2'-, 25-OAc
cimifoside A	Cimigenol	f	P	500	1.02	30	1.21	28	1.29	29	1.38	21	1.41	18	1.49	26	1.50	27		
cimifoside B	Cimigenol	f	P	500	0.99	21	1.21	30	1.22	18	1.23	29	1.30	28	1.44	26	1.47	27	2.01	25-OAc
2'- <i>O</i> -acetyl-25- <i>O</i> -ethylcimigenol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Cimigenol	as	C	300	0.80	30	0.88	21	0.94	28	0.94	29	1.07	18	1.07	26	1.16	27	1.13 2.14	25-OEt, 2'-OAc

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	...	Assgn						
					Assgn															
25- <i>O</i> -acetylcimigenol-3- <i>O</i> - $\beta$ -D-(2'- <i>O</i> -acetyl)xylopyranoside	Cimigenol	v	P	500	0.84	21	0.97	30	1.08	18	1.13	28	1.20	29	1.65	26	1.67	27	1.95 2.14	25-, 2'- OAc
12 $\beta$ -hydroxycimigenol-3- <i>O</i> - $\beta$ -D-xylopyranoside-(1 $\rightarrow$ 3)- $\beta$ -D-xylopyranoside	Cimigenol	v	P	500	1.02	30	1.21	28	1.29	29	1.39	21	1.42	18	1.48	27	1.49	26		
7,8-didehydrocimigenol-3- <i>O</i> - $\alpha$ -L-arabinopyranoside	Cimigenol	s	P	400	0.91	21	1.03	30	1.17	18	1.27	29	1.42	28	1.47	27	1.49	26		
25- <i>O</i> -acetyl-7,8-didehydrocimigenol-3- <i>O</i> - $\alpha$ -L-arabinopyranoside	Cimigenol	s	P	400	0.90	21	1.03	30	1.16	18	1.28	29	1.44	28	1.68	27	1.69	26	1.99	25- OAc
bugbanoside F	Cimigenol	s	P	500	0.98	30	1.22	29	1.37	21	1.45	28	1.47	18	1.51	26	1.55	27		
7,8-didehydrocimigenol-3- <i>O</i> - $\beta$ -D-galactopyranoside	Hydroxyshengmanol	s	P	400	0.90	21	1.03	30	1.16	18	1.30	29	1.42	28	1.46	27	1.49	26		
(23 <i>R</i> ,24 <i>S</i> )-16 $\beta$ ,23;16 $\alpha$ ,24-diepoxyart-7-en-3 $\beta$ ,11 $\beta$ ,25-triol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Cimigenol	<i>n/a</i>	P	500	0.84	21	1.16	30	1.27	18	1.40	29	1.43	28	1.46	27	1.52	26		
(23 <i>R</i> ,24 <i>S</i> )-16 $\beta$ ,23;16 $\alpha$ ,24-diepoxyart-7-en-3 $\beta$ ,12 $\beta$ ,15 $\alpha$ ,25-triol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Cimigenol	<i>n/a</i>	P	500	1.04	30	1.31	28	1.33	29	1.34	27	1.45	26	1.51	18	1.52	21		

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	...	Assgn						
					Assgn															
(23 <i>R</i> ,24 <i>S</i> )-16 $\beta$ ,23;16 $\alpha$ ,24-diepoxy-12 $\beta$ -acetoxy-cycloart-7-en-3 $\beta$ ,15 $\alpha$ ,25-triol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Cimigenol	<i>n/a</i>	P	500	1.04	30	1.07	21	1.22	28	1.28	29	1.32	27	1.40	18	1.44	26	2.16	12-OAc
25- <i>O</i> -acetyl-7,8-didehydrocimigenol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Cimigenol	<i>s</i>	P	400	0.89	21	1.06	30	1.15	18	1.30	29	1.42	28	1.66	27	1.67	26	1.96	25-OAc
cimicinol	Cimicinol	<i>h</i>	P	400	0.77	18	0.83	21	0.94	30	1.29	28	1.44	27	1.44	29	1.51	26		
1 $\alpha$ -hydroxy-24- <i>epi</i> -cimigenol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Cimigenol	<i>r</i>	P	900	0.95	21	1.15	30	1.19	28	1.25	18	1.26	26	1.41	27	1.43	29		
24- <i>epi</i> -7,8-didehydrocimigenol-3-xyloside	Cimigenol	<i>h</i>	P	400	0.98	21	1.05	30	1.16	18	1.27	28	1.28	29	1.31	27	1.41	26		
2',4'-di- <i>O</i> -acetyl-24- <i>epi</i> -7,8-didehydrocimigenol-3- <i>O</i> -xyloside	Cimigenol	<i>h</i>	C	400	0.85	30	0.89	21	0.97	29	1.02	18	1.05	28	1.22	27	1.33	26	2.11 2.13	
3'- <i>O</i> -acetyl-24- <i>epi</i> -7,8-didehydrocimigenol-3-xyloside	Cimigenol	<i>h</i>	P	400	0.98	21	1.00	30	1.18	18	1.26	28	1.28	27	1.28	29	1.42	26	1.98	3'-OAc
3- <i>O</i> -arabinosyl-24- <i>O</i> -acetylhydroshengmanol-15-glucoside	Hydroxyshengmanol	<i>d</i>	P	270 or 400	0.95	21	1.06	30	1.24	18	1.26	26	1.28	28	1.38	27	1.54	29	2.29	24-OAc

