



THE HONG KONG
POLYTECHNIC UNIVERSITY

香港理工大學

Pao Yue-kong Library

包玉剛圖書館

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

THE HONG KONG POLYTECHNIC UNIVERSITY

Department of Applied Biology and Chemical Technology

**Application of Near-Infrared Spectroscopy in the
Quality Control and Chemical Analysis of
Chinese Herbal Medicines**

LAU Ching Ching

**A thesis submitted in partial fulfillment of the requirements for the
degree of Doctor of Philosophy**

March, 2012

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.

LAU Ching-ching

Abstract

Abstract of the thesis entitled

“Application of Near-Infrared Spectroscopy in the Quality Control and Chemical Analysis of Chinese Herbal Medicines”

submitted by LAU Ching-ching

for the degree of Doctor of Philosophy in Chemistry

at The Hong Kong Polytechnic University in February, 2012

The use of herbal medicine in Asian countries has a long history, but the concern for quality and efficacy of herbal medicine in the light of modern science has just been started in the last few decades. Quality control of Chinese Herbal Medicine (CHM) aims to ensure their consistency, safety and efficacy. As the active ingredient(s) of the herbs are usually not known, current quality control practices focus on consistency of chemical compositions and identification of the herbs. Microscopic identification and chromatography analysis are the conventional methods used, but these methods are time demanding, and sometimes, the information obtained is not sufficient for identification beyond doubt. Near-infrared Spectroscopy (NIRS) may be an advantage as it is quick, simple and non-destructive. Validity of the NIR based method is established from statistical correlation extracted from the results of many samples. Moreover, NIR can be used for online monitoring during the manufacturing of herbal products to improve the quality control strategy.

The aim of this work is to examine the feasibility of using NIRS in analysing chemical composition of CHM and develop appropriate procedures and data analysis algorithms for quality control of CHM. In this work, NIRS was used to evaluate the quality of three CHM, *Purariae Radix*, *Coptidis Rhizoma* and *Ganoderma*, in terms of differentiation of the species, prediction of markers contents and biological effect. A systematic procedure for building up quantitative and classification models is proposed in the study of establishing NIR quantitation models for *Puerariae Radix* using conventional PLS. This improves the validity of the quantitation model being

built.

A major drawback for the application of NIRS in the analysis of CHM is that usually only those components of content higher than 1% can be detected. However, in many herbs, the content of the marker could be lower than this. In order to extend the applicability of the NIR to different CHM, a better algorithm, Genetic Algorithm – Partial Least Square regression (GA-PLS), is developed by selecting appropriate wavelength regions which correlate more with the parameters to be predicted. The case of *Ganoderma* shows the algorithm is capable to develop model for prediction of markers with low content (< 0.1%). The models developed for *Rhizoma Coptidis* are more robust when the sample size is small.

Generally speaking, people concerns more on the efficacy of CHM, while the current quality control practices focuses on consistency. To examine the possibility of predicting biological activity from the NIR spectra (which is, of course affected by components inside the herbs or herbal products), experiments were conducted to exam the possibility of predicting the anti-oxidant effect of *Ganoderma* methanolic extracts from the NIR spectra of the raw herbs. A good correlation model was established using GA-PLS.

This study clearly demonstrated that NIR spectroscopy with suitable chemometric techniques could be used for identification of the CHM, quantitative analysis of the selected chemical compounds in the CHM as well as evaluation of related biological effects. The GA-PLS modeling algorithm developed widens the scope of the application as more subtle correlations in the data can be extracted. As a conclusion, NIR spectroscopy could be an effective tool for better quality control of CHM.

List of Published and Presented Work from This Thesis

Journal Articles

1. C.C. Lau, C.O. Chan, F.T. Chau, Daniel K.W. Mok. “**Rapid analysis of Radix Puerariae by near-infrared spectroscopy.**” *Journal of Chromatography A*, 2009, 11, 2130-2135.
2. C.O. Chan, Z.D. Zeng, C.C. Lau, Y.S. Fung, D.J. Yang, H.X. Xu, F.T. Chau, Daniel K.W. Mok, “**Rapid analysis of protoberberine alkaloids in Rhizoma Coptidis using near-infrared spectroscopy**”. Manuscript submitted for publication.
3. C.C. Lau, C.O. Chan, F.T. Chau, Daniel K.W. Mok. “**Prediction of the anti-oxidant activity of *Ganoderma lucidum* using Near-infrared Spectroscopy (NIRS)**”. Manuscript under preparation.

Conference/ Symposium Presentation

1. C. C. Lau, Daniel W.M. Mok,, “**Classification and quantitative analysis of Radix Puerariae by Near-infrared Spectroscopy (NIRS) and high performance liquid chromatography (HPLC)**”, Poster presented at the 13th Symposium on Chemistry Postgraduate Research in Hong Kong, April 22, 2006
2. C. C. Lau, F.T. Chau and Daniel W.M. Mok, “**Classification of Radix Puerariae and quantitative analysis of isoflavonoids content by Near-infrared Spectroscopy (NIRS)**”, Poster presented at the 2nd Hong Kong–Macau Postgraduate Symposium on Chinese Medicine, Hong Kong, August 17, 2006
3. C. C. Lau, F.T. Chau and Daniel W.M. Mok, “**Classification of Radix Puerariae and quantitative analysis of isoflavonoids content by Near-infrared Spectroscopy (NIRS)**”, Poster presented at the 26th International Symposium on Chromatography, Copenhagen, Denmark, August 21-25 , 2006
4. C. C. Lau, C. O. Chan, F.T. Chau and Daniel W.M. Mok, “**The use of Near-infrared Spectroscopy (NIRS) in the qualitative and quantitative analysis of Gegen, Radix Puerariae**”, Poster presented at the 14th Symposium on chemistry postgraduate research in Hong Kong, April 28, 2007
5. C. C. Lau, F.T. Chau and Daniel W.M. Mok, “**Identification and quantitative analysis of Ganoderma using Near-infrared Spectroscopy (NIRS) and Liquid Chromatography Coupled with Electrospray Ionization Mass Spectrometry (LC-MS)**”, Poster presented at the 3rd Hong Kong–Macau Postgraduate Symposium on Chinese Medicine, Hong Kong, June 8, 2007
6. C.C. Lau, C.O. Chan, F.T. Chau and Daniel K.W. Mok, “**Rapid analysis of Radix Puerariae by Near-infrared Spectroscopy**”, Poster presentation at the 15th symposium on Chemistry Postgraduate Research in Hong Kong, Hong Kong, 26 April, 2008

7. C.O. Chan, C.C. Lau, F.T. Chau and Daniel K.W. Mok, “**Quality assessment of Radix Puerariae by HPLC-DAD**”, Poster presentation at the 15th symposium on Chemistry Postgraduate Research in Hong Kong, Hong Kong, 26 April, 2008.
8. C.C. Lau, F.T. Chau, and Daniel K.W. Mok, “**Classification of *Ganoderma lucidum* and *Ganoderma sinense* by using Near-infrared Spectroscopy (NIRS) and Liquid Chromatography Coupled with Electrospray Ionization Mass Spectrometry (LC-MS)**”, Poster presentation at the 4th Hong Kong-Macau Postgraduate Symposium on Chinese Medicine, Hong Kong, 15 August, 2008.
9. C.C. Lau, F.T. Chau, and Daniel K.W. Mok, “**Prediction of the anti-oxidant activity of *Ganoderma lucidum* using Near-infrared Spectroscopy (NIRS)**”, Poster presentation at the 16th symposium on Chemistry Postgraduate Research in Hong Kong, Hong Kong, 26 April, 2009.
10. C.O. Chan, Y.S. Fung, C.C. Lau, Daniel K.W. Mok, “**Application of response surface methodology: Determination of alkaloids contents in *Coptis chinensis* Franch**”, Poster presentation at the 16th symposium on Chemistry Postgraduate Research in Hong Kong, Hong Kong, 18 April, 2009.

Courses attended

1. **A one day training course on: TCM fingerprint techniques and chemometrics.** Lectured by Prof. F.T. Chau, Prof. O.M. Kvalheim, Prof. Y.Z. Liang and Prof. P.S. Xie. A course offered in the Hong Kong Polytechnic University, 22 February, 2006.
2. **Multivariate data analysis in chromatography.** Lectured by Prof. Rasmus Bro, in the 26th International Symposium on Chromatography, Denmark, 21 August, 2006.
3. **Workshop on recent development of TCM fingerprint analysis techniques and the New SFDA TCM stability Guidelines.** Lectured by Prof. F.T. Chau, Prof. P.S. Xie, Prof. Y.Z. Liang, Dr. Daniel K.W. Mok and Dr. C.O. Chan. The workshop offered by the Department of Applied Biology and Chemical Technology of the Hong Kong Polytechnic University, 7 June, 2007.
4. **Workshop on recent progress in chemical fingerprint-bioactivity analysis techniques of Chinese Medicine.** Lectured by Prof. F.T. Chau, Prof. O.M. Kvalheim, Prof. Y.Z. Liang and Prof. P.S. Xie. The workshop offered by the Department of Applied Biology and Chemical Technology of the Hong Kong Polytechnic University, 18 October, 2008.
5. **Workshop on recent advancement in linking bioactivities and chemical fingerprints of herbal medicine together for plant drug development.** Lectured by Prof. F.T. Chau and Prof. O.M. Kvalheim. The workshop offered by the Department of Applied Biology and Chemical Technology of the Hong Kong Polytechnic University, 27 October, 2009.

Committee for Symposium

1. Committee member of 3rd Hong Kong-Macau Postgraduate Symposium on Chinese Medicine, Hong Kong, 16 August, 2007.
2. Committee member of 4th Hong Kong-Macau Postgraduate Symposium on Chinese Medicine, Hong Kong, 15 August, 2008

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 support of this project including the financial support for me to attend the course “Method validation: key issues and tools” in the 25th International Symposium on Chromatography (2004) in France 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 my study.

My deepest-felt appreciation goes to my chief supervisor, Dr. Daniel Kam-Wah Mok, who spent his precious time and inspiration with me, shared his experience and provided me valuable suggestion not only on my study, but also in many other areas like personal development, social life etc. Dr Mok also gave me unlimited support and concern during the hard time in my research study, especially when I came across severe health problem.

I am greatly indebted to Prof. Foo-Tim Chau, my co-supervisor, for providing the resources and instruments for this project. Special thanks go to Dr. Chi-on Chan, for his guidance and consultancy in many challenging areas, and also Dr. Zhong-Da ZENG, for his great assistance on chemometrics techniques.

Finally, I am indebted my friend Dr. Wan-chi WONG Vincy for her constant encouragement and assistance during the course of the study especially in my hard times.

Table of Contents

Certificate of Originality	I
Abstract	II
List of Published and Presented Work from This Thesis.....	IV
Acknowledgements	IX
Table of Contents	X
List of Figures.....	XV
List of Tables	XIX
List of Abbreviations	XXII
Chapter 1: Introduction	1
1.1 Introduction	2
Chapter 2: Recent Developments on Quality Control of Chinese Herbal	
Medicines	5
2.1 Introduction	6
2.2 Sample preparation methods	9
2.2.1 Ultrasound-Assisted Extraction (UAE).....	9
2.2.2 Microwave-Assisted Extraction (MAE).....	10
2.2.3 Pressurized Liquid Extraction (PLE).....	10
2.2.4 Solid-Phase Extraction (SPE).....	10
2.2.5 Supercritical Fluid Extraction (SFE)	11
2.3 Analytical techniques	12
2.3.1 Thin Layer Chromatography (TLC).....	12
2.3.2 Gas Chromatography (GC).....	13
2.3.3 High Performance Liquid Chromatography (HPLC).....	14
2.3.4 Capillary Electrophoresis (CE).....	17
2.3.5 Spectroscopic methods	18
2.4 Fingerprint analysis for quality control of CHM	20
2.4.1 Multi-components approach.....	21

2.4.2 Pattern approach	22
2.4.3 Multi-pattern approach	23
2.4.4 Chemometrics techniques for fingerprint analysis	24
2.5 Summary	26
Chapter 3: Principles and Applications of Near-Infrared Spectroscopy in	
Quality Control	27
3.1 Introduction	28
3.2 Theory and features of NIR spectra	29
3.3 Working principle of Near-Infrared Reflectance Spectroscopy (NIRS)	33
3.4 Pre-processing methods for NIRS analysis	36
3.4.1 Mean Centering (MC)	37
3.4.2 Derivation	38
3.4.3 Multiplicative Scatter Correction (MSC)	39
3.4.4 Savitzky-Golay (SG)	40
3.4.5 Standard Normal Variate (SNV)	41
3.4.6 De-Trending (DT)	42
3.5 Chemometric tools for classification	43
3.5.1 Principal Component Analysis (PCA)	43
3.5.2 Correlation Coefficient	45
3.5.3 Discriminant Analysis (DA)	45
3.5.4 K-Nearest Neighbors (KNN)	46
3.5.5 Soft Independent Modeling of Class Analogy (SIMCA)	46
3.5.6 Artificial Neural Networks (ANN)	47
3.6 Chemometric tools for developing quantitative models	47
3.6.1 Multi-Linear Regression (MLR)	48
3.6.2 Principal Component Regression (PCR)	49
3.6.3 Partial Least Squares Regression (PLS)	50
3.6.4 Artificial Neural Networks (ANN)	50
3.7 Applications of NIRS	51
3.7.1 Moisture determination	51
3.7.2 Content determination	52
3.7.3 Quality control of CHM	52
3.8 Summary	53
Chapter 4: Classification and Quantitation Analysis of Puerariae Radix	55

4.1 Introduction	56
4.2 Background of Puerariae Radix	57
4.3 Chemical and biological studies of Puerariae Radix.....	58
4.3.1 Chemical studies.....	58
4.3.2 Biological studies of isoflavanoids of Puerariae Radix.....	60
4.4 Experimental methods and materials	62
4.4.1 Samples and chemicals.....	62
4.4.2 High Performance Liquid Chromatography (HPLC) analysis	63
4.4.3 Near-Infrared Spectroscopic (NIRS) analysis	64
4.5 Results and discussion.....	65
4.5.1 Purity determination of the chemical standards by quantitative H-NMR	65
4.5.2 Identification of Puerariae Radix by HPLC-DAD	66
4.5.2.1 Optimization of extraction and chromatographic conditions.....	67
4.5.2.2 Chromatographic studies of YG and FG by HPLC-DAD	70
4.5.3 Validation of the reference method.....	71
4.5.4 Quantitative analysis of YG and FG.....	73
4.5.5 Classification models development by NIRS.....	81
4.5.6 Quantitation models development by NIRS.....	87
4.5.7 Study of YG of different age	99
4.5.8 Implementation of NIRS method on quality control of CHM	102
4.6 Conclusion.....	105
Chapter 5: Variable Selection in Near-Infrared Spectroscopy: Analysis of	
Protoberberine Alkaloids in Coptidis Rhizoma.....	106
5.1 Introduction	107
5.2 Background of Coptidis Rhizoma	110
5.2.1 Chemical studies.....	112
5.2.2 Biological studies of alkaloids.....	113
5.3 Chemometric techniques used.....	115
5.3.1 Generic Algorithm - Partial Least Square (GA-PLS).....	115
5.3.2 Taguchi design.....	119
5.4 Experimental details	120
5.4.1 Samples and reagents	120
5.4.2 Rapid Resolution Liquid Chromatography (RRLC) analysis	122
5.4.3 Near-Infrared Spectroscopy (NIRS) and data analysis.....	124

5.5 Results and discussion.....	125
5.5.1 Optimization of extraction condition using Taguchi experimental design..	125
5.5.2 Chromatographic study of Coptidis Rhizoma by RRLC analysis.....	130
5.5.3 Method validation.....	132
5.5.4 Quantitation NIR models development by GA-PLS	136
5.5.5 Comparison of the NIR models developed by PLS and GA-PLS.....	143
5.5.6 Validation of quantitation models developed by GA-PLS	147
5.6 Conclusion.....	153
Chapter 6: Modeling the Biological Activity of CHM: Predicting the	
Anti-Oxidant Effect of <i>Ganoderma</i>	154
6.1 Introduction	155
6.2 Background of <i>Ganoderma</i>	156
6.3 Chemical and biological studies of <i>Ganoderma</i>	159
6.3.1 Chemical studies.....	159
6.3.2 Biological studies of <i>Ganoderma</i>	162
6.4 Experimental methods and materials	165
6.4.1 Samples and reagents	165
6.4.2 High Performance Liquid Chromatography coupled with Mass Spectrometry (LC/MS) analysis	166
6.4.3 The Ferric Reducing Antioxidant Power (FRAP) assay.....	168
6.4.4 Near-Infrared Spectroscopic (NIRS) analysis	169
6.5 Results and discussion.....	170
6.5.1 Optimization of extraction condition.....	170
6.5.2 Optimization of chromatographic conditions.....	173
6.5.3 Quantitative analysis of triterpenoids in <i>Ganoderma</i>	176
6.5.4 Identification of triterpenoids in <i>Ganoderma</i> by LC/MS.....	183
6.5.5 Antioxidant activity of <i>Ganoderma</i> samples.....	190
6.5.6 Calibration model for triterpenoids content using NIRS.....	193
6.5.7 Calibration model for anti-oxidant activity	198
6.6 Conclusion.....	200
Chapter 7: Conclusion.....	201
References.....	208
Appendices.....	248

Appendix 4.1 NMR spectrum of a) puerarin, b) daidzin, c) daidzein, d) genistin and
d) genistein249

List of Figures

Fig. 2.1 a) Typical flavonoids fingerprint pattern of Herba E. Sagittati, cf. ; b) the main characteristic ingredients disappeared in this sample.....	7
Fig. 4.1 The chemical structure of isoflavonoids a) puerarin, b) daidzin, c) daidzein, d) genistein, and e) genistin	59
Fig. 4.2 The Chemical structure of maleic acid	66
Fig. 4.3 The Peak area of puerarin and daidzin content in Puerariae Radix samples by using different extracting solvents with 1 hour sonication.....	68
Fig. 4.4 The Peak area of puerarin and daidzin content in Puerariae Radix samples using different extraction methods.....	68
Fig. 4.5 The Peak area of puerarin and daidzin content in Puerariae Radix samples using sonciation method with different duration.....	69
Fig. 4.6 HPLC/UV (254nm) chromatograms of the five studied iso-flavonoids components a) in standard solution, b) extract of <i>Pueraria thomsonii</i> (FG), c) extract of <i>Pueraria lobata</i> (YG): 1) puerarin; 2) daidzin; 3) genistin; 4) daidzein ; 5) genistein.....	71
Fig. 4.7 Histogram of the contents of the five isoflavonoids in Puerariae Radix obtained from HPLC.....	75
Fig. 4.8 Mean NIR spectra for <i>Pueraria thomsonii</i> (–) and <i>Pueraria lobata</i> (- - -) obtained from, a) raw data, b) with second derivative pre-treatment, c) Enlarged region between 1600nm and 1800nm with second derivative pre-treatment and d) Enlarged region between 2050nm and 2450 nm with second derivative pre-treatment	82

Fig. 4.9 Cooman's plot a) raw data, b) raw data with second derivative pretreatment, c) 1600 – 1800nm with second derivative pretreatment, d) 2050 – 2450nm with second derivative pretreatment and e) 1600 – 1800nm and 2050 – 2450nm with second derivative pretreatment for the square SIMCA distances obtained for FG and YG models	84
Fig. 4.10 A general PRESS against PLS factor graph showing the 'over-fitting' region	92
Fig. 4.11 NIRS predicted values against HPLC measurement for the content of a) puerarin, b) daidzin and c) total isoflavonoids in the Puerariae Radix samples ($n = 89$).....	97
Fig. 4.12 The scatter plots obtained by PCA of the YG samples with different cultivation years by using a) HPLC data and b) NIR data	102
Fig. 5.1 The chemical structure of the major protoberberine alkaloids a) berberine, b) palmatine, c) coptisine, d) jatrorrhizine, and e) epiberberine.....	113
Fig. 5.2 Schematic of a simple genetic algorithm.....	116
Fig. 5.3 The RRLC–DAD chromatogram of four successive extractions of <i>Coptidis Rhizoma</i> extract at 346 nm.....	129
Fig. 5.4 RRLC/UV (346 nm) chromatograms of the five alkaloid components studied in a) standard solution, b) extract of <i>Coptis teeta</i> Wall (YunL), c) extract of <i>Coptis chinensis</i> Franch (WeiL) and d) extract of <i>Coptis deltoidea</i> C.Y. Cheng et Hsiao (YaL): 1) coptisine 2) epiberberine 3) jatrorrhizine 4) berberine 5) palmatine.....	131
Fig. 5.5 The contents of five alkaloids in 47 <i>Coptidis Rhizoma</i> samples	134

- Fig. 5.6 47 NIR spectra of HL obtained from, a) raw data, b) treatment of SNV, second derivative with Savitsky-golay (SG) method, moving window average smoothing and MSC.....137
- Fig. 5.7 Prediction values against RRLC-DAD measured content of a) berberine b) coptisine c) palmatine d) epiberberine e) jatrorrhizine and f) total alkaloids content in all HL samples (n = 47).....142
- Fig. 5.8 Overview of the spectral variables selected by GA a) berberine, b) coptisine, c) palmatine, d) epiberberine e) jatrorrhizine and f) total alkaloids content overlayed with the average NIR spectrum of the *Coptidis Rhizoma*145
- Fig. 5.9 Raw NIR spectra of a) 30 *Phellodendri Cortex* samples, b) mean *Phellodendri Cortex* and *Coptidis Rhizoma*.....148
- Fig. 6.1 Structure of a) ganoderic acid A, b) ganoderic acid C2, c) ganoderic acid F, and d) ganoderic acid H.....160
- Fig. 6.2 The HPLC-DAD chromatogram of *Ganoderma* (C1 sample) extracted with a) chloroform b) methanol.....171
- Fig. 6.3 The HPLC chromatogram of *Ganoderma* (C1 sample) using different columns. a) Agilent Sorbax C18 b) Thermo Hypersil ODS C18 and c) Thermo Hypersil Gold C18.....175
- Fig. 6.4 a) The HPLC-UV typical chromatograms of the standards solution containing four standards mixture of ganoderic acid C2, A, H and F, b) the extract of *G. lucidum*, c) the extract of *G. sinense* (type I) and d) the extract of *G. sinense* (type II)177
- Fig. 6.5 Histogram of the four triterpenoids contents in *Ganoderma* samples obtained using HPLC-DAD.....183

