Applications of chemometrics techniques on chromatographic and spectroscopic methods to advance chemical analysis of *Radix Ligustici Chuanxiong*, *Radix Angelicae Sinensis*, *Cortex Phellodendri* and other Chinese Herbal Medicines

by

CHAN Chi-on

A thesis submitted in partial fulfillment of the requirements for the Degree of

DOCTOR OF PHILOSOPHY

in

the department of Applied Biology and Chemical Technology

at

THE HONG KONG POLYTECHNIC UNIVERSITY

July, 2006

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

CHAN Chi-on

ABSTRACT

Abstract of the thesis entitled

"Applications of chemometrics techniques on chromatographic and spectroscopic methods to advance chemical analysis for *Radix Ligustici Chuanxiong*, *Radix Angelicae Sinensis*, *Cortex Phellodendri* and other Chinese Herbal Medicines"

Submitted by CHAN Chi-on

For the degree of Doctor of Philosophy in Chemistry

at The Hong Kong Polytechnic University in July, 2006

Chromatography and spectroscopy are the most commonly used analytical methods in various fields of chemistry and their instrumental technologies have been well developed. Yet, Chinese Herbal Medicine (CHM) is not easy to be analyzed because of their complexity of chemical composition. Data processing techniques provides good opportunity for mining more useful chemical information from original information-rich data.

Chemometrics is a chemical discipline that uses mathematics, statistics to extract more useful chemical information from chemical data. Owing to its rapid development, several effective and precise techniques are invented to process data from advanced hyphenated instruments so as to characterize the chemical composition of CHM in more detail. In our works, conventional approach, 'marker' and recently introduced approaches including 'multi-component' and 'pattern' were devised and applied into

chromatographic and spectroscopic data sets together with chemometrics techniques for developing the chromatographic fingerprints of three CHMs, *Radix Ligustici Chuanxiong* (CX), *Radix Angelicae Sinensis* (DG), *Cortex Phellodendri* (HB) for both qualitative and quantitative analyses.

Chemometrics resolution method and spectral correlative chromatography were applied for finding out the common constituents of these CHMs while local least square method was used for chromatographic alignment during constructing the chromatographic fingerprint. The overall results indicated that pattern approach, with the use of the whole chromatogram, not only distinguishes geographical locations including Szechuan, Yunnan and Guizhou provinces of CX as well as *Phellodendron chinense Schneid* and *Phellodendron amurense Rupr* of HB through multivariate analysis, but is also feasible on classifying the two CHMs, CX and DG, with 100% correctness in which five-folds cross validations were involved via several pattern recognition methods.

Finally, the rapid and non-destructive near-infrared reflectance spectroscopy (NIRRS) was utilized in analyzing chemical compositions of CHMs. With the proper use of chemometric pre-treatment and processing techniques on the full NIR spectra, the feasibility of applying NIRRS to CHM for identification of different parts of *Herba menthae*, of species of HB and of five different CHMs and quantification of major ingredients were achieved.

In conclusion, we believe that using most chemical information in data analysis can advance the standard of quality control of CHM as demonstrated in this work.

ACKNOWLEDGEMENTS

I would like to thank the Research Committee of The Hong Kong Polytechnic University for providing a research fund, and the University Grant Council (UGC) of Hong Kong Special Administrative Region via the Area of Excellent (AoE) of "Chinese Medicine Research and Further Development" (AoE/B-10/01) for the kind support in granting this project including support for my attendance of the course "Method validation: key issues and tools" offered in the 25th International Symposium on Chromatography in France in 2004 and different symposia held in Hong Kong, Taiwan and Denmark. I also wish to express my gratitude to all those have helped in various ways during the course of this work for support of this project.

I would like to thank Professor Foo-tim CHAU, my supervisor, for his critical advice and supervision of this project as well as his comments on the final draft of this thesis and the publications. I am greatly indebted to Dr. Jian-yong WU, my co-supervisor, and Dr. Daniel Kam-wah MOK for the helpful discussions of the progress of this project. Special thanks due to Dr. Zhong-Da ZENG for giving his valuable suggestions on chemometrics techniques.

Finally, I am indebted all the members of my family and Miss Yu-ning WONG for their constant encouragement during the course of this work especially in the hard times.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviations

ASTM	American Society of Testing Materials
BH	*Herba menthae*
CA	Cluster Analysis
CE	Capillary Electrophoresis
CGE	Capillary Gel Eletrophoresis
CHM	Chinese Herbal Medicine
CRM	Chemometric Resolution Method
CX	*Radix Ligustici chuanxiong*
CZE	Capillary Zone Electrophoresis
D	Bias
DAD	Diode Array Detector
DCB	Dichlorobenzene
DG	*Radix Angelicae sinensis*
DT	Detrending
ED	Euclidean Distance
EFA	Evolving Factor Analysis
ELPG	Evolving Latent Projection Graph
FA	Factor Analysis
FSMWEFA	Fixed Size Moving Windows Evolving Factor Analysis
FSWM	Fixed Size Window Method
FT	Fourier Transform

GC	Gas Chromatography
GHX	*Herba Pogostemonis*
HB	*Cortex Phellodendri*
HCA	Hierarchical Cluster Analysis
HELP	Heuristic Latent Evolving Projection
HKCMMS	The Hong Kong Chinese Materia Medica Standards
HPLC	High-performance Liquid Chromatography
HPTLC	High-performance Thin Layer Chromatography
ITTFA	Iterative Target Transformation Factor Analysis
IT	Information Theory
KM	Kubelka-Munk
k-NN	k – nearest neighbour
LC	Liquid Chromatography
LDA	Linear Discriminant Analysis
LLS	Local Least Square
LOD	Limit of Detection
LOQ	Limit of Quantification
MAE	Microwave Assisted Extraction
MD	Mahalanobis Distance
MECC	Micellar Electrokinetic Capillary Chromatography
MIR	Mid Infrared
MLR	Multiple Linear Regression
MS	Mass Spectrometry
MSC	Multiplicative Scatter Correction
MSD	Mass Spectrometric Detection

MST	Minimal Spanning Tree
NIR	Near-infrared
NIRRS	Near-infrared reflectance spectroscopy
NIST	National Institute of Standards and Technology
NMR	Nuclear Magnetic Resonance
NN	Neural Network
PC	Principal Component
PCA	Principal Component Analysis
PCR	Principal Component Regression
PLE	Pressurized Liquid Extraction
PLSR	Partial Least Squares Regression
RG	*Cortex Cinnamomi*
PRESS	Prediction Residual Sum of Squares
PRM	Pattern Recognition Method
R	Correlation Coefficient
RMSECV	Root Mean Standard Error of Cross Validation
RMSEP	Root Mean Standard Error of Prediction
RSD	Relative Standard Derivation
SCC	Spectral Correlative Chromatography
SD	Standard Derivation
SFE	Supercritical Fluid Extraction
SG	Savitzky-Golay
SI	Similarity Index
SIMCA	Soft Independent Modeling of Class Analogy
SNV	Standard Normal Variate

S/N	Signal-to-noise ratio
TCM	Traditional Chinese Medicine
TIC	Total Ionic Current
TLC	Thin Liquid Chromatography
UV	Ultra-visible

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Introduction

1.1 Introduction

Traditional Chinese Medicine (TCM) has been used for curing and preventing human diseases for more than five thousands years in China and it has been recognized to be an effective and valuable health care alternative. However, people, especially in the western countries, have still not generally accepted. One of the reasons is that lack of accepted scientific methodology for evaluating the quality of Chinese herbal medicines (CHMs). This is one of key criteria needed to support its use worldwide due to its efficacy. There has been increasing trend in identifying and quantifying the active ingredients in crude plant material and hyphenated instruments like High Performance Liquid Chromatography – Diode Array Detector (HPLC-DAD) and Gas Chromatography – Mass spectrometry (GC-MS) are the commonly used methods for achieving such purposes (details in Chapter 2). Yet, huge amount of data acquired from these instruments look very complicated and is not easy to interpret. Thus, data handling, interpreting, predicting chemical data as well as extracting chemical information are desirable to make the best use of these data.

Owing to the rapid development in instrumental design and computation speed, huge amount of data is possible to be generated and recorded by instrument. To process and interpret these large amounts of data, efficient data analysis technique is required. Chemometrics is a new discipline in chemistry that applied mathematical and statistical tools and logic-based method to analyze chemical rich data, in particularly in analytical chemistry. Chemometrics has been now widely applied to the fields of pharmaceutical, food, industrial processing and medical study. The fundamental concepts of the chemometrics tools utilized in this investigation will be given in

Chapter 3.

In this research project, three different CHMs including *Radix Ligustici chuanxiong*, *Radix Angelicae sinensis* and *Cortex Phellodendri* were selected to investigate their chemical constituents via the high-performance liquid chromatography (HPLC) and gas chromatography mass spectrometry methods (GC-MS), coupled with chemometrics techniques together. There are several ways to analyze chromatographic data of herbal medicine and they are named as ‘marker approach’, ‘multi-component approach’ and ‘pattern approach’ in this thesis [1]. Marker approach, the commonly way, makes use of the contents of one or a few markers to characterize the quality of CHM while the multi-component approach extend the concept of marker approach with much more components, i.e. common constituents, being considered. As for pattern approach, the whole chromatogram is considered for quality characterization.

Radix Ligustici chuanxiong, known as Chuan Xiong, is one commonly used CHMs and its essential oil is believed to be responsible for the therapeutic effects. The ‘Chuan’ in its Chinese name referring to the species cultivated in the Szechuan province has the best medical value. We also present a comparison of the performance of, ‘marker approach’, ‘multi-component approach’ and ‘pattern approach’ in developing the chromatographic fingerprint (GC-MS) of *Radix Ligustici chuanxiong* on quality control purpose. Several chemometrics tools including chemometric resolution method, spectral correlative chromatography and local least square method were involved to extract more chemical information as well as eliminate instrumental interference. It is the first work of this kind with emphasis on the comparison of the performance of these three approaches for chemical characterization of Chuan Xiong

samples from different geographical sources. Detail of this investigation will be discussed in Chapter 4.

The essential oil chemical compositions between *Radix Ligustici chuanxiong* and *Radix Angelicae sinensis* are similar to each other in which z-ligustilide is the main component and phthalide components are the major class of chemicals. So far, it is difficult to distinguish them through the conventional marker approach. Based on the previous three different methodologies mentioned, another comparison study has been carried out on developing a classification model between *Radix Angelicae Sinensis* and *Radix Ligustici chuanxiong*. 20 samples of *Radix Angelicae Sinensis* and 32 samples of *Radix Ligustici chuanxiong* collected from different geographical sources were analyzed through GC-MS. Similarity index evaluations, hierarchical clustering analysis, principal component analysis and linear discriminant analysis have been tried on all 52 chromatograms for attempting to develop the classification model which five-folds cross validation were involved for model evaluation. Details of this study will be given in Chapter 5.

Cortex Phellodendri is one of sixty CHMs commonly used in Hong Kong and it comes from either *Phellodendron chinense Schneid* or *Phellodendron amurense Rupr*. The appearance of these two CHMs is similar to each other and both consist of alkaloids, but with different amount. In this work, thirty samples of *Cortex Phellodendri* with different geographical locations were analyzed by a new extraction method and a new HPLC chromatographic condition developed by us. In addition, multivariate analysis was also involved in differentiating these two species. Details will be provided in Chapter 6.

Near Infrared reflectance spectroscopy (NIRRS) has a rapid development in the last decade in food and pharmaceutical industry. The rapid and non-destructive nature of this spectroscopic method and the availablility of appropriate chemometrics techniques in data pre-processing and treatment are the main reasons for its tremendous expanding usage. The fundamental concepts and general operation procedures will be given in Chapter 7. In this investigation, we not only applied the NIRRS on authentication of CHMs including, differentiation between different parts of *Herba menthae*; different species of *Cortex Phellodendri*; different CHMs like *Rhizoma Chuanxiong*, *Radix Angelicae sinensis*, *Cortex Cinnamomi*, *Cortex Phellodendri* and *Herba Pogostemonis*, by using principal component analysis, but also carried out the quantitative analysis on active ingredient(s) in *Cortex Phellodendri* by using partial least square regression. Chapter 8 gives the details of the study.

In conclusion, in this thesis, the chemometrics techniques are the framework of the whole research works under quality control of Chinese Herbal Medicines. Combined with the use of chromatographic and spectroscopic methods, fingerprint concept has been introduced and successfully demonstrated how it works on for the analyses of CHMs including *Radix Ligustici chuanxiong*, *Radix Angelicae Sinensis* and *Cortex Phellodendri*. We believed that using much more chemical information for data analysis can advance in the quality control of CHM to a higher level.

Chapter 2: Review on quality control of Chinese Herbal Medicine

2.1 Introduction

Chinese herbal medicine (CHM) has been the most organized traditional medicine in the world and it has been utilized for the treatment of disease for thousands of years. Throughout China's long history, Chinese people have accumulated a rich empirical knowledge of the usage and properties of natural products and an enormous body of human clinical data on their efficacy and toxicity. The therapeutic effects and minimum side effects of many herbal remedies have recently been demonstrated or verified in multitudinous modern scientific investigations. These therapeutic effects are often harmonized to those of western drugs, which explain the recent explorer of enthusiasm worldwide on the study of CHMs. However, one of the characteristic on CHMs preparation in decoction process is that all CHMs, either a single herbs or a collection of herbs in composite formulae, are mainly extracted with boiling water only. It was totally different from preparation and usage of the western drugs, which required chemical standardization. Liang et al. pointed out that it may be one of main reason why quality control of herbal drugs is much difficult than that of western drugs [2].

In "General Guidelines for Methodologies on Research and Evaluation of Traditional Medicines', Zhang pointed out "Despite its existence and continued use over many centuries, and its popularity and extensive use during the last decade, traditional medicine has not been officially recognized in most countries. Consequently, education, training and research in this area have not been recorded due attention and support. The quantity and quality of the safety and efficacy data on traditional medicine are far from sufficient to meeting the criteria needed to support its use

worldwide. The reasons for the lack of research data are due to not only to health care policies, but also to a lack of adequate or accepted research methodology for evaluating traditional medicine" [3].

Although many CHMs are effective in treating diseases, their remedial mechanism is still not well understood. Analysis of active components in Chinese medicinal extracts is an main key to explore the secret of their effectiveness. The major compound types in CHMs consist of alkaloids, saponins, flavonoids, anthraquinones, terpenoids, coumarins, lignans, polysaccharides, polypeptides and proteins. Efficient detection and rapid characterization of these components on a molecular basis play an important role as analytical support in scientific studies aimed at a better understanding of the pharmacological basis of CHMs. The general approach to such studies is outlined in Fig 2.1 involves the preparation of herbal extracts, tests their pharmacological activity, isolating the individual components of the extracts by using chromatography or electrophoresis techniques such liquid chromatography (LC) and capillary electrophoresis (CE), and then performing structure elucidation by spectroscopic techniques such as nuclear magnetic resonance (NMR) and mass spectrometry (MS).

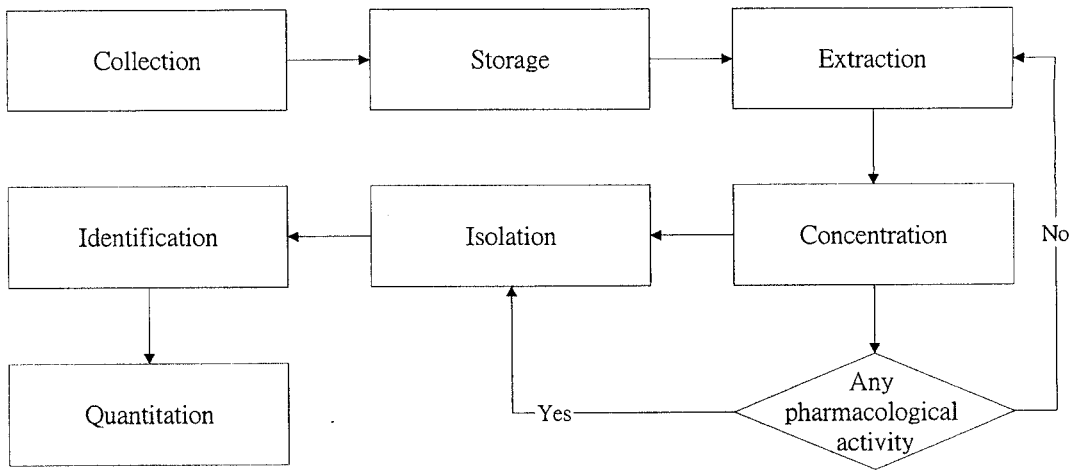


Fig 2.1 Flow diagram of the general procedures used to explore the remedial mechanism of CHM

However, CHM consists of hundreds of unknown components and many of them exist in low amount. Furthermore, even for the same CHMs materials, variability does commonly existed. Consequently, to obtain reliable approach for chemical characterization with pharmacological activities is a challenging job. Fortunately, chromatography offers very powerful separation ability, such that the complicated CHMs extracts can still be separated into relatively simple sub-fractions. Also, with the help of hyphenated instrument such as high-performance liquid chromatography – diode array detection (HPLC-DAD), gas chromatography – mass spectroscopy (GC-MS), capillary electrophoresis – diode array detection (CE-DAD), HPLC-MS and HPLC-NMR, it provides the additional spectral information, which is useful for the qualitative analysis and on-line structural elucidation. The spectral information obtained from hyphenated instruments provides alternative way to improve the performances in term of the elimination of instrumental interferences, retention time shift correction, selectivity, chromatographic separation abilities and measurement precision. These excellent properties are called dimension advantages proposed by Booksh and Kowalski [4].

Nowadays, quality related problems including lack of consistency, safety and efficacy seem to be overshadowing the potential genuine health benefits of various CHMs and their products, and a major cause of these problems seems to be related to the lack of simple and reliable analytical techniques and methodologies for the chemical analysis of herbal materials. A key factor in the widespread acceptance of TCM therapy by international community is “modernization” of CHM, i.e. quality control of CHMs is standardized by the use of modern science and technology.

In general, the methods for quality control of CHMs involve sensory inspection (macroscopic and microscopic examinations) and sophisticated analytical inspection using instrumental techniques such as thin layer chromatography (TLC), HPLC-DAD, GC-MS and LC-MS. Besides, sample preparation also play an important role in quality control of CHM, so it is possible to improve the efficiency, selectivity and sensitivity of the whole process by eliminating the interferences caused by herbal matrix. In this review chapter, how to develop an efficient evaluation for the purpose of quality control is the main section focused. Starting from sample preparation, chromatographic analysis, quality evaluation by ‘marker approach’, ‘multi-component approach’ and ‘pattern approach’ will be concisely discussed as follows [1].

2.2 Sample preparation

Sample preparation is often a neglected research area and it has received much less attention than the chromatographic separation or detection stages. However, getting the sample preparation stage correctly can be economically valuable as well as analytically important in the last decade. When separation methods provide high resolution of complicated mixture, the whole analytical process still can be wasted if an unsuitable sample preparation method has been utilized before chromatographic analysis, i.e. poor sample treatment will invalidate the whole assay, even using the most powerful techniques. Therefore, an inefficient or incomplete sample treatment can represent a considerable constraint on the throughput of any analytical methods [5]. For the quality control of CHMs, sample preparation is the first crucial step in the analysis of CHMs, because it is necessary to extract the desired chemical components

from the herbal materials for further separation and characterization.

2.2.1 Aims of sample preparation

The basic concept of a sample preparation method is to obtain a sub-fraction of the original sample enriched in most or even all the substances of analytical interest that is suitable for analysis by a separation or other analytical techniques. This can be achieved by employing a wide range of techniques. Their aims are nearly the same and list out as below [5]:

- To remove the potential interferences via the separation process from the sample for increasing the selectivity of the method.
- To increase the analyte concentration for enhancing the sensitivity of the assay.
- To convert the analyte into a more suitable form for detection or separation.
- To provide a robust and reproducible method that is independent of variations in the sample matrix

2.2.2 Traditional sample preparation

Sample preparation is one of the most important step in the development of analytical methods for the analysis of CHMs. For CHM, the target class components may be non-polar to polar and thermally labile so that the suitability of the extraction method must be considered carefully. In addition, the presence of analyte-matrix interaction in CHM may hinder the extraction process. Different methods of extraction with different conditions may often be required for the extraction of interested candidates from different CHMs. Even with the same techniques of extraction, for different

interested analytes in the same CHMs, different experimental conditions such as the solvent used, temperature applied and others may be considered.

The basic operation consists of several steps including pre-washing, drying, grinding to obtain a homogenous sample and often improving the kinetics of analyte extraction. For the monographies stated in the United States [6], Janpanese [7] and Chinese pharmacopeia [8], methods, such as sonication, heating under reflux, soxhlet extraction, steam distillation and others are commonly used. However, these methods can be time-consuming and require the use of large amount of organic solvent and have lower extraction efficiencies when the analytes of interest are not easily extractable.

2.2.3 New sample preparation

In order to reduce or eliminate the use of organic solvent and improve the extraction processes, newer sample preparation methods like microwave assisted extraction (MAE), supercritical fluid extraction (SFE) and pressurized liquid extraction (PLE) have been introduced. Both developed methods have been geared toward to the following trends [5]:

- Use smaller initial sample sizes even for trace analysis
- Have greater specificity or greater selectivity in extraction
- Increase potential for automation or on-line methods with less manual operations.
- Have more environmental friendly approach with less waste and the use of small volumes or no organic solvents.

The principle of all new extraction techniques are briefly discussed as follows. For MAE, the microwave energy not only uses for solution heating and results in significant reduction of extraction time, but also enables a significant reduction in the consumption of organic solvents. For SFE, with the unique properties of supercritical fluid such as higher diffusion coefficient and lower viscosity, it allows more efficient extraction with highly selectivity, even in the complicated mixture. For PLE, with the advantageous mass transfer properties of extraction medium under high pressure and temperature environment, lower viscosity properties of extracting solvent disrupts the solute-matrix interactions and increases the diffusion coefficients so that extraction efficient becomes higher [5, 9, 10].

Several successful studies demonstrated the sample preparation on CHMs. Pan et al. developed a MAE method for the root of *Salvia miltiorrhiza* with analysis by HPLC [11]. Gou et al. showed that the extraction of puerarin from *Radix puerariae* took 1 min with the aid of MAE [12]. Ma et al. optimized the SFE condition for extraction active ingredients from *Curcuma zedoaria* [13]while Li et al. used SFE to extract indirubin in *Strobilanthes cusia*, *Isatis tinctoria* & *Polygonum tinctorium* [14]. Ong et al. utilized PLE for extracting berberine and aristolochic acids I and II in *Coptidis rhizome* and *Radix aristolochiae*, respectively [15]; Lee et al. developed PLE for extracting ginsenosides in *Panax ginseng*, American ginseng [16].

2.3 Characterization of chemicals by hyphenated chromatography methods

Either a single CHM or a combination of several CHMs can be extracted with

hundreds of natural constituents during the sample preparation. Several chromatographic methods including TLC, GC-MS, HPLC-DAD / HPLC-MS and CE – DAD / CE – MS provide a powerful tool for the separation of those extracted components in order to characterize the chemical composition of CHM(s). A brief summary of these separation methods are discussed in the follow sections.

2.3.1 Thin Layer Chromatography

Thin layer chromatography (TLC) has been the most widely used classical method for chemical analysis in CHM. Chromatographic profiles of major components are used to evaluate herbal growers and suppliers, to standardize raw materials and to control formulation and tablet content uniformity. The Chinese Herbal Pharmacopoeia [8], American Herbal Pharmacopoeia [6] and British Herbal Pharmacopoeia [17] have had an emphasis on utilizing TLC profiles to characterize herbal materials, relying on the use of different spray reagents and TLC profiles to identify characteristic and active constituents of herbal materials.

Liang et al. also pointed out that TLC is used as an easier method of initial screening with a semi-quantitative evaluation together with other chromatographic methods [2]. TLC not only provides the multi-fold possibilities of detection in analyzing CHMs, but also employs for multiple sample analysis, i.e. more than 20 samples can be analyzed simultaneously in one time. Xie et al. developed high performance thin-layer chromatography (HPTLC) for authentication of various species of ginseng (*Panax ginseng*, *Panax quinquefolium*, *Panax noto-ginseng*) and stability of ginseng preparations [18]; Di. Et al. also applied HPTLC for differentiating different species

of *Ganoderma* [19]. In summary, the advantages of using TLC are its simplicity, versatility, specific sensitivity and simple sample preparation. Thus, it is also a convenient method for determining the quality and adulteration of CHMs [2].

2.3.2 Gas chromatography

Many pharmacologically active components in CHMs are volatile chemical compounds, so gas chromatography (GC) is suitable in analyzing volatile composition of CHMs. With the use of capillary column, it is able to separate many volatile compounds simultaneously within comparatively short durations, even in the complex system. Its high sensitivity of detection of almost volatile chemical compounds also serves as critical advantage. In addition, interfaced with mass spectrometric detector (MSD) and utilized the mass spectral database, qualitative and relatively quantitative chemical composition of CHMs investigated could be provided, which could be extremely valuable for further study in elucidating the relationship between chemical constituents in CHMs and its pharmacology. Liang et al. also pointed out that GC-MS is the most preferable tool for characterizing the volatile composition of CHMs [2].

In this last decade, GC is so popular in the research field in TCM and lot of GC-MS literatures have been reporting the chemical compositions of CHMs. Gong et al. studied the chemical compositions of *rhizoma chuanxiong, radix angelicae sinensis,* and si-wu decoction using GC-MS [20]. Guo et al. used GC-MS to characterize *Artemisia capillaries* from different geographical location [21]. Gong et al. characterized the chemical composition of *Cortex Cinnamomi* from different provinces by GC-MS [22]. Shen et al. reported quality assessment of *Flos*

Chrysanthemi Indici by GC-MS [23]. Li et al. analyzed of volatile fractions of *Schisandra chinensis* (Turcz.) Baill. using GC-MS [24].

The most serious disadvantage of GC, however, is that it is inconvenient for analyzing polar and non-volatile compounds. For this reason, it is necessary to use tedious sample preparation such as derivatization before hand. Therefore, the liquid chromatography becomes another important analyzing tool for comprehensive analysis of CHMs [2].

2.3.3 High-performance liquid chromatography

High performance liquid chromatography (HPLC) is a very popular analytical method for CHMs because the method is not limited by the volatility or stability of the analytes. The comparative advantage of HPLC lies in its versatility for the analysis of the chemical compounds in CHMs. Therefore, almost all compounds in CHMs can be analyzed through HPLC. Significantly increase in the number of HPLC conditions for CHMs in China Pharmacopoeia 2005 edition reflects the analytical trend that HPLC becomes a powerful tool on chemical analysis of CHM [25].

In the last decade, increasing usage of LC-MS and HPLC-DAD in analyzing CHMs became so obviously. Several good reviews have been published for the analysis of the bioactive chemical compounds in CHMs, which the method involved is hyphenated HPLC. With the use of diode array detector (DAD) and mass spectrometric detector (MSD), qualitative analysis of CHMs turns out to be much easier than before, because of the additional information of ultra-visible (UV) and

mass spectral data for individual peaks in a chromatogram. Through which, it is possible to identify the chromatographic peaks directly on-line by comparison with literature data or standard compounds. Also, it avoids the time-consuming isolation process before all compounds to be identified as that adapted in the traditional approach.

A lot of studies utilized HPLC for quality assessment and control of CHMs and their commercial products. For example, Yang et al. developed new HPLC-DAD method for quality control of *Tianjihuang* [26]. Ji et al. also studied the *Ginkgo biloba* extract by HPLC-DAD [27]. Luo et al. also used HPLC-DAD-MS to characterize the extracts of *Echinacea purpurea* [28]. Fan et al. also developed novel multiple HPLC method for quality control of *Danshen Dropping Pill* [29]. Lu el al. also reported the new HPLC-DAD-MS method for distinguishing *Chinese Angelica* from related umbelliferae herbs [30]. Zhang et al. analyzed the chemical compositions of *Danshen* injection products by HPLC-UV and HPLC-MS [31].

To sum up, the advantages of liquid chromatography include its high reproducibility, good linear range, ease of automation and its ability to analyze the number of constituents in CHM. However, for the analysis of marker compounds with more than two CHMs, co-eluting peak were commonly present in the chromatograms obtained due to the complexity of the matrix. The complexity of matrix may be reduced with additional sample preparation steps such as liquid-liquid partitioning and solid phase extraction.

2.3.4 Electrophoretic methods

Capillary electrophoresis (CE) as a powerful analytical and separation method was introduced by Jorgensen and Lukacs in early 1980s [32]. CE is promising for the separation and analysis of active ingredients in CHMs, as it required small amount of sample and can analyze sample rapidly with very good separation. The common techniques are capillary zone electrophoresis (CZE), capillary gel electrophoresis (CGE) and micellar electrokinetic capillary chromatography (MECC). Several studies have been reported for analyzing the chemical composition of CHMs. Yu et al. studied paeonol and paeoniflorin in *cortex moutan* by MECC-UV [33]. Liu et al. reported that four flavonoids in *Epimedium* were determined by CZE-DAD [34]. Sun el al. analyzed the chemical composition of *Flos Carthami* through CE-UV [35].

In general, CE is a versatile and powerful separation method with a high separation efficiency and selectivity. However, Shibabi and Hinsdale pointed out that CE has still been focused on the resolution improvement rather than the reproducibility and absolute precision [36]. Reproducibility and precision in CE technique hindered the development the quality control of CHMs.

2.4 Approaches for quality control CHMs

2.4.1 Traditional approaches

In the previous discussion, it clearly mentioned that there are many chromatographic methods, including the hyphenated chromatographies, available for analytical chemist to characterize the chemical composition of CHM through instrumental analysis. The

problem here is how could efficiently and reasonably evaluate all the obtained analytical results and how could one use the chemical information buried in these chromatographic profiles to address the problem of quality control for CHM.

2.4.1.1 Marker approach

When embarking on development of active ingredient in CHMs traditionally, it must first be decided which compounds have to be quantitated. If a principle active component is known, it is most logical to quantitate this compound. Where active ingredients contributing to therapeutic efficacy are known, botanical preparation should be standardized to these compounds. Where the active ingredients are not yet known, a marker substance which should be specific for the botanical could be chosen for analytical purpose, although it should only serve for internal batch control. In the traditional approach, single marker can be used to ensure that the concentrations and ratios of components in an herbal mixture are present at reproducible levels in the raw materials, manufacturing, intermediates and in the finished products. In this way, marker(s) by itself gives a very few chemical information only to assist manufacturing control and assures batch-to-batch consistency.

Ideally, these selected markers should be responsible for their medical functions. In fact, with most phytomedicines, their pharmacokinetics and metabolomics to the therapeutic efficacy is either non-specific or even wholly unknown. In these cases, the quality of CHM is adjusted using certain selected marker(s) that are usually the characteristic chemicals for specific CHM which may not be the active components. In addition, the marker is not unique to a particular herb in many cases. For example,

Z-ligustilide is an active ingredient of *Radix Angelicae sinensis* and very often served as marker for identifying the herb [37]. Yet, the compound also presents in other herbs such as *Radix ligustici chuanxiong* [38]. Moreover, the therapeutic efficacy of certain CHM may be owing to the synergistic effect of several constituents and they could hardly be separated individually. Examples included *Echinacea* [39] and St. John's wort [40]. The overall result is that this kind of quality assessment does not give a complete picture of the quality of CHM.

2.4.1.2 Multi-component approach

The multi-component approach is a natural extension of the marker approach and was proposed in last decade. It uses the relative compositions (e.g. are under the chromatographic profile) of many or even all identified components in CHM for quality assessment [1]. In an example illustrated by Schaneberg et al, alkaloid HPLC profiles of more than twenty *Ephedra* species were developed and a series of peaks ranged from 52 and 64 min was utilized to verify the presence of an *Ephedra* species in the sample [41]. This approach has been applied to CHM like single chemical entity pharmaceuticals in every way possible. The drug substance is the plant or extract and the drug product is the formulated herbal product. In this framework, CHM and its products contrast to single chemical entities. However, Luo noted that it is impossible and unnecessary to separate and clarify each component in the CHMs as the chemical composition of a CHM varies with different parts, geographical sources and collection periods [42].

Besides, the drawback of marker approach and multi-component is the involvement of

data reduction. Data reduction is a simple, but efficiency way to enhance the data processing for bringing out the relevant information. It is because much of the chromatographic information is irrelevant (instrumental noise) or redundant (several measurements on the same peak). However, data reduction has an intrinsic problem: some chemical information must be invariably lost. Fundamentally, irrelevant information and relevant chemical information should be removed and extracted as much as possible, respectively during data treatment. But in many real cases, it is hard to define correctly what information is relevant and what should be discarded. This is especially true for noisy data or complex data such as those coming from highly overlapping peaks and shoulder peaks.

2.4.2 New approach

As noted previously, one or a few markers or pharmacologically active components were employed for evaluating the quality and authenticity of CHMs. However, this kind of determination does not provide a complete picture of CHM and most probably, it will definitely give false identification. Xie use the extracts of *Ginkgo biloba* illustrating the false identification based on the marker approach [42]. In the last decade, another new concept, chromatographic fingerprint (categorized as pattern approach), was proposed and recommended as a potential and reliable approach for the quality control of CHMs.

2.4.2.1 Pattern approach

The basic concept of the pattern approach is to utilize the whole pattern like the entire

chromatographic profile or the full spectrum as chemical features [1]. It may be discovered that the relation between the pattern and chemical composition of the CHM extract is unclear. However, the pattern of chromatographic profile or spectrum is solely contributed by the chemical components presented in CHM extract under the same experimental condition. Yet for CHM such a complicated mixture, chromatographic profile becomes more useful on chemical characterization as chemical components are being separated. In the following, pattern approach using chromatography techniques (known as chromatographic fingerprinting) will be briefly discussed.

Chromatographic fingerprinting characterizes the chemical patterns consisting of a set of peaks that present unique composition of samples (e.g. CHMs) embedded in chromatograms, i.e. utilize all chemical components detected by the instrument for analyzing. It is well established that the samples with similar chemical patterns likely have similar properties. Such a similarity measurement between chromatograms (vectors) can be represented in terms of Euclidean distance, the Pearson correlation coefficient, or the cosine of the angle and similarity ratio and it is essential for objective comparison. Several analytical chemists also appraised and noticed the advantages of chromatographic fingerprint. Luo noted that the fingerprint concept should be encouraged to investigate the quality and quantity of the main effective fractions and key constituents as much as possible to understand the main pharmaceutical constituents or their combinations and the foundations of pharmacological action [42]. Also, Xie advocated the use of the chromatographic pattern approach as a means of quality control from the considerations of non-linear nature of the TCM theory and TCM herbs [42]. Nielsen also pointed out that the

comparative advantage of using pattern approach does not involve peak selections, extractions and integrations on the chromatographic profile subjectively. Furthermore, another advantage of taking the whole chromatographic profile to perform analysis directly is that the peak shape can be included in data analysis [43].

Many studies have been carried out on the fingerprint of CHMs using different chromatography techniques. A short list of examples include the TLC studies of ginseng [44], American ginseng [45], notoginseng [46], lingzhi [19] and ginkgo [47] and the HPLC studies of *panax ginseng*, *panax quinquefolium*, and *panax notoginseng* [48], *Chelidonium majus L*. [49], *Radix Angelica sinesis* [30], *Radix Paeoniae rubra* [50], *Hypericum japonicum thunb* [26], and *Paeonia lactiflora* [51]. Also, examples of GC fingerprint included *Houttuynia cordata* [52], *Pogostemon cablin* [53], and *Alpinin officinarum* [54].

2.4.2.2 Difficulties encountered in pattern approach

Construction of chromatographic fingerprint becomes one of the most powerful approaches for quality control of CHMs. A chromatographic fingerprint is, in practice, a chromatographic profile representing multiple chemical components with characteristics from CHMs investigated so as to identify and assess the stability of the chemical constituents observed by chromatography. However, since CHMs with large number of chemical components are very complex [i.e. a chromatographic fingerprint is usually high dimensional (>> 10 variables, here a variable means a peak and is considered as one dimension)]. Also the chromatographic instruments and experimental conditions are difficult to be reproducible, the development of

chromatographic fingerprints of CHMs for quality control is not a trivial work in fact.

As a result, extracting chemical fingerprints is an important step for representing and interpreting chromatographic data. Currently, perhaps the most commonly adopted method is to collect particular chemical fingerprints (e.g. retention time, amount and character of the eluting components, ratio of the eluting components) from specific peaks of interest in the chromatograms. However, these chemical fingerprints rely heavily on experimental variables such as the method of peak detection. They also depend on many human variables since the analyst must decide which peaks to select from the chromatograms for effective classification or identification. The retention times of chromatographic peaks always vary due to many interfering factors. These limitations impede fast, automated, and impartial analysis of the chromatographic fingerprints.

Chemometrics is a term applied to the generic discipline involving computer and mathematics to extract meaningful chemical information from samples of varying complexity, and to develop method and result. In early practice, the chemical patterns of CHM are usually assessed by visual comparison of the spectra, which is often subjective and not always persuasive. However, application of chemometrics methods in the development of chromatographic fingerprinting offers several benefits; it is not subjective, analysis is consistent, and the results are reproducible.

Comparative analysis of the highly dimensional vectors poses a challenge in implementation and interpretation. In general, transforming the high dimensional vector into a low dimensional space could greatly facilitate comparative analysis of

chromatographic fingerprint and also be beneficial for ensuring the quality of CHMs. However, it is important that the low dimensional features generated from mathematical transformation should retain the critical information in the original fingerprint. These so-called fingerprint features extracted from the original one are often low dimensional and high discriminatory power. A number of well-developed chemometrics tools have been utilized into quality control of CHM through data transformation. Cheng et al. used fractal fingerprint transformed from different wavelet analysis for characterizing different grade of *radix Angelica sinensis* [55]. Cheng et al also developed a novel approach called Fisher components for extracting most discriminatory information from the original fingerprint of shenmai injection in order to examine the commercial products from different manufacturers [56]. Lu et al. used principle component analysis (PCA) method on their corresponding chromatographic profile for distinguishing *China Angelica* from related umbelliferae herbs [30]. Li et al. used standard normal variate transformation on chromatographic profile for finding out the differences of chemical components among 33 samples of *Erigeron breviscapus* [57].

Another important feature using chemometrics tools is not only to eliminate certain experimental interferences, but also resolve part of the chromatographic resolution problem as well as work together with experimental design to optimize like extraction and chromatographic conditions. In the chemometrics field, in order to making good use the properties of two dimension data sets from hyphenated chromatography, lots of chemometrics tools have been innovated and applied into real complex system. In the last decade, several successful tools have been developed maturely and provided an efficient way to built up chromatographic fingerprint of CHM effectively. For

example, Liang et al created a novel method, known as heuristic latent evolving projection (HELP), to resolve overlapped peaks existed in chromatogram [58, 59]. Keller and Massart developed a fixed size moving windows evolving factor analysis (FSMWEFA), to detect low concentrations of impurities even at very low concentration in the process of resolving the overlapped peaks [60]. Gong el al. utilized a cubic spline interpolation method for correcting the retention shift of chromatographic peaks in a series of chromatograms [61]. Li et al. developed an efficient calculation method, called spectral correlative chromatography (SCC), for finding out the common peaks in a series of chromatograms [62]. Also, multi-simplex and response surface methodology, a sub-branch of chemometrics discipline utilized in experimental design, were commonly accepted and used in the chromatographic optimization process [63].

2.5 Concluding remarks

There is a growing interest in CHM in recent years because of the advantages of low toxicity, rare complications and pharmacological activity. Different chromatographic and electrophoretic methods coupled with various extraction techniques used in the instrumental measurement of CHM were reviewed. Chemical fingerprints obtained by chromatographic and electrophoretic techniques through ‘marker approach’, ‘multi-component approach’ and ‘pattern approach’ were comprehensively discussed with their pros and cons under the purpose for quality control of CHM. Pattern approach embraced by Chinese authorities in the last decade becomes widely accepted and served as an important step towards modernization and internationalization of CHM. We believe that using all chemical information for data analysis could be a benefit in quality control of CHM. Besides, rapidly establishment of chemometrics tools also solve a series of obstacles in developing the fingerprint of CHM. We deem that chemometrics techniques coupled with hyphenated instruments provide the best tools at this moment for establishing the chromatographic fingerprint of CHM and playing an important role after correlating their pharmacological activity, as well as developing the functional fingerprint.

Chapter 3: Introduction to chemometrics tools utilized in the chromatographic fingerprint of Chinese Herbal Medicines

3.1 Introduction

Chemometrics is a chemical discipline that uses mathematics, statistics and formal logic mainly to design or select optimal experimental procedures, to provide maximum relevant chemical information by analyzing chemical data, and to obtain knowledge about chemical systems. Many disciplines of chemistry like analytical chemists, environmental chemists, forensic scientists and organic chemists are benefits from the application of chemometrics because of the need for data interpreting and design of experiment. Analytical chemists are major users of chemometrics as tackling spectra or chromatograms are their majority work. Huge quantities of data collected from instruments confronted them to study, implement the chemometrics tools so as to obtain the conclusive results efficiently.

Different types of people are interested in chemometrics and most of them are application scientists. Their main interest is to define the need for analysis, to design experiments and to interpret results. In quality control of CHM, creating the experimental design on experimentation and extracting relevant chemical information from original data are the most interested ones. Due to complexity of analyzing CHM compositions, a large number of trial and error run in experimentation and huge amount of data with higher dimensions have to be carried out and interpreted, respectively, in order to characterizing the chemical compositions, especially in chromatography. It is time consuming and heavy work-loading on experimental part and not easy to insight and analyze efficiently large amount of data. Therefore chemometrics tools provide objective, effective and precise way to extract and analyze data so as to explore the chemical compositions of CHM in more details. In the

following section, chemometrics tools on quality control of CHMs via chromatographic analysis are briefly introduced and the fundamental concepts about these tools are also given.

3.2 Tools for extracting chemical information

3.2.1 Heuristic evolving latent projection (HELP)

Heuristic evolving latent projections (HELP) algorithm was proposed by Kvalheim and Liang in 1992 for analyzing data from liquid chromatography with diode array detection (LC-DAD) [58, 59]. It is classified as a Factor Analysis (FA) or Principal Component Analysis (PCA) technique for multivariate data analysis. This algorithm is a modification of fixed sized window moving – evolving factor analysis (FSMW-EFA) method and was shown to be better than the conventional methods such as evolving factor analysis (EFA), iterative target transformation factor analysis (ITTFA) and alternating regression (AR) [64, 65]. Within the concept of HELP, it has been applied to resolve system containing partial chromatographic selectivity [65], hidden minor chromatographic peaks [66] and baseline drift in the elution profiles [67]. HELP has been applied to an HPLC-DAD / GC-MS study involving chemical analysis of CHMs. F. Gong et al. used HELP to characterizing the chemical compositions of *Cortex Cinnamoni*, si-wu decotion and peptic powder under GC-MS [20, 22, 68]. A.K.M. Leung analyzed the water soluble constituents of *Cordyceps sinensis* through HELP under HPLC-DAD [69].

In order to apply the HELP algorithm successfully on data analysis of HPLC-DAD /

GC-MS, two simple assumptions are made. Firstly, the Beer's law is observed with non-zero chromatographic and spectral signal. Secondly, the rule first-in-first-out is obeyed. That is the first substance present in the system will be the first one to disappear, the second substance will disappear next and so on [60]. It is supposed the HPLC-DAD system generates a data set X with the rows and columns corresponding to spectra (s_a) measured at different times and chromatograms (c_a) acquired at a different wavelength respectively:

$$X = \sum_{n=1}^{N} c_n s_n^t + E = CS^t + E \qquad \text{(Eq. 3.1)}$$

Here, N and E represent the total number of components and noise in the sample, respectively. With the help of PCA treatment, X can be decomposed into a matrix of score T and matrix of loadings P as follow:

$$X = CS^t + E = TP^t + E \qquad \text{(Eq. 3.2)}$$

In the HELP treatment, the first step involves the construction of the eigenvalues plot which is a plot of the logarithm of eigenvalues against retention time. Instead of adopting an increasing size window to calculate the eigenvalues, fixed size moving window – evolving factor analysis is utilized. A number of rows of X, which is greater than 1, is selected and their eigenvalues are computed. Then, the window is moved downward by one row and the calculation is repeated until the last row of the matrix X is manipulated. At the end of the computation, all the first eigenvalues thus obtained are connected with a line and the same procedure is applied again to the other eigenvalues.

In the second step of HELP analysis, the zero concentration, zero component and selective regions of the corresponding components are determined from the

eigenvalues plot. The zero concentration region is defined as the chromatographic region where nothing elutes from a particular component [58, 70]; and the selective region is considered as the region with only one chemical component eluted [58]. The zero component region is defined as a region with no chemical component involved and where usually noise presents only. In this stage, the evolving latent projection graph can also provide information on the selective region of individual components. Such regions can be identified from the location of straight lines which pass through the origin in the latent projective graph. In the third step, with the information of the zero concentration, zero component and selective regions, chromatograms and spectra of individual components can be resolved through simple matrix computation. In the simplest cases such as a simple two components system, if the chromatographic profiles of each component are known, the corresponding spectral profiles can be determined by the following equation:

$$S^t = (C^tC)^{-1}\ C^tX \qquad \text{(Eq. 3.3)}$$

Resolved single peak of a particular component can be determined from the zero concentration region of that component. This can be achieved using the first principle component of a sub-matrix as derived from the selective region. Similarity, the chromatographic profiles can be resolved if the spectral profiles are known.

$$C = XS\ (S^tS)^{-1} \qquad \text{(Eq. 3.4)}$$

For a more complicated system, the component stripping approach is required [71]. Resolution can be performed by using a stepwise stripping procedure. From equation Eq. (3.2), we have

$$CS^t = UP^t \qquad \text{(Eq. 3.5)}$$

Then, a rotation matrix R can be introduced into the above expression so that

$$CS^t = U\,RR^{-1}\,P^t \qquad \text{(Eq. 3.6)}$$

Eq. (3.6) implies that

$$C = UR \qquad \text{(Eq. 3.7)}$$

In this way, the non-orthogonal rotation matrix R and its inverse R^{-1} are employed to transform the score vector U from the principal component decomposition into concentration profiles of the pure chemical components [58]. In order to determine r_i for the individual component i, both the zero concentration and selective regions are incorporated into Eq. (3.8)

$$c_{sel+zero,i} = U_{sel+zero,i}\,r_i \qquad \text{(Eq. 3.8)}$$

The subscript sel+zero,i implies the use of the selective region in addition to the zero concentration region for component i. By rearranging Eq. (3.8), r_i can be written as

$$r_i = (U^t_{sel+zero,i}U_{sel+zero,i})^{-1}\,U^t_{sel+zero,i}c_{sel+zero,i} \qquad \text{(Eq. 3.9)}$$

$C_{sel+zero,i}$ can be obtained as the score vector of the first principle component in the selective region of component i. After determination of r_i for a particular component, its pure chromatogram can be derived via Eq. (3.7). For a system with more than two components, the component stripping approach is applied in this stage by removing the contribution from resolved component in the following manner

$$X_{new} = X - C_iS_i^t \qquad \text{(Eq. 3.10)}$$

The HELP analysis is algorithm applied repeatedly to X_{new} until all components are resolved. Similarly, GC-MS chromatogram can also be resolved in the same way.

3.2.2 Spectral Correlative Chromatography (SCC)

Spectral Correlative Chromatography (SCC) was proposed by Li et al. in 2004 for identifying the same chemical component present in different chromatograms as acquired from hyphenated instrument like GC-MS or HPLC-DAD [62]. The idea of SCC is quite simple; the same chemical component should have the same spectrum no matter what or how they are eluted through diverse chromatographic columns. The spectral information is utilized to pick up the targeted component from the other two-way chromatograms. The procedure for carrying out SCC is given in the following steps.

1) Assess peak purity of a targeted component and then acquire its UV or MS spectrum from the chromatogram.
2) Identify this component in the other chromatogram(s) of interest through comparing by a series of the spectrum found in (1) to that at each scan point of the other chromatogram(s) via their correlation coefficients.
3) Get a curve of the correlation coefficient obtained vs scan point in the direction of retention time and further validate the result via the step 2 with consideration of the information in local chromatographic cluster where the target exists.

Suppose that A_1 and A_2 are GC-MS data sets of two different samples, and s_i, s_j are a spectra of the *i*th and *j*th component of A_1 and A_2, respectively, the correlation coefficient between these two spectra is given by,

$$r(i,j) = \frac{(s_i - \overline{s_i})^T \bullet (s_j - \overline{s_j})}{norm\left|(s_i - \overline{s_i})\right| \bullet norm\left|(s_j - \overline{s_j})\right|} \qquad \text{(Eq. 3.11)}$$

Here, $\overline{s_i}$ is the mean of spectral vector s_i, and the Euclidean norm is used. The $r(i, j)$ values are in the range $-1 \le r \le 1$. The larger the value of $r(i, j)$, the higher the correlation between the two components. When r equals 1, these two spectra are identical. In view of errors and interferences from noise and background, etc., in actual cases, r is rarely equal to 1. Owing to the uniqueness of mass spectrum, the chance of any two components with exactly the same spectrum and very close retention time is very rare. Therefore, the results obtained from applying SCC to GC-MS are highly accurate and this algorithm performs well in LC-MS data also. As for HPLC-UV chromatogram, the outcome may not be as good as that of GC-MS because of the feature uniqueness of UV spectrum.

3.2.3 Information theory (IT)

In information theory, the value of information content depends on the separation degree and concentration distribution of each chemical component in a chromatogram. The more the separated peaks with uniform concentration distribution in a chromatogram are, the higher the value of information content are, i.e. the more chemical information is provided in this chromatogram. As a result, both separation degree and concentration distribution of components involved in a chromatographic fingerprint should be considered in data treatment. An information content Φ of a chromatogram is calculated based on:

$$\Phi = -\int \frac{p_x}{sum(p_x)} \frac{\log(p_x)}{sum(p_x)} dx \qquad \text{(Eq. 3.12)}$$

where p_x was the real chromatographic response of all chemical components involved in the chromatographic fingerprint under study [72]. By comparing the magnitudes of

information content, maximal chemical information under certain chromatographic condition, that includes all the extraction parameters and detection parameter can be found out.

3.3 Approach for analyzing data of Chinese Herbal Medicines

The basic concepts of marker(s) approach, multi-component approach and pattern approach have been introduced from chapter 2. The critical difference among them is the proportion of chemical information being utilized in data analysis, i.e. marker(s) content, relative percentages of common constituents and the entire of chromatographic profile. In this thesis, those extracted chemical information serves as input parameters and were arranged as a form of matrix,

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,i-1} & x_{1,i} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,i-1} & x_{2,i} \\ x_{j-1,1} & x_{j-1,2} & \cdots & x_{j-1,i-1} & x_{j-1,i} \\ x_{j,1} & x_{j,2} & \cdots & x_{j,i-1} & x_{j,i} \end{bmatrix}$$

where i = number of markers, number of common constituents & number of scan data point in the whole chromatographic profile in number of j analyzed samples, under marker approach, multi-component approach & pattern approach, respectively. Then, the matrix was treated by different pre-processing techniques and multivariate analysis, depending on what kind of objective to be achieved.

3.4 Tools for pre-processing data

3.4.1 Normalization

Normalization of the chromatographic profiles is also an important step in pre-processing analytical data. The variation in sample concentration might affect the multivariate analysis of the entire chromatographic profile. Therefore, normalization of data was examined prior to multivariate analysis and was carried out in the following way: Each GC-MS / HPLC-DAD chromatogram consisted of N peaks with each peak area c_i of the *i*th component being utilized. The total area of all peaks was calculated by:

$$C = \sum_{i=1}^{N} c_i \qquad \text{(Eq.3.13)}$$

Since no internal standard being added in the study, the peak areas were normalized as follow:

$$c_i^{'} = \frac{c_i}{C} \qquad \text{(Eq.3.14)}$$

Then, each of the resulting peak was expressed as a percentage of the sum of peak areas.

3.4.2 Local least Square (LLS) method

One of the problems in analyzing analytical signals is the occurrence of signal shift. It causes a significantly influence in chromatography under multivariate analysis. To eliminate the effect of chromatography shift on the data analysis, Local Least Squares (LLS) method [57, 61] was utilized to correct the retention time shift found in chromatographic field so as to achieve the multi-point alignment (in this case, 10 points) before comparing the similarity or dissimilarity between different chromatographic profiles. It is necessary to correct the retention shift present in all

chromatograms because it causes a serious problem in data analysis under pattern approach. The basis of pattern approach is that chromatograms concerned must contain the same chemical information at the same positions. To match all chromatographic profiles with their retention times, all common constituents were selected and utilized for chromatographic alignment. Several parameters among them have been estimated using LLS method and the new-aligned chromatographic profiles can be re-constructed through the cubic spline data linear interpolation techniques. For more details, Gong and Li have concise explanations on the LLS method [57, 61].

3.5 Tools for displaying data

3.5.1 Hierarchical cluster analysis (HCA)

Cluster analysis is a group of multivariate techniques in which the main purpose is to assemble objects based on their characteristics. Cluster analysis classifies objects with those being similar one and others grouping in the same cluster under a pre-determined selection criterion. In the ideal situation, the resulting cluster of objects should exhibit high homogeneity (intra-group) and high heterogeneity (inter-group). Hierarchical agglomerative clustering is the most commonly used one, which provides intuitive similarity relationships between any one sample and the entire data set. It is typically expressed graphically by dendrogram, which provides a visual summary of the clustering processes, presenting a picture of the groups and their proximity, with a dramatic reduction in dimensionality of the original data. Interpretation of a similarity dendrogram between samples is based on intuition: two samples which are close to each other should have similar values under the measured

variables concerned, i.e. they should be numerically close to each other in the multi-dimensional space. Therefore, the greater the proximity between the samples is, the higher the similarity between them is. Therefore, the dendrogram imposes a hierarchy on this similarity, so that a two-dimensional vision of the similarity and dissimilarity of the whole set of samples used in the study is clearly illustrated [63, 73]. Given a set of *N* data to be clustered, the basic process of hierarchical clustering is carried out via the following procedures:

1. Start by assigning each set data (vector) to a cluster, so that there are N clusters with a data set for each one. The distances (similarities) between the clusters are the distances (similarities) between the data sets they contain.

2. Find the closest (most similar) pair of clusters and merge them into a single cluster, so that one larger cluster is being formed.

3. Compute distances (similarities) between the new cluster and each of the old clusters.

4. Repeat steps 2 and 3 until all data sets are clustered into a single cluster of size *N*. A clustering tree will be obtained finally after finishing all the steps.

It can be seen from Step 2, that the similarity criterion is distance. An important component of a clustering algorithm is the distance between data points. Usually Euclidean distance and Mahalanobis distance are utilized to measure the distance between two clusters. However, there are no general guidelines in selection which

kind of distance. If n_r is the number of objects in cluster r and n_s is the number of objects in cluster s, x_{ri} is the *i*th object in cluster r, Euclidean distance and Mahalanobis distance were employed in this work and their expression are given as follows [63]:

Euclidean distance, $ED_{rs} = \sqrt{(x_r - x_s)(x_r - x_s)'}$ (Eq. 3.15)

Mahalanobis distance, $MD_{rs} = \sqrt{(x_r - x_s)V^{-1}(x_r - x_s)'}$ (Eq. 3.16)

where *V* is covariance matrix

Step 3 can be done in different ways through the use of different hierarchical clustering methods such as single linkage, complete linkage, average linkage or Ward's linkage. Their definitions are listed below:

Single linkage

One way to select a cluster is to take the inter-group dis-similarity to be that of the smallest distance between objects in two groups.

$$d(r,s) = \min(dist(x_{ri}, x_{sj})), i \in (i,...,n_r),\ j \in (1,...,n_s) \quad \text{(Eq. 3.17)}$$

Complete linkage

Another way to select a cluster is to take inter-group dis-similarity to be that of the largest distance between objects in two groups.

$$d(r,s) = \max(dist(x_{ri}, x_{sj})), i \in (i,...,n_r),\ j \in (1,...,n_s) \quad \text{(Eq. 3.18)}$$

Average linkage

The other ways to select a cluster is to take inter-group dis-similarity to be that of the

average distance between all pairs of objects in cluster r and cluster s.

$$d(r,s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} dist(x_{ri}, x_{sj}) \qquad \text{(Eq. 3.19)}$$

Ward linkage

The Ward's method utilizes the incremental sum of squares within a cluster, i.e. the increase in the total within-group sum of squares as a result of joining groups r and s.

$$d(r,s) = n_r n_s d_{rs}^2 / (n_r + n_s) \qquad \text{(Eq. 3.20)}$$

Where d_{rs}^2 is the distance between cluster r and cluster s defined in the Centroid linkage. The within-group sum of squares of a cluster is defined as the sum of the squares of the distance between all objects in the cluster and the centroid of the cluster

3.5.2 Principal component analysis (PCA)

PCA, as multivariate exploratory, allows one to study the main sources of variability present in the data sets, to detect cluster formatting and to establish relationships between objects and variables [63, 73]. The samples variations in data with a minimum of latent variables and is given as below:

Principal component analysis (PCA): $X = U\Theta V^t = UP^t = \sum_{i=1}^{A} u_i p_i^t$ (Eq. 3.21)

where $U\Theta V^t$ is the singular value decomposition of the matrix X. The u_i and v_i values (i =1,2,…,A) are so-call score and loading vectors, respectively, and are orthogonal with each other. The diagonal matrix Θ collects the singular values, which are equal to the square root of the variance distributing on every orthogonal principal component axis.

Generally, score matrix (U) gives the relationship of samples while loading matrix (P) shows the importance of each variable (i.e. retention time and their corresponding mass spectra in GC-MS). When a fingerprint with unexpected features that differ from those of majority fingerprint is encountered, it would be diagnosed differently and displayed as outlier in the score plot with the first few principal components involved. Their differences in score plot can be expressed by a measure of distance like Euclidean distance or Mahalanobis distance.

Euclidean distance $ED_A = \sqrt{(x_{test} - \overline{x_A})(x_{test} - \overline{x_A})'}$ for herb A (Eq. 3.22)

$$ED_B = \sqrt{(x_{test} - \overline{x_B})(x_{test} - \overline{x_B})'} \quad \text{for herb B} \qquad \text{(Eq. 3.23)}$$

Mahalanobis distance $MD_A = \sqrt{(x_{test} - \overline{x_A})D_A^{-1}(x_{test} - \overline{x_A})'}$ for herb A (Eq. 3.24)

$$MD_B = \sqrt{(x_{test} - \overline{x_B})D_B^{-1}(x_{test} - \overline{x_B})'} \quad \text{for herb B} \qquad \text{(Eq. 3.25)}$$

where D_A & D_B are the covariance matrix and $\overline{x_A}$ and $\overline{x_B}$ are the mean of the training sets for herb A and B, respectively.

3.5.3 Linear discriminant analysis (LDA)

Linear discriminant analysis (LDA) is a supervised pattern recognition method. It seeks a linear function of the variables in multivariate space which maximizes the ratio between both variances compared to the within-group variance. Since multi-dimensional data arising when the number of variates are larger than that of observation cannot be directly used in LDA, principal component analysis or partial least squares analysis are employed for data compression, to transform the original

data set comprising of a large number of inter-correlated variates into a reduced new set of variates. The LDA builds up a discriminant function for each group, which operates on raw data and constructs a discriminant function for each group according to the equation below [63]:

$$F(G_i) = k_i + \sum_{j=1}^{n} w_{ij} p_{ij} \qquad \text{(Eq. 3.26)}$$

where i is the number of groups (G), k_i is the constant of each group, n is the number of parameters used to classify a set of data into a given group and w_j is the weight coefficient, assigned by LDA to a selected parameters (p_j).

3.5.4 Similarity index (SI)

Comparing the similarity index of different fingerprints is an intuitive method for discrimination analysis. The similarity index between fingerprints is determined by

$$\text{Similarity Index (SI)} = \frac{x^t \cdot \bar{X}}{\text{norm}|x^t| \cdot \text{norm}|\bar{X}|} \qquad \text{(Eq. 3.27)}$$

Here, the symbol "$\cdot$" represents the dot product of two vectors. The larger value of similarity index (SI) between the test fingerprint $\bar{x}$ and the mean vector of training fingerprint sets (X_A, X_B, or more) indicates that these fingerprints have more common features with each other. Then, the test fingerprint $\bar{x}$ can be discriminated based on comparing the magnitudes of the corresponding SI.

3.6 Conclusion

There is an increase in applying chemometrics tools to data analysis in chemistry disciplines. It not only provide alternative way on extracting much more chemical information in the data collected, but also provide effective way to insight, analyze data, that even exist in huge amount. Therefore, their role in data analysis becomes significant, especially in chromatography. In this chapter, a number of chemometrics tools utilized in quality control of CHM have been briefly introduced. They include heuristic evolving latent projection, spectral correlative chromatography, information theory, normalization, local least square method, similarity index, hierarchical cluster analysis, principal component analysis and linear discriminant analysis. Also, their advantages and usage will be illustrated in the following chapters. We believe that chromatography with the support of chemometrics techniques provides the best way to deal with quality control of CHM.

Chapter 4: A comparison of fingerprint analysis methodologies: Development of the chromatographic fingerprint of *Radix Ligustici chuanxiong* by gas chromatography-mass spectrometry and chemometric techniques

4.1 Introduction

Radix Ligustici chuanxiong (CX) was selected and investigated to develop its chromatographic fingerprint for quality control in this study. It is known as Chuan Xiong in China and is one of the most commonly used herbs in CHM prescriptions for menstrual disorder. Its essential oils are believed to be responsible for the therapeutic effects. The ‘Chuan’ in its Chinese name refers to the species being cultivated in the Szechuan province for medicinal purposes and several studies suggested that CX produced in Szechuan has the best medical value [74-76]. In this work, thirty-two samples of CX, from different geographical locations including Szechuan, Yunnan and Guizhou provinces in China were collected for GC-MS study. With the use of several chemometrics techniques including Chemometric Resolution Method (CRM) [68], Local Least Squares (LLS) [57, 61, 62] and Spectral Correlative Chromatography (SCC) [57, 62], which have been successfully applied to the chemical analysis of CHM, the whole chromatograms of all the CX samples have been fully utilized for data analysis without losing any chemical information. All these chemometrics techniques were mentioned in detail in the Chapter 3.

We also present in this work a comparison of the three methodologies used, ‘marker approach’, ‘multi-component approach’ and ‘pattern approach’ for developing the chromatographic fingerprint of *rhizoma chuanxiong* for quality control purpose [1]. The basis of these methods is that the process of data treatment on chromatograms through three different approaches by making use of different chemometrics techniques before using similarity index (SI) and principal component analysis (PCA) for clustering analysis. It is the first work with emphasis on the comparison of the

performance of these three approaches for chemical characterization of CHM with the samples from different geographical sources. The overall procedures developed can be used as a basis for further development on quality control of CHM.

4.2 Background of *Radix Ligustici chuanxiong*

'Chuanxiong', also known as 'Xiongqiong' or 'Huqiong', is the dried rhizome of *Ligusticum Chuanxiong Hort.*, family Umbelliferae. In China, Szechuan, Yunnan and Guizhou province provide sources to the herbal market. This herb is collected in summer when the node of stem becomes obviously swollen and purplish. Then it is removed from soil, baked gently to dryness after sun-drying, and removed from rootlet [25, 77].

For the appearance of Ligusticum Chuanxiong Hort., it is in irregular knotty and fist-like masses with 2~7cm in diameter. Its external is yellowish-brown, rough and shrunken, with many parallel and raised annulations, showing dent, subrounded stem scars on the summit and numerous tuberculous rootlet scars beneath the summit and at the annulations. Its texture is compact, uneasily broken, fracture yellowish-white or grayish-yellow, scattered with yellowish-brown oil cavities. In general, it is packaged in an undulate ring shape for daily life [25, 77].

Radix Ligustici chuanxiong. is pungent in flavour, warm in nature, acting on liver channel, gallbladder channel and pericardium channel. Its pungent flavour is good for dispersing blood stasis and the warm property frees the passage of blood. Its wandering function exerts its effect both on the blood division and the *qi* division,

hence, it is a herb for vital energy in blood with the actions of promoting blood circulation and flow of Qi and relieving pain. The herb is prescribed in headache due to apoplexy, rheumatism, stagnation of blood and vital energy, sores, ulcers, and irregular menstruation [25, 77].

The indications of *Radix Ligustici chuanxiong.* are given as follows [25, 77]:

1. For the treatment of irregular menstruation due to stagnation of Qi and blood, dysmenorrheal, amenorrhea, postpartum tormina, it is usually utilized in combination with other herbs for the functions of promoting the circulation of Qi and blood, regulating menstruation and relieving pain.

2. For the treatment of stagnation of Qi and blood in the Liver channel, chest and abdominal pain, or blockage of the Heart channel, it is often used in combination with other herbs for the action of promoting the circulation of Qi and blood and relieving pain.

3. For the treatment of headache due to the wind-cold, it can be used together other herbs having the action of dispersing cold and relieving pain; For the treatment of headache due to the wind-heat, it usually is worked together with other herbs having functions of dispelling wind and clearing away heat; For treatment of headache due to deficiency of blood, it is usually used in combination with other herbs for the action of nourishing the blood and relieving pain can be used together with it.

CX also shows pharmacological action of dilatating coronary arteries, increasing the blood flow of flow of coronary artery, decreasing myocardial oxygen consumption, increasing cerebral blood flow, and reducing resistance of peripheral blood vessels. It possesses the calming effect on central nervous system while it also inhibits the contraction of the uterus when used in large dose, resulting in cessation of contraction. It has been shown to have anti-cancer action [77, 78].

4.3 Literature review on chemical study of *Radix Ligustici chuanxiong*

4.3.1 Chemical Composition

The chemical compositions of *Radix Ligustici chuanxiong* contain several classes of components including phthalide, phenolic constituents, polyacetylenes and sterols etc [38]. Main constituents of *Radix Ligustici chuanxiong* are shown in Table 4.1 and chemical structures of some components from *Radix Ligustici chuanxiong* are also given in Fig 4.1:

Table 4.1: Main chemical constituents of *Radix Ligustici chuanxiong*

Alkalphthalides	Z-ligustilide, senkyunolide A, neocnidilide, butylphthalide, Z-butylidenephthalide, cnidilide,
Hydroxy alkylphthalide	senkyunolide B – M
Phthalide dimers	wallichilide, levistolide A, tokinolide B
Phenolic constituents	ferulic acid, coniferyl ferulate

Nitrogen containing substances	Ligustrazine, perlolyrine
Polyacetylenes	falcarindiol
Quinones	sekyunone
Steroles	pregneolone

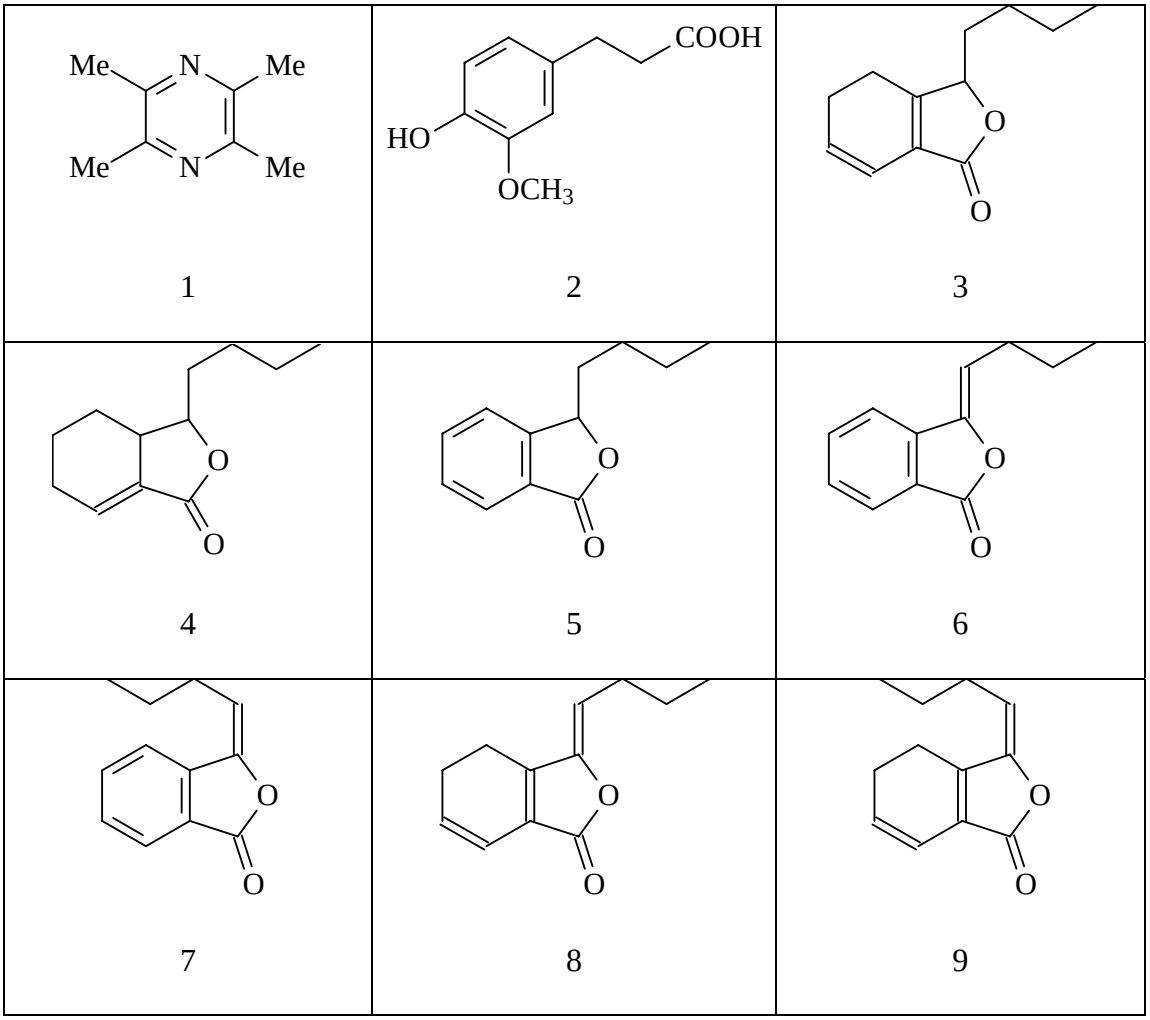


Fig 4.1 Chemical structures of the identified compounds in *Radix Ligustici chuanxiong*: 1) Ligustrazine; 2) Ferulic acid; 3) Senkyunolide A; 4) Neocnidilide; 5) Butylphthalide 6) Z-butylidenephthalide; 7) E-butylidenephthalide; 8) Z-ligustilide; 9) E-ligustilide

4.3.2 Marker components

The most important marker for *Radix Ligustici chuanxiong* is tetramethylpyrazine and it has been reported to be an effective component that affects heart, brain and vessel diseases. Tetramethylpyrazine, which is named as ligustrazine or chuanxiongzine, is nitrogen-containing compound and possesses anti-ulcers and anti-platelet activities. However, the concentration of ligustrazine is relatively low and not easy to be detected and isolated. In addition to the ligustrazine, some phthalide substances and organic acid were identified and served as commercial marker of *Radix Ligustici chuanxiong*. These include Z-ligustilide, L-ligustilide, butylphthalide, Z-butylidenephthalide, E-butylidenephthalide and ferulic acid. They have been reported to have certain pharmacological actions. Nevertheless, they are also presented in another famous herb, *Radix Angelicae Sinensis* and cannot act as unique markers for *Radix Ligustici chuanxiong* [38, 78].

4.3.3 Pharmaceutical value of the marker components

Ligustrazine is a major representative being extensively studied on its pharmacological action. It inhibits platelet aggregation and might be able to displace calcium ion from platelet membranes so it prevents arterial thrombus formation, probably by inhibiting platelet aggregation. Besides, the phthalides components from *Radix Ligustici chuanxiong*, butylphthalide, butylidene phthalide and ligustilide, also show anti-asthmatic and spasmolytic on tracheal muscle, exhibit antiarrhythmic effect, and dilatating activity on coronary arteries. In addition, ferulic acid also inhibits the platelet aggregation and the serotonin release from thrombocytes [37, 38, 78].