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	...	Assgn						
					Assgn															
24- <i>epi</i> -24- <i>O</i> -acetylhydroshengmanol-3- <i>O</i> - $\beta$ -D-galactopyranoside	Cimigenol	s	P	400	0.98	21	1.02	30	1.22	29	1.24	28	1.32	18	1.45	27	1.46	26	2.12	24-OAc
24- <i>O</i> -acetylhydroshengmanol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Cimigenol	r	P	900	0.97	21	1.05	30	1.23	28	1.25	18	1.33	29	1.49	27	1.52	26	2.13	24-OAc
3- <i>O</i> -xylosyl-24- <i>O</i> -acetylhydroshengmanol-15-glucoside	Hydroxyshengmanol	d	P	270 or 400	0.95	21	1.09	30	1.23	18	1.28	26	1.29	28	1.38	27	1.54	29	2.29	24-OAc
24- <i>epi</i> -7 $\beta$ -hydroxy-24- <i>O</i> -acetylhydroshengmanol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Hydroxyshengmanol	s	P	300	0.98	21	1.05	30	1.21	28	1.27	18	1.32	29	1.53	27	1.57	26	2.19	24-OAc
25- <i>O</i> -methyl-24- <i>O</i> -acetylhydroshengmanol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Hydroxyshengmanol	s	P	300	1.02	21	1.04	30	1.23	28	1.25	26	1.25	27	1.28	29	1.30	18	2.12	24-OAc
25- <i>O</i> -methyl-7 $\beta$ -hydroxy-24- <i>O</i> -acetylhydroshengmanol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Hydroxyshengmanol	s	P	300	1.01	21	1.04	30	1.22	28	1.26	26	1.28	27	1.29	18	1.31	29	2.11	24-OAc
25- <i>O</i> -methyl-1 $\alpha$ -hydroxy-24- <i>O</i> -acetylhydroshengmanol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Hydroxyshengmanol	s	P	300	0.98	21	1.07	30	1.22	28	1.28	27	1.29	18	1.35	29	1.37	26	2.12	24-OAc
cimifoside C	Hydroxyshengmanol	f	P	500	0.82	21	1.02	30	1.10	29	1.18	28	1.28	18	1.63	26	1.65	27	2.11	24-OAc