Fig. 6.6 The typical TIC chromatograms of the a) extract of <i>G. lucidum</i> , b) extract of <i>G. sinense</i> (type I) and c) extract of <i>G. sinense</i> (type II), under the optimized chromatographic and mass spectroscopic conditions.	185
Fig. 6.7 Results of antioxidant activities in <i>Ganoderma</i> samples determined by FRAP method	191
Fig. 6.8 A plot of total triterpenoid content against the antioxidant activity (FRAP value) of the 33 <i>Ganoderma</i> samples	192
Fig. 6.9 Typical NIR spectra for a) <i>G. lucidum</i> (—), b) <i>G. sinense</i> (Type I) (- - -) and c) <i>G. sinense</i> (Type II) (.....)	195
Fig. 6.10 Correlation plot of NIRS predicted value against HPLC-DAD measured content of a) ganoderic acid A, b) ganoderic acid C ₂ , c) ganoderic acid F, d) ganoderic acid H and e) total triterpenoids content (sum of four ganoderic acids) in <i>Ganoderma</i> samples (n = 33).....	197
Fig. 6.11 Predicted activity by NIRS against antioxidant activity measured by FRAP assay for <i>Ganoderma</i> samples (n = 33).....	198
Fig 6.12 Overview of the final spectral variables selected by GA for FRAP model.....	200

List of Tables

Table 4.1 Puerariae Radix samples collected of the study.....	63
Table 4.2 HPLC Method validations parameters for the five isoflavonoids.....	73
Table 4.3 Calibration curves for five isoflavonoids in Puerariae Radix	74
Table 4.4 The mean contents of the five isoflavonoids in Puerariae Radix obtained from HPLC.....	76
Table 4.5 Contents of the five isoflavonoids in the 89 Puerariae Radix samples.....	77-80
Table 4.6 Results of the classification of Puerariae Radix using LDA and SIMCA.....	85
Table 4.7 Sensitivity and specificity of different SIMCA models.....	86
Table 4.8 Statistics of the isoflavonoids contents of the samples determined by the HPLC in the calibration, validation and test sets.....	91
Table 4.9 a) Performance of different PLS regression models for puerarin.....	94
Table 4.9 b) Performance of different PLS regression models for daidzin.....	95
Table 4.9 c) Performance of different PLS regression models for total isoflavonoids content.....	96
Table 4.10 Variations in the content of individual and total isoflavonoids in YG.....	100
Table 5.1 The List of the Coptidis Rhizoma samples collected.....	121

Table 5.2 Sample extraction factors and variables for Taguchi's experimental design (L-0 orthogonal array).....	126
Table 5.3 Randomized runs of Taguchi experimental design	126
Table 5.4 Results for experimental responses in the (3 ²) matrix.....	128
Table 5.5 RRLC method validation for the five alkaloids and results of the alkaloids contents determined by RRLC-DAD in 47 <i>Coptidis Rhizoma</i> samples (n=4)	135
Table 5.6 Recoveries of the four alkaloids in <i>Coptidis Rhizoma</i> sample.....	136
Table 5.7 Parameters of the GA.....	140
Table 5.8 Performance of the models developed by PLS and GA-PLS.....	143
Table 5.9 Method selectivity for the <i>Coptidis Rhizoma</i> and <i>Phellodendri Cortex</i> samples.....	149
Table 5.10 Results of the validation of all the NIRS models.....	151
Table 6.1 The List of the <i>Ganoderma</i> samples collected.....	166
Table 6.2 The comparison of the integrated peak area of the ganoderic acid C2, A, H and F in chloroform and methanol extracts.....	170
Table 6.3 The peak area of ganoderic acid A and F of the sample using different solvent volume in different successive extraction cycles.....	172
Table 6.4 Calibration curves for four triterpenoids in <i>Ganoderma</i> where x is the content (ppm) of the standard solution and y is the peak area (mAU).....	178

Table 6.5 The content of the four triterpenoids in the 33 <i>Ganoderma</i> samples.....	179-180
Table 6.6 The average contents of the ganoderic acid markers in <i>Ganoderma</i> samples obtained using HPLC-DAD.....	182
Table 6.7 Identification of triterpenic acids from the chloroform extract of a) <i>G. lucidum</i> , b) <i>G. sinense</i> (Type I) and c) <i>G. sinense</i> (Type II).....	186-189
Table 6.8 The results of the antioxidant activity of 33 <i>Ganoderma</i> samples determined by FRAP assay.....	191
Table 6.9 The parameters of the GA-PLS used.....	193

List of Abbreviations

Abbreviations

AA	Aristolochic Acid
AMD	Automated Multiple Developments
ANN	Artificial Neural Networks
APCI	Atmospheric Pressure Chemical Ionization
CASES	Computer-Aided Similarity Evaluation System
CC	Correlation Coefficient
CE	Capillary Electrophoresis
CFS	Chronic Fatigue Syndrome
CHM	Chinese Herbal Medicine
CIR	Cerebral Ischemia Reperfusion
CMMs	Chinese Materia Medica
CZE	Capillary Zone Electrophoresis
DAD	Diode Array Detector
DBSCAN	Density Based Spatial Clustering of Applications with Noise
DPPH	2,2-diphenyl-1-picrylhydrazyl
DT	De-Trending
ECD	Electron Capture Detector
ELSD	Evaporative Light Scattering Detector
EMA	European Agency for the Evaluation of Medicinal Products
FD	Fluorescence Detection
Fe ³⁺ -TPTZ	Ferric 2,4,6-tripyridyl-s-triazine complex
FG	Radix Pueraria Thomsonii
FRAP	The Ferric Reducing Antioxidant Power assay
FT-IR	Fourier Transform Infrared Spectroscopy
GA	Genetic Algorithms
HD	Hydrodistillation
HDL	High Density Lipoprotein
HPLC	High Performance Liquid Chromatography
HPTLC	High Performance Thin Layer Chromatography
HRE	Heat-Reflux Extraction

ICH	International Conference on Harmonization
KM	Kubelka-Munk theory
KNN	K Nearest Neighbors
KS	Kennard-Stone algorithm
LCMS	Liquid Chromatography coupled with Mass Spectrometry
LDA	Linear Discriminant Analysis
LLS	Local Least Square
LOD	Limit of Detection
LOQ	Limit of Quantification
MAE	Microwave-Assisted Extraction
MC	Mean Centering
MEKC	Micellar Electrokinetic Chromatography
MIR	Middle Infrared Reflectance
MLR	Multi-Linear Regression
MS	Mass Spectrometry
MSC	Multiplicative Scatter Correction
NACE	Non-Aqueous CE
NIR	Near-Infrared
NIRS	Near-Infrared Reflectance Spectroscopy
NMR	Nuclear Magnetic Resonance
OA	Oleanolic acid
PASG	Pharmaceutical Analytical Sciences Group
PC	Principal Component
PCA	Principal Component Analysis
PCR	Principal Component Regression
PLE	Pressurized Liquid Extraction
PLS	Partial Least Square
PLSDA	Partial Least Squares Discriminant Analysis
PRESS	Predicted Residual Error Sum Square
QDA	Quadratic Discriminant Analysis
RMSEC	Root Mean Square Error of Calibration
RMSECV	Root Mean Square Error of Cross Validation
RMSEP	Root Mean Square Error of Prediction

RPA	Relative Peak Area
RRLC	Rapid Resolution Liquid Chromatography
RRT	Relative Retention Time
RSD	Relative Standard Derivation
SG	Savitsky-Golay
S/N	Signal-to-Noise ratio
SA	Simulated Annealing
SEC	Square Error of Calibration
SEP	Square Error of Prediction
SFDA	Chinese State Food and Drug Administration
SFE	Supercritical Fluid Extraction
SI	Similarity Indices
SIMCA	Soft Independent Modeling of Class Analogy
SNV	Standard Normal Variate
SPE	Solid-Phase Extraction
SVR	Support Vector Regression
TBARS	Thiobarbituric Acid Reactive Substances
TCM	Traditional Chinese Medicine
TIC	Total Ionic Current
TLC	Thin Layer Chromatography
TOF	Time of Flight
UA	Ursolic acid
UAE	Ultrasound-assisted Extraction
UP	Ultra-Performance
UV	Ultra-Violet
UVE	Uninformative Variable Elimination
WeiL	Rhizoma Coptis Chinensis
YaL	Rhizoma Coptis Deltoidea
YES	Yeast Screening Assay
YG	Radix Pueraria Lobata
YunL	Rhizoma Coptis Teeta

Chapter 1: Introduction

1.1 Introduction

Chinese Herbal Medicine (CHM) has been used in Traditional Chinese Medicine (TCM) to prevent and cure disease in China for centuries, and becoming popular around the world during the last decades. However, safety events of CHM and the lack of scientific data supporting their efficacy are still the major factors hindering the globalization of CHM. The major focus of research studies on CHM are identification methods of CHM, separation and chemical analysis of chemical components, and pharmaceutical activities. Quality control of CHM is still an important challenge and more works have to be done to ensure safe consumption and to form a solid background for further scientific studies on biological activities, pharmaceutical effects and mechanisms leading to understanding of their efficacy.

Conventionally, identification of CHM is based on morphological examination. Thin Layer Chromatography (TLC) identification and quantitation of one or two selected markers by modern analytical instruments like High Performance Liquid Chromatography (HPLC), Gas Chromatography (GC) are the conventional analytical methods for identification level on chemical composition. However, these conventional analytical methods are time-consuming, labor-intensive, expensive, and require large amount of organic solvents. Furthermore, the current approaches usually examine only one or two selected markers and usually ignore the chemical profile of the herbs. However, the marker compounds may not be specific, and most of the time, may not be responsible for the pharmaceutical actions of the herb. The characteristics of multi-target and synergistic action of CHM because of complex chemical composition are not addressed in the conventional quality control practice. Spectroscopic methods present a rapid and comprehensive assessment of the samples

which would complement the chromatographic techniques.

Near-Infrared Spectroscopy (NIRS) is a rapid and nondestructive technique which could be used to determine different quality parameters such as identity and chemical composition. The spectra contain information of all chemical constituents which align with the integrated or holistic philosophy of traditional Chinese medicine ¹. However, application of NIRS in quality control of CHM is limited because the technique is not very sensitive; hence, it is usually used to detect major compounds of content higher than 30% in the sample. However, the content of markers in a CHM is far lower than that level. The challenge here is to improve the sensitivity of the technique with proper data processing so that it could be used to detect or predict the naturally occurring compounds as well as other parameters in CHM.

Chemometrics is always a core component in NIRS analysis. It is a discipline studying the application of mathematical and statistical methods in chemical analysis including topics such as optimization, data preprocessing to remove noises and backgrounds, pattern recognition techniques and correlation models, data treatment of chemical analyses. Chemometrics include topics like experimental designs and information extraction methods (Modelling, classification and test of assumptions).

In our study, we tried to use NIRS in quality control of three multi-species CHM *Puerariae Radix*, *Coptidis Rhizoma*, and *Ganoderma*. Chromatographic methods are developed and validated to determine the corresponding markers in these three herbs to provide accurate information to establish classification and quantitation models. We developed systematic procedures for developing NIR analytical methods and new

chemometrics procedures to establish more reliable models to assess different quality parameters (chemical contents, biological activities) in different CHM.

Puerariae Radix was selected as the first subject of study because more than one species are involved and the markers are abundant (2-7%). This study suggested the application of NIR technique in classification and quantitation of CHM is feasible. However, the results also suggested for compound less than 1%, the prediction error of the NIR technique increases significantly. We assumed that the spectra did contain the information, yet, the presence of other signals affected the performance of the quantitation models. If only the spectra regions that correlate well with the parameters being modeled are used, then the sensitivity of the technique will be improved. A variable selection algorithm is being developed in this work and applied to study *Coptidis Rhizoma*. Finally, the ultimate goal of quality control is to make sure the biological activities of the CHM are actually there. Thus, the possibility of using the NIR spectroscopy to predict biological activities of the herb was examined. The testing case selected is the anti-oxidant power of *Ganoderma*.

A review of the conventional practices and approaches of quality control of CHM is given in Chapter 2. Then chemometrics methods that commonly used in NIR analysis are described in Chapter 3. The results on *Puerariae Radix*, *Coptidis Rhizoma*, and *Ganoderma* are presented in Chapter 4, 5 and 6 respectively. Our results strongly suggested that NIR is a powerful technique that could supplement other analytical techniques for chemical analysis and quality control.

Chapter 2: Recent Developments on Quality Control of Chinese Herbal Medicines

2.1 Introduction

Traditional Chinese Medicine (TCM) is an empirical healthcare system based on experiences dating back to several thousand years ago. Chinese Herbal Medicine (CHM) was an important part of it, and they have been used in China for prevention and treatment of diseases. Each herb contains many chemical compounds that possess concerted actions ^{2, 3}. The herbs are usually prescribed in formulae. A formula is designed to organize these concerted actions derived from different herbs to create holistic, multi-targets, multi-dimensional pharmacological actions to achieve personalized therapy.

The prevalent use of herbal medicines has raised concerns over their quality, efficacy and safety due to their easy availability. The practice of CHM is popular in China and some Asian countries; and getting more available in western world. It was estimated that more than 1.5 billion people all over the world trust in the efficacy and safety of Chinese medicine ⁴. As the increasing popularity of Chinese medicine, the reports on the adverse reactions of CHM were also increased ⁵. Therefore, the public concern on the quality, safety and efficacy of the CHM and herbal remedies have grown extremely.

Quality control is crucial to ensure the safety, and hopefully, the efficacy of herbal medicines. Species authentication and quality consistency of the CHM material used is of the utmost importance because therapeutic actions based on unauthenticated or inconsistent materials are not reproducible, rendering the manufactured products ineffective. Herb, although the same drug from the same plant species, may contain different amount of the constituents or even different chemical profiles ⁶, the variation

depends on various factors such as often cited geographic location, season and time of harvest, post-harvest handling, processing and storage (Fig. 2.1).

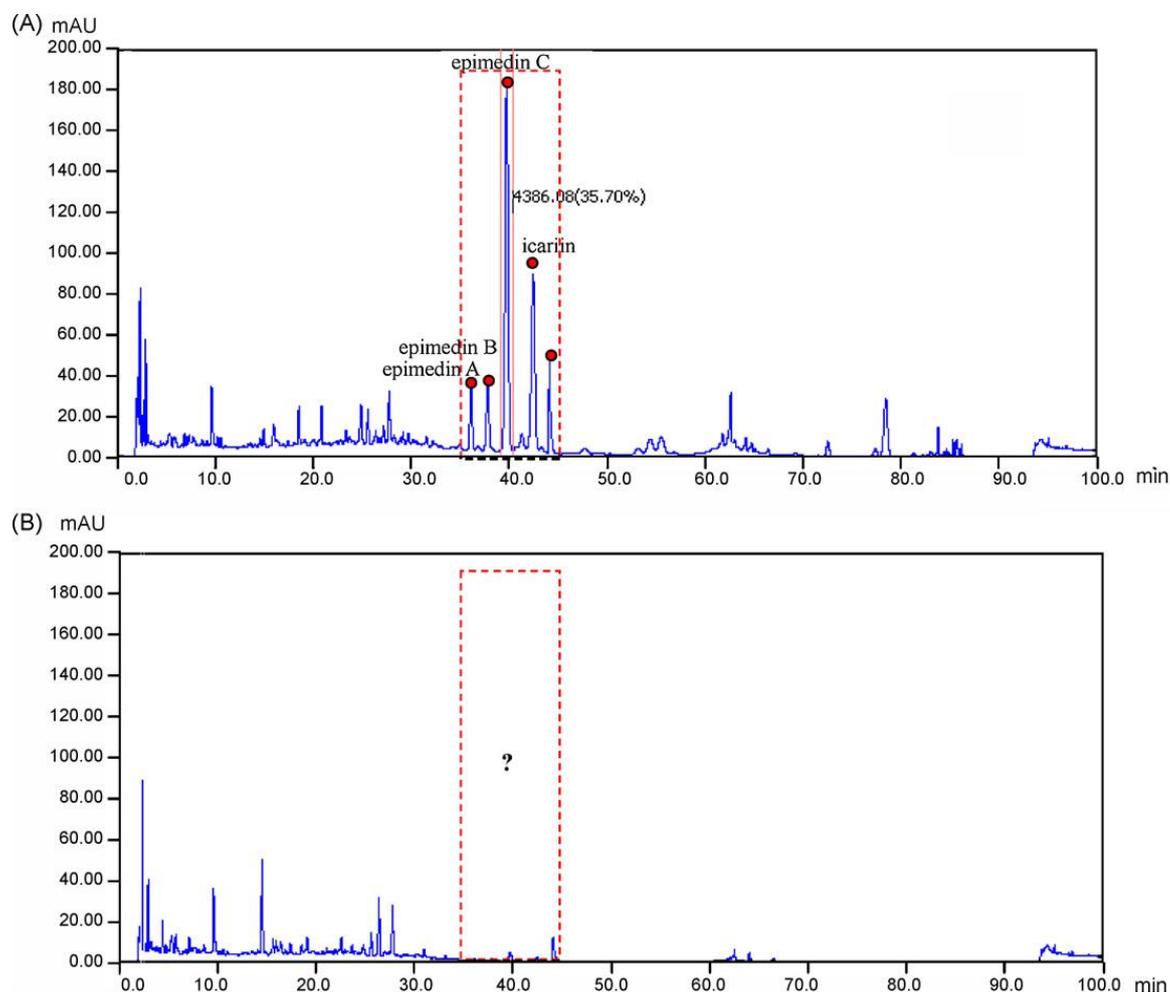


Fig. 2.1 a) Typical flavonoids fingerprint pattern of *Herba E. Sagittati*, cf. ; b) the main characteristic ingredients disappeared in this sample ⁶.

The exact part of the plant used and post-harvest processing could lead to totally different drugs with distinctly different chemical profiles as discovered in the recent decades by the scientific community. Because of the natural variations and inherent complexities of CHM, traditional identification techniques such as microscopic identification and morphological examination are far from satisfactory for identifying its safety and/or efficacy to safeguard consistent biological effects and

research results. The same hold true for the use of identification techniques using one or two marker compounds with chromatographic methods such as TLC⁷⁻¹⁰, HPLC, LC/MS etc. A multi-technique approach is necessary, which not only define the components comprehensively in CHM but also identify it from its confusable variety.

CHM consists of hundreds of unknown components and many methods do not provide a complete chemical profile of the drug, and it might be difficult to distinguish herbs with similar appearance or chemical markers. The characteristic of multi-target and synergistic actions of CHM are one of the major characteristic that distinguishes them from modern chemical drugs. This comes from their complicated constituents. Thus, in recent years, systematic research on CHM has begun to emphasize more on the integral and holistic property of the CHM¹¹. Many comprehensive methods, such as fingerprint and multi-component quantification have been used for the quality control of CHM in recent years. The challenge is to combine the appropriate techniques so that the quality of CHM could be fully evaluated. Chromatographic techniques (especially HPLC and HPTLC) are the most commonly used techniques, as they can separate and identify chemical components of the broadest range (from non-polar to polar), thus contributing to the maximum assurance of material equivalence. Currently, pursuing stringent quality control measure to safeguard consistency in CHM should be of highest priority. The common sample preparation methods, analytical methods and different approaches: multi-components, pattern and multi-pattern approach used in quality assessments of CHM will be discussed in this chapter.

2.2 Sample preparation methods

Sample preparation is one of the key steps which greatly influences the repeatability and accuracy of the analysis of CHM. It is reported that 70–80% of analysis time is spent on sample preparation and more than 60% of analysis error derived from non-standard sample pretreatment¹². Therefore, a proper sample preparation approach is vital for quality control of CHM.

Basic extraction methods such as sonication, reflux, liquid-liquid extraction, distillation and others were commonly used. However, these methods are time consuming and require large amount of organic solvents. Recent years, considerable efforts have been made to develop more efficient and faster extraction methods. The technological development for sample preparation of CHM has been reviewed^{9, 13, 14} and some recent development like Ultrasound-Assisted Extraction (UAE), Microwave-Assisted Extraction (MAE), Solid-Phase Extraction (SPE) are discussed.

2.2.1 Ultrasound-Assisted Extraction (UAE)

Ultrasound-Assisted Extraction (UAE) is a fast, inexpensive and efficient alternative to liquid-solid extraction methods. UAE has been reported for extraction of chlorogenic acid from *Eucommia ulmoides* Oliv.¹⁵ and isoflavonoids from *Radix puerariae*¹⁶. Under optimal conditions, UAE was highly efficient and performed much better than Heat-Reflux Extraction (HRE) and Pressurized Liquid Extraction (PLE). These studies proved that using UAE should be the most economic way.

2.2.2 Microwave-Assisted Extraction (MAE)

Microwave-Assisted Extraction (MAE) use microwave radiation to heat the sample in a very short time and accelerate the extraction. MAE has been widely used in extracting volatile compounds from CHM^{17, 18}. MAE with an adsorption-wave technique was used to extract trace-metal elements¹⁹ in medicine plants.