4.3.4 Chemical studies of *Radix Ligustici chuanxiong*

Different chromatographic and spectroscopic techniques have been reported for characterizing the chemical compositions of *Radix Ligustici chuanxiong*. Several chromatographic studies on phthalide components have been reported under HPLC-DAD / MS [38, 79-83], HPLC-UV / LC-MS [38, 83-86]. In addition, lot of analytical studies has been explored by using GC-MS for chemical compositions of volatile oil in *Radix Ligustici chuanxiong* [83, 87-91]. Furthermore, H.B. Li et al. isolated chuanxiongzine using high speed counter current chromatography [92] while Y. Cao et al. developed HPLC-UV method for finding out chuanxiongzine [93].

4.4 Methods of Investigation

The main feature in this work is to minimize the chemical information lost in the whole set of chromatograms by the use of different chemometrics techniques during data processing. Then, similarity index or principal component analysis was employed for quality assessment, depending on what kinds of purpose(s) or objective(s) to be achieved. In this work, the goal is to characterize CHM chemical compositions via comparing samples from different geographical sources.

Qualitative determination of all components in the chromatogram is achieved by CRM, where all interested peaks (markers), highly-overlapped peaks and shoulder peaks were resolved out. For CRM technique, the following symbols are used: The data obtained from a GC-MS analysis are represented by a two dimensional matrix $A_{m,n}$ with m chromatographic scan points (retention time) and at each scan point links with

a mass spectrum of the range 1 to n mass to charge ratio (m/e). Assuming there are N components in the sample under study, the GC-MS data can be expressed as

$$A_{m,n} = CS^t + E$$

$$= \sum_{i=1}^{N} c_i s_i^t + E \qquad \text{(Eq. 4.1)}$$

where c_i and s_i are column vectors representing the chromatographic profiles and the spectra of the i-th component, respectively. E represents the instrumental noise while the superscript t denotes the transpose of a matrix. Through CRM, the unique resolution of a two-dimensional data gives peaks and spectra of the pure chemical constituents. A well known method of this type is heuristic evolving latent projection (HELP) [58, 59]. By determining the information obtained like the number of components, the eluting order of the components, the selective regions, and zero concentration regions of all the constituents in the certainly overlapped and shoulder peaks within chromatograms, the pure profile and pure spectrum of each component has been resolved and obtained.

Afterward, based on the result obtained by CRM, Spectral Correlative Chromatography (SCC) provides an efficient way to identify all the common constituents among all the samples through the treatment, which is the score between two single vectors (mass spectra). One is the resolved mass spectra from the first chromatogram while the other is the mass spectra from one scan point of the other chromatograms. After the calculations, common peaks are digged out, based on comparison of the scores determined by SCC of each resolved component. Similar computations are performed with all GC-MS data repeatedly.

Next, the extracted data obtained from the combination use of CRM and SCC serve two purposes. They are not only used either in the marker approach or multi-components approach directly, but also in chromatographic alignment. Local Least Squares (LLS) method [57, 61] is utilized to correct the retention time shift found in chromatograms so as to achieve the multi-point alignment before comparing the similarity or dissimilarity between different chromatographic profiles. It is necessary to correct the retention shift present in all chromatograms because it causes a serious problem in data analysis under pattern approach. To match all chromatographic profiles with their retention times, all common constituents were selected and utilized for chromatographic alignment. Several parameters among them have been estimated using LLS method and the new-aligned chromatographic profiles can be re-constructed through the cubic spline data linear interpolation techniques.

As a final step, several input parameters, including the concentrations of the three selected markers, the concentrations of the common constituents and their chromatographic profile from different sources were arranged as a matrix, $x=[\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_{i-1}, \mathbf{x}_i]$, based on those approaches mentioned previously. Similarity index and principal component analysis are computed using the formula mentioned previously.

4.5 Experimental

4.5.1 Herbal samples

Thirty-two CX samples were obtained from China Academy of Traditional Chinese Medicine, Beijing, P.R. China in 2001-2002 with authenticated certificates provided.

The chuanxiong samples were mainly collected from Szechwan Province (genuine type), the recommended cultivation region for this herb. The geographical sources with their collection periods among all the CX samples are summarized in Table 4.2.

Table 4.2: CX samples collected in this study

No.	Sample codes	Geographical sources	Collection period
1	CX1	Pengzhou, Szechuan, China	April, 2001
2	CX2	Xindou, Szechuan, China	April, 2001
3	CX3	Xindou, Szechuan, China	April, 2001
4	CX4	Pengzhou, Szechuan, China	April, 2001
5	CX5	Xindou, Szechuan, China	April, 2001
6	CX6	Pengzhou, Szechuan, China	April, 2001
7	CX7	Doujiang, Szechuan, China	January, 2001
8	CX8	Doujiang, Szechuan, China	January, 2001
9	CX9	Doujiang, Szechuan, China	January, 2001
10	CX10	Doujiang, Szechuan, China	January, 2001
11	CX11	Doujiang, Szechuan, China	January, 2001
12	CX12	Chougzhou, Szechuan, China	February, 2001
13	CX13	Chougzhou, Szechuan, China	February, 2001
14	CX14	Chougzhou, Szechuan, China	February, 2001
15	CX15	Chougzhou, Szechuan, China	February, 2001
16	CX16	Chougzhou, Szechuan, China	February, 2001
17	CX17	Daili, Yunnan, China	January, 2001
18	CX18	Zunyi, Guizhou, China	April, 2001
19	CX19	Daxian, Szechuan, China	April, 2001
20	CX20	Danxian, Szechuan, China	April, 2001
21	CX21	Danxian, Szechuan, China	March, 2001
22	CX22	Pengzhou, Szechuan, China	March, 2001
23	CX23	Doujiang, Szechuan, China	April, 2001
24	CX24	Xindou, Szechuan, China	April, 2001
25	CX25	Pengzhou, Szechuan, China	April, 2001

26	CX26	Danxian, Szechuan, China	April, 2001
27	CX27	Pengzhou, Szechuan, China	April, 2001
28	CX28	Pengzhou, Szechuan, China	April, 2001
29	CX29	Pengzhou, Szechuan, China	April, 2001
30	CX30	Xindou, Szechuan, China	April, 2001
31	CX31	Panxian, Guizhou, China	April, 2001
32	CX32	Panxian, Guizhou, China	April, 2001

4.5.2 Chemical and reagents

The three chemical standards, z-ligustilide, z-butylidenephthalide and ferulic acid were purchased from ChromaDex, Inc. (UAS), Wako Pure Chemical Industries, Ltd. (Japan), and Lancaster Synthesis (England), respectively. Solvents for GC-MS experimentation were of analytical grade. Double deionized water was used in the sample preparation.

4.5.3 Apparatus and chromatographic conditions

GC-MS experiments were performed on all the CX extracts using Agilent 6890 coupled with Agilent 5973 series mass selective detector equipped with a fused silica capillary column HP-5MS (5% phenyl methyl siloxane, 0.25mm × 30m × 0.25μ m). The column temperature was maintained at 40℃ for 0.5 minute after injection, then increased at 3℃ min^{-1} to 175℃ and 5℃ min^{-1} to 220℃; maintained at this temperature for 3 minutes. Split injection was conducted with a split ratio of 1:100 and helium was utilized as the carrier gas with 1.0ml min^{-1} flow rate. The spectrometer was operated in the electron-impart (EI) mode with the scan range 45-500amu, the ionization energy

70eV and the scan rate 1.66scans per second. The inlet and ionization source temperature were 250℃ and 280℃, respectively.

4.5.4 Sample preparation

4.5.4.1 Sample pre-treatment

CX sample were ground into powdered and passed through 100 mesh stainless sieve. Then the sieved powder was stored in polypropylene containers at 20°C.

4.5.4.2 Sample preparation

Essential oils of the CX samples were extracted by the standard method recommended by the Pharmacopoeia of the People's Republic of China [25]. About 80g dried herbs and 500mL double-deionized water were pre-mixed into 1000mL round bottom flask. The solution was shaken and mixed well. Then the apparatus was set up and shown in Fig 4.2. The flask was heated gently, and continued heating for about 5 hours after boiling begun, until the volume of oil did not increase anymore. The heater was turned off and the setup was allowed to stand for cooling. After cooling, the oily layer was centrifuged and the water layer was run off. Then the volume of oil in the graduate tube was read and recorded. The volatile oil contents of Chuanxiong ranged from 0.4% to 1.2% by weight in the investigation. In the GC-MS study, samples were firstly diluted 100 times by hexane and aliquots of 1μl were injected.

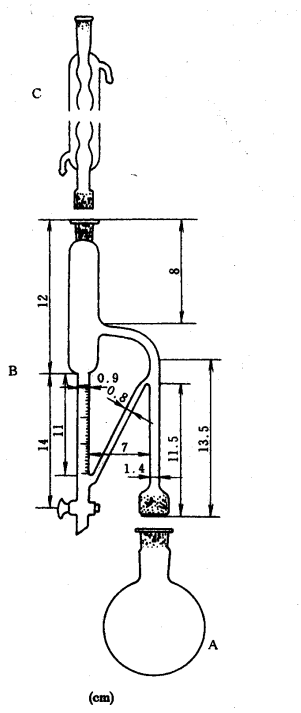


Figure 4.2 The set up to extract volatile oil (The State Pharmacopoeia Commission of P.R. China, 2005).

4.5.5 Data analysis

All Data analysis were performed on a Pentium IV personal computer. All the calculations and chemometrics data treatments were carried out using MATLAB 6.5. The National Institute of Standards and Technology (NIST147) MS database, which contains about 107000 compounds and the Wiley138 library were used for matching the resolved pure spectra to identify the chemical constituents.

4.6 Results and Discussion

4.6.1 Qualitative analysis of CX by the application of CRM

Fig. 4.3 depicts the TIC chromatograms of the essential oil of 32 chuanxiong (CX) samples. As can be seen from the chromatograms, there are a lot of peaks and their contents do not vary, quite significantly, from sample to sample. Although the experimental condition for the chromatographic separation was optimized, overlapping peaks still existed. Without any data processing, this may lead to incorrect identification of the components present. In addition, components with low contents are also difficult to identify as the background spectra became more significant. Certainly, qualitative analysis of components should be more reliable if overlapped peaks and peaks of those components with low content are resolved into separated chromatographic profiles and pure spectra. CRM is a fast and effective tool to give more reliable and accurate identification of components present.

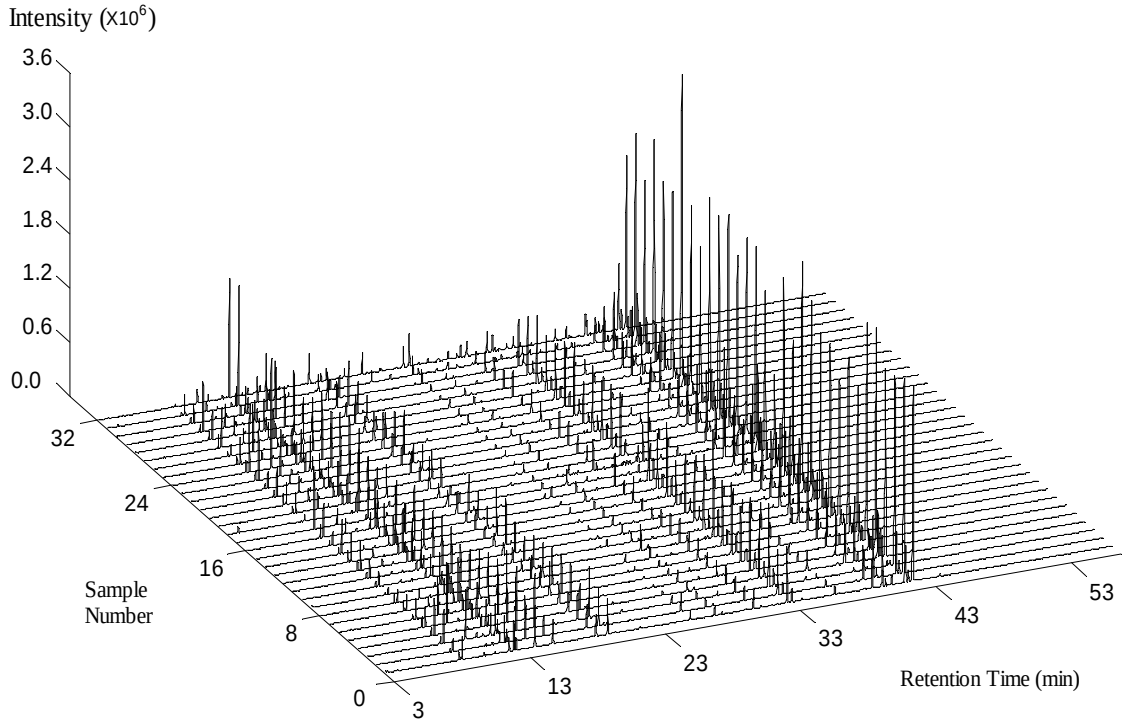


Fig. 4.3 GC-MS Chromatograms of 32 samples of CX

Background subtraction is conventional performed subjectively by subtracting a

manually selected background spectrum at one retention point throughout the whole data set. As for CRM, with the use of locating the zero-component regions and local rank analysis, the information available is sufficient for correcting the background and drifting accurately. After a much better background subtraction was done, the resolution of the overlapping peaks become possible. Now the peak cluster I of sample CX1 within 18.65 – 19.05 min (see Fig. 4.4) is illustrated as demonstration.

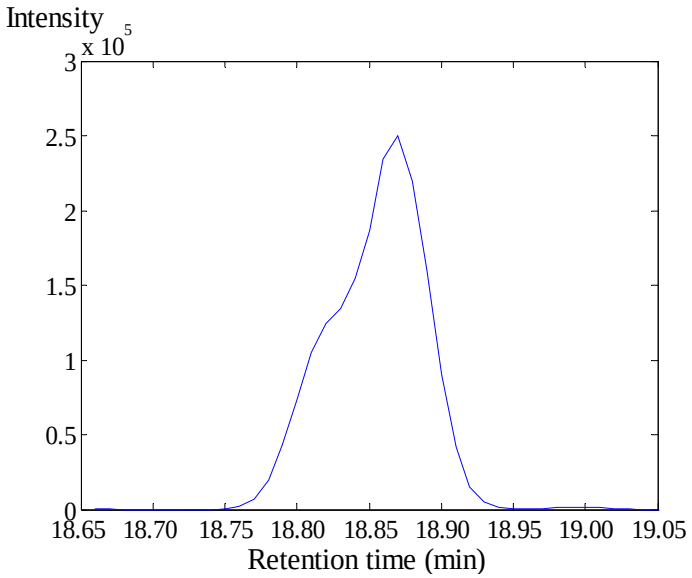


Fig. 4.4 TIC curve for peak cluster I (within 18.65 – 19.05 min) of sample CX1

Peak purity can be found out by a FSWMEFA or so called eigenstructure tracking analysis. In the fixed size window method (FSWM) plot, the noise level is characterized by eigenvalue curves which have similar values and appears together at the bottom. Eigenvalue curves higher than the noise level represent the existence of new components. If a peak contains only one component, there is only one eigenvalue curve higher than the noise level in its FSWM plot. From the FSWM plot peak cluster I as shown in Fig. 4.5, there are two eigenvalue curves higher than the noise level within the peak region. In this situation, one may conclude that there exists two components where the region of 1 + 2 is the overlapping region of the first and second

components while the regions 1 and 2 are possibly the regions of the pure first and second components, respectively.

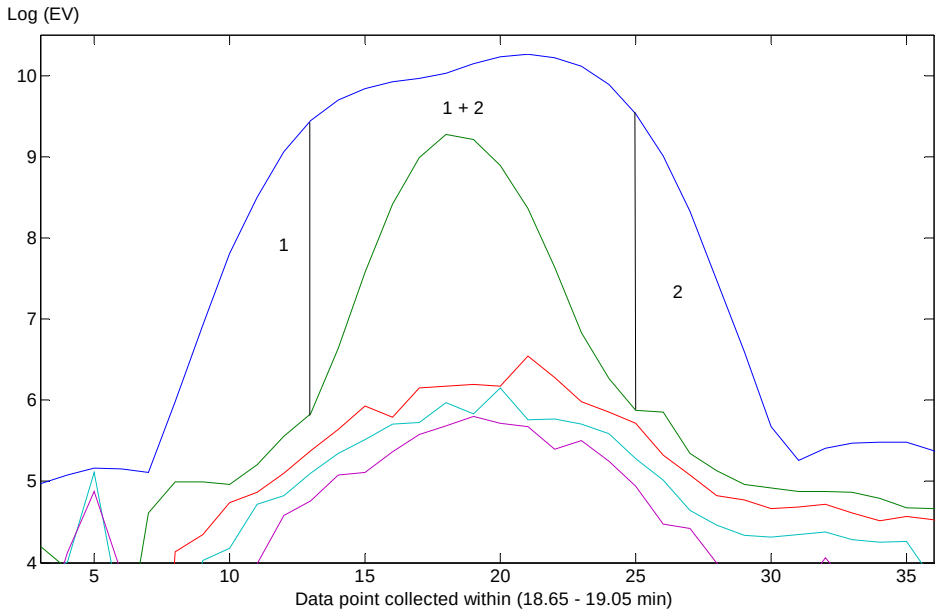


Fig. 4.5 FSWM plot for peak cluster I after correcting background. The regions 1, 1 + 2, and 2 represent the regions of the pure component 1, the overlapping region of components 1 and 2 and the region of the pure component 2, respectively.

The stepwise eluting information of chemical components in peak cluster studied can further be confirmed by evolving latent projection graph (ELPG). The ELPG is essentially a principal component projective curve from chromatographic or spectral space. Several advantages using ELPG are listed as below [58, 59]:

1. A straight line segment pointing to the origin suggests selective information in the chromatographic direction, spectral direction under a bivariate score plot and loading plot, respectively.

2. The evolving in and out information of the chemical components in retention time direction can be provided in ELPG. In the ELPG from the chromatographic space,

the straight line section represents the pure selective region of one component while the curving section indicates the overlapping region of at least two components.

3. Shifts of the chromatographic baseline and instrumental background are able to be detected under ELPG.

4. ELPG is a good diagnostic tool to insight the structure of two-way data, i.e. the embedded peaks in chromatogram can be identified.

Fig. 4.6 shows the ELPG of peak cluster I (see Fig 4.4). It indicates that this peak cluster is a two-component system. The marks, a, b and c in Fig. 4.6 represent respectively the pure regions of the first component, the overlapping region of the first and second components and the pure region of the second component in the chromatographic direction. This is consistent with the previous result obtained from the FSWM.

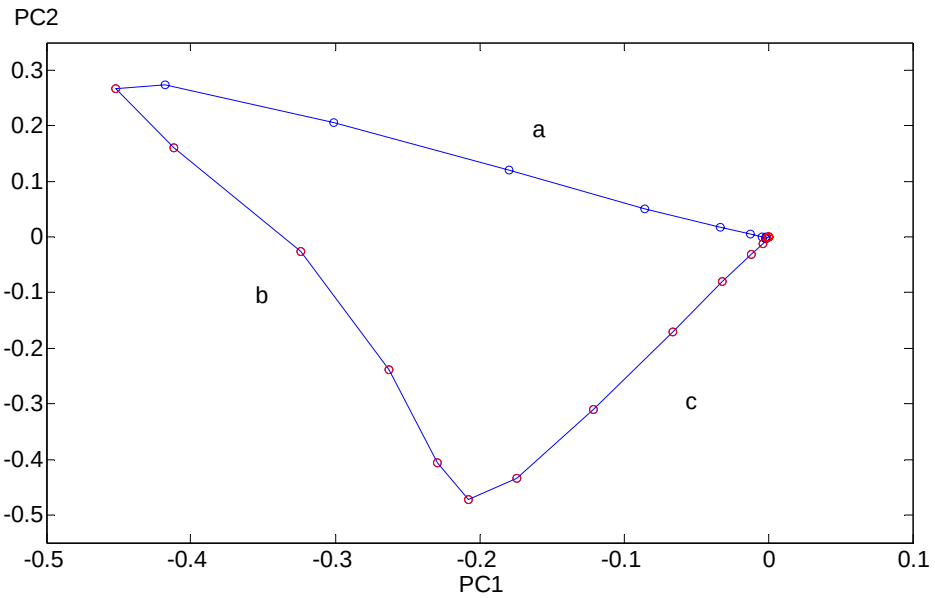


Fig. 4.6 Evolving latent projection graph (ELPG) for peak cluster I. The regions a, b and

c represent the regions of the pure component 1, the overlapping region of components 1 and 2 and the region of the pure component 2, respectively.

From the above results, the chromatographic eluting order can be determined and so the number of components in the system, the selective regions and zero-concentrations regions of all the constituents. With all the information determined, the two-dimensional data matrix can be uniquely resolved into pure chromatographic profiles and mass spectra of all components.

Since the pure chromatographic profile and mass spectrum of each component have been resolved, the qualitative analysis can be directly performed by similarity searches in the MS library. The result shows that these two components in peak cluster I were pentyl benzene and 6-butyl-1,4-cycloheptadiene. Their chromatographic profiles are shown in Fig. 4.7 and the resolved mass spectra together with the standard spectrum of each component from the NIST MS library are also depicted in Fig. 4.8 and 4.9.

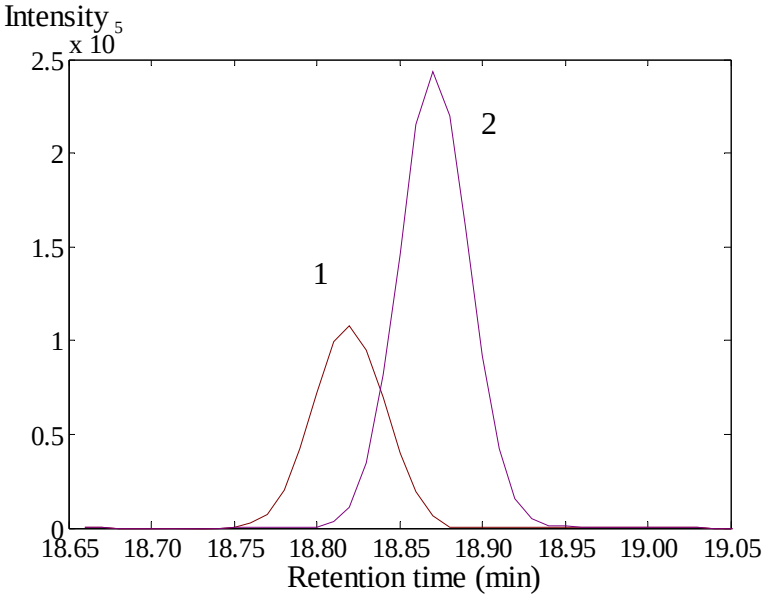


Fig. 4.7 Resolved chromatograms for peak cluster I (see Fig 4.4) containing component 1 and 2.

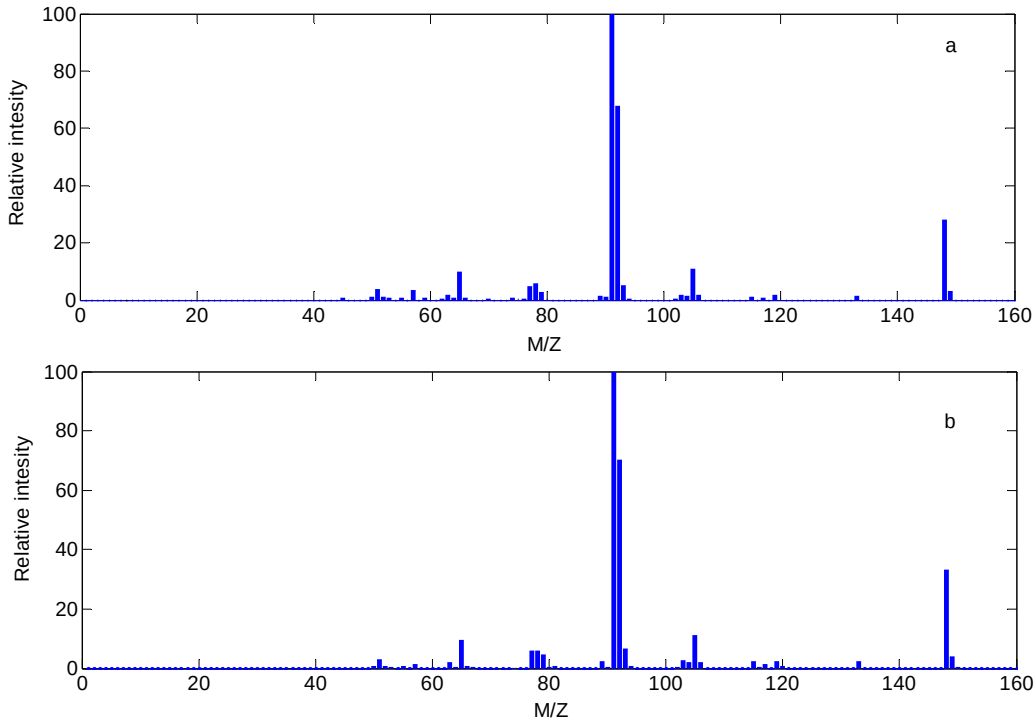


Fig. 4.8 Resolved mass spectrum of component 1(a) and standard mass spectrum of pentyl benzene (b)

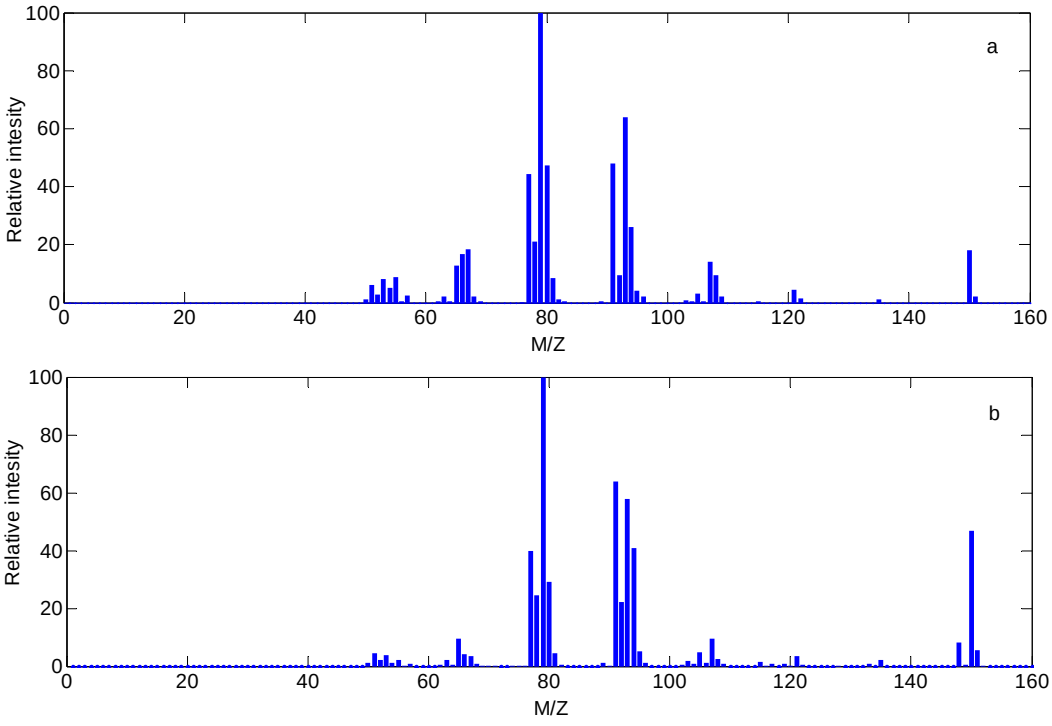


Fig. 4.9 Resolved mass spectrum of component 2 (a) and standard mass spectrum of 6-butyl-1,4-cycloheptadiene (b)

From theses two figures, it is easily indicated that the resolved results are quite appreciated and worth noted that the use of CRM to work with is strongly determined by the reproducibility of the intensities in the MS. In similar way, all the overlapped peak clusters of the CX1 chromatogram were resolved with the peak profiles and spectra of all components found. GC-MS chromatograms of other CX samples were also treated in the same way if required.

4.6.2 Marker approach

z-ligustilide, z-butylidenephthalide and ferulic acid are widely accepted standards for *Radix Ligustici chuanxiong* and they are available in the market for quality control purpose. Before quantitative determination of these standards in all the CX extracts, their stability was one of the major technical concerns. Previous studies shown that z-ligustilide is readily isomerized in air and/or in some solvent systems [79, 82, 83]. Consequently, all CX extracts were stored under -20^{o}C refrigerator without light in this study and stability of z-ligustilide in such storage condition was investigated by two independently tests. First, the amount of z-ligustilide in the CX extract was determined, after extraction. Another test was performed on the same CX extract after two day of storage. Our results show that the level of z-ligustilide was found to be nearly the same in both two independent tests. This says that z-ligustilide was relatively stable under our storage condition and our results of the chromatographic analysis should be reliable.

To quantify the amount of z-ligustilide, z-butylidenephthalide and ferulic acid within

all the CX samples, the external standard method was employed for setting up the calibration curves in this work. Their linearity was determined in the concentration ranges of 2-1000mg/L, 0.1-1000mg/L and 2-100mg/L, with six different concentrations of standard solutions for each one, respectively while their corresponding regression coefficients were found to be 0.9993, 0.9989 and 0.9986. The regression equations obtained were utilized for quantifying their contents in all the CX samples and Table 4.3 lists out the amounts of each standard within all samples. Moreover, method reproducibility was also examined by four successive injections of both samples and standard solutions. Precision of replicated injections were determined and the relative standard deviations of z-ligustilide, z-butylidenephthalide and ferulic acid in all samples and standards were less than 5%.

The mean values of the absolute contents of the three markers, Z-ligustilide, Z-butylidenephthalide and ferulic acid of all the samples, were 19.05 ± 5.17 mg/g, 5.95 ± 2.44 mg/g, 2.06 ± 1.09 mg/g respectively. These values are closed to or even higher than the literature value reported in previous HPLC studies [85, 94]. Hence, this indicates that the quality of our 32 CX samples is consistent with the samples reported earlier by other workers and the results of this work reflect the generality feature of CX.

Table 4.3: The absolute contents of Z-ligustilide, Z-butylidenephthalide and ferulic acid in the 32 CX samples

Sample	Content (mg/g)[a]		
	Z-ligustilide	Z-butylidenephthalide	Ferulic acid
CX1	21.52 ± 0.56	5.85 ± 0.18	2.69 ± 0.13
CX2	20.60 ± 0.68	7.10 ± 0.21	5.81 ± 0.27

CX3	17.17 ± 0.60	1.99 ± 0.80	2.27 ± 0.06
CX4	19.78 ± 0.91	7.64 ± 0.33	1.57 ± 0.07
CX5	23.18 ± 0.54	3.03 ± 0.08	0.77 ± 0.02
CX6	27.75 ± 0.10	3.67 ± 0.01	1.15 ± 0.02
CX7	15.13 ± 0.38	6.78 ± 0.20	1.57 ± 0.03
CX8	17.01 ± 0.46	6.60 ± 0.19	2.94 ± 0.10
CX9	11.92 ± 0.50	8.84 ± 0.42	0.61 ± 0.02
CX10	17.45 ± 0.42	9.09 ± 0.23	2.67 ± 0.08
CX11	16.83 ± 0.58	10.39 ± 0.34	0.91 ± 0.03
CX12	18.25 ± 0.49	7.93 ± 0.19	1.26 ± 0.04
CX13	24.45 ± 0.72	6.42 ± 0.19	0.94 ± 0.04
CX14	11.82 ± 0.22	6.09 ± 0.13	1.83 ± 0.06
CX15	21.04 ± 0.29	10.85 ± 0.17	1.54 ± 0.04
CX16	7.54 ± 0.25	5.12 ± 0.18	3.33 ± 0.11
CX17	14.30 ± 0.19	6.36 ± 0.08	2.10 ± 0.08
CX18	17.91 ± 0.28	3.79 ± 0.07	1.75 ± 0.03
CX19	20.38 ± 0.78	4.80 ± 0.18	2.70 ± 0.07
CX20	18.44 ± 0.59	7.42 ± 0.24	0.97 ± 0.03
CX21	21.73 ± 0.20	6.66 ± 0.09	1.94 ± 0.05
CX22	24.90 ± 0.05	7.68 ± 0.15	1.70 ± 0.05
CX23	20.24 ± 0.43	3.28 ± 0.08	1.58 ± 0.06
CX24	12.32 ± 0.30	1.47 ± 0.05	1.50 ± 0.02
CX25	18.10 ± 0.59	6.46 ± 0.23	2.55 ± 0.12
CX26	34.12 ± 0.74	2.43 ± 0.06	2.12 ± 0.02
CX27	18.10 ± 0.06	5.17 ± 0.18	2.77 ± 0.07
CX28	18.30 ± 0.36	6.64 ± 0.16	1.52 ± 0.07
CX29	24.20 ± 0.72	1.87 ± 0.06	2.40 ± 0.11
CX30	17.18 ± 0.69	7.00 ± 0.30	4.28 ± 0.14
CX31	18.40 ± 0.84	4.88 ± 0.15	2.30 ± 0.03
CX32	11.20 ± 0.26	4.38 ± 0.60	1.21 ± 0.03

a The value is mean ± S.D. (n=4)

4.6.3 Multi-component approach

It can be seen in Fig. 4.3 that the thirty-two chromatographic profiles were rather consistent with one another with only minor variation on the relative contents of some individual components. It would be time-consuming and unimaginable to extract the pure component spectra for all chromatograms without making use of the results deduced from the previous analysis, in this case, say that sample CX1. Therefore, SCC provides another effective tool to determine and compare the common constituents among all samples by using of the CRM result (pure spectra) obtained above for matching. An example is given below to demonstrate how it works.

Fig 4.10a depicts the total ionic current (TIC) chromatograms of CX1 and CX2 in which components U in CX1 was selected for SCC demonstration while Fig. 4.10b shows the spectral correlative curve of CX2 with respect to component U in CX1. As it can be seen, the largest value of 0.9985 of the correlation coefficients r (U,U') occurs at retention time 25.70min, which corresponds to component U' in CX2. Based on the fact, the same chemical component should have the same spectrum no matter when they eluted out through diverse chromatographic system. Together with the excellent agreement in retention time, it is almost certain that components U' and U are identical. The mass spectra of component U and U' are given in Fig. 4.10c and it shows the success in SCC application.

By using CRM together with SCC, 35 common components (see Table 4.4) among the 32 CX samples have been found and 33 of them were tentatively identified by matching spectra in the NIST mass library and mass spectra published previously. The

mass spectra data of the resolved components, which match with previous literature data [83, 95], are listed in the Table 4.5. It should be noted that the newly identified components provides more choice of markers for CX in future analysis.

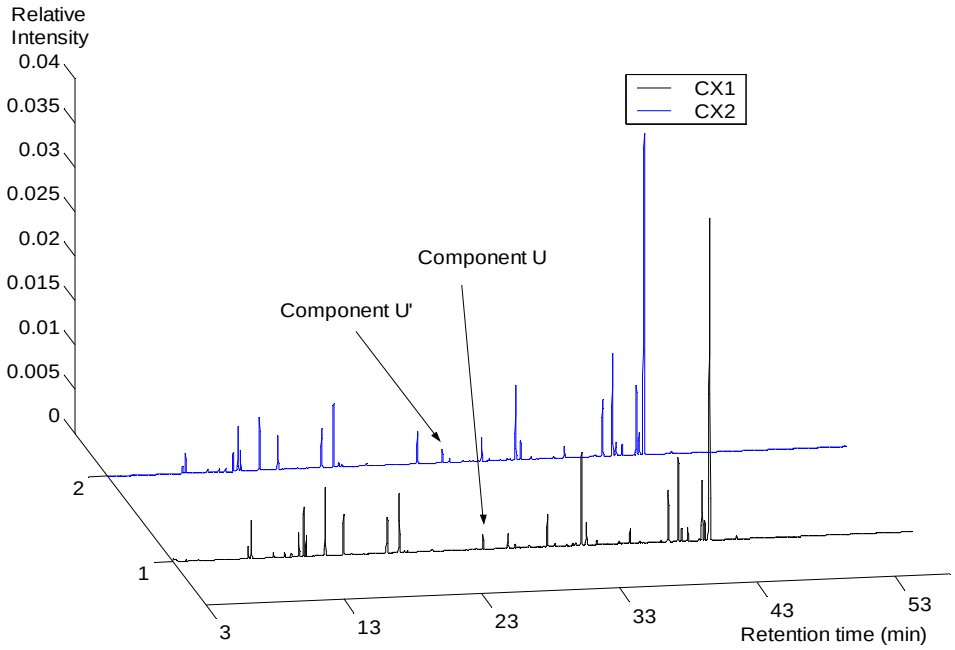


Fig.4.10a Total ionic current chromatograms of CX1 and CX2 obtained from GC-MS, where components U and U' are the correlative compounds as identified by SCC

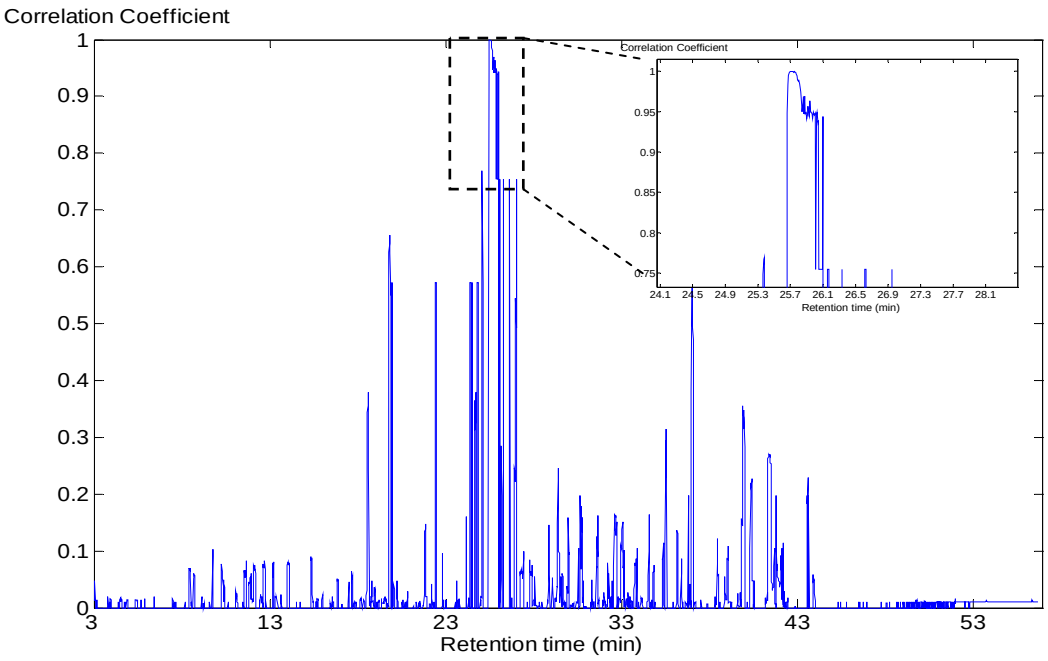


Fig. 4.10b Correlative chromatogram of component U of CX1 against that of CX2. Top right is the amplified profile where compound U' exists in CX 2. The correlation

coefficient of component U and U' is r(U,U')

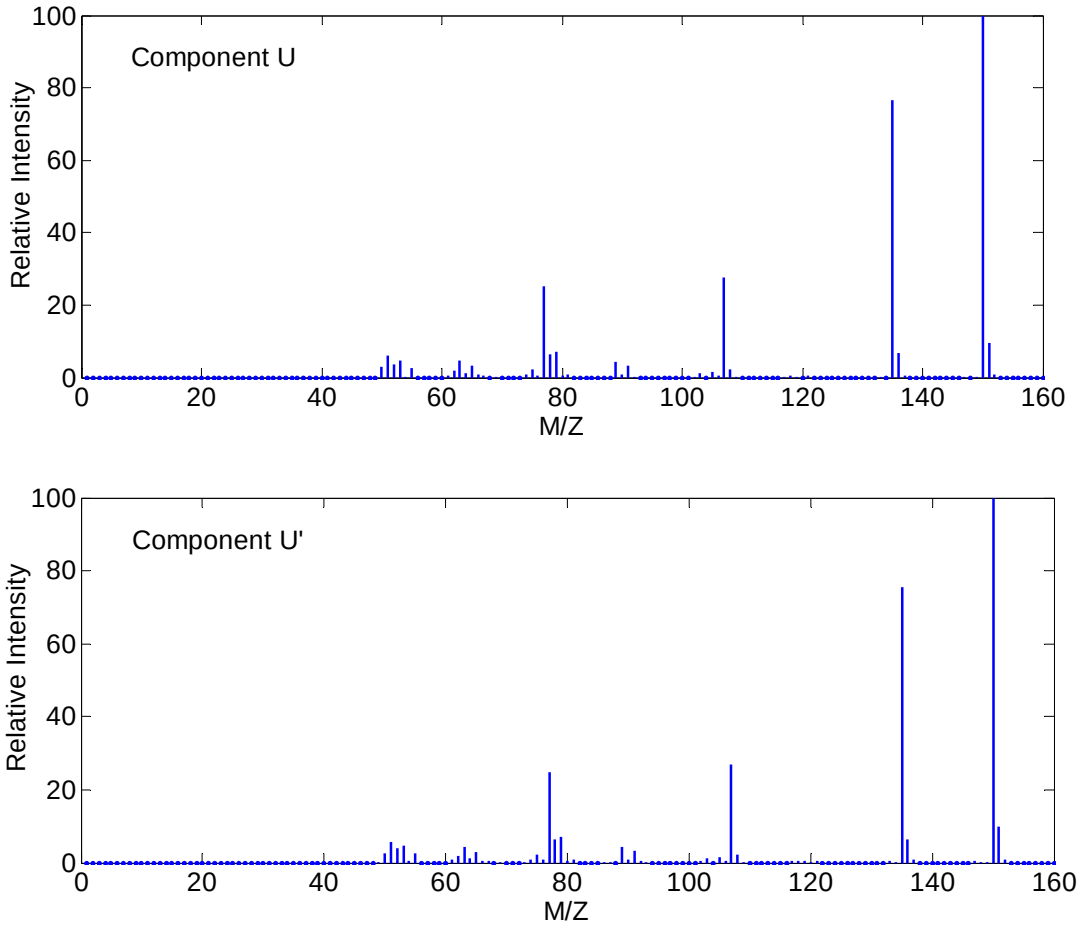


Fig. 4.10c Mass spectrum of component U and component U'

Fig. 4.10a) Total ionic current chromatograms of CX1 and CX2, b) Correlative chromatogram of component U of CX1 against that of CX2, & c) Mass spectrum of component U and component U'

Table 4.4: Identification and quantification of common constituents in the essential oil among the 32 CX samples investigated

Series No.	Retention Time (min)	Compound Name	Molecular structure	Content range (%)
1	8.600	Alpha-thujene	$C_{10}H_{16}$	0.06% - 1.05% (0.06% – 1.28%)

2	8.863	Alpha-pinene	$C_{10}H_{16}$	0.22% - 5.04% (0.22% - 5.61%)
3	10.485	Sabinene	$C_{10}H_{16}$	0.05% - 9.44% (0.05% - 11.2%)
4	10.586	beta-pinene	$C_{10}H_{16}$	0.03% - 1.46% (0.03% - 1.63%)
5	11.306	beta-myrcene	$C_{10}H_{16}$	0.09% - 0.62% (0.10% - 0.63%)
6	11.793	Alpha-phellandrene	$C_{10}H_{16}$	0.04% - 0.20% (0.04% - 0.23%)
7	12.056	Benzene 1,4-dichloro	$C_6H_4Cl_2$	0.09% - 20.61%
8	12.360	Alpha-terpinene	$C_{10}H_{16}$	0.15% - 2.11% (0.17% - 2.41%)
9	12.705	Benzene 1-methyl-(1-methylethyl)	$C_{10}H_{14}$	0.17% - 4.42% (0.17% - 5.41%)
10	12.887	Cyclohexene, 1-methyl-5(1-methylethenyl)	$C_{10}H_{16}$	0.51% - 11.09% (0.86% – 11.14%)
11	14.286	Gamma-Terpinene	$C_{10}H_{16}$	0.45% - 6.22% (0.51% - 7.37%)
12	15.635	Alpha-terpinolene	$C_{10}H_{16}$	1.00% - 5.37% (1.11% - 6.33%)
13	18.807	Benzene pentyl	$C_{11}H_{16}$	0.36% -1.38% (0.42% -1.64%)
14	18.858	6-butyl-1,4-cycloheptadiene	$C_{11}H_{18}$	0.44% - 3.09% (0.78% - 3.86%)
15	19.730	4-Terpineol	$C_{10}H_{18}O$	0.82% - 5.27% (0.82% - 6.07%)
16	20.125	p-cymen-8-ol	$C_{10}H_{14}O$	0.07% - 1.25% (0.08% - 1.25%)
17	20.368	Alpha-terpineol	$C_{10}H_{18}O$	0.03% - 0.16% (0.03% - 0.18%)
18	25.903	ferulic acid	$C_{10}H_{10}O_4$	0.03% - 2.22% (0.03% - 2.45%)

19	27.718	1-Pentanone 1-phenyl	$C_{11}H_{14}O$	0.32% - 2.74% (0.32% - 3.04%)
20	30.404	Trans-caryophyllene	$C_{15}H_{24}$	0.02% - 3.24% (0.05% - 3.48%)
21	32.087	beta-farnesene	$C_{15}H_{24}$	0.01% - 0.94% (0.01% - 1.01%)
22	32.725	gamma-selinene	$C_{15}H_{24}$	0.02% - 0.67% (0.03% - 0.67%)
23	33.131	beta-selinene	$C_{15}H_{24}$	0.23% - 12.23% (0.23% - 13.44%)
24	33.516	Alpha-selinene	$C_{15}H_{24}$	0.03% - 3.17% (0.03% - 3.51%)
25	34.267	gamma-cadinene	$C_{15}H_{24}$	0.04% - 0.49% (0.04% - 0.54%)
26	36.719	Spathulenol	$C_{15}H_{24}O$	0.04% - 2.89% (0.04% - 3.2%)
27	39.531	3-butylphthalide	$C_{12}H_{14}O_2$	0.52% - 13.60% (0.6% - 15.13%)
28	40.277	(Z)-3-butylidenephthalide	$C_{12}H_{12}O_2$	1.77% - 10.52% (2.09% - 12.33%)
29*	40.511	Unknown 1	MW = 192	0.19% - 2.43% (0.36% - 2.71%)
30*	40.977	Unknown 2	MW = 150	0.03% - 2.92% (0.03% - 2.93%)
31	41.849	(E)-3-butylidenephthalide	$C_{12}H_{12}O_2$	0.01% - 0.57% (0.01% – 0.58%)
32	42.021	Senukyunolide A	$C_{12}H_{16}O_2$	0.46% - 12.37% (0.5% - 13.76%)
33	42.203	Neocuidilide	$C_{12}H_{18}O_2$	0.45% - 4.60% (0.53% - 5.12%)
34	42.588	Z-Ligustilide	$C_{12}H_{14}O_2$	10.88% - 52.71% (13.32% – 56.17%)
35	44.586	E-Ligustilide	$C_{12}H_{14}O_2$	0.05% - 0.62%

				(0.06% - 0.62%)

* represents that the resolved pure component has low matching value with library searching

Bracket value represents the relative content value without counting DCB area.

Table 4.5: A summary of the mass spectral data of the compounds obtained in this study work with these data identified by comparing with literature data

Peak No.	Compound	Retention time (min)	Mass data
1	Ferulic acid	25.90	150(100), 135(76), 107(27), 77(25), 51(6)
2	3-butylphthalide	39.51	190(M^+,3), 134(11), 133(100), 105(25), 77(10)
3	Z-Butylidenephthalide	40.28	188(M^+,20), 160(13), 159(100), 146(32), 131(19), 103(18), 77(13)
4	E-Butylidenephthalide	41.85	188(M^+,20), 159(100), 146(33), 131(21), 104(13), 103(19), 77(12)
5	Senkyunolide A	42.02	192(M^+,24), 163(2), 135(5), 107(100), 79(22)
6	Neocnidilide	42.20	194(M^+,1), 137(4), 108(100), 79(29)
7	Z-Ligustilide	42.59	190(M^+,64), 161(100), 148(80), 134(16), 106(36), 105(48), 77(26), 55(38)
8	E-Ligustilide	44.59	190(M^+,64), 161(100), 148(73), 134(15), 106(38), 105(55), 77(32), 55(44)

Common peaks are one of the important indices for comparing the constituents from different samples and Table 4.4 lists the common components of volatile oil among 32 CX samples found. These common constituents included mono-terpene, di-terpene and phthalide-type compounds. Phthalide-type compounds of (Z)-ligustilide, (Z)-3-butylidenephthalide, neocuidilide and senkyunolide A were the major volatile

components existing in most CX samples and Z-ligustilide was the major constituent among them (27.78 ± 7.13%). The mean relative content of the total phthalide-type compounds in all samples was 45.75 ± 8.82% and 80% of CX samples contained more than 40% of total phthalide-type content. On the other hand, by looking into the relative contents of the three commonly used markers, Z-ligustilide, (Z)-3-butylidenephthalide and ferulic acid in all samples, the variations were quite significant. This may be attributed to different factors such as geographical location of cultivation, the nutrition level of the soil, the harvesting season, the processing method and the genetic factor.

It should be pointed out that dichlorobenzene (DCB), one of the frequently used pesticides which could cause damage to the kidney and liver [96, 97], has been found in all the CX samples. Owing to their significantly high relative content (13.19% ± 4.55%), over-use of pesticide during cultivation is notified and should be avoided. Since DCB does not exist naturally in herb, the DCB peak was removed from all 32 CX chromatograms before the establishment of CX chromatographic fingerprints.

Tetramethyl-pyrazine, one of the active ingredients, was found in *Radix Ligustici chuanxiong* via HPLC-DAD and HPLC-MS. Up to now, there is no previous report on the discovery of tetramethyl-pyrazine in chuanxiong through GC-MS analysis. In this study, we succeeded in discovering this ingredient with relative content of 0.39% in theCX5 sample. We speculated that sample CX5 has a high a concentration of tetramethyl-pyrazine so that an observable amount could still be extracted via our method.

Another finding in this work is that there were two unknown compounds 1 and 2 present in all the CX chromatographic profiles and their mass spectra are shown in Fig 4.11a. Here, compound 1 is used as example to illustrate how to obtain its pure mass spectrum in this work. Peak purity can also be identified through fixed size moving window evolving factor analysis (FSWMEFA). Its fixed size window method (FSWM) plot and evolving latent projection graph (ELPG) plot (see Fig. 4.11b) strongly suggested that only one component presents in that peak. Afterward, by using CRM with background noise being removed properly, its pure mass spectrum was obtained. Similarly, mass spectrum of another unknown component 2 was acquired. Further structure elucidation of these two unknown components is needed before taking them as markers of CX though their mass spectrums are known now.

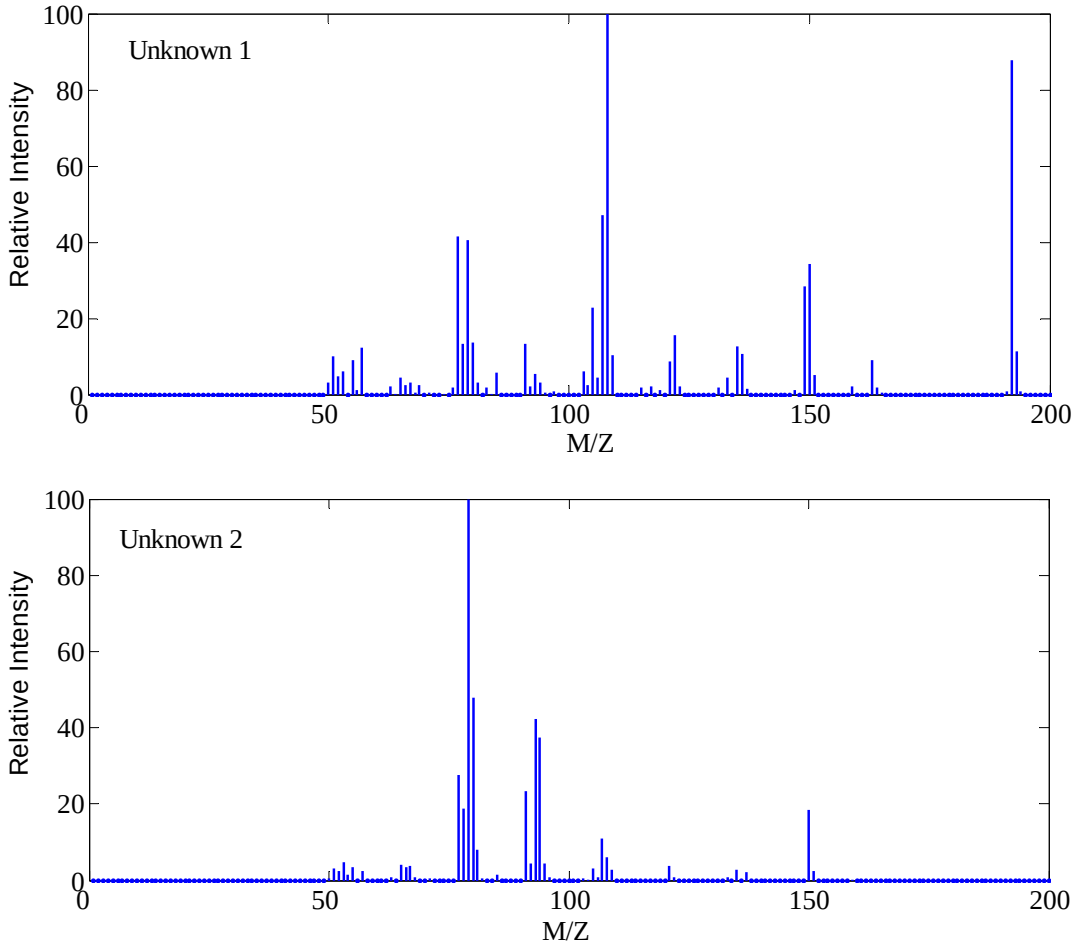


Fig 4.11a Resolved pure mass spectra of unknown compounds 1 and 2 using CRM

FSWM plot ELPG plot

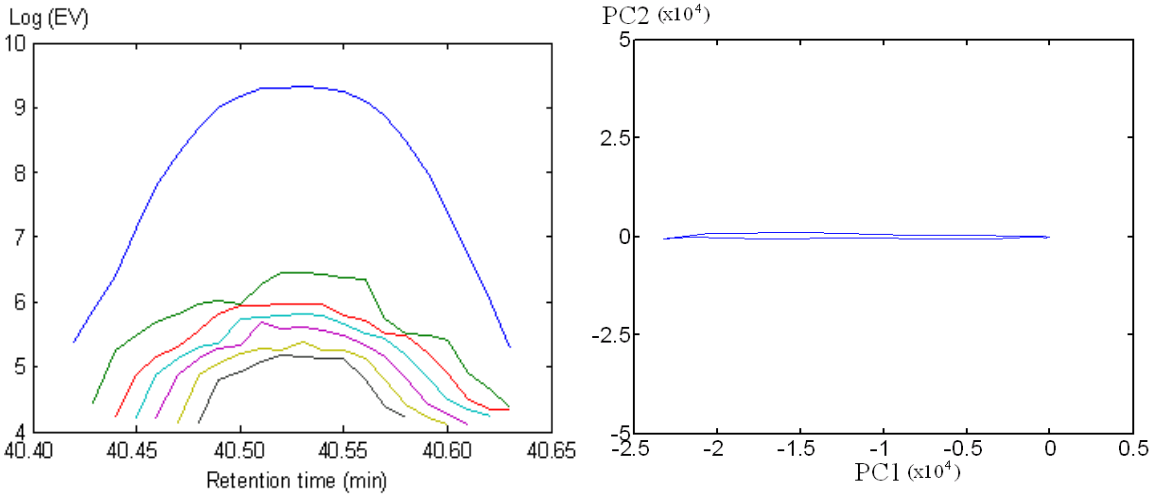


Fig 4.11b FSWM (left) and ELPG (right) plot of Unknown Component 1

Fig.4.11a) Resolved pure mass spectra of unknown compounds 1 and 2 using CRM, b) FSWM (left) and ELPG (right) plot of Unknown Component 1

4.6.4 Pattern approach

The concept of pattern approach has been innovated and utilized in CHM assessment in the last decade. The unique feature of pattern approach provides an excellent assessment criterion for herbs, herbal extract and prescriptions in quality control of CHM and it performs well in identification and classification. The basic idea of the pattern approach is to consider the whole chromatographic profile as an input feature. The advantage of pattern approach is that it does not involve the subjective selection of peaks, peak integration parameters as well as knowledge of the chemical to corresponding peak. The last chemical information is particularly difficult to get for CHM as the chemical composition of CHM is complex and chemical standards of CHM are usually not easy to get. Also, all the detectable components from the chromatographic instrument are utilized for assessment.

Before constructing the chromatographic fingerprint of CHM using pattern approach in this investigation, it is necessary to correct the retention times of the common peaks as identified by the combination of SCC and LLS methods on each chromatogram of each sample as the multivariate analysis with the whole chromatographic fingerprint being very sensitive to the retention time shift. The LLS-corrected or the aligned chromatograms were conducted and Fig.4.12 provides comparison of the chromatographic segments within 38 to 43 min before and after LLS treatment.

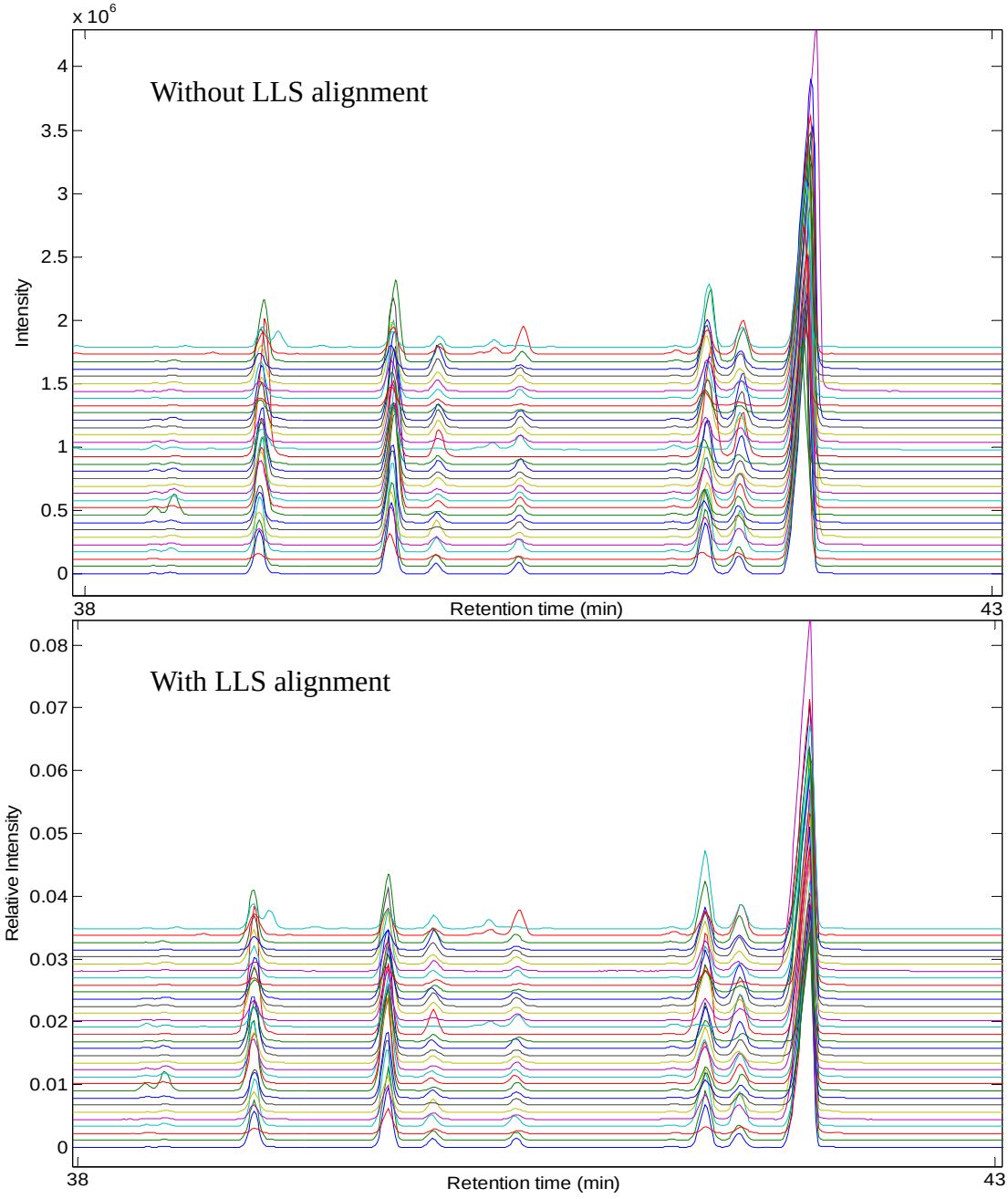


Fig. 4.12 A segment of original chromatographic profiles of 32 samples of CX from GC-MS before LLS treatment (top) and after LLS treatment (bottom)

Normalization of the chromatographic profiles is also an important step in pre-processing analytical data. The variation in sample concentration might affect multivariate analysis of the entire chromatographic profile. Therefore, normalization must be carried out before multivariate analysis.

By investigating the influences of LLS-treatment on the similarity index values based on the mean chromatograms among 32 CX samples, we found that there was a significant improvement on their SI values (see Table 4.6). The average SI value for non-LLS treatment was 0.909 ± 0.088 while the value after chromatographic alignment was 0.948 ± 0.060. The higher the average SI value is, the closer the similarity of the chemical compositions among all the CX samples is. Such improved result exhibited the fair consistency of 32 CX samples from different locations and the mean chromatographic profile thus established for authentication of CX herbal sample through SI assessment.

Table 4.6: Similarity index of each CX sample with their mean chromatogram before and after LLS treatment

	Without LLS treatment	With LLS treatment
Sample	Similarity Index (SI)	
CX 1	0.9922	0.9971
CX 2	0.9867	0.9871
CX 3	0.9167	0.9317

CX 4	0.9724	0.9830
CX 5	0.9182	0.9652
CX 6	0.8350	0.9741
CX 7	0.9238	0.9688
CX 8	0.9493	0.9874
CX 9	0.8630	0.8666
CX 10	0.9781	0.9710
CX 11	0.9654	0.9706
CX 12	0.9875	0.9874
CX 13	0.9071	0.9829
CX 14	0.8775	0.9549
CX 15	0.9600	0.9811
CX 16	0.7685	0.8002
CX 17	0.8451	0.8377
CX 18	0.8730	0.8825
CX 19	0.9839	0.9863
CX 20	0.9853	0.9917
CX 21	0.9508	0.9945
CX 22	0.8797	0.9897
CX 23	0.9622	0.9611
CX 24	0.8402	0.8381
CX 25	0.9910	0.9933
CX 26	0.6848	0.9376
CX 27	0.9901	0.9892
CX 28	0.9965	0.9929
CX 29	0.9141	0.9681
CX 30	0.9517	0.9815
CX 31	0.7432	0.8714
CX 32	0.6833	0.7990

4.6.5 Assessment of CX samples by Similarity Index

The mean chromatograms of the samples from three provinces of P. R. China (Fig.4.13) show that there are variations in their chemical contents, especially in the region with retention time between 27min and 35min. For some, one can easily distinguish them even by visual inspection. However, a better way is to use a systematic and objective measure to differentiate them for quality control of CHM. In order to quantify the differences among the samples, similarity index between these chromatographic profiles were adopt to reveal their differences.

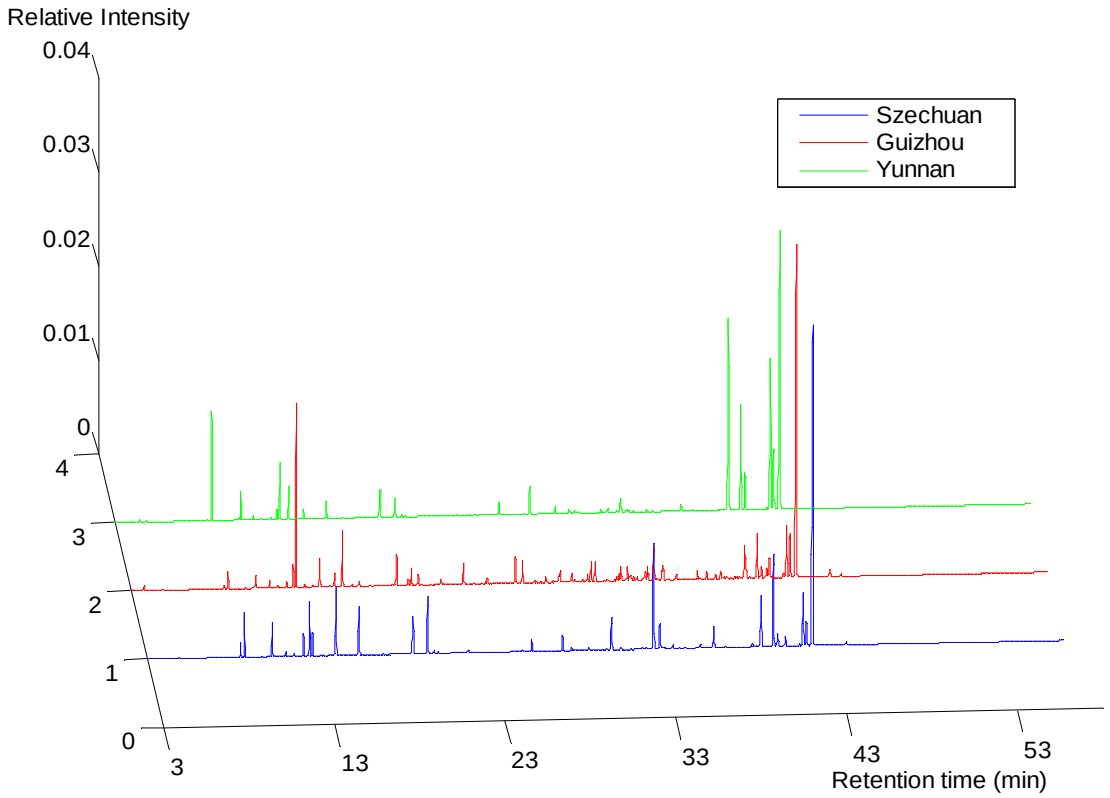


Fig. 4.13 The mean GC-MS chromatogram of the CX sample(s) from three different provinces of Szechuan, Guizhou, and Yunnan

Table 4.7 lists the results obtained using the similarity index indicator based on approaches mentioned previously. It involved two types of similarity index: a) SI of each chromatogram to the mean chromatogram of its own group and b) SI between their mean chromatograms. From the results of the marker approach, the similarity

index of individual chromatograms from Szechuan, Yunnan and Guizhou to their mean chromatogram from Szechuan were 0.9999 ± 0.0002 (mean ± S.D., n = 28), 0.9985 (n = 1), 0.9994 ± 0.0003 (n = 3), respectively. Since the relative contents of these CX samples are so similar, it does not reveal the differences in constituents in such a detail among all the CX samples. However, the mean SI values with their standard deviations under multi-component approach among three different provinces were 0.9715 ± 0.0415 (n = 28), 0.8129 (n = 1), 0.8824 ± 0.0393 (n = 3), respectively, while those values with their standard deviations under pattern approach were 0.9643 ± 0.0417 (n = 28), 0.8270 (n = 1), 0.8434 ± 0.0446 (n = 3), respectively. A more detailed inspection on the input parameters of the different approaches shows that it is the composition of the monoterpene and diterpene that differentiate the CX herbs from different geographic origin. In fact, from chemical composition point of view, the critical one relies on the difference between the genuine or non-genuine one was the relative content ratio between the monoterpene and diterpene components. We found out that both the multi-component approach and pattern approach were able to distinguish two classes of *rhizoma chuanxiong*, genuine herb and non-genuine herb. Besides, it can be seen that their SI values and their standard derivation of SI become lower and larger, respectively, when the input data features of chromatogram increased by comparing the different results of the SI value among the three different approaches. It may be attributed to that as more chromatographic information is utilized for SI calculation, the variations among chromatograms can be revealed better by the SI values obtained.

Table 4.7: A summary of the results of similarity index assessment based on the 'marker approach', 'multi-component approach' and 'pattern approach' among all the CX samples

Similarity Index (Marker approach)			
Geographical location	Szhechuan	Yunnan	Guizhou
Szhechuan (28)	0.9999 ± 0.0002[a]	0.9979 ± 0.0012[b]	0.9993 ± 0.0005[b]
Yunnan (1)	0.9985[b]	1.0000[a]	0.9956[b]
Guizhou (3)	0.9994 ± 0.0003[b]	0.9955 ± 0.0010[b]	1.0000 ± 0.0001[a]
Similarity Index (Multi-component approach)			
Szhechuan (28)	0.9715 ± 0.0415[a]	0.7900 ± 0.0776[b]	0.8666 ± 0.0734[b]
Yunnan (1)	0.8129[b]	1.0000[a]	0.8010[b]
Guizhou (3)	0.8824 ± 0.0393[b]	0.7889 ± 0.0246[b]	0.9826 ± 0.0130[a]
Similarity Index (Pattern approach)			
Szhechuan (28)	0.9643 ± 0.0417[a]	0.7987 ± 0.0685[b]	0.8356 ± 0.0595[b]
Yunnan (1)	0.8270[b]	1.0000[a]	0.7963[b]
Guizhou (3)	0.8434 ± 0.0446[b]	0.7728 ± 0.0129[b]	0.9699 ± 0.0179[a]

[a] Similarity index of each chromatogram to its own group of mean chromatogram, mean ± SD

[b] Similarity index of their mean chromatograms.

From the previous discussion, it seems that multi-component approach and pattern approach performed well in classifying genuinity of the CX samples through SI assessment. However, there were still some variations in their constituents among all the 28 Szechuan samples even though their mean SI was larger than 0.9. When the SI

values of individual samples (see Table 4.8) were investigated, we found out that the SIs of some CX samples such as CX9, CX16 and CX24 are down to 0.9. The two samples, CX16 and CX24, comprise relative low phthalide content while CX9 contains diterpene content, which was significantly different from their mean diterpene content.

4.6.6 Assessment of CX samples by Principal Component Analysis

To visualize and rationalize the chromatographic information from the three approaches, principal component analysis (PCA) was applied to all the above data (matrix). Its methodology involves transformation of the original variables into new uncorrelated variables called principal components and each principal component is a linear combination of the original variables. This new coordinate system is linked to variation of all the data sets. In our studies, only two or three largest principal components are sufficient to explain all of the variations present in all data sets so that classification of samples and identification of outliers in high dimensional data is easier to detect using PCA.