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	...	Assgn						
					Assgn															
24- <i>epi</i> -24- <i>O</i> -acetyl-7,8-didehydrohydroshengmanol-3- <i>O</i> - $\alpha$ -L-arabinopyranoside	Hydroxyshengmanol	s	P	400	1.01	21	1.02	30	1.25	18	1.29	29	1.43	28	1.45	27	1.49	26	2.12	24-OAc
24- <i>O</i> -acetyl-7,8-didehydrohydroshengmanol-3- <i>O</i> - $\alpha$ -L-arabinopyranoside	Hydroxyshengmanol	s	P	400	1.04	21	1.09	30	1.28	18	1.29	29	1.47	27	1.47	28	1.48	26	2.00	24-OAc
24- <i>epi</i> -24- <i>O</i> -acetyl-7,8-didehydrohydroshengmanol-3- <i>O</i> - $\beta$ -D-galactopyranoside	Hydroxyshengmanol	s	P	400	1.01	21	1.03	30	1.25	29	1.33	18	1.44	28	1.45	27	1.49	26	2.12	24-OAc
24- <i>O</i> -acetyl-7,8-didehydrohydroshengmanol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Hydroxyshengmanol	s	P	400	1.03	21	1.04	30	1.29	18	1.31	29	1.46	28	1.47	26	1.47	27	2.00	24-OAc
		<i>h</i>		400	0.96	21	1.04	30	1.18	18	1.29	29	1.34	28	1.49	27	1.69	26	2.02	24-OAc
24- <i>epi</i> -24- <i>O</i> -acetyl-7,8-didehydrohydroshengmanol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Hydroxyshengmanol	s	P	400	1.03	21	1.06	30	1.27	18	1.34	29	1.44	28	1.47	27	1.50	26	2.14	24-OAc
24- <i>O</i> -acetyl-25- <i>O</i> -methyl-7,8-didehydroshengmanol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Dahurinol	s	P	500	1.04	30	1.05	21	1.21	27	1.26	18	1.27	26	1.31	29	1.46	28	2.02	24-OAc
24- <i>epi</i> -24- <i>O</i> -acetyl-7,8-didehydroshengmanol-3- <i>O</i> -(2'- <i>O</i> -malonyl)- $\beta$ -D-xylopyranoside	Dahurinol	s	P	500	1.01	21	1.02	30	1.15	29	1.24	18	1.42	28	1.48	26	1.52	27	2.15	24-OAc

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	...	Assgn						
					Assgn															
1 $\alpha$ -hydroxydahurinol-3-O- $\alpha$ -L-arabinopyranoside	Cimigenol	<i>r</i>	P	900	0.96	21	1.07	29	1.17	28	1.25	18	1.38	30	1.59	27	1.65	26		
actaeaepoxide-3-O- $\beta$ -D-xylopyranoside		<i>r</i>	P	600	1.03		1.07		1.31	21	1.33		1.41		1.69		1.77		2.15	12-OAc
cimiracemoside E	Dahurinol	<i>r</i>	P	500	0.90	21	1.00	28	1.02	30	1.15	18	1.30	29	1.61	26	1.61	27	2.15	24-OAc
24-O-acetylisdahuriny-3-O- $\beta$ -D-(2'-O-acetyl)xylopyranoside	Dahurinol	<i>v</i>	P	500	0.90	21	0.92	28	1.00	30	1.07	18	1.15	29	1.59	26	1.59	27	2.13 2.14	24-, 2'-OAc
24-O-acetyldahuriny-3-O- $\beta$ -D-(2'-O-acetyl)xylopyranoside	Dahurinol	<i>v</i>	P	500	0.92	21	1.00	30	1.07	28	1.17	18	1.30	29	1.48	26	1.48	27	2.09	24-OAc
cimicifugadine	Alkaloid	<i>f</i>	P	600	1.06	28	1.16	30	1.35	18	1.42	29	1.51	26	1.64	27	2.20	21		
23-O-acetylshengmanol-3-O- $\alpha$ -L-arabinopyranoside	23-O-Acetylshengmanol	<i>r</i>	P	500	1.05	30	1.21	28	1.25	26	1.26	21	1.30	29	1.37	18	1.40	27	2.06	23-OAc
shengmanol-3-O- $\alpha$ -L-arabinopyranoside	Shengmanol	<i>r</i>	P	400 or 500	1.02	30	1.05	21	1.25	18	1.26	28	1.27	26	1.28	29	1.29	27		