2.2.3 Pressurized Liquid Extraction (PLE)

Pressurized Liquid Extraction (PLE) combines elevated temperature and higher pressure to achieve fast, efficient extraction of the analytes from the solid matrix. Elevated pressure allows solvents to be heated above their normal atmospheric boiling points to improve their dissolution capacity and the rate of mass transport. Recently, Chen and co-workers determined 15 flavonoids in different species of Epimedium using PLE coupled to HPLC and UPLC^{20, 21}. PLE followed by GC-MS²² or HPLC-DAD-MS²³ was applied to quantitative determination of active components in *Curcuma longa*, *Folium Ginseng* and *Radix Ginseng*.

2.2.4 Solid-Phase Extraction (SPE)

Solid-Phase Extraction (SPE) is considered a well-established, widely-accepted and increasingly-used sample-preparation technique based on solid-phase adsorption. Many commercial SPE cartridges are routinely used for clean up and pre-concentration of CHM samples. Clarkson and co-workers proposed a post-column, online SPE device to concentrate flavonol glycosides in medicinal plants for HPLC-NMR. This hyphenation technique dramatically enhanced the sensitivity of

HPLC-NMR and speed up the structural determination of complex natural products ²⁴.

2.2.5 Supercritical Fluid Extraction (SFE)

Supercritical Fluid Extraction (SFE) is an environmentally-benign, efficient extraction technique that has been introduced and extensively studied for extracting organic compounds from medicinal plants. It can perform rapid extraction, and reduce or avoid the use of hazardous organic solvents by using carbon dioxide as the extraction solvent. The dissolving power of the supercritical fluid could be varied by adjusting experimental condition such as pressure and temperature. Furthermore, using SFE could avoid the potential degradation of chemical components in CHM resulted from long exposure to elevated temperature and atmospheric oxygen. Because of these advantages, SFE has been employed to extract volatile components from *Angelica* species ²⁵, ginsenosides from North American Ginseng ²⁶, and paeonol from Ji Sheng Shen Qi Wan (濟生腎氣丸) ²⁷.

Five methods, Ultrasound-Assisted Extraction (UAE), Pressurized Liquid Extraction (PLE), Supercritical Fluid Extraction (SFE), Hydrodistillation (HD) and Decoction (DC), were applied in the extraction of coniferyl ferulate, ferulic acid, Z/E-ligustilide and Z/E-butylidenephthalide, from *Angelica sinensis* Radix and compared ^{28, 29}. The results showed that the order of extraction efficiency was PLE≈UAE > SFE > HD, DC. The compositions of the UAE, PLE and SFE extracts, which had a high ratio of coniferyl ferulate, were very similar, while coniferyl ferulate was not obtained by HD and DC. However, HD and DC had higher selectivity for the extraction of ligustilide and ferulic acid.

2.3 Analytical techniques

In the last decades, there are many attempts in developing methods and techniques for analysis and quality control of herbal medicines. Chromatography techniques such as as Thin Layer Chromatography (TLC), Gas Chromatography (GC)/GC–MS, High Performance Liquid Chromatography (HPLC)/LC–MS, and Capillary Electrophoresis (CE) have been extensively applied for quality control of CHM, some of the developed methods are novel which coupled the more advanced instruments such as such as LC–DAD–ESI–MS_n, LC–UV–SPE–NMR, Capillary HPLC–1H NMR, 2D GC–TOFMS^{6, 7, 9-11, 30, 31}.

2.3.1 Thin Layer Chromatography (TLC)

TLC is a simple, low-cost, versatile and specific method for the identification of herbal medicines. The unique feature of picture like image of TLC supplies an intuitive visible profiling of CHM. Especially after combination with digital scanning and documentation software, TLC and HPTLC provide much more information and parameters for comprehensive identification and assessment of CHM. With this technique, authentications of many CHM are possible, and it can be used to evaluate the stability and consistency of their preparations from different manufacturers^{32, 33}. HPTLC is mainly used to study the compounds with low or moderate polarities, but Di et al. established a TLC method of fungal polysaccharide acid hydrolyzates by using automated multiple developments (AMD)³⁴.

2.3.2 Gas Chromatography (GC)

GC and GC-MS are well known methods for the analysis of volatile constituents in CHM, due to higher sensitivity, stability and efficiency than other instruments. The use of MSD provides reliable information for the qualitative identification of the compounds present in the complex compositions. With the help of the sample extraction methods (SPE, PLE, MAE)^{35,36} developed in recent years, the analysis of GC became the more rapid, low-cost, and environmental-friendly.

A novel technique – Comprehensive Two-dimensional Gas Chromatography (GC×GC) is considered to be the most powerful and versatile separation tool for volatile components. This technique couples two columns with different separation mechanisms via a modulator, so it has the advantages of enhanced sensitivity and superior resolution, capacity and separation. These facilitate the identification of unknown compounds³⁷. With this technique, a total of 769 compounds were tentatively identified by TOFMS and quantified by FID in *Notopterygii Rhizome et Radix*, which provide useful information to distinguish plants from different geographical sources³⁸. The development and application of the comprehensive two-dimensional Gas Chromatography (GC×GC) during 2003–2005 were summarized in a review³⁹. It could be used for the detection of constituents at trace level, establishment of the comprehensive fingerprint, and exploration of unknown volatile compounds in CHM or TCM preparation, etc. However, up to now, this powerful tool has not yet been widely used for the analysis of CHM.

2.3.3 High Performance Liquid Chromatography (HPLC)

HPLC is the most popular analytical methods among all, due to its easy operation, wide suitability and high accuracy for the qualitative and quantitative analyses of CHM. It is widely used in analyzing and quantifying compounds, very often with similar structures, in CHM ⁴⁰⁻⁴⁶. One of the main advantages of HPLC is that many detectors can be employed, such as Ultra-Violet Detector (UV), Diode-Arraye Detector (DAD), Evaporative Light Scattering Detector (ELSD), Flourecence Detector (FLD), Mass Spectrometry (MS), and Nuclear Magnetic Resonance Spectrometry (NMR), etc., which extends the possibilities for detecting different classes of constituents.

The UV detector is most frequently used. It is very convenient and sensitive for determining components with chromophore groups such as flavonoids, phenolic acids and alkaloids in CHM ⁴⁷. Compounds like saponins, which have very few chromophore groups, have a poor response with the UV detector. ELSD is an alternative detector for determination of non-chromophoric compounds in CHM ^{42, 48, 49}. For example, Chai et al. determined seven major saponins in *Lonicerae Japonicae Caulis* ⁴⁹. Recently, UV combined with ELSD in series has been successfully used for simultaneous determination of multi-components with different structures in a single chromatographic run. Li and coworkers determined six major active isoflavonoids and four main saponins in *Astragali Radix* by HPLC using UV and ELSD ⁴². The combination of UV and ELSD can extend classes of components that could be detected in an optimal condition to obtain higher sensitivity for each analyte target. The general implementation of ELSD for quantitative analysis is hampered by several factors, such a poor reproducibility and mobile phase-dependent characteristic of the detector. A new method, post-column mobile phase compensation, has been developed

to eliminate the strong dependence of detector response on the mobile phase composition during gradient elution⁵⁰.

Fluorescence Detection (FD) is usually more sensitive and can meet the sensitivity requirement for the limit testings of toxic compounds in CHM^{51, 52}. Cai and coworkers determined nephrotoxic and carcinogenic aristolochic acid (AA) in six CHM with FD by using pre-column derivatization with zinc powder⁵². This method provides a significant increase in sensitivity compared with other common analytical methods for AA. The limits of detection were 0.39 ng/mL for AA-I, and 0.52 ng/mL for AA-II, which were two orders of magnitude lower than those obtained from MS or Electron Capture Detector (ECD), and three-to-four orders of magnitude lower than the detection limits of UV detector. FD is sensitive, but its application is restricted because not many compounds in CHM show fluorescence, and derivatization procedure is usually time-consuming and less reproducible.

MS is becoming more and more popular as a detector for LC⁵³⁻⁶⁴. HPLC-MS provides higher selectivity and sensitivity for assaying minor components, isomeric compounds, or compounds without chromophore groups. A HPLC-ESI/Time of Flight TOF-MS was developed to identify and quantify 32 bioactive compounds in the flowers of *Lonicerae Japonicae Caulis* simultaneously (Fig. 2.2), with high selectivity, sensitivity and accuracy⁴⁶. Chen and coworkers developed a quantitative method, using SFE followed by Liquid Chromatography-Atmospheric Pressure Chemical Ionization-mass spectrometry (LC-APCI-MS) analysis for identifying isomeric compounds, oleanolic acid (OA) and ursolic acid (UA), in *Anoectochilus roxburghii* (wall.) Lindl⁶⁵. When MS is used for quantitative analysis, the experimental

conditions have to be optimized, such as ionization methods (ESI or APCI), detection mode (positive or negative ion mode) and composition of the mobile phase, in order to get stable and maximal signal to noise ratio. Tolonen et al. compared ESI and APCI techniques in the analysis of the main constituents from *Rhodiola rosea* L. extracts by LC-MS. The positive ion ESI mode was found to be the most sensitive ionization method for the analyte compounds studied, resulting two-five times higher signal-to-noise ratios for the extracted ion chromatograms than positive ion mode in APCI. Totally ten compounds were determined by Selected Ion Monitoring (SIM) mode using ESI in both negative and positive ion mode without mobile phase additives, while only seven were detected in negative ion mode using APCI⁶⁶.

Except availability of more sensible detectors for HPLC, the appearance of capillary HPLC and Ultra Performance (UPLC) have increased the analysis efficiency, so that shorter analytical time and better separation can be achieved^{67, 68}. More and more applications of UPLC in analysis of CHM have been reported^{21, 69, 70}. Another promising development in HPLC technique is the introduction of comprehensive two-dimensional LC (2D HPLC), resulting is much higher capacities, resolution and separation efficiency⁷¹.

Liquid Chromatography-Nuclear Magnetic Resonance (LC-NMR) can produce comprehensive information for the structure elucidation of novel compounds. LC-NMR skips the laborious and time-consuming isolation process, and therefore, it is particularly suitable for the identification of isomeric pairs and unstable compounds. In most LC-NMR protocols, Solid-Phase Extraction (SPE) is used to trap target compounds, deuterated reagents are used, and the stoped-flow mode is used to give a

longer scan time to record NMR spectra. In recent years, online LC–NMR combined with LC–UV/MS has attracted attention due to more qualitative information are obtained in a single run, including UV spectra, molecular weight, mass fragments, and NMR spectra⁷²⁻⁷⁵. If a suitable bio assay is combined with LC–NMR–MS analysis in parallel, biological screening and structure identification of the components can be done simultaneously⁷⁶.

2.3.4 Capillary Electrophoresis (CE)

Compared to HPLC, CE requires lower consumption of solvent and sample. Capillary Electrophoresis (CE) coupled with Mass Spectrometry (MS), is also very powerful in separating of characterizing compounds in CHM^{45, 46, 67-69}. It is a valuable alternative to LC-MS methods for the analysis of some special secondary plant metabolites (such as, alkaloids). There are different separation modes in the CE technique: Capillary Zone Electrophoresis (CZE), Micellar Electrokinetic Chromatography (MEKC), Non-Aqueous CE (NACE). CZE is suitable for neutral compounds. CZE uses an uncoated capillary column and requires less maintenance. This is the usual and simplest separation mode within CE which makes optimization of the experimental conditions easy. Li and coworkers successfully established a CZE method for simultaneous determination of four flavonoids, including icariin, epimedin A, epimedin B and epimedin C in *Herbra Epimedii*⁷⁷. Alkaloids are easy to be charged and separated by CZE, as described by Zhu and coworkers, who determined tropane alkaloids in *Daturae metel* L.⁷⁸. Wang et al. determined four alkaloids in <7 min, and this method was successfully applied to determine the amounts of opium alkaloids in *Papaver somniferum* L. samples⁷⁹. Analysis of alkaloids by CE avoids the peak

tailing problem usually occur in the HPLC method, while MEKC can separate charged and neutral analytes. A simple and rapid MEKC method was developed to separate and determine four toxic pyrrolizidine alkaloids in *Senecionis scandentis* Buch. Ham. and *Tussilago farfara* L. by Yu et al.⁸⁰.

In recent years, NACE, which is based on electrolyte solutions prepared from pure organic solvents, has become the current interest because many target analytes have good solubility in organic solvent. Chen and coworkers simultaneously separated three bioactive triterpenes in six CHM extractions by NACE⁸¹. NACE also offers additional advantages, such as better selectivity, smaller electrophoretic currents, and improved compatibility with mass spectrometry detection.

2.3.5 Spectroscopic methods

In addition to the above-mentioned chromatographic methods, there are also many spectroscopic methods used widely in the quality control of CHM, such as FT-IR, NIR and NMR. In comparison with traditional chromatographic methods, spectroscopic methods emphasize more on the integrative and holistic characteristics of CHM. Moreover, these techniques are simple, rapid with little or even no pre-preparation of sample.

Fourier Transform Infrared Spectroscopy (FT-IR) is originally a spectroscopic technique to identify the functional groups of the chemical constituents, now it has been used widely for the identification, quality control and manufacturing process of CHM in recent years. The techniques included conventional FT-IR, second derivative

spectroscopy (SD-IR), and Two-Dimensional Infrared (2D-IR) correlation spectroscopy^{82 83}. Among them, 2D-IR Spectroscopy was proposed by Noda in 1986. The technique was then introduced to the quality control of CHM by Sun⁸⁴. The 2D correlation analysis simplified the complex spectra consisting of overlapped peaks, by enhancing the resolution for interpretation in second dimension⁸². CHM powder samples are milled with potassium bromide (KBr) and then compressed into a thin pellet for analysis. Sun et al. had demonstrated the characterization of the chemical constituents in *Herba Epimedii*, *Angelica* and their extracts by using FT-IR and 2D-IR^{85, 86}. Based on these techniques, discriminations of *Chrysanthemums*⁸⁷, Licorice⁸⁸ and Danshen⁸⁹ from different geographic sources, as well as different species of *Fritillaria*⁹⁰, *Puerariae Radix*⁹¹ and animal drugs⁹² have been reported by this group of researchers. They have also used the same technique for injections from different manufacturers^{84 93}. FTIR was used to distinguish whether the cellular wall of *Ganoderma lucidum* spores samples have been broken⁹⁴. FTIR were used to identify the different origins or species of CHM, while quantitation of chemical markers in herbal samples is difficult.

Compared with FT-IR, Near-Infrared Spectroscopy (NIR) has a higher precision and easier sample preparation. As a result, in recent years, there has been an increasing trend on the use of NIR for the qualitative and quantitative analysis of CHM⁹⁵. However, the limitation of NIR is that the classification and quantitation models have to be established with a reference method like HPLC or GC before analysis of test samples, and the time required for model development is long. So, in recent years, different multivariate calibration regression methods have been compared in order to establish a rapid and reliable NIR method⁹⁶⁻⁹⁸. Through NIR method, the

identification of *Puerariae Radix*⁹⁶, *Paeoniae Radix*⁹⁹, *Pseudostellariae Radix*¹⁰⁰, *Scutellariae Radix*¹⁰¹, *Lonicerae Japonicae*¹⁰², and *Corydalis Rhizoma*¹⁰³ from different geographical sources, and the representative compounds in these plants have been quantified by the partial least squares (PLS) regression^{99, 101, 102}. Moreover, based on the properties of allowing rapid analysis of solid samples, NIR was applied as a tool for real-time and online monitoring of extraction process^{104, 105}, pharmaceutical blending process¹⁰⁶ and intermediates of a TCM injection⁹⁷.

Although NIR is impossible to replace the chromatographic LC method, our work reported in later part had demonstrated that the major component(s) (>0.1%) in CHM can be analyzed quickly and simultaneously by NIR⁹⁶.

Nuclear Magnetic Resonance (NMR) spectroscopy is a conventional structural identification technique of natural products, but recently, it is developed into an important tool for the qualitative and quantitative analysis of traditional herbal medicines, especially for the metabolomics and metabonomics studies of plants¹⁰⁷⁻¹⁰⁹.

2.4 Fingerprint analysis for quality control of CHM

Fingerprint method emphasizes the integrated chromatographic or spectroscopic characteristics of samples. It addresses the systematic and comprehensive nature of CHM, so it has been internationally accepted as one of the efficient methods to control the quality of CHM¹¹⁰. Today, The Chinese State Food and Drug Administration (SFDA) tried to standardize CHM injections and their corresponding raw materials

made from the CHM, by developing chromatographic fingerprinting for formulation from different manufactures. In Hong Kong, the Department of Health has published the “Hong Kong Chinese Materia Medica Standards” as a guideline to the quality requirement of the common CHM prescribed in Hong Kong and fingerprints are used as a mean to describe the quality of the herbs. These promote the use of the fingerprint techniques. There are several approaches in the fingerprint studies of CHM, the “multi-components approach”, the “pattern approach” and the “multi-pattern approach”.

2.4.1 Multi-components approach

The conventional analysis of herbs by identifying one or two markers in CHM can be considered as the simplest kind of fingerprint. If the bioactivity of the compounds were not known, a marker which is specific to the CHM could be selected for the analysis. Conventionally, the content of the marker is used as a reference to ensure the concentration of other compounds presented in the herbal mixture assuming that the ratio between the content of marker and other components are about the same. This approach is used in the quality control of the raw materials, intermediates as well as finished products. Marker approach is commonly applied to determine the batch-to-batch consistency of herbal products in manufacturing process, but in fact, only little chemical information was considered and monitored.

The multi-component approach is a natural extension of the marker approach, which assay as many bioactive markers as possible. Chemical constituents of CHM comprise of different classes of compounds (e.g., saponins, flavonoids, alkaloids and

anthraquinones). Many analytical methods have been developed for the simultaneous determination of compounds of certain class or different classes. Twelve polar and non-polar ginsenosides in red ginseng were analyzed using the RP-HPLC-PAD method simultaneously by Lee ¹¹¹.

However, the chemical compositions of CHM and the bioactive compounds are usually of low content. This made the isolation and identification of pure compounds difficult, and most reference components are not commercially available or very expensive. These disadvantages limit the multi-component approach to laboratory research.

2.4.2 Pattern approach

The basic concept of the fingerprint pattern is to consider the whole chromatographic profile or the whole spectrum as a feature for identification. It uses the relative compositions of many or even all identified components, i.e. the chemical profile of the sample, for identification or quality control purposes. Although the compounds present are not known, this approach appears to be more logical for evaluating the consistency of CHM and products.

Chromatographic fingerprinting characterizes the chemical composition of the CHM by a set of peaks. The samples with similar chemical patterns are likely to have similar compositions leading to similar biological activities. The similarity between profiles could be quantified or evaluated in terms of Euclidean distance, Pearson correlation coefficient, the cosine of the angle between the two patterns in the

multi-dimensional space, similarity ratio and etc. The similarity of the whole chromatogram serves as a measure to verify the composition of CHM herbal extracts or preparations. The chromatographic fingerprint is considered to be practical way to monitor CHM extracts and preparations as the consistency of the CHM products can be guaranteed ¹¹².

There are enormous examples in the literature reporting fingerprints for complex systems. The use of spectroscopic techniques coupled with chromatographic methods (e.g., high-performance thin layer chromatography (HPTLC), LC, GC or CE coupled with UV, DAD, ELSD, MS or MS²) provides spectral information which would be very helpful for the qualitative analysis and even structural elucidation of constituents of CHM ¹¹³. For example, Cai et al. developed the chemical fingerprint for *Rhizoma Gymnadenia* by HPLC–DAD–MSn ¹¹⁴. The fingerprint was constructed from UV chromatograms; similarity between chromatograms was evaluated with the Computer-Aided Similarity Evaluation System (CASES). Seven main peaks in the fingerprint were identified from their mass spectrum. We have illustrated the use of fingerprinting in the similarity and classification studies of *Puerariae Radix* in Chapter 4.

2.4.3 Multi-pattern approach

Some CHM, especially for complex prescriptions with many herbs, might have many different classes of compounds as active components. It is therefore almost impossible to use a single detector to reveal all chemical constituents in the CHM. It would be better to develop a fingerprint by combining various chromatographic approaches

with different detections to provide comprehensive information about the quality of CHM. This approach is referred as multi-pattern. Li and coworkers¹¹⁵ successfully developed a multi-pattern fingerprint of the total alkaloids from *Caulophyllum robustum* based on various chromatographic analysis using HPLC/DAD and HPLC/GS/MS. Van Nederkassel et al. established a *Ginkgo biloba* fingerprint chromatogram based on UV and ELSD detections¹¹⁶.

Another way to include more information is to include all spectroscopic data in the fingerprint. Chen developed a 2D fingerprint to analyze various Qingkailing injections by analyzing HPLC/DAD data with PCA¹¹⁷. This fingerprint compiled of chromatographic data at all wavelengths, much more data was used. In addition to 2D HPLC fingerprint, some other binary chromatographic fingerprints were developed by combination of two different separation principles, such as HPLC/DAD for aporphinoid alkaloids and GC/MS for quinolizidine alkaloids in *Caulophyllum robustum*, both fingerprints were combined to represent the total alkaloids contained in the herb¹¹⁵,

2.4.4 Chemometrics techniques for fingerprint analysis

Fingerprint, no matter the multi-component approach, pattern approach or the multi-pattern approach, is a way to represent the chemical information of the CHM extract or preparation. However, development of the chromatographic fingerprint could be difficult for some CHM which contain a large number of chemical components. Also, natural variations in the CHM and instrumental variations in the data are also unpredictable factors. Therefore, extracting useful chemical information

from the chromatographic data is important in defining the fingerprint.