PCA utilized in this study to identify the key features is the characteristics of the volatile oil profile of each CX sample. Fig. 4.14a, b & c give the three plots of the three largest principal components of the 32 CX samples using the marker, multi-component and pattern approaches. Each sample is represented as a point in the principal component plot: Blue circle (●), green triangle (▲) and red square (■) represent samples from Szechuan, Yunnan and Guizhou, respectively. It should be noted that there was an overlap zone between Szechuan and Guizhou samples in the

Table 4.8: The SI values of each of the 32 CX samples by comparison of its mean chromatograms through the ‘multi-components approach’ and ‘pattern approach’

	Similarity index							Similarity index						
	Multi-component approach			Pattern approach				Multi-component approach			Pattern approach			
	Mean chromatogram of							Mean chromatogram of						
Sample	Yunnan	Guizhou	Szechuan	Yunnan	Guizhou	Szechuan	Sample	Yunnan	Guizhou	Szechuan	Yunnan	Guizhou	Szechuan
CX1	0.8225	0.9096	0.9971	0.8356	0.8771	0.9941	CX17	1.0000	0.8010	0.8129	1.0000	0.7963	0.8270
CX2	0.8464	0.9146	0.9902	0.8485	0.8750	0.9806	CX18	0.7744	0.9826	0.9160	0.7586	0.9645	0.8731
CX3	0.6542	0.8387	0.9573	0.6771	0.8090	0.9481	CX19	0.7681	0.9033	0.9946	0.7785	0.8720	0.9884
CX4	0.8659	0.8957	0.9900	0.8661	0.8584	0.9819	CX20	0.8483	0.8812	0.9938	0.8543	0.8514	0.9912
CX5	0.7360	0.8896	0.9824	0.7411	0.8451	0.9699	CX21	0.8391	0.9116	0.9955	0.8457	0.8757	0.9931
CX6	0.7830	0.9149	0.9852	0.7843	0.8738	0.9762	CX22	0.8504	0.9073	0.9911	0.8503	0.8697	0.9868
CX7	0.7734	0.8412	0.9810	0.7820	0.8128	0.9744	CX23	0.7460	0.9016	0.9750	0.7608	0.8671	0.9685
CX8	0.7658	0.8774	0.9924	0.7823	0.8495	0.9900	CX24	0.5890	0.7505	0.8829	0.6173	0.7187	0.8658
CX9	0.7446	0.6992	0.8890	0.7488	0.6944	0.8772	CX25	0.8120	0.8921	0.9966	0.8224	0.8591	0.9913
CX10	0.8390	0.8566	0.9810	0.8381	0.8239	0.9720	CX26	0.7633	0.9227	0.9678	0.7463	0.8680	0.9334
CX11	0.8492	0.8567	0.9727	0.8520	0.8289	0.9699	CX27	0.8269	0.9032	0.9959	0.8318	0.8656	0.9889
CX12	0.7895	0.8840	0.9909	0.8057	0.8513	0.9877	CX28	0.8370	0.8888	0.9967	0.8447	0.8547	0.9914
CX13	0.8020	0.9132	0.9903	0.8090	0.8724	0.9790	CX29	0.7746	0.9086	0.979	0.7842	0.8687	0.9713
CX14	0.8650	0.8233	0.9588	0.8628	0.8048	0.9578	CX30	0.8950	0.8997	0.9801	0.8964	0.8692	0.9754
CX15	0.8548	0.8846	0.9791	0.8588	0.8484	0.9752	CX31	0.7750	0.9956	0.8920	0.7762	0.9898	0.8650
CX16	0.5800	0.5935	0.8154	0.6388	0.6307	0.8224	CX32	0.8173	0.9696	0.8391	0.7837	0.9555	0.7922

PCA plot of the marker approach (Fig. 3.14a). On the other hand, PCA plot of the multi-component and pattern approaches can successfully classify samples from three different provinces. Genuine herb type and non-genuine herb type of CX were also clearly separated into two separate zones (Fig 3.14b & c). Therefore, this may imply that the mean chromatograms of the 28 CX Szechuan samples under study (Fig. 4.15) can be considered as a reference standard of the GC-MS fingerprint of CX and also be used for differentiation of CX from other sources of CX samples and CX from other related species under PCA.

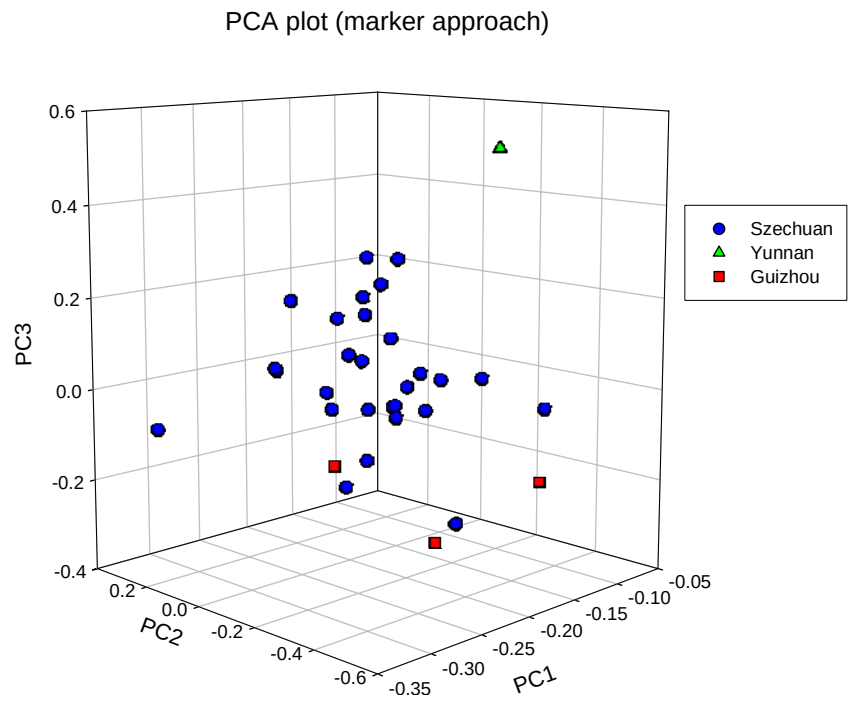


Fig. 4.14a Three projection plots on the three principal components of data from the 32 CX samples using the marker approach

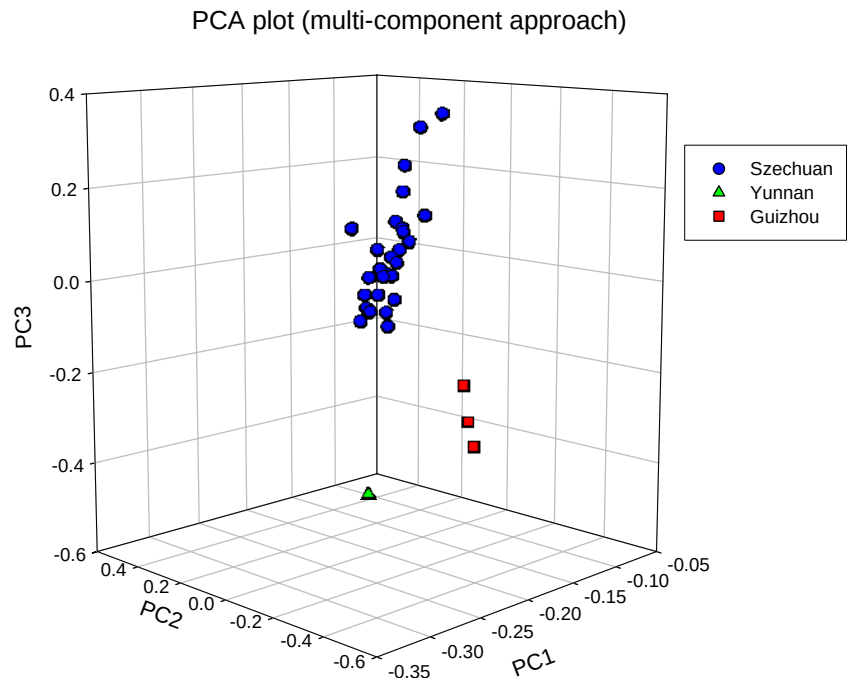


Fig. 4.14b Three projection plots on the three principal components of data from the 32 CX samples using the multi-component approach

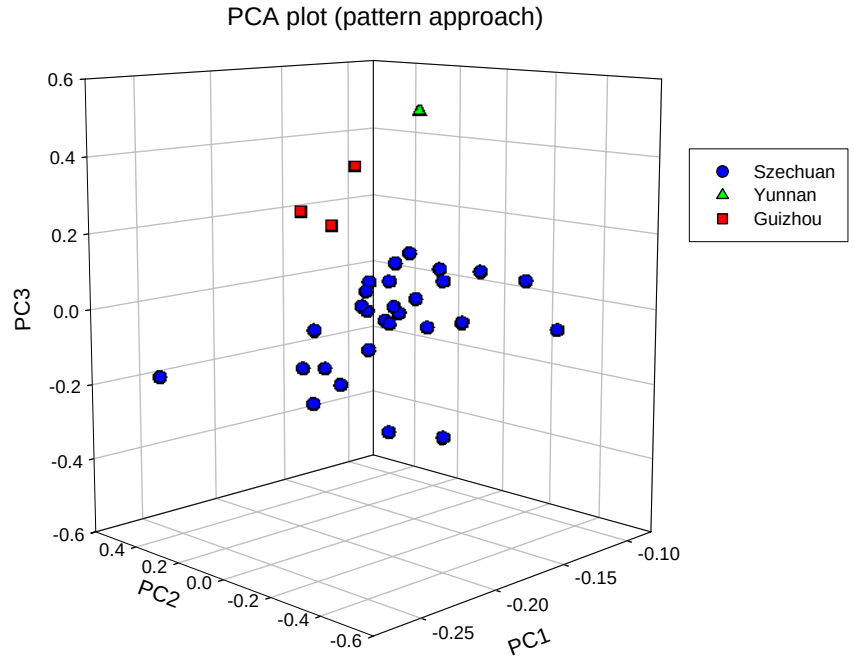


Fig. 4.14c Three projection plots on the three principal components of data from the 32 CX samples using the pattern approach

Fig. 4.14 Three projection plots on the three principal components of data from the 32 CX samples using the a) marker approach, b) multi-component & c) pattern approach

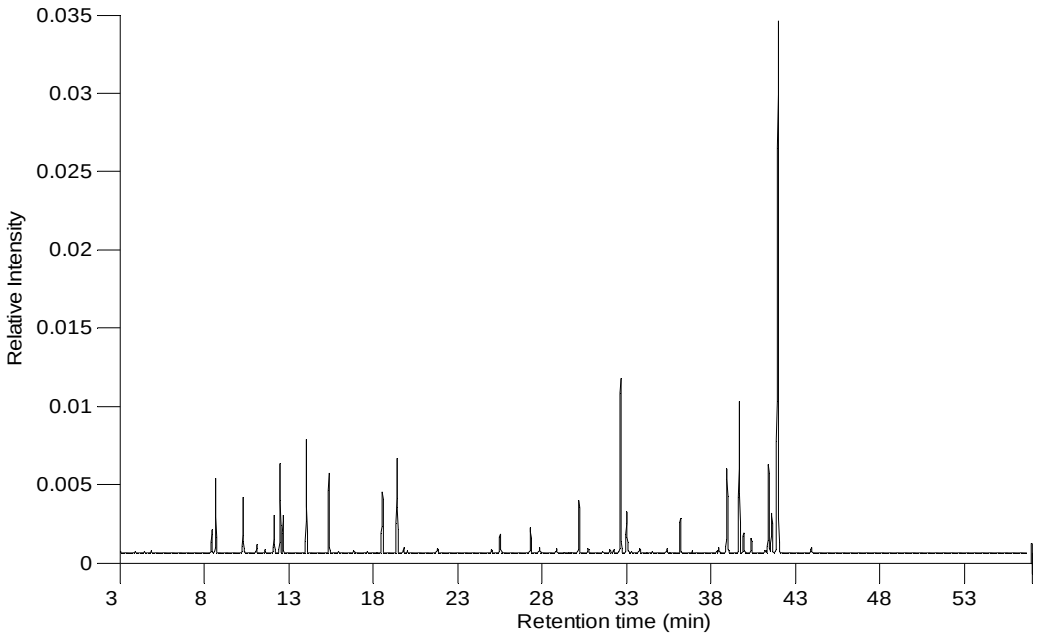
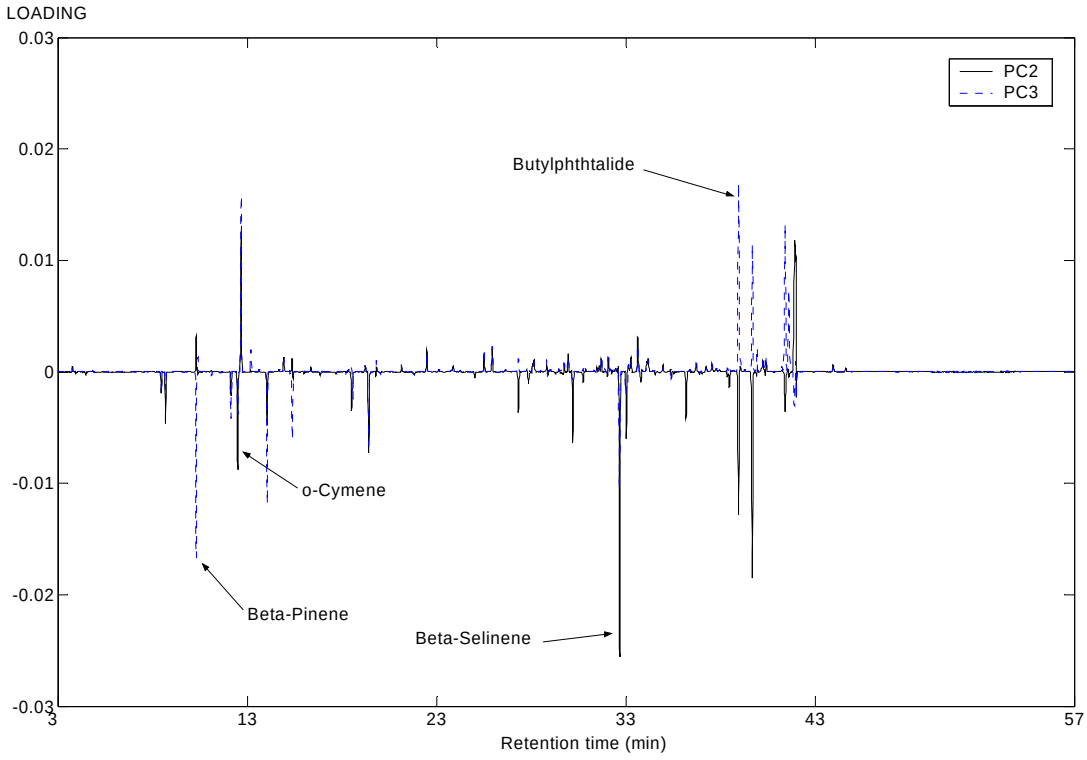


Fig. 4.15 The reference chromatographic fingerprinting standard of *Radix Ligustici chuanxiong* from the mean of 28 analyzed samples in szechuan province



4.16 The loading plot of PC 2 and 3 via PCA under pattern approach

By comparison of the relative percentage of common constituents present in each

province, several findings were discovered. Szechuan samples and Guizhou samples consisted of comparative larger content of beta-Selinene and o-cymene, respectively, to the other samples while Yunnan samples contained significant larger contents of beta-pinene and butylphthalide. It is the main reason why the multi-component and pattern approach can successfully distinguish samples from the three different provinces under the PCA plot. By evaluating the performance among the three approaches on the quality control of CHM, multi-component approach and pattern approach perform well in classifying of CX samples through PCA pattern recognition. Nevertheless, in respect of the effectiveness and efficiency on the quality control of CHM in real case, especially in the analysis duration, pattern approach is more preferable and cost effective as more chemical information as detected and shown in the chromatographic profile was utilized through the SI assessment or PCA analysis with no subjective judgment needed. The disadvantages of peak detections and integration, and the presence of a subjectively selected peak(s) that usually takes longer time to get can be avoided through the use of the whole chromatographic profile under the pattern approach. A further disadvantage using multi-component approach is that a large amount of data not selected in the chromatograms was not considered. For example, the presence of 3-methyl phenol, beta-bisabolene, cyclo-dodecane in the ‘Guizhou’ samples only have been neglected in previous part.

In summary, the variations among CHM samples from different sources may show the differences in curative effect on medical treatment. It also illustrate why quality control is so important. For further study of development of CX chromatographic fingerprint, a larger set of samples from different provinces and countries is certainly needed to come up a complete picture of CX with a very creditable conclusion.

4.7 Conclusion

This work demonstrated that it is possible to fully characterize the chemical composition of chunanxiong from different geographical sources, based on the three different data analysis methodologies, ‘marker approach’, ‘multi-component approach’ and ‘pattern approach’. The difference in nature among these three approaches is only the amount of chemical or chromatography information being used. We utilized several types of chemometrics techniques to explore how they worked, and they include Chemometrics Resolution Methods (CRM), Spectral correlative chromatography (SCC) and Local Least Squared (LLS) method to provide rapid, efficient and effective ways for the development of GC-MS chromatographic fingerprints of the chunaxiong herb.

Under these three approaches, several findings on the 32 chuanxiong chromatographic profiles were concluded. Z-ligusitilide, the commonly used marker, is the major constituents of volatile oil in *rhizoma chuanxiong* and the quantities of z-ligustilide, z-butylidenephthalide and ferulic acid among CX samples have been determined. With the help of CRM, the overlapped peaks can be resolved accurately and reliably and the components involved can be identified. In this manner, more than 30 common constituents were tentatively identified from their resolved MS spectra through library and literature matching. In addition, the resolved pure mass spectrum of two unknown components 1 and 2 were determined and further investigated. Finally, similarity index and principal component analysis through ‘marker approach’, ‘multi-component approach’ and ‘pattern approach’ have been explored and demonstrated in the quality control of *rhizoma chuanxiong*. 32 CX samples from different provinces are well

classified under PCA through ‘pattern approach’. We recommend that the mean chromatogram and the average content of the common volatile components from the Szechuan samples could be used as reference standards for quality control of this herb.

The entire work in this chapter is considered to be the first attempt on investigating the influence of CHM quality control based on different chromatographic information used. It seems that quality assessment of CHM works successfully under the ‘pattern approach’ in this case. In the future, the whole methodology needs to be implemented on the methods of identification and component analysis in the other CHM samples. Computers and data processing algorithms are now powerful enough to allow us to process the entire chromatographic profiles. We believe that us1ing all chemical information for data analysis could be a benefit in the quality control of CHM.

Chapter 5: A comparison of fingerprint analysis: Development of classification model between *Radix Angelicae sinensis* and *Radix Ligustici chuanxiong* by gas chromatography-mass spectrometry

5.1 Introduction

Radix Angelicae sinensis (DG) and *Radix Ligustici chuanxiong* (CX) were selected and investigated in developing the classification model in this study. According to reports in the literature, chemical compositions of these two CHMs in essential oil part are quite similar and z-ligustilide, which has been served as markers, co-existed in both CHMs. Therefore, it is difficult to differentiate them in the authentication process based on the conventional approach.

The present study has been focused on developing a classification model to identify CX and DG. Based on three different methodologies, including ‘marker approach’, ‘multi-component approach’ and ‘pattern approach’, different portions of chemical information detected by GC-MS chromatogram were extracted and used to construct the classification model in which five-folds cross validation have been utilized in testing their overall performance. Multi-samples collections of these CHMs with different geographical sources were involved in order to generate a representative picture. This is the first attempt to compare the three different approaches for characterizing two CHMs with similar chemical compositions in order to be able to differentiate them objectively in a reliable way. The procedures developed in this work can be used as a basis for further development on the authentication process in CHM.

5.2. Background of Radix Angelicae sinensis

Danggui, is the dried root of *Angelica sinensis* (Oliv.) Diels., perennial herb of family *Umbelliferae*. It is produced mainly in Min County in the southeast part of Gansu and

also in Shannxi, Sichuan, Yunnan, Guizhou and Hubei provinces and collected in the late autumn. After removal of rootlets and drying slightly by evaporation, the roots are tied up in bundles, placed on a shelf and smoke-dried. It also placed on frame, baked dry over slow fire, sliced and used unprepared or stir-baked with wine [25, 77].

For the appearance of *Radix Angelicae sinensis*, it is somewhat cylindrical root with 15 - 25cm long where 5 or more branched at the lower part. Its externally color is yellowish-brown to brown externally and its surface longitudinally wrinkled and transversely lenticellate. It can be sub-divided into three parts, root stocks, main roots and branching roots. Root stocks (Guitou) is 1.5 - 4cm in diameter, annulated, apex obtuse and rounded, showing purple or yellowish-green remains of stem and leaf sheaths; main roots (Guishen) is lumpy on the surface; branching roots (Guiwei) is 0.3 - 1cm in diameter, with thick upper portion and thin lower portion where mostly twisted and exhibiting a few rootlet scars [25, 77].

Its texture is flexible, fracture yellowish-white or yellowish-brown while its bark is thick and shown in some clefts and numerous brown dotted secretory cavities. The odour is strongly aromatic and the taste is sweet, pungent and slight bitter. If the roots become woody withered and not oily, or greenish-brown on the fracture, it cannot use as medicinal treatment [25, 77].

Radix Angelicae sinensis is pungent and sweet in flavour, warm in nature, it is tropistic to the heart, liver and spleen channels. Being sweet, warm and moist properties, this herb serves function of replenishing blood. Being pungent, warm and volatile in nature, it also promotes blood circulation, as a main herb for enriching

blood and promoting the circulation of blood to regulate menstruation and relieve pain. It is often used to treat the syndromes of blood-deficiency, various kinds of pain due to blood stasis, menoxenia, sores, ulcers, swellings and traumatic ecchymoma, etc. It also functions in nourishing blood and lubricating the bowel, serving for constipation due to blood deficiency and dryness in the bowel [25, 77].

The indications of *Radix Angelicae sinensis* are given as follows [25, 77]:

1. For any syndromes of blood deficiency, it is often utilized with other herbs to nourish the blood.

2. For menoxenia, menorrhalyia and amenorrhea in the type of blood deficiency, it is worked together with other herbs to replenish blood and regulate menstruation; for those in the type of blood stasis, it is utilized with other herbs for promoting blood circulation and for regulation menstruation; and for those in the type of deficiency-cold, it is used with other herbs for warming channels and dispelling cold pathogen.

3. For chest and abdominal pain due to blood stasis, it is often utilized with other herbs for resolving blood stasis to relieve pain; for obstruction of the heart channel manifested as oppressive pain in the chest, it can be used with other herbs for promoting blood circulation and Qi flow to relieve pain.

4. For traumatic ecchymoma and pain, and for sores, carbuncles and swelling, it is often used with other herbs for promoting blood circulation and tissue regeneration

and relieving pain.

Besides, it is also combined with other herbs to treat constipation due to blood deficiency and dryness in the bowel.

Radix Angelicae sinensis shows a dual action on the uterus. Its volatile oil components inhibit uterus to decreases the rhythmic contraction and to relax the smooth muscle in a rapid and long-standing way; and its water or alcohol soluble components excite uterus to strengthen its contraction. Its liquid extract functions in inhibiting isolated heart to lower cardiac excitability and to prolong atrial refractory period; dialating the coronary artery and markedly increasing the coronary flow; lowering myocardial oxygen consumption, blood pressure, platelet aggregation and blood-fat; resisting thrombosis and preventing atherosclerosis; protecting the liver and preventing degradation of liver glycogen; resisting absence of vitamin E, malignant anemia and inflammation; and inducing diuresis, sedation and analgesia [78].

In addition, it also not only shows immunosuppressant action, i.e. significant inhibition of the production of antibodies, cellular immunity and humoral immumity, serving for nephritis, neonatal hemolytic disease (an autoimmune disease), but also raises the phagoctosis of macrophagocyte and reticuloendothetiol cells and strengthen nonspecific immunity [78].

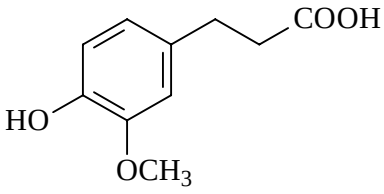
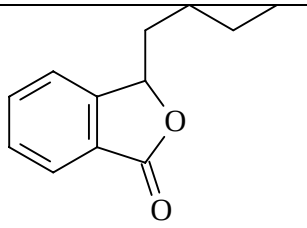
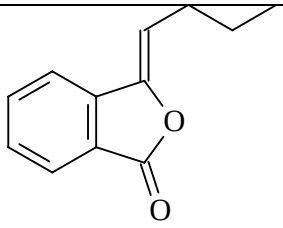
5.3 Literature review on chemical studies of *Radix Angelicae sinensis*

5.3.1 Chemical Composition

The chemical compositions of *Radix Angelicae Sinensis* contain several classes of components such as essential oil, phthalide dimers, organic acids and their ester, polyacetylenes and sterols etc [37]. Main constituents of *Radix Angelicae Sinensis* are shown in Table 5.1 and chemical structures of some identified compounds from *Radix Angelicae Sinensis* are also listed in Fig. 5.1:

Table 5.1: Main constituents of *Radix Angelicae Sinensis*

Essential oil	Monomeric phthalides: E- and Z-ligustilide, butylphthalide, E- and Z-butylidenephthalide, carvacrol, isoeugenol, vanillin, α-pinene, β-bisabolone, mycene, etc
Phthalide dimers	riligustilide, levistolide A, tokinolide B
Phenolic constituents	ferulic acid, coniferyl ferulate, linoleic acid, palmitic acid, etc
Vitamins	Vitamins A, B1, B12 and E, etc
Polyacetylenes	Falcarindiol, falcarinol, etc
Others	Sterols, polysaccharides

		
1	2	3

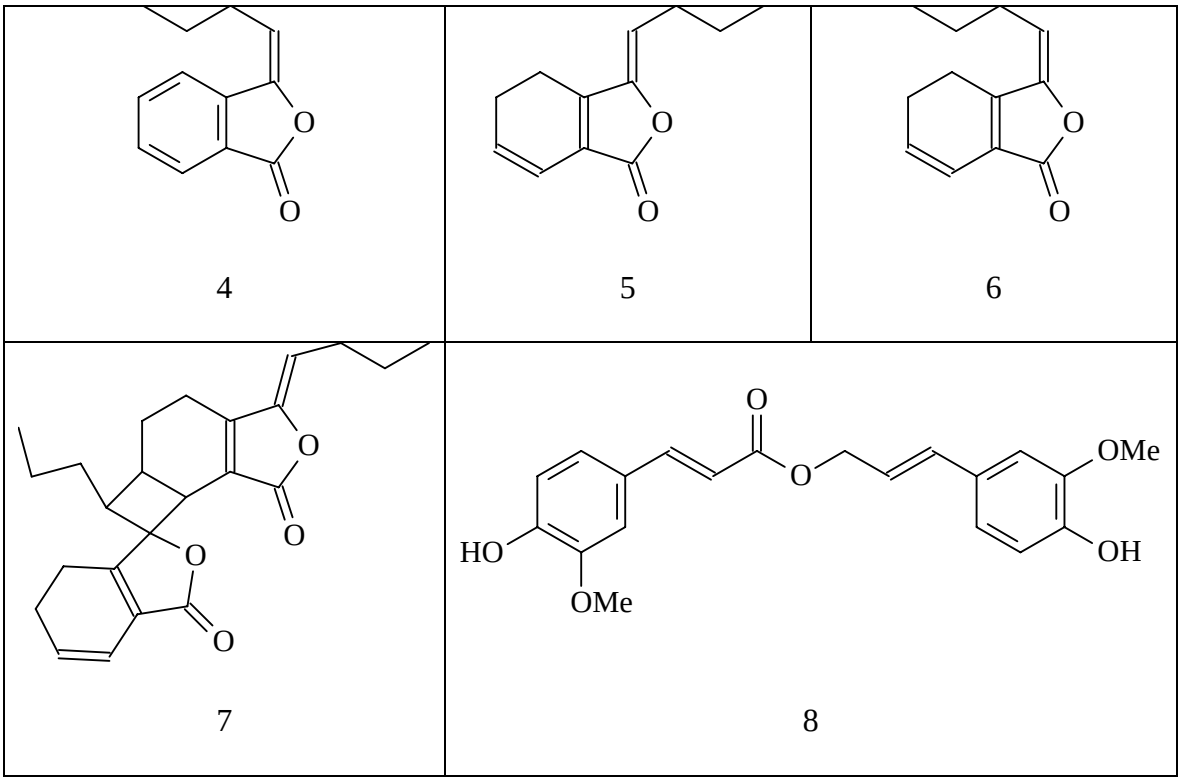


Fig. 5.1 Chemical structures of the identified compounds in *radix angelicae sinensis*: 1) Ferulic acid; 2) Butylphthalide; 3) Z-butylidenephthalide; 4) E-butylidenephthalide; 5) Z-ligustilide; 6) E-ligustilide; 7) Riligustilide; and 8) coniferyl ferulate;

5.3.2 Marker components

Z-ligustilide and ferulic acid were identified and have been served as commercial marker of *radix angelicae sinensis*. Both components have been reported and showed some pharmacological actions. Therefore, *the Hong Kong Chinese Materia Medica Standards* uses both compounds served as markers of *radix angelicae sinensis* in which the actual content of z-ligustilide must contain more than 0.6% while the Pharmacopoeia of the People's Republic of China provides the reference standardization on the actual content of 0.05% ferulic acid [25, 98].

A number of homologues of ligustilide, including butylphthalide, Z-butylidenephthalide, E-butylidenephthalide and vinylphthalide were isolated and their biological activities also were investigated. Those components have been reported and showed certain pharmacological actions, indicating that the phthalide moiety is the principal anti-asthmatic component of phthalide derivatives of *Radix Angelicae Sinensis*. Nevertheless, those reported phthalide components and ferulic acid are still presented in another famous herb, *Radix Ligustici chuanxiong* and cannot act as a unique marker for *Radix Angelicae Sinensis* [37, 38, 78].

5.3.3 Pharmaceutical value of the marker components

The phthalides components from *Radix Angelicae sinensis* including butylphthalide, butylidenephthalide and ligustilide show anti-asthmatic and spasmolytic on tracheal muscle, exhibit antiarrhythmic effect, and dilatating activity on coronary arteries. Also, ferulic acid shows anti-flammatory and analgesic effects, i.e. increase of phagocytosis activity of macrophages. Furthermore, it inhibits the platelet aggregation and the serotonin release from thrombocytes [37, 78].

5.3.4 Chemical studies of *Radix Angelicae Sinensis*

Different chromatographic and spectroscopic techniques have been reported for characterizing the chemical compositions of *Radix Angelicae Sinensis*. Several chromatographic studies on phthalide components have been reported under HPLC-DAD / MS [30, 91, 94], HPLC-UV / LC-MS [30, 94, 99-102]. In addition, lot of analytical studies using GC-MS has been explored for chemical compositions of

volatile oil in *Radix Angelicae Sinensis* [20, 87, 91, 95, 103]. Furthermore, W. Cao et al & Y. Sun et al characterized the polysaccharides compounds of *Radix Angelicae Sinensis* using Fourier Transform – Mid Infrared FT-MIR [104-106].

5.3.5 Chemical studies for classifying between *Radix Angelicae Sinensis* & *Radix Ligustici chuanxiong*

Only a few chromatographic and spectroscopic studies have been reported in the literature about classifying *Radix Angelicae Sinensis* & *Radix Ligustici chuanxiong*. Z. Zschocke et al. discovered one unknown component (retention time at 8.6min) found in *Radix Ligustici chuanxiong* via HPLC chromatographic technique [91]. G.H. Lu et al. found out that large amount of senkyunolide A existed in *Radix Ligustici chuanxiong* is qualitatively different those components existed in *Radix Angelicae Sinensis* [30]. F. Gong et al. completely characterized the chemical compositions of essential oil between *Radix Angelicae Sinensis* and *Radix Ligustici chuanxiong* through CRM techniques [20]. Z. Zschocke et al. developed TLC method to differentiate *Radix Angelicae Sinensis* & *Radix Ligustici chuanxiong* where the latter TLC chromatographic profile appears a red-pink zone at Rf = 0.35 with anisealdehyde sulphuric acid spray reagent [91]. The approach in almost previous studies utilized a ‘unique marker’ for differentiation. Subjective judgments including peaks detection, peaks integration, peaks elimination in data reduction are involved. Also, sampling size is not large enough to fully represent their corresponding species.

5.4 Methods of Investigation

The main feature of this entire algorithm in this work is the ability to maximize the chemical information obtained in the whole set of chromatograms by the use of different chemometrics techniques during data processing. Then, similarity index, hierarchical clustering analysis (HCA), principal component analysis and linear discriminant analysis (LDA) were utilized for classifying *Radix Angelicae Sinensis* and *Radix Ligustici chuanxiong*. The objective is to characterize the chemical compositions of *Radix Angelicae Sinensis* and *Radix Ligustici chuanxiong* and to develop a classification model for them

Several chemometrics techniques have been employed into characterizing the essential oil of chemical compositions for *Radix Angelicae Sinensis & Radix Ligustici chuanxiong* (completed in Chapter 4) in this work. Chemometrics Resolution Method was fully utilized into resolving all the overlapping peak of chromatograms obtained in GC-MS; Spectral Correlative Chromatography was made good use of finding out all the common components in a series of chromatograms; The Local least Squares method was also applied to accomplish the chromatographic alignment before developing the classification model.

In this work, marker approach, multi-component approach and pattern approach were investigated to develop a classification model for DG and CX. In the markers approach, the actual contents of three commonly used markers, ferulic acid, z-butylidenephthliade and z-ligustilide in all samples were worked altogether distinguishing between DG and CX. For the multi-component approach, the corresponding common constituents obtained from CX and DG were applied for classifying them. For pattern approach, all the chromatographic profiles including CX

and DG were utilized for differentiating them. Afterward, cross-validation process was exploited in evaluating the performance of classification models where k-nearest neighbour (k-NN), which is non-parametric classification technique without hypothesis formulation on distribution of the input variables, acts as a classifier. In this investigation, five-folds cross validation were employed for testing performance of the classification models under those proposed approaches. The whole data set are divided into five groups and each group consists of CX and DG samples randomly. Four of five groups were used to develop the classification model while the remaining group was used to validate the model. The process has to be repeated until all five groups have been used for prediction, and the prediction results of the tested samples will be compared with its known result.

5.5 Experimental

5.5.1 Herbal samples

Twenty DG samples were obtained from China Academy of Traditional Chinese Medicine, Beijing, P.R. China in 2001-2002 with authenticated certificates provided in 2002. All danggui samples were collected from Gansu Province (mainly in Minxian), the recommended cultivation region for this herb. The geographical sources with their collection periods among all the DG samples are summarized in Table 5.2.

Table 5.2: A list of DG samples collected for this study

No.	Sample codes	Geographical sources	Collection period
1	DG1	Minxian, Gansu, China	April, 2001

2	DG2	Weiyuan, Gansu, China	April, 2002
3	DG3	Weiyuan, Gansu, China	April, 2002
4	DG4	Minxian, Gansu, China	May, 2001
5	DG5	Minxian, Gansu, China	May, 2002
6	DG6	Zhangxian, Gansu, China	April, 2001
7	DG7	Zhangxian Gansu, China	April, 2002
8	DG8	Minxian, Gansu, China	May, 2002
9	DG9	Minxian, Gansu, China	May, 2002
10	DG10	Longxi, Gansu, China	April, 2002
11	DG11	Minxian, Gansu, China	January, 2002
12	DG12	Minxian, Gansu, China	May, 2001
13	DG13	Minxian, Gansu, China	May, 2001
14	DG14	Minxian, Gansu, China	April, 2002
15	DG15	Minxian, Gansu, China	May, 2002
16	DG16	Minxian, Gansu, China	May, 2001
17	DG17	Minxian, Gansu, China	May, 2002
18	DG18	Minxian, Gansu, China	May, 2002
19	DG19	Minxian, Gansu, China	May, 2001
20	DG20	Minxian, Gansu, China	May, 2002

5.5.2 Chemical and reagents

The three standards, z-ligustilide, z-butylidenephthalide and ferulic acid were purchased from ChromaDex, Inc. (UAS), Wako Pure Chemical Industries, Ltd. (Japan), and Lancaster Synthesis (England), respectively. Solvents for GC-MS experimentation were of analytical grade. Double deionized water was used in the sample preparation.

5.5.3 Apparatus and chromatographic conditions

GC-MS experiments were performed on all the DG extracts using Agilent 6890 coupled with Agilent 5973 series mass selective detector equipped with a fused silica capillary column HP-5MS (5% phenyl methyl siloxane, 0.25mm × 30m × 0.25μ m). The column temperature was maintained at 40℃ for 0.5 minute after injection, then increased at 3℃ min^{-1} to 175℃ and 5℃ min^{-1} to 220℃; maintained at this temperature for 3 minutes. Split injection was conducted with a split ratio of 1:100 and helium was utilized as the carrier gas with 1.0ml min^{-1} flow rate. The spectrometer was operated in the electron-impart (EI) mode with the scan range 45-500amu, the ionization energy 70eV and the scan rate 1.66scans per second. The inlet and ionization source temperature were 250℃ and 280℃, respectively.

5.5.4 Sample preparation

5.5.4.1 Sample pre-treatment

DG being examined were ground into powder and passed through 100 mesh stainless sieve. Then the sieved powder was stored in polypropylene containers at 20°C.

5.5.4.2 Sample extraction

Essential oils of the DG samples were extracted by the standard method recommended by the Pharmacopoeia of the People's Republic of China [25]. About 80g dried herbs and 500mL double-deionized water were pre-mixed into 1000mL round bottom flask. The solution was shaken and mixed well. Then the apparatus was set up and shown in

Fig 5.2. The flask was heated gently, and continued heating for about 5 hours after boiling begun, until the volume of oil did not increase anymore. The heater was turned off and the setup was allowed to stand for cooling. After cooling, the oily layer was centrifuged and the water layer was run off. Then the volume of oil in the graduate tube was read and recorded. The volatile oil contents of danggui ranged from 0.2% to 0.7% by weight in the investigation. In the GC-MS analysis, samples were firstly diluted 100 times by hexane and aliquots of 1μl were injected.

5.5.5 Data analysis

All Data analysis were performed on a Pentium IV personal computer. All the calculations and chemometrics data treatments were carried out using MATLAB 6.5. The National Institute of Standards and Technology (NIST147) MS database, which contains about 107000 compounds and the Wiley138 library were used for matching the resolved pure spectra to identify the chemical constituents.

5.6 Results and Discussion

5.6.1 Development of danggui fingerprint

5.6.1.1 Qualitative analysis of DG by the application of CRM

Fig. 5.2 depicts the TIC chromatograms of the essential oil of 20 danggui (DG) samples. As can be seen from the chromatograms, there are a lot of peaks and their contents do not vary, quite significantly, among all the samples. Although the

experimental condition for the chromatographic separation was optimized, overlapping clusters still existed. Therefore, CRM provide an alterative way to resolve all the overlapped peaks. The peak cluster I within 33.30-34.51min of DG14 Fig. 5.3 is illustrated as demonstration.

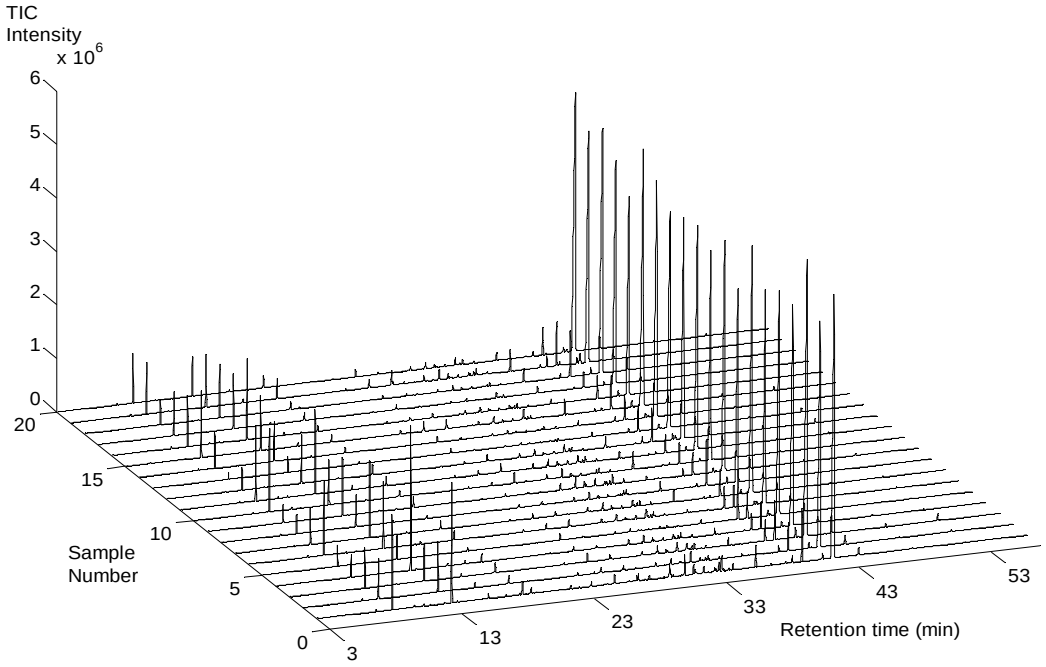


Fig. 5.2 GC-MS chromatogram of 20 samples of DG

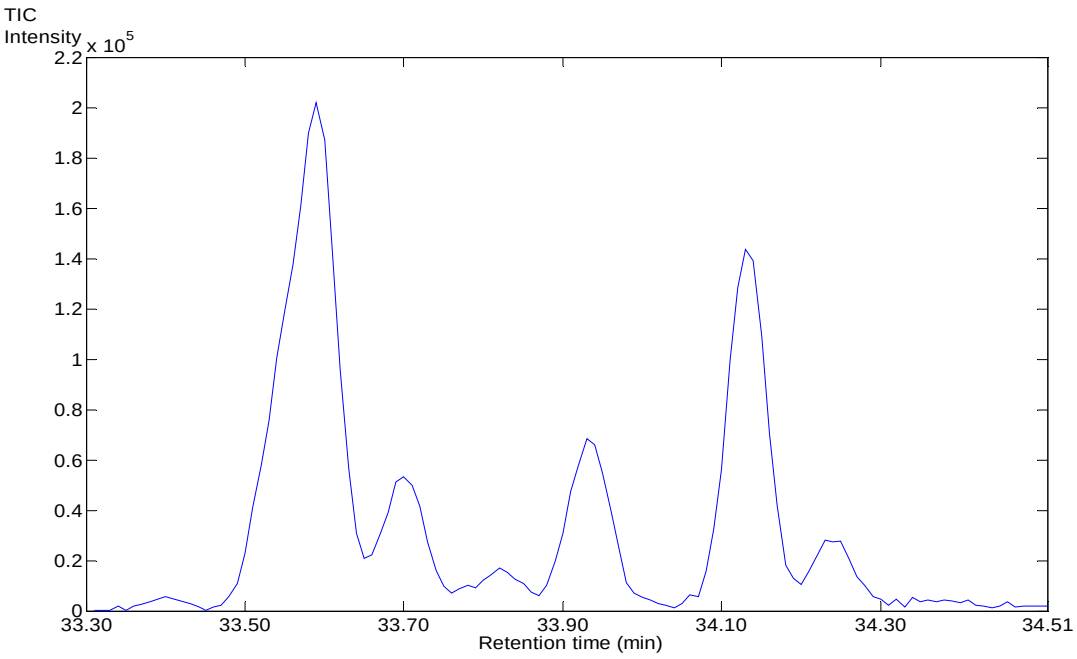


Fig. 5.3 TIC curve for peak cluster I (within 33.30 – 34.51 min) of sample DG14

In the peak cluster I, it seems that there are seven components in the segment by visual inspection, but the diverse spectra of different parts of the cluster indicates that there would be more. Routine direct search can definitely not achieve the satisfactory result. Thus, CRM is used to extract the pure spectra and pure chromatogram. Based on the result of eigenstructure tracking analysis, nine components were found in this cluster and they were marked as components 1-9 according to their elution sequence. By determining the eluting order of each component, individual selective and zero concentration regions of all the constituents in this peak cluster I, the pure profile of each component with its respectively pure spectra have been resolved out. By similarity searches in the mass library, each component in this cluster can be identified. These components 2, 3, 4, 7 & 8 were ledene, elixene, beta-himachalene, cuparene and beta-bisabolene with the similarity of matching results 0.9123, 0.9590, 0.9189, 0.9737 and 0.8686, respectively. Components 1, 5, 6 and 9 cannot be identified because of the low signal-to-noise ratio or the absence of the compound from the mass spectra database. With the help of the pure mass spectra of the components obtained from CRM, the accuracy and reliability of the results were improved greatly. The resolved mass spectra and the resolved pure chromatograms in peak cluster I are shown in Fig. 5.4 - 5.8 and Fig. 5.9, respectively. In the same way, the components in other peak clusters of DG chromatograms were resolved, if needed.

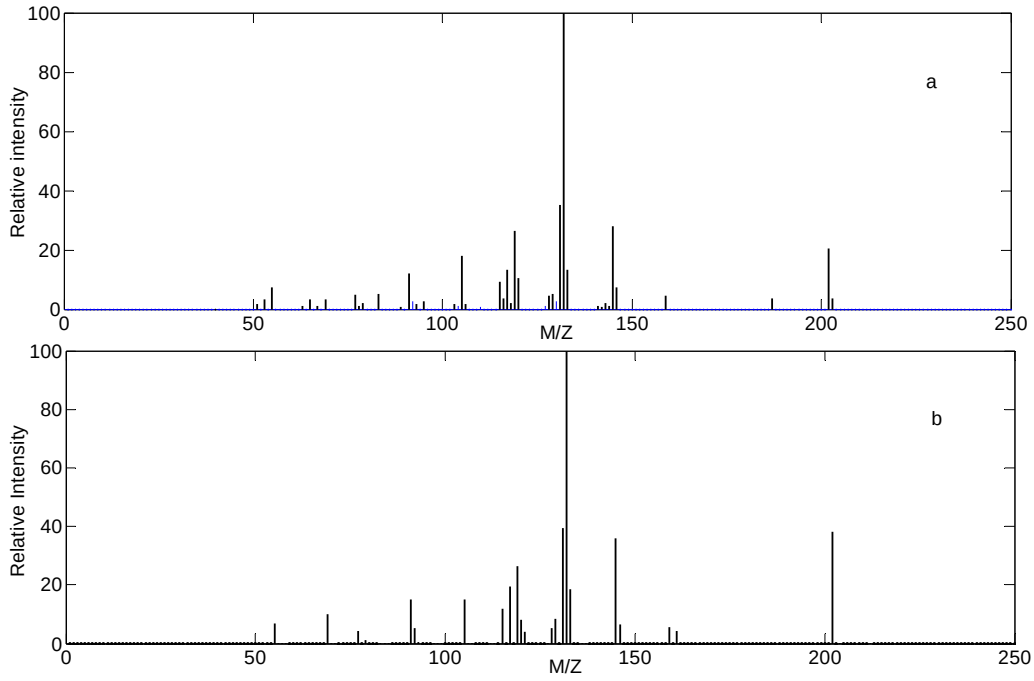


Fig. 5.4 Resolved mass spectrum of component 7(a) and the standard mass spectrum of cuparene (b)

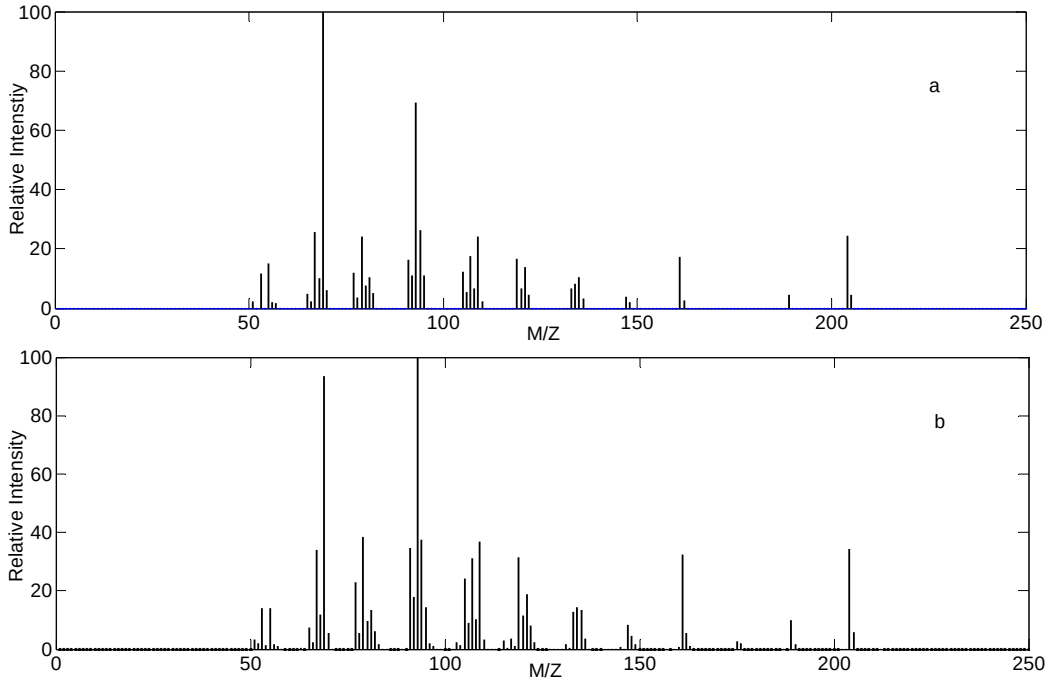


Fig. 5.5 Resolved mass spectrum of component 8 (a) and the standard mass spectrum of beta-bisabolene (b)

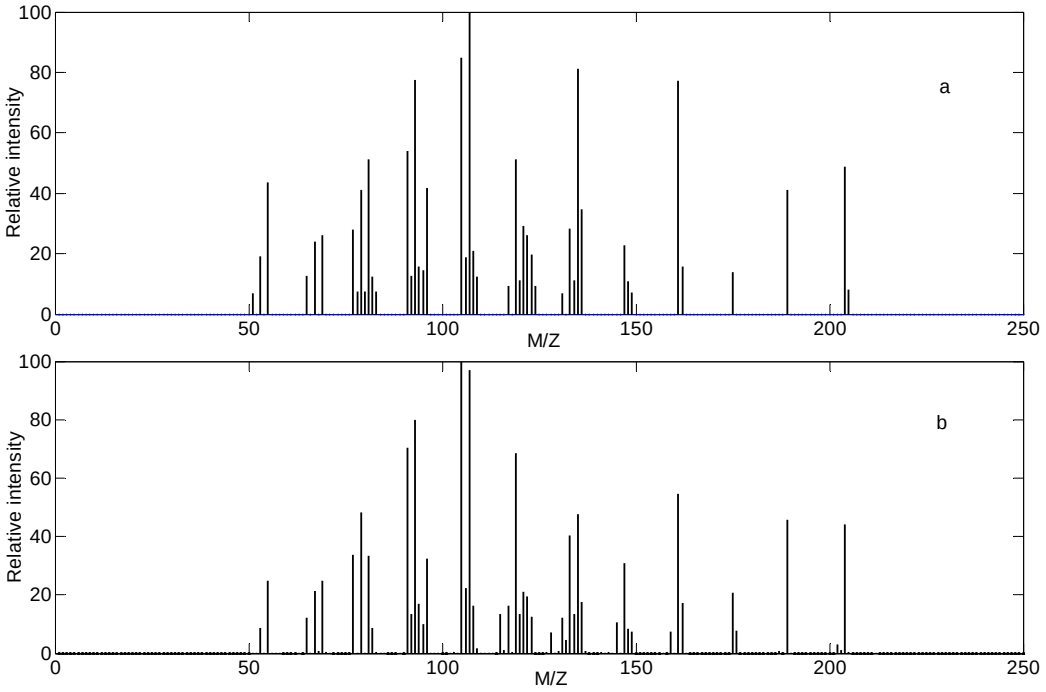


Fig. 5.6 Resolved mass spectrum of component 2 (a) and the standard mass spectrum of ledene (b)

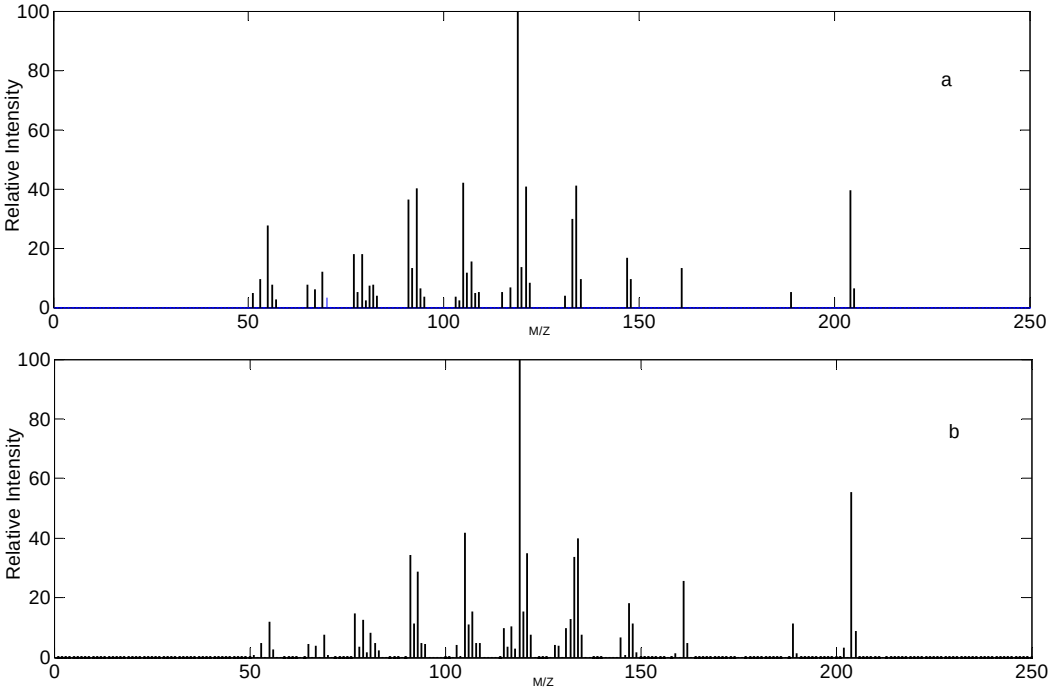


Fig. 5.7 Resolved mass spectrum of component 4 (a) and the standard mass spectrum of beta-himachalene (b)

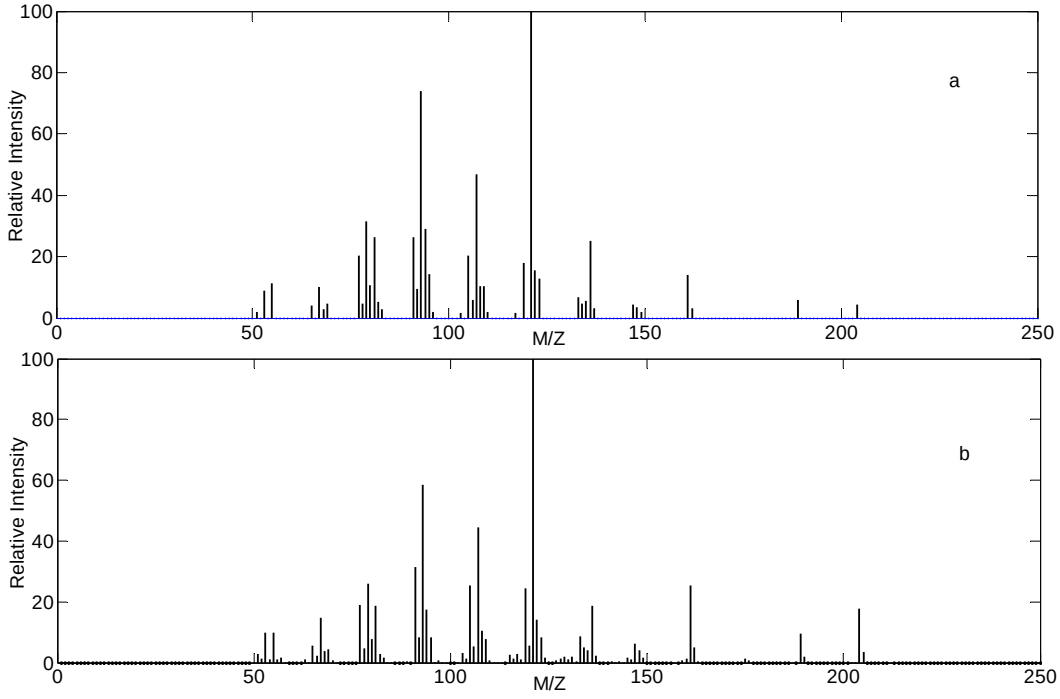


Fig. 5.8 Resolved mass spectrum of component 3 (a) and the standard mass spectrum of elixene (b)

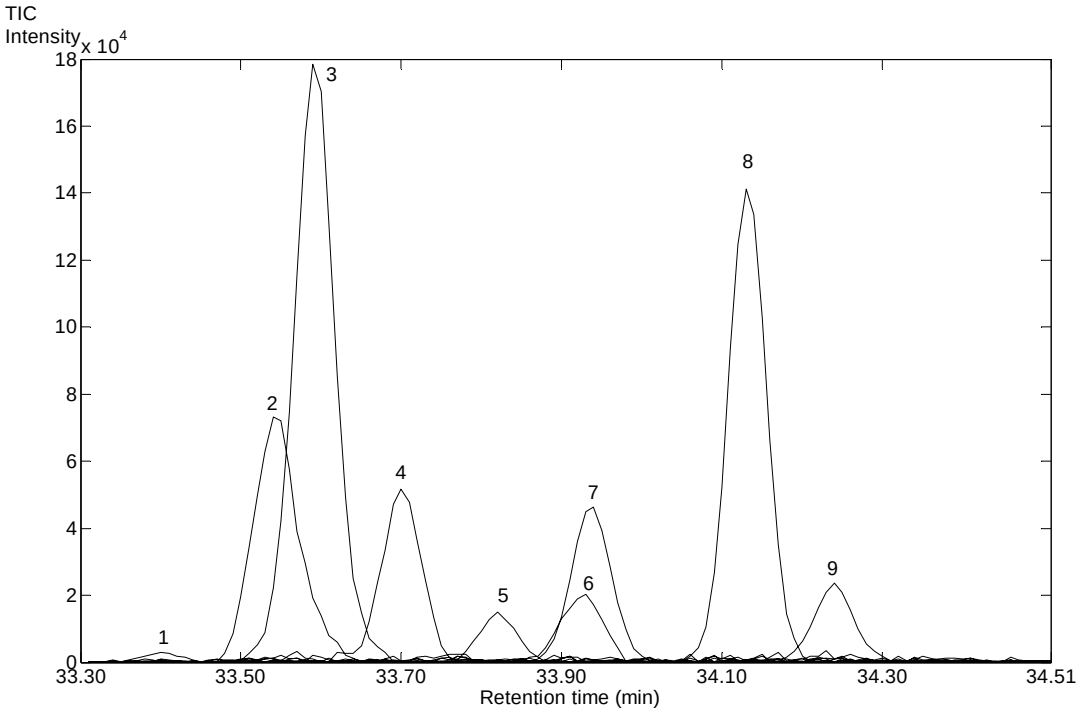


Fig. 5.9 TIC resolved profiles of peak cluster I (within 33.30 – 34.51 min)

5.6.1.2 Marker approach

z-ligustilide, z-butylidenephthalide and ferulic acid are commonly accepted standards for *Radix Angelicae Sinensis* and they are commercially available for quality control purpose. Before quantitative determination of these standard chemicals in all the DG extracts, their stability was one of the major technical concerns. In this work, all DG extracts were stored under -20^{o}C refrigerator without light and stability of z-ligustilide in such storage condition was investigated by two independently tests which are same as those tests used in CX extracts. Our outcome is that the level of z-ligustilide was nearly the same in both independent tests. This indicates that z-ligustilide was relatively stable under our storage condition.

To quantify the amount of z-ligustilide, z-butylidenephthalide and ferulic acid within all the DG samples, the external standard method was employed for building up their calibration curves, respectively, in this work. Their linearity was determined in the concentration ranges of 2-1000mg/L, 0.1-1000mg/L and 2-100mg/L, respectively, with six different concentrations of standard solutions for each one, while their corresponding regression coefficients were found to be 0.9993, 0.9989 and 0.9986. The regression equations thus obtained were utilized for quantifying their contents in all the DG samples and Table 5.3 lists out the amounts of these three standard components within all samples. Moreover, method reproducibility was also examined by four successive injections of both samples and standard solutions. Precision of replicated injections were determined and the relative standard deviations of z-ligustilide and z-butylidenephthalide in all samples and standards are less than 5% while that of ferulic acid is less than 7.5%.

The mean values of the absolute contents of the three markers, Z-ligustilide, Z-butylidenephthalide and ferulic acid of all the samples, were 23.54 ± 3.73 mg/g, 6.27 ± 2.30 mg/g, 3.37 ± 1.29 mg/g respectively. These values are closed to or even higher than the literature value reported in previous HPLC studies [94]. Hence, this indicates that the quality of our 20 DG samples is consistent with the samples reported earlier by other workers and the results obtained from this investigation also reflect the generality feature of DG.

Table 5.3: The absolute contents of Z-ligustilide, Z-butylidenephthalide and ferulic acid in the 20 DG samples

Sample	Content (mg/g)[a]		
	Z-ligustilide	Z-butylidenephthalide	Ferulic acid
DG1	27.88 ± 0.45	11.45 ± 0.22	5.30 ± 0.12
DG2	21.86 ± 0.67	5.59 ± 0.15	3.34 ± 0.07
DG3	32.16 ± 1.04	8.42 ± 0.28	5.23 ± 0.28
DG4	21.97 ± 0.79	5.57 ± 0.23	2.74 ± 0.17
DG5	22.03 ± 0.44	5.37 ± 0.13	4.38 ± 0.24
DG6	20.07 ± 0.32	4.32 ± 0.08	3.37 ± 0.11
DG7	26.18 ± 0.37	5.24 ± 0.07	3.15 ± 0.09
DG8	17.44 ± 0.83	9.75 ± 0.54	1.99 ± 0.15
DG9	23.69 ± 0.86	4.44 ± 0.17	1.76 ± 0.10
DG10	20.94 ± 0.67	4.34 ± 0.14	2.05 ± 0.10
DG11	22.72 ± 0.68	3.53 ± 0.15	2.57 ± 0.11
DG12	22.42 ± 0.37	7.17 ± 0.13	1.87 ± 0.08
DG13	21.05 ± 0.85	5.35 ± 0.16	3.19 ± 0.10
DG14	26.00 ± 0.93	5.29 ± 0.21	3.34 ± 0.13
DG15	29.38 ± 0.51	7.16 ± 0.16	5.76 ± 0.27
DG16	18.26 ± 0.80	4.53 ± 0.22	2.05 ± 0.06
DG17	22.89 ± 0.49	3.25 ± 0.09	3.49 ± 0.10

DG18	27.45 ± 0.85	9.69 ± 0.36	2.01 ± 0.05
DG19	25.28 ± 0.81	9.17 ± 0.33	5.15 ± 0.11
DG20	31.12 ± 1.04	5.75 ± 0.21	4.62 ± 0.32

a The value is mean ± S.D. (n=4)

5.6.1.3. Multi-component approach

It can be seen in Fig. 5.3 that the chromatographic profiles of twenty DG samples are rather consistent with only minor variation on the relative contents of some individual components. It would be time-consuming and unimaginable to extract the pure component spectra for all chromatograms through CRM. Therefore, SCC as mentioned in Chapter 3 provides another effective method to determine and compare the common constituents among all samples. By using CRM together with SCC, 29 common components (see Table 5.4) of the 20 DG samples have been found and tentatively identified by matching their spectra with those in the NIST mass library and mass spectra published previously. It should be noted, these newly identified components provides more choice of markers for DG in future analysis.

Table 5.4: Identification and quantification of common constituents in the essential oil among 20 DG samples

Series No.	Retention Time (min)	Compound Name	Molecular structure	Content range (%)
1	5.001	Acetic acid butyl ester	$C_6H_{12}O_2$	0.04% – 0.07%
2	7.636	Nonane	C_9H_{20}	0.04% – 0.19%
3	8.924	Alpha-pinene	$C_{10}H_{16}$	1.77% – 7.61%
4	10.617	Beta-pinene	$C_{10}H_{16}$	0.05% – 0.25%
5	13.425	Trans-ocimene	$C_{10}H_{16}$	2.44% – 17.10%

6	13.881	3-Carene	$C_{10}H_{16}$	0.02% – 0.19%
7	16.263	Undecane	$C_{11}H_{24}$	0.06% – 0.59%
8	17.591	2,4,6-Octariene,2,6,dimehtyl	$C_{10}H_{16}$	0.03% – 0.40%
9	18.817	Benzene pentyl	$C_{11}H_{16}$	0.04% – 0.51%
10	18.818	6-butyl-1,4-cycloheptadiene	$C_{11}H_{18}$	0.36% – 2.20%
11	24.150	2-Undecanone	$C_{11}H_{22}O$	0.09% - 0.37%
12	25.924	Ferulic acid	$C_{10}H_{10}O_4$	0.47% – 1.26%
13	27.626	Benzaldehyde, 2,4,5, trimethyl	$C_{10}H_{12}O$	0.17% – 1.79%
14	27.758	1-Pentanone 1-phenyl	$C_{11}H_{14}O$	0.11% – 0.68%
15	31.276	Gymnomitrene	$C_{15}H_{24}$	0.42% – 1.96%
16	32.735	Beta-Chamigerene	$C_{15}H_{24}$	0.13% – 1.80%
17	33.496	Ledene	$C_{15}H_{24}$	0.06% – 0.44%
18	33.557	Elixene	$C_{15}H_{24}$	0.16% – 1.34%
19	33.668	Beta-himachalene	$C_{15}H_{24}$	0.06% – 0.44%
20	33.932	Cuparene	$C_{15}H_{24}$	0.16% – 1.01%
21	34.124	Beta-bisabolene	$C_{15}H_{24}$	0.33% – 1.46%
22	36.719	Spathulenol	$C_{15}H_{24}O$	0.35% – 2.80%
23	39.571	3-butylphthalide	$C_{12}H_{14}O_2$	0.27% – 8.65%
24	40.247	(Z)-3-butylidenephthalide	$C_{12}H_{12}O_2$	2.50% – 9.11%
25	41.838	(E)-3-butylidenephthalide	$C_{12}H_{12}O_2$	0.58% – 1.39%
26	42.041	Senukyunolide A	$C_{12}H_{16}O_2$	0.07% – 1.76%
27	42.588	Z-Ligustilide	$C_{12}H_{14}O_2$	51.64% – 78.15%
28	44.586	E-Ligustilide	$C_{12}H_{14}O_2$	0.72% – 2.05%
29	48.782	Palmitic acid, methyl ester	$C_{17}H_{34}O_2$	0.02% – 0.39%

Common components are one of the important features for evaluating the constituent differences among all the samples and Table 5.4 lists the tentatively identified common components of essential oil of the 20 DG samples. These common constituents included mono-terpene, di-terpene and phthalide-type compounds. Phthalide-class compounds of (Z)-ligustilide, (Z)-3-butylidenephthalide, and 3-butylphthalide were the major essential oil components in all the DG samples and

the mean relative content of the total phthalide-type compounds in all samples was 74.52 ± 7.06%. In addition, Z-ligustilide was the major constituent (66.01 ± 6.33%) among all the DG samples. This further supports why z-ligustilide has been served as one of the prominent markers in Danggui in many previous studies.

5.6.1.4 Pattern approach

Before construction of the chromatographic fingerprint of CHM through the pattern approach in this investigation, retention time correction of the common peaks identified by the combination of CRM and SCC had to be performed by local least squares (LLS) methods on each chromatogram of each DG sample. It is because the multivariate analysis on the entire chromatographic fingerprint is very sensitive to the retention time shift. Through the LLS treatment, the aligned chromatograms were obtained. Finally, normalization of all DG chromatographic profiles has to be done before quality assessment through similarity index evaluation (see Section 4.5) in order to eliminate the concentration effect on TIC chromatograms.

By investigating the influences of LLS-treatment on the similarity index from their mean chromatograms among 20 DG samples, we found that there was a significant improvement on their SI values (see Table 5.5). The average SI value for non-LLS treatment was 0.9335 ± 0.0591 while this value after chromatographic alignment was 0.9869 ± 0.0102. The higher the average SI value is, the closer the similarity of the chemical compositions among all the DG samples is. Such improved result exhibited the good consistency of 20 DG samples from different locations and the mean chromatographic profile thus established for authentication of DG herbal sample

through SI assessment (see Fig. 5.10).

Table 5.5: The similarity index of each DG sample with their mean chromatogram before and after LLS treatment

	Without LLS treatment	With LLS treatment
Sample	Similarity Index (SI)	
DG 1	0.9005	0.9772
DG 2	0.9870	0.9981
DG 3	0.8296	0.9905
DG 4	0.9468	0.9543
DG 5	0.9739	0.9842
DG 6	0.9393	0.9884
DG 7	0.9694	0.9973
DG 8	0.8677	0.9798
DG 9	0.9964	0.9962
DG 10	0.9715	0.9877
DG 11	0.9899	0.9771
DG 12	0.9974	0.9963
DG 13	0.9818	0.9913
DG 14	0.9364	0.9876
DG 15	0.8579	0.9927
DG 16	0.9139	0.9797
DG 17	0.9928	0.9967
DG 18	0.8775	0.9946
DG 19	0.9349	0.9848
DG 20	0.8057	0.9837

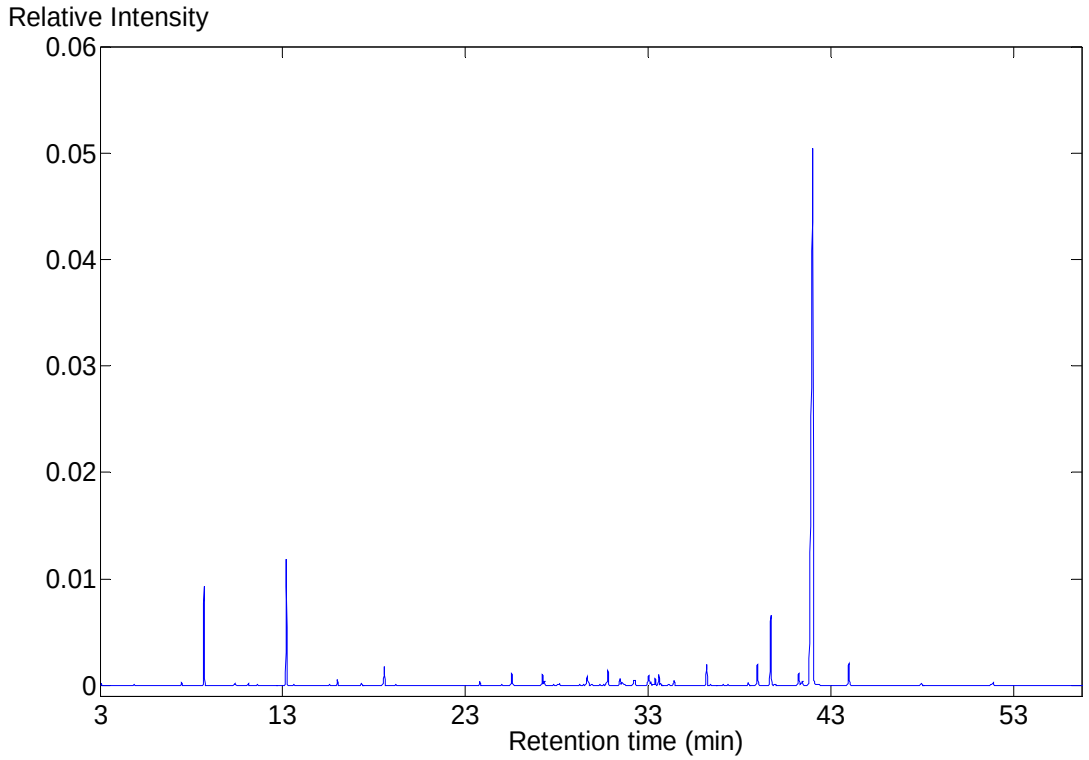


Fig. 5.10 The reference chromatographic fingerprinting standard of *Radix Angelicae Sinensis* obtained from the mean of 20 analyzed samples in Gansu province

5.6.2 Development of classification model between *Radix Angelicae Sinensis* and *Radix Ligustici chuanxiong*

The number of total fingerprint samples of the two medicinal herbs, CX and DG, from different sources are 32 and 20, respectively and they are randomly divided into five groups. Then, chemical information from chromatographic profiles based on 'marker approach', 'multi-component' and 'pattern approach' was extracted for developing the classification models. Five-folds cross validation methods were also utilized for testing the performance of classification model during the process. Finally, comparisons of the developed classification models based on different approaches were investigated and evaluated. All CX and DG samples for training and testing sets

are listed in Table 5.6.