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	...	Assgn						
					Assgn															
cimiracemoside L	23- <i>O</i> -Acetylshengmanol	<i>r</i>	P	300	1.06	30	1.22	28	1.26	26	1.27	21	1.31	29	1.38	18	1.40	27	2.06 2.12	23-, 4'- OAc
23- <i>O</i> -acetyl-1 $\alpha$ -hydroxyshengmanol-3- <i>O</i> - $\beta$ -D-xylopyranoside	Hydroxyshengmanol	<i>s</i>	P	500	1.11	30	1.22	21	1.24	26	1.27	28	1.37	27	1.39	29	1.40	18	2.06	23- OAc
7 $\beta$ -hydroxy-23- <i>O</i> -acetylshengmanol-3- <i>O</i> - $\beta$ -D-xylopyranoside	23- <i>O</i> -Acetylshengmanol	<i>s</i>	P		1.07	30	1.22	28	1.23	26	1.28	21	1.35	29	1.38	27	1.40	18	2.09	23- OAc
cimiracemoside M	23- <i>O</i> -Acetylshengmanol	<i>r</i>	P	300	1.07	30	1.22	28	1.26	26	1.27	21	1.34	29	1.38	18	1.41	27	1.99 2.07	23-, 4'- OAc
23- <i>O</i> -acetylshengmanol-3- <i>O</i> - $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 3)- $\beta$ -D-xylopyranoside	23- <i>O</i> -Acetylshengmanol	<i>s</i>	P		1.05	30	1.21	28	1.26	21	1.29	26	1.33	29	1.37	18	1.44	27	2.12	23- OAc
23- <i>O</i> -acetylshengmanol-3- <i>O</i> -(2'- <i>O</i> -malonyl)- $\beta$ -D-xylopyranoside	23- <i>O</i> -Acetylshengmanol	<i>s</i>	P	500	1.03	30	1.17	29	1.20	28	1.26	21	1.27	26	1.36	18	1.41	27	2.07	23- OAc
bugbanoside D	Cimicidanol	<i>s</i>	P	500	1.02	30	1.27	21	1.30	27	1.31	29	1.35	26	1.37	28	1.55	18	2.27	12- OAc
23- <i>O</i> -acetyl-7,8-didehydroshengmanol-3- <i>O</i> - $\alpha$ -L-arabinopyranoside	Hydroxyshengmanol	<i>s</i>	P	400	1.05	30	1.23	21	1.28	29	1.30	27	1.32	18	1.42	26	1.43	28	2.01	23- OAc

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	...	Assgn						
					Assgn															
cimicidanol-3-O-arabinoside	Cimicidanol	<i>h</i>	P	400	1.00	21	1.07	30	1.17	18	1.17	28	1.31	27	1.33	29	1.37	26		
bugbanoside C	Cimicidol	<i>s</i>	P	500	1.02	30	1.30	29	1.32	28	1.37	21	1.52	26	1.54	27	1.60	18	2.27	12-OAc
bugbanoside E	Cimicidanol	<i>s</i>	P	500	1.01	30	1.21	28	1.24	21	1.30	27	1.30	29	1.36	26	1.38	18	2.28	12-OAc
23-O-acetyl-7,8-didehydroshengmanol-3-O-β-D-galactopyranoside	23-O-Acetylshengmanol	<i>s</i>	P	500	1.05	30	1.24	21	1.29	18	1.29	26	1.35	29	1.43	27	1.45	28	2.05	23-OAc
23-acetoxy-3,15,24,25-tetrahydroxycycloart-7-en-16-one-3-O-β-D-xylopyranoside	23-O-Acetylshengmanol	<i>h</i>	P	500	1.03	30	1.27	18	1.28	21	1.32	28	1.32	29	1.59	26	1.59	27	2.00	23-OAc
cimicidanol-3-O-β-xyloside	Cimicidanol	<i>h</i>	P	400	1.13	21	1.15	30	1.16	28	1.22	18	1.41	29	1.54	27	1.66	26		
cimicifugoside H1	Cimicidanol	<i>n/a</i>		270 or 400	1.04	21	1.15	30	1.21	18	1.21	28	1.35	26	1.36	27	1.41	29		
cimicifugoside H2	Cimicidol	<i>n/a</i>	P	270 or 400	1.13	21	1.14	30	1.16	28	1.22	18	1.42	29	1.55	26	1.67	27		

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	...	Assgn						
					Assgn															
cimicifugoside H5	Cimicidanol	<i>n/a</i>	P	270 or 400	1.05	21	1.25	30	1.30	18	1.31	26	1.36	27	1.36	29	1.37	28		
cimicifol	Cimicidanol	<i>h</i>	P	400	1.03	30	1.20	28	1.22	21	1.29	26	1.33	27	1.33	29	1.36	18	2.26	12-OAc
15 $\alpha$ -hydroxycimicidol-3- <i>O</i> - $\beta$ -xyloside	Cimicidol	<i>h</i>	P	400	1.13	21	1.15	30	1.28	28	1.38	29	1.40	18	1.50	26	1.63	26		
23- <i>O</i> -acetyl-7,8-dehydroshengmanol-3- <i>O</i> - $\beta$ -D-xylopyranoside	23- <i>O</i> -Acetylshengmanol	<i>s</i>	P	500	1.09	30	1.24	21	1.30	18	1.30	26	1.36	29	1.43	27	1.45	28	2.06	23-OAc
23,24-diacetoxy-3,15,25-trihydroxycycloart-7-en-16-one-3- <i>O</i> - $\beta$ -D-xylopyranoside	23- <i>O</i> -Acetylshengmanol	<i>h</i>	P	500	1.05	30	1.24	18	1.25	21	1.32	28	1.32	29	1.51	27	1.52	26	2.09 2.16	
23- <i>O</i> -acetyl-7,8-dehydroshengmanol-3- <i>O</i> -(2'- <i>O</i> -malonyl)- $\beta$ -D-xylopyranoside	23- <i>O</i> -Acetylshengmanol	<i>s</i>	P	500	1.06	30	1.19	29	1.24	21	1.28	26	1.29	18	1.42	27	1.43	28	2.05	23-OAc
(24 <i>S</i> )-15,23,24-triacetoxy-25-hydroxycycloart-7-en-16-one-3- <i>O</i> - $\beta$ -D-(2',3',4'-tri- <i>O</i> -acetyl)xylopyranoside	23- <i>O</i> -Acetylshengmanol	<i>h</i>	P	500	0.94	30	1.05	29	1.23	21	1.27	18	1.27	28	1.51	27	1.53	26	1.99 2.06 2.09 2.14 2.15 2.23	