Chemometrics is the application of mathematical and statistical techniques to retrieve information from the data. Thus, the application of chemometrics techniques could greatly improve the quality of the fingerprint obtained. The chromatographic profile of a complex mixture such as CHM extracts or preparations almost always contained overlapped peaks. These overlapped peaks hinder the identification of chemical components present as pure spectra of the corresponding components cannot be obtained. The Heuristic Evolving Latent Projections (HELP) technique, sometimes being referred as the Chemometrics Resolution Method (CRM) developed by our collaborator Liang and co-workers^{118, 119} is particularly useful in resolving overlapping peaks and extract chemical information in chromatographic fingerprint of CHM.

Chemometrics techniques are also very useful in building pattern fingerprint. First of all, as we have mentioned previously, a measure for similarity between chromatograms is essential for objective comparison. The similarity could be measured in terms of Euclidean distance, the Pearson correlation coefficient, or the cosine of the angle and similarity ratio. For more advanced comparison, pattern recognition, classification or clustering techniques such as Principal Component Analysis (PCA), Soft Independent Modeling of Class Analogy (SIMCA) and Artificial Neural Networks (ANN) can be used^{120, 121}. Details of the chemometrics techniques will be discussed in next chapter, and the application of the classification techniques is discussed in the study of *Radix puerariae* in Chapter 4.

2.5 Summary

Quality control plays an important role in the researches and developments of CHM. Different modern extraction methods for CHM, as well as chromatographic and spectroscopic methods for separation and characterization used in the quality control of CHM are reviewed. The different approaches of the chemical fingerprint analysis are also described, which include ‘multi-components approach’, ‘pattern approach’ and ‘multi-pattern approach’. Pattern approach was widely accepted by Chinese authorities in last decade, which served an important step towards standardization and modernization of CHM. We believe that including all chemical information in data analysis is very important to the quality control of CHM. This presents the characteristic of multi-target and synergistic action of CHM as different constituents are represented by the pattern. In the development of fingerprint analysis, data acquired from hyphenated instruments analyzing with chemometrics techniques provide the most powerful tools at the moment for establishing chromatographic fingerprint of CHM and serve as important foundation to examine their pharmacological activity.

Chapter 3: Principles and Applications of Near-Infrared Spectroscopy in Quality Control

3.1 Introduction

Near-Infrared Reflectance Spectroscopy (NIRS) is relatively speaking, a new spectroscopic technique that offers many advantages with a broad range of applications. The near infrared bands are severely overlapping and difficult to interpret by inspection. However, by using suitable mathematical and chemometric techniques, NIRS could be a powerful analytical tool for chemical analysis and quality control in agriculture^{122,123}, food¹²⁴, chemical¹²⁵ and oil industry¹²⁶. In the last decade, NIRS together with chemometrics techniques have proven their effectiveness for both qualitative and quantitative analyses in different fields.

NIRS examines transitions between the visible and mid-infrared spectral regions (800–2500 nm or 12821–4000 cm^{-1})¹²². The transitions involved are mainly vibrations of –CH, –OH, –SH and –NH bonds. All the absorption bands are the results of overtones or combinations of the fundamental vibrational transitions¹²⁷. There are many references and papers which describe the theory of NIRS^{128,129}.

NIRS is generally recognized for its speed, low cost and non-destructive characteristic. Thanks to the improvements in the instrumentation, and the advancement in the computer power, the development of fiber optics that allow chemometric methods and the applications of NIRS increased exponentially. Chemometrics plays a very crucial role in the data interpretation in NIR. Chemometrics¹³⁰⁻¹³² is a discipline developing suitable mathematical and statistical methods for optimization of the experimental procedures and data treatments of chemical analysis. Major topics in chemometrics include experiments design, and information extraction methods (modeling, classification and test of assumptions). In this chapter, commonly used chemometric

techniques relevant to the analysis of NIR spectra are discussed. The three main classes of techniques are the following:

- **Mathematical pretreatments**
- **Classification methods**
- **Regression methods**

A summary on the previous studies on the application of the NIRS in CHM are discussed in this chapter.

3.2 Theory and features of NIR spectra

NIRS examines spectral features in the range from 780 to 2500 nm (12,500 - 4,000 cm^{-1}) which provides information related to the complex vibrational behavior of the molecules. Transitions in the NIR region mainly involve the vibrations of the O-H, C-H, C-O and N-H bonds of a molecule. These bonds are subjected to vibrational energy changes when irradiated with NIR radiations.

For diatomic molecule, the vibrational frequency f can be determined assuming the harmonic oscillator behaviour in which the force on an atom is proportional to the shift from the equilibrium position of the atom in the molecule (Hooke's Law) ¹²⁸ and the frequency is given by,

$$f = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad (\text{Eq. 3.1})$$

where k is the force constant which is a measure of the strength of the bonding in the molecule, and μ is the reduced mass.

In the harmonic oscillator model, the potential energy with respect to the bond distance is represented by a parabola centred at the equilibrium distance. After solving the equation of motion as described by quantum mechanics, the allowed energy states of a harmonic oscillator are evenly spaced. These energy states are labelled by a non-negative vibrational quantum number, ν , and the energy, E_n , of the n -th level is given by the following equation ¹²⁸,

$$E_n = hf\left(n + \frac{1}{2}\right) \quad (\text{Eq. 3.2})$$

where h is the Planck's constant and f is vibrational frequency defined above

The selection rule of the transitions between vibrational energy states of a harmonic oscillator is $\Delta n = \pm 1$. As the vibrational energy levels are evenly spaced and the energy difference between two successive levels is always equal to $E_{n+1} - E_n = f$, which is the vibrational frequency. At room temperature, most of the molecules stay in the vibrational ground state (i.e $n = 0$), the transition from $n = 0$ to 1 is the major vibrational transitions which is usually called the “fundamental”.

For polyatomic molecules, the vibrational motions became more complex. As more atoms are present, there are different modes of vibrations corresponds to oscillations of different atoms. For a non-linear molecule with N atoms, there are $3N - 6$ modes of vibrations, the vibrational motions of a polyatomic molecule can be considered as a series of independent harmonic oscillators. Thus, the vibrational energy for each mode is described by Eq 3.2. The vibrational energy of the molecule, E , is then given by the

following equation,¹²⁸

$$E(n_1, n_2, n_3, \dots) = \sum_{i=1}^{3N-6} \left(n_i + \frac{1}{2}\right) h f_i \quad (\text{Eq. 3.3})$$

where $n_1, n_2, \dots, n_{3N-6}$ and $f_1, f_2, \dots, f_{3N-6}$ are the vibrational quantum number and vibrational frequency of different normal modes, respectively.

Similar to the case of diatomic molecule, most of the polyatomic molecules reside in ground vibrational state (all vibrational quantum numbers equal to zero) at room temperature. For transitions involved only one vibrational mode, transitions from $n = 0$ to 1 is known as the fundamental and transition to higher vibrational levels are known as overtones. Transitions involved more than one vibrational mode which is known as combination bands. In the harmonic oscillator model, only fundamental transition is allowed, overtones and some combinations bands are forbidden. However, these transitions do observed in practice as harmonic oscillator model is only an approximation and real molecules behaves differently from harmonic model, which usually known as the anharmonicity effect.

In reality, the vibrational motion of molecules deviate from the harmonic oscillator model as the potential energy curve is only roughly parabolic at low energy. The nuclei charge on the nuclei limits the approach of the atom to each other when the bond distance is smaller than the equilibrium distance and the molecules will dissociate at large bond distance. Because the potential energy curve differs from a parabola centred at the equilibrium distance, energy differences between two consecutive energy levels decrease with increasing energy and quantum number. The

energy level of an anharmonic oscillator are expressed as polynomial of $(n + 1/2)$ as shown in the following ¹²⁸ :

$$E_n = f\left(n + \frac{1}{2}\right) - \chi f^2\left(n + \frac{1}{2}\right)^2 \quad (\text{Eq. 3.4})$$

where χ the an-harmonicity constant and f 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. The selection rule $\Delta n = \pm 1$ is no longer hold for anharmonic oscillator, thus, overtone at energy position approximately two, three, or more times higher than the fundamental frequency can be observed. The intensities of these higher energy bands decay abruptly because the transition probability decreases markedly with increasing Δn . In practice, only the first two or three overtones can be observed. Combination and subtraction bands arise from transition involve more than one vibration mode. The transition energy observed could be approximately a sum of the fundamental frequencies involved ($f_1 + f_2, 2f_1 + f_2, \text{etc.}$) for combination bands or involving their difference ($f_1 - f_2, 2f_1 - f_2, \text{etc.}$) for subtraction bands. These bands also have lower intensities comparing to the fundamentals. Furthermore, anharmonicity results in combination bands occurs at a frequency smaller than the sum of corresponding fundamental frequencies involved.

The energy absorption of organic molecules in NIR region occurs when molecules changes its vibrational state. The NIR region is divided into the short-wave NIR (SW-NIR) (780 – 1300 nm) and the common NIR (1300 - 2500 nm). The SW-NIR

region is considered as the absorption band of higher overtones and the latter corresponded to first or second overtone transitions. Since the absorption intensity decreases as overtone increases, thus, SW-NIR spectra is usually recorded in transmission mode with long path length, and common NIR is used in diffuse reflection analysis.

3.3 Working principle of Near-Infrared Reflectance Spectroscopy (NIRS)

Although the low molar absorptivity of absorption bands in the NIR region (typical between 0.01 and $0.1 \text{ mol}^{-1}\text{cm}^{-1}$) severely limits the sensitivity of this technique, but this permits operation in reflectance mode to record spectra of solid samples. For reflectance spectroscopy, it measures the light reflected by the sample surface, which contains a specular component and a diffuse component^{127, 133}. 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 contains more chemical information and can be described by the Kubelka-Munk (KM) theory which is the theoretical basis for measurements using this technique¹²⁸.

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 of propagation may be changed as a result of reflection, refraction and random diffraction at the surfaces of various particle boundaries. Combination of these effects is called light scattering. Since scattered light encounters more

boundaries of particles, further scattering occurs in all directions and some energy is absorbed, which diminishes the intensity of the NIR beam. Scattering and absorption take place simultaneously until the remaining attenuated light re-emerges eventually from the entry surface. This light is called diffuse reflection¹²⁸. 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 from a general radiation transfer equation to describe the diffuse reflection of radiation^{128, 134}. The Kubelka-Munk (KM) theory has been the most successful for NIRS with simplified solution to the radiation transfer equation. Kubelka-Munk theory assumed the particles in the layer are randomly distributed so that the scattered radiation is distributed iso-tropically and the particle layer is very much smaller than the thickness of the layer so that the layer is subject only to diffused reflection¹²⁸. With these assumptions, the following relationship can be obtained,

$$f(R_{\infty}) = \frac{(1 - R_{\infty})^2}{2R_{\infty}} = \frac{k}{s} \quad (\text{Eq. 3.5})$$

where $f(R_{\infty})$ is the KM function, and R_{∞} is the absolute diffuse reflectance of the sample, k is absorption coefficient of the sample and s is its scattering coefficient.

From the KM equation, it can be seen that the reflectance, which is measurable, is

only the ratio of two constants, k and s . In NIRS measurement, the relative reflectance, rather than the absolute diffuse reflectance is usually measured.

In practice, the 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 NIRS 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 sulphate, magnesium oxide, Teflon and ceramic plate are the standards commonly used in NIRS ¹³⁵.

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

$$f(R) = \frac{(1 - R_\infty)^2}{2R_\infty} = \frac{k}{s} = 2.303a \frac{c}{s} \quad (\text{Eq. 3.6})$$

where a is the absorptivity and $k = 2.303ac$, c is the concentration of the absorbing analyte.

A plot of $f(R)$ against c for sample conforming to this relationship will be a straight line.

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 an analyte with strong absorption bands, the diffuse reflectance of the sample does not fit the KM equation and the plot of $f(R)$ against concentration tends to be non-linear. Since the KM equation is only applicable to weak absorption bands, or when the product of absorptivity and concentration is small, this is true only within the NIR region. If the matrix absorbs strongly at the same wavelength region as the analyte, absorption by the latter cannot be resolved and deviations from the previous equation becomes much more significant¹³³.

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

$$A = \log \frac{1}{R} = a \cdot c \quad (\text{Eq. 3.7})$$

Where A is apparent absorbance, R is relative reflectance, c is analyte concentration and a is the proportionality constant.

Even though the KM equation is an empirical expression, it provides highly satisfactory results and can be used in many diffused reflectance spectroscopic applications¹³³.

3.4 Pre-processing methods for NIRS analysis

The NIR spectra acquired are affected by many factors such as the variation of particle sizes in the sample, different optical paths and crystalline forms. Therefore, a well defined and systematic sample preparation and analysis protocol is required to

control the particle size of the sample. To avoid or minimize these interferences, data pretreatments are applied on the spectra.^{136, 137} Certain transformation techniques are useful to remove baseline shifts, slope changes and curvilinearity of spectra, which arises from the physical effects of particle size, scattering and other influence factors.

The most commonly used data pretreatments are normalization methods like Mean Centring¹³⁸, derivation, Standard Normal Variate¹³⁹ (SNV) and Multiplicative Scatter Correction¹⁴⁰ (MSC), the derivative methods (for example the Savitzky-Golay method and¹⁴¹ the detrending^{140, 142, 143}). The spectral pretreatments and the wavelength ranges should be carefully chosen before the pattern recognition in order to optimize the performance of the model.

3.4.1 Mean Centering (MC)

Mean centering is applied by subtracting the mean spectrum of the data set from every spectra in the data set. For a data set $R(I \times J)$ of I samples with J discrete digitized wavelengths, the mean centered j -th wavelength of the i -th sample is defined by

$$R_{i,j}^{MC} = R_{i,j} - \left(\sum_{j=1}^J \frac{R_{i,j}}{J} \right) \quad (\text{Eq. 3.8})$$

The practical consequence of the mean centering is often a more simple and interpretable regression model. Estimated analyte concentration may be more precise following mean centering of the data. It should be noted that mean centering does not always yield the most precise calibration model. Each calibration method should be tested on mean centered and non-mean centered data.

3.4.2 Derivation

The general characteristic of the NIR reflectance spectrum is a rising baseline with small number of bands and little fine structures. The baseline depends on the physical properties, like the particle size of the sample. Second derivative of the spectra are used in order to minimize the background. As a result, sloping baselines and offset can be eliminated, while differences in signal height are retained. Also, the fine structures in the spectrum are improved. However, the derivation transformation would have the disadvantage of enhancing the noise in the original spectrum so and the signal-to-noise ratio of the resulting signal is reduced.

The most usual way to determine the derivative of spectra is the segment-gap method. Firstly, the spectrum 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 the signals that are constant at every wavelength while application of the second derivative removes the signals that are proportional to the wavelengths. In practice, second derivative spectra are used much more frequently used than the first-derivative spectra and higher orders derivation have no additional advantages with regard to the quality of the signals obtained.

3.4.3 Multiplicative Scatter Correction (MSC)

Multiplicative Scatter Correction (MSC) is a preprocessing tool developed to correct for the light-scattering which is significant in reflectance spectroscopy. MSC assumes that all the samples have the same scattering coefficients at all wavelength. An “ideal” or reference spectrum is required and. For that purpose, the average spectrum of the whole sample set is determined in order to estimate the scattering present in the spectra. Each individual spectrum is expressed in terms of the set-mean spectrum,

$$R_{i,j} = a_i + b_i \bar{R}_j + e_{i,j} \quad (\text{Eq. 3.9})$$

where $R_{i,j}$ is spectral intensity of the i -th sample at wavelength j , \bar{R}_j is the mean spectrum of the data set, $e_{i,j}$ is the residual spectrum at wavelength j , which ideally contains the chemical information of the data, a_i is the fitted offset or intercept, b_i is the fitted slope

In other words, 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, $e_{i,j}$ are the errors or residual spectrum, which are representative of the difference between $R_{i,j}$ and \bar{R}_j and are mainly attributable of the chemical information. Now, the MSC processed spectrum is given as follows:

$$R_{i,j}^{MSC} = \frac{(R_{i,j} - a_i)}{b_i} \quad (\text{Eq. 3.10})$$

Since the scattering and particle size are independent of chemical information, one often defines a sub-region of the spectrum, which represents explicitly the baseline and contains no other 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 the interpretation.

In pattern recognition, MSC is typically applied to each class separately because MSC includes the determination of the corresponding reference spectrum, and the definition of the correction terms for each class.

3.4.4 Savitzky-Golay (SG)

Savitzky-Golay (SG) is a smoothing technique commonly used in NIRS. This is a moving window averaging method; a suitable sized 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 ¹⁴⁴.

The basic concept of SG is a least square fit to of a polynomial of a given order to $2n+1$ consecutive points of the signal that are within in the moving window. $A_n, A_{(n-1)} \dots, A_{n-1}, A_n$ could be derived and used as Savitzky-golay filter coefficients to carry out the smoothing operation. The use of these weighting coefficients, known as convolution integers, turns out to be exactly equivalent to fitting the data to a polynomial, as just described and it is computationally more effective and much faster.

Therefore, the smoothed data point R_j^{SG} by the Savitzky-Golay algorithm is given by the following equation:

$$R_j^{SG} = \sum_{i=-n}^n A_i R_{j+i} \quad (\text{Eq. 311})$$

where R_j is the raw data at j -th wavelength, A_i is the Savitzky-golay filter coefficient when $i = -n, \dots, n$

The width of the smoothing window and the order of the polynomial could be chosen according to width of the peak in the signal and amount of smoothing desired.

3.4.5 Standard Normal Variate (SNV)

This transformation first centers the spectral intensity ($R_{i,j}$), and then subtracts the mean of the individual spectrum (\bar{R}_i) from the spectral values obtained at each wavelength j . These centered values are then scaled by the standard deviation calculated from the individual spectrum values. The SNV transformed spectrum then becomes

$$R_{i,j}^{SNV} = (R_{i,j} - \bar{R}_i) / \sqrt{\frac{\sum (R_{i,j} - \bar{R}_i)^2}{p-1}} \quad (\text{Eq. 3.12})$$

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 knowledge of an “ideal” or reference spectrum. In general, SNV can be combined with de-trending to remove the curvilinearity of spectra.

3.4.6 De-Trending (DT)

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

$$R_{i,j}^{DT} = R_{i,j} - \widehat{R}_{i,j} \quad (\text{Eq. 3.13})$$

This transformation also has the advantage that it can be applied to each individual spectrum, without the knowledge of an “ideal” or a reference spectrum. Normally, de-trending is carried out in combination with the SNV transformation.

3.5 Chemometric tools for classification

Classification of samples according to their NIR spectra is an important way for identification. This is also known as pattern recognition methods, and many important applications have been developed in chemistry ¹⁴⁵, biology ¹⁴⁶, and food sciences ¹⁴⁷ are important. The classification techniques can be divided into two categories: the unsupervised and the supervised methods. In the unsupervised classification, samples are classified without a prior knowledge on the identity of the samples. Classification is done by examining the natural clustering of the samples. Supervised pattern recognition is techniques in which a prior knowledge of the samples is needed. Thus, the classification model is developed on a training set of samples with known categories ¹⁴⁸. Then the model performance is evaluated by comparing the classification predictions to the true categories of the validation samples. This section briefly describes the different classification methods.

3.5.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is an unsupervised classification method. It is a feature reduction method which forms the basis for multivariate data treatment. PCA is used to visualize the multidimensional data. The most important PCA application is the reduction of the number of variables in multivariate data for better representation in a low dimensional space ¹⁴⁹. Thus, the new variables (loadings) are linear combinations of the original ones and can be interpreted like spectra. For example, drugs of different manufactures were analyzed using NIRS ¹⁵⁰. The PCA score plot confirmed statistical differences between the production sites, and the loadings identified the wavelengths region with most variation and showed that the excipients

were responsible of the differences.

Generally speaking, there are two classes of unsupervised clustering methods, the hierarchical and the non-hierarchical methods. Hierarchical methods proceed by successive divisions of the data set and result in a cluster sequence which can be represented as a tree. Non-hierarchical methods included Gaussian mixture models, K-means¹⁵¹, Density Based Spatial Clustering of Applications with Noise (DBSCAN)^{152, 153} or Kohonen neural network¹⁵⁴. The unsupervised methods are usually for exploratory studies. Supervised methods described in the next section are usually better, and found more practical applications.

There are three major differences between supervised pattern recognition algorithms¹⁵⁵. The first distinction between methods is whether the method focuses on discrimination, such as Linear Discriminant Analysis (LDA), or those that emphasis on similarity within a class, for example Soft Independent Modeling of Class Analogy (SIMCA). The second difference concerns linear and non-linear methods like neural methods. The third distinction is the parametric or non-parametric computations. In the parametric techniques such as LDA, statistical parameters of the normal distribution of samples are used in the decision rules. The classical methods for the supervised classification are correlation based methods, distance based methods, LDA, SIMCA, and Partial Least Squares Discriminant Analysis (PLSDA).

3.5.2 Correlation Coefficient

To categorize samples into clusters, similarity or dissimilarity between samples have to be measured. Commonly used quantities in NIRS to indicate similarity are Correlation Coefficient (R) and distances¹⁵⁶. R is defined as the cosine of the angle between the vector for the sample spectrum and the one for the average spectrum for each class in the data set. Different types of distance can be defined, such as the Euclidean or the Mahalanobis distance¹⁴⁸.

In supervised methods, a threshold is usually defined for classification. For example, if the correlation coefficient is higher than a certain threshold, the two spectra are considered as belonging to the same class. EMEA (European Agency for the Evaluation of Medicinal Products) guidance encourages the application of wavelength correlation threshold of 95% or the maximum wavelength distance¹⁵¹. However, the choice of threshold is somewhat arbitrary and based on experiences with the data.