Table 5.6: The group lists of two different CHM samples of CX and DG

Group 1	Group 2	Group 3	Group 4	Group 5
CX-02	CX-01	CX-05	CX-03	CX-16
CX-04	CX-07	CX-08	CX-06	CX-18
CX-09	CX-10	CX-15	CX-12	CX-21
CX-11	CX-13	CX-25	CX-14	CX-23
CX-26	CX-19	CX-30	CX-17	CX-24
CX-28	CX-31	CX-32	CX-20	CX-27
DG-07	DG-01	DG-03	CX-22	CX-29
DG-12	DG-02	DG-10	DG-04	DG-06
DG-19	DG-05	DG-11	DG-08	DG-14
DG-20	DG-15	DG-13	DG-09	DG-16
			DG-17	DG-18

5.6.2.1 Marker approach

From the previous studies, z-ligustilide, z-butlidenephthalide and ferulic acid were served as markers components of *Radix Angelicae Sinensis* & *Radix Ligustici chuanxiong* in quality control. In this study, both samples contained all three markers and their actual contents have been found out. By comparing with their corresponding marker amounts only in conventional approach (Table 5.7), it is ambiguous to distinguish which is Danggui or Chuanxang.

Table 5.7: Comparison of average actual content among three markers between CX and DG

Marker	Average actual contents (mg/g)	
	CX	DG
Ferulic acid	2.06 ± 1.09	3.37 ± 1.29
z-butylidenephthalide	5.95 ± 2.44	6.27 ± 2.30
z-ligustilide	19.05 ± 5.17	23.54 ± 3.73

Multivariate statistical analysis including hierarchical clustering analysis (HCA), similarity index, principal component analysis and linear discriminant analysis have been also applied to the entire data sets extracted from marker approach, however both clustering analysis were failure to classify Danggui and Chuanxiong. HCA dendrogram plot based on the single-linkage approach with their corresponding euclidean distance and three-dimensional principal component analysis plot were depicted in Fig. 5.11 and Fig 5.12 respectively. It is obvious that two clusters between DG and CX mixed up together.

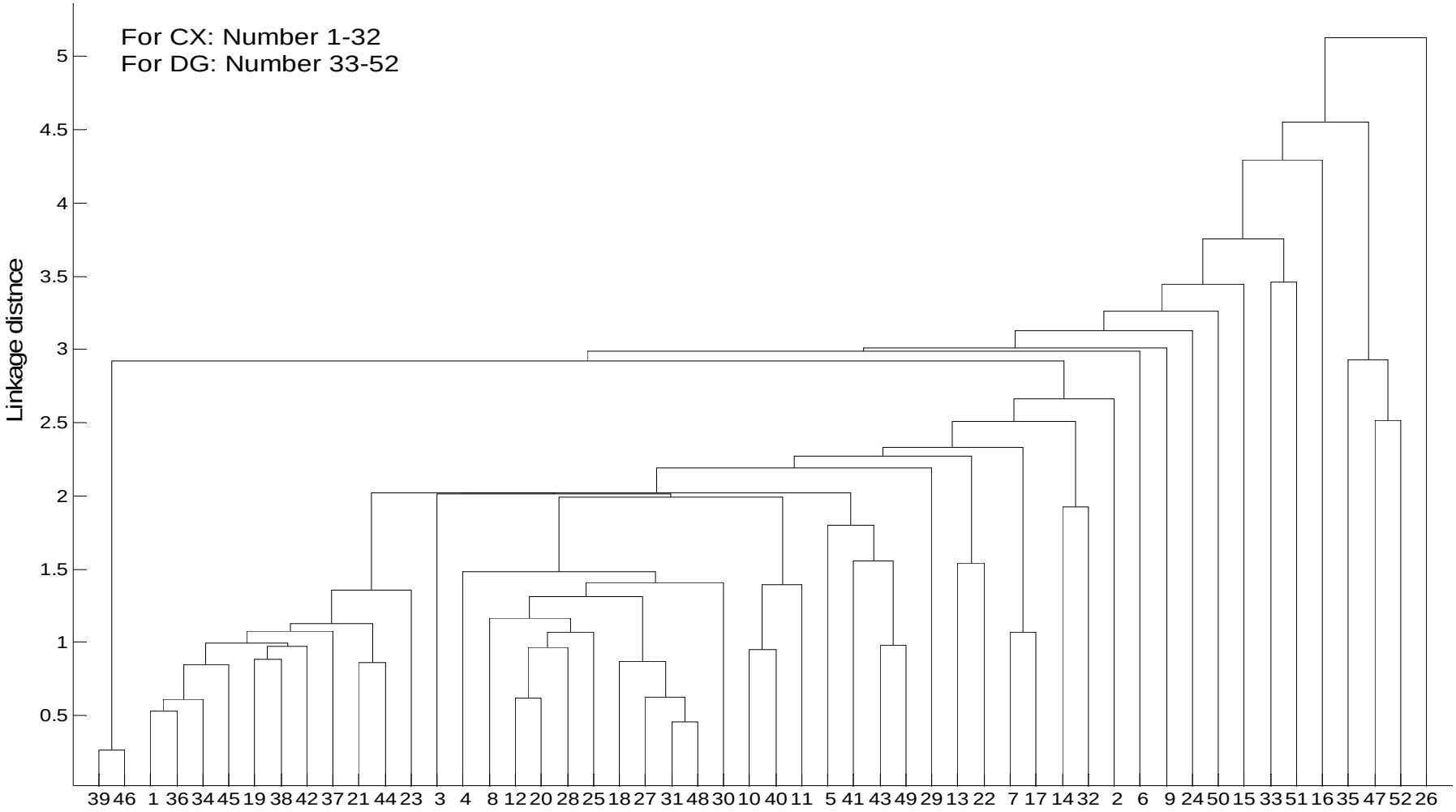


Fig. 5.11 Dendrogram plot of the CX and DG samples obtained from chromatographic profiles through marker approach

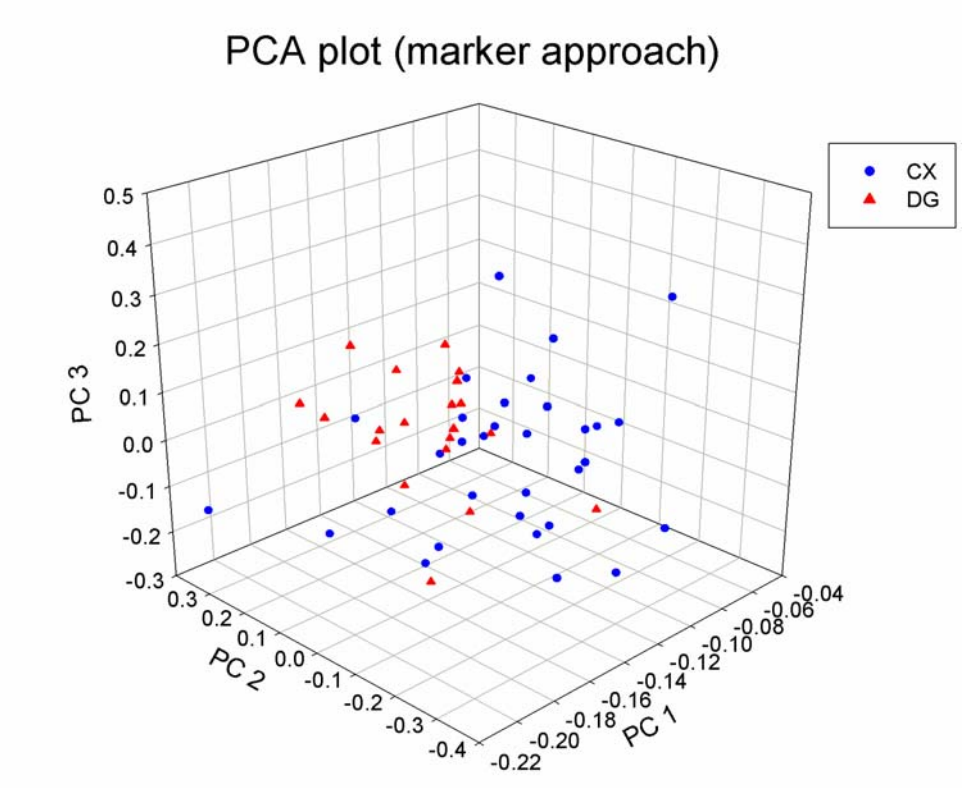


Fig. 5.12 Three projection plots on the three principal components of data from the 32 CX and 20 DG samples using the marker approach

5.6.2.2 Multi-component approach

Based on the result of common compositions obtained (Table 4.4 and Table 5.4), it was found that there were thirteen components co-existed in both CX and DG samples and majority of them are phthalide components. As a consequence, arrangement of the relative contents of common constituents into matrix form must be done before multivariate analysis have been applied. There were 52 different samples with 51 constituents included and they were to be arranged to [51 × 52] matrix form with the same constituent being put into the same row.

To perform an agglomerative clustering, a method of measuring the distance between two groups has to be decided. Euclidean distance, city block metric, cosine distance and correlation distance were selected for measurement. In addition, simple linkage, average linkage, complete linkage and wards method were also utilized into linkage clustering algorithm. Table 5.8 lists out the results of hierarchical agglomerative clustering analysis for classifying DG and CX. It was found that a few agglomerative cluster methods succeeded in classifying CX and DG. Cityblock metric was the best measuring distance while Ward's method performed relatively well on the linkage of cluster under the multi-component approach. Fig. 5.13 shows the dendrogram plot of the CX and DG samples obtained from city block metric with the use of ward's linkage through multi-component approach.

Table 5.8: The results of hierarchical agglomerative clustering analysis for classifying DG and CX under the multi-component approach

	Single linkage	Average linkage	Complete linkage	Ward's linkage
Euclidean distance	×	×	×	×
City block metric	×	✓	✓	✓
Cosine distance	×	×	×	✓
Correlation distance	×	×	×	×

✓ represents successful classification

× represents unsuccessful classification

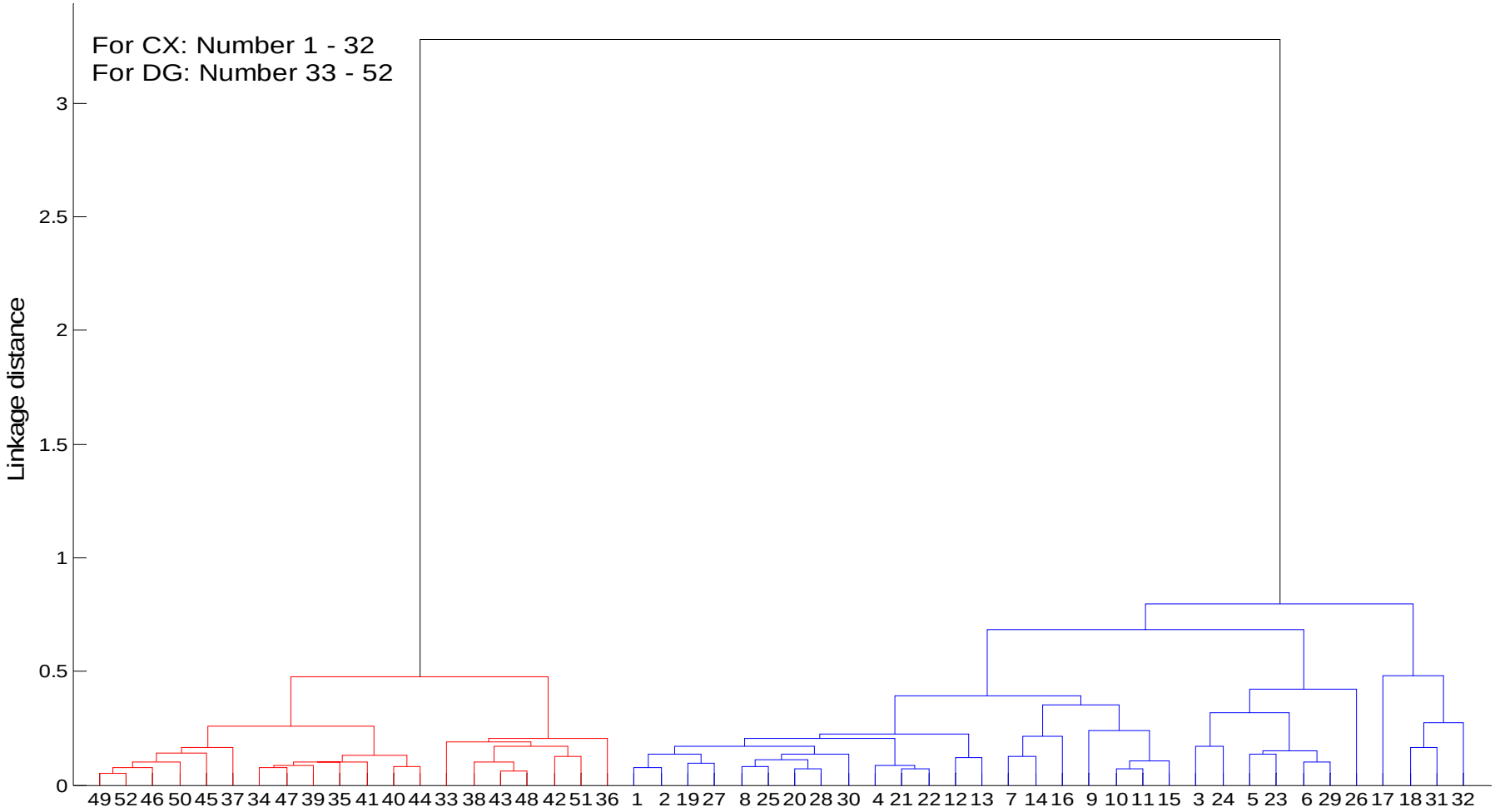


Fig. 5.13 Dendrogram plot of the CX and DG samples obtained from chromatographic profiles through multi-component approach

Table 5.9 shows the cross validation result of similarity index evaluation through multi-component approach. Since the compositions of essential oil in CX and DG are phthalide components existed in majority, the influence of other components with respect to phthalide components on the overall similarity index is insignificant. Therefore, the performance of similarity index evaluations was not adequate on classifying CX and DG under multi-component approach.

Table 5.9: Results of the classification model based on several clustering analyses through multi-component approach

Methods of Clustering analysis	Successful identification (%)	
	CX	DG
Similarity index	84.38%	100%
PCA with Euclidean distance*	96.88%	100%
PCA with mahalanobis distance*	96.88%	100%
LDA with Euclidean distance*	96.88%	100%
LDA with mahalanobis distance*	96.88%	100%

* represent that k-NN (k = 5) method being validated in the classification model under five-foldss cross validation

The results of principal component analysis and linear discriminant analysis are also listed in Table 5.9 and it is obvious that the classification models performance of PCA and LDA were better than that of similarity index. An insight as to the common components extracted from chromatographic profile could assist in discrimination between CX and DG obtained by investigating the percentage of variance account on principal components (PCs) in PCA. With the first three principle components, 95.61% of the desired variance was extracted out and it also explains why almost CX and DG samples showed high percentage of correctly classified cases. Thus, the

results can be considered satisfactory and acceptable and the select variables, i.e. common constituents of CX and DG, were useful in differentiation of CX and DG by their chemical compositions variety. The graphical results of the classification model (groups 2 – 5 (training set); group 1 (testing set)) developed by PCA and LDA through multi-component approach are depicted in Fig 4.14 and Fig. 4.15, respectively.

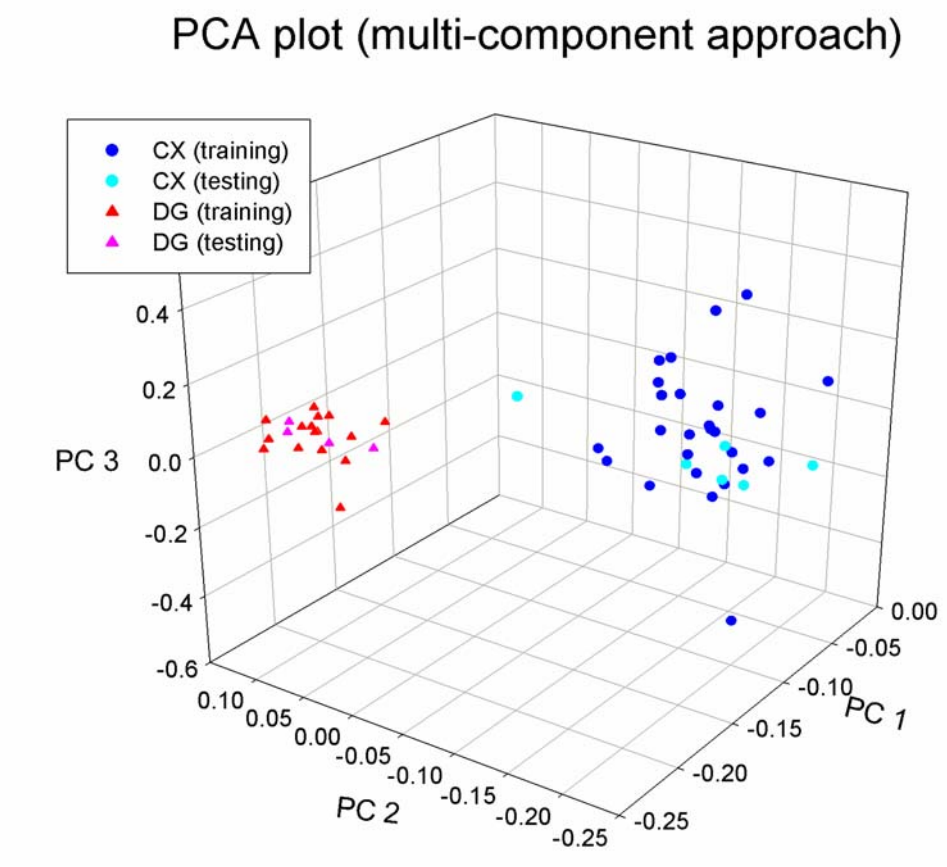


Fig 5.14 Three projection plots on the three principal components of extracted data from the 32 CX and 20 DG samples through the multi-component approach

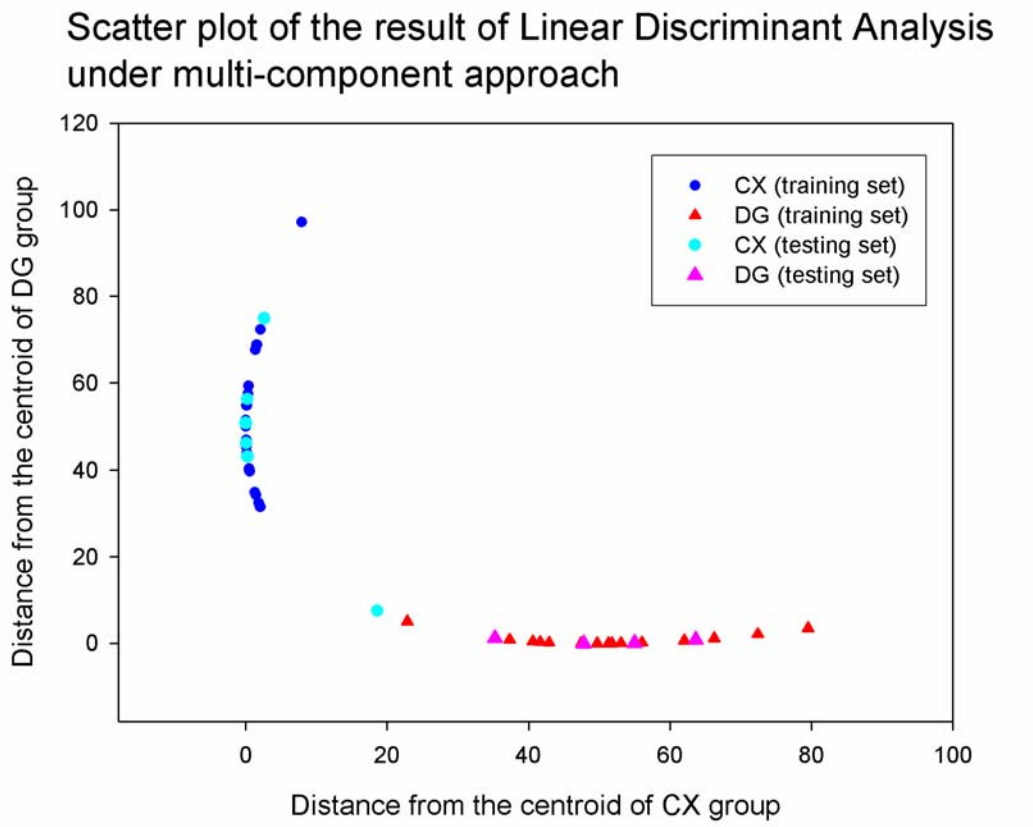


Fig 5.15 Distance from DG centroid vs Distance from CX centroid scatter plot of the result of linear discriminant analysis under multi-component approach

5.6.2.3 Pattern approach

In this investigation, retention time correction of the common peaks co-existed in CX and DG samples must be performed by local least squares (LLS) methods on each chromatogram of each CX and DG sample. Through the treatment of LLS, the aligned chromatograms were obtained. Afterward, normalization of all the chromatographic profiles of CX and DG have to be done before clustering analysis (see Section 4.5) in order to eliminate the concentration effect on TIC chromatograms.

Table 5.10 lists out the results of hierarchical agglomerative clustering analysis in

classifying DG and CX under the pattern approach. It was found that majority of these methods can be succeeded in classifying the two CHMs under pattern approach. City block metric gave the best measuring distance while average linkage, complete linkage and Ward's method achieved relatively well on the linkage of cluster. Fig. 5.16 shows the dendrogram plot of the CX and DG samples obtained from city block metric with the use of ward's linkage through the pattern approach.

Table 5.10: The results of hierarchical agglomerative clustering analysis on classifying DG and CX under the pattern approach

	Single linkage	Average linkage	Complete linkage	Ward's linkage
Euclidean distance	✓	×	✓	×
City block metric	✓	✓	✓	✓
Cosine distance	×	✓	✓	✓
Correlation distance	×	✓	✓	✓

✓ represents successful classification

× represents unsuccessful classification

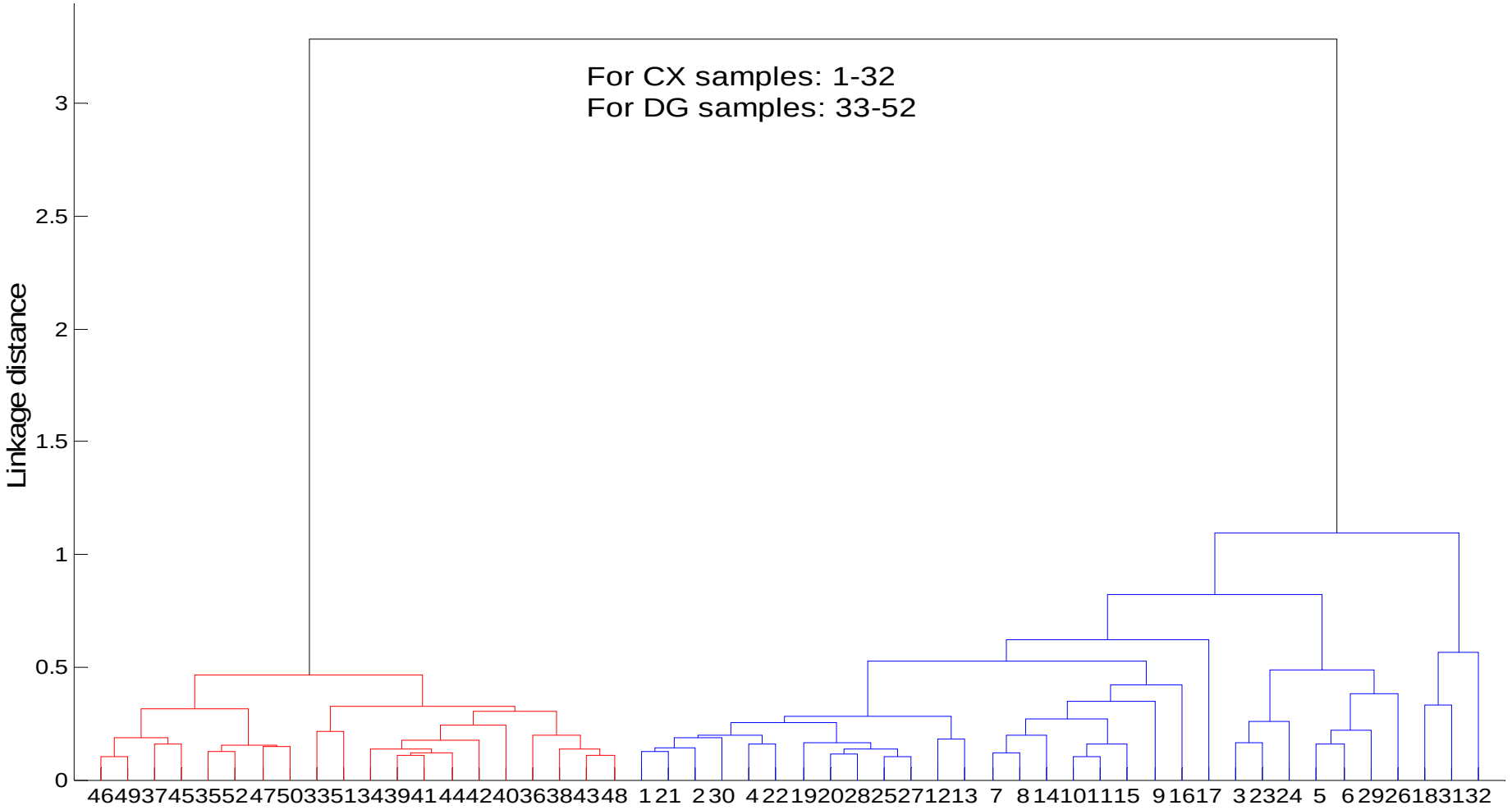


Fig. 5.16 Dendrogram plot of the CX and DG samples obtained from chromatographic profiles through pattern approach

Table 5.11 shows the cross validation result of similarity index evaluations through pattern approach. Since the composition differences of essential oil in CX and DG are ratios of mono-terpene and di-terpene components being existed, all of the chemical information are contributed into similarity index calculation so that the influence of such differences with respect to phthalide components on the overall similarity index becomes significant. This explains why the performance of similarity index under pattern approach was good and SI was feasible on classifying CX and DG.

Table 5.11: Results of the classification model based on several clustering analyses through pattern approach

Methods of Clustering analysis	Successful identification (%)	
	CX	DG
Similarity index evaluations	100%	100%
PCA with Euclidean distance	100%	100%
PCA with mahalanobis distance	100%	100%
LDA with Euclidean distance	100%	100%
LDA with mahalanobis distance	100%	100%

* represent that k-NN (k = 5) method being validated in the classification model under five-foldss cross validation

The results of principal component analysis and linear discriminant analysis are also summarized in Table 5.11 and the graphical results of the classification model (groups 2 – 5 (training set); group 1 (testing set)) developed by PCA and LDA through pattern approach are depicted in Fig 5.17 and Fig. 5.18, respectively. It is obvious that the classification models based on all these methods gave excellent performance with 100% successful identification. The existence of two different groups CX and DG that reflects the sample variation was clear. The correct classification thus obtained by the

use of entire chromatogram between CX and DG is in agreement with the statement that retaining all chemical information for data analysis could be a benefit in the quality control of CHM.

An insight as to the reason why the CX and DG chromatographic profiles succeed in discriminating these two CHMs can further be understood by investigating the percentage of variance obtained on principal components (PCs) in PCA. With the first three principle components, 89.73% of the desired variance was extracted out. Since all chemical information detected from GC-MS chromatograms are input the clustering analysis, almost composition and content differences between CX and DG are possibility projected into the first three PCs. It can also explain why all CX and DG samples s correctly classified in all cases. Then the results were very appreciable and illustrated as a good example that pattern approach is feasible in classifying CHMs with similar chemical compositions.

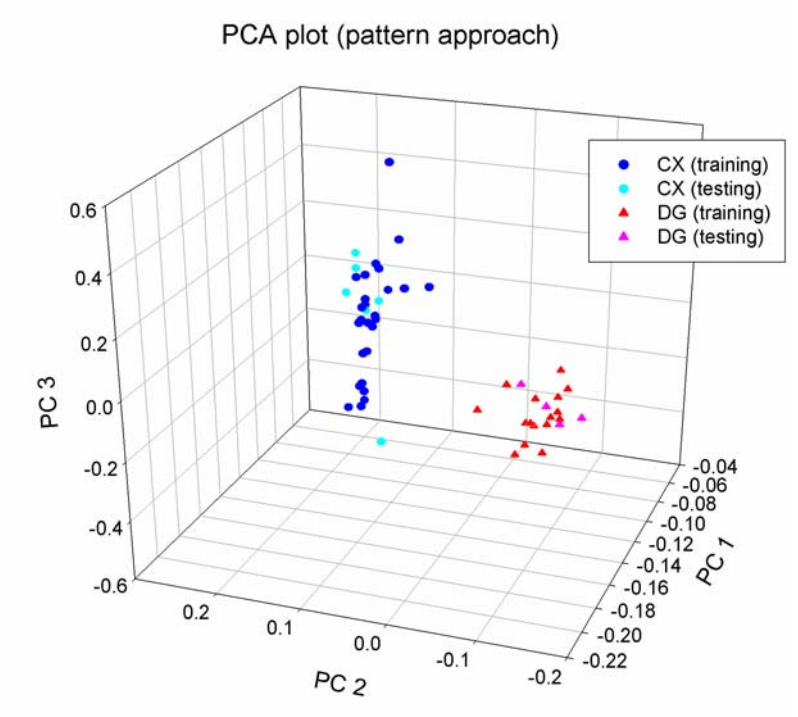


Fig 5.17 Three projection plots on the three principal components of the entire data from the 32 CX and 20 DG samples through pattern approach

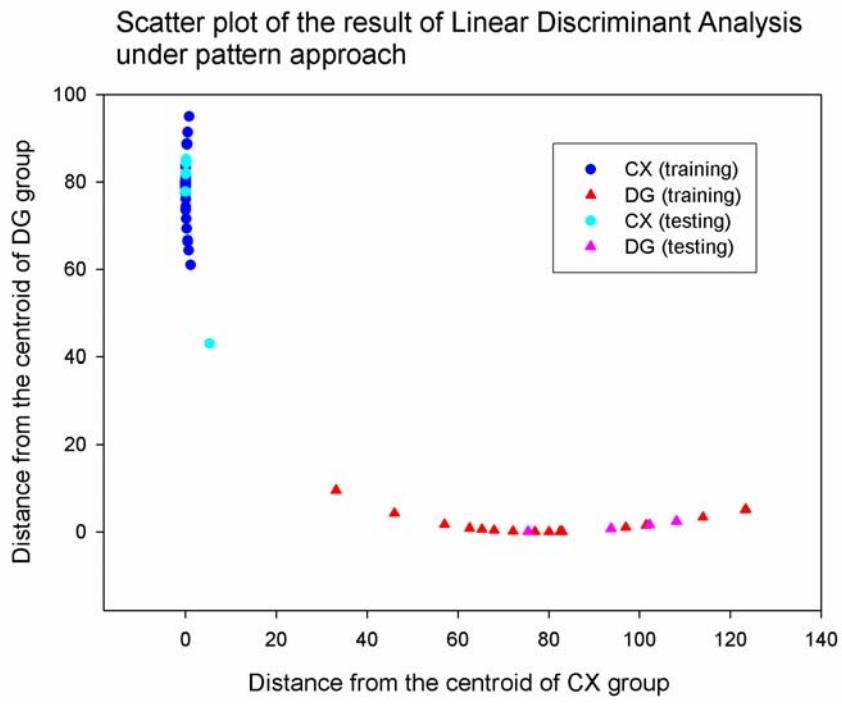


Fig 5.18 Distance from DG centroid vs Distance from CX centroid scatter plot of the result of linear discriminant analysis under pattern approach

In summary, ‘multi-component approach’ and ‘pattern approach’ were explored on the possibility in classifying CX and DG coupled with several multivariate analysis methods. At the same time, the disadvantage of ‘marker approach’ was also revealed from the results of classification modeling. It seems that using as much as chemical information for data analysis could be an advantage in the quality control of CHM. For further study on developing the classification model between CX and DG, a larger set of samples from different provinces and countries is certainly needed to come up an accurate and reliable model.

5.7 Conclusion

This work demonstrated that 20 *Radix Angelicae Sinensis* and 32 *Radix Ligustici chuanxiong* from different geographical sources have been classified, based on the three different data sets used through, ‘marker approach’, ‘multi-component approach’ and ‘pattern approach’. With the help of several chemometics techniques including Chemometrics Resolution Methods (CRM), Spectral correlative chromatography (SCC) and Local Least Squared (LLS) method, development of their chromatographic fingerprints have been achieved efficiently before building up the classification model. Similarity index evaluations, hierarchical clustering analysis (HCA), principal component analysis and linear discriminant analysis (LDA) have been utilized in developing the classification model which five-folds cross validation were tested in evaluating the overall model performance. The findings are that pattern approach is feasible on classifying CX and DG in which 100% correct identifications in the validation sets have been achieved.

The entire work in this chapter is considered to be a first attempt at the comparison on investigating how the nature of the chromatographic data set used on affects the performance of classifying CHMs with similar chemical compositions. It seems that authentication of CHM works successfully under the ‘pattern approach’ in this case. In the future, the whole methodology needs to be implemented on the methods of authentication in the other CHM samples. Computers and data processing algorithms are now powerful enough to allow us to process the entire chromatographic profiles. We believe that using all chemical information for data analysis could be a benefit in the quality control of CHM.

Chapter 6: Application of chromatographic fingerprint to the quality control of *Cortex Phellodendri*

6.1 Introduction

C*ortex Phellodendri* (HB) was selected and investigated to develop its chromatographic fingerprint in this study. It has been used as an anti-phlogistic, antibacterial, anti-inflammatory agent for the treatment of diarrhea, icterus, ulcer, carbuncle and eczema [78]. From Pharmacopoeia of the People's Republic of China (2000 edition) [8], *Cortex Phellodendri* is from either *Phellodendron chinense Schneid* or *Phellodendron amurense Rupr*.. Both species consist of similar active ingredient e.g. alkaloids like berberine, jatrorrhizine, palmatine etc, but in different amount. They are, therefore, different in pharmaceutical value. As a result, *cortex Phellodendri* has been separated into two entries in the latest edition of Pharmacopoeia of the People's Republic of China (2005 edition) [25].

A number of analytical studies for *cortex Phellodendri* including CE-DAD/MS, HPLC-DAD, LC-MS have been reported, they still used the marker approach to characterize the chemical composition differences between the two. In this work, a novel extraction method, with the use of information theory [72], and a new chromatographic condition for alkaloids were developed before developing its chromatographic fingerprint. Furthermore, the established chromatographic fingerprints were applied to detecting frauds through principle component analysis (PCA) and similarity index (SI). This is a first paper to use the whole chromatographic profile (pattern approach) for authentication of *cortex Phellodendri*.

6.2 Background of *Cortex Phellodendri*

Cortex Phellodendri, known as Huangbo (黃柏) in Chinese name, is one of the fifty fundamental herbs in Chinese herbalism. According to the Pharmacopoeia of the People's Republic of China, 2000 edition [8], it originates from the dried bark of *Phellodendron chinense* Schneid and *Phellodendron amurense* Rupr. The former is frequently called "Chuan huangbo (川黃柏)" while the latter is called "Guan huangbo (關黃柏)". They all belong to the genus Phellodendron in the family Rutaceae. They are harvested in early April. The bark is stripped off, dried in the sunlight, flattened and soaked thoroughly, then sliced or cut into shreds. It is used as raw or stir-baked with saline. The major sources of *Phellodendron chinense* Schneid are Sichuan (四川) and Guizhou (貴州). Other provinces like Hubei (湖北), Hunan (湖南) and Yunnan (雲南) also produce *P. chinense*. The major sources of *Phellodendron amurense* Rupr are Liaoning (遼寧) and Jilin (吉林). Other provinces like Heilongjiang (黑龍江), Inner Mongolia (內蒙古) also produce *P. amurense* [107].

The appearance of *Phellodendron chinense* Schneid is tabular or shallowly channeled with different length and width, but usually 2-6mm thick. Its outer surface is yellowish-brown in color, and even or longitudinally furrowed. Some have scars of lenticels, and remains of grayish-brown coarse bark. Its inner surface is yellow or yellowish-brown, with fine longitudinal ridges. It is light and hard in texture with fracture fibrous, showing lobe-like layers. It also has slight odor, very bitter in taste and viscous on chewing [98].

The appearance of *Phellodendron amurense* Rupr is tabular or shallowly channeled with different length and width, but usually 2-8mm thick. Its outer surface is brownish-yellow or pale brownish-yellow in color and is relatively even. It contains

remnants of grayish-brown coarse bark. Its inner surface is yellow or yellowish–brown, with longitudinal ridges. It has relatively hard and light texture with fracture fibrous and lobe-like layers. It also has slight odor, very bitter in taste and viscous on chewing [98].

The properties of *Cortex Phellodendri* are bitter in flavour, cold in nature. It acts on the kidney, bladder, large intestine channels. The bitter and cold properties also purge the heat so that it is capable of clearing away the damp-heat in the lower-jiao and purging away the ministrial fire. It is indicated for the syndromes of jaundice, dysentery and diarrhea due to damp-heat, and the disease caused by deficiency of kidney-yin and hyperactivity of the ministrial fire. The action of Cortex *Phellodendri* is effective in removing damp-heat (清濕熱), quenching fire (瀉火), counteracting toxicity (解毒), and relieving consumptive fever and night sweating due to exuberant fire secondary to deficiency of yin [77].

The indications of *Cortex Phellodendri* are given as follows and the main effects are for clearing heat, dampness and fire, and detoxicating [77].

1. For treating the disease of fever, night sweating, spermatorrhea caused by deficiency of kidney-yin, and hyperactivity of ministrial fire, it is utilized in combination with other herbs to nourish Yin and relieve fire.

2. For treating jaundice due to damp-heat, it is used in combination with other herbs to normalize the function of the gallbladder and clear the jaundice. In cases of diarrhea due to damp-heat in the large intestine, it is mixed with other herbs to

clear heat, and dampness, eliminate heat from the blood and stop diarrhea. For treating the disease of stranguria, profuse leucorrhea, pruritus vulvae and arthritis caused by the damp-heat in the lower-jiao, it works together with other herbs to remove the damp-heat in the lower-jiao.

3. For treating suppurative infections on the body surface, such as boils, sores, eczema caused by noxious heat, it is good to combine with other herbs for clearing away heat, detoxicating, removing dampness and relieving itching.

The antibacterial properties of *Cortex Phellodendri* are similar to that of *Rhizoma Coptidis*. Many of the pathogenic bacteria, such as *Shigella dysenteriae, Salmonella typhi, Staphylococcs aureaus*, etc. are inhibited by the herb as found by in vitro study. It is also effective against dermatophytes though its potency is less than that of the *Radix scutellariae*. It possesses the protective effect on blood platelets. When used externally, it enhances the absorption of subcutaneous purpure. Besides, the herb is cholagogic, diuretic, hypotensive and antipyretic though less potency than the *Rhizoma Coptidis* [77].

6.3 Literature review on chemical studies of *Cortex Phellodendri*

6.3.1 Chemical Composition

Cortex Phellodendron chinense contains compound like obacunone, limonin, γ-sitosterol and β-sitosterol and alkaloids such as berberine, magnoflorine and

phellodedrine [108]. One of the major alkaloids is berberine which should not be less than 3.0% according to the Pharmacopoeia of the People's Republic of China [25].

Cortex Phellodendron amurense contains compound like obacunone, obacuonic acid, dictamnolide, lumicaeruliec acid, campesterol 7-Dehdrostigmasterol and β-sitosterol [108]. Alkaloid is also present and it includes berberine of not less than 0.3% according to the Pharmacopoeia of the People's Republic of China [25], and other alkaloids including jatrorrhizine, palmatine, magnoflorine, phellodedrine, candicine, and others.

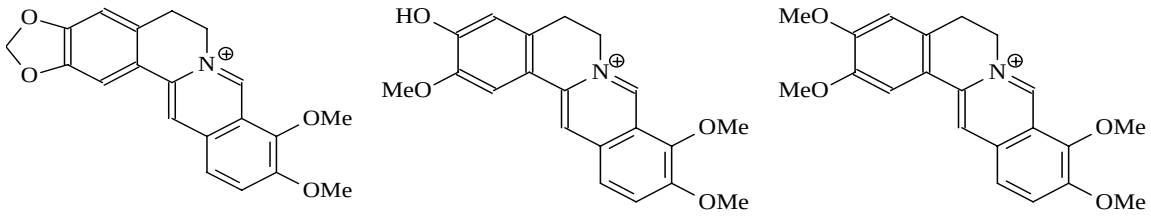
Chemical compositions of essential oil in *Cortex Phellodendri* have also been studied. *Phellodendron chinense* contain volatile oils like limonene, cis-limonene oxide, B-elemene and (+)–carvone [109]; *Phellodendron amurense* contains volatile oil of myrcene, β-citronellol and α-pinene [110]. Other compounds are also found such as kihadalactone A, B; nitloticin, dihydronilotcin, niloticin acetate, Piscidinol A, Hispidol B, bourjotinolone A, and Hispidone [111].

6.3.2 Marker components

As mention previously, one major class of chemical component found in *Cortex Phellodendri* is alkaloids. They are cyclic organic molecules containing nitrogen which are typically bitter in taste and in most cases toxic. They are usually derivatives of amino acids and are weak bases. Alkaloids can be classified in terms of pharmacological action, chemical property or biogenetic property. Also, Alkaloids like berberine, palmatine and jatrorrhizine, found in *Cortex Phellodendri* are

bis-benzylisoquinoline, which are usually classified as protoberberine alkaloid. It is a structural class of organic cations found in many plants, for example, the families Ranunculaceae (e. g., *Rhizoma Coptidis*), Berberidaceae (e. g., *Cortex Berberidis*, *Caulis Mahoniae*), Papaveraceae (e. g., *Herba Chelidonii*), and Rutaceae (e. g., *Cortex Phellodendri*) [112].

According to some studies, *P. chinense and P. amurense* contain different composition of alkaloids, in which palmatine and jatrorrhizine are found in *P. amurense* but not in *P. chinense*. In addition, the berberine content of *P. chinense* is higher than *P. amurense*. These differences are good indicators for the classification of the two species of *Cortex Phellodendri* with the use of these compounds as markers. Since, Berberine, palmatine and jatrorrhizine are commercially available, these three alkaloids are used as markers in this study. Their chemical structures are shown below:



(a) (b) (c)

Fig. 6.1: Chemical structures of (a) berberine, (b) jatrorrhizine, (c) palmatine

6.3.3 Pharmaceutical value of the marker components

Berberine is a major representative of the protoberberine alkaloids and its pharmacological actions has been studied extensively. It shows clinical efficacy in respiratory stimulation, transient hypotension, convulsion and has sedative activity as well as uterine stimulant activity. Berberine is antianaemic agent showing antibacterial

and antifungal propertoies. In clinical treatment of gastrointestinal disorders, berberine has been used for cholera and infantile diarrhea. Extracts of *Rhizoma Coptidis* exhibited in vitro antibacterial has been shown principally due to berberine [113].

Palmatine also has bactericidal and analgesic activity, and it may be a vasodilator. It shows antiarrhythmic, inotropic, adrenocorticotropic, anticholinestrase and analgesic effects in experimental animals. Jatrorrhizine is butyrylochlinesterinase inhibitor and antibacterial to *Mycobacterium smegmatis* [113, 114].

6.3.4 Chemical studies of *Cortex Phellodendri*

Different chromatographic and spectroscopic techniques have been reported for characterizing the chemical compositions of *Cortex Phellodendri*. A few chromatographic and electrophoretic studies on the alkaloids component have been studied, in which berberine, palmatine and jatrorrhizine were served as specific marker under HPLC-UV [115], HPLC-DAD [116, 117], CE-DAD [118], LC-MS [119] and CE-MS [120] methods. Li et al. created a novel method using H-NMR for quantitative determination of protoberberine alkaloids in *Phellodendron* species [121]. Lee et al. developed a quick clean-up process for qualitative determination of berberine and palmatine using cobalt thiocyanate reagent [122]. In addition, *Pharmacopoeia of the People's Republic of China* (2005) [25] and *The Hong Kong Chinese Materia Medica Standards (HKCMMS)* (2005) [98] both listed out the extraction method and chromatographic condition.

6.4 Experimental

6.4.1 Herbal samples

Fourteen *Phellodendron chinense* samples, five *Phellodendron amurense* authenticated samples were collected from different provinces in China (see Table 6.1). Eleven *Cortex Phellodendri* samples were purchased from different commercial brands in the local market and China crude drug market.

Table 6.1: A list of *Cortex Phellodendri* samples collected for this study

No.	Sample codes	Geographical sources	Class
1	CHB1	Hanyuan, Sichuan, China	*Phellodendron chinense*
2	CHB2	Lushan, Sichuan, China	*Phellodendron chinense*
3	CHB3	Nachong, Sichuan, China	*Phellodendron chinense*
4	CHB4	Yingjing, Sichuan, China	*Phellodendron chinense*
5	CHB5	Guanxian, Sichuan, China	*Phellodendron chinense*
6	CHB6	Pixian, Sichuan, China	*Phellodendron chinense*
7	CHB7	Yongshun, Hunan, China	*Phellodendron chinense*
8	CHB8	Sangzhi, Hunan, China	*Phellodendron chinense*
9	CHB9	Mayang, Hunan, China	*Phellodendron chinense*
10	CHB10	Renhuai, Guizhou, China	*Phellodendron chinense*
11	CHB11	Douyun, Guizhou, China	*Phellodendron chinense*
12	CHB12	Zhengan, Guizhou, China	*Phellodendron chinense*
13	CHB13	Fangxian, Hubei, China	*Phellodendron chinense*
14	CHB14	Nazhang, Hubei, China	*Phellodendron chinense*
15	GHB01	Heilongjiang, China	*Phellodendron amurense*
16	GHB02	Heilongjiang, China	*Phellodendron amurense*
17	GHB03	Jilin, China	*Phellodendron amurense*
18	GHB04	Liaoning, China	*Phellodendron amurense*
19	GHB05	Dongbei, China	*Phellodendron amurense*
20	CP1	Guangzhou drug market	*Unknown*

21	CP2	Guangzhou drug market	*Unknown*
22	CP3	Guangzhou drug market	*Unknown*
23	CP4	Guangzhou drug market	*Unknown*
24	CP5	Guangzhou drug market	*Unknown*
25	CP6	Hong Kong drug store	*Unknown*
26	CP7	Hong Kong drug store	*Unknown*
27	CP8	Hong Kong drug store	*Unknown*
28	CP9	Hong Kong drug store	*Unknown*
29	CP10	Hong Kong drug store	*Unknown*
30	CP11	Hong Kong drug store	*Unknown*

6.4.2 Chemical and reagents

Berberine chloride was purchased from Wako while jatrorrhizine chloride and palmatine chloride were purchased from International Laboratory. HPLC grade acetonitrile, methanol and were purchased from Tedia (USA). Tetrabutyl ammonium chloride was purchased from Aldrich. Phosphoric acid and hydrochloric acid were of analytical-reagent grade. Double deionized water was purified by Milli-Q water system (Millipore Corp., Bedford, MA, USA) and used to prepare all buffer and sample solutions

6.4.3 Apparatus and chromatographic conditions

Aglient 1100 series HPLC apparatus, equipped with a quaternary solvent delivery system, an auto-sampler and UV detector, was used. Chromatographic runs was carried out on a Therom ODS Hypersil column (250 X 4.6mm, 5.0μm) at a room temperature Detection wavelength was set at 346 nm. The mobile phase consisted of (A) 20mM Tetrabutylammonium chloride (TBA) at pH 2.2 and (B) acetonitrile (v/v)

using a gradient elution. The gradient program is shown as follows: initial 100% A at 0–30 min, linear change from 100% to 90% A at 30–42 min, keep 90% A at 42-55 min linear change from 90% to 70% A at 55-60min, keep 70% A at 60-72min. Re-equilibration duration was 10 min between individual runs. The flow rate was 0.7 ml/min and aliquots of 10µl were injected.

6.4.4 Sample preparation

6.4.4.1 Sample pre-treatment

All samples were ground in powder form and passed through 100 mesh stainless sieve. Then, the sieved powder was stored in polypropylene containers at 20°C.

6.4.4.2 Sample extraction

0.5g of pre-treated sample, accurately weighted, was added into a round-bottom flask containing 30 ml acidified (1%HCl (2M)) methanol. The mixture was heated by reflux 15 minutes. After reflux, the mixture was filtered and acidic methanol was vaporized to dryness through rotary evaporator. The residue was re-dissolved in 5ml methanol and was filtrated through a 0.45 μm syringe filter. Finally, the solution was diluted into four times and transferred into an HPLC vial prior to HPLC-UV analysis.

6.5 Results and Discussion

6.5.1 Optimization of extraction condition

An important task for chemists in quality control of CHM is to obtain a high quality chromatographic fingerprint which contain as much as chemical information. In this work, the influences of the extraction solvent, extraction method, extraction time and number of successive step for determining optimum extraction conditions on alkaloid were investigated. As mentioned above, concentrations of several markers including berberine chloride, jatrorrhizine chloride and palmatine chloride were served as assessment indicators to find out the best extraction condition. In addition, another criterion, information content (Φ) was utilized to optimize the extraction condition.

For information theory, the value of information content of a chromatogram depends on the separation degree and concentration distribution of each chemical component. The more the separated peaks with uniform concentration distribution in a chromatogram are, the higher the value of information content are. As a result, both separation degree and concentration distribution of components involved in a chromatographic fingerprint should be involved in this evaluation. There might be, at least, several advantages of calculating the information content. First, the whole chromatogram instead of the retention time, peak intensity, peak width, peak area and / or peak height of each peak are taken into consideration. Second, it is unnecessary to identify the peaks in chromatogram before hand. Also, the influence of noise in chromatogram on the value of information content is very small. Therefore, optimization of chromatographic profiles was done by finding out the value of information content through investigating the influences of the solvents used for extraction, extraction method, extraction duration and successive steps.

6.5.1.1 Optimization of extraction solvent

Methanol was often used for extracting alkaloid in previous reports. In this study, six different solvent systems, water, acidic water, 70% methanol, methanol, acidic methanol and ethanol, coupled with sample CP06 were selected and utilized in extracting alkaloid components in order to obtain optimal extraction medium. The actual amount among three alkaloids (area counts), the chromatographic profiles of six different solvents and the results for these two factors are shown in Fig 6.2, Fig 6.3 and Table 6.2. It is obvious that acidic methanol was the best extracting solvent medium.

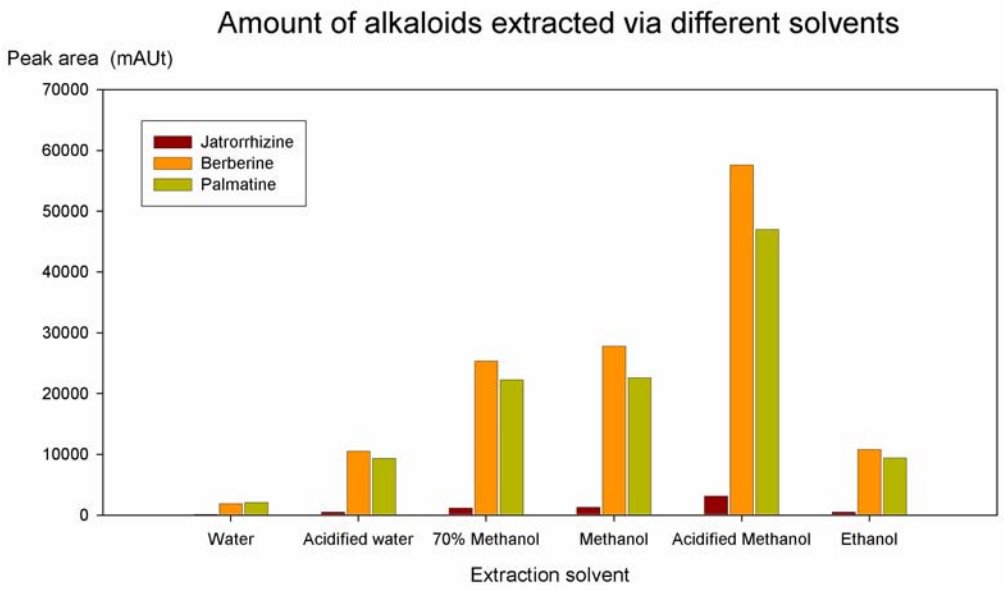


Fig 6.2 Comparison of the actual amount (area counts) of Jatrorrhizine, Berberine and Palmatine in different solvent systems

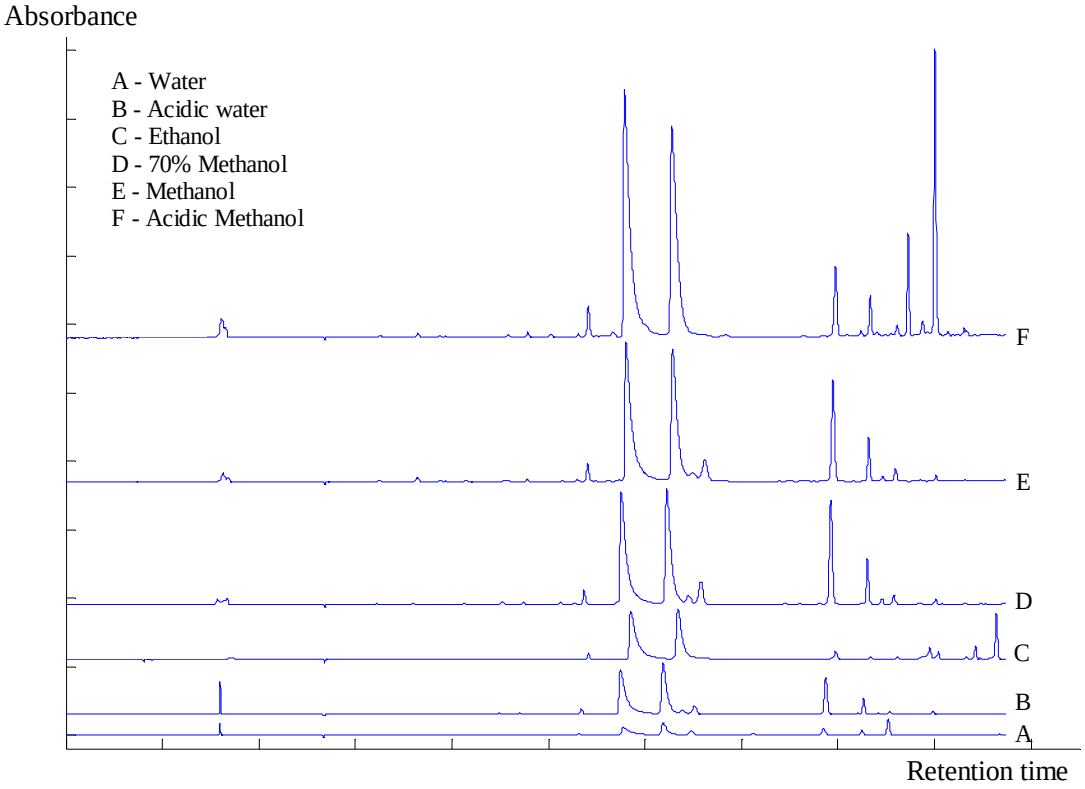


Fig.6.3: Comparison of chromatographic profiles obtained from six different extraction solvents.

6.5.1.2 Optimization of extraction method

Two different methods, ultrasonic and reflux were carried out with different extraction durations of 15, 30, 45, 120min. The results (see Table 6.2 and Fig. 6.4) showed that the reflux method was more efficient on extracting berberine and jatrorrhizine while ultrasonic method had high efficiency on extracting palmatine. Furthermore, successive steps on the extraction did not give a significant increase on the actual amount of alkaloid (Fig. 6.5 & 6.6). Since CHM often comprises a complex mixture of different phytochemicals and these ingredients work ‘synergistically’ for the therapeutic effect, determination of concentration of one or a few marker

concentrations does not give a complete picture of the herb. Therefore, we tried to explore how good the information content (Φ) served as indicators for determining an optimal extraction method. In summary, 15min reflux in acidic methanol provides the best extraction result on the alkaloid components.

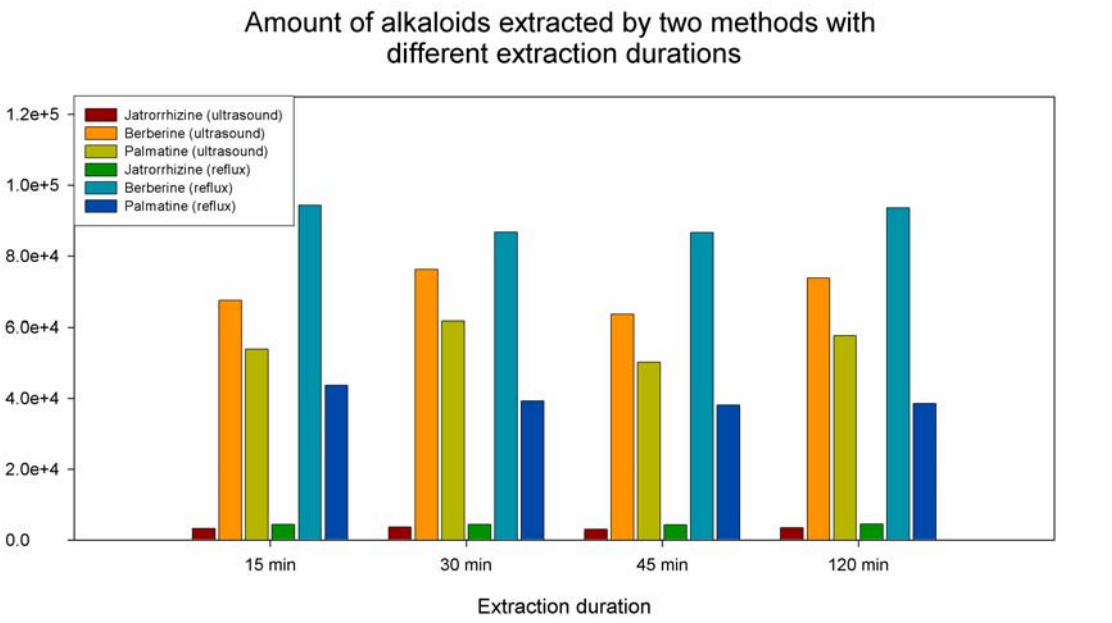


Fig 6.4 Comparison of the actual amount (area counts) of Jatrorrhizine, Berberine and Palmatine via different extraction methods with different durations

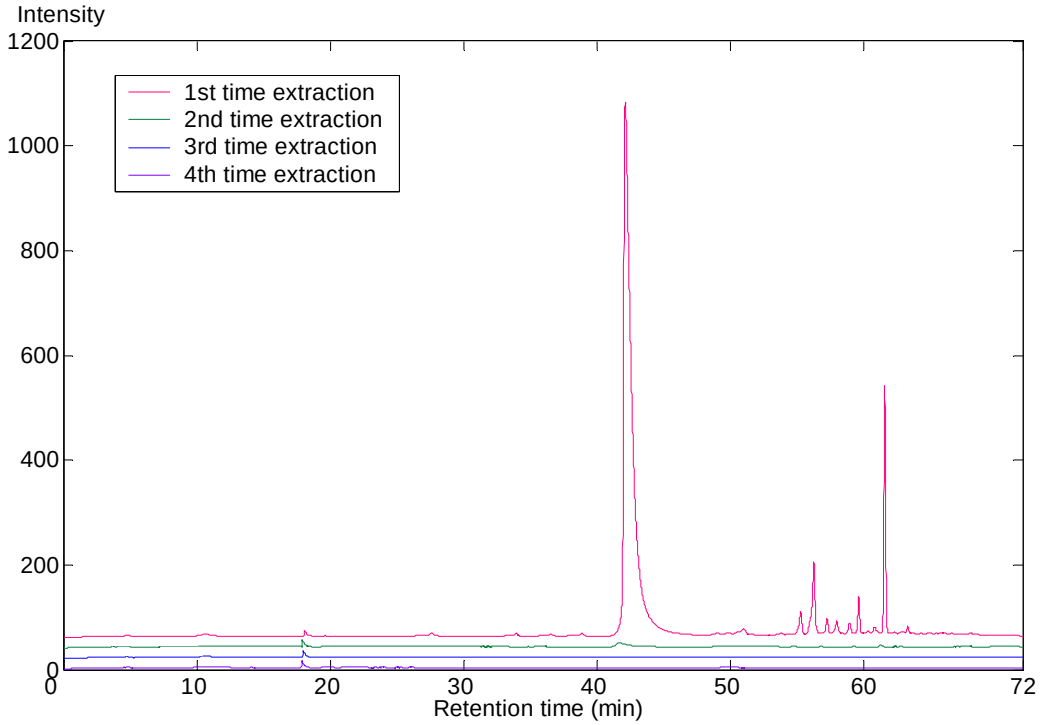


Fig 6.5 Overlap of HPLC-DAD chromatogram of *Cortex Phellodendron chinense* at 346nm for (—, 1st time extraction), (—, 2nd time extraction), (—,3rd time extraction) and (—, 4th time extraction)

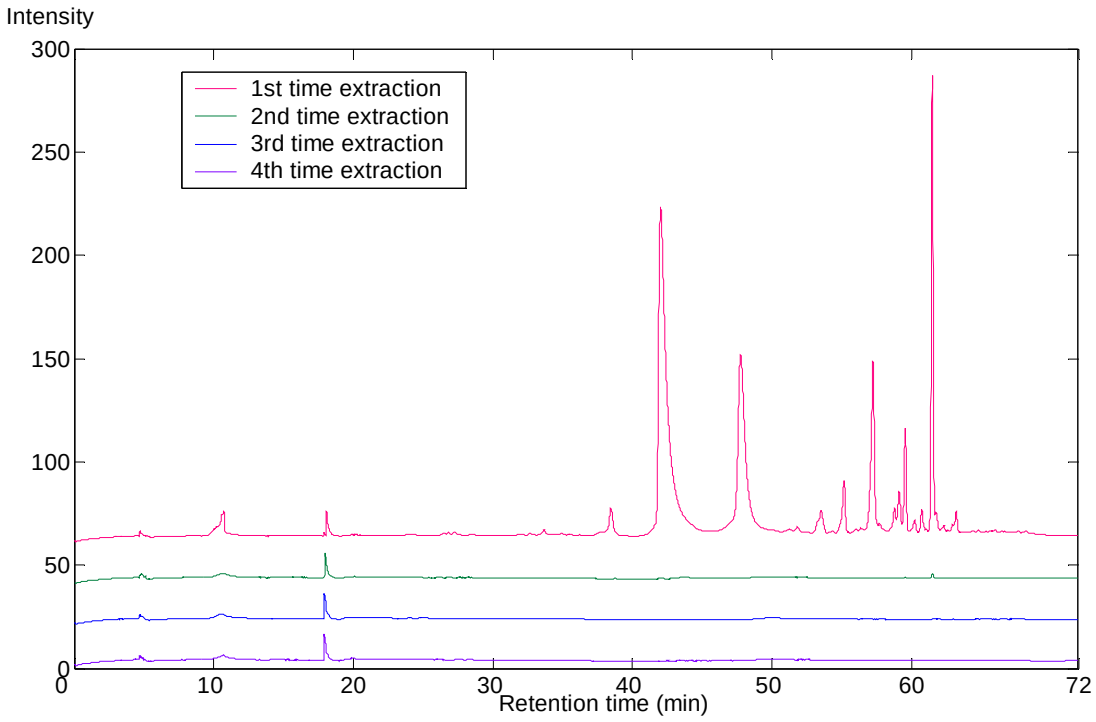


Fig 6.6 Overlap of HPLC-DAD chromatogram of *Cortex Phellodendron amurense* at

346nm for (—, 1st time extraction), (—, 2nd time extraction), (—,3rd time extraction) and (—, 4th time extraction)

Table 6.2: Effects of different extraction methods on the total alkaloids contents

Optimization of extraction medium				
	Area of Berberine (mAUt)	Area of jatrorrhizine (mAUt)	Area of palmatine (mAUt)	Information content (Φ)
Water	1895	112	2078	6.19
Acidic water	10521	4801	9375	6.61
Ethanol	10826	505	9435	6.73
70% Methanol	25357	1161	22286	6.95
Methanol	27831	1307	22620	6.78
Acidic Methanol	**57576**	**3131**	**46990**	**7.04**
Optimization of Extraction method (using acidic methanol)				
Reflux15	**94340**	4365	43645	**7.11**
Reflux30	86782	4441	39223	7.07
Reflux45	86635	4294	38066	7.05
Reflux120	93627	4465	38487	7.06
Ultrasonic15	67471	3198	53861	6.99
Ultrasonic30	76205	3709	**61721**	7.05
Ultrasonic45	63589	2994	50168	7.00
Ultrasonic120	73818	3443	57640	7.06

6.5.1.3 Optimization of chromatographic condition

The chromatographic conditions were optimized in order to obtain chromatograms with a good resolution of the analyte peaks within a short analysis time. Different mobile phase compositions were tried to optimize the alkaloid components and the best separation was achieved by adjusting mobile phase compositions, the flow-rates and gradient programs as given in Section 6.4.3. Also, two different types of chromatographic column, Lichrosorb RP and ODS Hypersil, were tested and compared in order to achieve maximum retention and separation of the alkaloid components. Both columns showed similar retention behavior and the analysis time did not vary significantly. Since an ODS Hypersil can perform better and less damage under low pH environment, an ODS Hypersil was used for developing chromatographic analysis throughout this work.

6.5.2 Samples analysis by marker approach

Fig. 6.7 represents typical chromatograms of three alkaloid components in standard solution, extract of *Phellodendron chinense* and extract of *Phellodendron amurense*. Under the experimental conditions, peaks in the chromatograms were identified by comparing the retention times and UV spectra with those of the standards. Retention time for compounds 1,2 and 3 were 38.24, 41.94, 47.41 min, respectively (see Fig. 6.7). The content of each analyte was calculated from the corresponding calibration curve (see Table 6.3) and their corresponding UV spectra of alkaloids components were shown in Fig. 6.8.

Based on standardization of berberine content from *Pharmacopoeia of the People's*

Republic of China (2005 edition) [25], its actual content in *Phellodendron chinense* and *Phellodendron amurense* are 3.0% and 0.3%, respectively. Table 6.4 clearly proved that all our 19 authenticated samples achieved to the reference standard and their chromatographic profiles were shown in Fig.6.9. Furthermore, all *Phellodendron amurense* samples contained three alkaloids while all samples composed of berberine only. It is consistent with the past literature reports.

Table 6.3: Calibration curves for the three alkaloids in *Cortex Phellodendri*

Components	Calibration curve*	r^2	Test range (μg)	LOQ(ng/mL)	LOD(ng/mL)
Berberine	y= 2778.9x - 44.702	0.9999	0.5-40	286	102
Jatrorrhizine	y=3141.2x + 29.865	0.9999	0.04-16	52	17
Palmatine	y= 2931.9x - 193.92	0.9991	0.5-32	142	43
* y: Peak area; x: actual amount (μg)					

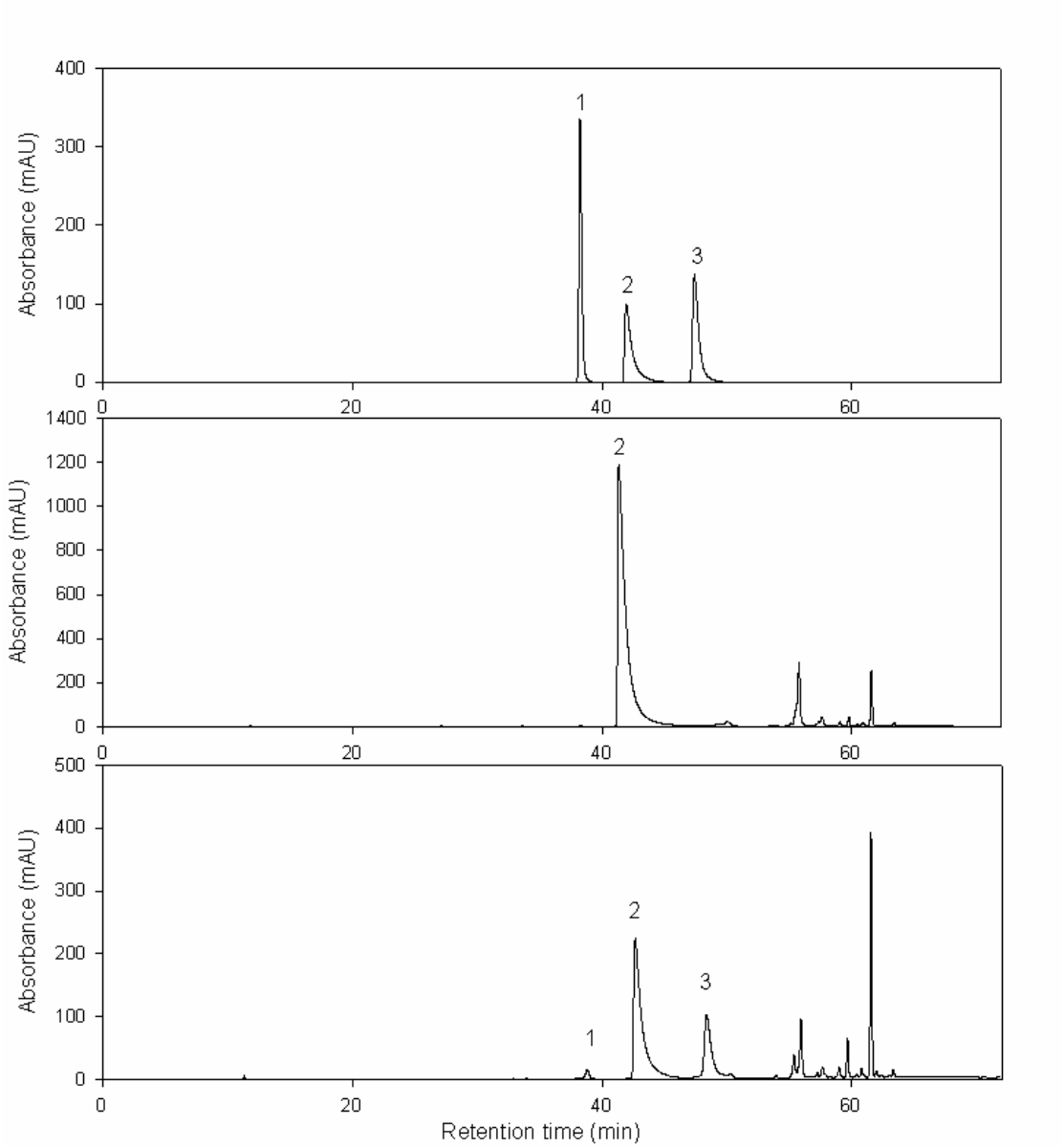


Fig. 6.7 HPLC/UV chromatograms of the three alkaloid components in the standard solution (upper, sample CHB1), extract of *Cortex Phellodendron chinense* (middle, sample GHB1) and extract of *Cortex Phellodendron amurense* (lower): (1) jatrorrhizine; (2) berberine; (3) palmatine

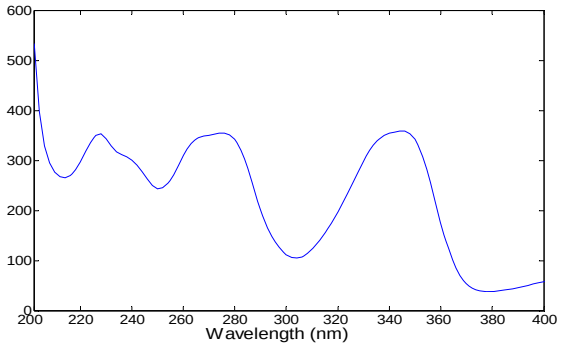


Fig 6.8a: UV spectrum of jatrorrhizine

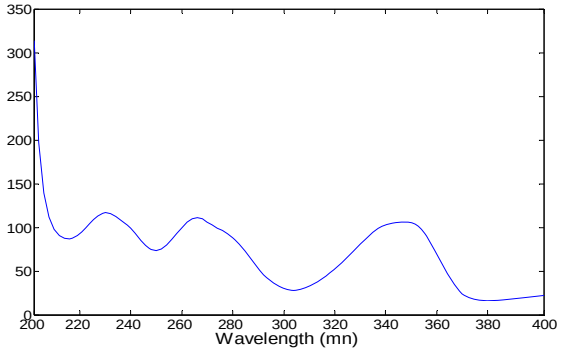


Fig 6.8b: UV spectrum of berberine

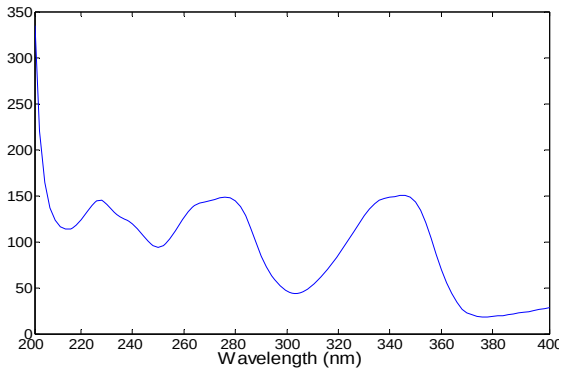


Fig. 6.8c: UV spectrum of palmatine

Fig. 6.8 UV spectra of a) jatrorrhizine, b) berberine & c) palmatine

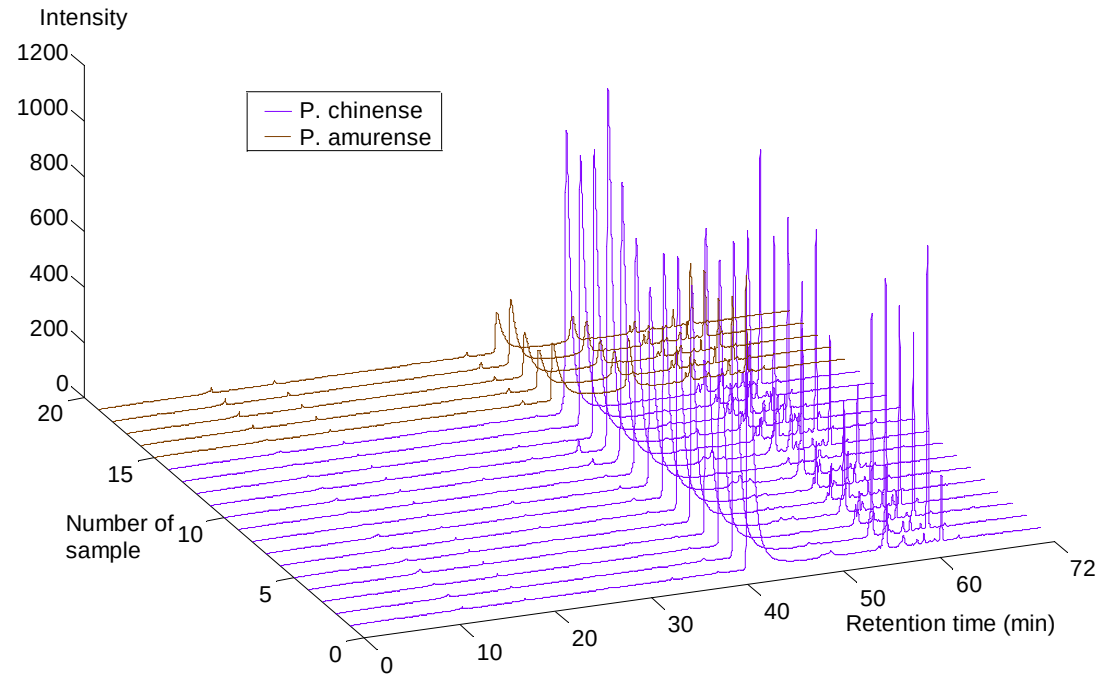


Fig. 6.9: The nineteen chromatographic profiles of *cortex Phellodendri*

6.5.3 Method Validation

6.5.3.1 Calibration curves

Standard solutions of three alkaloids were prepared and diluted to the appropriate concentrations for the construction of calibration curves. Seven concentrations of each of the three alkaloids were analyzed in five replicate injections. The calibration curves were constructed by plotting the peak areas versus the actual amount (μg) of each alkaloid. The results were given in Table 6.3. Under the chromatographic conditions used in this study, all calibration curves of the three standards gave good linearity ($r^2 >$ 0.999) within the test ranges.