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	...	Assgn					
					Assgn														
(24 <i>R</i> )-15,23,24-triacetoxy-25-hydroxycycloart-7-en-16-one-3- <i>O</i> - $\beta$ -D-(2',3',4'-tri- <i>O</i> -acetyl)xylopyranoside	23- <i>O</i> -Acetylshengmanol	<i>h</i>	<i>P</i>	500	0.94	30	1.05	29	1.22	21	1.29	28	1.33	18	1.46	27	1.59	26	2.00 2.07 2.08 2.08 2.15 2.26
podocarpaside G	Podocarposiide	<i>p</i>	<i>C</i>	400	1.00	28	1.03	21	1.04	18	1.08	30	1.44	29	1.45	26	1.45	27	
podocarpaside A	Podocarposiide	<i>p</i>	<i>C</i>	400	0.82	18	0.87	28	1.03	21	1.17	30	1.37	29	1.49	26	1.50	27	
podocarpaside B	Podocarposiide	<i>p</i>	<i>C</i>	400	0.91	30	0.93	18	0.93	28	1.04	21	1.38	29	1.47	26	1.49	27	
podocarpaside C	Podocarposiide	<i>p</i>	<i>C</i>	400	0.93	30	1.06	28	1.08	21	1.09	18	1.35	29	1.45	26	1.46	27	
podocarpaside D	Podocarposiide	<i>p</i>	<i>C</i>	400	0.86	28	0.90	18	1.01	21	1.37	30	1.39	29	1.46	26	1.47	27	
podocarpaside E	Podocarposiide	<i>p</i>	<i>C</i>	400	0.99	18	1.04	21	1.10	28	1.24	30	1.32	29	1.48	26	1.48	27	

## Appendix E (Continued)

Literature Name	Previous Type	Sp	Sv	F. S.	Me1	Me2	Me3	Me4	Me5	Me6	Me7	Me8	Assgn							
					Assgn															
podocarpaside F	Podocarposiide	<i>p</i>	C	400	0.87	30	0.96	28	1.04	18	1.05	21	1.26	29	1.45	26	1.46	27		
24- <i>epi</i> -15-carboxy-7,8-dehydro-16-oxo-15,16-seco-cimiracemoside E	15,16-Secocimicidol	<i>n/a</i>	P	500	1.05	21	1.06	30	1.33	29	1.61	27	1.67	26	1.96	28	1.98	18	2.10	24-OAc
24- <i>epi</i> -24-deacetyl-7,8-dehydro-16-oxo-15,16-secocimiracemoside E-15-aldehyde	15,16-Secocimicidol	<i>n/a</i>	P	500	1.02	21	1.05	30	1.31	29	1.56	18	1.62	28	1.69	26	1.74	27		

## VITA

NAME: Feng Qiu

EDUCATION: Ph.D., Pharmacognosy, University of Illinois at Chicago, Chicago, IL, 2012  
B.S., Pharmaceutical Engineering, Chengdu University of Traditional Chinese Medicine, Chengdu, China, 2005

TEACHING: Teaching Assistant, University of Illinois at Chicago, 2007–2009

RESEARCH: Research Assistant, University of Illinois at Chicago, 2009–2012

HONORS: W.E. van Doren Scholarship, University of Illinois at Chicago (2011)  
Travel Award, American Society of Pharmacognosy (2010)

PROFESSIONAL MEMBERSHIP: American Society of Pharmacognosy (ASP)  
American Chemical Society (ACS)  
American Association for the Advancement of Science (AAAS)

PRESENTATIONS: Qiu, F. (Invited Speaker); Imai, A.; McAlpine, J. B.; Lankin, D. C.; Burton, I.; Karakach, T.; Farnsworth, N. R.; Chen, S.-N.; Pauli, G. F. NMR Pattern Recognition Enables Dereplication of Natural Products, International Small Molecule NMR Conference, Providence, RI (2012).