3.5.3 Discriminant Analysis (DA)

LDA is a linear parametric method with discriminating characteristics¹⁵⁷. LDA focuses on finding optimal boundaries between classes. LDA like PCA is a feature reduction method. However, while PCA selects a direction that retains maximal variation in a lower dimension among the data, LDA selects the directions that achieve a maximum separation among the different classes¹⁵⁸. LDA uses Euclidean distance as the measure for discrimination while Quadratic Discriminant Analysis (QDA) is a non-linear classification method¹⁴⁸. It is generally believed that QDA can perform better in practical pattern recognition applications compared to linear

discriminant analysis (LDA) method. This is due to the fact that QDA relaxes the assumption made by LDA-based methods that the covariance matrix for each class is identical.

3.5.4 K-Nearest Neighbors (KNN)

K-Nearest Neighbors ^{159, 160} (KNN) is a non-parametric method. An unknown sample of the validation set is classified according to the class of the majority of its K nearest neighbors in the training set ¹⁶¹. The matrix of distances of the validation set samples to all other spectra of the training set is computed. The neighbors of an unknown sample are the samples having the lowest Euclidean norms. The predicted class is the class featuring the largest number of objects among the K nearest neighbors.

3.5.5 Soft Independent Modeling of Class Analogy (SIMCA)

SIMCA ¹⁶² is a parametric method which uses the modeling properties of Principal Component Analysis (PCA). For each class, a PCA is performed which leads to a Principal Component (PC) model for this class. The validation set is used with every class models. An unknown sample is assigned to the class described by the model that produces the smallest residue during the prediction.

SIMCA puts more emphasis on similarity within a class than on discrimination between classes.

3.5.6 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) is inspired by the structure and functional aspects of biological neural networks, which consisted of several layers of neurons: input, hidden and output layers. A neuron is a processing unit which transforms an input by an activation function into an output data ¹⁶³. In this case, the output data correspond to the class of the sample. Modern neural networks are non-linear statistical data modeling tools. They are usually used to model complex relationships between inputs and outputs or to find patterns in data.

3.6 Chemometric tools for developing quantitative models

NIR spectroscopy cannot be considered as a primary method. Therefore, to develop a quantitative model it is necessary to have a reference analytical method to evaluate the property of samples of interest such as the content of certain compounds. The value of the properties determined by accurate and robust analytical methods are known as the reference value of the samples. Chemometrics tools are then applied to determine the correlation between the spectral data and the reference value.

To predict the property of unknown samples from their NIR spectra, a model is built and validated by using at least two distinct sample sets. The first one is the calibration set used to establish the model. Another sample set is the validation set used to evaluate the prediction ability of the model to unknown samples.

The range of the validity of the models is affected by the samples collected and the variations that have been included in the samples ¹⁶⁴⁻¹⁶⁶. The most common partition is

that two third of the samples are used in the calibration set and one third is used as the internal validation set. Some studies on sample selection are comparing the use of Kennard-stone, successive projections algorithm, random sampling and full cross-validation on modeling with MLR and PLS ^{167, 168}. Wu et al. made a study on the influence of sample selection in the sets on neural networks models ¹⁶⁹.

Once the model has been constructed and validated, the NIR analysis become routine. It can be used for prediction from the spectrometer device which was used during the development step or on another device.

Knowing the identity of a sample is important. However, that is not sufficient for quality control. Quantitative analysis is the next step forward to verify the sample. Historically the first quantitative application of NIRS is determining the moisture of samples. NIRS is quite sensitive in this case because the two strong water absorption bands at 1450 and 1940 nm ¹²⁸. The following are the chemometric tools commonly used to construct a quantitative model.

3.6.1 Multi-Linear Regression (MLR)

The Multi-Linear Regression (MLR) ¹⁷⁰ is the oldest methods of this kind and better methods are available nowadays as the improvement of computation power. MLR establishes a linear relationship between a reduced number of wavelengths (or wavenumber) and a property of the samples (for example, content of a compound).

The prediction value y_i of the property can then be described with the formula:

$$y_i = b_0 + \sum_{j=1}^J b_j R_{i,j} + e_i \quad (\text{Eq. 3.14})$$

where b_j is the regression coefficient at j -th wavelength, $R_{i,j}$ the absorbance at the j -th wavelength of i -th sample and e_i is the error. Each wavelength is studied one after the other and correlated with the studied property. The selection is based on the predictive ability of the wavelength. There are three modes of selection: forward, backward, and stepwise. When the correlation at a particular wavelength reaches a threshold value set by the investigator, it is selected as one of the model calibration wavelengths. The model is then computed between this set of calibration wavelengths and the reference values of the studied property.

3.6.2 Principal Component Regression (PCR)

The Principal Component Regression (PCR) is divided into two steps. First the spectral data are treated with a PCA. Then a MLR is performed on the scores as predictive variables¹⁷¹. The prediction equation is written

$$y_i = b_0 + \sum_{j=1}^P b_j Z_{i,j} + e_i \quad (\text{Eq. 3.15})$$

where b_j is the regression coefficient at j -th wavelength, $Z_{i,j}$ is the principle component of i -th sample at j -th wavelength, p is the number of the principle component Z included in the regression, e is the residual.

The method has advantage that the PCA suppresses the spectral colinearity.

Meanwhile there is no guarantee that the computed principal components are correlated to the studied property.

3.6.3 Partial Least Squares Regression (PLS)

In PLS¹⁷², the regressions are computed with least squares algorithms. There are some similarities with the PCR. In both methods, they attempt to find the factors that will be regressed to Y variables. The major difference is, PLS is to establish a linear relationship between two matrices, the spectral data **X** and the reference values **Y**. PLS uses both the variations of **X** and **Y** to find out the new factors as the variables, while PCR only consider the variation of X only. The intention of PLS is to form factors that capture most of the information in the X variables, which is useful in predicting y_i , while reducing the dimensionality of the regression by using fewer factors than the number of X variables. This can be visualized by the representation of the spectra in the wavelength space to determine directions, represented by linear combinations of wavelengths called factors, which best correlate with the property under interested (say the content of a compounds).

3.6.4 Artificial Neural Networks (ANN)

ANN can also used for developing correlation models. As described before, ANN consists of input, hidden and output layers. The first one, called the inputs, where reflectance at specific wavelengths are feed in. In this case, the output is the content of the compounds. The hidden layers need to be trained by back-propagation, i.e., computing the transfer functions from the output to the inputs to achieve the necessary predictions.

The use of ANN requires the optimization of a lot of parameters such as the number of hidden neurons or the number of iteration to train the network, and of course the selection of the data pretreatment and the selection of the wavelengths are also important ^{173, 174}.

3.7 Applications of NIRS

This section highlights some of the application of NIRS in pharmaceutical fields and quality control of CHM. A brief summary on application of NIRS to quantitation analysis is given below.

3.7.1 Moisture determination

The moisture determination is one of the very first applications of NIRS in the pharmaceutical field. Water is a critical parameter that has to be controlled in a lot of pharmaceuticals because that affect for the stability of the product. There are strong water signal in the NIR spectral range with two bands centered at 1450 nm and 1940 nm. Thus, the technique could be applied even the instrument is not very sensitive. NIR spectroscopy is used to determine the water content in powders or granulates ¹⁷⁵⁻¹⁷⁷, tablets or capsules ^{178, 179}, as well as in lyophilized vials or in solutions ¹⁸⁰. This has been developed for a long time; most of the recent applications of NIRS in moisture determination are on-line detection in production environment.

3.7.2 Content determination

NIR spectroscopy is used for the quantitative determination of active compounds, excipients, or coating thickness in pharmaceuticals. Samples can be of various types, including powders, granulates, tablets, liquids, gels, films or lyophilized vials. NIRS methods were used to determine the content of an antimalarial-antibiotic (artesunate and azithromycin) co-formulation in hard gelatin capsule¹⁸¹. The content of aspirin and phenacetin were determined simultaneously in the aspirin tablets by NIRS¹⁸². Furthermore, NIRS is capable to determine the active components API ketoprofen as well as two preservatives in low concentration of a pharmaceutical gel, with the hydro gel exhibiting a strong matrix absorption in the NIR region¹⁸³.

More and more studies are published on the use of NIR spectroscopy to follow the process of tablets production, from the raw materials, either coated or uncoated, and on packaged tablets¹⁸⁴⁻¹⁸⁶.

3.7.3 Quality control of CHM

In the recent years, there were a few publications which are successful on establishing the quantitation models for active ingredient of CHM by using NIRS. For example, Chen and Sorensen developed that the quantitation models of glycyrrhizin and ginsenosides, the marker constituents of Radix Notoginseng. The correlation coefficients are 0.94 and 0.98 respectively through a modified PLS¹⁸⁷. Ren and Chen established the quantitation model of the total ginsenoside in American Ginseng with correlation coefficient of 0.99 by PLS¹⁸⁸. The quantitation models of the markers, hyperforin and biapigenin of St John's Wort, have been established by Rager et al.

using PLS with correlation coefficient 0.97 and 0.91 respectively¹⁸⁹. A SIMCA model was developed to identify the *Lonicera japonica*, six organic acids were determined by PLS model using NIRS¹⁰². NIRS was also applied on the quality control of Chinese medicine injection - *Tanreqing* injection, four markers baicalin, chlorogenic acid, ursodeoxycholic acid, chenodeoxycholic acid, and the total solid contents were determined¹⁹⁰. The application of NIRS on CHM is more difficult than in other fields, it is because CHM contain hundreds of chemical constituents which hinder the signal of the marker of interest, which is called matrix effect. The matrix effect could vary from sample to sample with different geographic origin. For the common cases, the markers present in the CHM are usually low (<1%), which is not high enough to be detected by NIRS.

We had attempted to perform quality control of CHM by using NIRS. Two species of *Phellodendri Cortex* were identified and differentiated by PCA, and the content of berberine and total alkaloid content in all *Phellodendri Cortex* samples were determined by NIRS⁹⁸. Two species of *Puerariae radix* were classified by using linear discriminant analysis (LDA) and soft independent modeling class analogy (SIMCA). Five isoflavonoids, puerarin, daidzin, daidzein, genistin and genistein were analyzed by NIRS with HPLC⁹⁶.

3.8 Summary

The basic principles of NIRS are described in this chapter. Data pre-processing is important to eliminate the background and enhance the differences in signals. With appropriate chemometric tools, NIRS is capable to perform classification of different types of samples and quantitative analysis of the interested compounds in the sample.

The chemometric techniques used in interpretation and analyzing the complex NIRS data were described. In detail, the commonly used pre-processing (MSC, SG, SNV...), classification (SIMCA, LDA...) and regression (MLR, PCR, PLS...) methods were discussed.

NIRS has proved its suitability for the analysis of pharmaceutical formulations and food products in western country long ago. However, its limited sensitivity can severely restrict its scope of application. That is the reason the application of NIRS in quality control of CHM were only found in recent. A systematic procedure of building the NIRS models with appropriate chemometric techniques will be the key topic in this study to increase the sensitivity of the models built.

**Chapter 4: Classification and
Quantitation Analysis of Puerariae
Radix**

4.1 Introduction

Conventional identification of Chinese Herbal Medicines (CHM) is based on morphological examination. Identification and quantitation of one or two markers by Thin Layer Chromatography (TLC), High Performance Liquid Chromatography (HPLC), and Gas Chromatography (GC) etc are the commonly accepted quality assessment methods as described in Chapter 2. Yet, these commonly used analytical methods are time-consuming, labor-intensive, expensive, and require large amount of organic solvents¹⁹¹. While in comparison with these methods, Near-Infrared Spectroscopy (NIRS) emphasize much more on the integrative and holistic characteristics of CHM, which could serve as a potential tool to correlate and determine different quality parameters of CHM from its NIR spectra as it is a rapid, accurate and nondestructive technique.

NIRS is commonly utilized for prediction of content of major chemical components (30-95%) in the samples and found many applications in pharmaceutical and other industrial fields. In this chapter, the feasibility of employing the NIRS technique to access the quality of an herb is being examined. *Puerariae Radix* has been chosen for study because more than one species are used, and it consists of high content of markers (2-7%). *Radix Puerariae* was obtained from the root of two species *Pueraria lobata* (Wild.) Ohwi (YG) and *Pueraria thomsonii* benth (FG) previously, but the two species have been listed as two different entries in Chinese pharmacopeia since 2005 due to the large difference in the isoflavonoids contents. The high content of the markers and large variation in the two species is favorable to NIRS analysis. Eighty-nine samples were collected and analyzed. The classification of the two species were done by using Linear Discriminant analysis (LDA) and Soft independent

modeling of class analogy (SIMCA), Partial least square (PLS) regression was utilized to develop correlation models to predict its isoflavonoids contents.

Some NIRS studies on CHM have been reported recently^{98, 189, 192-195} and there is an increasing trend on the use of NIRS for quality control of CHM. In this work, we tried to establish reliable quantitative and classification models based on NIRS spectra. A systematic and objective procedure for data analysis, selection of training sets, model construction and evaluation have been developed.

4.2 Background of Puerariae Radix

Puerariae Radix is known as Gegen in Chinese. The roots of *Pueraria lobata* (Willd.) Ohwi (YG) and *Pueraria thomsonii* Benth (FG) were both used as Puerariae Radix in the past. Since 2005, the Chinese Pharmacopeias only considered *Pueraria lobata* (Willd.) Ohwi as Puerariae Radix, the puerarin, the major isoflavonoid presents, in YG could be 10 times more than that of FG. This may lead to a significantly difference in the pharmacological effects. Traditionally, YG is used in treating diseases while FG is usually consumed as food.

Puerariae Radix is the dried root of the *Pueraria lobata* (Willd.) Ohwi, is collected in autumn and winter. YG is often cut into thick slices or pieces when fresh and dried. FG, known as “starchy Puerariae Radix”, is the root of *Pueraria thomsonii* Benth with the outer bark removed, dried for a while, then cut into sections or cut again longitudinally into two parts and dried¹⁹⁶.

The appearances of YG and FG are different. YG is longitudinally cut rectangular,

thick slices or small square pieces, 5-35cm long, 0.5-1cm thick. The outer bark pale brown, with longitudinal wrinkles, rough, cut surface yellowish-white, striations indistinct. Texture is pliable and strongly fibrous. FG, cylindrical, almost fusiform or semi-cylindrical, 12-15 cm long, 4-8 cm in diameter; some are longitudinal or obliquely cut thick slices, varying in size. It is externally yellowish-white or pale brown or greyish-brown when unpeeled. Transverse section shows pale brown concentric ring formed by fibers. The texture is hard and starchy¹⁹⁷.

According to TCM description, the properties of Puerariae Radix are sweet and pungent in flavor, cool in nature and attributive to the spleen and stomach meridians. Puerariae Radix is used to expel pathogenic factors from the muscles, to abate heat, let out the skin eruptions, promote the production of body fluid to relieve thirst and uplift yang to relieve diarrhea. It is used internally as spasmodic, antipyretic, secretory, anti-diarrheal and for the treatment of alcohol addiction, angina pectoris and hypertension¹⁹⁷.

4.3 Chemical and biological studies of Puerariae Radix

4.3.1 Chemical studies

Puerariae Radix contains rich isoflavonoids components like puerarin, genistin, genistein, daidzin, daidzein, 3'-hydroxy-puerarin and daidzein-4',7'-diglucoside

One major class of constituents in Puerariae Radix is isoflavonoids¹⁹⁷ which is also pharmaceutically active. The herbal extract is rich in isoflavonoids, which are also believed to be responsible for most of the biological activities of the herb. Several

isoflavonoids, such as puerarin and daidzin, have been isolated¹⁹⁸⁻²⁰². The puerarin content in YG is 5–10-fold higher than that in FG. Puerarin content should not be less than 2.4% and 0.3% for YG and FG, respectively according to Chinese Pharmacopeia

203

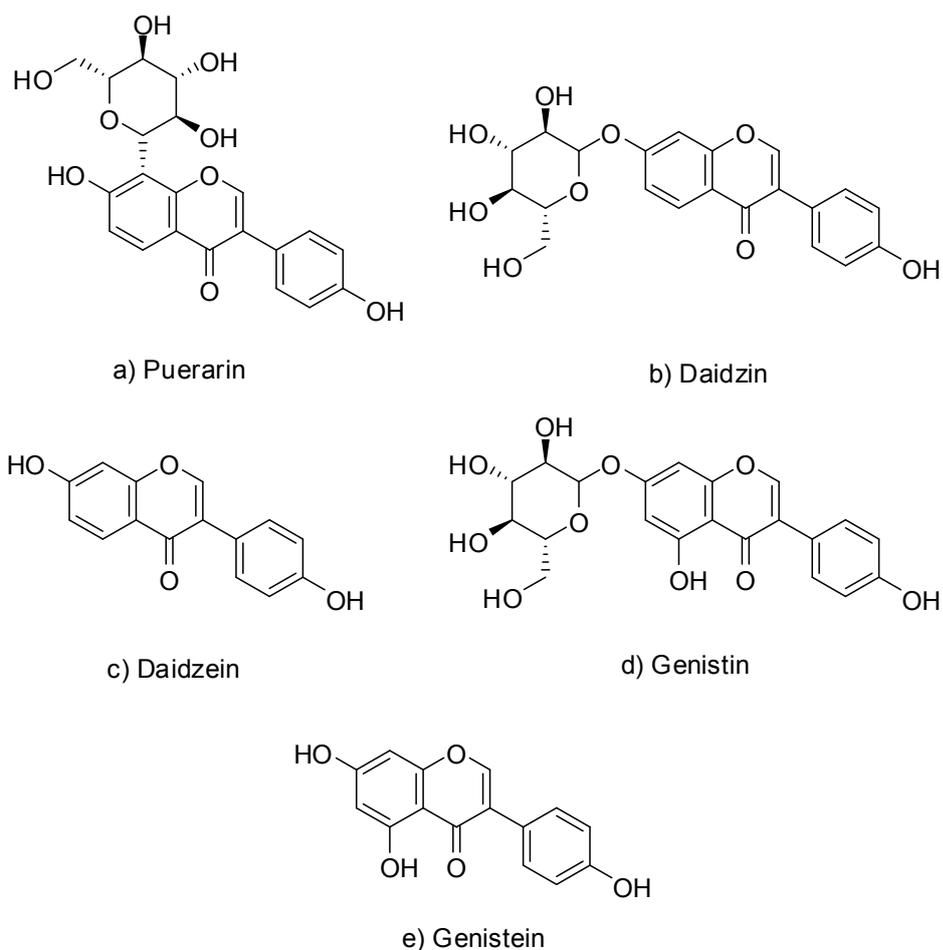


Fig. 4.1 The chemical structure of isoflavonoids a) puerarin, b) daidzin, c) daidzein, d) genistein, and e) genistin¹⁹⁷.

Besides, triterpene compounds are present in *Puerariae Radix*, which included kudzusapogenol A, kudzusapogenol B, kudzusapogenol C, kudzusapogenol B methylveter, sapogenol, sophoradiol, and cantoniensistro¹⁹⁷.

Several analytical methods were reported to distinguish the two species of *Puerariae Radix* by determination of the isoflavonoid content using High Performance Liquid Chromatography (HPLC) ^{33, 197, 204-206} and High Performance Thin Layer Chromatography (HPTLC) ^{33, 197}. However, both methods are destructive and time-consuming.

4.3.2 Biological studies of isoflavonoids of *Puerariae Radix*

Modern studies of *Puerariae Radix* had shown that its extract exhibits antidipsopic effect ^{200, 201, 207, 208}, antioxidant effect ^{202, 209-212}, estrogenic activity ^{213, 214}, anti-cancer effect ^{215, 216}, and neuro-protective effect ²¹⁷⁻²²⁰.

There are many reports on the effect of *Puerariae Radix* to reduce alcohol addition, Keung ²⁰⁷ demonstrated that a crude extract of *Puerariae Radix* suppresses the free-choice ethanol intake of ethanol-preferring golden Syrian hamsters and the results suggested that daidzein and daidzin account for this effect, a similar study using rat also reach the same conclusion ²⁰⁸. A clinical study had been conducted on some alcohol drinkers ²⁰⁰. Significant reduction in beer consumed has been observed in the drinkers after *Puerariae Radix* treatment as a result of a decrease in the volume of each sip. The underlying mechanism of alcohol consumption reduction effect of daidzin had been examined ²⁰¹, correlation between daidzin's capacity to reduce alcohol consumption and its ability to increase the liver mitochondrial monoamine oxidase (MAO): aldehyde dehydrogenase (ALDH-2) activity ratio has been established.

Several studies²⁰⁹⁻²¹¹ on Puerariae Radix extract confirmed its protective effect on the oxidative stress induced by hydrogen peroxide in vitro and in vivo. Puerarin has been found to be the active component. Thus, the aqueous extract of YG showed much stronger antioxidant activity than that of FG²⁰². Other than puerarin, it is found that quercetin and rutin inhibit the oxidation of High-density Lipoprotein (HDL) significantly²¹².

The isoflavonoids of YG showed estrogen-like effect on lipid metabolism in liver and adipose tissue²¹³. The estrogenic activity of Puerariae Radix was investigated by a recombinant yeast screening assay (YES)²¹⁴. The higher the isoflavonoid content in the Puerariae Radix extract, the stronger the estrogenic activity. This suggested that Puerariae Radix extract could be a good candidate for hormone replacement therapy.