6.5.3.2 Limit of detection and quantification

The limit of detection and (LOD) and quantification (LOQ) under the chromatographic conditions were determined at signal-to-noise ratio (S/N) of 3 and 10, respectively, for each compounds. LOD and LOQ were reported in Table 6.3 for each compound. The LOD and LOQ of all analyte were in the range 43-102 ng/mL and 52-286 ng/mL, respectively.

6.5.3.3 Precision and accuracy

Measurement of intra- and inter-day variability was utilized to determine the repeatability of our measurements and working procedures. The intra-day repeatability was examined on five replicate injections within one day while the inter-day repeatability was determined for three consecutive days. The relative standard derivation (R.S.D.) was calculated as a measurement of method repeatability and determined by the formula:

$$\text{RSD (\%)} = \left(\frac{\text{SD}}{\text{mean}}\right) \times 100\% \qquad \text{(Eq. 6.1)}$$

Recovery tests were utilized to determine the accuracy of the developed assay. Accurate amounts of three standard alkaloids were added to approximate 0.5g pre-treated *Phellodendron chinense and Phellodendron amurense* samples with respectively, and then extracted and analyzed as described in section 6.4.4.2. The recoveries were calculated by the formula:

$$\text{Recovery (\%)} = \frac{\text{(amount found - original amount)}}{\text{amount spliked}} \times 100\% \qquad \text{(Eq. 6.2)}$$

As the results given in Tables 6.4, 6.5, 6.6 and Fig. 6.10, the analytical method developed by us had good accuracy with overall recovery high than 100% ± 7.5% for all three analytes concerned in *Phellodendron chinense and Phellodendron amurense* samples. The overall intra- and inter-day variations of berberine and palmatine in almost samples were less than 4%. However, the overall intra- and inter-day of jatrorrhizine was relatively large (<8%). It was because the actual content of jatrorrhizine in almost samples was significantly small. Stability was test with both standard and sample extract that stored at room temperature within 2 day. The analytes were found to be rather stable. So far a conclusion could be generalized that this assay showed good reproducibility.

Table 6.4: Contents of alkaloids in different *Cortex Phellodendri* samples (Calibration set) (n=3)

	berberine				**Jatrorrhizine**				**Palmatine**			
	intra-day		**inter-day**		**intra-day**		**inter-day**		**intra-day**		**inter-day**	
sample	Actual amount(mg/g)	RSD(%)	Actual amount(mg/g)	RSD(%)	Actual amount(mg/g)	RSD(%)	Actual amount(mg/g)	RSD(%)	Actual amount(mg/g)	RSD(%)	Actual amount(mg/g)	RSD(%)
CHB01	78.20 ± 0.73	0.94	83.09 ± 3.79	4.57								
CHB02	71.92 ± 1.13	1.58	71.73 ± 1.63	2.28								
CHB03	62.56 ± 0.98	1.58	66.06 ± 3.18	5.47								
CHB04	67.24 ± 0.76	1.13	67.09 ± 1.12	1.66								
CHB05	50.54 ± 0.77	1.53	50.81 ± 0.77	1.52								
CHB06	56.76 ± 0.62	1.09	57.23 ± 1.31	2.28								
CHB07	48.44 ± 0.64	1.31	49.05 ± 0.76	1.55								
CHB08	55.54 ± 0.74	1.34	53.86 ± 1.96	3.64								
CHB09	63.67 ± 0.86	1.36	63.63 ± 0.92	1.45								
CHB10	72.15 ± 1.08	1.50	73.88 ± 1.54	2.08								
CHB11	64.63 ± 0.64	0.99	66.04 ± 1.36	2.06								
CHB12	64.67 ± 0.84	1.30	66.43 ± 1.46	2.20								
CHB13	65.80 ± 1.07	1.63	66.49 ± 0.90	1.35								
CHB14	61.13 ± 0.11	0.17	60.70 ± 0.72	1.19								
GHB01	14.08 ± 0.30	2.11	13.96 ± 0.36	2.61	0.30 ± 0.01	3.54	0.28 ± 0.02	5.84	5.23 ± 0.12	2.30	5.20 ± 0.17	3.25
GHB02	8.63 ± 0.07	0.81	8.38 ± 0.28	3.34	0.29 ± 0.02	6.10	0.28 ± 0.03	6.27	4.87 ± 0.07	1.37	4.80 ± 0.13	2.74
GHB03	9.34 ± 0.04	0.40	9.28 ± 0.18	1.95	0.24 ± 0.01	5.91	0.24 ± 0.01	5.12	4.72 ± 0.07	1.55	5.06 ± 0.27	5.35
GHB04	13.69 ± 0.20	1.47	13.69 ± 0.24	1.73	0.41 ± 0.01	2.81	0.40 ± 0.02	4.95	5.54 ± 0.09	1.59	5.47 ± 0.19	3.42
GHB05	8.63 ± 0.07	0.81	8.38 ± 0.31	3.72	0.24 ± 0.01	3.18	0.24 ± 0.01	5.21	4.43 ± 0.07	1.57	4.59 ± 0.18	3.89

Table 6.5: Contents of alkaloids in different *Cortex Phellodendri* samples (testing set) (n=3)

	berberine				**Jatrorrhizine**				**Palmatine**			
	intra-day		**inter-day**		**intra-day**		**inter-day**		**intra-day**		**inter-day**	
sample	Actual amount(mg/g)	RSD(%)	Actual amount(mg/g)	RSD(%)	Actual amount(mg/g)	RSD(%)	Actual amount(mg/g)	RSD(%)	Actual amount(mg/g)	RSD(%)	Actual amount(mg/g)	RSD(%)
CP01	56.03 ± 0.69	1.22	56.34 ± 0.68	1.21								
CP02	8.72 ± 0.17	1.96	8.53 ± 0.26	3.03	0.23 ± 0.01	3.20	0.23 ± 0.01	5.66	4.80 ± 0.08	1.75	4.85 ± 0.16	3.31
CP03	12.86 ± 0.17	1.36	12.72 ± 0.23	1.84	0.30 ± 0.02	5.42	0.31 ± 0.02	5.29	5.16 ± 0.09	1.78	5.22 ± 0.12	2.36
CP04	12.82 ± 0.12	0.95	12.40 ± 0.36	2.91	0.32 ± 0.01	1.82	0.32 ± 0.01	4.44	6.55 ± 0.04	0.67	6.36 ± 0.18	2.77
CP05	11.74 ± 0.11	0.91	11.50 ± 0.22	1.89	0.42 ± 0.02	5.78	0.41 ± 0.02	5.46	6.53 ± 0.26	4.02	6.88 ± 0.33	4.78
CP06	7.90 ± 0.13	1.63	7.75 ± 0.20	2.61	0.26 ± 0.02	7.90	0.27 ± 0.01	4.96	3.50 ± 0.12	3.49	3.50 ± 0.13	3.76
CP07	39.54 ± 0.09	0.22	39.99 ± 0.69	1.74	0.27 ± 0.01	2.51	0.28 ± 0.01	4.39	2.77 ± 0.07	2.44	2.81 ± 0.14	4.87
CP08	4.15 ± 0.16	3.94	4.21 ± 0.15	3.59	0.17 ± 0.01	4.36	0.17 ± 0.01	4.40	2.79 ± 0.10	3.56	2.83 ± 0.10	3.59
CP09	10.71 ± 0.11	1.01	10.53 ± 0.22	2.07	0.27 ± 0.01	2.28	0.27 ± 0.01	2.49	8.22 ± 0.42	5.12	8.12 ± 0.30	3.69
CP10	60.15 ± 0.86	1.43	60.02 ± 0.79	1.31								
CP11	6.83 ± 0.12	1.72	6.87 ± 0.13	1.86	0.22 ± 0.01	6.74	0.22 ± 0.01	5.64	6.72 ± 0.08	1.26	6.66 ± 0.17	2.59

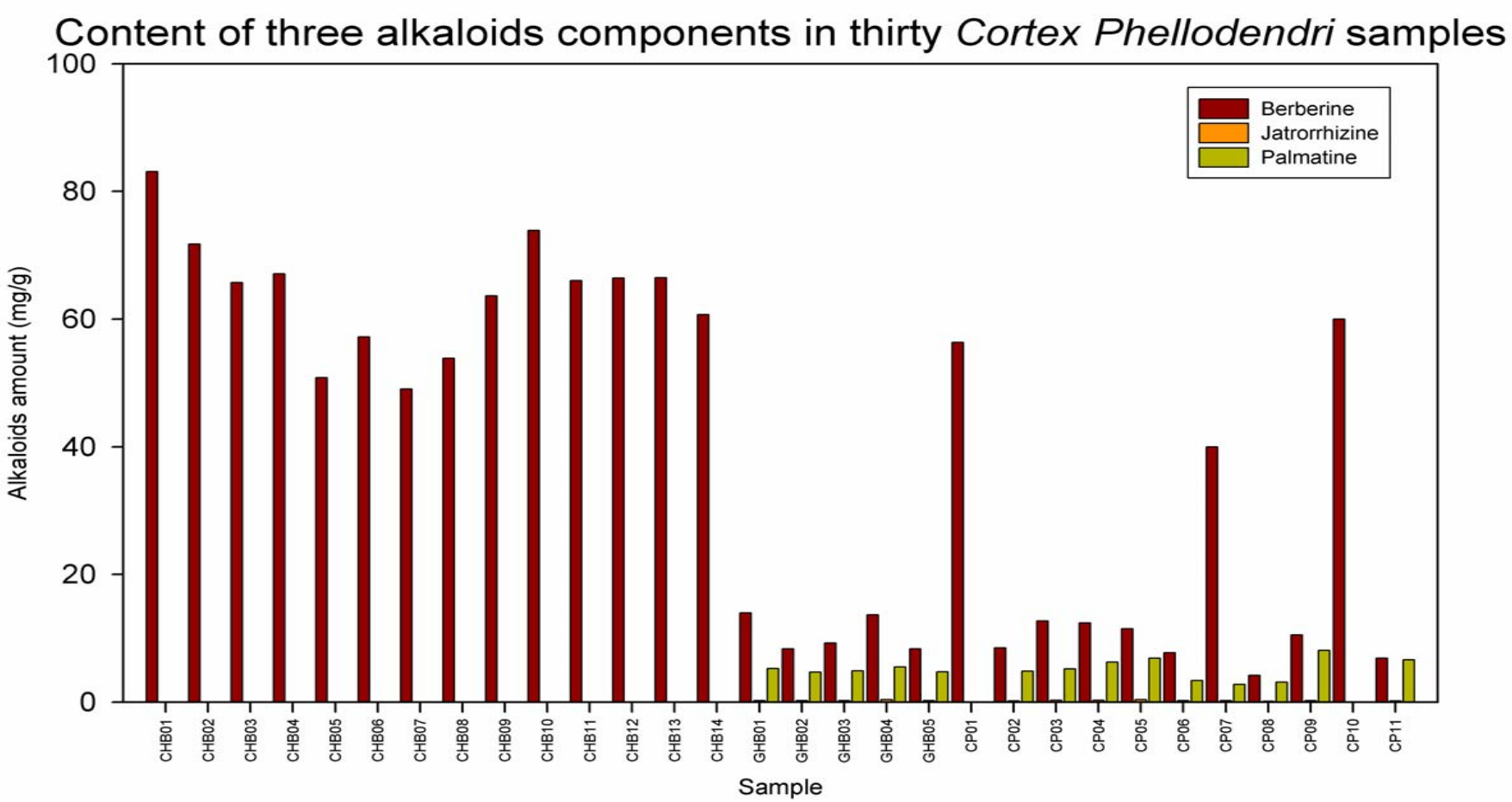


Fig. 6.10: Content of the three alkaloids components in thirty *Cortex Phellodendri* samples

Table 6.6a: Recoveries of the three alkaloids in *Phellodendron chinense* (n=4)						
Analyte	Original (mg/g)	Spiked (mg/g)	Found (mg/g)	Recovery (%)	Mean (%)	R.S.D. (%)
Berberine	50.542	3.693	54.175	98.38	100.02	7.35%
			54.001	93.63		
			54.533	108.05		
Table 6.6b: Recoveries of the three alkaloids in *Phellodendron amurense* (n=4)						
Analyte	Original (mg/g)	Spiked (mg/g)	Found (mg/g)	Recovery (%)	Mean (%)	R.S.D. (%)
Jatrorrhizine	0.205	0.959	1.219	105.75	103.86	3.25
			1.220	105.87		
			1.163	99.97		
Berberine	8.581	1.857	10.393	97.59	104.13	7.29
			10.669	112.46		
			10.481	102.33		
Palmatine	4.246	2.075	6.203	94.32	101.96	6.71
			6.476	107.50		
			6.405	104.05		

Table 6.6: Recoveries of three alkaloids in a) *Phellodendron chinense*, b) *Phellodendron amurense*

6.5.4 Samples assessment under fingerprint approach

Before construction of the chromatographic fingerprint in this investigation, it is necessary to correct the retention time shift on each chromatogram concerned by the local least squared (LLS) methods as the multivariate analysis with the whole chromatographic fingerprint are very sensitive to the retention time shift. The LLS-corrected or the aligned chromatograms were obtained using algorithm mentioned in Chapter 3.

To visualize the chemical differences between *Phellodendron chinense* and *Phellodendron amurense*, principal component analysis (PCA) and similarity index (SI) was applied to reflect their differences. In this work, the entire chromatographic profiles of all authenticated samples with LLS treatment were processed and or evaluated by PCA and SI. Then, eleven samples of *Cortex Phellodendri* bought from Hong Kong drug store and Guangzhou drug market were assessed by PCA and SI results to find out which one is *Phellodendron chinense* and or which one is *Phellodendron amurense*. The results of actual amount among three alkaloids for these eleven samples were found and shown in Table 6.5 and Fig. 6.10.

The SI and PCA results were given in table 6.6 and Fig. 6.11, respectively. It indicated that each group of authenticated samples is similar to its own group from its corresponding SI mean value with standard deviation. In addition, SI value showed that their differences of chemical composition between two groups were large while PCA plot also indicated that two clusters corresponding to *Phellodendron chinense* (red circle, ●) and *Phellodendron amurense* (green triangle, ▲), were well-separated. It seems that there is no difference on the results between marker approach and fingerprint approach. However, we used one validated sample, CP07 as typical representative to illustrate that fingerprint approach performs better than marker approach and is feasible on real application.

Sample CP07 contain three alkaloids with 3.954% berberine content (see table 6.5 and Fig. 6.10). From the previous mentioned, *Phellodendron chinense* must contain at least 3% berberine content while *Phellodendron amurense* must compose of three

alkaloid components. It is ambiguous that it belongs to which type of *Cortex Phellodendri* under marker approach. However, PAC plot and SI value (shown in Table 6.7 and Fig. 6.11), under fingerprint approach, can be able to distinguish CP07 sample as *Phellodendron chinense* type. Therefore, fingerprint approach offers a powerful for authentication of CHMs

Table 6.7a: A summary of the results of similarity index assessment among all the samples in the calibration set based on pattern approach

Calibration set	*Mean chromatogram of Phellodendron chinense*	*Mean chromatogram of Phellodendron amurense*
CHB1	0.9742	0.7852
CHB2	0.9921	0.8759
CHB3	0.9969	0.8438
CHB4	0.9924	0.8573
CHB5	0.9852	0.8663
CHB6	0.9984	0.8538
CHB7	0.9692	0.7745
CHB8	0.9903	0.8435
CHB9	0.9943	0.8565
CHB10	0.9896	0.8122
CHB11	0.9923	0.8450
CHB12	0.9910	0.8395
CHB13	0.9917	0.8627
CHB14	0.9800	0.877
GHB1	0.7964	0.9792
GHB2	0.8210	0.9945
GHB3	0.8651	0.9899
GHB4	0.9013	0.9873
GHB5	0.8263	0.989
Phellodendron chinense	0.9884 ± 0.0084^{a} (n=14)	0.9880 ± 0.0056^{a} (n=5)

Phellodendron amurense	0.8424 ± 0.0312[b]	0.8420 ± 0.0413[b]

[a] Similarity index of each chromatogram to the mean chromatogram of its own group, mean ± SD

[b] Similarity index of each chromatogram to the mean chromatograms of the other group, mean ± SD

Table 6.7b: A summary of the results of similarity index assessment based on pattern approach among all the samples in the testing set

Validation set	*Mean chromatogram of Phellodendron chinense*	*Mean chromatogram of Phellodendron amurense*
CP1	0.9565	0.8178
CP2	0.7937	0.973
CP3	0.9041	0.977
CP4	0.7754	0.9647
CP5	0.8387	0.9845
CP6	0.805	0.9494
CP7	0.982	0.8978
CP8	0.6441	0.9248
CP9	0.6115	0.8949
CP10	0.9959	0.8494
CP11	0.5317	0.8652

Table 6.7: A summary of the results of similarity index assessment based on pattern approach among all the samples in the a) calibration set, b) validation set

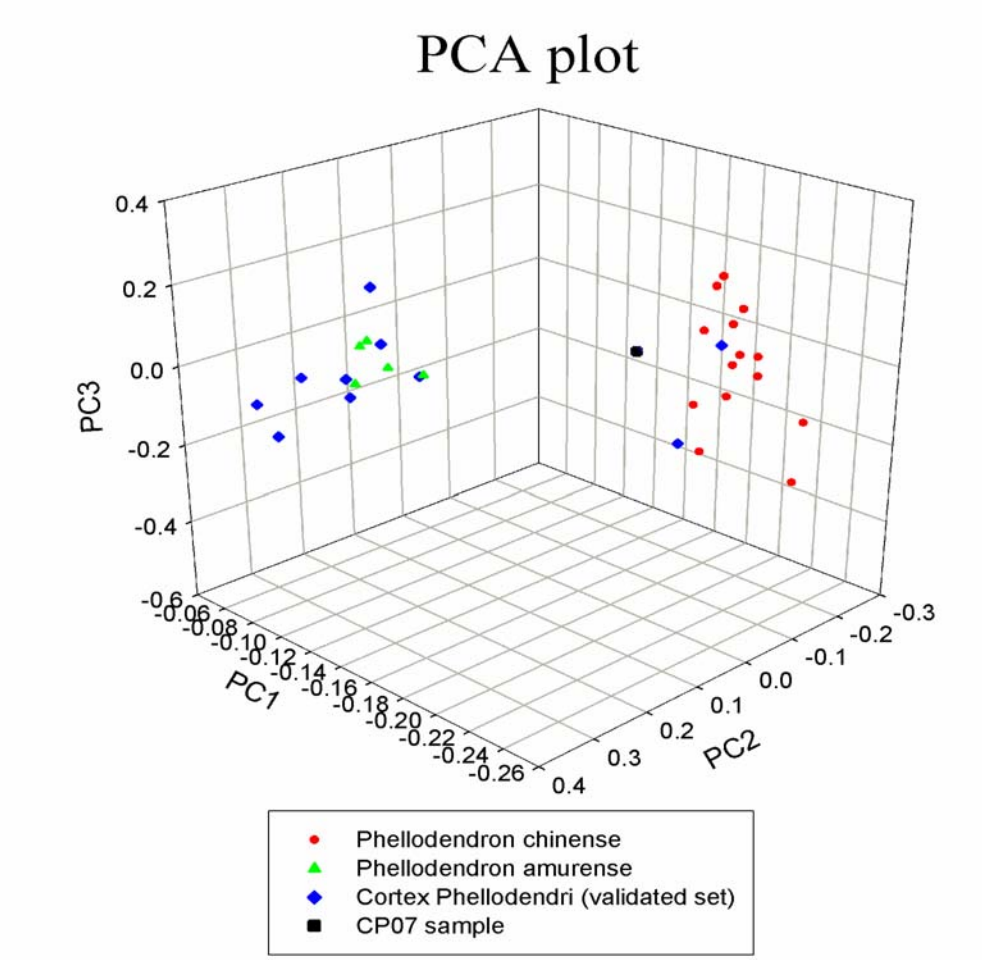


Fig. 6.11: Three projection plots on the three principal components of data from the *Cortex Phellodendri* samples using the pattern approach

6.6 Conclusion

Pattern approach, proposed as a strategy for quality control of Chinese herbal medicine, is demonstrated in developing the chromatographic fingerprint of *Cortex Phellodendri* in this study. Chemometrics tool was also found playing an important role under this approach in order to eliminate the instrumental disturbance and improve the efficiency of experimental work. Our work also shows that chromatographic fingerprint could show the promising prospect for chemists to deal with the difficult problems in quality control of CHMs. Finally, we believe that using all chemical information for data analysis could be an advantage in providing better quality control of CHMs.

Chapter 7: Review of near-infrared reflectance spectroscopy and its application in analyzing Chinese Herbal Medicine

7.1 Introduction

7.1.1 Background

Near Infrared Reflectance Spectroscopy (NIRRS) is a new technique involves the multi-discipline approach of the analytical chemist, statistician, and computer programmer in the past decade. Knowledge of reflectance and optical spectroscopy are also the essential in the advanced application of this technology. The methodology of NIRRS contains certain data processing techniques not often addressed in those common spectroscopic applications. Hydrogen bondings dominate the near infrared (NIR) spectrum while interaction and non-linearities give rise to nearly total disregard for the tenants of Beer's Law. Its applications are quite different from those of the well-known ultraviolet / visible (UV / Vis) and mid-infrared (MIR), with significant on statistics and chemometrics which is a term applied to the generic discipline involving computer and mathematics to derive meaningful chemical information from samples of varying complexity, to develop method and get results. It is important for NIR users to familiarize themselves with basic concepts from these disciplines. In order to introduce the NIRRS technique to readers, the first section will provide a basic knowledge of the methodology of NIRRS and diffuse reflectance measurements in the NIR region. The later section will give a general application of NIRRS in Chinese herbal medicine

7.1.2 Principle of Near Infrared spectroscopy

NIR region was discovered by Herschel in 1800. It lies between the visible and

mid-infrared regions of the electromagnetic spectrum and is defined by the American Society for Testing Materials (ASTM) as the spectral region spanning the wavelength range of 780-2526nm (or in the wavenumber range of 12820-3959 cm^{-1}). Light absorption in this region is primarily due to overtones and combinations of fundamental vibration bands occurring in the MIR region [123]. For an infrared ray to be adsorbed, its frequency has to be exactly the same as the fundamental vibration frequency of the molecule interested. Also the molecule should undergo a change in its dipole moment by virtue of its fundamental vibration.

The vibrational frequency f of a diatomic molecule can be determined via the assumption of the harmonic oscillator model (Hooke's Law), where an atom shifts from its equilibrium position with a strength proportional to the shift [124].

$$f = \frac{1}{2\pi c}\sqrt{\frac{k}{\mu}} \qquad \text{(Eq. 7.1)}$$

Here, c is the speed of light, k is the bonding force constant, a measure of the rigidity of a chemical bond in its normal equilibrium position, and μ is the reduced mass. With regard to the relationship of potential energy with respect to the bond distance, it is represented by a parabola centred about the equilibrium distance with evenly spaced vibrational energy levels and its energy E_v of each level is given by the following equation [124].

$$E_v = f\left(v + \frac{1}{2}\right) \qquad \text{(Eq. 7.2)}$$

where v is the vibrational quantum number

For harmonic oscillator, the selection rule of vibrational transitions is $\Delta v = \pm 1$. The vibrational energy levels are evenly spaced and the energy difference between two successive levels is always given by $E_{v+1} - E_v = f$, which is called the 'fundamental

frequency' of the band. In fact, vibrations in polyatomic molecules involve complex movements of their constituent atoms and the number of energy levels becomes numerous. To the first approximation, a molecule acts as a series of diatomic treating with independent harmonic oscillators in the normal coordinate mode. The equation for this case can be generalized as the follows [124].

$$E(v_1,v_2,v_3,...) = \sum_{i=1}^{3N-6} (v_i + 1/2)hv \qquad \text{(Eq. 7.3)}$$

$$(v_1,v_2,v_3,... = 0,1,2,...)$$

In any case with the transition from v = 0 to 1 in any one of the vibrational states $(v_1,v_2,v_3,...)$, the transition is considered as a fundamental. As for the transitions from the ground state $(v_i = 0)$ to $(v_i = 2,3,...)$ with all others v_i being zero, it is known as an overtone. Transitions from the ground state to a state with which $v_i = 1$ and $v_j = 1$ simultaneously are known as combination bands. In the harmonic oscillator model, only fundamental transition is allowed while overtones and some combinations transition are not allowed. However, the later one does appear as weak bands due to the reality of anharmonicity in the real molecule.

In reality, molecular vibrations tend to be anharmonic, i.e., vibrations about the equilibrium position are non-symmetric. The potential energy curve for a real bond is only roughly parabolic. The respective electron clouds of the two bound atoms within a molecule, as well as the same charges on the nuclei limit the approach of the nuclei closed to the equilibrium bond distance. Thus, an energy barrier is created while the bond eventually breaks when the respective vibrational energy level reaches the dissociation energy during the bond extension. Furthermore, energy differences between two consecutive energy levels are not identical but rather decrease with

increasing energy. In order to improve the agreement between theoretical and experimental data, harmonic oscillator model must be corrected by including additional terms of higher order. Hence, the energy E_v for each level is now given as [124]:

$$E_v = w_e(v+\tfrac{1}{2}) - w_e x_e(v+\tfrac{1}{2})^2 + \text{higher-order terms} \qquad \text{(Eq. 7.4)}$$

where x_e the an-harmonicity constant and w_e is the harmonic frequency

Usually, the quadratic term in the above equation is sufficient to account for the observed transition energies for levels with low quantum numbers. One consequence of this anharmonicity is that it is no longer true that only the selection rule fundamental $\Delta v = \pm 1$ applies and it is possible to observe transitions with overtones $\Delta v = \pm 1, \pm 2$, etc. Hence, in addition to the fundamental band ($\Delta v = \pm 1$) observed in the MIR range, other, higher frequency transitions called overtones or harmonics appear at energy position approximately two, three, or more higher than the fundamental frequency. The intensities of these higher energy bands decay abruptly because the transition probability decreases markedly with increase in Δv. In practice, each fundamental band only exhibits its first two or three overtones.

For the case of polyatomic molecules, it possesses several fundamental frequencies so they may exhibit simultaneous energy changes two or more vibrational modes. The energy observed will be the sum of individual fundamental frequencies ($f_1 + f_2, 2f_1 + f_2, etc.$) or their difference between ($f_1 - f_2, 2f_1 - f_2, etc.$) and the corresponding infrared profiles are called combination and subtraction bands respectively. All these bands have low intensities. Furthermore, anharmonicity results in combination bands are smaller than those combined fundamental frequencies involved.

Many NIR bands caused by the overtones or combination bands for hydrogen bonds such as C-H, N-H, O-H and S-H (The small mass of hydrogen and large force constants for bonds containing hydrogen are the origin of the high fundamental frequencies in this atom) while C=O, C-C, C-F and C-Cl groups usually exhibit very weak or even no bands observed in the NIR region (fundamental vibrations of these groups occur at low frequencies in the MIR region). Therefore, the former group consisting of the first few overtones usually appears in the NIR region while the latter group consisting of first few overtones appears in the MIR region. Owing to lower transition probabilities, the intensities usually decrease by a factor of 10-100 for each step from the fundamental to the next overtone. Thus, the intensities of absorption bands decrease from the MIR to the visible region and broad absorption bands are present in the NIR spectrum because of overlapping of various overtones and combination bands.

7.1.3 Fundamental of NIR Diffuse Reflectance Spectroscopy

Although the low molar absorptivity of absorption bands in the NIR region (typical between 0.01 and 0.1 $mol^{-1}cm^{-1}$) severely limits its sensitivity, it permits operation in the reflectance mode and recording of spectra for solid samples. For reflectance spectroscopy, it measures the light reflected by the sample surface, which contains a specular component and a diffuse component [125, 126]. Specular reflectance, described by Fresnel's law, contains little information about composition; consequently, its contribution to measurements is minimized by adjusting the detector's position relative to the sample. On the other hand, diffuse reflectance,

which consists of more chemical information can be described by the Kubelka-Munk (KM) theory, which is the basis for measurements by this technique [124].

When an NIR beam incidents on a powdery material of a weakly-absorbing medium, with certain thickness in order to prevent transmission, it will penetrate the layer and its direction will be changed each time as a result of reflection, refraction and random diffraction at the surfaces of various particle boundaries. Combination of these effects is called light scattering. Light propagates through a diffusely reflecting material by scattering. Since scattered light encounters more boundaries of particles, further scattering occurs in all directions and part of it is absorbed, which diminishes its intensity. Scattering and absorption take place simultaneously in the layer until finally the remaining attenuated light re-emerges from the entry surface; this light is called diffuse reflectance [124]. Diffuse reflectance highly depends on the particulate nature of the medium and on the effective depth of penetration to provide a spectrum with representative information of the entire sample.

In past centuries, many theories such as Lambert cosine law, Mie theory and Schuster's theory that have been developed to describe the diffuse reflection of radiation evolved from a general radiation transfer equation [124, 127]. The KM theory has been most successfully utilized in NIRRS with simplified solution to the radiation transfer equation. Kubelka & Munk have several simplifying assumptions of which the particles in the layer are randomly distributed so that the scattered radiation is iso-tropically distributed and the particle layer is very much smaller than the thickness of the layer so that the layer is subject only to diffused reflection [124]. Through these assumptions, the theory derives the relationship,

$$f(R_\infty) = \frac{(1-R_\infty)^2}{2R_\infty} = \frac{k}{s} \qquad \text{(Eq. 7.5)}$$

where f(R) is the KM function, and R_∞ is the absolute diffuse reflectance of the sample. In addition k is its absorption coefficient and s is its dispersion coefficient.

From the KM equation, it is seen that the reflectance, which is measurable, is a function only of the ratio of two constants, k and s, and not of their absolute values. In NIRRS measurement, the relative reflectance, rather than the absolute diffuse reflectance is usually measured.

In practice, relative reflectance (R), which is the ratio of the intensity of the light reflected by the sample to that by a standard, is measured. In general, the standard for NIRRS is a material that does not absorb any NIR at any wavelength but reflects light at an angle identical with the incidence angle. However, no single material can meet these requirements, the standards used in NIRS must be a stable, homogeneous and non-transparent material with a high to fairly constant absolute reflectance. Barium sulfate, magnesium oxide, teflon and ceramic plate are the common standards used in NIRRS [128].

The KM equation can be re-written in terms of the relative reflectance and the concentration of the absorbing analyte (c):

$$f(R_\infty) = \frac{(1-R_\infty)^2}{2R_\infty} = \frac{k}{s} = \frac{\varepsilon c \ln 10}{s} = \frac{c}{a} \qquad \text{(Eq. 7.6)}$$

where ε is the molar absorptivity and $a = \frac{s}{2.303\varepsilon}$

Thus a plot of f(R) against c for sample conforming to this relationship will be linear

with a straight line with slope 1/a.

The KM equation is usually applied to diffuse reflection spectra of dilute dispersions of absorbing material in a non-absorbing powdered matrix. In case, the matrix absorbs or the analyte exhibits strong absorption bands, the diffuse reflectance of the sample will not fit the KM equation and the $f(R_\infty)$ plot concentration tends to be non-linear. Since the KM equation is only applicable to weak absorption bands, or when the product of absorptivity time concentration is small, this is true only within the NIR region. If the matrix frequently absorbs strongly at the same wavelength as the analyte, absorption by the latter cannot be resolved and deviations from the previous equation becomes significant [126].

In practical, one widely used alternative is a direct relationship between concentration and relative reflectance similar to that of the Beer's law:

$$A = \log\frac{1}{R} = a' c \qquad \text{(Eq. 7.7)}$$

where A is apparent absorbance, R is relative reflectance, c is analyst concentration and a' is a proportionality constant.

Even though there is no theoretical basis on the KM equation, it provides highly satisfactory results and uses in many diffused reflectance spectroscopic applications [126].

7.1.4 Mathematical processing of NIR signals

NIR spectroscopic theory does not have to assume a linear relationship between the

optical data and constituent concentrations, as data transformation or pretreatment is used to linearize the reflectance data. This entails convert recorded data into apparent absorbance values [A=log(1/R)] or KM units when the measurements are performed in the reflectance mode, and into absorbance units [A=log(1/T)] in the transmission mode.

When an NIR ray incidents on the powdered sample, three different phenomenons have to be considered in diffuse reflectance NIR spectroscopy. They are the multicollinearity among variables, light scattering and particle size. The multicollinearity of the variables is typical for spectroscopic data because of the continuous signal in the data. Some transformation methods are available to reduce the correlation between these variables. In addition, scattering occurs on the surface of a material and depends therefore on the physical nature of the material and the particle size. Interaction of the incident light and the medium occurs within the material, e.g. within the particles of the powder. Furthermore, the particle size defines the spectral path length and varying particle sizes resulting in a baseline shift in the spectra. Besides, some other factors may also influence the NIR spectrum, for example, the particle size distribution, the density of a powder and consequently the packing of the material inside a measuring cup, the moisture content of a material, the instrument itself and temperature. Multiplicative interferences of these effects are responsible for the off baseline shift, slope changes and curvilinearity observed in the spectra. As a result, the analytical signal obtained in NIR spectroscopy is a complex function that depends on not only both the physical and chemical properties of the sample, but also the measuring conditions, which may be irrelevant to the problem under study [128]. To exclude such undesired information in the data analysis, pre-processing the

spectrum must be done, in order to extract more relevant chemical information [129-131]. Certain transformation techniques are useful to remove baseline shifts, slope changes and curvilinearity of spectra, i.e. they reduce the physical effects of particle size, scattering and other influence factors.

In order to minimize the effects of those physical properties of the sample that influences an NIR spectrum and introduce variability that gives the irrelevant chemical information, an important decision is whether data pre-processing method (mathematical processing of the spectrum) is necessary and, if yes, which pre-processing method(s) is/are utilized. Selection of a suitable pre-processing method is therefore important in the method development. By application of suitable pre-processing methods, it is possible to minimize the contribution of physical effects to the NIR spectra for data treatment.

Some of the widely used mathematical treatments for scattering in NIR spectra include normalization [132], derivation [133-135], multiplicative scatter correction [136], piecewise multiplicative scatter correction [137, 138], extended multiplicative signal correction [139], optimized scaling [140, 141], Savitsky-Golay [142], standard normal variate transformation and detrending [143-145]. Here, derivation, multiplicative scatter correction, savitsky-golay, detrend and standard normal variate are briefly introduced in the following sections.

Derivation

The characteristic of the NIR reflectance spectrum is a rising baseline and relatively few bands with little fine structure. The baseline offset depends on the physical

properties of the sample (particle size). The spectral data are often preprocessed by calculating the second derivative of the spectra in order to minimize these effects. As a result, sloping baselines and offset can be eliminated, while differences depending on signal height are retained. Also, the fine structure of the spectrum is also improved. However, the derivation transformation has the disadvantage of enhancing the noise in the original spectrum so that the signal-to-noise ratio of the resulting signal is reduced.

The most usual way in getting derivative spectra is by using the segment-gap method. Firstly, the spectrum to be derived is split into segments (of 10 and 20nm for first- and second-derivative spectra, respectively) that are separated by a gap, and the mean absorbance in each segment is calculated. The first derivative, measure of the slope for absorbance data, is equal to the difference between the mean absorbance for adjacent segments while the second derivative, measure of slope change comes from the difference between consecutive first-derivative values. Application of the first derivative suppresses those terms that are constant at every wavelength while application of the second derivative cancels those that change by a constant amount at all wavelengths. In practice, second derivative spectra are much more frequently used than the first-derivative spectra and higher spectral orders outwardly have no added advantages with regard to the quality of the results.

Multiplicative scatter correction (MSC)

Multiplicative scatter correction is a preprocessing tool developed to correct for the significant light-scattering problems in reflectance spectroscopy. When MSC is used, it carried one assumption, that all samples have the same scatter coefficient at all NIR wavelength. For MSC, an “ideal” or reference spectrum is required and for that

purpose the average spectrum of the whole representative set is taken in order to estimate the scatter of the spectra. Each individual spectrum is regressed on the set-mean spectrum. The fit for the individual and the mean spectrum is achieved by least squares.

$$x_{i,k} = a_i + b_i r_k + e_{i,k} \quad \text{(Eq. 7.8)}$$

Where $x_{i,k}$ is individual spectrum (spectral value), r_k is the mean spectrum of the data set, $e_{i,k}$ is the residual spectrum, which ideally contains the chemical information the data, a_i is the fitted constants for offset and intercept, b_i is the fitted constants for slope

* a_i represents the "common shift" in the spectrum relative to the mean, which is related to proportional additive effect, b_i represents the "common amplification" of the spectrum, which is related to multiplicative effect, $e_{i,k}$ are the errors or residual spectrum, which are representative of the difference between $x_{i,k}$ and r_k and are mainly attributable of the chemical information.

Now, the spectrum processed by MSC is given as follows:

$$x_{i,k}^{MSC} = (x_{i,k} - a_i) / b_i \quad \text{(Eq. 7.9)}$$

Since the scattering and particle size are independent of chemical information, one often normally defines a sub-region of the spectrum, which represents explicitly the baseline and no chemical information. This sub-region is utilized to get the parameters a_i and b_i and they are applied to process the entire spectrum. One comparative advantages of MSC over the derivative methods is that the preprocessed spectrum

resembles the original spectrum, which aids in the interpretation.

In pattern recognition, MSC is typically applied to each class separately. This includes the determination of the corresponding ideal spectrum, i.e. the mean spectrum, and the definition of the correction terms for each class.

Savitzky-Golay (SG)

Savitzky-Golay is smoothing technique commonly used in NIRRS. In this algorithm, a moving window averaging method is utilized, i.e. a window is selected where the data are fitted by a polynomial (either second or three degrees polynomial). The central point in the window is replaced by the value of polynomial [146]. In general, it usually applies together with the derivation method.

Standard Normal Variate (SNV) transformation

This transformation first centers the spectral value ($x_{i,k}$), and then subtracts the mean of the individual spectrum ($\overline{x_i}$) from the spectral values obtained at each wavelength k. These centered values are then scaled by the standard deviation calculated from the individual spectrum values. The SNV transformed spectrum then becomes

$$x_{i,k}^{SNV} = (x_{i,k} - \overline{x_i}) / \sqrt{\frac{\sum (x_{i,k} - \overline{x_i})^2}{p-1}} \qquad \text{(Eq. 7.10)}$$

where p is the number of variables in the spectrum.

SNV is able to remove the multiplicative interferences of scatter and particle size. The comparative advantage of SNV over MSC is that the former can be applied to each individual spectrum without the presence of an “ideal” or reference spectrum. In general, SNV can be combined with de-trending in order to remove the curvilinearity

of spectra.

De-trending (DT)

De-trending is another baseline correction method that is able to remove offset and curvilinearity which occurs powdered, densely packed samples. This transformation corrects the spectral baseline on the basis of its nonlinearity. It models the spectral values ($x_{i,k}$) to a quadratic function ($\hat{x}_{i,k}$) as $\hat{x}_{i,k} = a + b \cdot k + c \cdot k^2$ and subtracts this function (quadratic baseline) from the spectral value by

$$x_{i,k}^{DT} = x_{i,k} - \hat{x}_{i,k} \qquad \text{(Eq. 7.11)}$$

This transformation also has the advantage that it can be applied to each individual spectrum, with no need for an “ideal” or a reference spectrum. Normally, de-trending is carried out in combination with SNV transformation.

7.1.5 Requirement of NIRRS technique

Powdered sample form is one of the easiest and most reproducible solids to measure by NIR in NIRRS technique. Samples whose particle size is small, whose variation in particle size is small, and can be compacted relatively uniformly, provide the most reproducible type of solid sample to measure. Powdered sample cells can be rotated while the NIR scan is collected so as to present a greater area of sample to the spectrometer, presenting more of an ‘average’ of the light being scattered from the sample. This type of approach becomes more necessary as sample morphology in order to achieve acceptable analytical results by NIR.

Since NIRRS is an inferential method, it requires a reliable reference method to be

developed and maintained. For example, when NIRRS is utilized in quantitative analysis, a reference method with good accuracy for developing the calibration model is a must. In NIRRS qualitative analysis, a suitable reference method for characterizing the chemical information to build up the classification model must be needed. Furthermore, the calibration / classification model of NIRRS requires typically more samples to adequately characterize thr various chemical information in order to develop a stable, accurate calibration / classification model. If only a few samples contain in the calibration/ classification set, the variability in the samples may not be good enough for characterizing the whole set of samples and the accuracy of the calibration / classification will be poor.

7.1.6 Advantages and disadvantages of NIRRS

NIRRS offers a number of important advantages over traditional chemical methods. It is a physical, non-destructive method, requires minimal or no sample preparation and its precision can be high. In contrast with traditional chemical analysis, no reagents are required and no wastes are produced. It is a multi-analytical technique that several measurements can be made simultaneously. The method offers the possibility of measuring physical and chemical properties [128]. Once calibrated, the NIR spectrometer is simple to operate. NIR instruments have a very high signal to noise ratio which is typically 10000:1. In addition, NIRRS also has the potential to be used for developing on-line methods, leading to time control systems. This advantage has been well used in the pharmaceutical or chemistry industry to give real-time information about processes [147].

The main disadvantage of NIRRS is its dependence on time-consuming and laborious calibration procedures and the complexity in the choice of data treatment. Much work has to be done to develop as NIRRS working procedure. Furthermore, NIRRS lacks robustness, i.e. the calibrations often need to be updated. The greatest disadvantage is probably its low sensitivity to minor constituents and impurities. The sensitivity limit is about 0.1% for most constituents.

There are also some minor disadvantages. First, in contrast to MIR spectra, NIR spectrum exhibits low specificity so that it do not show clear peaks characteristic to a specific compound of interest. Thus extensive statistical treatments are required to extract useful qualitative or quantitative information. Second, NIRRS is developed using the reference analysis results of the calibration samples. Thus, the accuracy of the NIR method cannot be better than the accuracy of the reference method. Another limitation may be the portability of calibrations between different instruments due to their variation in optical properties and others [148].

7.2 Application of NIRRS to study on Chinese Herbal Medicines

7.2.1 Qualitative Analysis: Identification and qualification of CHM

In order to ensure the identity of the CHMs, from the time CHM being collected to, develop suitable procedures for analyses have been implemented under quality control. NIRRS is an advantageous alternative to wet chemical methods and other instrumental techniques such as high performance liquid chromatography (HPLC), thin layer

chromatography (TLC) and gas chromatography (GC) because most of these analyzing CHM methods required sampling extraction.

However, identification of a CHM by the mere inspection of its NIR spectrum is usually difficult since it consists of broad and highly overlapped bands. With the use of statistical treatment for enhancing the characteristics of spectra under pattern recognition methods (PRMs), it is possible to develop an accurate identification process [63]. Essentially, the identification process involves two steps of recording a series of analytical signals for the real CHMs authenticated by another reference method(s) and also generating a so-called ‘true spectral library’ by archiving the sample signal and comparing it with those in the previously complied spectral library which has been set up on the basis of mathematical criteria for parametrizing spectral similarity. If the similarity level exceeds a pre-set threshold which can be adjusted manually, then the spectra are considered to be identical and the sample is identified with the corresponding CHM in the library.

In the field of chemistry, PRM has been used to classify different chemical compounds into different classes on the basis of their spectra or other properties as obtained from the analytical instruments such as HPLC-UV and GC-MS. Most PRMs rely on similarity measurements. Similarity is a common criterion with, which a spectrum of an unknown substance or chemical is used to identify with another one. In general, similarity is expressed quantitatively in terms of correlation or distance depending on whether the spectrum is known to belong to specific class or not. There are two types of PRMs called supervised and unsupervised methods [73].

Unsupervised methods search for clustering in an N-dimensional space without knowing the class to which the sample to be identified belongs. Cluster Analysis (CA), the minimal spanning tree (MST) and unsupervised neural networks (NNs) are the most commonly utilized in the PRMs unsupervised one. On the other hand, supervised methods rely on the prior training of the system using a set of objects belonging to specific, known classes. These methods can be of the discriminant or modelling type. Discriminant methods split the space pattern into as many regions as possible into classes included in the training set, thereby creating bounds that are shared by the spaces; Modelling methods create volumes in the pattern space in which each class possesses different boundaries. Such boundary can be established in the form of correlation coefficients, distances such as Euclidean Distance or Mahalanobis distance, and the residual variance or supervised artificial neural networks. Most often, the supervised method are discriminant analysis, supervised neural network and soft independent modeling of class analogy (SIMCA) [63]. However, one important thing must be mentioned that not all pattern recognition methods are suitable for constructing product libraries for identification.

Reliable identification of a product totally relies on correct spectra being used in the library. The spectra complied for each CHM should contain every possible source of variability associated with the spectra recording and the samples selected. In setting up NIR spectral library of a CHM, the two variabilities, spectra and sample, must be included. Spectral variability is considered by including spectra for the same sample acquired by different operators on different days while sample variability is regarded for different parts from different locations / sources with different collection durations. One of the big problem on CHM analysis is the sample variability as the chemical

composition of a CHM varies with parts, sources, collection periods and processing procedures. In additional, it is difficult to anticipate the exact number of spectra to be included in a ‘comprehensive’ library. In general, it is ideal to have a large amount of CHM samples available. Yet, it is almost impossible to meet to this demand. For a CHM only, it is necessary to collect at least 40 to 50 samples with different sources and collection time in order to cover the sample variability.

Finally, with the availability of the reliable spectra of several CHMs with samples collected from different geographical sources, it is possible to build up the NIR library for identification and classification of the CHMs concerned. Construction of a spectral library basically consists of two stages. A spectral library is composed of spectra of different CHMs referring to as individual type. Each CHM requires a training set, i.e. a set of spectra of the same CHM meeting the defined specifications. During the training stage, several training set spectra are recorded, and the average spectrum for each CHM is calculated from these spectra. The average spectra are then used to construct the spectral library and to train the algorithm to recognize unknown CHM. The efficiency of a library mainly depends on the good choice of a representative training set. Since a PRM can only recognize variability to which it is accustomed to, the training set must include examples of all expected sources of spectral variability. Both research planning and recording technique must ensure that account is taken of on both the expected physicochemical variation in the CHMs and the instrument-dependent uncertainty of measurement in getting of the spectra for the training set. As the packing density influences the depth of penetration of the radiation beam in the sample and, hence, the intensity of the absorption bands, repeated measurements are required.

During the test stage, the performance of a pattern recognition method for different CHMs in the library is evaluated. Internal validation, i.e. cross-validation, is performed to check whether the library can correctly identify samples of the training set first. External validation is then carried out with test set spectra not included in the library. Test set spectra are mathematically compared with the average spectrum of each CHM in the library, and any corresponding lack of fit is utilized to judge the identity of the sample. Apart from correct identifications, false rejection of acceptable samples (type I errors) and false acceptance of unacceptable samples (type II errors) must be analyzed [149]. External validation enables one to select a classification method that yields both satisfactory recognition and rejection rates. Remember that one other important consideration in building up a spectral library is to ensure that all the spectra included for each CHM must be corrected. Any uncontrolled factor such as incompletely filled cuvettes, voltage drops at the time of recording may result in spectral differences not ascribable to natural variability. Any such spectrum should be discarded [128].

In the past few years, there are several examples of successful application of NIRRS in analyzing CHM. For examples, Woo et al. reported that China *Ginseng Radix* and Korea *Ginseng Radix* as well as China *Ganoderma* and Korea *Ganoderma* have been geographically classified by using PCA under NIRRS [150, 151]. Woo et al. also discovered that Korea and Chinese *Angelicae gigantis Radix* have been clearly identified using SIMCA [152]. Laasonen et al. also developed a fast identification of Echinacea purpurea dried roots model under NIRRS [153].

7.2.2 Quantitative Analysis: Setting up the calibration curve for quantitative analysis of marker(s) in Chinese Herbal Medicines

NIR spectrum typically contains broad, highly overlapping bands that cannot always be ascribed to an individual sample component, especially in CHMs. As a result, whenever the NIRRS technique is applied for quantitative purposes - whether the physical or chemical property of the sample to be analyzed – a calibration curve must be set up by using an existing multivariate procedure. Essentially, the general process for quantification using multivariate calibration involves the following steps [154, 155]:

1) Selecting a representative sample set.
2) Acquiring the analytical signals and obtaining the reference values.
3) Processing of the signals using appropriate mathematical means.
4) Selecting the model that relates properly the property to be determined and the signals.
5) Validating the model.

The starting point for almost all calibration techniques of chemical analysis is a set of samples which has previously been analysed by a reference method with wide working concentration range and are representative of the variability sources that are bound to influence the NIR spectra. The reference methods for CHMs analysis are HPLC, GC, TLC and capillary electrophoresis (CE) coupled with ultra-visible or mass spectrometry detection. One of the difficulties encountered in using NIRRS for the quantitative analysis of CHMs is the need to establish a calibration model with wide working range in order that the equation obtained is reboost and reliable. As

mentioned before, chemical analysis of each CHM is a difficult task because it contains many components. It is not easy to set up a good chemical method for direct analysis. In addition, even though a CHM comes from the same species, the quality and efficacy varies with growing conditions and geographical origin. Furthermore, the chemical composition of a CHM fluctuates with the collection period or the parts of CHM. On the other hand, the CHM sample set must be as large as possible in order to get a wide range calibration curve. Yet, in reality, it is difficult to collect the sample with high variability.

Once the calibration set has been established, then it is divided into two sub-sets of the calibration set and the prediction set as mentioned before. The calibration set consists of a small number of samples that are representative of the entire set and allows the determinant to be related to the analytical measurement, and the prediction set contains of the remaining data samples for assessing how good the predictive ability of the established model is. How to split the original set? It depends on what is the optimum number of samples to be included in the calibration set and how such sample should be chosen.

The use of a small number of samples in the calibration set may result in some sources of variability in the CHMs being excluded and giving spurious results in analysing new samples. In fact, the optimal number of samples depends on the complexity of the system, on the concentration range to be spanned and on the calibration method used [156]. Thus, when the objective is to quantify 1 to 4 components and the samples exhibit no large differences in their physical and chemical properties, a calibration set consisting of a minimum of 15 to 20 samples will be more than enough. However,

owing to their complexity, CHM sample set needs to be as large as possible to reduce the variability.

So as to relate the property to be measured to the analytical signals acquired by NIRRS, the calibration methods most frequently used are multiple linear regression (MLR) [157], principal component regression (PCR) and partial least-squares regression (PLSR) [155]. Most of the earliest quantitative applications of NIRRS rely on MLR because spectra were then acquired on filter instruments, in which NIR measurement recorded with a relatively small number of wavelengths only. Applications of PCR and PLSR (full-spectrum regression techniques) have proliferated after introduction of commercially available instrument that the whole NIR region is allowed to be scanned. The choice of the calibration depends on the nature of the sample, the number of components to be determined simultaneously, the a priori knowledge of the system studied, and data available.

The advantage of estimating a model with stepwise MLR rather than with the full-spectrum techniques such as PLS and PCR is that the MLR model is simple. It does not add variables whose variability is described by previously input variables or that are not linearly related to the analyte of interest. Because, with the full spectrum regression techniques, all sources of variation are implicitly accounted for in the model, this is a more complicated way of dealing with the variation not related to the analyte concerned.

However, because of the elimination of variables during regression, one disadvantage of MLR over PLS and PCR is that the ability to detect unusual samples is limited.

Also, the benefits of multivariate signal averaging are largely eliminated. Lastly, it can only be applied to limited number of wavelengths, which, if incorrectly selected, may result in over-fitting while the precision of the results deteriorate appreciably when the spectral data are highly collinear.

For PCR and PLSR, their comparative advantage is that each single wavelength in a recorded spectrum is considered without missing any one. They permit the simultaneously determination of several components in the same sample and avoid the problems associated with collinearity among spectral data and with noise-related variability. As mentioned earlier, non-linearity in NIR signals is recognized to non-linear detector responses that result in curved signal-concentration plots, as well as to physical and / or chemical factors giving rise to shifts and width changes in spectral bands. Non linearity properties in NIR spectra are seems to be marked in a non-linear calibration methodology [158]. In general, multivariate calibration is more preferable in NIRRS techniques.

For complex composition such as CHM, PCR and PLSR are usually more suitable for quantitative analysis. It offers comparative advantages for complex sample matrices through making use of all the spectral data rather than just discrete wavelengths so that chemical interaction effects can in this way be modelled into the calibration.

Several indicators of statistics are used to describe the quality of calibration and prediction equations in NIR calibration model. They are listed in Table 7.1.

Statistic	Definition
Root mean standard error of cross validation (RMSECV)	Variability in the difference between predicted values and reference values when the equation is applied to a subset of data coming from the calibration data set

Root mean standard error of prediction (RMSEP)	Variability in the difference between predicted and reference values when the equation is applied to the prediction data set
Correlation coefficient (R)	Extent to which reference method values and predicted values by NIRRS at particular calibration / prediction model are correlated
Bias (D)	The mean difference between the predicted and the reference values

Table 7.1: Lists of the definitions of the NIR criteria in quantitative analysis

In the past few years, there are several cases which are successful on establishing the calibration for active ingredient by using NIRRS in CHMs. For example, Chen & Sorensen developed that the calibration curves of ,Glycyrrhizin and Ginsenosides, the marker constituents of radix Notoginseng, with correlation coefficients having value of 0.94 and 0.98, respectively, through a modified PLS [159]. Ren & Chen established the calibration curve of the total ginsenoside in American ginseng with correlation coefficient of 0.99 by PLS [160]; The calibration curve of the markers, Hyperforin and Biapigenin of St John's wort has been set up by Rager et al. with correlation coefficient 0.97 and 0.91, respectively, via PLS [161].

7.3 Conclusion

Near-Infrared technique is gaining acceptance in the Traditional Chinese Medicine Industry and there are many other applications beyond those described in this thesis. But the work reported so far has demonstrated the success of the technique to the CHMs identification and determination of active ingredients assay only. Furthermore, NIRRS method is very fast in routine application, though it is necessary in careful setting up and validating spectral libraries or quantitative calibration during method development. Routine analyses can just take less than 5 min per sample and can simultaneously provide identification without any sample preparation. Also, the technique is non-destructive and non-invasive and allows the same CHMs to be used further for other tests. In addition, it can also be applied directly to CHMs without the need to disrupt the CHMs by extraction or dissolution as required by conventional analysis. This allows NIRRS analysis to take full account of matrix effects, particularly in CHMs which is almost in solid form, and so to provide information relating either to a particular active ingredient which exists in CHM.

Finally, it is worthwhile to reiterate that it is a pre-requisite to establish accurate, robust and validated NIR spectral libraries or calibrations which incorporate all variations to be encountered in real CHMs in practice so as to make good use of NIRRS in rapid, convenient and efficient routine or quality control testing.

Chapter 8: Application of near infrared reflectance spectroscopy to chemical studies on Chinese Herbal Medicine

8.1 Introduction

Near Infrared reflectance spectroscopy (NIRRS) has been becoming a potential alternative method for simultaneous chemical analysis and is being studied extensively in a number of different fields such as process monitoring, biotechnology and the pharmaceutical and food industries in the last decade. However, its application on chemical studies of CHMs has been just started at the initial stage. A larger population of samples with different geographical sources and complete chemical characterization of CHMs are the main difficulties to be overcome.

In this work, the aim is to investigate the feasibility of NIRRS to quality control of CHMs. With the support of chemical information of previous chapters, five different objectives including differentiation of different parts in the same CHM (*Herba menthae* (Bohe, BH), stem and leaf), different species in the same CHM (*Phellodendron chinense Schneid* and *Phellodendron amurense Rupr* (Huangbai, HB)), classification of different CHMs (*Rhizoma Chuanxiong* (Chuanxiong, CX), *Radix Angelicae sinensis* (Danggui, DG), *Cortex Cinnamomi* (Rougui, RG), *Cortex Phellodendri* (Huangbai, HB) and *Herba Pogostemonis*, (GuangHuoXiang, GHX)), quantitative analysis of berberine and total alkaloids contents, in Huangbai, respectively, were explored in the following sections.

8.2 Experimental Procedure

8.2.1 Herbal samples

All herbal samples (36 Bohe (stem and leaf), 31 Chuanxiong, 28 Danggui, 29 Rougui, 24 Guanghuoxiang and 30 Huangbai) were obtained from China Academy of Traditional Chinese Medicine, Beijing, P.R. China with authenticated certificates provided. The geographical information of the samples was lists in the Table 8.1.

8.2.2 Sample treatment

All the samples were grinded by blender and the powder was sieved with a width of 100 mesh. Then this sieved powder was examined through the freeze dryer for 2 days in order to remove the water content in the sample before NIRRS analysis.

8.2.3 Near Infrared Reflectance Spectroscopy

NIR reflectance spectra of all CHM samples were collected over the spectral region 1100 to 2500 nm with an NIRSystems Model XDS spectrometer (Foss NIRSystems, Silver Spring, MD, USA) equipped with a quartz halogen lamp and PbS detector. The spectra were collected at 0.5-nm data intervals. The spectra were acquired with a circular sample cup with a quartz window (38 mm in diameter and 10 mm in thickness). Approximately 2 gram of sample was utilized in the reflectance mode. Each sample was measured in three times after rotation of the cup for 120^{o}, and the corresponding spectra were averaged. Each sample spectrum was obtained by averaging 32 scans. All spectra were recorded as the logarithm of the reciprocal, log (1/R), with respect to ceramic reference standard.

Table 8.1: A list of CHMs samples for NIRRS analysis

BH		CX		DG		RG		GHX		HB	
Label	Location	Label	Location	Label	Location	Label	Location	Label	Location	Label	Location
BH01	Heibi, China	CX01	Sichuan, China	DG01	Gansu, China	RG01	Guangxi, China	GHX01	Guangdong, China	HB01	Sichuan, China
BH02	Heibi, China	CX02	Sichuan, China	DG02	Gansu, China	RG02	Guangxi, China	GHX02	Guangdong, China	HB02	Sichuan, China
BH03	Heibi, China	CX03	Sichuan, China	DG03	Gansu, China	RG03	Yunnan, China	GHX03	Guangdong, China	HB03	Sichuan, China
BH04	Heibi, China	CX04	Sichuan, China	DG04	Gansu, China	RG04	Yunnan, China	GHX04	Guangdong, China	HB04	Sichuan, China
BH05	Henan, China	CX05	Sichuan, China	DG05	Gansu, China	RG05	Sichuan, China	GHX05	Guangdong, China	HB05	Sichuan, China
BH06	Henan, China	CX06	Sichuan, China	DG06	Gansu, China	RG06	Yuenan	GHX06	Guangdong, China	HB06	Sichuan, China
BH07	Henan, China	CX07	Sichuan, China	DG07	Gansu, China	RG07	Guangxi, China	GHX07	Guangdong, China	HB07	Sichuan, China
BH08	Henan, China	CX08	Sichuan, China	DG08	Gansu, China	RG08	Yunnan, China	GHX08	Guangdong, China	HB08	Sichuan, China
BH09	Henan, China	CX09	Sichuan, China	DG09	Gansu, China	RG09	Yuenan	GHX09	Guangdong, China	HB09	Sichuan, China
BH10	Sichuan, China	CX10	Sichuan, China	DG10	Gansu, China	RG10	Miandian	GHX10	Guangdong, China	HB10	Hunan, China
BH11	Sichuan, China	CX11	Sichuan, China	DG11	Gansu, China	RG11	Fujian, China	GHX11	Guangdong, China	HB11	Hunan, China
BH12	Sichuan, China	CX12	Sichuan, China	DG12	Gansu, China	RG12	Guizhou, China	GHX12	Hainan, China	HB12	Hunan, China
BH13	Sichuan, China	CX13	Sichuan, China	DG13	Gansu, China	RG13	Yunnan, China	GHX13	Guangdong, China	HB13	Guizhou, China
BH14	Sichuan, China	CX14	Sichuan, China	DG14	Gansu, China	RG14	Guangdong, China	GHX14	Guangdong, China	HB14	Guizhou, China
BH15	Sichuan, China	CX15	Sichuan, China	DG15	Gansu, China	RG15	Miandian	GHX15	Hainan, China	HB15	Guizhou, China
BH16	Sichuan, China	CX16	Sichuan, China	DG16	Gansu, China	RG16	Guangxi, China	GHX16	Guangdong, China	HB16	Hubei, China
BH17	Sichuan, China	CX17	Yunnan, China	DG17	Gansu, China	RG17	Guangxi, China	GHX17	Hainan, China	HB17	Hubei, China

Label	Location	Label	Location	Label	Location	Label	Location	Label	Location	Label	Location
BH18	Sichuan, China	CX18	Guizhou, China	DG18	Gansu, China	RG18	Guangxi, China	GHX18	Guangdong, China	HB18	Sichuan, China
BH19	Sichuan, China	CX19	Sichuan, China	DG19	Gansu, China	RG19	Guangxi, China	GHX19	Guangdong, China	HB19	Sichuan, China
BH20	Sichuan, China	CX20	Sichuan, China	DG20	Gansu, China	RG20	Guangxi, China	GHX20	Guangdong, China	HB20	Sichuan, China
BH21	Sichuan, China	CX21	Sichuan, China	DG21	Gansu, China	RG21	Guangxi, China	GHX21	Guangdong, China	HB21	Sichuan, China
BH22	Guizhou, China	CX22	Sichuan, China	DG22	Gansu, China	RG22	Guangxi, China	GHX22	Guangdong, China	HB22	Sichuan, China
BH23	Shanghai, China	CX23	Sichuan, China	DG23	Gansu, China	RG23	Guangxi, China	GHX23	Guangdong, China	HB23	Sichuan, China
BH24	Shanghai China	CX24	Sichuan, China	DG24	Gansu, China	RG24	Guangxi, China	GHX24	Guangdong, China	HB24	Sichuan, China
BH25	Fujian, China	CX25	Sichuan, China	DG25	Gansu, China	RG25	Guangxi, China			HB25	Sichuan, China
BH26	Anhui, China	CX26	Sichuan, China	DG26	Gansu, China	RG26	Guangxi, China			HB26	Heilongjiang, China
BH27	Anhui, China	CX27	Sichuan, China	DG27	Gansu, China	RG27	Guangdong, China			HB27	Heilongjiang, China
BH28	Anhui, China	CX28	Sichuan, China	DG28	Gansu, China	RG28	Guangdong, China			HB28	Jilin, China
BH29	Jiangsu, China	CX29	Sichuan, China			RG29	Guangdong, China			HB29	Liaoning, China
BH30	Jiangsu, China	CX30	Sichuan, China							HB30	Dongbei, China
BH31	Jiangsu, China	CX31	Guizhou, China								
BH32	Jiangxi, China										
BH33	Jiangxi, China										
BH34	Jiangxi, China										
BH35	Jiangxi, China										
BH36	Jiangxi, China										

8.2.4 Data analysis

All computations, including selection of spectra and wavelengths, mathematical pretreatment and statistical analysis of cluster analysis and partial least squares regression (PLS), were performed using a VISION (3,0,1,0), FOSS Nirsystems and MATLAB (version 6.5).

8.3 Results and Discussion

8.3.1 Differentiation of different parts of *Herba menthae*

To explore the possibility of NIRRS on quality control of CHM, different parts of Herba menthae were illustrated through the use of principal component analysis (PCA) on their NIR spectra. In this work, in order to eliminate the existing water interference, from the first overtone of the OH stretch at 1450nm and the combination at 1940nm [162], freeze drying process of all samples must be examined prior to NIRRS analysis. The reason for selecting freeze drying is that loss of volatile components due to vaporization can be minimized. The influence of freeze drying process is shown in Fig. 8.1.

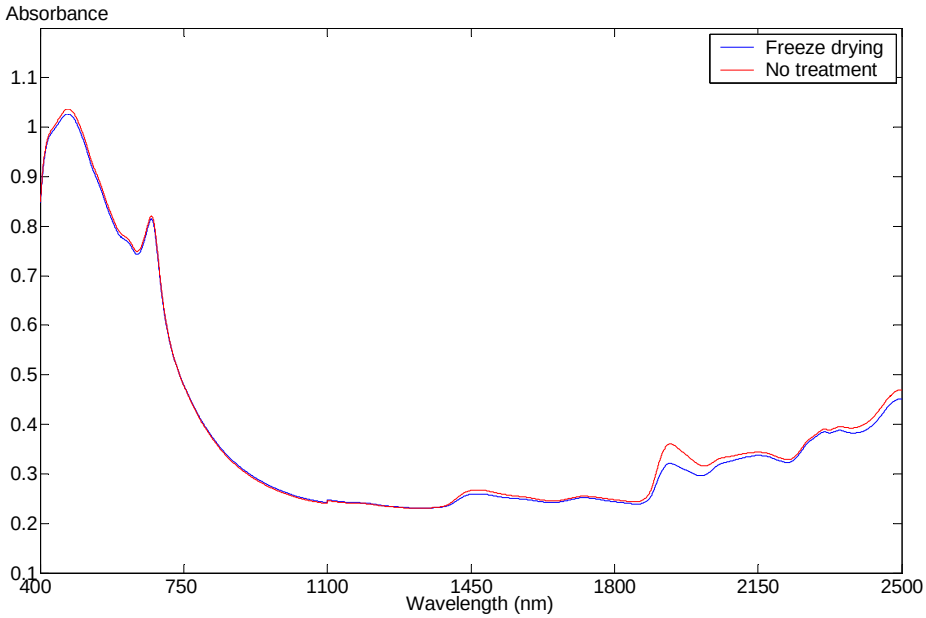


Fig. 8.1 NIR spectrum of *Herba menthae* leaf with (—) no treatment, (—) freeze drying

All samples were expected to exhibit spectral variations throughout the whole NIR range because of compositional variations among different species. However, the NIR spectra ranged between 1100 – 1500nm relating to the stem and leave of Herba menthae were similar to each others. It implies that the chemical information under this range may not be useful for characterizing the chemical differences (see Fig. 8.2).

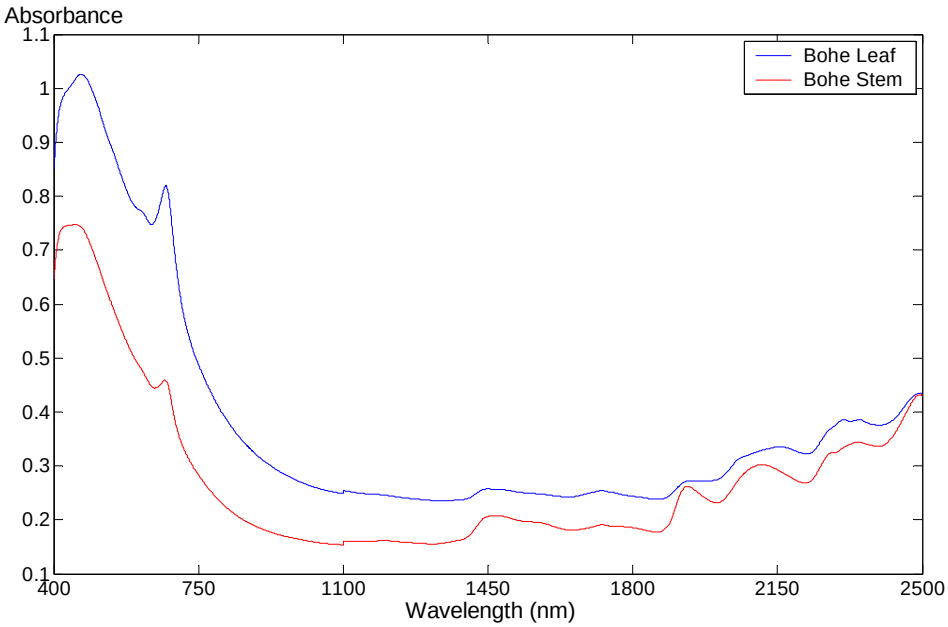


Fig. 8.2 Mean NIR spectra of *Herba menthae* leaf and stem

Principal component analysis (PCA), one of the most powerful and commonly used tools in explorative data analysis, has been commonly utilized in this NIRRS study. In this work, the spectral range of NIR region used for PCA was 1500nm – 2500nm. Before performing PCA, second derivative pre-treatment was done in all spectra in order to eliminate the baseline shift and enhance spectral differences [128]. With the first three principle components, it was possible to extract about 93.8% of the desired variance. The three-dimensional score plot depicted in Fig. 8.3 shows that two different parts, stem of *Herba menthae* (brown square, ■) and leave of *Herba menthae* (green circle, ●), were well separated in the principal components spaces. Hence, NIRRS can be utilized to study composition differences in different parts of *Herba menthae* in this case.