Qiu, F.; Friesen, J. B.; McAlpine, J. B.; Lankin, D. C.; Pauli, G. F. Targeted Purification of Natural Products by Orthogonal *K*-Based Countercurrent Separations, International Congress on Natural Products Research, New York, NY (2012).

Qiu, F.; Friesen, J. B.; Chen, S.-N.; McAlpine, J. B.; Pauli, G. F. Chromatographic Orthogonality Resolves Structurally Similar Natural Products, 52<sup>nd</sup> Annual Meeting of American Society of Pharmacognosy, San Diego, CA (2011).

Qiu, F.; Friesen, J. B.; Cho, S.; Becker, B.; Franzblau, S. G.; Pauli, G. F. EECCC Fractionation and Biochemometric Evaluation of Anti-Tuberculosis Active Principles in Lipophilic Hops Extract, 6<sup>th</sup> International Conference on Countercurrent Chromatography, Lyon, France (2010).

Qiu, F.; Gödecke, T.; Chen, S.-N.; Li, J.-H.; Nikolic, D.; van Breemen, R. B.; Farnsworth, N. R.; Pauli, G. F. Phytochemistry of *Cimicifuga* Triterpenes with P450 Inhibitory Activity, 48<sup>th</sup> Annual MIKI Meeting, Chicago, IL (2010).

Qiu, F. Chemical and Biological Properties of *Apiaceae* Natural Product Z-ligustilide, Literature Seminar, UIC, College of Pharmacy, Chicago, IL (2009).

Chadwick, L. R.; Pro, S.; Friedl, W.; Qiu, F.; Rodríguez Brasco, M. F.; Friesen, J. B.; Pauli, G. F. Optimization of Countercurrent Chromatography Operational Parameters for Natural Products Analysis, 50<sup>th</sup> Annual Meeting of American Society of Pharmacognosy, Honolulu, HI (2009).

PUBLICATIONS:

Qiu, F.; Imai, A.; McAlpine, J. B.; Lankin, D. C.; Burton, I.; Karakach, T.; Farnsworth, N. R.; Chen, S.-N.; Pauli, G. F. Barcoding and Differential Analysis of 2D-NMR Spectra for Chemical Identification of Natural Product Mixtures, in preparation.

Qiu, F.; Cai, G.-P.; Jaki, B. U.; Franzblau, S. G.; Pauli, G. F. Quantitative Purity–Activity Relationships of Natural Products: The Case of Anti-Tuberculosis Active Triterpenes from *Oplopanax horridus*, *Journal of Natural Products*, in press.

Dong, S.-H.; Nikolic, D.; Simmler, C.; Qiu, F.; van Breemen, R. B.; Soejarto, D. D.; Pauli, G. F.; Chen, S.-N. Diarylheptanoids from *Dioscorea villosa*, *Journal of Natural Products* 2012, 75, 2168–2177.

Qiu, F.; Imai, A.; McAlpine, J. B.; Lankin, D. C.; Burton, I.; Karakach, T.; Farnsworth, N. R.; Chen, S.-N.; Pauli, G. F. Dereplication, Residual Complexity and Rational Naming: The Case of *Actaea* Triterpenes, *Journal of Natural Products* 2012, 75, 432–443.

Qiu, F.; Friesen, J. B.; McAlpine, J. B.; Pauli, G. F. Design of Countercurrent Separation of *Ginkgo biloba* Terpene Lactones by Nuclear Magnetic Resonance, *Journal of Chromatography A* 2012, 1242, 26–34.

Zhang, M.; Qiu, F.; Xie, X.-J.; Liu, J.; Chen, X. Metabolomic Analysis of the Extracts of *Pueraria lobata* Using HPLC/UV-DAD, *China Journal of Traditional Chinese Medicine and Pharmacy* 2008, 23, 215–217.

Zhang, M.; Qiu, F. Establishment and Application of Analytical Chemistry Test Database for Pharmacy Undergraduates. *Pharmaceutical Education* 2007, 23, 49–51.