The anti-cancer activity of Puerariae Radix had also been studied. A fraction PE-D purified from the ethanol-extract of Puerariae Radix has significant anti-proliferative effects on breast cancer cell lines MCF-7, MDA-MB-231, ovarian cell line (2774) and cervical cancer cells lines (HeLa). The anti-cancer activity is due to the component exhibits a similar behavior as that of 17β - estradiol, which activates both estrogen receptor α (ER α) and β (ER β)²¹⁵. Another finding showed that puerarin promote the apoptosis of colon cancer call line (HT-29) and reduce cell viability, which suggesting it could be a chemopreventive and/or chemotherapeutic agent for colon cancer cells (HT-29) by reducing cell viability and inducing apoptosis²¹⁶.

Puerariae Radix had been claimed to have neuro protective effect. Yan²¹⁷ reported the Puerariae Radix extract exhibited anti-depressant effect to neuronal damage caused by

CIR (Cerebral Ischemia Reperfusion), these damages play an important role in the development of post stroke depression (a common and severe complication after stroke)²¹⁸⁻²²⁰.

From the above information, it is quite clear that the isoflavonoids in *Puerariae Radix* are the active constituents which account for the major biological effects for the therapeutic actions of the herb. Therefore, the content of isoflavonoids in *Puerariae Radix* is an important parameter to evaluate the quality of *Radix Puerariae*.

4.4 Experimental methods and materials

4.4.1 Samples and chemicals

Seventy-five YG and fourteen FG samples were collected from different provinces of China. All samples were authenticated by Dr Si-Bao Chen (State Key Laboratory of Chinese Medicine and Molecular Pharmacology, Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China). Table 4.1 shows the species and the geographic origins of the samples. All samples were grinded into powder and passed through a 100 mesh (150 μ m) stainless sieve. The sieved powdered samples were stored in plastic containers at 20°C. Samples were freeze dried for 2 days before analysis.

Puerarin, daidzin, daidzein, genistin and genistein, were purchased from International Laboratory (USA) to serve as standards or markers. HPLC grade acetonitrile and acetic acid were purchased from Tedia (USA) and deuterated dimethyl sulfoxide

(99.8%) was obtained from Armar. Double deionized water was prepared by a Milli-Q water-purification system (Millipore, MA, USA).

Sample No.	Sample Code	Latin Name	Origin
1	FG-02	<i>Pueraria thomsonii</i> Benth	Hunan, China
2 – 4	FG-03 – 05	<i>Pueraria thomsonii</i> Benth	Guangdong, China
5	FG-06	<i>Pueraria thomsonii</i> Benth	Yujiang, Jiangxi, China
6	FG-07	<i>Pueraria thomsonii</i> Benth	Nanjing, Jiangsu, China
7	FG-08	<i>Pueraria thomsonii</i> Benth	Baoan, Gunagdong, China
8	FG-09	<i>Pueraria thomsonii</i> Benth	Shenyang, Liaoning, China
9	FG-10	<i>Pueraria thomsonii</i> Benth	ChengDu, SiChuan, China
10	FG-11	<i>Pueraria thomsonii</i> Benth	Shanxi, China
11	FG-12	<i>Pueraria thomsonii</i> Benth	Guangxi, China
12 – 14	YG-01 – 03	<i>Pueraria lobata</i> Ohwi	Huoshan, Anhui, China
15 – 18	YG-04 – 07	<i>Pueraria lobata</i> Ohwi	Putian, Fuhian, China
19 – 21	YG-08 – 10	<i>Pueraria lobata</i> Ohwi	Hunan, China
22 – 23	YG-11 – 12	<i>Pueraria lobata</i> Ohwi	Hangzhou, Zhejiang, China
24	YG-14	<i>Pueraria lobata</i> Ohwi	Shanghai, China
25	YG-16	<i>Pueraria lobata</i> Ohwi	Shenyang, Liaoning, China
26	YG-17	<i>Pueraria lobata</i> Ohwi	Nanjing, Jiangsu, China
27 – 50	YG-18 – 41	<i>Pueraria lobata</i> Ohwi	Huoshan, Anhui, China
51 – 52	YG-42 – 43	<i>Pueraria lobata</i> Ohwi	Shanxi, China
53	YG-44	<i>Pueraria lobata</i> Ohwi	Sichuan, China
54 – 59	YG-45 – 50	<i>Pueraria lobata</i> Ohwi	Huoshan, Anhui, China (1 Year old)
60 – 65	YG-51 – 56	<i>Pueraria lobata</i> Ohwi	Huoshan, Anhui, China (2 Years old)
66 - 71	YG-57 – 62	<i>Pueraria lobata</i> Ohwi	HuoShan, AnHui, China (3 Years old)
72 - 77	YG-63 – 68	<i>Pueraria lobata</i> Ohwi	Huoshan, Anhui, China (4 Years old)
78 - 83	YG-69 – 74	<i>Pueraria lobata</i> Ohwi	Huoshan, Anhui, China (5 Years old)
84 - 89	YG-75 – 80	<i>Pueraria lobata</i> Ohwi	Huoshan, Anhui, China (Above 5 Years old)

Table 4.1 Puerariae Radix samples collected for the study.

4.4.2 High Performance Liquid Chromatography (HPLC) analysis

Chromatographic analysis was carried out on a Thermo ODS hypersil column (250mm×4.6mm,5µm) at ambient condition using an Agilent 1100 liquid

chromatography system, equipped with a quaternary solvent deliver system, an auto-sampler and a DAD system. The detection wavelength was 254 nm. The mobile phase consists of (A) acetonitrile and (B) 0.3% (v/v) acetic acid and the gradient elution program is as follows: 0–25% (A) at 0–30 min; from 25% to 45% (A) at 35–40 min and from 45% to 95% (A) at 40–45 min. There was 10 min re-equilibrium between injections. The flow rate was 1.0mL/min and aliquots of 10 μ L were injected into the HPLC.

0.2 g sample, accurately weighed, was sonicated with 20mL 50% ethanol for 30 min. Then, the mixture was filtered and evaporated to dryness with a rotary evaporator. Afterward, the residue was re-dissolved in 10mL and 20mL methanol for FG and YG samples, respectively. Finally, the sample solution was filtrated through a 0.45 μ m syringe filter before HPLC-DAD analysis.

4.4.3 Near-Infrared Spectroscopic (NIRS) analysis

All NIR spectra were recorded by a NIRSystems Model XDS spectrometer (Foss NIRSystems, Silver Spring, MD, USA) equipped with a quartz halogen lamp and PbS detector. NIR spectra of the all samples were collected in reflectance mode over the spectral region 1100 to 2500 nm at a resolution of 0.5-nm. The spectra were acquired with a circular sample holder with a quartz window (38 mm in diameter and 10 mm in thickness). Each sample spectrum was obtained by averaging 32 scans. All spectra were recorded as the logarithm of the reciprocal reflectance, $\log(1/R)$, with respect to the ceramic reference standard. Each sample measurement was repeated three times after rotation of the sample holder by 120 $^{\circ}$, and the three spectra collected were

averaged.

Several data preprocessing methods, including derivation, Multiplicative Scatter Correction (MSC), Savitsky-Golay (SG), Standard Normal Variate transformation (SNV) and detrend have been applied to all NIR spectra to minimize the interferences such as scattering effect and baseline shift, as well as to enhance the spectral differences and to smooth the spectra etc. All data analysis was performed using a VISION (3.0.1.0), FOSS Nirsystems, and MATLAB (version 6.5)

4.5 Results and discussion

4.5.1 Purity determination of the chemical standards by quantitative H-NMR

The purity of the standard compounds affects the subsequent quantification analysis. To verify the purity of the standard compounds, about 1.5mg of markers and 1.2mg of internal standard (maleic acid 99%, Aldrich) were weighed accurately and dissolved in 0.3mL of deuterated dimethyl sulfoxide. All ¹H NMR experiments were performed at 400MHz on a Bruker DPX400 spectrometer at 298K. Typical acquisition conditions for ¹H NMR were 45°-pulse length, 5.0 s relaxation delay, 32,768 data points and 32 transients. Spectra were processed using an exponential function with a line-broadening coefficient of 0.3 Hz. The peak area of H3' and H5' of the isoflavonoids were compared with that of the internal standard signal at 6.26 ppm to obtain the content of the isoflavonoids. The purities of puerarin, daidzin, daidzein, genistin and genistein were determined to be 97%, 99%, 99%, 99% and 99%,

respectively by quantitative H-NMR using maleic acid as a reference (Fig. 4.2). An appendix 4.1 shows the NMR spectra of the five standards.

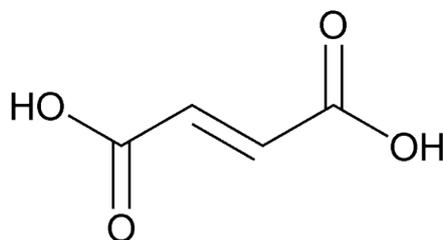


Fig. 4.2 The chemical structure of maleic acid

4.5.2 Identification of Puerariae Radix by HPLC-DAD

The NIRS spectra consist of overlapping bands which arises from many chemical constituents. Assuming each chemical constituent has a distinct NIR spectrum; the whole spectrum is a sum of all and depends on the chemical compositions. By correlating the spectra with samples of different chemical compositions and contents, one can express the contents of certain compounds as weighted sums of spectral intensities at different wavelengths. The weighting factors can be obtained from appropriate mathematical analysis on spectra from samples containing known amounts of certain compounds. For example, if there are n compounds of interests, spectra collected from n samples contain different amount of these n compounds are required. The spectral intensities can be expressed as a weighted sum of the spectra of these n compounds plus a matrix spectrum. This results in simultaneous equation of n unknown (the unknown are the contents of n compounds). A unique solution for the weighting factor can be obtained by solving the simultaneous equation if the contents of the n compounds are known. In other words, an analytical method that can determine the identity and/or contents of the compound selected for study is required

in order to determine the weighting factors for calculating the content from spectral intensities. This analytical method is known as the reference method. In this work, a HPLC-DAD method was developed as the reference method to obtain the contents of the isoflavonoid in the samples. Thus, prior to NIRS quantitative analysis, the isoflavonoids compositions of *Puerariae Radix* samples were examined by HPLC-DAD. The details of the HPLC method are described below. The method validation data are given in Table 4.2 and the results are discussed in 4.5.3.

4.5.2.1 Optimization of extraction and chromatographic conditions

Before the chromatographic analysis, the effects of the different experimental conditions on the amount of puerarin and daidzin being extracted were examined. Methanol and 95% ethanol were often used as the extracting solvent for isoflavonoids^{204, 221-223}. In this work, other different solvent systems, including 30%, 50% & 70% methanol, 30%, 50% & 70% ethanol and water were also investigated. It was found that 30%, 50% methanol and 30%, 50% & 70% ethanol have similar performance on isoflavonoid extraction (See Fig.4.3). 50% ethanol was selected for our subsequent experiments as the extraction efficiency of this solvent is higher than 95% for all the five isoflvaonoids being studied.

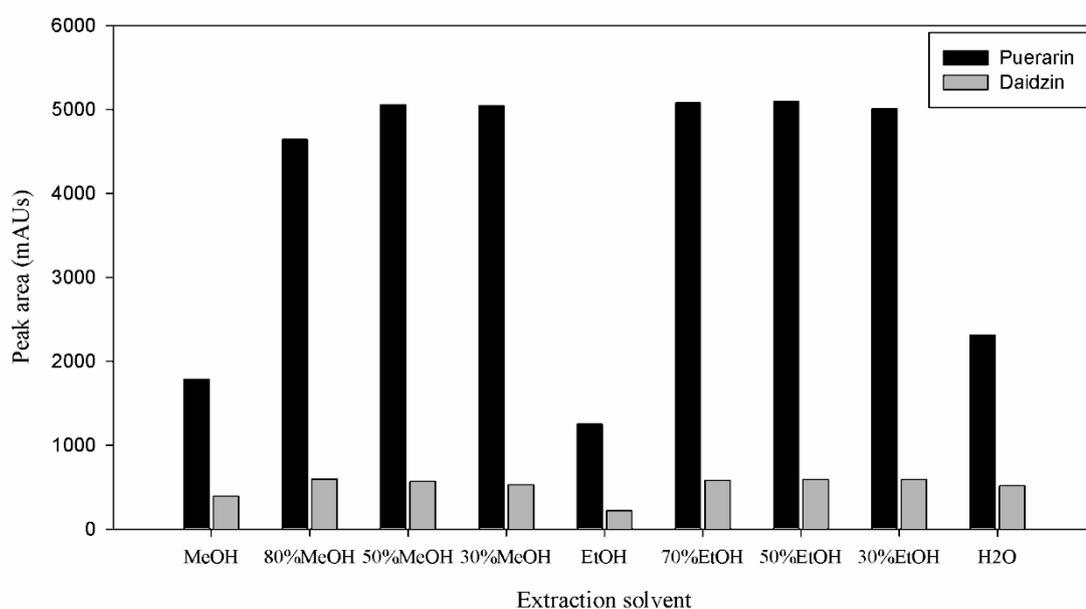


Fig. 4.3 Peak area of puerarin and daidzin content of Puerariae Radix sample by using different extraction solvents with 1 hour sonication

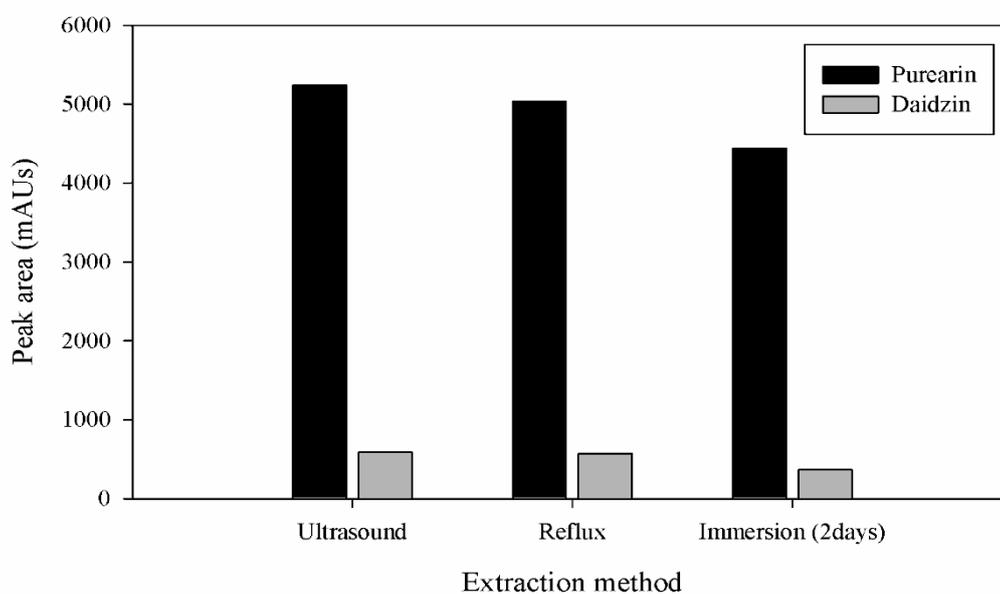


Fig. 4.4 Peak area of puerarin and daidzin content in Puerariae Radix samples using different extraction methods

Afterward, three different extraction methods, including ultrasonic, reflux and soaking were carried out and the results indicated that the ultrasonic method was the best (See Fig. 4.4).

Finally, the extraction durations for ultrasonic method (from 15 min to 2 hrs) were also examined and the results suggested 30 min was optimal (See Fig.4.5). Thus, the best extraction procedure for isoflavonoids in *Puerariae Radix* is sonification with 50% ethanol for 30 min.

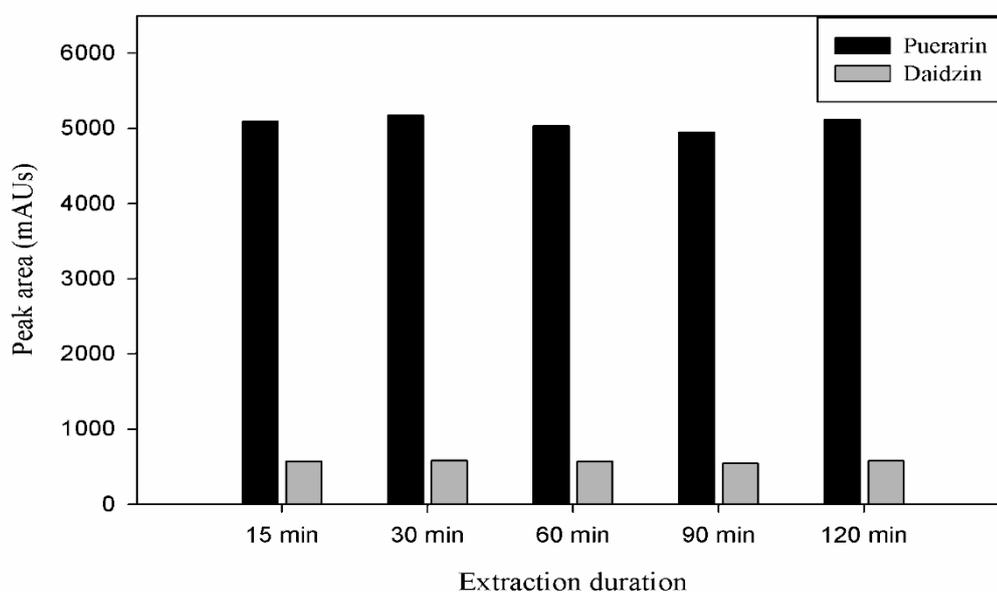


Fig. 4.5 Peak area of puerarin and daidzin content in *Puerariae Radix* samples using sonication method with different duration

In this work, acetic acid was added to minimize the tailing effect and to suppress the ionization of the analytes in the mobile phase during the separation process. The best separation after adjusting mobile phase compositions, flow-rate and gradient programs were given Section 4.4.3. The relative retention time (RRT) and relative peak area of the major peaks detected (i.e. puerarin and daidzin > 5% of total peak area) were used to evaluate the quality of chromatogram. The precision was determined by five replicated injection of the same sample solution within a day and their RSD of RRT and RPA obtained were less than 5% for all 89 samples. This suggested that the

chromatographic condition gives stable separation for isoflavonoid components.

4.5.2.2 Chromatographic studies of YG and FG by HPLC-DAD

Fig. 4.6 shows the typical chromatograms of the standards solution containing the five isoflavonoid components, the extract of FG and the extract of YG, respectively. The presences of the five isoflavonoids were confirmed by comparing the retention time and UV spectrum of the corresponding peak with those of the standards. The retention time of puerarin, daidzin, genistin, daidzein and genistein are 23.71, 27.40, 32.30, 37.63 and 39.78 min respectively. Overall result suggested that all five isoflavonoids were found in YG samples, while genistin was absent in FG samples.

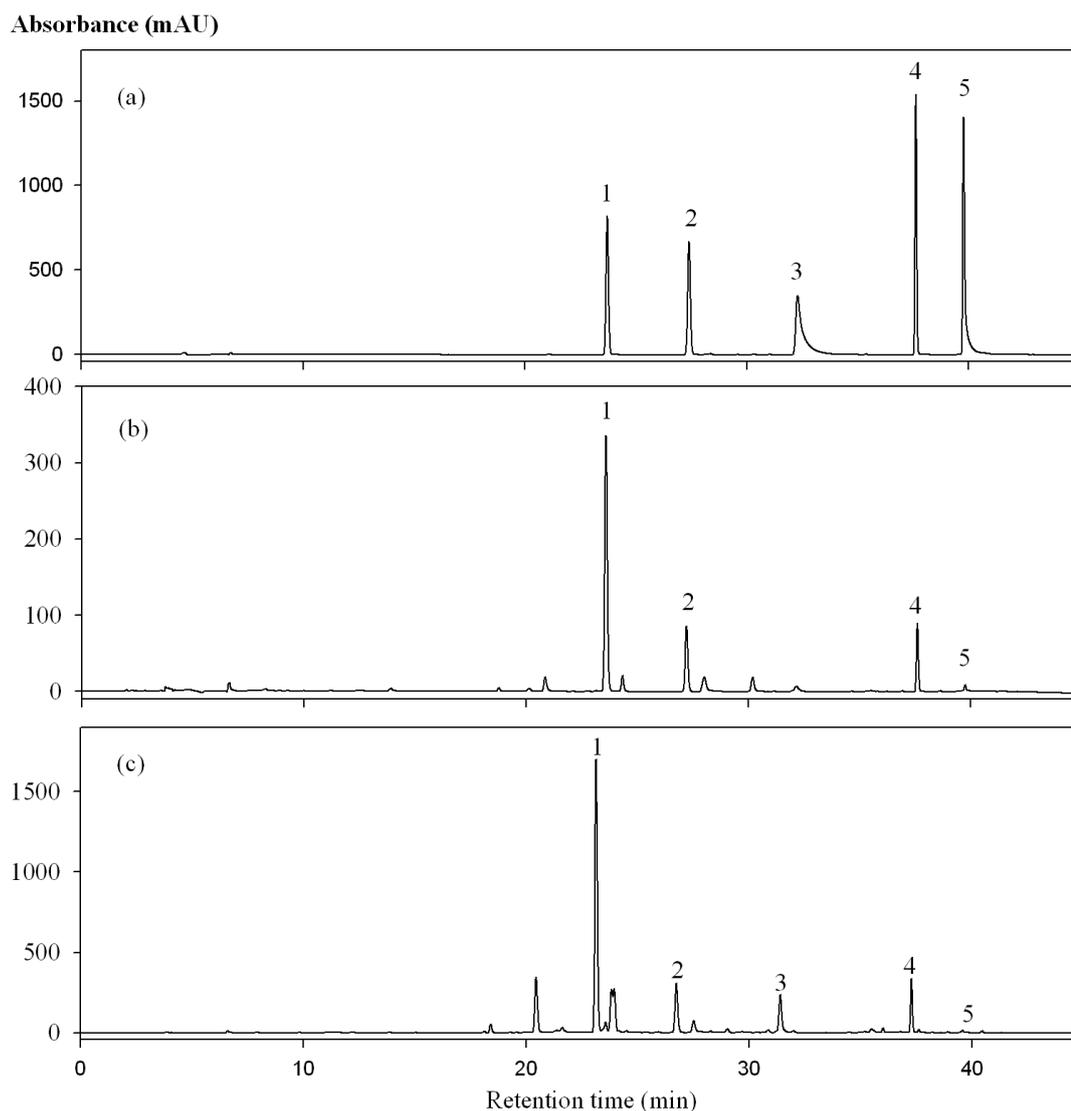


Fig. 4.6 HPLC/UV (254nm) chromatograms of the five studied iso-flavonoids components a) in standard solution, b) extract of *Pueraria thomsonii* (FG), c) extract of *Pueraria lobata* (YG): 1) puerarin; 2) daidzin; 3) genistin; 4) daidzein ; 5) genistein

4.5.3 Validation of the reference method

The limit of detection (LOD) and quantification (LOQ) under the chromatographic conditions were determined at signal-to-noise ratio (S/N) of 3 and 10, respectively, for each compound.