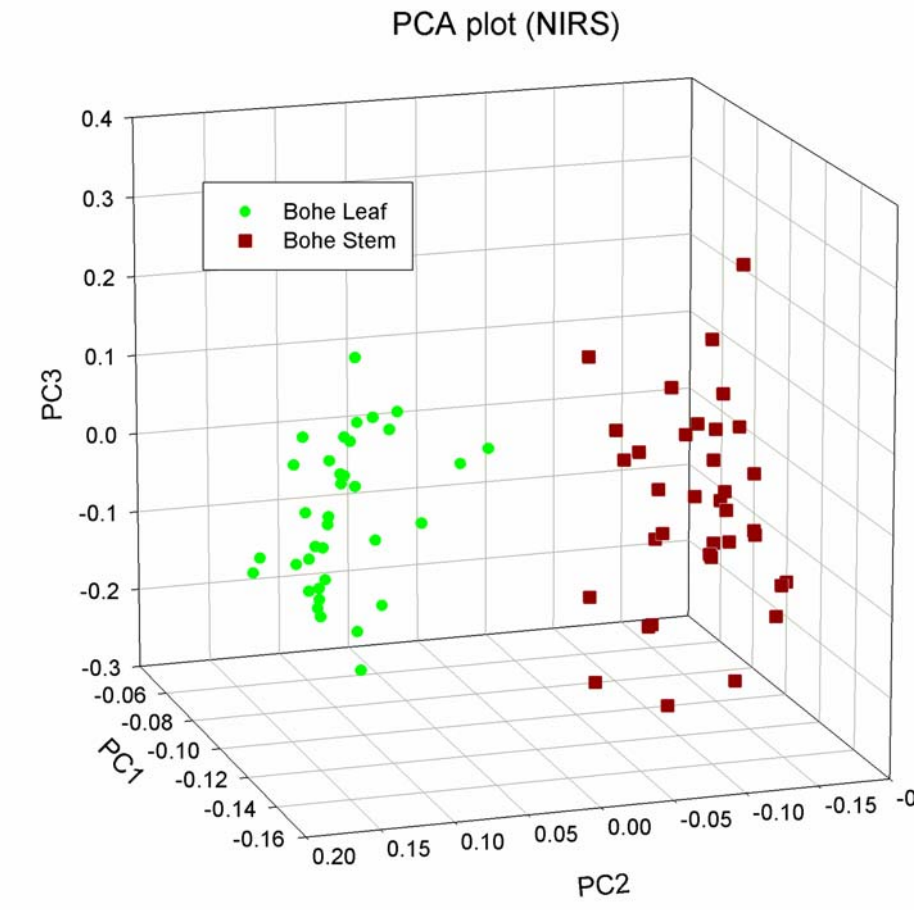


Fig. 8.3 Three-dimensional PCA score plot showing the two parts, *Herba menthae* stem (brown square, ■) and *Herba menthae* leave (green circle, ●).

8.3.2 Differentiation of different species of *Cortex Phellodendri*

Cortex Phellodendri, known as 'Huangbai', has commonly used as a medicinal herb in China for the treatment of diseases like bacillary dysentery, liver cirrhosis, meningitis, pneumonia and tuberculosis. From Pharmacopoeia of the People's Republic of China (2000 edition) [8], *Cortex Phellodendri* is from either *Phellodendron chinense Schneid* or *Phellodendron amurense Rupr*.. Both species consist of similar active ingredient e.g. alkaloids like berberine, jatrorrhizine, palmatine etc, but in different amount. They are, therefore different in pharmaceutical value. As a result, *cortex Phellodendri* has been separated into two entries in the latest 2005 edition of Chinese Pharmacopoeia [25].

In order to differentiate the two species of *Cortex Phellodendri*, PCA was applied for their NIR spectra in this study. All samples were expected to explore spectral variations in the whole NIR range because of compositional variations among these species. Therefore, the NIR spectra of *Phellodendron chinense Schneid* or *Phellodendron amurense Rupr* were shown in Fig. 8.4 and their entire spectral range of NIR region was used for PCA. Before performing PCA, second derivative pre-treatment was carried out in all spectra for eliminating the baseline shift and enhancing of spectral differences. With the first three principle components, it is possible to extract about 95.6% of the desired variance.

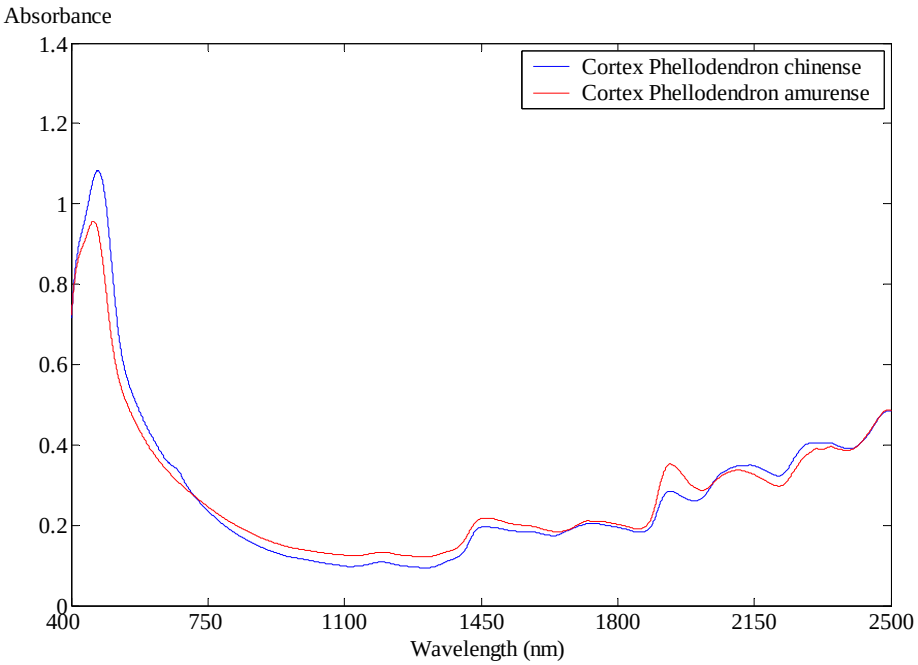


Fig. 8.4 Mean NIR spectra of *Cortex Phellodendron chinense* (blue line, —) and *Cortex Phellodendron amurense* (red line, —)

The three-dimensional score plot depicts in Fig. 8.5 and shows that two species, *Phellodendron chinense* (blue circule, ●) and *Phellodendron amurense* (red triangle, ▲), were well distributed in the principal components spaces. It is possible to explore the existence of composition differences among them from their NIR spectra.

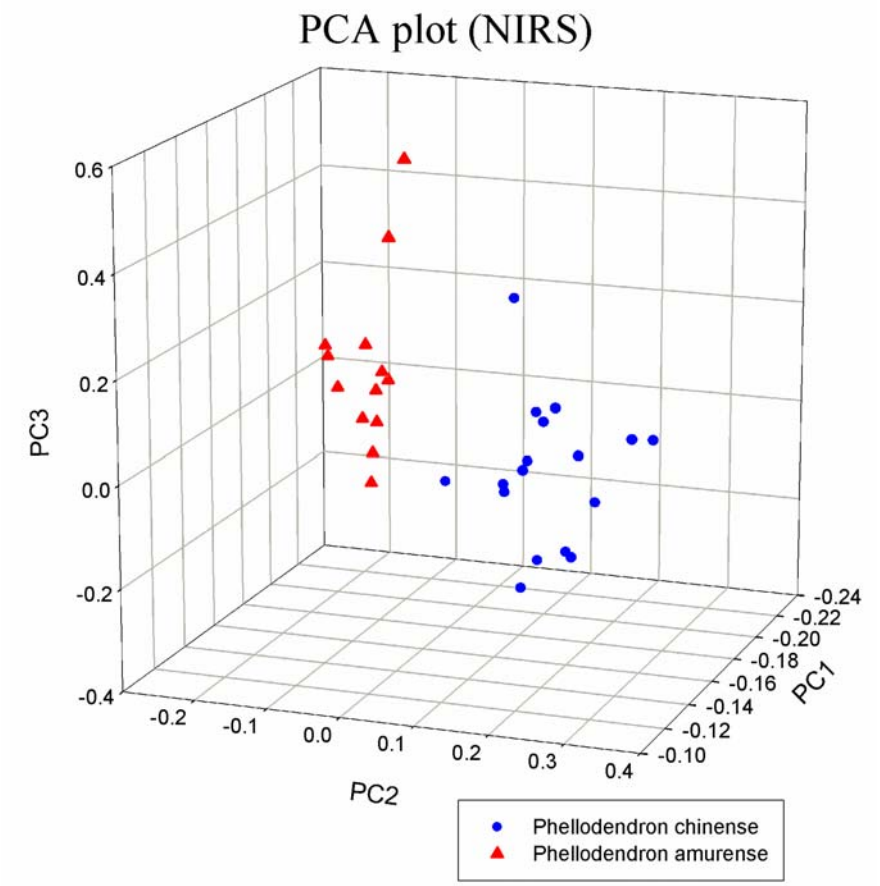


Fig. 8.5 Three-dimensional PCA score plot showing the two species, *Cortex Phellodendron chinense* (blue circule, ●) and *Cortex Phellodendron amurense* (red triangle, ▲).

8.3.3 Classification of different CHMs

Ideally, each CHM has its own characteristic NIR spectrum. To accomplish a fast data acquisition and carry out the data treatment accurately, NIRRS combined with chemometrics methods is a good way to achieve these premises. In this work, all five commonly used CHMs including CX, DG, RG, GHX and HB were analyzed in order to create a database with the characteristic NIR signals of each CHM. The strategy proposed is to investigate how good NIRRS can be used as a screening process to be carried out each CHM, i.e. a prior step used to check the sample first before a deeper

analysis when a positive result was obtained by this proposed methodology.

From the results, it was observed that these CHMs have similar spectral profiles (see Fig 8.6). Hence, it was necessary to utilize patterns recognition method to classify and characterize each group. Before performing PCA, application of the second derivative calculation on all NIR spectra must be carried out so as to enhance their spectral differences.

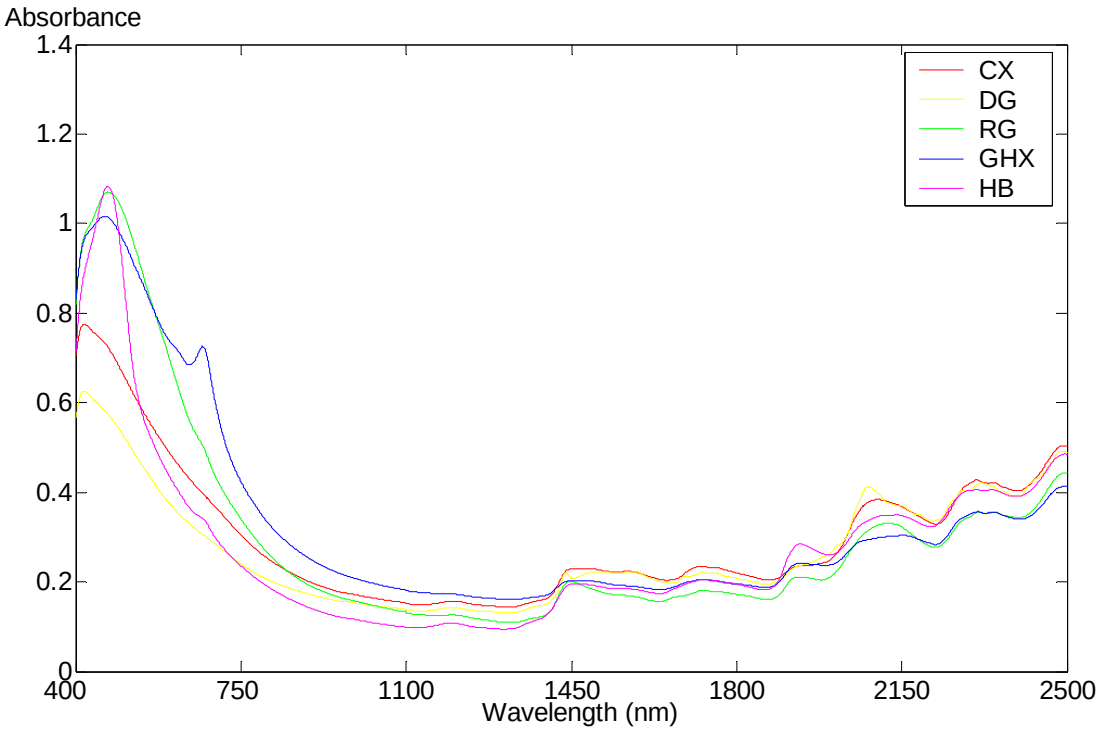


Fig 8.6 Mean NIR spectra of chuanxiong (—), danggui (—), rougui (—), guanghuoxiang (—) and huangbai (—)

Fig. 8.7 depicts the results obtained from PCA of five different CHMs. It is easily seen that there is no overlapping classes when 3 PC's are used. From Fig. 8.7, it supports the PC1 promoted the separation two large groups, danggui and guanghuoxiang; rougui and chuanxiong. In addition, the PC2 indicated the variance differences between chuanxiong and danggui where the essential part of these two CHMs have

similar chemical compositions. Also, the PC3 differentiated between rougui and huangbai in which all PC3 scores of rougui are smaller than -0.05, vice versa. The overall results for classification of the five CHMs were achieved by using NIRRS coupled with pattern recognition tool are appreciated.

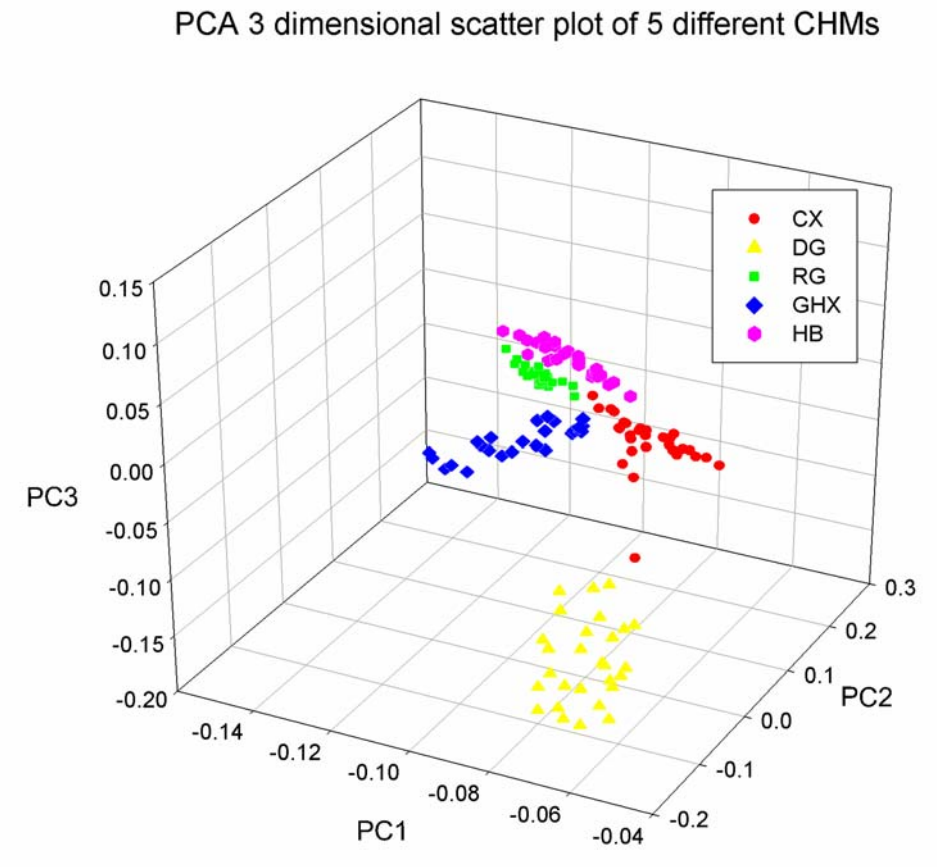


Fig. 8.7 Three-dimensional PCA score plot showing the distribution of the five different CHMs, chuanxiong (●, red circule), danggui (▲ ,yellow triangle), rougui (■, green square), guanghuoxiang (◆, blue diamond) and huangbai (⬢,pink hexagon)

8.3.4 Quantitative analysis of alkaloid content in *Cortex Phellodendri*

The NIRRS technique was explored to seek for a rapid quantitative determination of berberine and total alkaloids in all *Cortex Phellodendri* samples. Multivariate

calibration by PLS regression was also involved. Content of berberine and total alkaloids contents in all *Cortex Phellodendri* are the critical parameters in quality control of all these samples according to Chinese Pharmacopoeia 2005 edition and were mentioned in Chapter 7.

The robustness of the PLS model chosen was evaluated according to the root mean square error of cross-validation (RMSECV), the root mean square error of prediction (RMSEP) and the correlation coefficient (R). In this work, 60% of samples were used for calibration while 40% of samples for validation. The contents of berberine and total alkaloids were equally distributed over the entire concentration range in both calibration and test sets. The optimum number of factors employed for validation was determined by cross-validation.

In this work, the number of PLS factors included in the model was determined by taking the minimum prediction residual sum of squares (PRESS) value as criterion. This procedure was repeated for each of the preprocessing spectra. Besides, so as to pre-process the NIR spectra, we utilized several methods, including derivation, multiplicative scatter correction (MSC), Savitsky-Golay, standard normal variate transformation (SNV) and detrend were applied to all NIR spectra in order to minimize the interference effect such as correction of scatter effect, elimination of baseline shift, enhancement of spectral differences and smoothing spectrum etc.

The performance of the final partial least squares PLS regression model chosen was evaluated by root mean square error of cross-validation (RMSECV), the root mean square error of prediction (RMSEP) and the correlation coefficient (R). For RMSECV,

a leave-one-sample-out cross-validation was performed: the spectrum of one sample of the training set was removed from this set and a PLS regression model was built with the remaining spectra of the training set. The left-out sample was predicted by this model and this procedure was repeated by leaving out each of the samples in the training set. The RMSECV was calculated as follows,

$$\text{RMSECV}=\sqrt{\frac{\sum_{i=1}^{n}(\hat{c}_i - c_i)^2}{n - A - 1}} \qquad \text{(Eq. 8.1)}$$

where c_i is reference measurement result for sample *i*, $\hat{c}_i$ are predicted values of the sample *i* when all samples are included in the calibration model, n is the number of calibration samples, A is the number of PLS components.

For the test set, the root mean square error of prediction (RMSEP) was computed as follows,

$$\text{RMSEP}=\sqrt{\frac{\sum_{i=1}^{n}(\hat{c}_i - c_i)^2}{n}} \qquad \text{(Eq. 8.2)}$$

where n is the number of samples in the test set, $\hat{c}_i$ is the estimated value of the model for test sample *i*, and c_i is reference measurement result for test sample *i*.

Finally, correlation coefficients between the predicted and the measured value were calculated for both the training and test set, which are obtained via the following equation,

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n}(\hat{c}_i - c_i)^2}{\sum_{i=1}^{n}(c_i - \overline{c}_i)^2}} \qquad \text{(Eq. 8.3)}$$

with $\bar{c}$ being mean of the reference measurement results for all samples in the training and test sets.

A model for the quantitative analysis was established in this way for berberine and total alkaloids content in the NIR spectra. In PLS algorithm, it is generally recognized that the spectral pre-processing treatments and the number of PLS factors are critical parameters. The optimum number of factors is determined by the lowest PRESS value. A good rule for choosing the number of PLS components to be retained is not to include additional components unless they improve the PRESS by at least 2%. Fig. 8.8 shows PRESS plotted as a function of PLS factors for determining the berberine content by different spectral pre-processing methods.

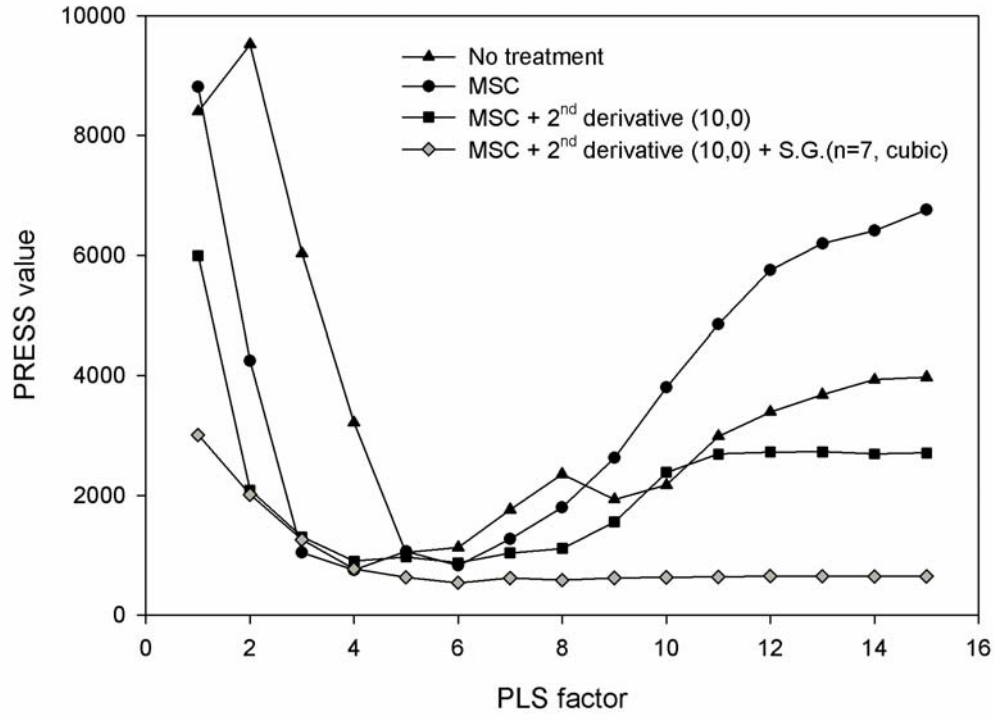


Fig. 8.8 Effect of the number of PLS factors on PRESS values in the berberine calibration model.

Table 8.2 shows the values obtained for RMSECV, RMSEP and the correlation coefficient between the measured and NIRS predicted values of berberine content for the training set and the test set by different preprocessing methods. For each one, only the results for the model with the lowest RMSEP value were given. The pre-treatment including MSC, 2 derivative (10,0) and S.G. (n=7,cubic) in the full range of NIR (1100 – 2500nm) give the lowest RMSECV value among all these methods. Deriving spectra with smoothing treatment usually enhances small differences between spectra and splits overlapping band. This could be a reason for the good correlation results of berberine content.

Fig. 8.9 is a scatter plot showing a correlation between prediction values based on NIR and the reference HPLC measurement for berberine content by the spectral pre-processing method including MSC, second derivative and Savitsky-Golay,. Red circles and blue triangles represent calibration and prediction data, respectively. It can be seen that the calibration and prediction data have very good correlation with reference measurement data and many data points fall on or close to the unity line. Berberine content in the test set was predicted with RMSEP of value 2.6571 (6.907%). The correlation coefficients for this calibration model for the training and test set were 0.9980 and 0.9950, respectively.

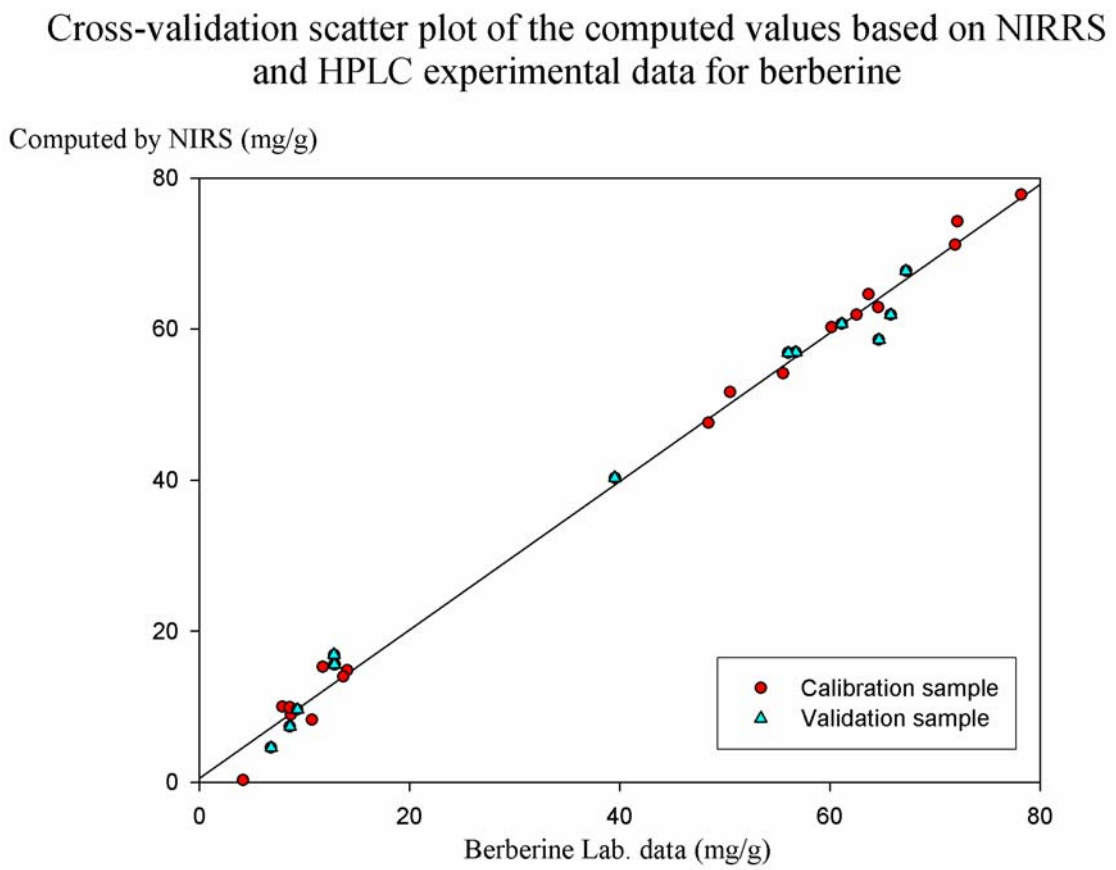


Fig. 8.9 Prediction values vs HPLC measurement for the determination of the actual content of berberine in all *Cortex Phellodendri* samples (n=30). R (training) = 0.9980; R (test) = 0.9950.

Table 8.2: Performance of the PLS regression models for berberine content with different pretreatment methods

Pretreatments Method used[b]	PLS Factor	RMSECV	RMSEP	R (training set)	R (testing set)
No treatment	1	19.4933	14.1461	0.7370	0.8489
MSC	4	5.3053	6.7020	0.9862	0.9673
MSC+ SNV	4	5.1978	6.8301	0.9867	0.9665
MSC+ detrend	4	5.3272	6.5831	0.9861	0.9685
MSC + 2nd derivative (10,0)	4	4.1368	4.5753	0.9916	0.9852
MSC + 2nd derivative (10,0) + S.G. (n=5, quadratic)	6	2.1553	2.6793	0.9981	0.9980
MSC + 2nd derivative (10,0) + S.G. (n=7, cubic)[a]	**6**	**2.2020**	**2.6571**	**0.9980**	**0.9950**
MSC + 2nd derivative (10,0) + S.G. (n=9, quartic)	6	2.1224	2.6996	0.9981	0.9950
MSC + 2nd derivative (10,0) + S.G. (n=11, quintic)	6	2.1358	2.6918	0.9981	0.9950

[a] Results from models with the best performance are marked in bold

[b] MSC, multiplicative scatter correction; S.G., Savitsky-Golay (data point, polynomial order); SNV, standard normal variate transformation; derivative (segment, gap); RMSECV, root mean of square error of cross validation; RMSEP, root mean of square error of prediction set; R, correlation coefficient.

Table 8.4 shows for each of the preprocessing methods the RMSECV, RMSEP and the correlation coefficient between the measured and NIRS predicted values of total alkaloids content for the training set and the test set. For each of the preprocessing methods, only the results for the model with the lowest RMSEP value were shown. The pre-treatment including MSC, 2 derivative (10,0) and S.G. (n=7,cubic) in the range of whole NIR spectra (1100 – 2500nm) gave the lowest RMSECV value of all pre-processing methods. Fig. 8.10 is a scatter plot depicting a correlation between NIR prediction value and reference HPLC measurement for total alkaloids content by the MSC, second derivative and Savitsky-Golay, spectral pre-processing method. Red circles and blue triangles represent calibration and prediction data, respectively. It shows that good correlation with HPLC reference measurement in calibration and prediction sets was developed. Total alkaloids content in the test set was predicted with RMSEP of value 2.5797 (6.627%). The correlation coefficients for this calibration model for the training and test set were 0.9975 and 0.9937, respectively. The overall statistical results of berberine and total alkaloids contents in calibration and validation set are summarized in Table. 8.3

Table 8.3: Statistical results of calibration and validation sets for berberine and total alkaloids contents

	Samples	Berberine content (mg/g)	Mean	Standard deviation
Calibration set	18	4.15 – 78.20	39.30	27.98
Validation set	12	6.83 – 67.24	38.47	26.06
	Samples	Total alkaloids content (mg/g)	Mean	Standard deviation
Calibration set	18	11.26 – 78.20	41.96	25.07
Validation set	12	13.77 – 67.24	41.17	23.18

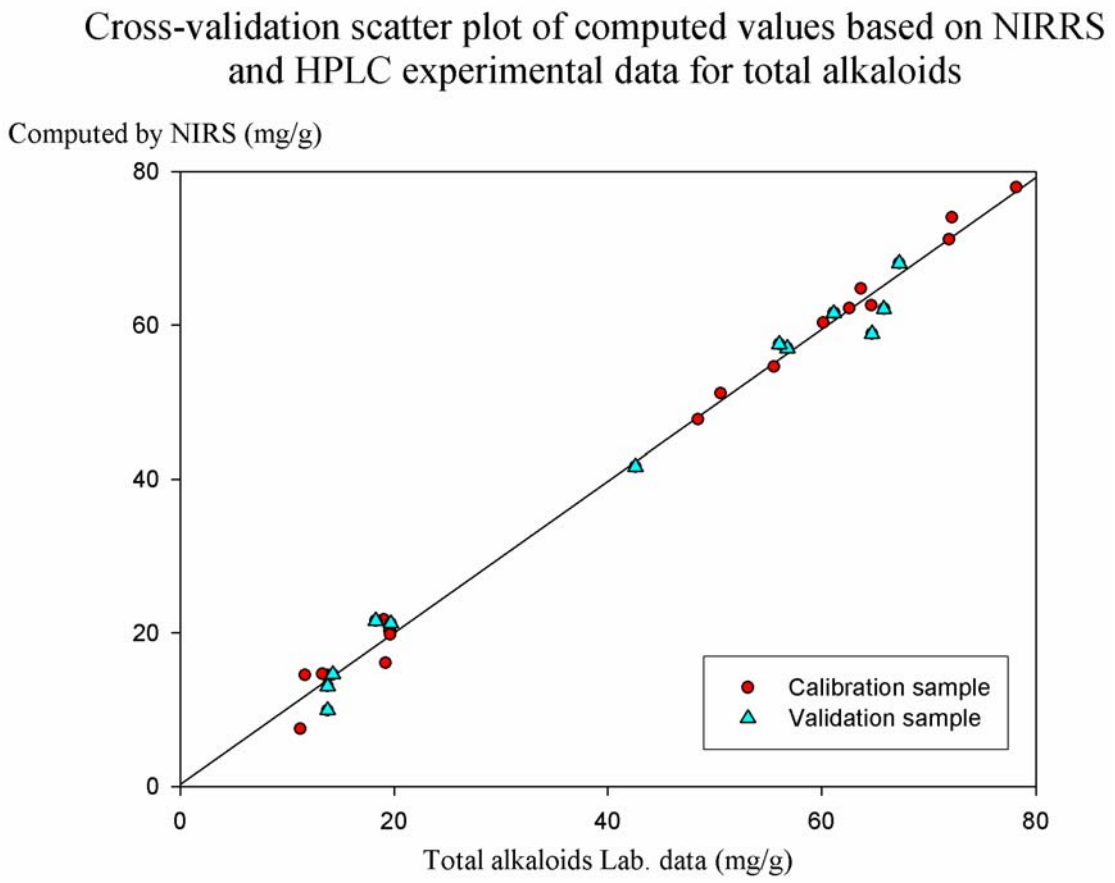


Fig. 8.10 Prediction values vs HPLC results for the determination of total content of alkaloids in all *Cortex Phellodendri* samples (n=30). R (training) = 0.9975; R (test) = 0.9937.

Table 8.4: Performance of the PLS regression model for total alkaloids contents with different pretreatment methods

Pretreatments Method used[b]	Factor	RMSECV	RMSEP	R (training set)	R (testing set)
No treatment	1	17.6032	12.6559	0.7320	0.8471
MSC	4	4.9495	5.9442	0.9850	0.9671
MSC+ SNV	4	4.8539	6.0846	0.9855	0.9660
MSC+ detrend	6	4.3556	4.9632	0.9902	0.9774
MSC + 2nd derivative (10,0)	4	3.4447	4.1027	0.9932	0.9853
MSC + 2nd derivative (10,0) + S.G. (n=5, quadratic)	6	2.1486	2.6128	0.9976	0.9936
MSC + 2nd derivative (10,0) + S.G. (n=7, cubic) [a]	**6**	**2.1958**	**2.5797**	**0.9975**	**0.9937**
MSC + 2nd derivative (10,0) + S.G. (n=9, quartic)	6	2.1155	2.6403	0.9977	0.9935
MSC + 2nd derivative (10,0) + S.G. (n=11, quintic)	5	2.9002	3.1375	0.9953	0.9914

[a] Results from models with the best performance are marked in bold

[b] MSC, multiplicative scatter correction; S.G., Savitsky-Golay (data point, polynomial order); SNV, standard normal variate transformation; derivative (segment, gap); RMSECV, root mean of square error of cross validation; RMSEP, root mean of square error of prediction set; R, correlation coefficient.

8.4 Conclusion

In summary, we demonstrated in this study that the potential of near infrared reflectance spectroscopy (NIRRS) for quality control of CHMs is promising. Rapid differentiation between different part of *Herba menthae*; different species of *Cortex Phellodendri*; different species of CHMs like *Rhizoma Chuanxiong, Radix Angelicae sinensis*, *Cortex Cinnamomi*, *Cortex Phellodendri* and *Herba Pogostemonis* have been carried out while rapid quantitative detection of berberine and total alkaloids content in *Cortex Phellodendri* have been performed. Even though developing the calibration and identification models in NIRRS needs a lot of time, cost and effort, NIRRS provides the great advantage of non-destructive, short time measurement. Also, once the model is built up, routine analysis can be carried out very quickly in contrast to that of HPLC analysis. For development of a more reliable and accurate NIRRS model, a larger sample size from different provinces and countries is certainly needed to come up a robustness result with higher creditability.

Chapter 9: Overall summary

Chemometrics techniques have been developed an useful tool on chromatographic and spectroscopic studies. Huge amount of multi-dimentional data obtained from some of these advanced instruments are richer in chemical information that the two-dimensional ones. Introduction of chemometrics techniques provides alternative way to objectively analyze the data, even in huge amount, so as to obtain more reliable, reproducible results than those from traditional ones. With their benefits being used in chromatographic and spectroscopic methods, the analysis not only is shortened significantly, but also more chemical information from original data can be extracted. The most important one, human error, are minimized by avoiding repetitive manual measurements. In this study, the application of chemometrics techniques on chromatographic and spectroscopic methods to advance chemical analysis of *Radix Ligustici chuanxiong*, *Radix Angelicae sinensis*, *Cortex Phellodendri* and other Chinese herbal medicines, were demonstrated to work successfully in analyzing complex data, extracting useful chemical information effectively as well as minizing instrumental interferences. A brief summary of the finding in this work is given in the following paragraphs.

In Chapter 2, different chromatographic and electrophoretic methods coupled with various extraction techniques used in the instrumental measurement of CHM were briefly introduced. In addition, three different methodologies including ‘marker approach’, ‘multi-component approach’ and ‘pattern approach’ for mining chemical information from original data obtained from chromatographic and electrophoretic techniques were comprehensively discussed with their pros and cons mentioned under the purpose for quality control of CHM. Furthermore, the fundamental concepts of chemometrics tools utilized in this thesis have been briefly given in Chapter 3. They

consist of heuristic evolving latent projection, spectral correlative chromatography, information theory, normalization, local least square method, similarity index, hierarchical cluster analysis, principal component analysis and linear discriminant analysis.

A comparison of the three methodologies used, ‘marker approach’, ‘multi-component approach’ and ‘pattern approach’ for developing the chromatographic fingerprint of *Radix Ligustici chuanxiong* on quality control purposed is presented in Chapter 4. The basis for comparison is that the process of data treatment on GC-MS chromatograms through these approaches before using similarity index (SI) and principal component analysis (PCA) for multivariate analysis. It is the first work with emphasis on the comparison of these three data analytical approaches for chemical characterization of CHM in order to be able to compare the samples from different geographical sources. Two chemometrics tools, heuristic evolving latent projection and spectral correlative chromatography were utilized extensively for mining more chemical information effectively while a local least square method was used in chromatographic alignment before multivariate analysis. From the results obtained, pattern approach was explored on it their feasibility in differentiating different geographical locations and the overall procedures processed in constructing the herbal fingerprint can be used as a basis for further development on authentication process in CHM.

Based on the three different methodologies mentioned above, another comparison approach study on developing a classification model between *Radix Angelicae Sinensis* and *Radix Ligustici chuanxiong* from different geographical sources has been worked out and mentioned in Chapter 5. Different portions of chemical information

detected from GC-MS chromatogram, with the help of chemometrics tools used in developing herbal fingerprint, were extracted for constructing the classification model in which multi-samples collections with different geographical sources were involved in order to generate a representative picture. Then, similarity index evaluations, hierarchical clustering analysis, principal component analysis and linear discriminant analysis have been utilized in developing the classification model which five-folds cross validation were tested in evaluating the overall model performance. The results are that pattern approach is feasible on classifying CX and DG in which 100% correct identifications in the validation sets have been obtained.

Cortex Phellodendri comes from either *Phellodendron chinense Schneid* or *Phellodendron amurense Rupr* and both species consist of alkaloids with different amount. In Chapter 6, a new extraction method, with the use of information theory, and a new chromatographic HPLC condition for alkaloids in *Cortex Phellodendri* were developed. The alkaloids contents of thirty samples from different geographical sources were analyzed and method validations like inter-variability, intra-variability, LOD, LOQ and recovery were carried out. Finally, a fingerprint concept has been introduced to differentiate these two species through similarity index and principal component analysis.

In the past decade, near infrared reflectance spectroscopy (NIRRS) was introduced as another useful analytical tool to investigate CHMs including qualitative and quantitative analysis because of its rapid, accurate and non-destructive properties. The fundamental concept of NIRRS including theory, advantages and disadvantages, limitations and general procedures in qualitative and quantitative analysis were

described in Chapter 7.

Finally, in Chapter 8, several outcomes using NIRRS, combined with chemometrics data processing, were obtained from the spectral feature of NIR. By the use of principal component analysis with second-derivative algorithm for data processing, it is possible to differentiate different parts of the same CHM (*Herba menthae* (Bohe), stem and leaf), different species in the same CHM (*Phellodendron chinense Schneid* and *Phellodendron amurense Rupr* (Huangbai)), different CHMs (*Rhizoma Chuanxiong* (Chuanxiong), *Radix Angelicae sinensis* (Danggui), *Cortex Cinnamomi* (Rougui), *Cortex Phellodendri* (Huangbai) and *Herba Pogostemonis*, (GuangHuoXiang)). Besides, with the use of partial least square regression, quantitative analysis of berberine and the total alkaloids content in *Cortex Phellodendri* via NIRRS were carried out based on the results of reference method obtained from chromatographic analysis. The results showed good correlation between NIRRS and chromatographic method using partial least-squares regression.

Prospects and future work

Prospects and future work

Chemometrics techniques are well applied to the chemical analysis of CHM with good results and this study demonstrates the applications of new technical and computational approaches to set up criteria and standards for the quality control of CHM. After this investigation, there are some suggestions for future research to make this project to be more complete. First of all, more works can be done on the analysis of *Cortex Phellodenri*. In the conventional preparation of the herbal material, water or ethanol is used to extract the constituents of the herbs. However, acidified methanol was employed in this project to extract the components. Therefore, some modifications can be done to relate the chemical profile of the acidified methanol extracts and the water / ethanol extracts. Moreover, further work can be performed to implement the combination use between multi-component and pattern approach on herbal authentication of other CHM via chromatography. Later on, this novel approaches can be implemented on the herbal formulation and commercial products to set up objective criteria for advancing the standard of the quality control.

Finally, more works on the NIRS techniques can be done on the industrial production process of CHM commercial product. We believe that implementation the NIRS techniques provides on-line monitoring on the production process and NIRS can be possibly examined the change of chemical compositions of commercial product during production.

References

References

1. Mok, D. K. W. and Chau, F. T., *Chemical information of Chinese medicines: A challenge to chemist.* Chemometrics and Intelligent Laboratory Systems, 2006. 82(1-2): p. 210-217.

2. Liang, Y. Z., Xie, P. S., and Chan, K., *Quality control of herbal medicines.* Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 2004. 812(1-2): p. 53-70.

3. WHO, *General Guidelines for Methodologies on Research and Evaluation of Traditional Medicines*. 2001.

4. Booksh, K. S. and Kowalski, B. R., *Theory of Analytical-Chemistry.* Analytical Chemistry, 1994. 66(15): p. A782-A791.

5. Smith, R. M., *Before the injection - modern methods of sample preparation for separation techniques.* Journal of Chromatography A, 2003. 1000(1-2): p. 3-27.

6. United States Pharmacopeial Convention, *United States Pharmacopeia and National Formulary*. Vol. USP (25), NF(19). 2002.

7. The Society of Japanese Pharmacopeia, *The Japanese Pharmacopeia*. Vol. XIII. 2001.

8. The Pharmacopeia Commission of PRC, *Pharmacopoeia of the People's Republic of China (Chin.)*. 2000, Beijing.

9. Camel, V., *Recent extraction techniques for solid matrices-supercritical fluid extraction, pressurized fluid extraction and microwave-assisted extraction: their potential and pitfalls.* Analyst, 2001. 126(7): p. 1182-1193.

10. Zygmunt, B. and Namiesnik, J., *Preparation of samples of plant material for chromatographic analysis.* Journal of Chromatographic Science, 2003. 41(3): p. 109-116.

11. Pan, X. J., Niu, G. G., and Liu, H. Z., *Comparison of microwave-assisted extraction and conventional extraction techniques for the extraction of tanshinones from Salvia miltiorrhiza bunge.* Biochemical Engineering Journal, 2002. 12(1): p. 71-77.

12. Guo, Z. K., Jin, Q. H., Fan, G. Q., Duan, Y. P., Qin, C., and Wen, M. J., *Microwave-assisted extraction of effective constituents from a Chinese herbal medicine Radix puerariae.* Analytica Chimica Acta, 2001. 436(1): p. 41-47.

13. Ma, X. Z., Yu, X. B., and Han, J., *Application of Off-Line Supercritical-Fluid Extraction-Gas Chromatography for the Investigation of Chemical-Constituents in Curcuma-Zedoaria.* Phytochemical Analysis, 1995. 6(6): p. 292-296.

14. Li, L., Chen, Z. Q., and Li, X. L., *Application of supercritical fluid extraction in quality control of Chinese medicinal materials.* Yaoxue Xuebao, 1995. 30(2): p. 133-137.

15. Ong, E. S., Woo, S. O., and Yong, Y. L., *Pressurized liquid extraction of berberine and aristolochic acids in medicinal plants.* Journal of Chromatography A, 2000. 904(1): p. 57-64.

16. Lee, H. K., Koh, H. L., Ong, E. S., and Woo, S. O., *Determination of ginsenosides in medicinal plants and health supplements by pressurized liquid extraction (PLE) with reversed phase high performance liquid chromatography.* Journal of Separation Science, 2002. 25(3): p. 160-166.

17. British Herbal Medicine Association, *British Herbal Pharmacopoeia*. 1996.

18. Xie, P. S., Chen, S. B., Liang, Y. Z., Wang, X. H., Tian, R. T., and Upton, R., *Chromatographic fingerprint analysis - a rational approach for quality assessment of traditional Chinese herbal medicine.* Journal of Chromatography A, 2006. 1112(1-2): p. 171-180.

19. Di, X., Chan, K. V. K. C., Leung, H. W., and Huie, C. W., *Fingerprint profiling of acid hydrolyzates of polysaccharides extracted from the fruiting bodies and spores of Lingzhi by high-performance thin-layer chromatography.* Journal of Chromatography A, 2003. 1018(1): p. 85-95.

20. Gong, F., Liang, Y. Z., and Chau, F. T., *Combination of GC-MS with local resolution for determining volatile components in si-wu decoction.* Journal of Separation Science, 2003. 26(1-2): p. 112-122.

21. Guo, F. Q., Liang, Y. Z., Xu, C. J., Huang, L. F., and Li, X. N., *Comparison of the volatile constituents of Artemisia capillaris from different locations by gas chromatography-mass spectrometry and projection method.* Journal of Chromatography A, 2004. 1054(1-2): p. 73-79.

22. Gong, F., Liang, Y. Z., Xu, Q. S., and Chau, F. T., *Gas chromatography-mass spectrometry and chemometric resolution applied to the determination of essential oils in Cortex Cinnamomi.* Journal of Chromatography A, 2001. 905(1-2): p. 193-205.

23. Li, X. N., Cui, H., Song, Y. Q., Liang, Y. Z., and Chau, F. T., *Analysis of volatile fractions of Schisandra chinensis (Turcz.) baill. using GC-MS and chemometric resolution.* Phytochemical Analysis, 2003. 14(1): p. 23-33.

24. Shen, S., Sha, Y. F., Deng, C. H., Zhang, X. M., Fu, D. X., and Chen, J. K., *Quality assessment of Flos Chrysanthemi Indici from different growing areas in China by solid-phase - microextraction-gas chromatography-mass spectrometry.* Journal of Chromatography A, 2004. 1047(2): p. 281-287.

25. The Pharmacopeia Commission of PRC, *Pharmacopoeia of the People's Republic of China (Chin.)*. 2005: Beijing.

26. Yang, L. W., Wu, D. H., Tang, X., Peng, W., Wang, X. R., Ma, Y., and Su, W. W., *Fingerprint quality control of Tianjihuang by high-performance liquid chromatography-photodiode array detection.* Journal of Chromatography A, 2005. 1070(1-2): p. 35-42.

27. Ji, Y. B., Xu, Q. S., Hu, Y. Z., and Heyden, Y. V., *Development, optimization and validation of a fingerprint of Ginkgo biloba extracts by high-performance liquid chromatography.* Journal of Chromatography A, 2005. 1066(1-2): p. 97-104.

28. Luo, X. B., Chen, B., Yao, S. Z., and Zeng, J. G., *Simultaneous analysis of caffeic acid derivatives and alkamides in roots and extracts of Echinacea purpurea by high-performance liquid chromatography-photodiode array detection-electrospray mass spectrometry.* Journal of Chromatography A, 2003. 986(1): p. 73-81.

29. Fan, X. H., Cheng, Y. Y., Ye, Z. L., Lin, R. C., and Qian, Z. Z., *Multiple chromatographic fingerprinting and its application to the quality control of herbal medicines.* Analytica Chimica Acta, 2006. 555(2): p. 217-224.

30. Lu, G. H., Chan, K., Liang, Y. Z., Leung, K., Chan, C. L., Jiang, Z. H., and Zhao, Z. Z., *Development of high-performance liquid chromatographic fingerprints for distinguishing Chinese Angelica from related umbelliferae herbs.* Journal of Chromatography A, 2005. 1073(1-2): p. 383-392.

31. Zhang, J. L., Cui, M., He, Y., Yu, H. L., and Guo, D. A., *Chemical fingerprint and metabolic fingerprint analysis of Danshen injection by HPLC-UV and HPLC-MS methods.* Journal of Pharmaceutical and Biomedical Analysis, 2005. 36(5): p. 1029-1035.

32. Jorgenson, J. W. and Lukacs, K. D., *Zone Electrophoresis in Open-Tubular Glass-Capillaries.* Analytical Chemistry, 1981. 53(8): p. 1298-1302.

33. Yu, K., Wang, Y. W., and Cheng, Y. Y., *Determination of protocatechuic aldehyde, danshensu, salvianolic acid B and gallic acid in Chinese medicine 'SHUANGDAN' granule by MEKC.* Chromatographia, 2006. 63(7-8): p. 389-393.

34. Liu, J. J., Li, S. P., and Wang, Y. T., *Optimization for quantitative determination of four flavonoids in Epimedium by capillary zone electrophoresis coupled with diode array detection using central composite design.* Journal of Chromatography A, 2006. 1103(2): p. 344-349.

35. Sun, Y., Guo, T., Sui, Y., and Li, F., *Fingerprint analysis of Flos Carthami by capillary electrophoresis.* Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 2003. 792(2): p. 147-152.

36. Shihabi, Z. K. and Hinsdale, M. E., *Sample Matrix Effects in Micellar Electrokinetic Capillary Electrophoresis.* Journal of Chromatography B-Biomedical Applications, 1995. 669(1): p. 75-83.

37. Zschocke, S., Classen-Houben, D., and Bauer, R., *Radix Angelicae sinensis: Danggui*. 2001: Verlag fur Ganzheitliche Medizin, Wald (Germany).

38. Zschocke, S., Classen-Houben, D., and Bauer, R., *Radix Ligustici chuanxiong: Chuanxiong*. 2001: Verlag fur Ganzheitliche Medizin, Wald (Germany).

39. Foster, S. and Tyler, V. E., *Tyler's Honest Herbal*. 1999: Haworth Herbal Press.

40. Murray, M. T., *Common Questions about the St John's Wort Extract.* American Journal of Natural Product, 1997. 4: p. 14-19.

41. Schaneberg, B. T., Crockett, S., Bedir, E., and Khan, I. A., *The role of chemical fingerprinting: application to Ephedra.* Phytochemistry, 2003. 62(6): p. 911-918.

42. Jordan, J. R., *The Big Picture in Traditional Chinese Medicine*, in *AOAC International*. 2005. p. 22-24.

43. Nielsen, N. P. V., Carstensen, J. M., and Smedsgaard, J., *Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping.* Journal of Chromatography A, 1998. 805(1-2): p. 17-35.

44. Xie, P. S., *中醫葯傳統文化與現代質量控制.* 中國中西醫結合雜誌, 1998. 18: p. 645.

45. Xie, P. S. and Yan, Y. Z., *HPTLC fingerprint identification of commercial ginseng drugs: Reinvestigation of HPTLC of ginsenosides.* Journal High Resolution chromatography & chromatography Communication, 1987. 10: p. 607-613.

46. Xie, P. S., *中草藥劑型改進與質量控制.* Traditional Chinese Drug Research and Clinical Pharmacology, 1998. 9: p. 5.

47. Sticher, O., *Quality of Ginkgo Preparations.* Planta Medica, 1993. 59(1): p. 2-11.

48. Zhai, W. M., Yuan, Y. S., Zhou, L. X., and Wei, L. X., *HPLC fingerprints identification of Panax ginseng (C.A. Mey), P. quinquefolin L. and P. notoginseng (Burk. F. H. Chen).* China Journal of Chinese Materia Medica, 2001. 26(7): p. 481-482.

49. Tao, L., Qu, C. H., Huang, J. W., Zhang, Z. L., and Wu, H. Q., *白屈菜注射液 HPLC 指紋圖譜研究.* 中葯材, 2004. 27: p. 376-377.

50. Zhang, K. R. and Tu, P. F., *Study on fingerprints of Radix Paeonice Rubra by HPLC.* Chinese Traditional and Herbal Drugs, 2003. 34: p. 1048-1051.

51. Zhang, T. M., Liang, Y. Z., Li, B. Y., Cui, H., and Gong, F., *Systemic analysis of structures and contents of nitrogen-containing compounds and other non-hydrocarbons in crude oils in conjunction with chemometric resolution technique.* Analytical science, 2004. 20: p. 717-724.

52. Lu, H. M., Liang, Y. Z., and Chen, S., *Identification and quality assessment of Houttuynia cordata injection using GC-MS fingerprint: A standardization approach.* Journal of Ethnopharmacology, 2006. 105(3): p. 436-440.

53. Hu, L. F., Li, S. P., Cao, H., Liu, J. J., and Gao, J. L., *GC-MS fingerprint of Pogostemon cablin in China.* Journal of Pharmaceutical and Biomedical Analysis, 2006. in press.

54. Qian, H. Q., Li, C. J., and Xie, P. S., *On GC Fingerprint of Rhizome of aplinia officinarum.* Traditional Chinese Drug Research and Clinical Pharmacology, 2001. 12: p. 179-184.

55. Cheng, Y. Y., Chen, M. J., and Welsh, W. J., *Fractal fingerprinting of chromatographic profiles based on wavelet analysis and its application to characterize the quality grade of medicinal herbs.* Journal of Chemical Information and Computer Sciences, 2003. 43(6): p. 1959-1965.

56. Cheng, Y. Y., Chen, M. J., and Tong, W. D., *An approach to comparative analysis of chromatographic fingerprints for assuring the quality of botanical drugs.* Journal of Chemical Information and Computer Sciences, 2003. 43(3): p. 1068-1076.

57. Li, B. Y., Hu, Y., Liang, Y. Z., Xie, P. S., and Du, Y. P., *Quality evaluation of fingerprints of herbal medicine with chromatographic data.* Analytica Chimica Acta, 2004. 514(1): p. 69-77.

58. Kvalheim, O. M. and Liang, Y. Z., *Heuristic Evolving Latent Projections - Resolving 2-Way Multicomponent Data .1. Selectivity, Latent-Projective Graph, Datascope, Local Rank, and Unique Resolution.* Analytical Chemistry, 1992. 64(8): p. 936-946.

59. Liang, Y. Z., Kvalheim, O. M., Keller, H. R., Massart, D. L., Kiechle, P., and Erni, F., *Heuristic Evolving Latent Projections - Resolving 2-Way Multicomponent Data .2. Detection and Resolution of Minor Constituents.* Analytical Chemistry, 1992. 64(8): p. 946-953.

60. Keller, H. R. and Massart, D. L., *Peak Purity Control in Liquid-Chromatography with Photodiode-Array Detection by a Fixed Size Moving Window Evolving Factor-Analysis.* Analytica Chimica Acta, 1991. 246(2): p. 379-390.

61. Gong, F., Liang, Y. Z., Fung, Y. S., and Chau, F. T., *Correction of retention time shifts for chromatographic fingerprints of herbal medicines.* Journal of Chromatography A, 2004. 1029(1-2): p. 173-183.

62. Li, B. Y., Hu, Y., Liang, Y. Z., Huang, L. F., Xu, C. J., and Xie, P. S., *Spectral correlative chromatography and its application to analysis of chromatographic fingerprints of herbal medicines.* Journal of Separation Science, 2004. 27(7-8): p. 581-588.

63. Massart, D. L., Vandegiste, B. G. M., Buydens, L. M. C., De Jong, S., Lewi, P. J., and Symeyers-verbeke, J., *Handbook of Chemometrics and qualimetrics Part A & B.* 1998.

64. Toft, J. and Kvalheim, O. M., *Multiarray Resolution Parameter for Multidetection Chromatography - Performance of Alternating Regression, Iterative Target Transformation Factor-Analysis and Heuristic Evolving Latent Projections.* Chemometrics and Intelligent Laboratory Systems, 1994. 25(1): p. 61-75.

65. Grung, B. and Kvalheim, O. M., *Resolution of Multicomponent Profiles with Partial Selectivity - a Comparison of Direct-Methods.* Chemometrics and Intelligent Laboratory Systems, 1995. 29(1): p. 75-87.

66. Liang, Y. Z. and Kvalheim, O. M., *Unique Resolution of Hidden Minor Peaks in Multidetection Chromatography by 1St-Order Differentiation and Orthogonal Projections.* Analytica Chimica Acta, 1993. 276(2): p. 425-440.

67. Liang, Y. Z., Kvalheim, O. M., Rahmani, A., and Brereton, R. G., *A 2-Way Procedure for Background Correction of Chromatographic Spectroscopic Data by Congruence Analysis and Least-Squares Fit of the Zero-Component Regions - Comparison with Double-Centering.* Chemometrics and Intelligent Laboratory Systems, 1993. 18(3): p. 265-279.

68. Gong, F., Liang, Y. Z., Cui, H., Chau, F. T., and Chan, B. T. P., *Determination of volatile components in peptic powder by gas chromatography-mass spectrometry and chemometric resolution.* Journal of Chromatography A, 2001. 909(2): p. 237-247.

69. Leung, A. K. M., Gong, F., Liang, Y. Z., and Chau, F. T., *Analysis of the water soluble constituents of Cordyceps sinensis with heuristic evolving latent projections.* Analytical Letters, 2000. 33(15): p. 3195-3211.

70. Maeder, M., *Evolving Factor-Analysis for the Resolution of Overlapping Chromatographic Peaks.* Analytical Chemistry, 1987. 59(3): p. 527-530.

71. Liang, Y. Z. and Kvalheim, O. M., *Heuristic Evolving Latent Projections - Resolving Hyphenated Chromatographic Profiles by Component Stripping.* Chemometrics and Intelligent Laboratory Systems, 1993. 20(2): p. 115-125.

72. Gong, F., Liang, Y. Z., Xie, P. S., and Chau, F. T., *Information theory applied to chromatographic fingerprint of herbal medicine for quality control.* Journal of Chromatography A, 2003. 1002(1-2): p. 25-40.

73. Beebe, K. P., Pell, R. J., and Seasholtz, M. B., *Chemometrics: A practical guide*. 1998, New York.

74. Huang, Y. Z. and Pu, F. D., *Studies on the chemical components of the essential oil from the rhizome of Ligusticum sinense Oliv. cv. Chuanxiong Hort.* . 藥學學報, 1988. 23: p. 426-429.

75. Zhong, F. L., Yang, L. J., Ji, L., Hu, S. L., and Fu, G. F., *Studies on the essential oils in Ligusticum chuanxiong Hort. of different Habitats and species.* China Journal of Chinese Materia Medica, 1996. 21: p. 147-151.

76. Song, P. S., Ma, X., Zhang, Q. Z., and Wang, Q. Z., *芎劳(川芎)的本草考証及歷史演變*. China Journal of Chinese Materia Medica, 2000. 25: p. 434.

77. Beijing Universtiy of Traditional Chinese Medicine, *The Chinese Materia Medica*. 1998, Beijing.

78. Tang, W. and Eisenbrand, G., *Chinese Drugs of Plant Origin: Chemistry, Pharmacology, and Use in Traditional and Modern Medicine*. 1992: Springer-Verlag.

79. Chen, X. G., Kong, L., Su, X. Y., Fu, H. J., Ni, J. Y., Zhao, R. H., and Zou, H. F., *Separation and identification of compounds in Rhizoma chuanxiong by comprehensive two-dimensional liquid chromatography coupled to mass spectrometry.* Journal of Chromatography A, 2004. 1040(2): p. 169-178.

80. Yi, T., Leung, K. S. Y., Lu, G. H., Zhang, H., and Chan, K., *Identification and comparative determination of senkyunolide A in traditional Chinese medicinal plants Ligusticum chuanxiong and Angelica sinensis by HPLC coupled with DAD and ESI-MS.* Chemical & Pharmaceutical Bulletin, 2005. 53(11): p. 1480-1483.

81. Yi, T., Leung, K. S. Y., Lu, G. H., Chan, K., and Zhang, H., *Simultaneous qualitative and quantitative analyses of the major constituents in the rhizome of Ligusticum chuanxiong using HPLC-DAD-MS.* Chemical & Pharmaceutical Bulletin, 2006. 54(2): p. 255-259.

82. Li, S. L., Chan, S. S. K., Lin, G., Ling, L., Yan, R., Chung, H. S., and Tam, Y. K., *Simultaneous analysis of seventeen chemical ingredients of Ligusticum chuanxiong by on-line high performance liquid chromatography-diode array detector-mass spectrometry.* Planta Medica, 2003. 69(5): p. 445-451.

83. Li, H. X., Ding, M. Y., and Yu, J. Y., *Separation and identification of the phthalic anhydride derivatives of liqusticum Chuanxiong Hort by GC-MS, TLC, HPLC-DAD, and HPLC-MS.* Journal of Chromatographic Science, 2002. 40(3): p. 156-161.

84. Yan, R., Li, S. L., Chung, H. S., Tam, Y. K., and Lin, G., *Simultaneous quantification of 12 bioactive components of Ligusticum chuanxiong Hort. by high-performance liquid chromatography.* Journal of Pharmaceutical and Biomedical Analysis, 2005. 37(1): p. 87-95.

85. Kong, L., Yu, H., Zou, N., Sun, L., Wu, L., and Ni, J., *Determination of the active ingredients in Rhizoma chuanxiong by high performance liquid chromatography-mass spectrometry.* Fenxi Huaxue, 2004. 32: p. 1501-1504.

86. Zhang, X. Z., Xiao, H. B., Xu, Q., Li, X. L., Wang, L. N., and Liang, X. M., *Characterization of phthalides in Ligusticum chuanxiong by liquid chromatographic-atmospheric pressure chemical ionization-mass spectrometry.* Journal of Chromatographic Science, 2003. 41(8): p. 428-433.

87. Deng, C. H., Ji, J., Wang, X. C., and Zhang, X. M., *Development of pressurized hot water extraction followed by headspace solid-phase microextraction and gas chromatography-mass spectrometry for determination of ligustilides in Ligusticum chuanxiong and Angelica sinensis.* Journal of Separation Science, 2005. 28(11): p. 1237-1243.

88. Liang, M. J., He, L. C., and Li, Y. M., *Analysis of the component and fingerprint of effective part of Ligusticum Chuanxiong Hort. by gas chromatography-mass spectrometry.* Zhipu Xuebao, 2004. 25: p. 150-154.

89. Hong, G., Wang, X., Le, J., Zhang, D., Chai, Y., and Liu, Z., *Supercritical fluid extraction of essential oil from dry rhizome of Ligusticum chuanxiong Hort and their characterization by GC/MS.* Journal of Chinese Pharmaceutical Sciences, 2002. 11: p. 31-34.

90. Li, H. X., Ding, M. Y., Lv, K., and Yu, J. Y., *Determination of the active ingredients in Chuanxiong by HPLC, HPLC-MS, and EI-MS.* Journal of Liquid Chromatography & Related Technologies, 2001. 24(13): p. 2017-2031.

91. Zschocke, S., Liu, J. H., Stuppner, H., and Bauer, R., *Comparative study of roots of Angelica sinensis and related umbelliferous drugs by thin layer chromatography, high-performance liquid chromatography, and liquid chromatography mass spectrometry.* Phytochemical Analysis, 1998. 9(6): p. 283-290.

92. Li, H. B. and Chen, F., *Preparative isolation and purification of chuanxiongzine from the medicinal plant Ligusticum chuanxiong by high-speed counter-current chromatography.* Journal of Chromatography A, 2004. 1047(2): p. 249-253.

93. Cao, Y., Wang, T., Wang, Y. T., Yu, Y. Z., and Bi, K., *HPLC determination of tetramethylpyrazine in Ligusticum chuanxiong Hort.* . Yaowu Fenxi Zazhi, 2005. 25: p. 278-280.

94. Lu, G. H., Chan, K., Chan, C. L., Leung, K., Jiang, Z. H., and Zhao, Z. Z., *Quantification of ligustilides in the roots of Angelica sinensis and related umbelliferous medicinal plants by high-performance liquid chromatography and liquid chromatography-mass spectrometry.* Journal of Chromatography A, 2004. 1046(1-2): p. 101-107.

95. Lao, S. C., Li, S. P., Kan, K. K. W., Li, P., Wan, J. B., Wang, Y. T., Dong, T. T. X., and Tsim, K. W. K., *Identification and quantification of 13 components in Angelica sinensis (Danggui) by gas chromatography-mass spectrometry coupled with pressurized liquid extraction.* Analytica Chimica Acta, 2004. 526(2): p. 131-137.

96. Budavari, S., *Merck Index: an encyclopedia of chemicals, drugs, and biologicals* 11 ed. 1989, New York: Rahway.

97. Agency for toxic substances and disease registry (2004) *Public Health statement for dichlorobenzenes*.

98. Department of Health in Hong Kong Special Administrative Region, *The Hong Kong Chinese Materia Medica Standards (HKCMMS)*. Vol. 1. 2005, Hong Kong.

99. Li, S. L., Li, P., Sheng, L. H., Li, R. Y., Qi, L. W., and Zhang, L. Y., *Live cell extraction and HPLC-MS analysis for predicting bioactive components of traditional Chinese medicines.* Journal of Pharmaceutical and Biomedical Analysis, 2006. 41(2): p. 576-581.

100. Zhao, K. J., Dong, T. T. X., Tu, P. F., Song, Z. H., Lo, C. K., and Tsim, K. W. K., *Molecular genetic and chemical assessment of radix angelica (Danggui) in China.* Journal of Agricultural and Food Chemistry, 2003. 51(9): p. 2576-2583.

101. Lin, L. Z., He, X. G., Lian, L. Z., King, W., and Elliot, J., *Liquid chromatographic electrospray mass spectrometric study of the phthalides of Angelica sinensis and chemical changes of Z-ligustilide.* Journal of Chromatography A, 1998. 810(1-2): p. 71-79.

102. Li, P., Li, S. P., Lao, S. C., Fu, C. M., Kan, K. K. W., and Wang, Y. T., *Optimization of pressurized liquid extraction for Z-ligustilide, Z-butylidenephthalide and ferulic acid in Angelica sinensis.* Journal of Pharmaceutical and Biomedical Analysis, 2006. 40(5): p. 1073-1079.

103. Huang, L. F., Li, B. Y., Liang, Y. Z., Guo, F. Q., and Wang, Y. L., *Application of combined approach to analyze the constituents of essential oil from Dong quai*. Analytical and Bioanalytical Chemistry, 2004. 378(2): p. 510-517.

104. Cao, W., Li, X. Q., Liu, L., YANG, T. H., Li, C., Fan, H. T., Jia, M., Lu, Z. G., and Mei, Q. M., *Structure f anti-tumor polysaccharide from Angelica sinensis (Oliv.) Diels.* Carbohydrate Polymers, In press.

105. Cao, W., Li, X. Q., Liu, L., Wang, M., Fan, H. T., Li, C., Lv, Z., Wang, X., and Mei, Q., *Structural analysis of water-soluble glucans from the root of Angelica sinensis (Oliv.) Diels* Carbohydrate Research, 2006. 341: p. 1870-1877.

106. Sun, Y. L., Tang, J., Gu, X. H., and Li, D. Y., *Water-soluble polysaccharides from Angelica sinensis (Oliv.) Diels: Preparation, characterization and bioactivity.* International Journal of Biological Macromolecules, 2005. 36(5): p. 283-289.

107. 康廷國, *中藥鑒定學*. 2003, 北京: 中國中醫藥出版社.

108. 中國醫學科學院及中國協和醫科大學葯物研究所、日本大正制葯株式會社, *常用中草葯效相色譜分析*. 1999.

109. Guo, S. H., Zhou, M. H., and Li, T. C., *A study on chemical constituents of the essential oil from the fruit of Phellodendron Chinense Schneid.* Journal of Jinan University (Natural Science & Medicine Edition), 1998. 61-63.

110. Hou, D. Y., Hui, R. H., and Li, T. C., *Study on volatile component of the fruit from Phellodendron Amurense Rupr.* Journal of Chinese Mass Spectrometry society 2001. 22: p. 61-65.

111. 王衡奇, 秦民堅, and 余國奠, *黃柏的化學成分及葯理學研究進展*. 中國野生植物資源, 2001: p. 6-8.

112. Gao, W. H., Lin, S. Y., Jia, L., Gu, X. K., Chen, X. G., and De Hu, Z., *Analysis of protoberberine alkaloids in several herbal drugs and related medicinal preparations by non-aqueous capillary electrophoresis.* Journal of Separation Science, 2005. 28(1): p. 92-97.

113. Cordell, G. A., Saxton, J. E., Shamma, M., and Smith, G. F., *Dictionary of alkaloids*. 1989, New York.

114. 鄭虎占, 董澤宏, and 佘靖, *中藥現代研究與應用*. 1997, 北京.

115. Zhu, C. C., Mo, J. X., and Lin, C. Z., *Fingerprint analysis of Cortex Phellodendri by RP-HPLC.* Traditional Chinese Drug Research and Clinical Pharmacology, 2003. 14: p. 324-327.

116. Ding, Q. and Xu, D., *Determination of jatrorrhizine, palmatine and berberine in Cortex phellodendri by HPLC.* Xibei Zhiwu Xuebao, 2004. 24: p. 2143-2145.

117. Tsai, P. L. and Tsai, T. H., *Simultaneous determination of berberine in rat blood, liver and bile using microdialysis coupled to high-performance liquid chromatography.* Journal of Chromatography A, 2002. 961(1): p. 125-130.

118. Liu, Y. M. and Sheu, S. J., *Determination of Quaternary Alkaloids from Phellodendri Cortex by Capillary Electrophoresis.* Journal of Chromatography, 1993. 634(2): p. 329-333.

119. Wu, W., Song, F. R., Yan, C. Y., Liu, Z. Q., and Liu, S. Y., *Structural analyses of protoberberine alkaloids in medicine herbs by using ESI--FT-ICR-MS and HPLC-ESI-MSn.* Journal of Pharmaceutical and Biomedical Analysis, 2005. 37(3): p. 437-446.

120. Chen, Y. R., Wen, K. C., and Her, G. R., *Analysis of coptisine, berberine and palmatine in adulterated Chinese medicine by capillary electrophoresis-electrospray ion trap mass spectrometry.* Journal of Chromatography A, 2000. 866(2): p. 273-280.

121. Li, C. Y., Lu, H. J., Lin, C. H., and Wu, T. S., *A rapid and simple determination of protoberberine alkaloids in cortex phellodendri by H-1 NMR and its application for quality control of commercial traditional Chinese medicine prescriptions.* Journal of Pharmaceutical and Biomedical Analysis, 2006. 40(1): p. 173-178.

122. Lee, H. S., Eom, Y. E., and Eom, D. O., *Narrowbore high performance liquid chromatography of berberine and palmatine in crude drugs and pharmaceuticals with ion-pair extraction using cobalt thiocyanate reagent.* Journal of Pharmaceutical and Biomedical Analysis, 1999. 21(1): p. 59-63.

123. ASTM, *Standard Practice for Near Infrared Qualitative Analysis*. 2004. p. E1790-00.

124. Burns, D. A. and Ciurczak, E. W., *Handbook of Near-Infrared Analysis*. 2001, New York. 81.

125. Siester, M. C., Ozaki, Y., Kawata, S., and Heise, H. M., *Near Infrared Spectroscopy: Priniciples, Instruments, applications*. 2002.

126. Pasikatan, M. C., Steele, J. L., Spillman, C. K., and Haque, E., *Near infrared reflectance spectroscopy for online particle size analysis of powders and ground materials.* Journal of near Infrared Spectroscopy, 2001. 9(3): p. 153-164.