The intra-day variability was examined in five replicated injections within one day

and the inter-day repeatability was determined by repeating the analysis for three consecutive days. The relative standard derivation (RSD) was calculated with the formula:

$$\text{R.S.D. (\%)} = \frac{\text{S.D.}}{\text{Mean}} \times 100\%$$

To examine the accuracy of the extraction method, the recovery test was performed. Accurate amounts of the five standard isoflavonoids were added to 0.2g dried FG and YG powder samples respectively, and then extracted and analyzed as described in Section 4.4.3. The recoveries were calculated by the formula:

$$\text{Recovery (\%)} = \frac{(\text{Amount found} - \text{original amount})}{\text{Amount spiked}} \times 100\%$$

The HPLC method developed was validated by linearity, precision and accuracy according to the guideline in International Conference on Harmonization (ICH) and the results were shown in Table 4.2. All the calibration curves of the five standards exhibited good linearity ($r^2 > 0.999$) within the test range. The LOD and LOQ were in the range of 0.2-1.9ng and 0.7-15ng, respectively. The developed analytical method had good recovery of $100\% \pm 7.5\%$ ($n=3$) for all five analytes concerned in YG and FG samples and the relative recovery for all analytes are larger than 95%. In addition, the intra-day and inter-day variability of puerarin, daidzin in herbal extracts are less than 5%. All the results suggested that the HPLC assay is reproducible.

Components	r^2	Test range (ng)	Recovery in	Recovery in	LOQ (ng)	LOD (ng)
			YG (%) (n=3)	FG (%) (n=3)		
Puerarin	0.999	10 – 6000	98.3 ± 0.08	102.1 ± 1.67	1.7	0.5
Daidzin	0.999	5 – 4000	99.8 ± 1.05	107.5 ± 1.78	6.4	1.9
Daidzein	1.000	10 – 1000	102.0 ± 0.66	106.3 ± 0.81	15	0.5
Genistin	0.999	5 – 1000	107.3 ± 4.50	109.4 ± 3.00	3.1	0.9
Genistein	0.999	5 – 200	101.6 ± 3.04	102.4 ± 1.80	0.7	0.2

Table 4.2 HPLC method validation parameters for the five isoflavonoids

4.5.4 Quantitative analysis of YG and FG

To determine the actual contents of the isoflavonoids present in the *Puerariae Radix* samples, calibration of the five isoflavonoids were prepared. Standard solutions of five isoflavonoids were prepared and diluted to appropriate concentrations. A range of six concentrations of the five isoflavonoids were analyzed by four replicate injections. The calibration curves were constructed by plotting the peak areas versus the actual amount (ng) of each isoflavonoid. The contents were calculated from the equation of the calibration curve. The concentration range and the retention times of the corresponding isoflavonoids were stated in the table 4.3:

Components	Retention Time	Calibration curve*	r^2	Test range (ppm)
Puerarin	23.786	$y = 35.24x + 35.85$	0.9997	1- 600
Daidzin	27.481	$y = 31.76x + 20.52$	0.9999	0.5 - 400
Daidzein	37.659	$y = 65.16x + 8.94$	1	1-100
Genistin	32.387	$y = 42.64x + 6.70$	0.9998	0.5-100
Genistein	39.797	$y = 52.86x - 12.47$	0.999	0.5-20
y: Peak Area (mAU), x: Concentration (ppm)				

Table 4.3 Calibration curves for the five isoflavonoids in Puerariae Radix

Five isoflavonoids in eighty-nine samples of Puerariae Radix were quantitatively determined by the developed HPLC-DAD method. It was observed that the entire HPLC chromatographic patterns of these two species were similar except that more peaks were detected in YG samples. However, as shown in Fig. 4.7, it is worth noticing that the actual contents of all the five isoflavonoids found in YG are far more than that in FG.

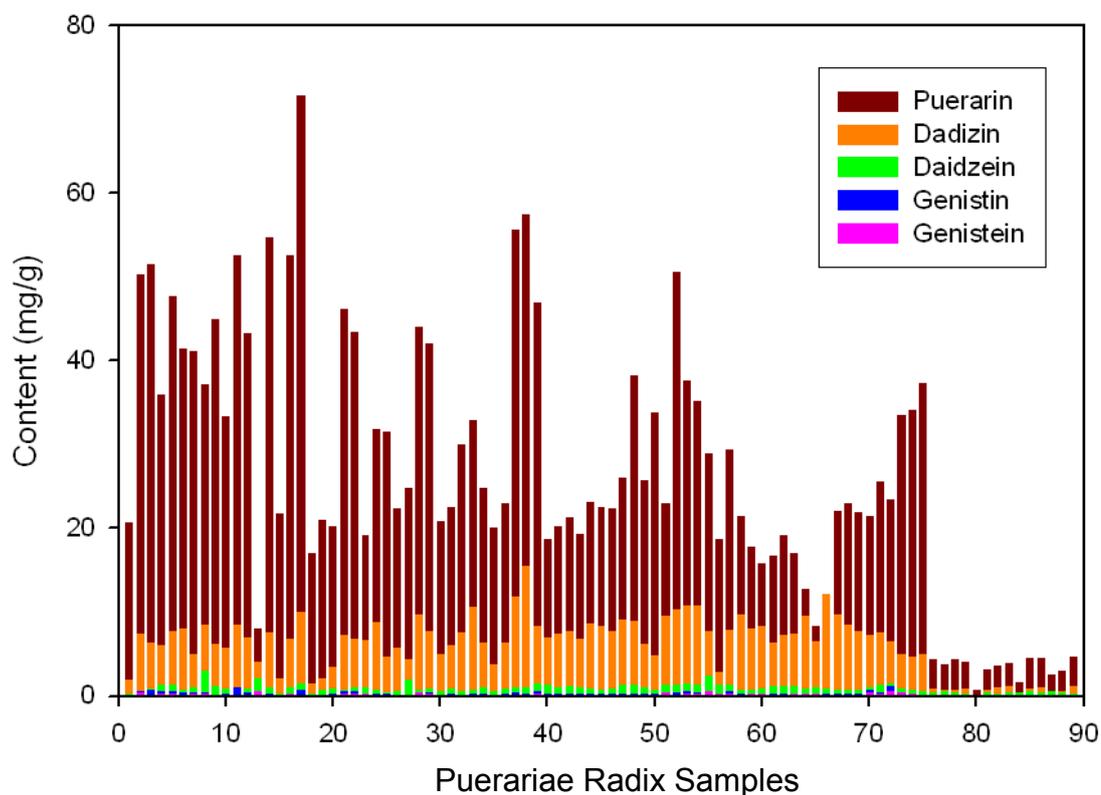


Fig. 4.7 Histogram of the contents of the five isoflavonoids in Puerariae Radix obtained from HPLC (Sample 1-75, 76-89 are YG and FG respectively)

From the Table 4.4, the puerarin content of YG is nearly ten times of that in FG. Thus, YG and FG are listed as separated entry in Chinese Pharmacopoeia since 2005²⁰³. Based on the standard derived from Chinese Pharmacopoeia, in which the puerarin content in YG and FG should be higher than 2.4% and 0.3%, respectively. Base on these criteria, seventy-five samples collected were classified as YG and fourteen samples were categorized as FG.

Sample	Content (mg/g)					
	Puerarin	Daidzin	Genistein	Daidzein	Genistin	Total iso-flavonoids
<i>Pueraria lobata</i> (n=75)	30.42	7.15	0.29	0.96	0.14	38.96
<i>Pueraria thomsonii</i> (n=14)	3.46	0.71	0	0.34	0.05	4.57

Table 4.4 The mean contents of the five isoflavonoids in *Puerariae Radix* obtained from HPLC

Table 4.5 listed the content of the five isoflavonoids in the 89 *Puerariae Radix* samples. The RSD of puerarin and daidzin among all samples are almost less than 2% while, that of daidzein, genistein and genistin are almost less than 8%. Since puerarin and daidzin are the two most abundant components (>0.5%) with low RSD, therefore, we will only focus on puerarin, daidzin and the total isoflavonoids in the NIRS analysis.

	Puerarin (mg/g)	RSD (%)	Daidzin (mg/g)	RSD (%)	Daidzein (mg/g)	RSD (%)	Genistein (mg/g)	RSD (%)	Genistin (mg/g)	RSD (%)
YG-01	20.58 ± 0.22	1.08	1.92 ± 0.07	3.76	0.26 ± 0.02	7.89	0.10 ± 0.01	1.44	0.11 ± 0.00	3.38
YG-02	50.22 ± 0.52	1.04	7.42 ± 0.09	1.20	0.46 ± 0.02	4.88	0.37 ± 0.05	1.45	0.58 ± 0.02	3.28
YG-03	51.46 ± 0.48	0.93	6.38 ± 0.31	4.79	0.63 ± 0.24	3.79	0.14 ± 0.03	1.77	0.68 ± 0.06	3.22
YG-07	35.85 ± 0.50	1.39	6.04 ± 0.12	1.93	1.28 ± 0.01	1.14	0.15 ± 0.00	1.93	0.48 ± 0.02	4.23
YG-08	47.61 ± 0.94	1.98	7.68 ± 0.16	2.14	1.27 ± 0.06	4.78	0.20 ± 0.02	9.77	0.46 ± 0.02	4.30
YG-09	41.40 ± 0.25	0.61	8.07 ± 0.07	0.83	0.63 ± 0.00	0.68	0.08 ± 0.01	9.00	0.32 ± 0.00	1.26
YG-10	41.17 ± 0.65	1.58	4.95 ± 0.17	3.52	0.97 ± 0.14	1.46	0.15 ± 0.02	1.40	0.36 ± 0.01	2.38
YG-11	37.11 ± 0.30	0.82	8.45 ± 0.11	1.26	2.98 ± 0.02	0.77	0.26 ± 0.04	1.54	0.31 ± 0.01	2.75
YG-12	44.96 ± 0.69	1.54	6.10 ± 0.11	1.77	1.10 ± 0.02	2.00	0.10 ± 0.00	1.73	0.13 ± 0.01	4.43
YG-14	33.33 ± 0.44	1.31	5.67 ± 0.12	0.02	0.80 ± 0.21	2.68	0.09 ± 0.03	3.83	0.29 ± 0.02	5.98
YG-15	0.04 ± 0.00	2.16	0.64 ± 0.01	0.02	0.33 ± 0.00	1.20	0.23 ± 0.00	0.95	0.43 ± 0.02	3.58
YG-16	52.56 ± 0.60	1.14	8.52 ± 0.09	0.01	0.78 ± 0.08	1.02	0.08 ± 0.00	3.29	0.98 ± 0.07	7.25
YG-17	43.17 ± 0.34	0.78	6.93 ± 0.07	0.01	0.83 ± 0.01	1.64	0.07 ± 0.00	4.69	0.39 ± 0.00	1.20
YG-18	7.94 ± 0.06	0.79	4.04 ± 0.05	1.14	2.06 ± 0.03	1.58	0.49 ± 0.02	5.01	0.40 ± 0.01	2.96
YG-19	54.73 ± 0.09	0.17	7.57 ± 0.13	1.66	1.01 ± 0.09	9.09	0.11 ± 0.01	8.32	0.22 ± 0.02	4.57
YG-20	21.70 ± 0.07	0.32	1.99 ± 0.05	2.57	0.25 ± 0.03	12.44	0.06 ± 0.01	10.92	0.07 ± 0.01	3.66
YG-21	52.58 ± 0.43	0.82	6.85 ± 0.09	1.26	0.96 ± 0.02	1.63	0.18 ± 0.00	2.38	0.18 ± 0.01	3.63
YG-22	71.62 ± 0.35	0.48	9.94 ± 0.08	0.81	1.41 ± 0.01	0.44	0.11 ± 0.01	9.98	0.63 ± 0.02	3.98
YG-23	17.06 ± 0.14	0.85	1.42 ± 0.03	1.84	0.23 ± 0.02	7.00	0.06 ± 0.01	11.84	0.08 ± 0.01	7.38
YG-24	20.99 ± 0.07	0.34	1.97 ± 0.07	3.31	0.68 ± 0.01	1.85	0.06 ± 0.01	7.56	0.11 ± 0.01	10.16
YG-25	20.18 ± 0.16	0.81	3.49 ± 0.08	2.31	0.77 ± 0.04	5.21	0.10 ± 0.00	0.69	0.17 ± 0.00	1.87
YG-26	46.12 ± 0.34	0.75	7.29 ± 0.06	0.80	0.67 ± 0.02	2.83	0.23 ± 0.02	9.77	0.46 ± 0.01	2.26

YG-27	43.39 ± 0.43	1.00	6.72 ± 0.13	1.92	0.92 ± 0.02	2.20	0.21 ± 0.02	8.93	0.44 ± 0.01	1.89
YG-28	19.18 ± 0.02	0.13	6.62 ± 0.09	1.34	0.90 ± 0.08	8.41	0.17 ± 0.01	7.57	0.24 ± 0.02	6.72
YG-29	31.81 ± 0.27	0.86	8.74 ± 0.12	1.41	0.60 ± 0.02	3.57	0.11 ± 0.02	7.16	0.15 ± 0.01	6.79
YG-30	31.54 ± 0.27	0.87	4.58 ± 0.05	1.11	0.42 ± 0.00	1.05	0.07 ± 0.00	5.08	0.21 ± 0.00	2.14
YG-31	22.32 ± 0.21	0.93	5.76 ± 0.07	1.20	0.49 ± 0.05	9.90	0.06 ± 0.00	5.68	0.07 ± 0.00	3.20
YG-32	24.71 ± 0.29	1.17	4.28 ± 0.05	1.25	1.90 ± 0.02	0.96	0.08 ± 0.00	5.63	0.07 ± 0.00	6.85
YG-33	44.05 ± 0.38	0.86	9.64 ± 0.10	1.02	0.68 ± 0.01	1.69	0.31 ± 0.03	8.43	0.41 ± 0.01	2.31
YG-34	41.97 ± 0.30	0.71	7.68 ± 0.13	1.64	0.78 ± 0.06	8.52	0.16 ± 0.01	6.64	0.32 ± 0.02	7.16
YG-35	20.77 ± 0.21	1.03	4.91 ± 0.06	1.26	0.55 ± 0.01	1.10	0.03 ± 0.00	5.47	0.08 ± 0.00	1.67
YG-36	22.52 ± 0.24	1.05	6.05 ± 0.04	0.61	0.80 ± 0.02	2.38	0.05 ± 0.00	1.15	0.15 ± 0.02	9.77
YG-37	29.95 ± 0.39	1.30	7.54 ± 0.13	1.68	0.54 ± 0.03	5.46	0.09 ± 0.01	14.30	0.13 ± 0.01	5.69
YG-38	32.91 ± 0.43	1.30	10.52 ± 0.16	1.52	0.73 ± 0.01	1.49	0.03 ± 0.00	5.58	0.28 ± 0.01	2.84
YG-39	24.81 ± 0.24	0.98	6.25 ± 0.08	1.36	0.95 ± 0.01	0.96	0.05 ± 0.00	2.24	0.18 ± 0.01	7.02
YG-40	19.99 ± 0.29	1.47	3.71 ± 0.08	2.09	0.51 ± 0.01	1.35	0.12 ± 0.00	2.23	0.07 ± 0.01	11.88
YG-41	23.00 ± 0.33	1.45	6.26 ± 0.11	1.79	0.84 ± 0.02	1.94	0.06 ± 0.00	2.44	0.22 ± 0.01	5.80
YG-42	55.57 ± 0.49	0.88	11.77 ± 0.13	1.13	0.96 ± 0.01	1.38	0.03 ± 0.00	5.09	0.30 ± 0.02	6.43
YG-43	57.44 ± 0.64	1.12	15.5 ± 0.23	1.46	1.01 ± 0.02	1.51	0.04 ± 0.00	9.51	0.30 ± 0.01	4.95
YG-44	46.85 ± 0.12	0.25	8.3 ± 0.06	0.71	1.39 ± 0.01	0.61	0.15 ± 0.02	10.96	0.55 ± 0.08	14.69
YG-45	18.66 ± 0.04	0.24	6.86 ± 0.09	1.33	1.24 ± 0.03	2.22	0.07 ± 0.01	8.43	0.25 ± 0.01	3.85
YG-46	20.23 ± 0.30	1.47	7.32 ± 0.19	2.60	0.98 ± 0.02	1.59	0.05 ± 0.00	5.72	0.18 ± 0.01	7.16
YG-47	21.27 ± 0.06	0.28	7.64 ± 0.08	0.98	1.13 ± 0.05	3.99	0.10 ± 0.02	16.29	0.16 ± 0.00	2.42
YG-48	19.25 ± 0.28	1.48	6.7 ± 0.16	2.34	1.05 ± 0.02	1.61	0.02 ± 0.00	2.75	0.21 ± 0.00	1.80
YG-49	23.03 ± 0.03	0.13	8.59 ± 0.12	1.42	0.79 ± 0.04	5.44	0.07 ± 0.01	19.35	0.24 ± 0.01	5.43

YG-50	22.55 ± 0.26	1.17	8.33 ± 0.14	1.65	0.72 ± 0.01	1.22	0.05 ± 0.00	1.65	0.23 ± 0.01	2.36
YG-51	22.34 ± 0.24	1.07	7.66 ± 0.10	1.27	0.78 ± 0.02	2.10	0.12 ± 0.01	9.96	0.21 ± 0.00	2.10
YG-52	26.01 ± 0.34	1.30	9.02 ± 0.19	2.10	1.33 ± 0.12	9.16	0.12 ± 0.00	3.56	0.22 ± 0.00	2.08
YG-53	38.24 ± 0.44	1.15	8.87 ± 0.16	1.84	1.23 ± 0.02	1.92	0.09 ± 0.01	12.09	0.26 ± 0.02	6.43
YG-54	25.64 ± 0.02	0.09	6.0 ± 0.08	1.38	1.04 ± 0.01	0.86	0.07 ± 0.00	3.30	0.17 ± 0.00	0.77
YG-55	33.70 ± 0.05	0.14	4.85 ± 0.05	1.07	0.70 ± 0.01	2.07	0.07 ± 0.01	9.34	0.15 ± 0.01	6.52
YG-56	22.91 ± 0.01	0.06	9.51 ± 0.03	0.34	1.29 ± 0.01	0.62	0.33 ± 0.04	10.89	0.29 ± 0.01	3.00
YG-57	50.53 ± 0.98	1.94	10.33 ± 0.05	0.48	1.21 ± 0.01	1.18	0.09 ± 0.01	6.02	0.40 ± 0.02	5.32
YG-58	37.54 ± 0.53	1.40	10.81 ± 0.25	2.29	1.36 ± 0.09	6.27	0.19 ± 0.02	13.03	0.55 ± 0.05	9.55
YG-59	35.22 ± 0.23	0.67	10.69 ± 0.29	2.71	1.31 ± 0.07	5.09	0.17 ± 0.01	3.40	0.36 ± 0.03	7.25
YG-60	28.81 ± 0.06	0.20	7.77 ± 0.06	0.76	2.30 ± 0.03	1.49	0.47 ± 0.04	8.81	0.54 ± 0.04	8.13
YG-61	18.66 ± 0.07	0.38	2.73 ± 0.02	0.77	1.32 ± 0.13	9.75	0.18 ± 0.03	18.68	0.07 ± 0.01	12.60
YG-62	29.36 ± 0.48	1.64	7.86 ± 0.16	2.10	1.29 ± 0.05	3.61	0.26 ± 0.02	6.95	0.51 ± 0.05	9.91
YG-63	21.44 ± 0.37	1.84	9.69 ± 0.18	1.83	0.71 ± 0.05	6.27	0.09 ± 0.01	9.28	0.19 ± 0.01	4.59
YG-64	17.82 ± 0.34	1.90	8.05 ± 0.20	2.44	0.70 ± 0.01	1.99	0.14 ± 0.00	2.52	0.19 ± 0.00	1.28
YG-65	15.77 ± 0.04	0.28	8.26 ± 0.09	1.15	0.86 ± 0.01	0.64	0.16 ± 0.01	3.50	0.20 ± 0.01	5.86
YG-66	16.74 ± 0.16	0.97	6.37 ± 0.14	2.13	1.10 ± 0.06	5.33	0.08 ± 0.01	10.46	0.25 ± 0.03	9.88
YG-67	19.06 ± 0.08	0.43	7.26 ± 0.11	1.56	1.13 ± 0.05	4.31	0.10 ± 0.01	9.64	0.15 ± 0.01	6.03
YG-68	16.94 ± 0.25	1.50	7.33 ± 0.15	2.05	1.09 ± 0.01	1.32	0.15 ± 0.00	2.79	0.18 ± 0.00	1.19
YG-69	12.68 ± 0.19	1.48	9.50 ± 0.12	1.31	0.84 ± 0.05	6.37	0.20 ± 0.01	5.49	0.25 ± 0.01	5.81
YG-70	8.29 ± 0.13	1.56	6.50 ± 0.16	2.47	0.92 ± 0.01	1.28	0.15 ± 0.01	6.38	0.17 ± 0.01	5.58
YG-71	11.68 ± 0.13	1.13	12.18 ± 0.09	0.75	0.70 ± 0.01	1.71	0.13 ± 0.01	7.91	0.19 ± 0.01	2.86
YG-72	21.99 ± 0.19	0.84	9.68 ± 0.17	1.72	0.69 ± 0.01	1.48	0.04 ± 0.00	8.54	0.20 ± 0.00	2.08