127. Kortüm, G., *Reflectance Spectroscopy*. 1969, New York.

128. Blanco, M., Coello, J., Iturriaga, H., Maspoch, S., and de la Pezuela, C., *Near-infrared spectroscopy in the pharmaceutical industry.* Analyst, 1998. 123(8): p. 135R-150R.

129. Wu, W., Walczak, B., Massart, D. L., Prebble, K. A., and Last, I. R., *Spectral Transformation and Wavelength Selection in near-Infrared Spectra Classification.* Analytica Chimica Acta, 1995. 315(3): p. 243-255.

130. Salamin, P. A., Cornelis, Y., and Bartels, H., *Identification of Chemical-Substances by Their near-Infrared Spectra.* Chemometrics and Intelligent Laboratory Systems, 1988. 3(4): p. 329-333.

131. Hailey, P. A., Doherty, P., Tapsell, P., Oliver, T., and Aldridge, P. K., *Automated system for the on-line monitoring of powder blending processes using near-infrared spectroscopy .1. System development and control.* Journal of Pharmaceutical and Biomedical Analysis, 1996. 14(5): p. 551-559.

132. Downey, G., Robert, P., and Bertrand, D., *Qualitative analysis in the NIR region: a whole spectrum approach.* Analytical Process, 1992. 8: p. 29.

133. McClure, W. F. and Davies, A. M. C., *More on derivatives: Part 1: Segments, gaps and ghosts.*, in *NIR News*. 1993. p. 12.

134. McClure, W. F. and Davies, A. M. C., *More on derivatives. Part 2: Band shifting and noise.*, in *NIR News*. 1994. p. 14-16.

135. McClure, W. F. and Davies, A. M. C., *More on derivatives: Part 3. Computing derivatives with Fourier coefficients.* , in *NIR News*. 1994. p. 14-15.

136. Isaksson, T. and Naes, T., *The Effect of Multiplicative Scatter Correction (Msc) and Linearity Improvement in Nir Spectroscopy.* Applied Spectroscopy, 1988. 42(7): p. 1273-1284.

137. Naes, T., Isaksson, T., and Kowalski, B., *Locally Weighted Regression and Scatter Correction for near-Infrared Reflectance Data.* Analytical Chemistry, 1990. 62(7): p. 664-673.

138. Isaksson, T. and Kowalski, B., *Piece-Wise Multiplicative Scatter Correction Applied to near-Infrared Diffuse Transmittance Data from Meat-Products.* Applied Spectroscopy, 1993. 47(6): p. 702-709.

139. Martens, H. and Stark, E., *Extended multiplicative signal correction and spectral interference subtraction: New preprocessing methods for near infrared spectroscopy.* Journal of Pharmaceutical and Biomedical Analysis, 1991. 9: p. 625-635.

140. Karstang, T. V. and Manne, R., *Optimized Scaling - a Novel-Approach to Linear Calibration with Closed Data Sets.* Chemometrics and Intelligent Laboratory Systems, 1992. 14(1-3): p. 165-173.

141. Isaksson, T., Wang, Z., and Kowalski, B., *Optimised scaling (OS-2) regression applied to near infrared diffuse spectroscopy data from food products.* Journal of near Infrared Spectroscopy, 1993. 1: p. 85-97.

142. Candolfi, A., De Maesschalck, R., Jouan-Rimbaud, D., Hailey, P. A., and Massart, D. L., *The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra.* Journal of Pharmaceutical and Biomedical Analysis, 1999. 21(1): p. 115-132.

143. Barnes, R. J., Dhanoa, M. S., and Lister, S. J., *Standard Normal Variate Transformation and De-Trending of near-Infrared Diffuse Reflectance Spectra.* Applied Spectroscopy, 1989. 43(5): p. 772-777.

144. Dhanoa, M. S., Lister, S. J., and Barnes, R. J., *On the Scales Associated with near-Infrared Reflectance Difference Spectra.* Applied Spectroscopy, 1995. 49(6): p. 765-772.

145. Dhanoa, M. S., Lister, S. J., Sanderson, R., and Barnes, R. J., *The link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) transformations of NIR spectra.* Journal of near Infrared Spectroscopy, 1994. 2: p. 43-47.

146. Savitzky, A. and Golay, M. J. E., *Smoothing and Differentiation of Data by Simplified Least Squares Procedures.* . Analytical Chemistry, 1964. 36: p. 1627-1639.

147. Philip, D. L. and Lanny, D. V., *Pharmaceutical Process Control by NIR Spectroscopy.* American Pharmaceutical Review, 2001. 3: p. 99-103.

148. Hildrum, K. I., Isaksson, T., Naes, T., and Tandberg, A., *Near Infrared Spectroscopy (briding Gap between Data Analysis and NIR Applications)*. 1992, Chichester.

149. Svensson, O., Josefson, M., and Langkilde, F. W., *Classification of chemically modified celluloses using a near-infrared spectrometer and soft independent modeling of class analogies.* Applied Spectroscopy, 1999. 51: p. 1826-1835.

150. Woo, Y. A., Kim, H. J., Cho, J., and Chung, H., *Discrimination of herbal medicines according to geographical origin with near infrared reflectance spectroscopy and pattern recognition techniques.* Journal of Pharmaceutical and Biomedical Analysis, 1999. 21(2): p. 407-413.

151. Woo, Y. A., Kim, H. J., and Chung, H., *Classification of cultivation area of ginseng radix with NIR and Raman spectroscopy.* Analyst, 1999. 124(8): p. 1223-1226.

152. Woo, Y. A., Kim, H. J., Ze, K. R., and Chung, H., *Near-infrared (NIR) spectroscopy for the non-destructive and fast determination of geographical origin of Angelicae gigantis Radix.* Journal of Pharmaceutical and Biomedical Analysis, 2005. 36(5): p. 955-959.

153. Laasonen, M., Harmia-Pulkkinen, T., Simard, C. L., Michiels, E., Rasanen, M., and Vuorela, H., *Fast identification of Echinacea purpurea dried roots using near-infrared spectroscopy.* Analytical Chemistry, 2002. 74(11): p. 2493-2499.

154. Workman, J. J., Mobley, P. R., Kowalski, B. R., and Bro, R., *Review of chemometrics applied to spectroscopy: 1985-95.* Applied Spectroscopy Review, 1996. 31(1-2): p. 73-124.

155. Martens, H. and Naes, T., *Multivariate Calibration*. 1989, New York.

156. Buchanan, B. and Honigs, D., *Trends in near-Infrared Analysis.* Trac-Trends in Analytical Chemistry, 1986. 5(6): p. 154-157.

157. Thomas, E. V. and Haaland, D. M., *Comparison of Multivariate Calibration Methods for Quantitative Spectral-Analysis.* Analytical Chemistry, 1990. 62(10): p. 1091-1099.

158. Vogt, N. B., *Polynomial Principal Component Regression - an Approach to Analysis and Interpretation of Complex Mixture Relationships in Multivariate Environmental Data.* Chemometrics and Intelligent Laboratory Systems, 1989. 7(1-2): p. 119-130.

159. Chen, Y. X. and Sorensen, L. K., *Determination of marker constituents in radix Glycyrrhizae and radix Notoginseng by near infrared spectroscopy.* Fresenius Journal of Analytical Chemistry, 2000. 367(5): p. 491-496.

160. Ren, G. X. and Chen, F., *Simultaneous quantification of ginsenosides in American ginseng (Panax quinquefolium) root powder by visible/near-infrared reflectance spectroscopy.* Journal of Agricultural and Food Chemistry, 1999. 47(7): p. 2771-2775.

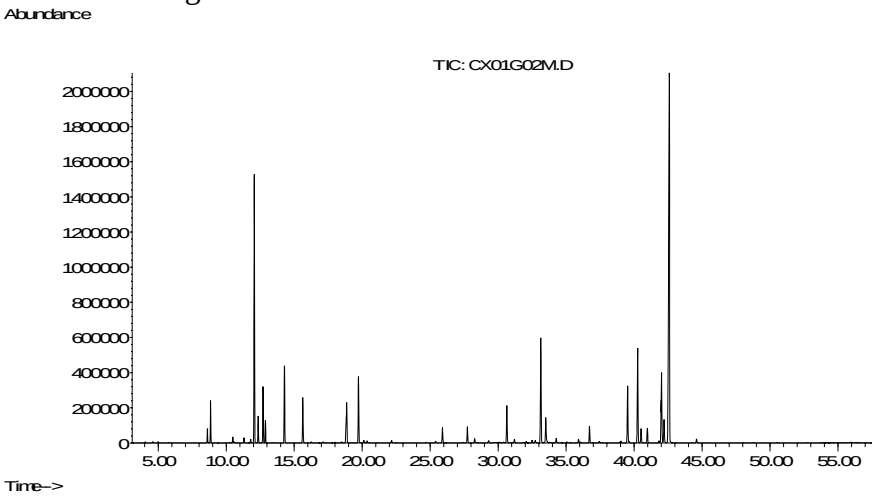
161. Rager, I., Roos, G., Schmidt, P. C., and Kovar, K. A., *Rapid quantification of constituents in St. John's wort extracts by NIR spectroscopy.* Journal of Pharmaceutical and Biomedical Analysis, 2002. 28(3-4): p. 439-446.

162. Curcio, J. A. and Petty, C. C., *The Near Infrared Absorption Spectrum of Liquid Water.* Journal of the Optical Society of America, 1951. 41: p. 302-304.

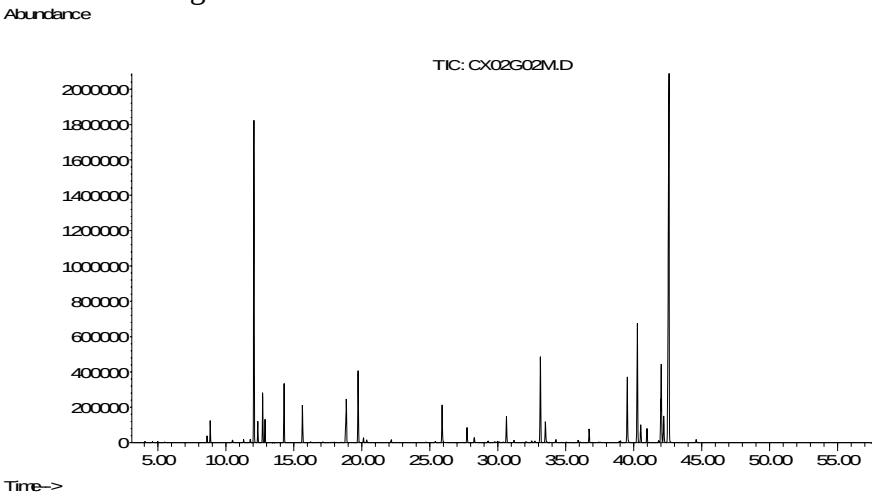
Appendices

Appendix 4.1 Chromatograms of thirty-two *Radix Ligustici chuanxiong* samples obtained from GC-MS study

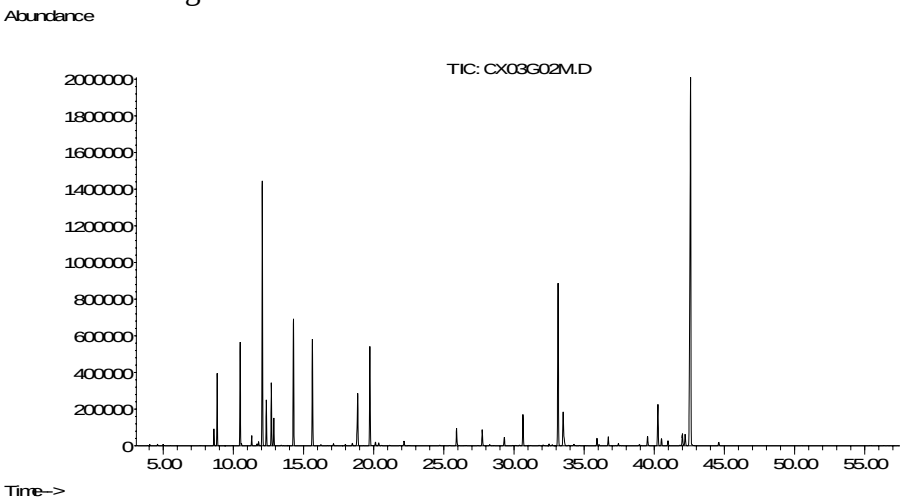
TIC chromatogram of CX1



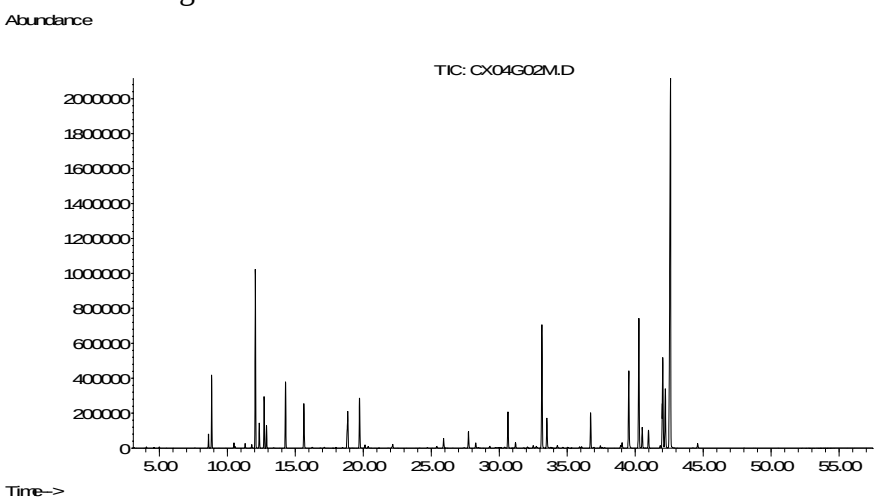
TIC chromatogram of CX2



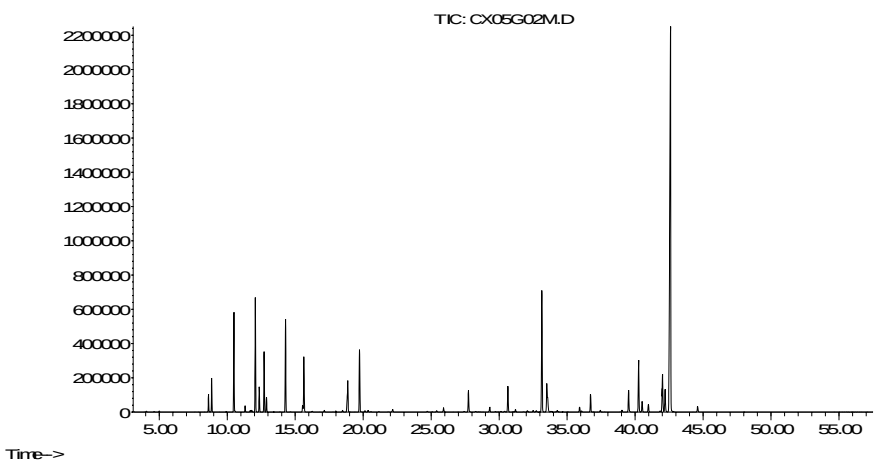
TIC chromatogram of CX3



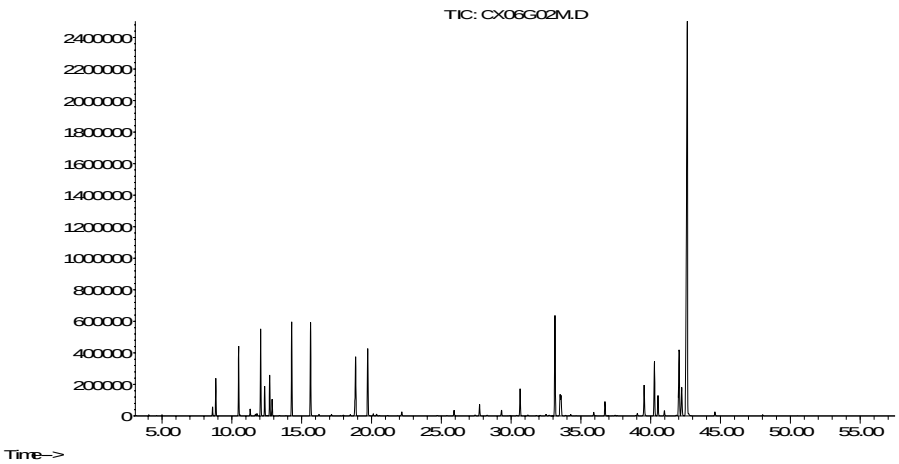
TIC chromatogram of CX4



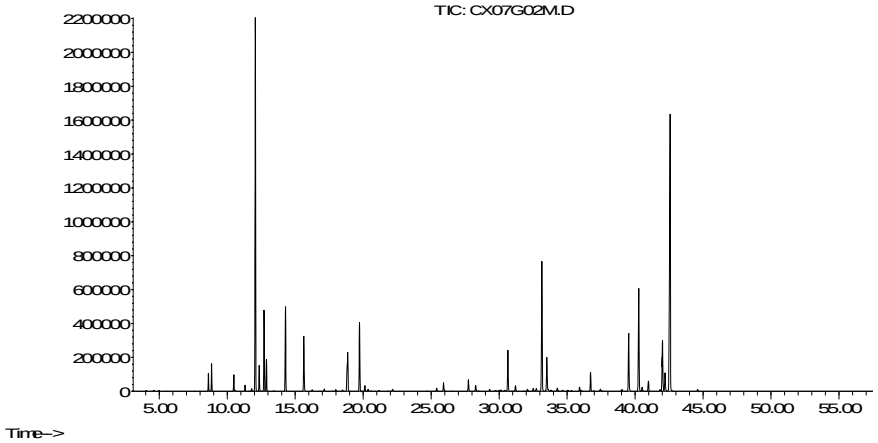
TIC chromatogram of CX5



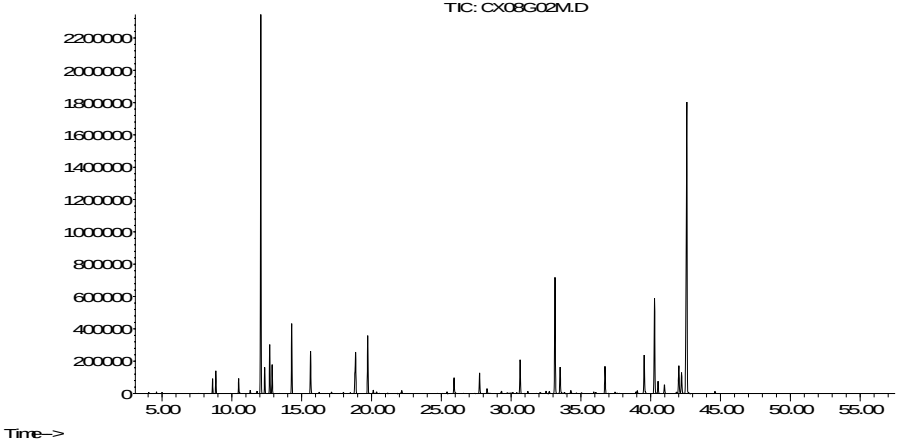
TIC chromatogram of CX6



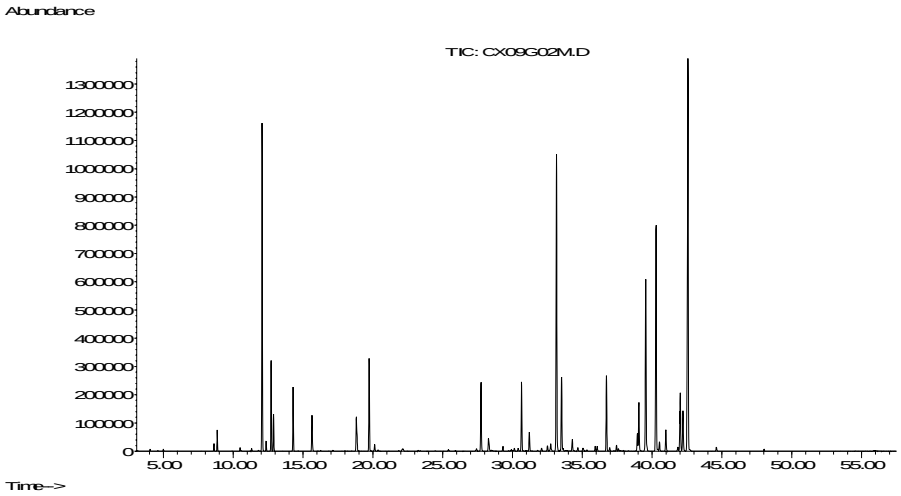
TIC chromatogram of CX7



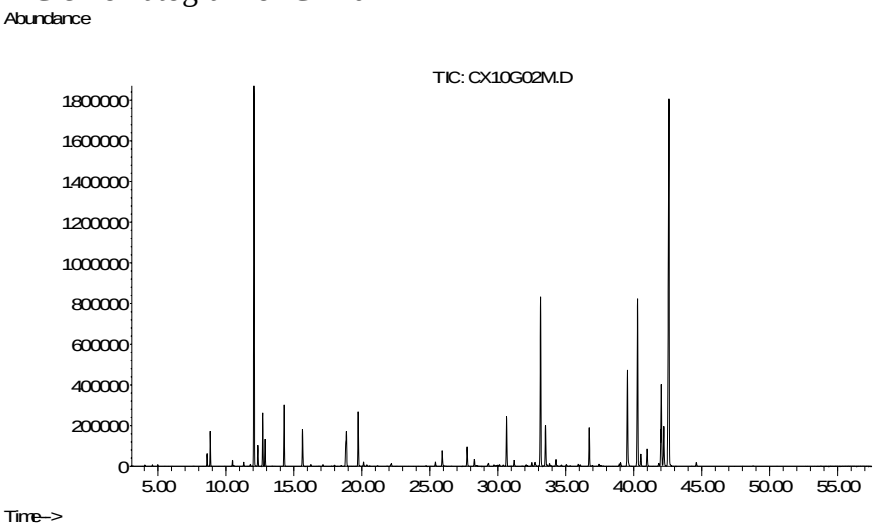
TIC chromatogram of CX8



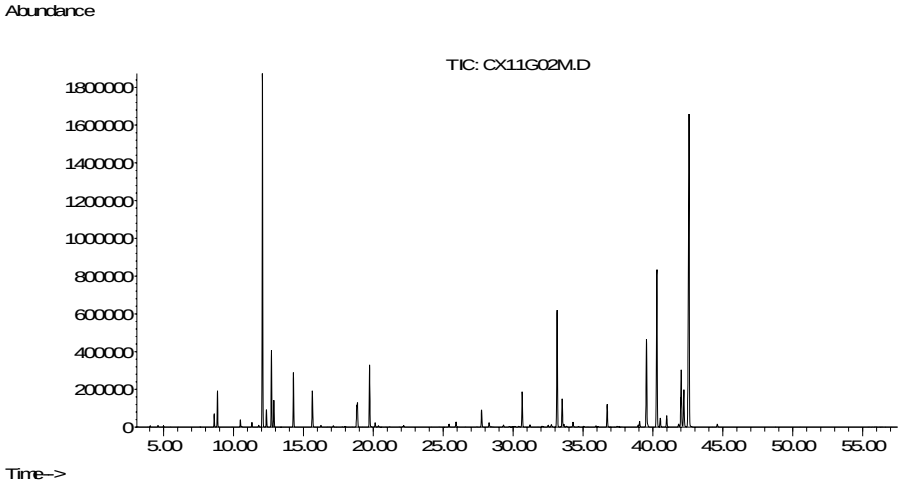
TIC chromatogram of CX9



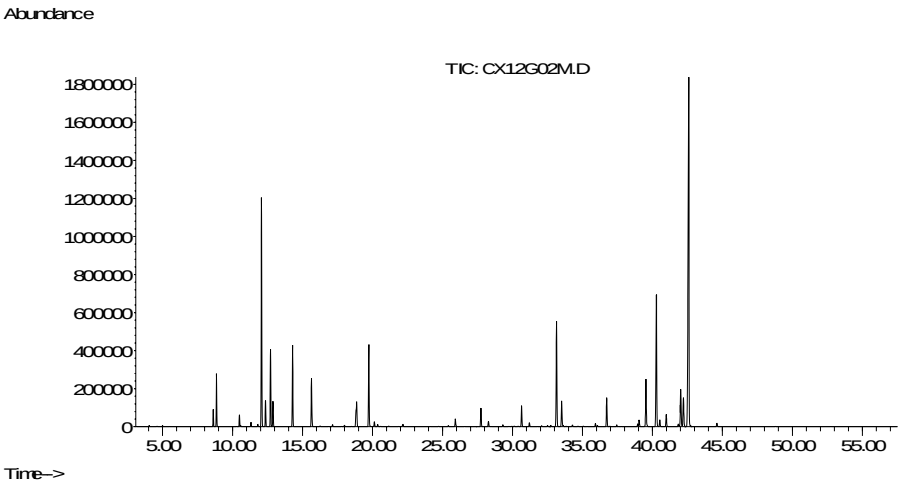
TIC chromatogram of CX10



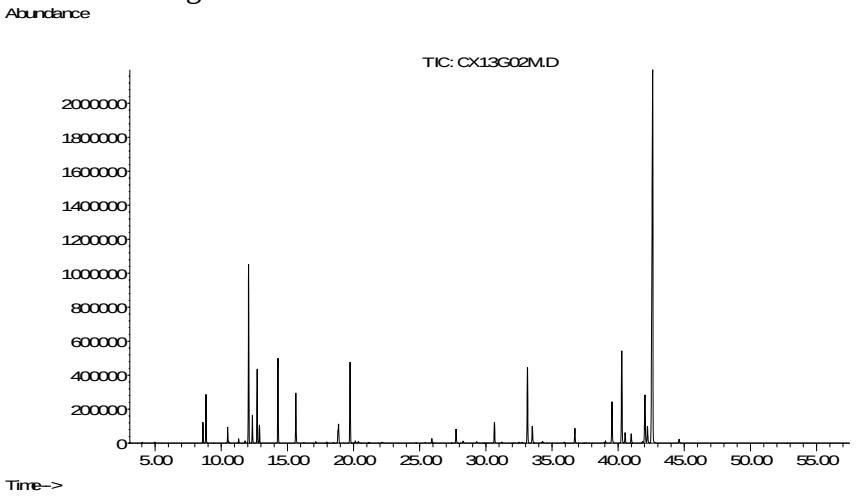
TIC chromatogram of CX11



TIC chromatogram of CX12



TIC chromatogram of CX13

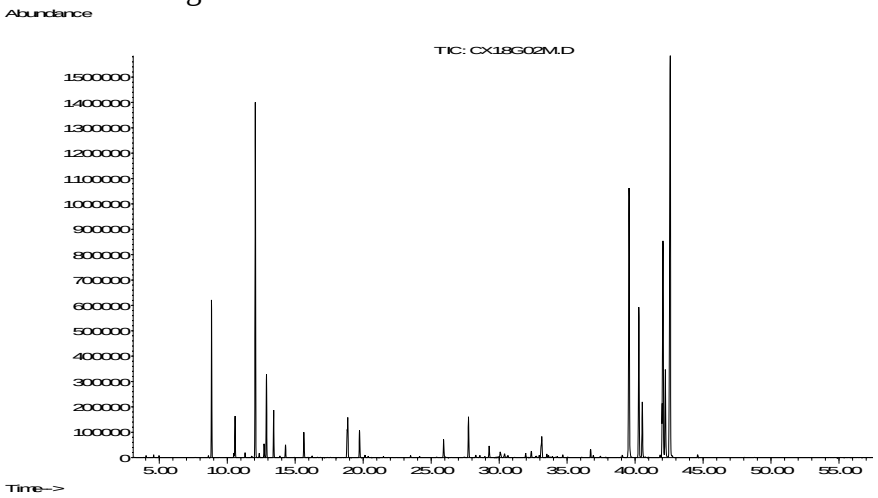


TIC chromatogram of CX14

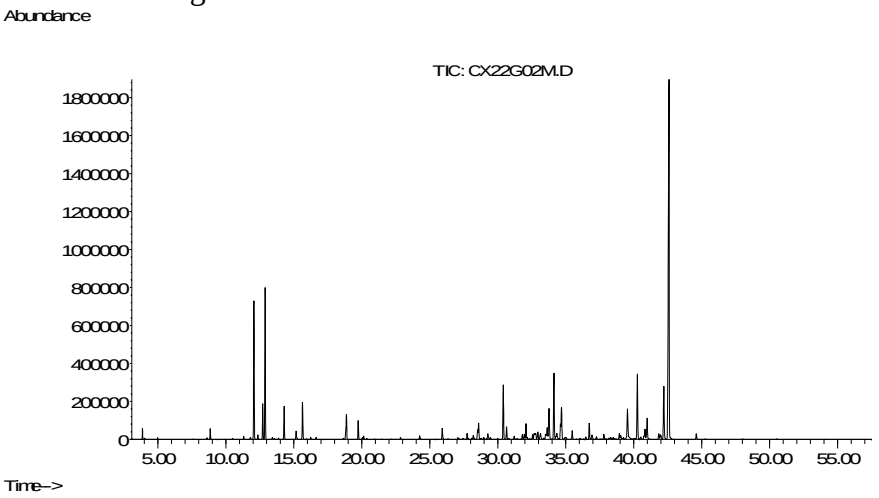
TIC chromatogram of CX15

TIC chromatogram of CX16

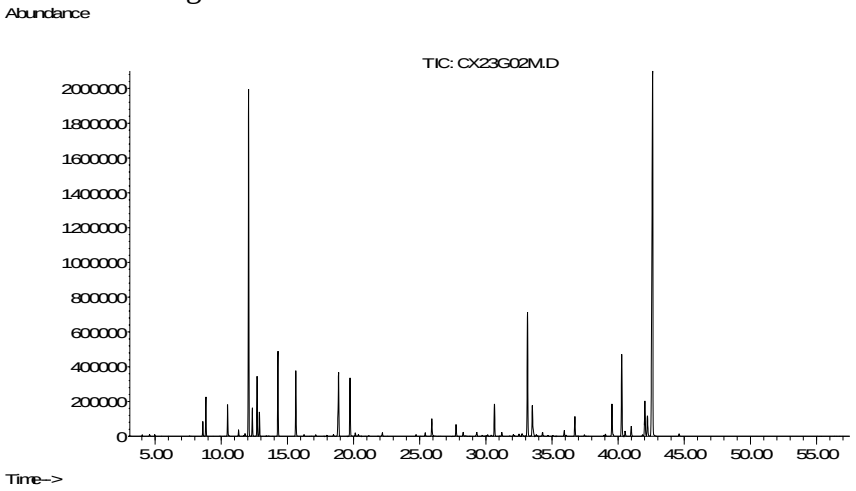
TIC chromatogram of CX17



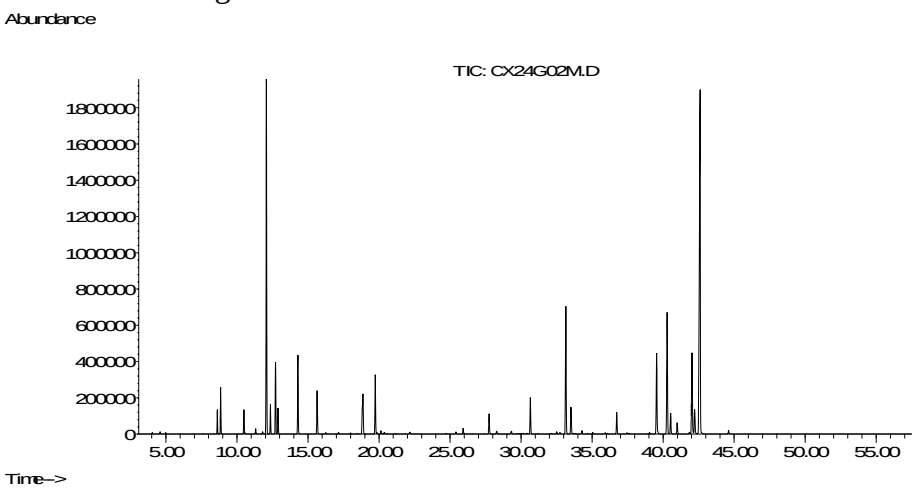
TIC chromatogram of CX18



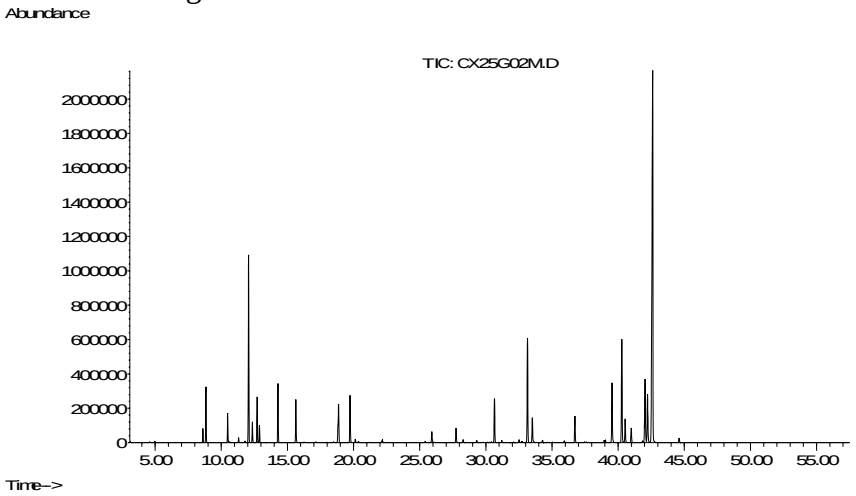
TIC chromatogram of CX19



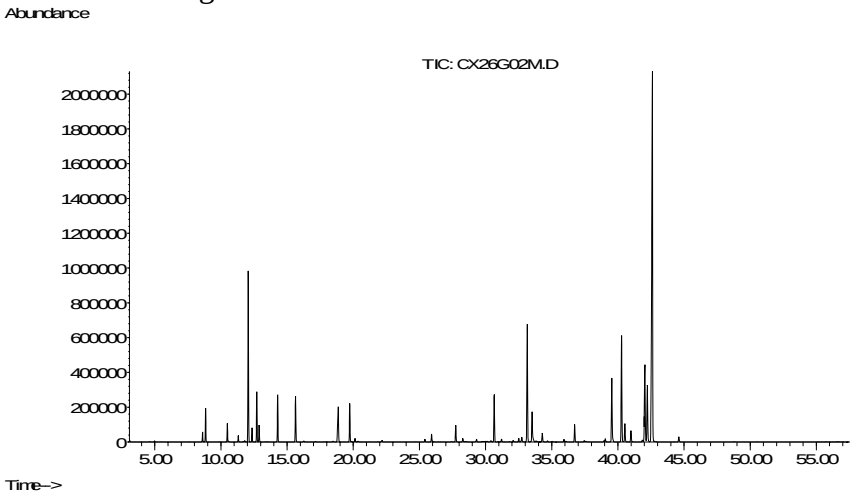
TIC chromatogram of CX20



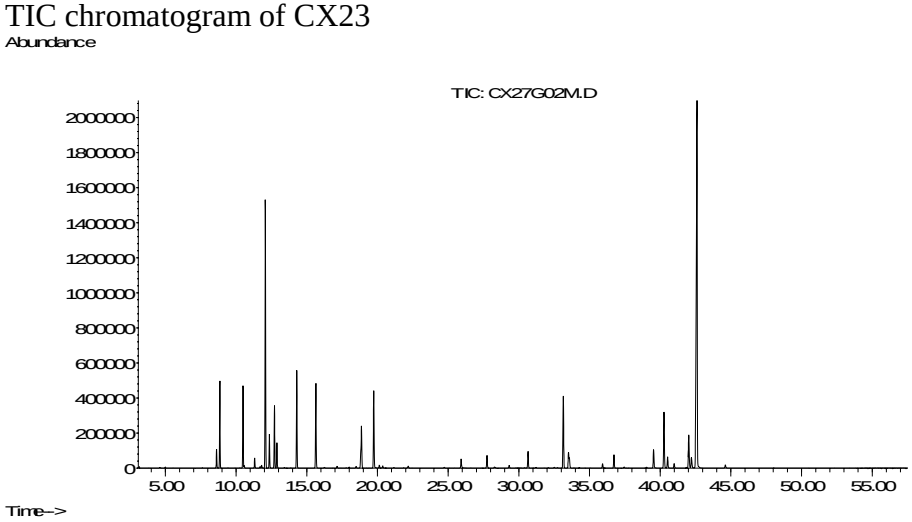
TIC chromatogram of CX21



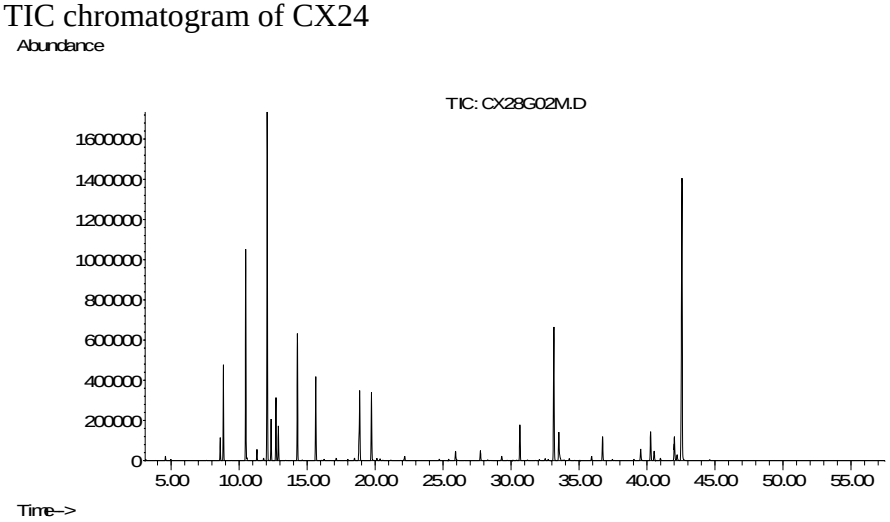
TIC chromatogram of CX22



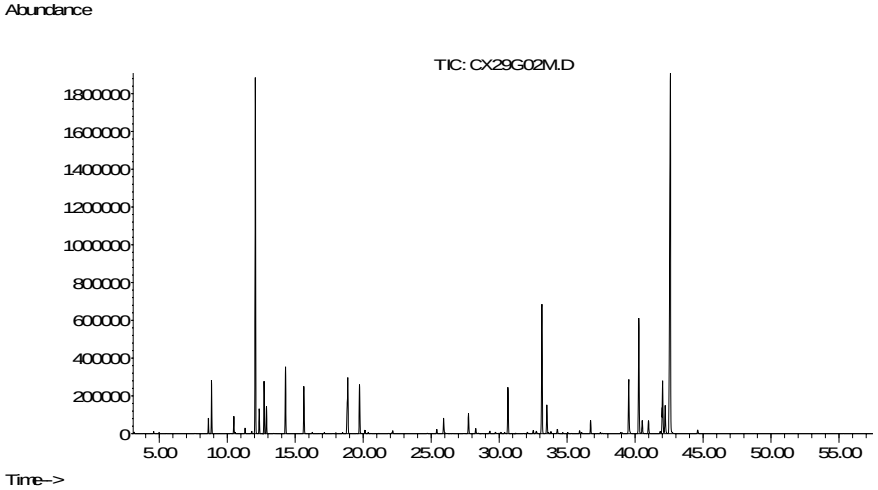
TIC chromatogram of CX23



TIC chromatogram of CX24

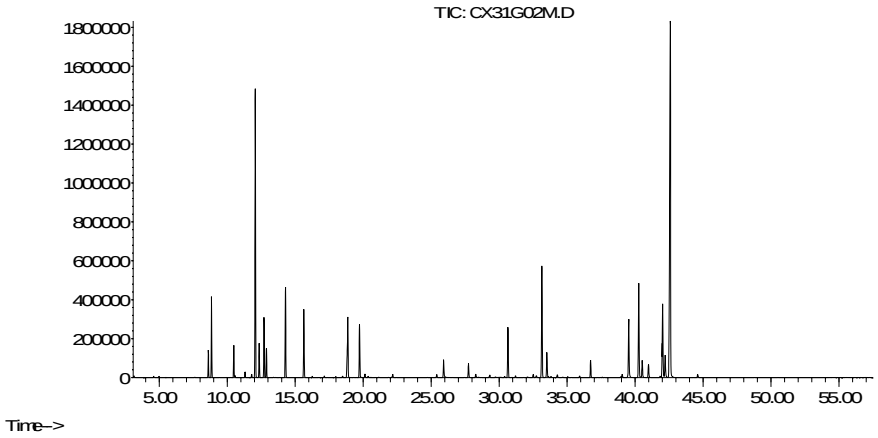


TIC chromatogram of CX25

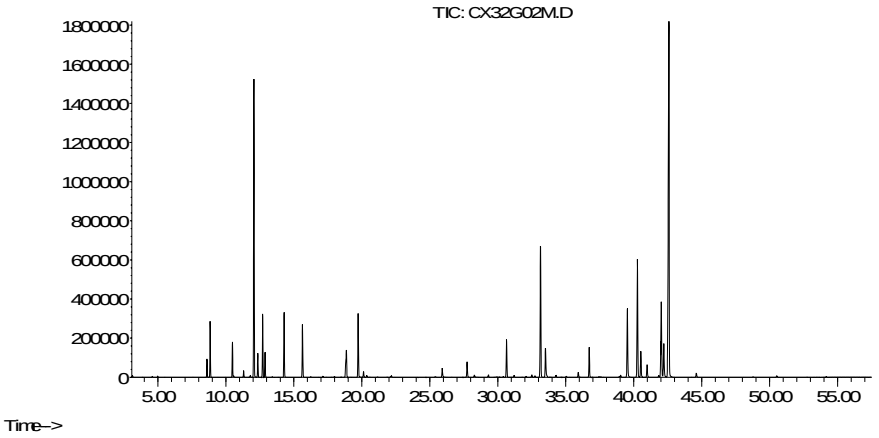


TIC chromatogram of CX26

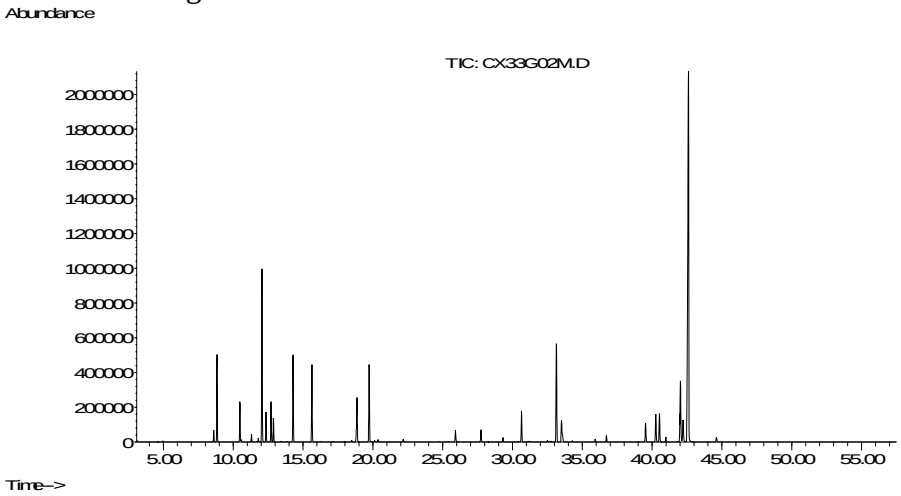
TIC chromatogram of CX27



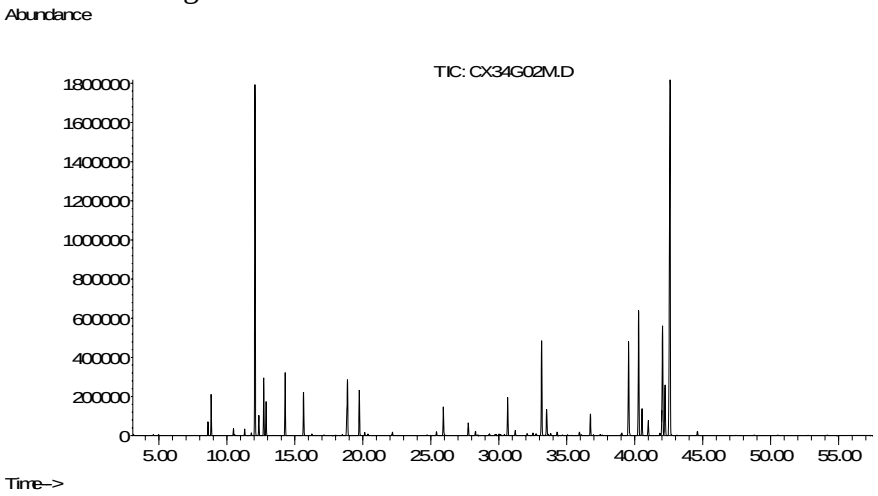
TIC chromatogram of CX28



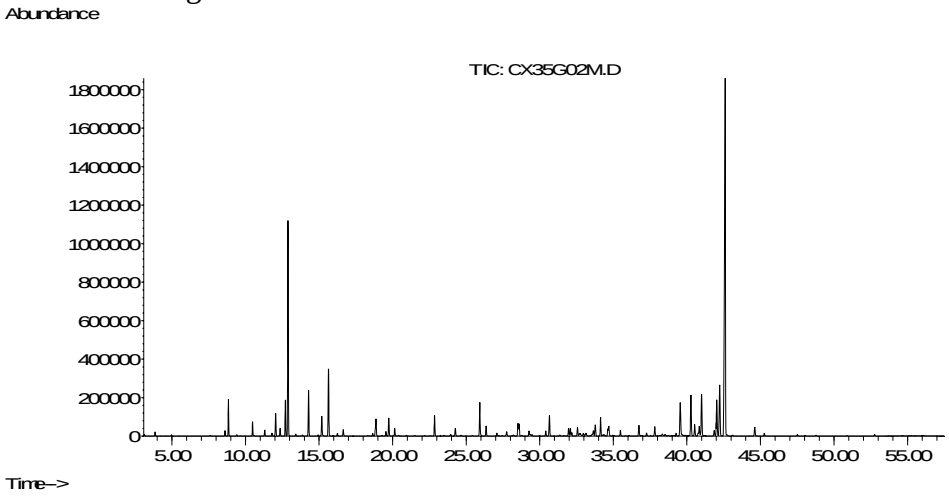
TIC chromatogram of CX29



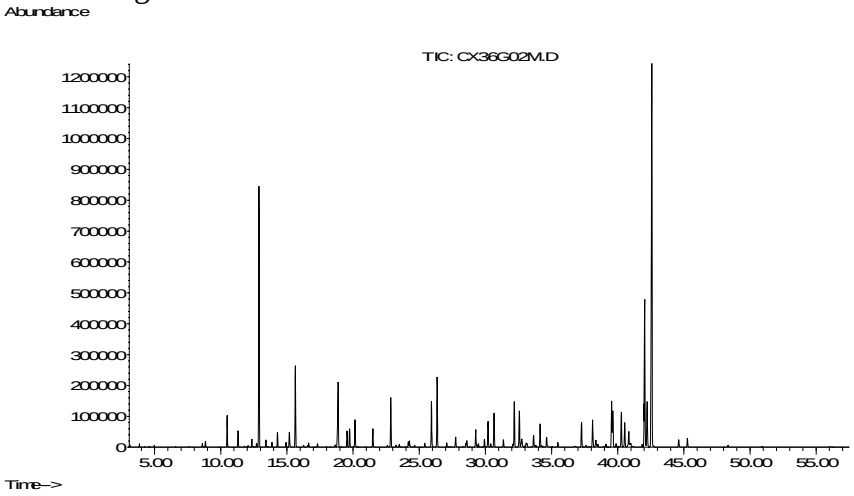
TIC chromatogram of CX30



TIC chromatogram of CX31

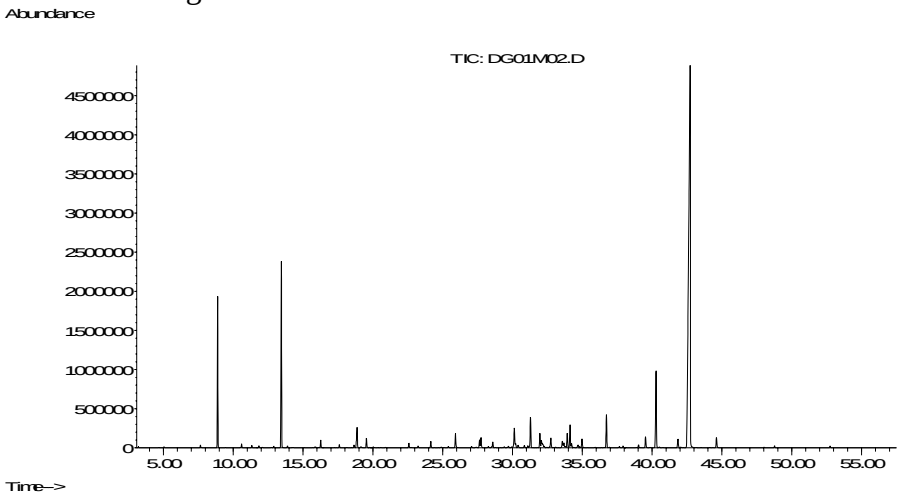


TIC chromatogram of CX32

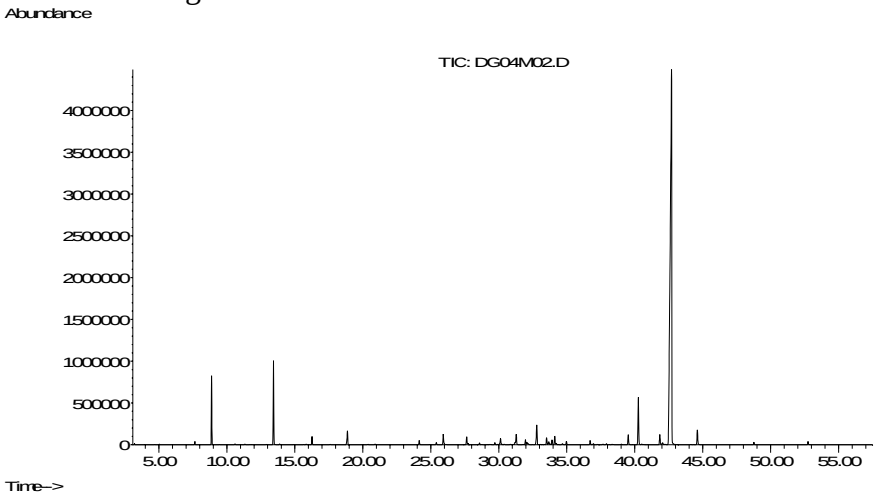


Appendix 5.1 Chromatograms of twenty *Radix Angelicae sinensis* samples obtained from GC-MS study

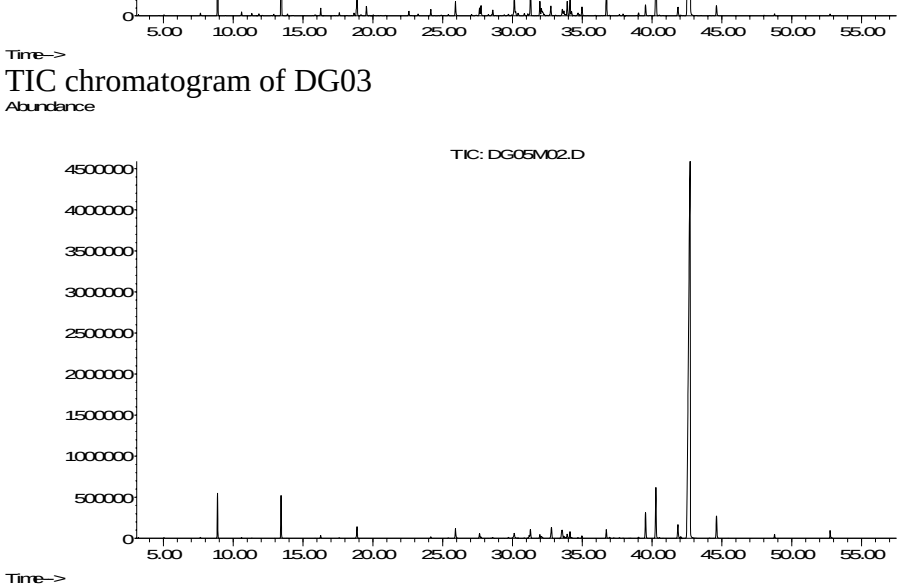
TIC chromatogram of DG01



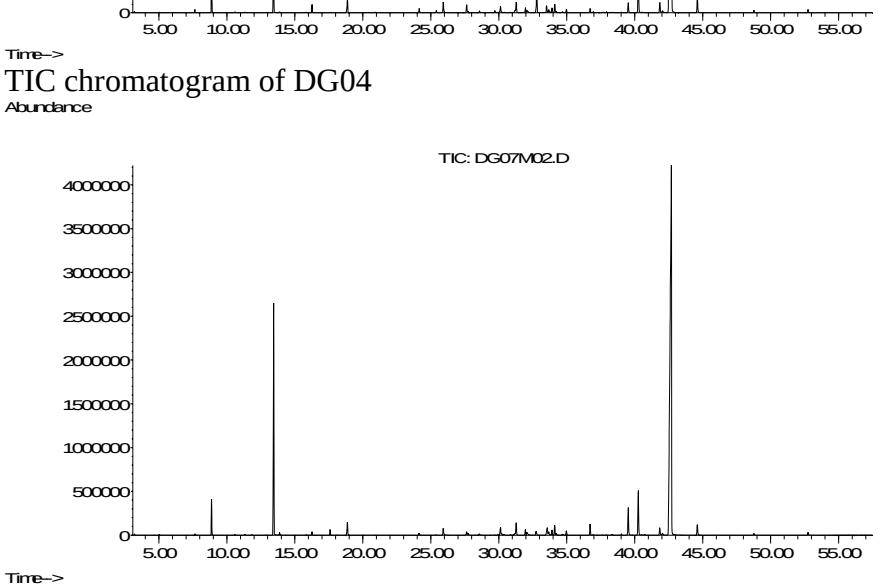
TIC chromatogram of DG02



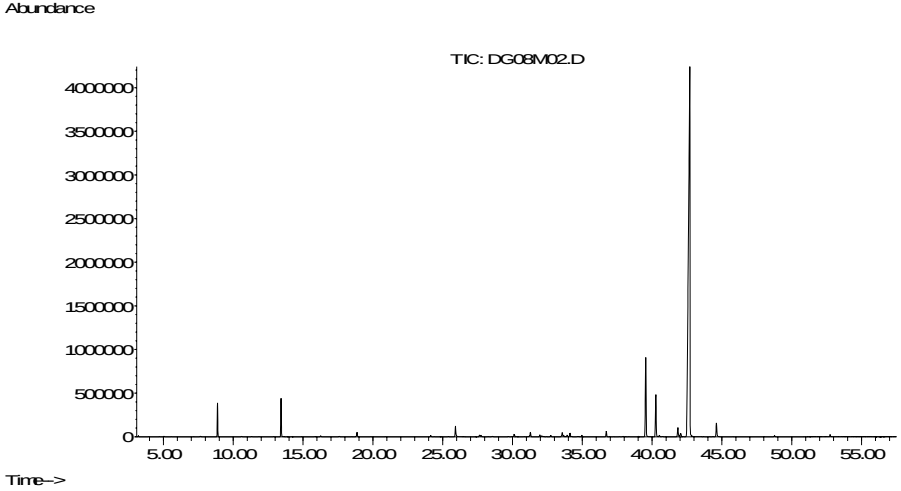
TIC chromatogram of DG03



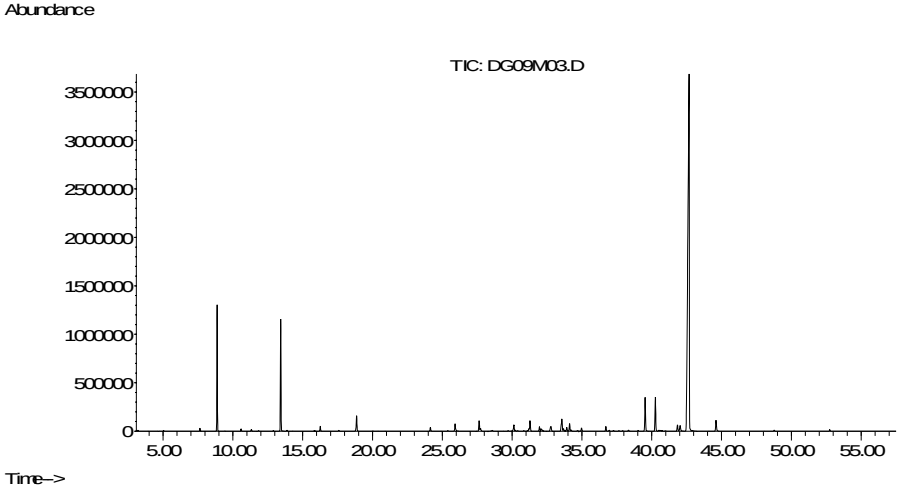
TIC chromatogram of DG04



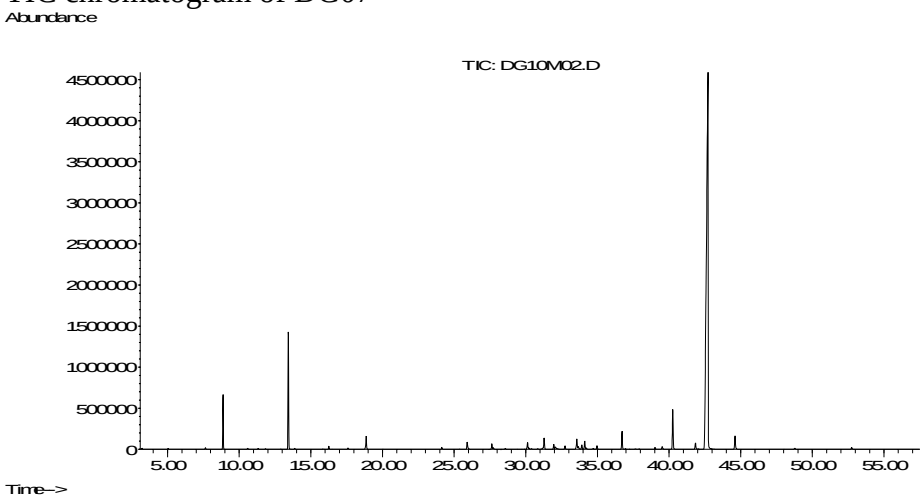
TIC chromatogram of DG05



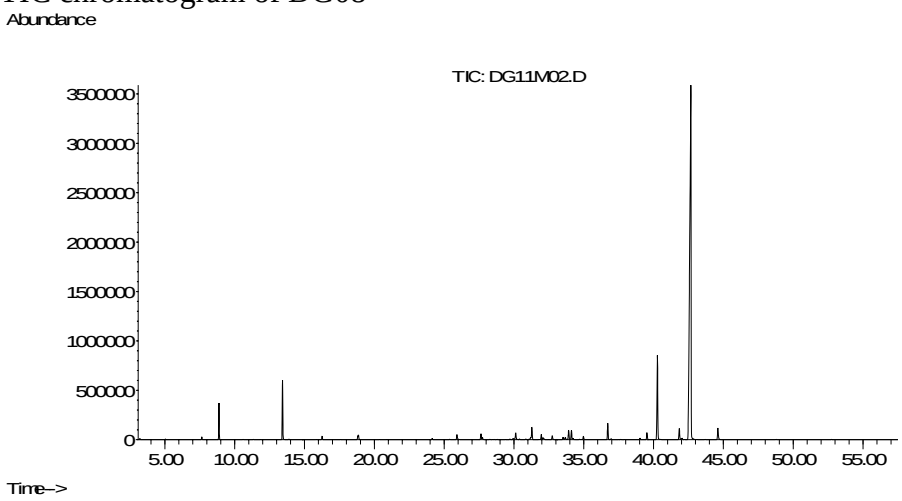
TIC chromatogram of DG06



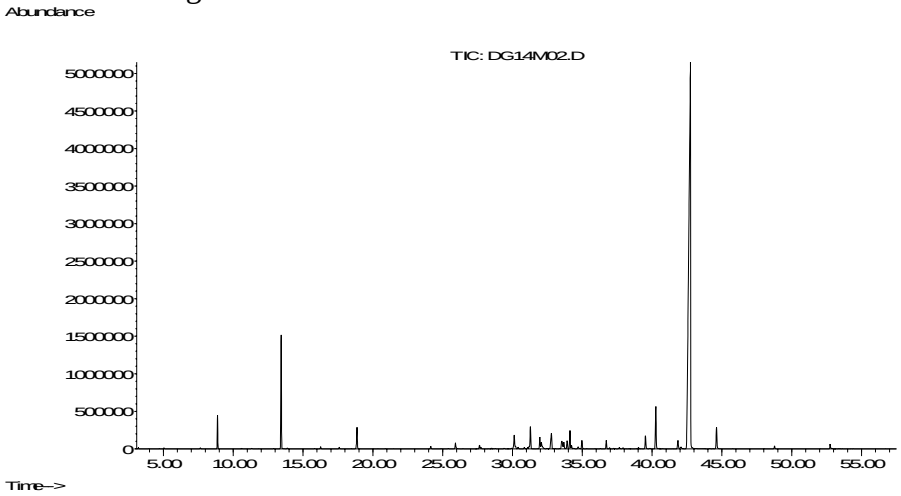
TIC chromatogram of DG07



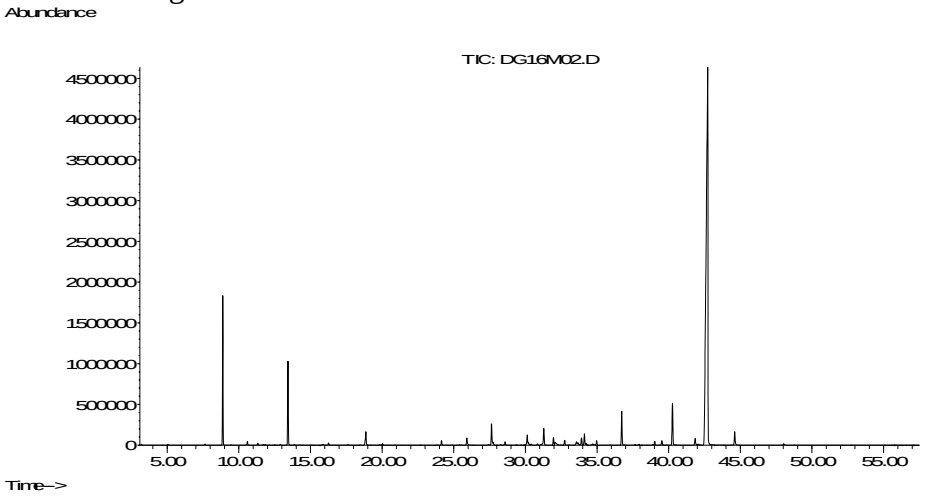
TIC chromatogram of DG08



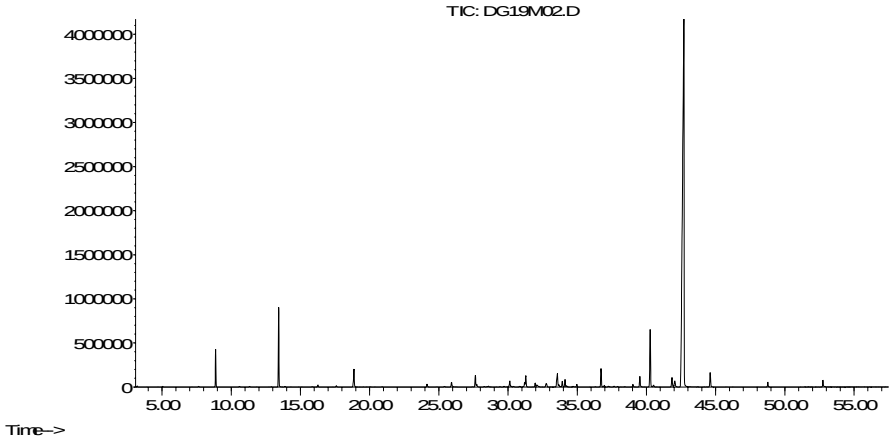
TIC chromatogram of DG09



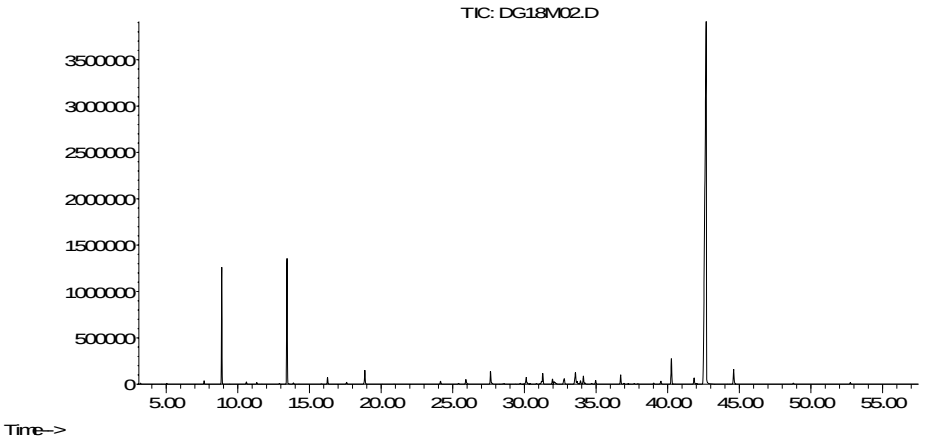
TIC chromatogram of DG10



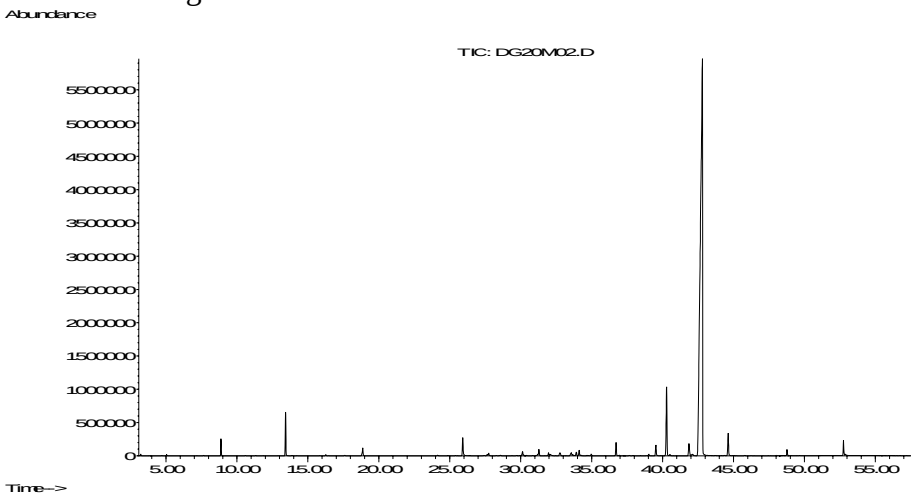
TIC chromatogram of DG11



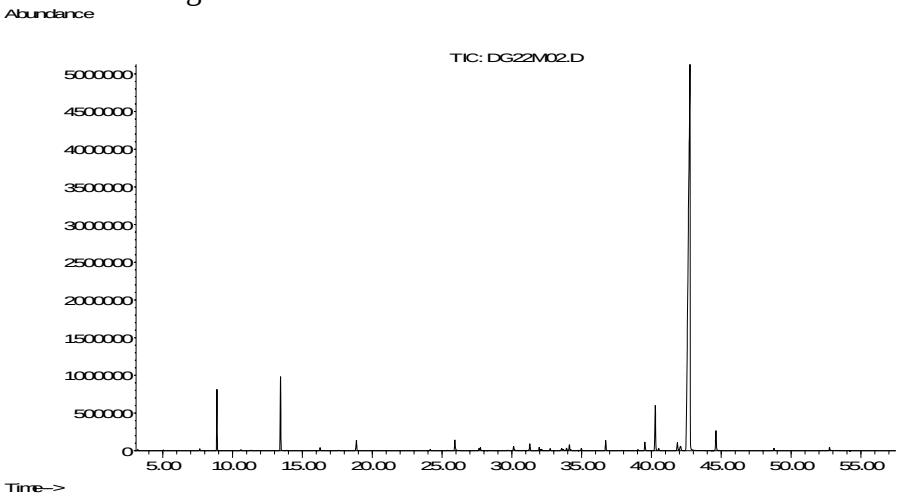
TIC chromatogram of DG12



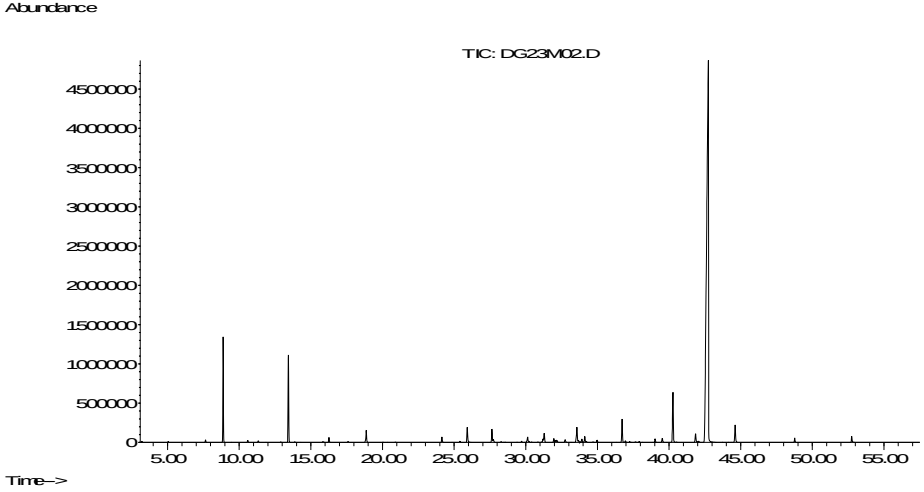
TIC chromatogram of DG13



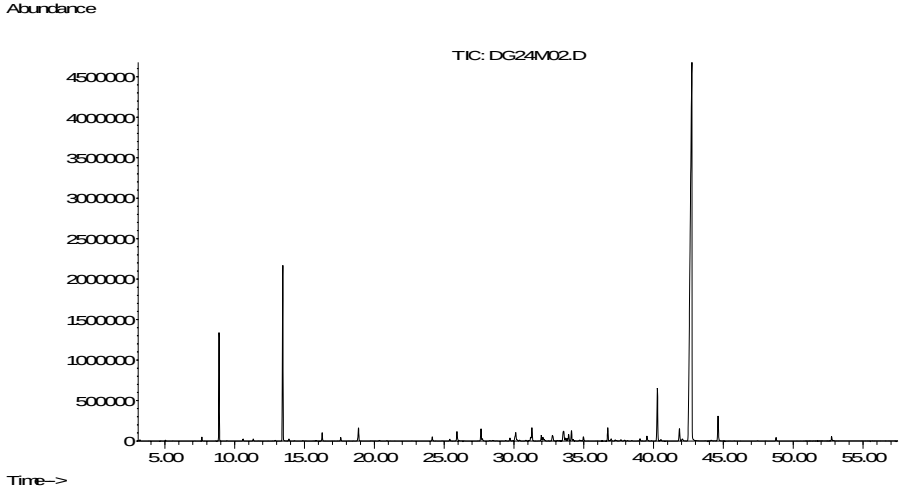
TIC chromatogram of DG14



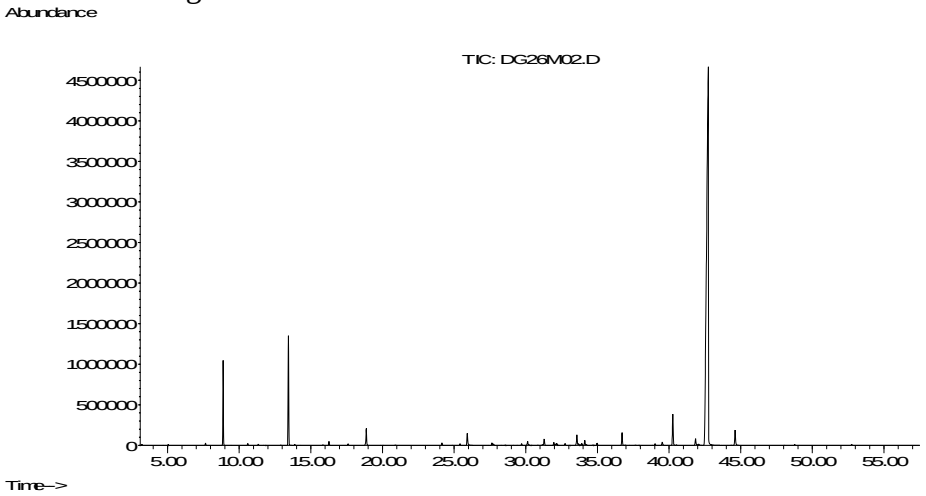
TIC chromatogram of DG15



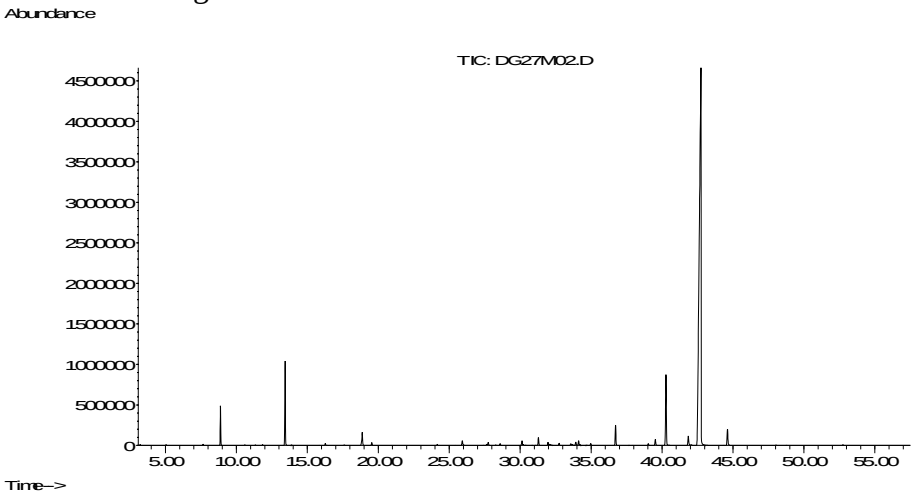
TIC chromatogram of DG16



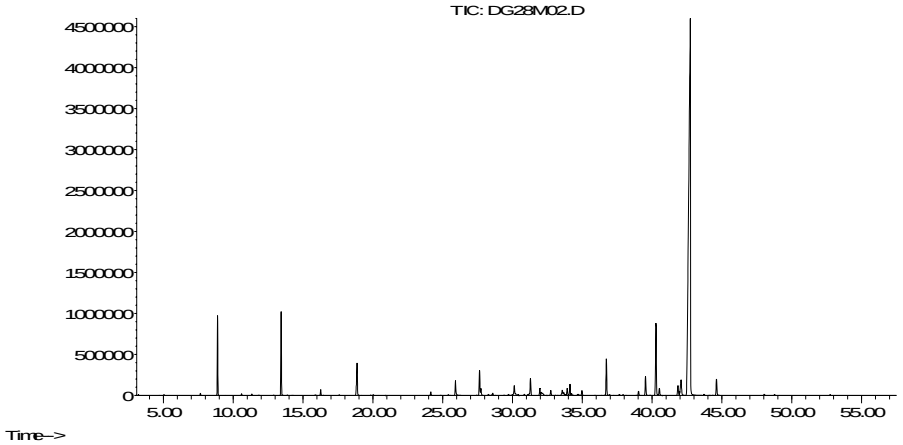
TIC chromatogram of DG17



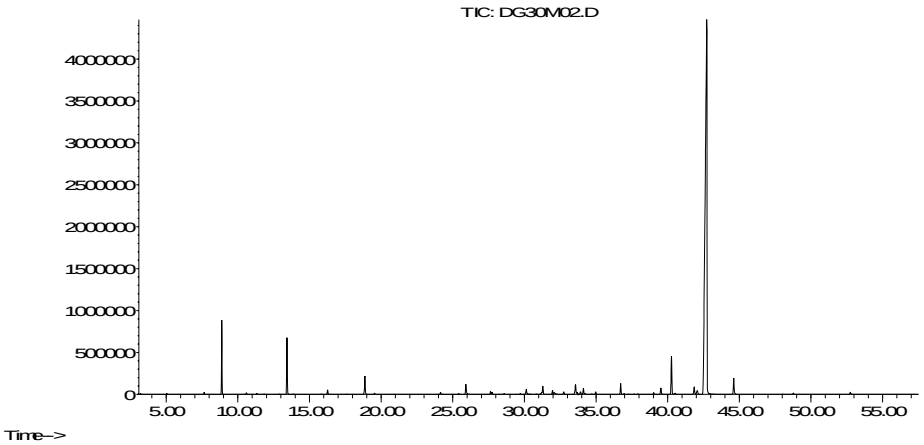
TIC chromatogram of DG18



TIC chromatogram of DG19

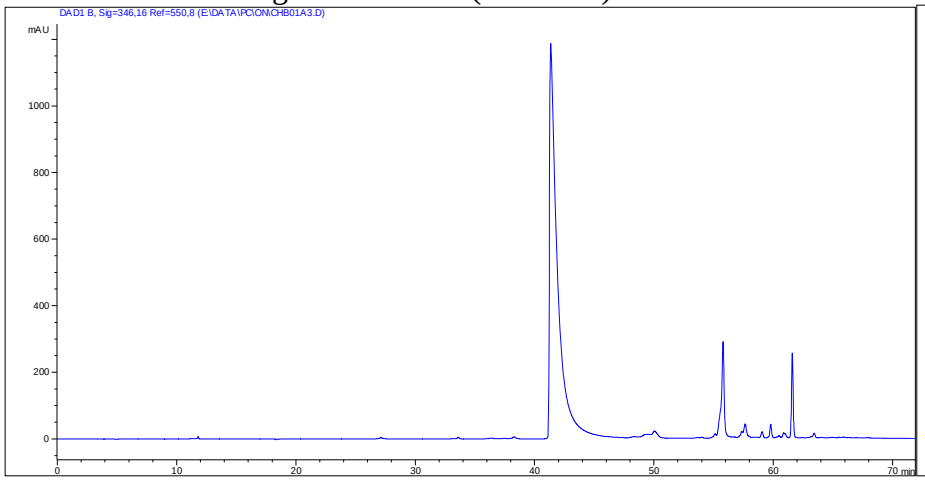


TIC chromatogram of DG20

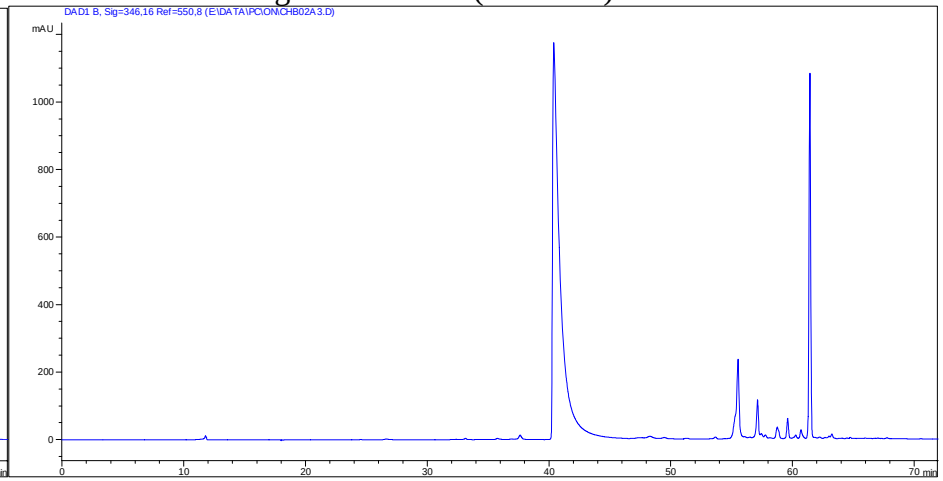


Appendix 6.1 Chromatograms of thirty *Cortex Phellodendri* samples obtained from HPLC-DAD study

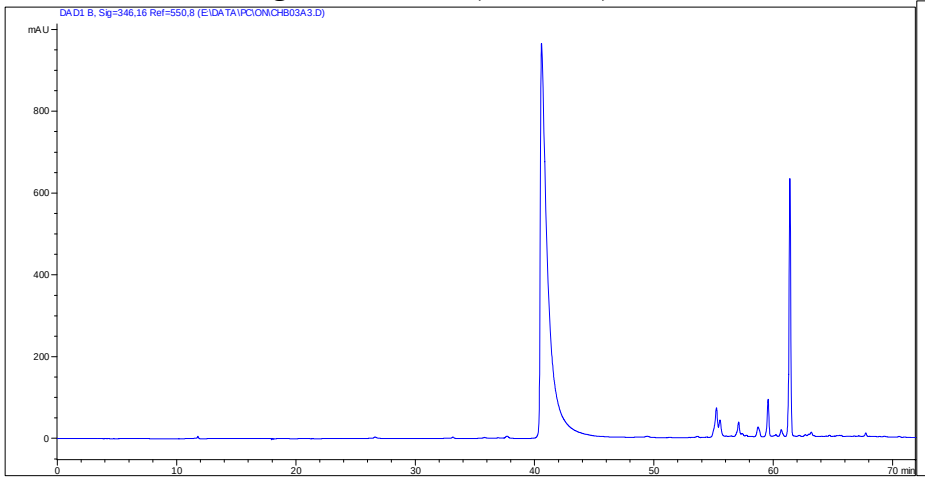
HPLC-UV chromatogram of CHB1 (at 346nm)



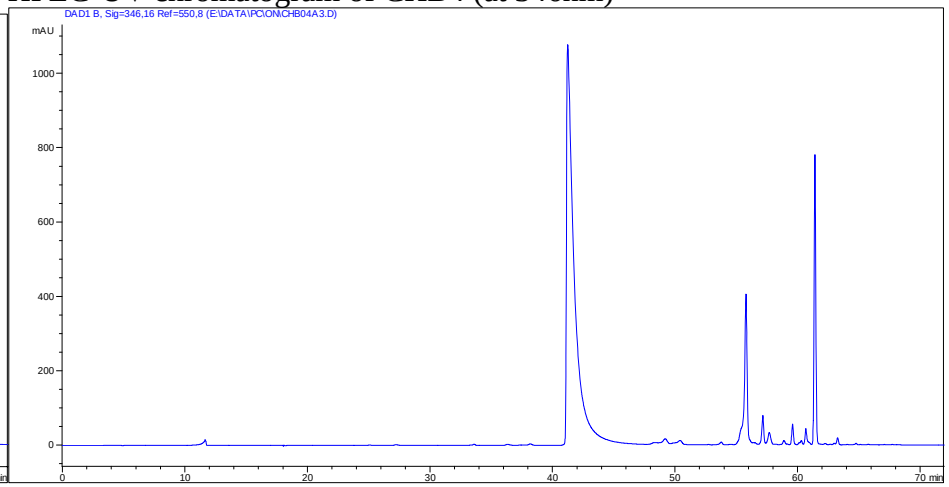
HPLC-UV chromatogram of CHB2 (at 346nm)



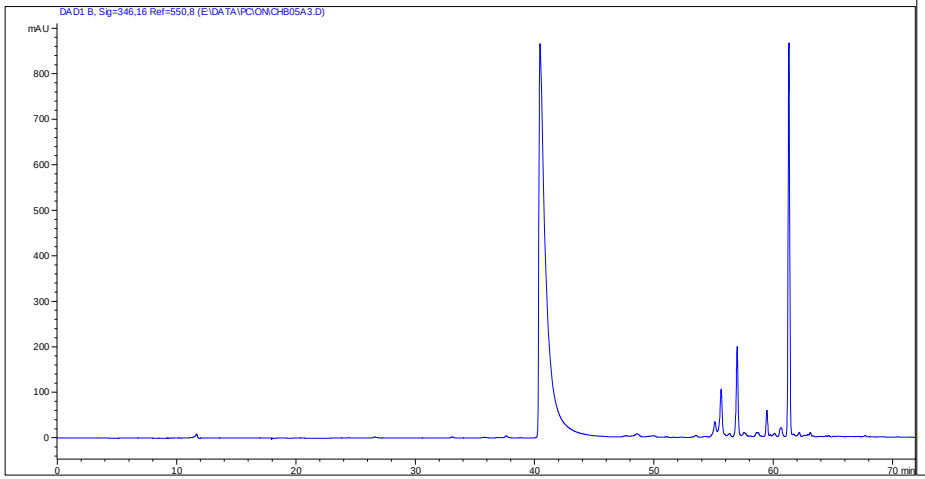
HPLC-UV chromatogram of CHB3 (at 346nm)



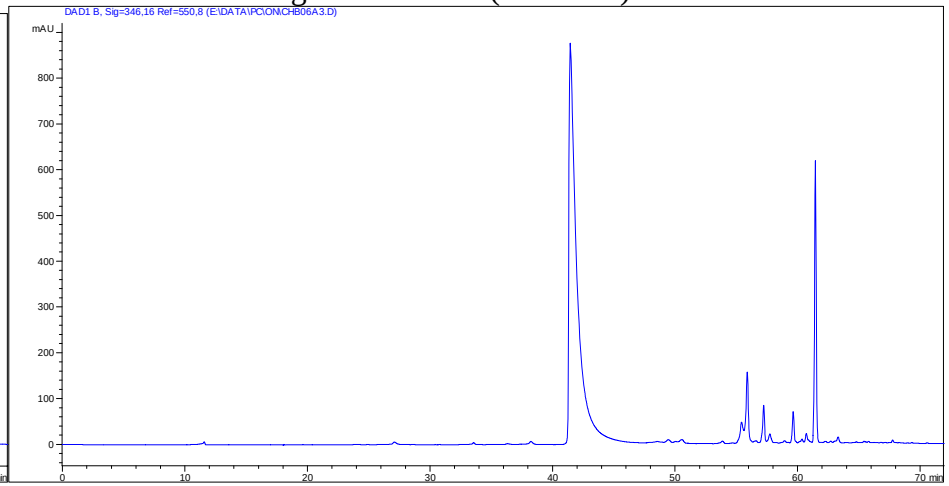
HPLC-UV chromatogram of CHB4 (at 346nm)



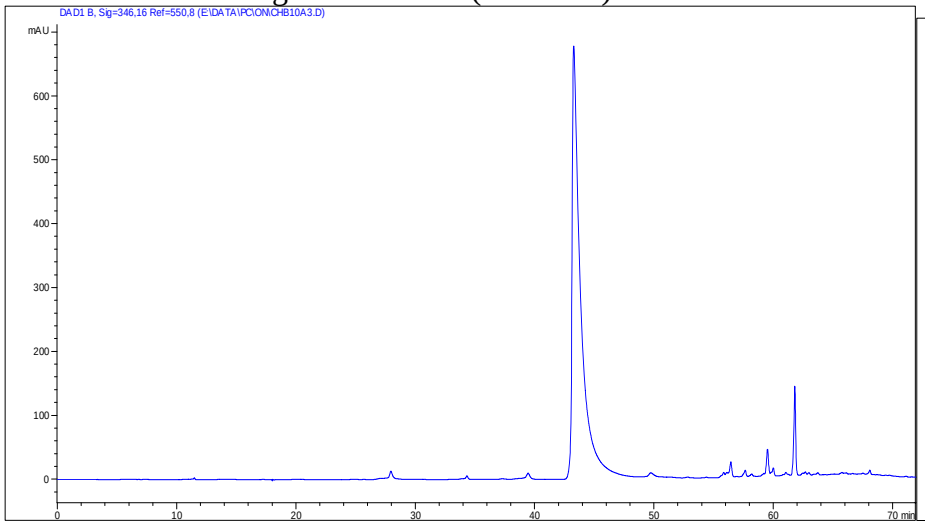
HPLC-UV chromatogram of CHB5 (at 346nm)



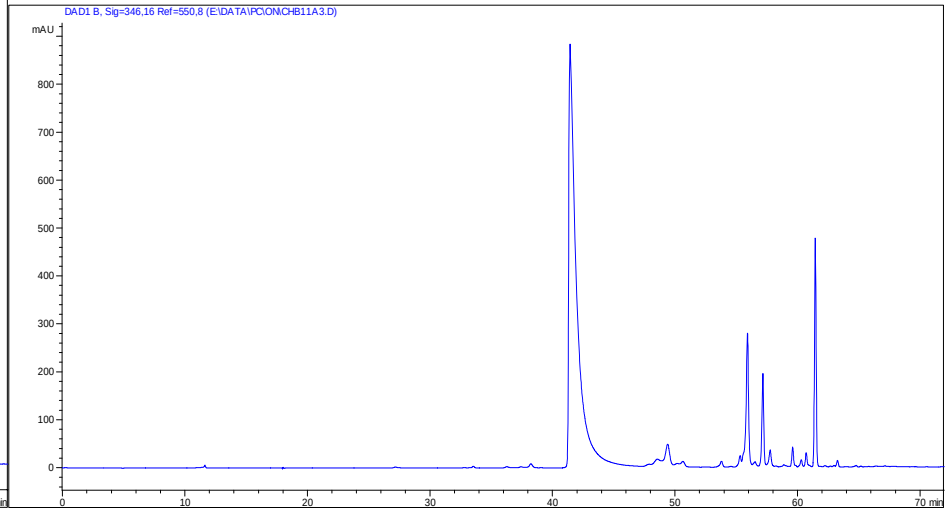
HPLC-UV chromatogram of CHB6 (at 346nm)



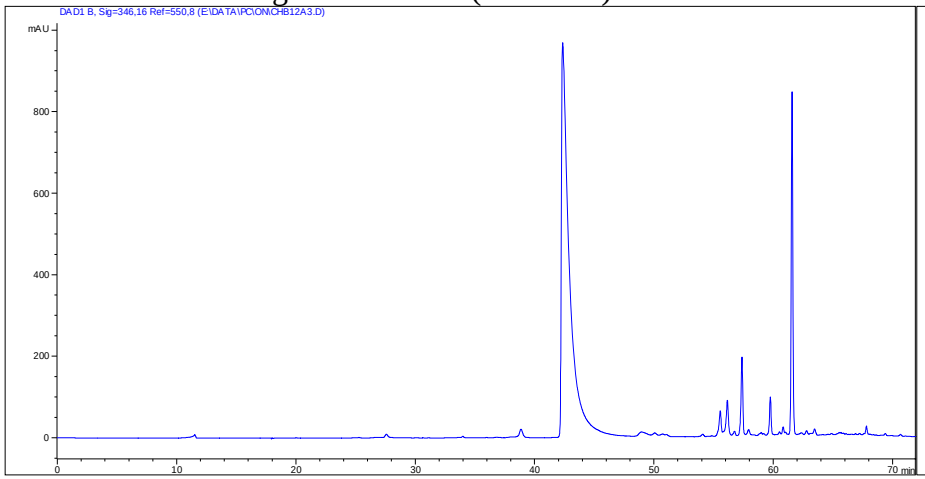
HPLC-UV chromatogram of CHB7 (at 346nm)



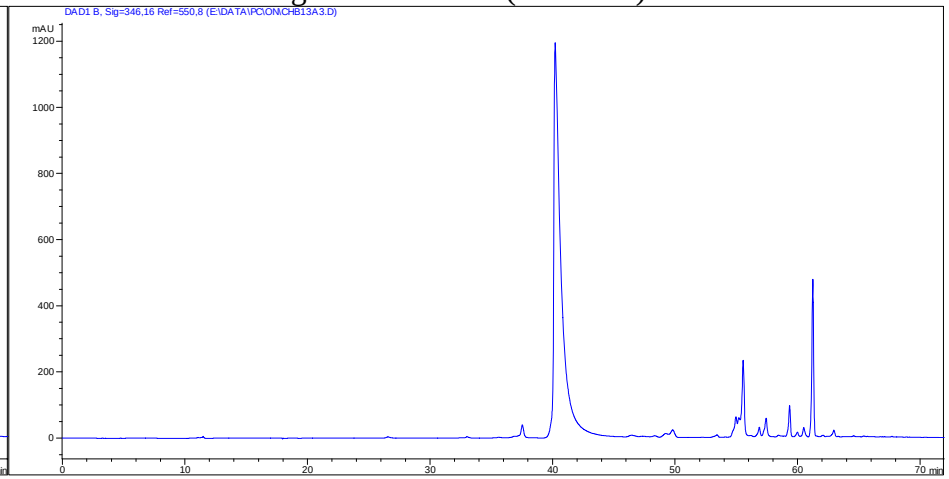
HPLC-UV chromatogram of CHB8 (at 346nm)



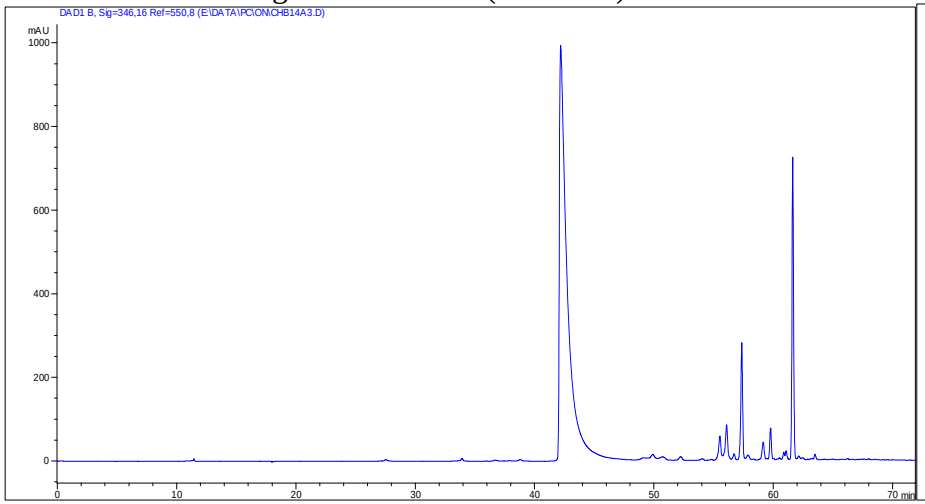
HPLC-UV chromatogram of CHB9 (at 346nm)



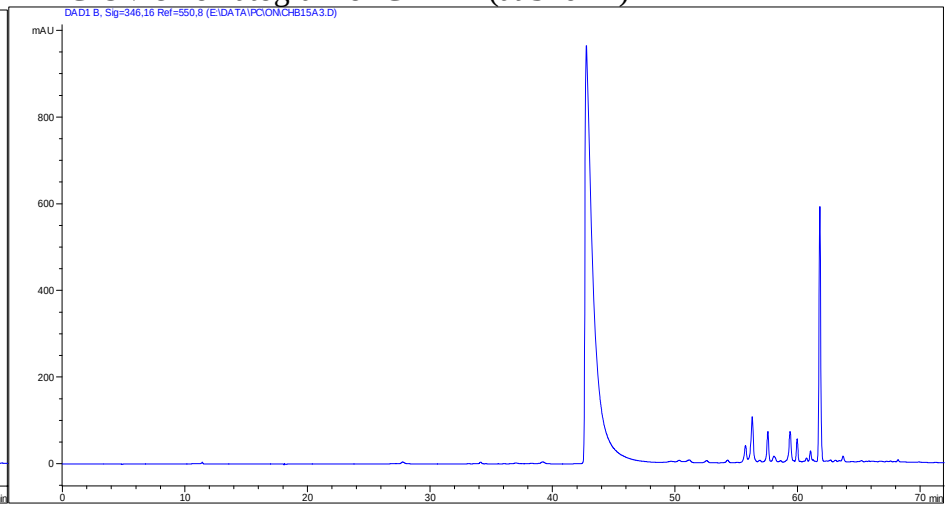
HPLC-UV chromatogram of CHB10 (at 346nm)



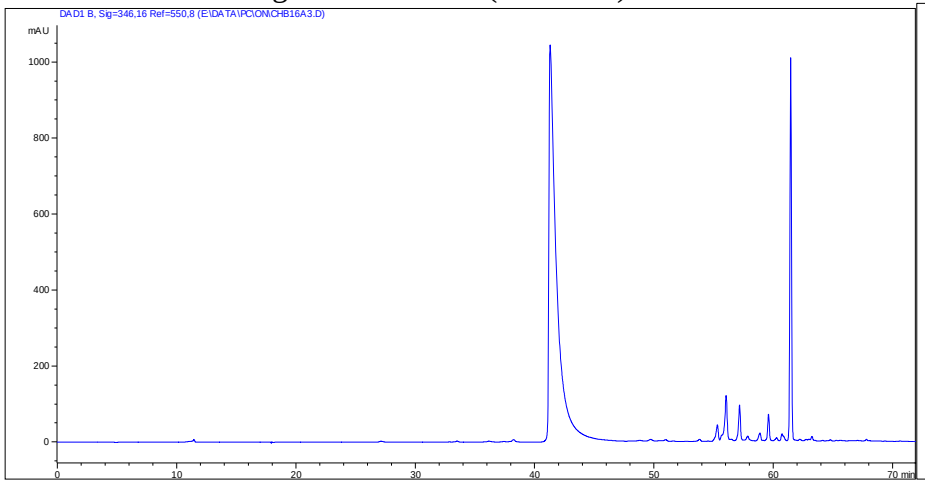
HPLC-UV chromatogram of CHB11 (at 346nm)



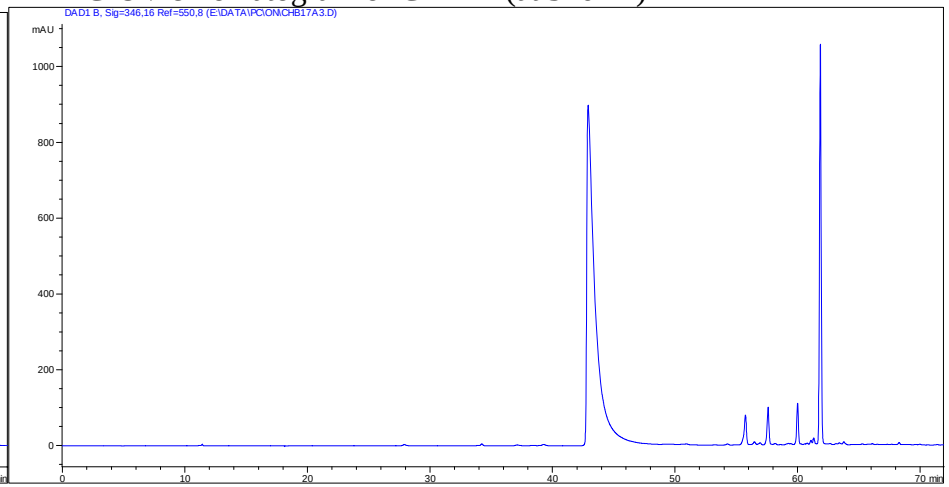
HPLC-UV chromatogram of CHB12 (at 346nm)



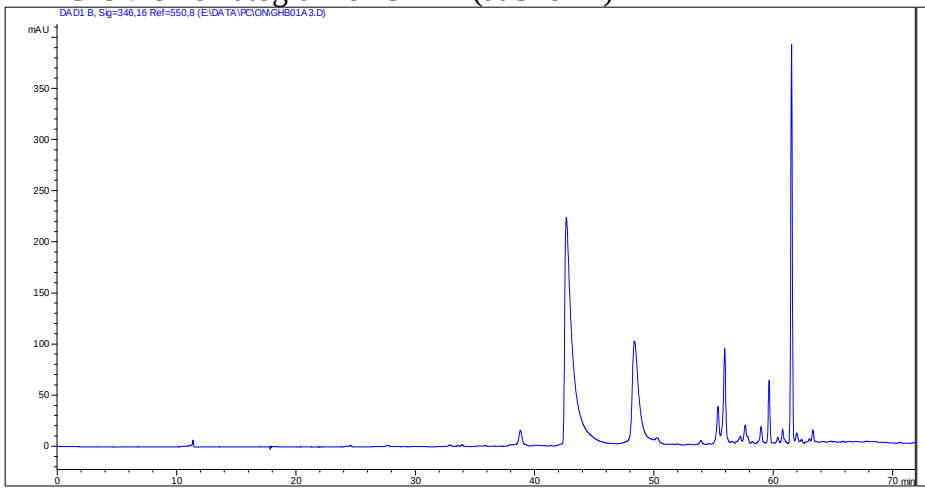
HPLC-UV chromatogram of CHB13 (at 346nm)



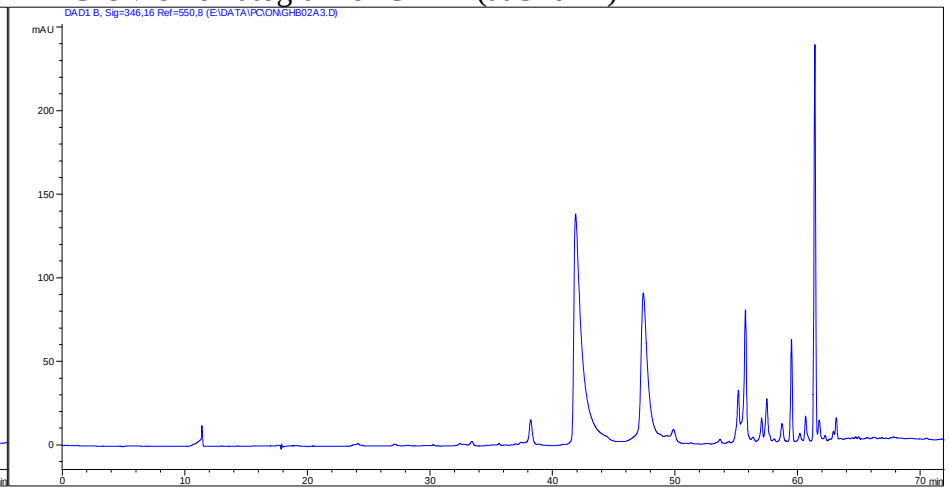
HPLC-UV chromatogram of CHB14 (at 346nm)



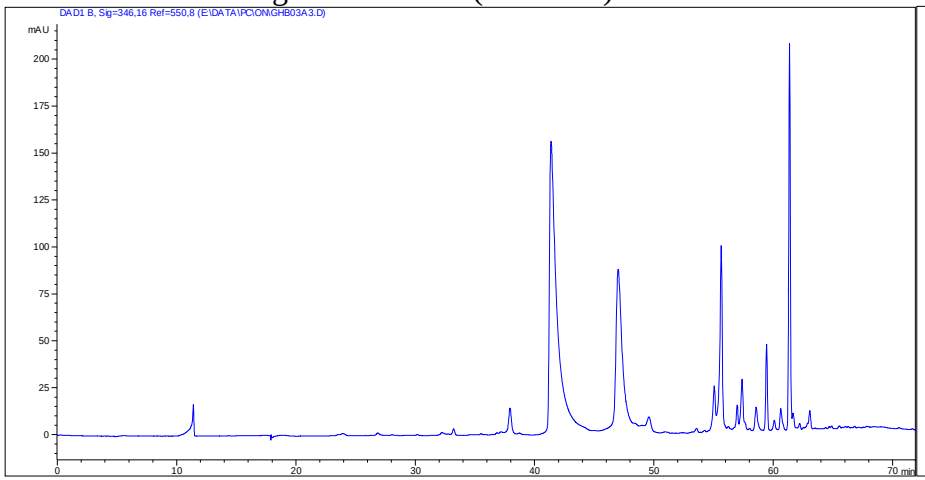
HPLC-UV chromatogram of GHB1 (at 346nm)



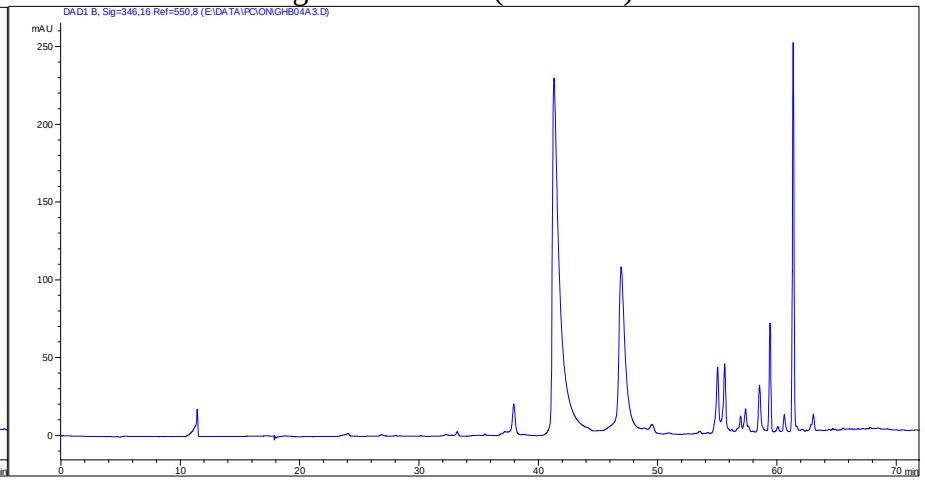
HPLC-UV chromatogram of GHB2 (at 346nm)



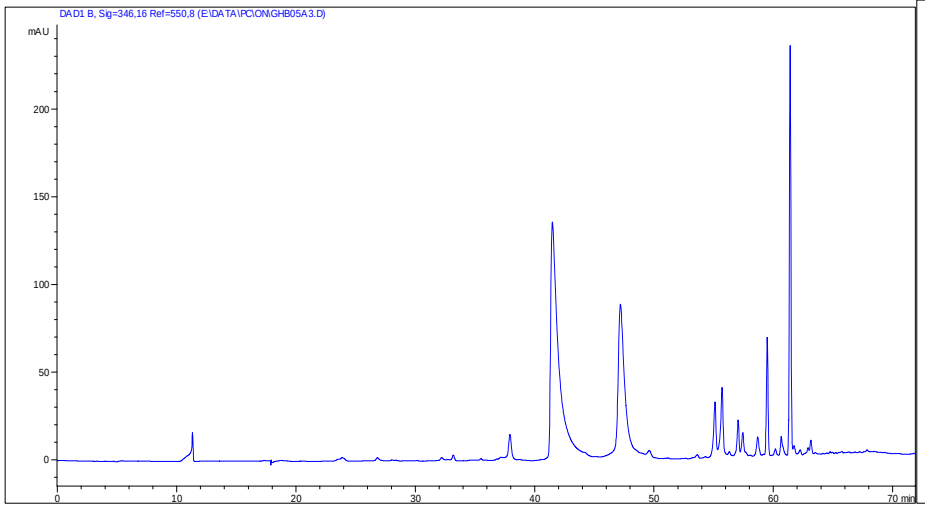
HPLC-UV chromatogram of GHB3 (at 346nm)



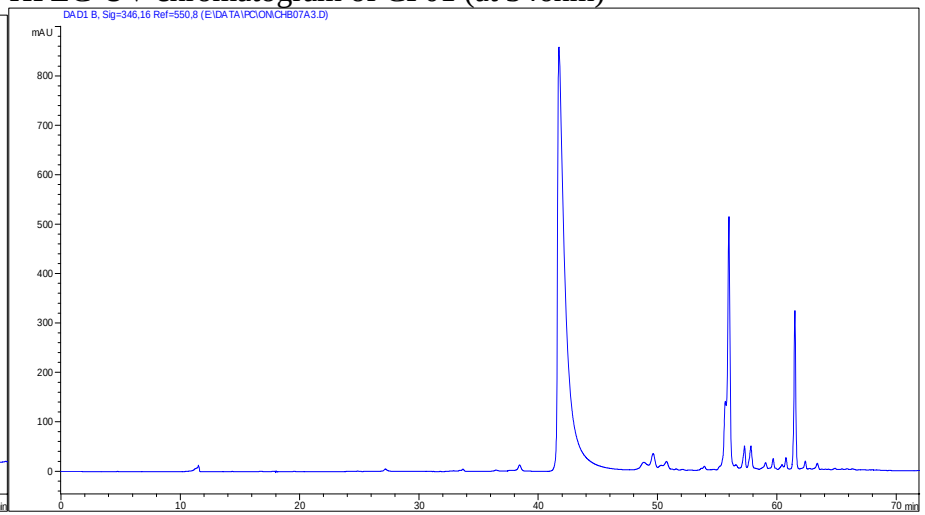
HPLC-UV chromatogram of GHB4 (at 346nm)



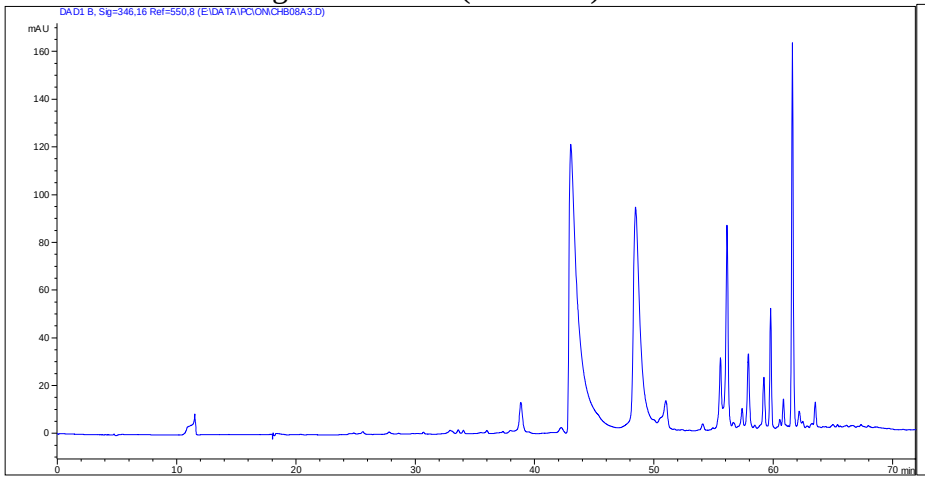
HPLC-UV chromatogram of GHB5 (at 346nm)



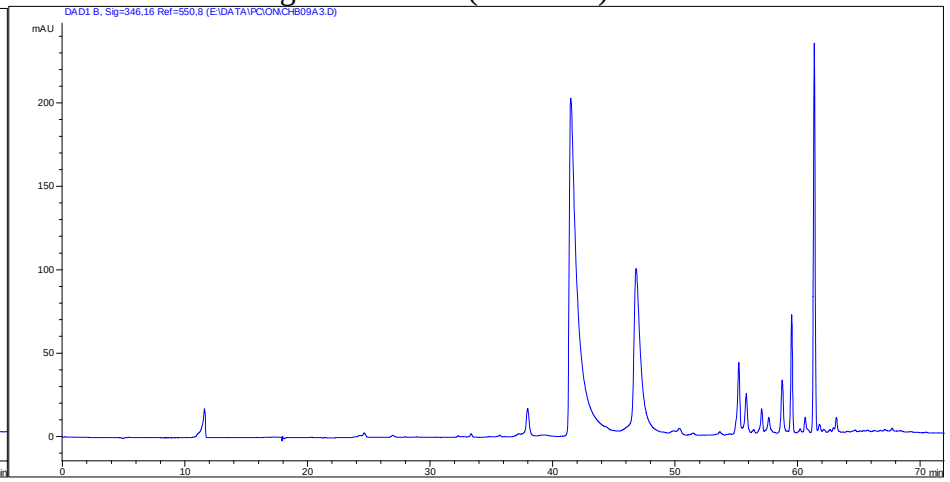
HPLC-UV chromatogram of CP01 (at 346nm)



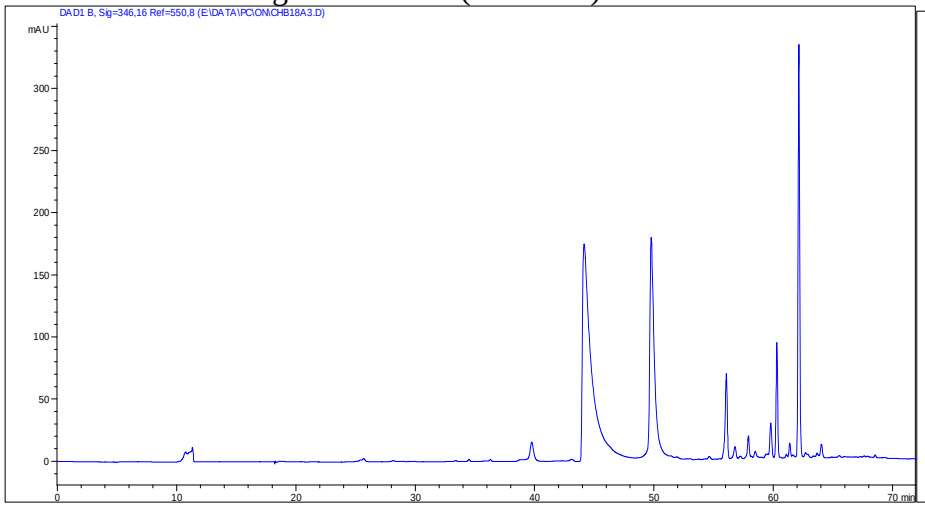
HPLC-UV chromatogram of CP02 (at 346nm)



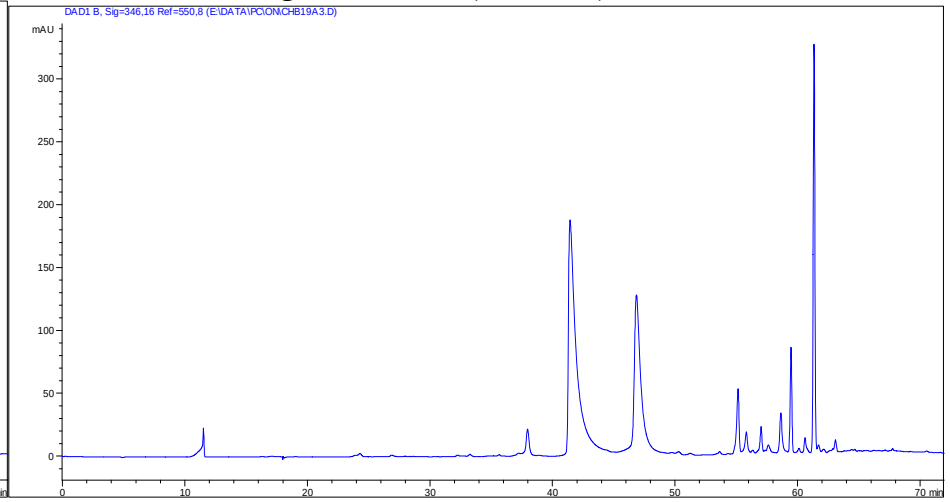
HPLC-UV chromatogram of CP03 (at 346nm)



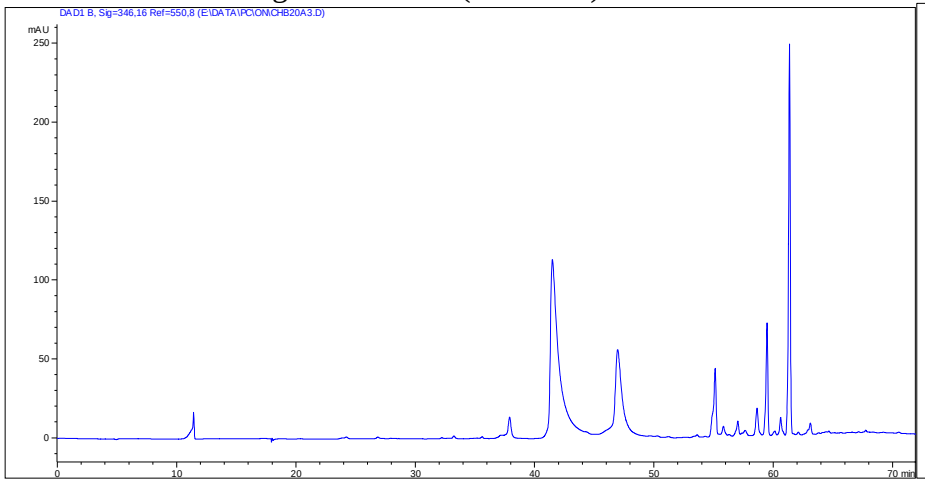
HPLC-UV chromatogram of CP04 (at 346nm)



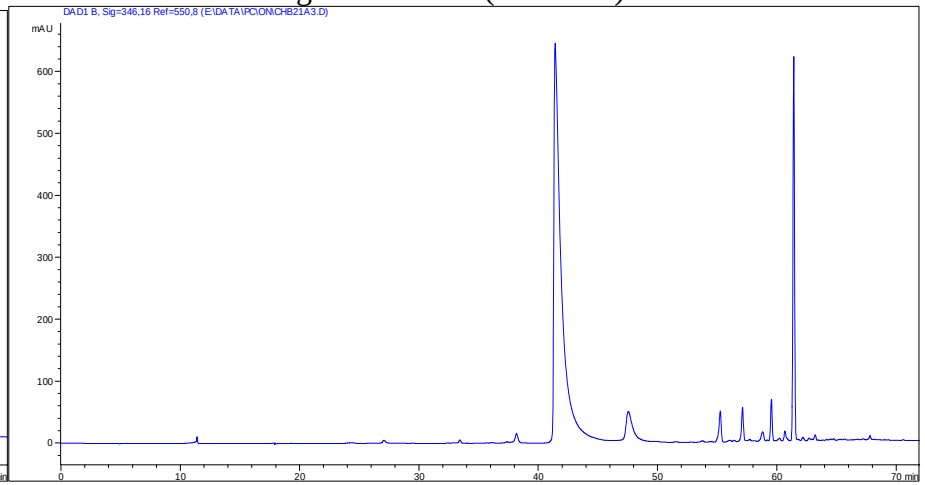
HPLC-UV chromatogram of CP05 (at 346nm)



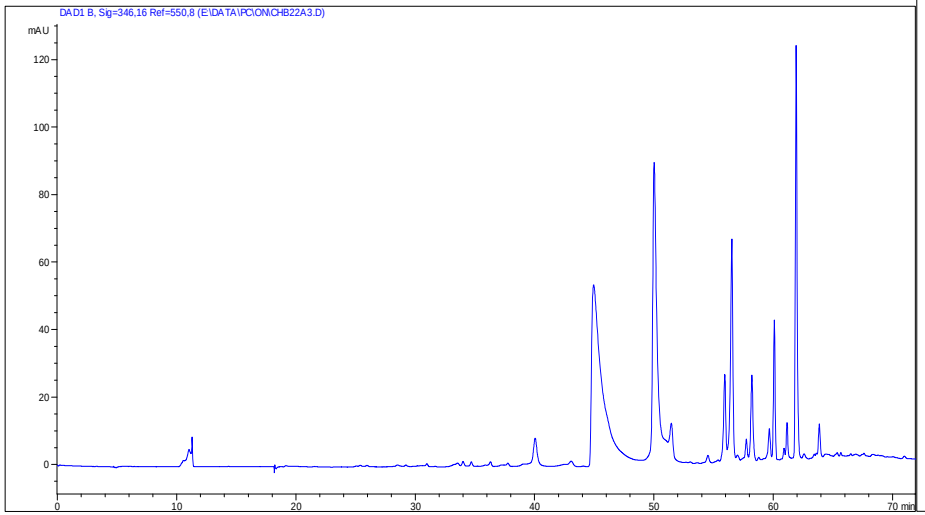
HPLC-UV chromatogram of CP06 (at 346nm)



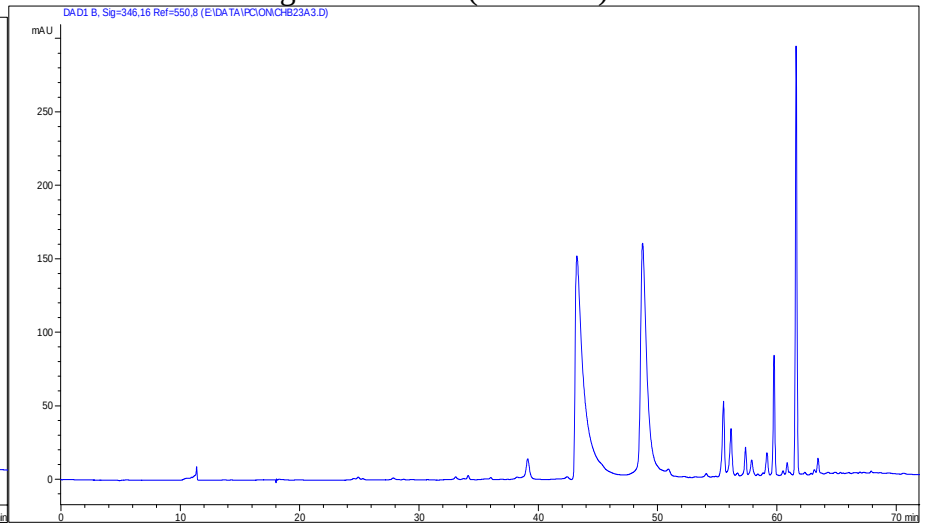
HPLC-UV chromatogram of CP07 (at 346nm)



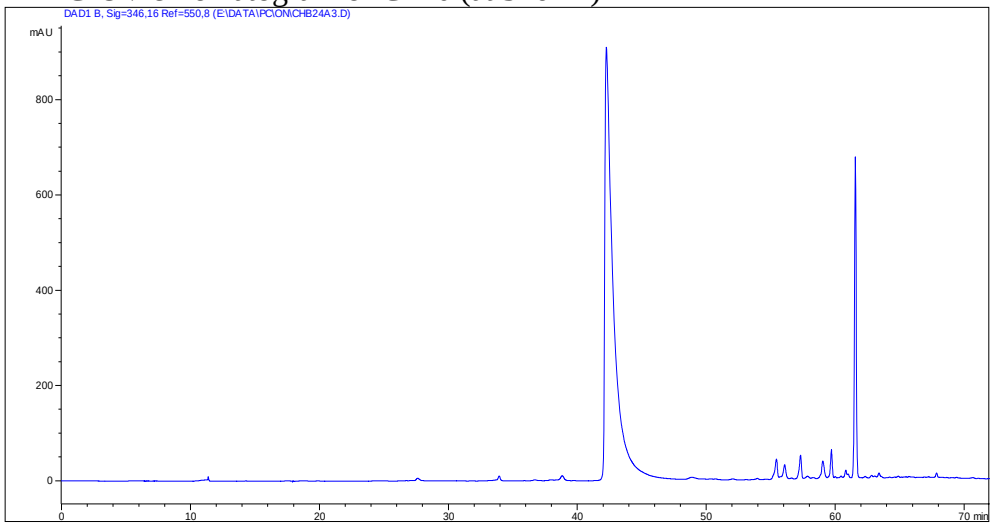
HPLC-UV chromatogram of CP08 (at 346nm)



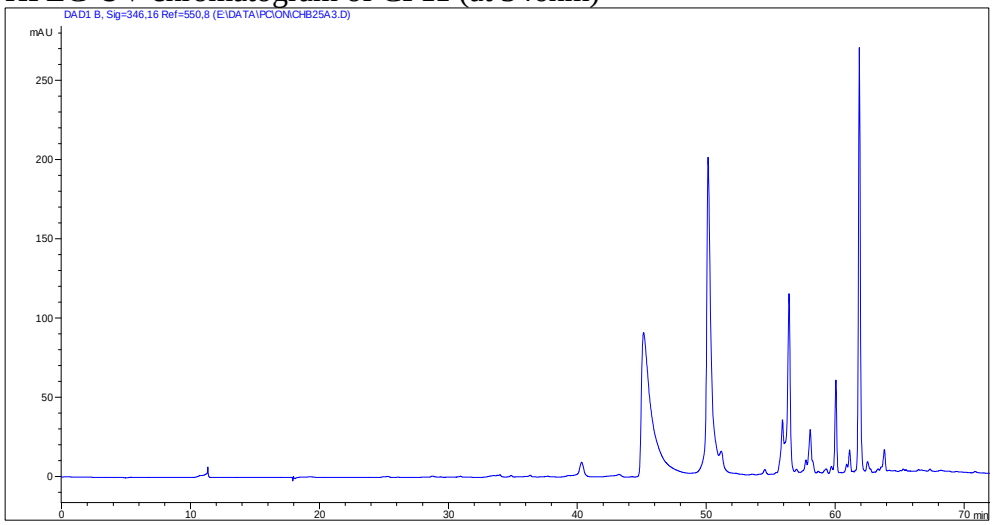
HPLC-UV chromatogram of CP09 (at 346nm)



HPLC-UV chromatogram of CP10 (at 346nm)



HPLC-UV chromatogram of CP11 (at 346nm)



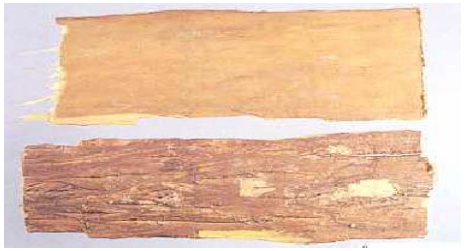
Appendices 8.1: Pictures of *Herba menthae*, *Radix Ligustici chuanxiong*, *Radix Angelicae sinensis*, *Cortex Cinnamomi*, *Phellodendron chinense Schneid*, *Phellodendron amurense Rupr*, *Pogostemon cablin (Blanco) Benth*.

Herba menthae

Radix Ligustici chuanxiong

Radix Angelicae sinensis

Cortex Cinnamomi



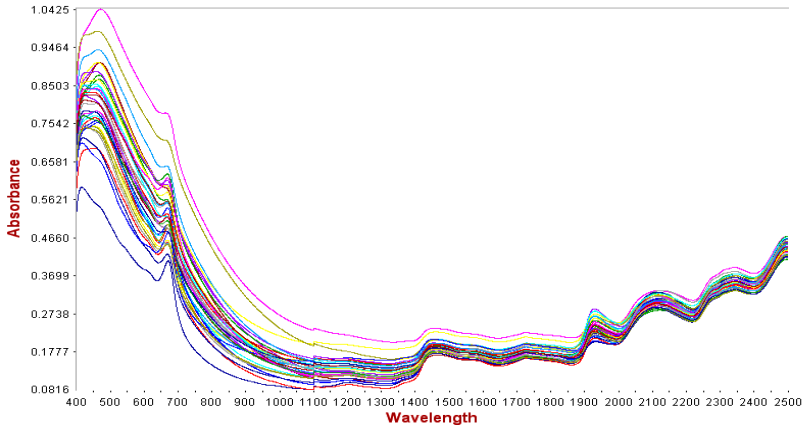
Phellodendron amurense Rupr

Phellodendron chinense Schneid

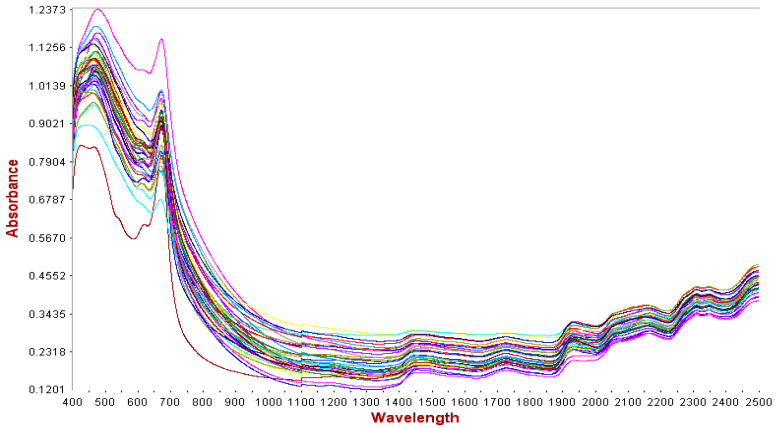
Pogostemon cablin (Blanco) Benth.

Appendix 8.2 Near Infrared Spectra of *Herba menthae*, *Radix Ligustici chuanxiong*, *Radix Angelicae sinensis*, *Cortex Cinnamomi*, *Pogostemon cablin (Blanco) Benth.*, *Phellodendron chinense Schneid*, *Phellodendron amurense Rupr* obtained from NIRRS study

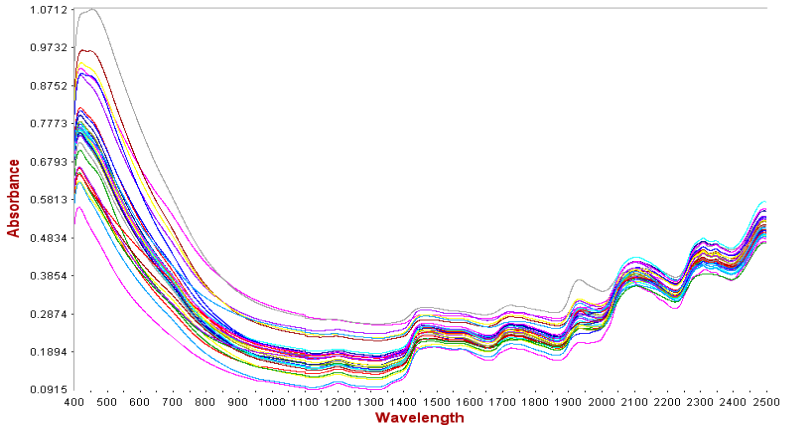
36 NIR spectra of *Herba menthae* (Stem)



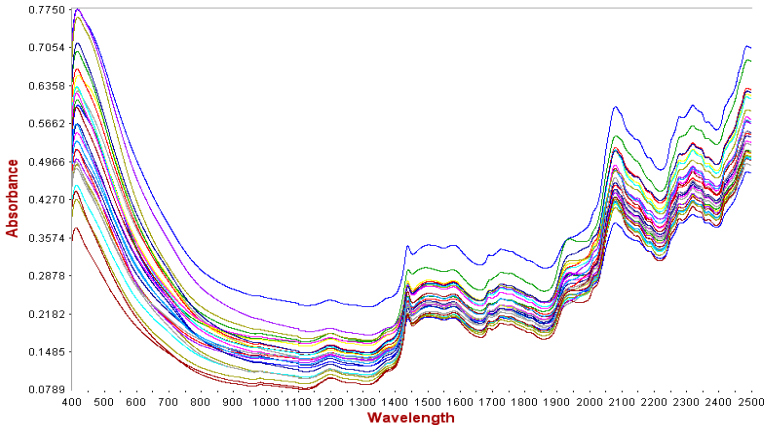
36 NIR spectra of *Herba menthae* (Leave)



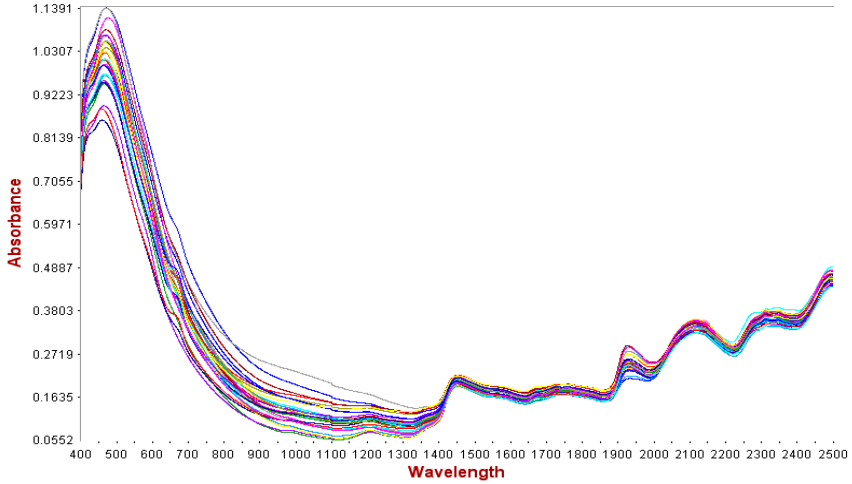
31 NIR spectra of *Radix Ligustici chuanxiong*



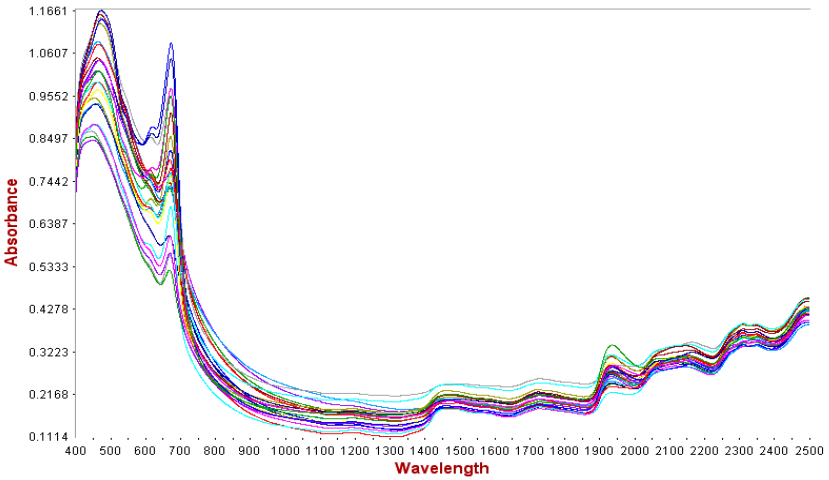
29 NIR spectra of *Radix Angelicae sinensis*



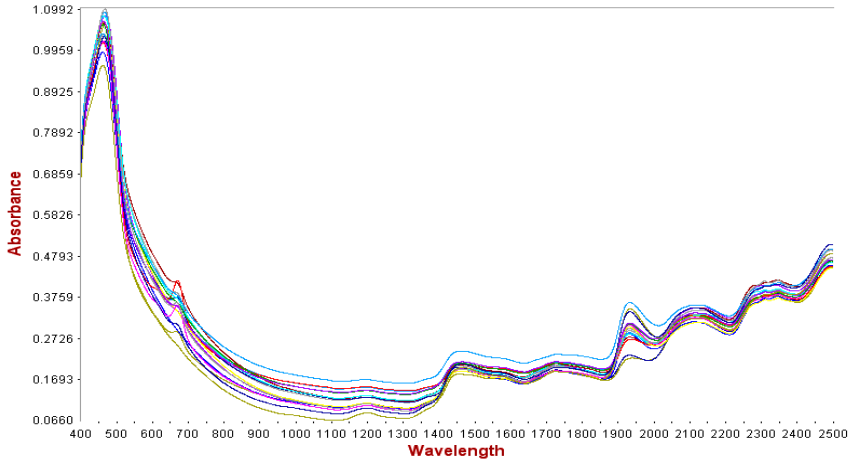
28 NIR spectra of *Cortex Cinnamomi*



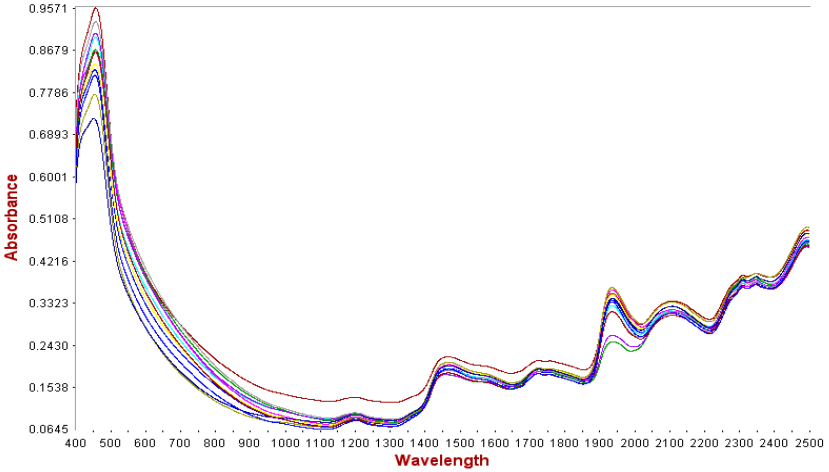
24 NIR spectra of *Pogostemon cablin (Blanco) Benth.*



17 NIR spectra of *Phellodendron chinense Schneid*



13 NIR spectra of *Phellodendron amurense Rupr*



Publications and Presentations

List of Publications and Conference / Symposium Presentations

Journal of Articles

1. C.O. Chan, F.T. Chau, Daniel W.M. Mok, 'A comparison of marker approach, multi-component approach and pattern approach: Development of the chromatographic fingerprint of *Radix Ligustici chuanxiong* by gas chromatography mass spectrometry', manuscript submitted for publication

2. C.O. Chan, F.T. Chau, Daniel W.M. Mok, 'A comparison of marker approach, multi-component approach and pattern approach for classifying *Radix Ligustici chuanxiong* & *Radix Angelicae sinensis* by gas chromatography mass spectrometry', manuscript under preparation

3. C.O. Chan, F.T. Chau, Daniel W.M. Mok, 'Development of the chromatographic fingerprint of *Radix Angelicae sinensis* by gas chromatography mass spectrometry', manuscript under preparation

4. C.O. Chan, F.T. Chau, Daniel W.M. Mok, 'Development of the chromatographic fingerprint of *Pogostemon cablin (Blanco) Benth* by gas chromatography mass spectrometry', manuscript under preparation

5. C.O. Chan, C.C. Chu, F.T. Chau, Daniel W.M. Mok, 'Analysis of berberine and total alkaloids content in *Cortex Phellodendri* by near infrared spectroscopy (NIRS) compared to high-performance liquid chromatography coupled to spectrometric detection', manuscript submitted for publication

6. C.O. Chan, C.C. Chu, F.T. Chau, Daniel W.M. Mok, 'Application of chromatographic fingerprint to the quality control of *Cortex Phellodendri*', manuscript submitted for publication

7. C.O. Chan, Y.Y. Fung, F.T. Chau, Daniel W.M. Mok, C. Hui, 'Analysis of patchouli alcohol content in *Pogostemon cablin (Blanco) Benth.* by near infrared spectroscopy (NIRS) compared to gas chromatography coupled to mass

spectrometric detection' manuscript under preparation

Conference / Symposium Presentations

1. Chi-On Chan, Daniel Kam-Wah Mok, Foo-Tim Chau 'Development of the chromatographic fingerprint of *Rhizoma Chuanxiong* by gas chromatography-mass spectrometry coupled with Chemometrics Resolution Methods (CRM). Poster presented at the Eleventh Symposium on Chemistry Postgraduate Research in Hong Kong, The Hong Kong Polytechnic University, April 17, 2004.

2. C.O. Chan, F.T. Chau, Daniel W.M. Mok, 'Development of the chromatographic fingerprint of *Rhizoma Chuanxiong* by gas chromatography-mass spectrometry coupled with Chemometrics Resolution Methods', Poster presented at the Seventh Asian Conference on Analytical Sciences, The Hong Kong Baptist University, Hong Kong, July 28, 2004.

3. C.O. Chan, F.T. Chau, Daniel W.M. Mok, 'Development of the chromatographic fingerprint of *Rhizoma Chuanxiong* by gas chromatography-mass spectrometry coupled with Chemometrics Resolution Methods', Poster presented at the 25th International Symposium on Chromatography, Paris, France, October 4-8, 2004

4. C.O. Chan, F.T. Chau, Daniel W.M. Mok, 'Development of the chromatographic fingerprint of *Rhizoma Chuanxiong* by gas chromatography-mass spectrometry coupled with Chemometrics Resolution Methods', Poster presented at the International Conference on Chemometrics and Bioinformatics in Asia (CCBA) Shanghai, P. R. China, October 16, 2004

5. C.O. Chan, F.T. Chau, Daniel W.M. Mok, 'Application of Marker Approach and All-information Approach on quality control of Chinese Herbal Medicine (CHM)', Poster presented at the Twelfth Symposium on Chemistry Postgraduate Research in Hong Kong, City University of Hong Kong, April 23, 2005

6. C.O. Chan, K.W. Ma, F.T. Chau, Daniel W.M. Mok, ‘Application of Near Infrared Reflectance Spectroscopy to the quantification of the volatile components of *Herba Menthae* with the help of chemometrical resolution method (CRM)’, Poster presented at the Twelfth Symposium on Chemistry Postgraduate Research in Hong Kong, City University of Hong Kong, April 23, 2005

7. C.O. Chan, F.T. Chau, Daniel W.M. Mok, ‘Development of the chromatographic fingerprint of *Rhizoma Chunaxiong* by gas chromatography mass spectrometry’, Oral presented at 2005 Hong Kong-Macau Postgraduate Symposium on Chinese Medicine, Hong Kong Convention and Exhibition Centre, August 12, 2005

8. C.O. Chan, K.W. Ma, Daniel W.M. Mok, F.T. Chau, C. Hui, Application of Near Infrared Reflectance Spectroscopy to the quantification of the volatile components of *Herba Menthae* with the help of chemometrical resolution method (CRM)’, Poster presented at 2005 Hong Kong-Macau Postgraduate Symposium on Chinese Medicine, Hong Kong Convention and Exhibition Centre, August 12, 2005

9. C.O. Chan, F.T. Chau, Daniel W.M. Mok, ‘Development of the chromatographic fingerprint of *Rhizoma Chunaxiong* by gas chromatography mass spectrometry’, Poster presented at the Eighth Asian Conference on Analytical Sciences, National Taiwan University, Taipei, Taiwan, October 16-20, 2005

10. Y.Y. Fung, C.O Chan, Daniel K.W. Mok, F.T. Chau, C. Hui, ‘Application of Near Infrared Reflectance Spectroscopy on the development of the classification model between two different chemotypes of *Herba Pogostemonis*’, Poster presented at the Thirteenth Symposium on Chemistry Postgraduate Research in Hong Kong, The Hong Kong University of Science and Technology, April 22, 2006

11. C.C. Chu, C.O Chan, F.T. Chau, Daniel K.W. Mok, ‘Differentiation of two types of *Cortex Phellodendri* using Near Infrared Reflectance Spectroscopy’, Poster presented at the Thirteenth Symposium on Chemistry Postgraduate Research in Hong Kong, The Hong Kong University of Science and Technology, April 22, 2006

12. C.O Chan, Daniel K.W. Mok, F.T. Chau, 'Development of the chromatographic fingerprint of *Cortex Phelldendri* by HPLC-DAD', Poster presented at the Thirteenth Symposium on Chemistry Postgraduate Research in Hong Kong, The Hong Kong University of Science and Technology, April 22, 2006

13. C.O. Chan, Daniel W.M. Mok, F.T. Chau, 'Quantitative Analysis of berberine and total alkaloids content in *Cortex Phellodendri* by near infrared spectroscopy (NIRRS) compared to HPLC-DAD detection', Oral presented at 2006 Hong Kong-Macau Postgraduate Symposium on Chinese Medicine, Hong Kong Convention and Exhibition Centre, August 17, 2006

14. C.O. Chan, Daniel W.M. Mok, F.T. Chau, 'A comparison of fingerprint analysis: Development of classification model between *Radix Angelicae Sinensis* and *Radix Ligustici Chuanxiong* by gas chromatography-mass spectrometry', Poster presented at 2006 Hong Kong-Macau Postgraduate Symposium on Chinese Medicine, Hong Kong Convention and Exhibition Centre, August 17, 2006

15. C.O. Chan, Daniel W.M. Mok, F.T. Chau, 'Application of information theory to developing the extraction condition of alkaloids in *Cortex Phellodendri*', Poster presented at 2006 Hong Kong-Macau Postgraduate Symposium on Chinese Medicine, Hong Kong Convention and Exhibition Centre, August 17, 2006

16. C.O Chan, Daniel K.W. Mok, F.T. Chau, 'Analysis of berberine and total alkaloids content in Cortex Phellodendri by near infrared spectroscopy (NIRS) compared to HPLC-DAD detection', Oral presented at the 26th International Symposium on Chromatography, Copenhagen, Denmark, August 21-25 , 2006

Award

1. Best Oral Presenter, 2005 Hong Kong-Macau Postgraduate Symposium on Chinese Medicine, Hong Kong Convention and Exhibition Centre, August 12, 2005

2. Young Fellowship, the Eighth Asian Conference on Analytical Sciences, National Taiwan University, Taipei, Taiwan, October 16-20, 2005
3. Conference Grants, 26th International Symposium on Chromatography, Copenhagen, Denmark, August 21-25 , 2006
4. Young Scientist Presentation Award Winner, 26th International Symposium on Chromatography, Copenhagen, Denmark, August 21-25 , 2006

Course Attendance

1. 'Method validation: key issues and tools', offered by Jérôme Vial & Jean-Michel Menet, A course offered by 25th International Symposium on Chromatography, Paris, France, October 4-8, 2004