YG-73	22.92 ± 0.31	1.35	8.48 ± 0.19	2.19	0.73 ± 0.03	3.84	0.05 ± 0.00	5.91	0.27 ± 0.02	5.80
YG-74	21.91 ± 0.29	1.31	7.70 ± 0.22	2.89	0.67 ± 0.01	1.90	0.05 ± 0.00	5.13	0.24 ± 0.01	2.44
YG-75	21.44 ± 0.13	0.60	7.20 ± 0.27	3.81	1.04 ± 0.07	7.06	0.38 ± 0.03	8.37	0.65 ± 0.01	1.16
YG-76	25.47 ± 0.35	1.38	7.56 ± 0.22	2.89	1.26 ± 0.06	4.88	0.25 ± 0.02	6.48	0.39 ± 0.04	9.06
YG-77	23.43 ± 0.31	1.33	6.41 ± 0.25	3.95	1.48 ± 0.04	2.75	0.56 ± 0.02	4.18	1.06 ± 0.04	3.80
YG-78	33.41 ± 0.45	1.35	4.99 ± 0.09	1.83	0.78 ± 0.01	1.03	0.30 ± 0.00	0.90	0.15 ± 0.01	3.88
YG-79	34.08 ± 0.18	0.54	4.71 ± 0.05	1.10	0.71 ± 0.60	7.84	0.25 ± 0.02	5.41	0.13 ± 0.02	10.40
YG-80	37.21 ± 0.18	0.48	4.97 ± 0.04	0.83	0.45 ± 0.00	0.86	0.08 ± 0.00	3.12	0.10 ± 0.00	3.67
FG-02	4.37 ± 0.01	0.25	0.83 ± 0.00	0.26	0.36 ± 0.00	0.25	0.05 ± 0.00	1.58		
FG-03	3.71 ± 0.04	0.99	0.75 ± 0.01	0.86	0.47 ± 0.00	0.84	0.07 ± 0.00	1.24		
FG-04	4.35 ± 0.06	1.31	0.66 ± 0.01	1.18	0.36 ± 0.00	1.15	0.04 ± 0.00	9.02		
FG-05	4.10 ± 0.06	1.36	0.86 ± 0.01	1.38	0.25 ± 0.00	1.35	0.04 ± 0.00	1.25		
FG-06	0.64 ± 0.00	0.50	0.03 ± 0.00	0.96	0.13 ± 0.00	0.49	0.04 ± 0.00	6.34		
FG-07	3.12 ± 0.05	1.73	0.65 ± 0.01	1.82	0.34 ± 0.01	1.68	0.06 ± 0.00	1.62		
FG-08	3.51 ± 0.05	1.49	0.92 ± 0.01	1.42	0.15 ± 0.00	1.45	0.03 ± 0.00	1.42		
FG-09	3.84 ± 0.02	0.58	1.18 ± 0.01	0.55	0.42 ± 0.00	0.59	0.06 ± 0.00	0.85		
FG-10	1.59 ± 0.02	1.18	0.20 ± 0.00	1.22	0.35 ± 0.00	1.12	0.05 ± 0.00	1.27		
FG-11	4.54 ± 0.07	1.55	0.87 ± 0.01	1.50	0.49 ± 0.01	1.42	0.08 ± 0.01	8.53		
FG-12	4.44 ± 0.05	1.21	0.92 ± 0.01	1.32	0.34 ± 0.00	1.18	0.05 ± 0.00	1.76		
YG-04	2.56 ± 0.01	0.36	0.47 ± 0.00	0.43	0.45 ± 0.00	0.57	0.08 ± 0.00	3.53		
YG-05	2.96 ± 0.03	0.84	0.53 ± 0.00	0.91	0.36 ± 0.00	0.95	0.06 ± 0.00	0.84		
YG-06	4.69 ± 0.08	1.65	1.13 ± 0.02	1.64	0.29 ± 0.00	1.63	0.06 ± 0.00	1.44		

Table 4.5 Contents of the five isoflavonoids in the 89 Puerariae Radix samples

4.5.5 Classification models development by NIRS

In this work, LDA and SIMCA are used to establish the classification models in the NIR analysis³³. LDA is a supervised pattern recognition technique and determines a linear function in multivariate space which maximizes the ratio between both inter-group variances compared to the intra-group variances. As LDA cannot deal with the large number of variables in the spectra, principal component analysis (PCA) was applied in order to reduce the number of variables. In this work, LDA was developed on a PCA sub-space which accounts for 95% variance of the data.

SIMCA is a supervised method for data classification that requires a training set with known attributes and class membership. It has the advantage of being able to handle collinear X-variables, missing data, noisy variables and overlapped classes. In this work, the model parameters (95%, $\alpha = 0.05$) were applied and the number of principal components used was determined by cross validation. The species of unknown samples in the validation set were determined by calculating the shortest distances between the centroids of two groups in the calibration model developed.

Fig 4.8 shows the mean spectra of the original NIR data of YG and FG. Visual inspection suggested that the NIR spectra of the two species are similar. The most intensive band is contributed by combination bands, which including C–H stretching, deformation vibration in CH₃ group (2280 nm), C–H deformation vibration in CH₂ group (2320 nm), and O–H stretching and deformation vibration in CH₃ group (2075 nm). These vibrational modes can be found in isoflavonoids as well as other constituents such as protein and organic acids.

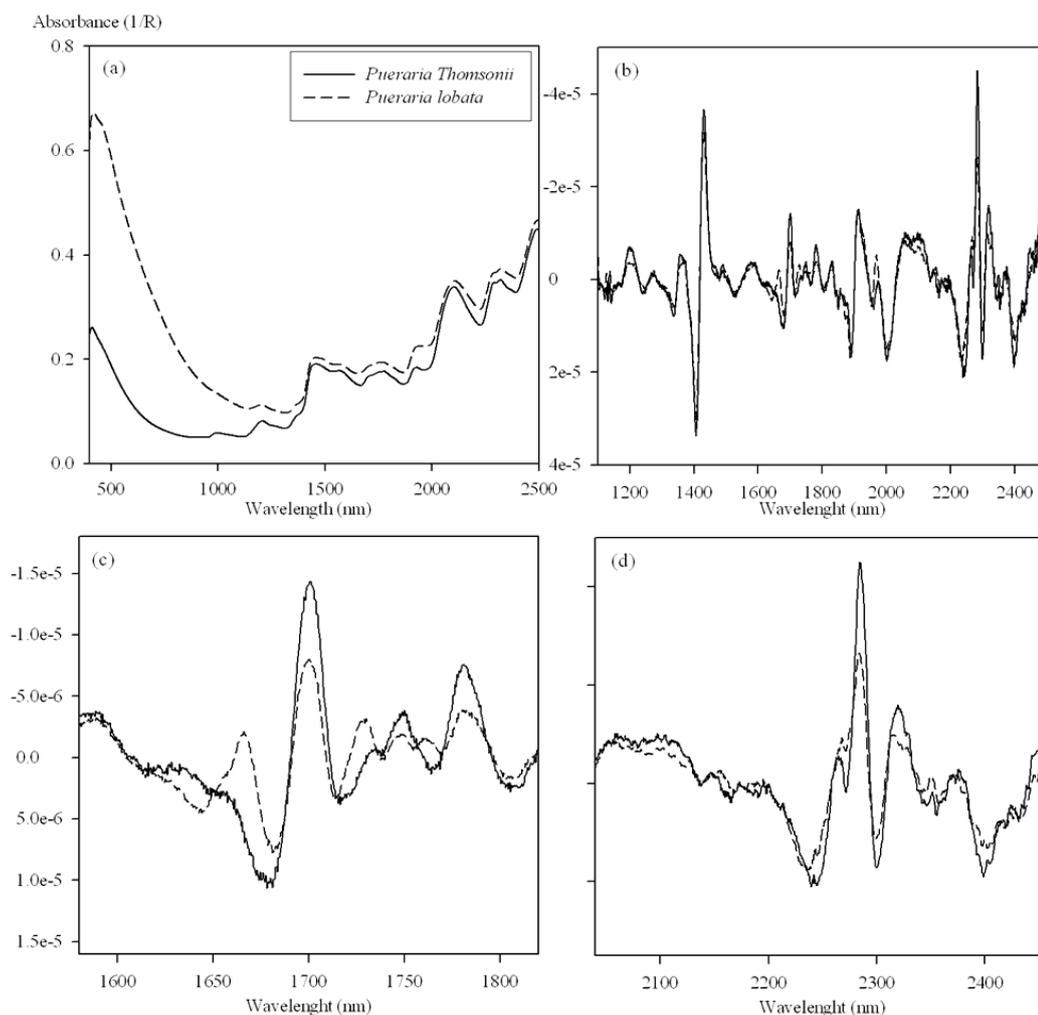


Fig. 4.8 Mean NIR spectra for *Pueraria thomsonii* (—) and *Pueraria lobata* (- - -) obtained from, a) raw data, b) with second derivative pre-treatment, c) Enlarged region between 1600nm and 1800nm with second derivative pre-treatment and d) Enlarged region between 2050nm and 2450 nm with second derivative pre-treatment

In order to build robust classification models, the samples were divided into three groups and a three-fold cross-validation was performed. Kennard–Stone (KS) algorithm was used to divide the samples into groups objectively. The advantage of KS algorithm is capable of creating a unique list of objects that is selected from the cluster borders in a multi-dimensional space. This enhances the discriminatory power, which is particularly important in developing model with overlapping clusters²²⁴. Fig 4.8 shows the mean NIR spectra of the two species and the second derivative of the

spectra in 1600–1800 nm and 2050–2450nm where the differences between two species are the most significant.

Classification models using the full spectrum and the two selected regions (1600–1800 nm and 2050–2450 nm) were established by applying LDA and SIMCA. The average percentage of the correct classification of the testing group is shown in Table 4.6. The average percentage of correct prediction from models using raw spectrum are significantly less than that of the models using second derivative. This suggests that the pre-processing step is important in establishing reliable classification models. Models based on the intensities from spectral regions among 1600–1800 nm and 2050–2450 nm are superior using either LDA or SIMCA methods. The models obtained provide a clear discrimination between these two species. The success in classification was possibly a result of the large differences in total isoflavonoid content between the two species.

For SIMCA, the sensitivity and specificity of classification models are tabulated in Table 4.7. Sensitivity is the proportion of samples belonging to a certain category, which are correctly identified by the model, while specificity is the proportion of samples not belonging to a certain class, which are correctly identified as foreign by the model. These two parameters indicated the extent of overlapping between classes. Fig. 4.9 depicts the Cooman's plot for the SIMCA distances obtained for YG and FG classes. A clear boundary and clear differentiation between two species was noticed by using 1600–1800 nm and 2050–2450 nm wavelength regions with second derivative pretreatment and the developed SIMCA model possessed a good recognition and prediction ability.

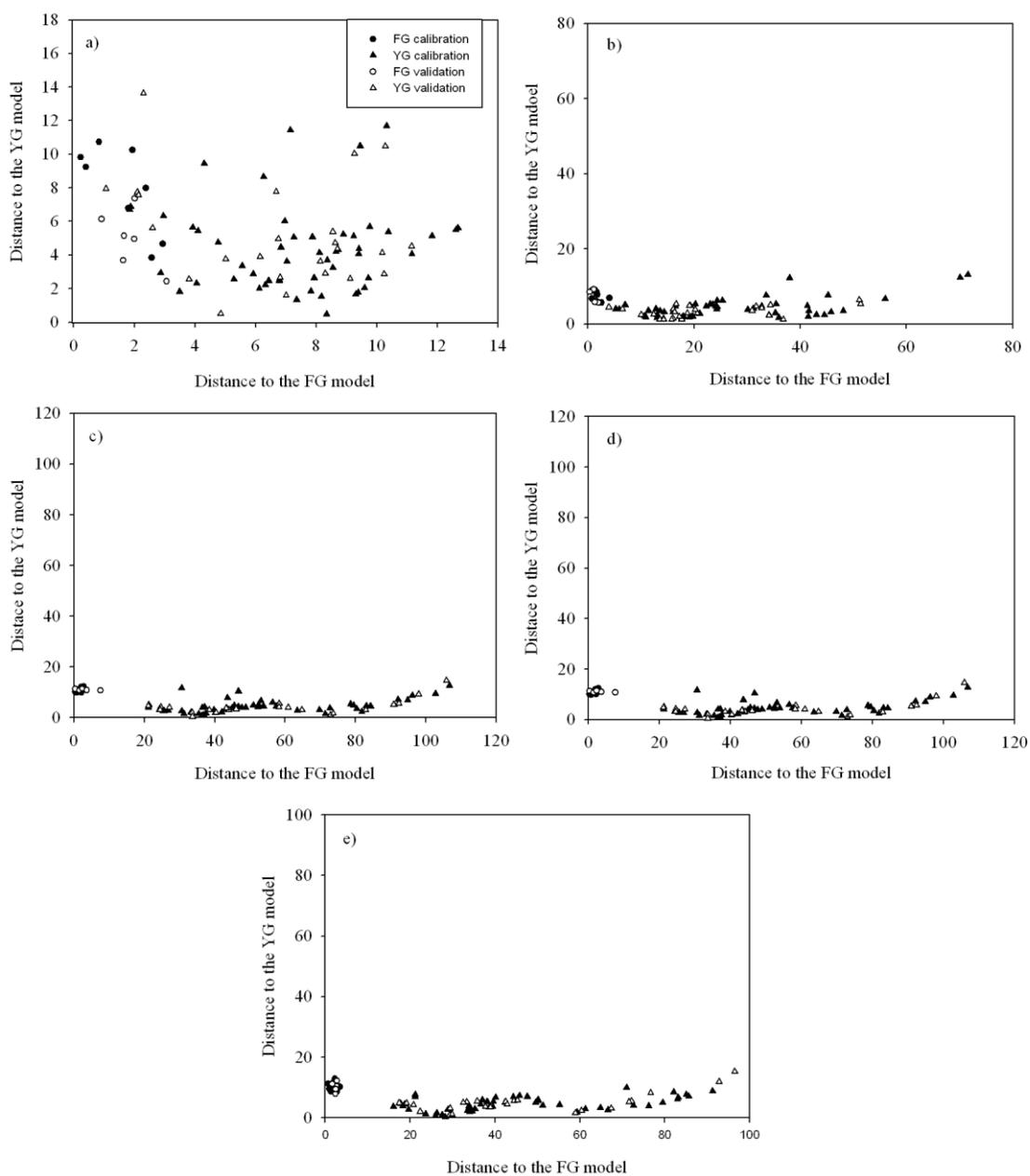


Fig. 4.9 Cooman's plot a) raw data, b) raw data with second derivative pretreatment, c) 1600–1800 nm with second derivative pretreatment, d) 2050–2450 nm with second derivative pretreatment and e) 1600–1800 nm and 2050–2450 nm with second derivative pretreatment for the square SIMCA distances obtained for FG and YG models

		No. of Principal Components used	Successful rate (Calibration)*	Successful rate (Validation)*
LDA [#]	1100-2500nm	1	73.6%	71.8%
	1100-2500nm with 2 nd derivative	2	93.8%	94.3%
	1600-1800nm with 2 nd derivative	2	100%	98.9%
	2050-2450nm with 2 nd derivative	2	98.9%	98.9%
	1600-1800, 2050-2450 with 2 nd derivative	2	98.9%	98.9%
SIMCA [#]	1100-2500nm	1	82.6%	83.1%
	1100-2500nm with 2 nd derivative	2	97.8%	98.9%
	1600-1800nm with 2 nd derivative	2	100%	100%
	2050-2450nm with 2 nd derivative	2	100%	100%
	1600-1800, 2050-2450 with 2 nd derivative	2	100%	100%

*The percent classification for the training set objects has been computed considering the average over the values obtained for each cross-validation run.

For LDA, model is developed by taking the PCA scores when total percentage variance accounted is over 95%; for SIMCA, the normal models were applied and the stop criterion used for the model being developed was determined by cross validation

Table 4.6 Results of the classification of Puerariae Radix using LDA and SIMCA

SIMCA	Sensitivity		Specificity	
	<i>Pueraria lobata</i>	<i>Pueraria thomsonii</i>	<i>Pueraria lobata</i>	<i>Pueraria thomsonii</i>
Spectral region selected				
1100-2500nm	84.0% (83.6%)	92.8% (74.4%)	69.8% (39.4%)	91.3% (83.6%)
1100-2500nm with 2 nd derivative	88.0% (88.0%)	93.3% (83.3%)	87.2% (86.7%)	97.3% (100%)
1600-1800 with 2 nd derivative	84.7% (84.2%)	76.6% (77.8%)	100% (100%)	100% (100%)
2050-2450 with 2 nd derivative	80.7% (82.4%)	90.24% (57.78%)	100% (100%)	100% (100%)
1600-1800, 2050-2450 with 2 nd derivative	88.0% (88.0%)	93.26% (83.33%)	100% (100%)	100% (100%)

Values in bracket are the results for the test set

Table 4.7 Sensitivity and specificity of different SIMCA models

4.5.6 Quantitation models development by NIRS

Partial Least Squares PLS was used to establish the regression models, 55% of the samples were used as the calibration set and 25% and 20% of the samples were utilized for validation and test sets, respectively. In here, as many samples were collected, some samples can be spared as test set to evaluate the performance of the model and the spectra were not used in the model development. To avoid bias to a particular concentration range in forming the sample sets, the division was performed using the Kennard-Stone (KS) algorithm. The entire concentration range of puerarin, daidzin and total iso-flavonoids were covered in all the calibration, validation and test sets.

The Leave-One-Out Cross-Validation was used to establish the PLS regression model utilizing the training set to measure how well the calibration models with different factors fit. The spectrum of a sample was removed from the calibration set and a PLS regression model was built with the remaining spectra in the calibration set. Puerarin, daidzin and total iso-flavonoids contents of the one left-out sample was predicted by the PLS model obtained from the rest of the samples. The same procedure was repeated by leaving out each of the sample of the calibration set. The number of PLS factors used in the models were determined by minimizing the Predicted Residual Error Sum Square (PRESS) value. PRESS is the sum of square of deviation between the predicted and reference values of all the samples in the validation set. The final correlation model is the one with the lowest PRESS developed so far. The PRESS was calculated as follows,

$$\text{PRESS} = \sum_{i=1}^{n_v} (\hat{c}_i - c_i)^2 \quad (\text{Eq. 4.1})$$

where c_i is reference measurement result for sample i , \hat{c}_i is estimated value of the sample i when the model is constructed with sample i removed, n_c is the number of calibration samples, n_v is the number of validation samples.

Once the final calibration model was developed, the validity of the Partial Least Squares PLS regression model was evaluated by The Root Mean Square Error of Cross-Validation (RMSECV) representing the prediction error of the one left-out sample by the calibration model developed by Leave-One-Out Cross-Validation and Root Mean Square Error of Validation (RMSEV) representing the prediction error of the samples in validation set, The RMSECV and RMSEV were calculated as follow,

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n_c} (\hat{c}_i - c_i)^2}{n_c}} \quad (\text{Eq. 4.2})$$

$$\text{RMSEV} = \sqrt{\frac{\sum_{i=1}^{n_v} (\hat{c}_i - c_i)^2}{n_v}} \quad (\text{Eq. 4.3})$$

Generally speaking, the RMSECV is usually slightly smaller than RMSEV as the validation set only used to determine the number of PLS factor used.

To evaluate the performance of the model developed in predicting unknown samples in the test set, the Root Mean Square Error of Prediction (RMSEP), BIAS, and the Standard Error of Prediction (SEP) were used and calculated from the prediction error of the samples in test set. The RMSEP, BIAS and SEP were calculated as follows,

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n_t} (\hat{c}_i - c_i)^2}{n_t}} \quad (\text{Eq. 4.4})$$

$$\text{BIAS} = \frac{\sum_{i=1}^{n_t} (\hat{c}_i - c_i)}{n_t} \quad (\text{Eq. 4.5})$$

$$\text{SEP} = \sqrt{\frac{\sum_{i=1}^{n_t} (\hat{c}_i - c_i - \text{BIAS})^2}{n_t}} \quad (\text{Eq. 4.6})$$

where n_t is the number of test samples, \hat{c}_i is the predicted value from the model for test sample i and c_i is the reference measurement result for test sample i .

Both RMSEP and SEP estimate the precision and accuracy of the developed NIRS model. The SEP is always smaller than the RMSEP because the bias term has been subtracted. When the bias becomes insignificant, the RMSEP approaches SEP with increasing sample size. Generally, RMSEP gives a more realistic estimate of the prediction capability of a calibration model. RMSEP should be normally larger than RMSECV and RMSEV, since the test set samples are independent from the model development.

Finally, the correlation coefficient (R) of the predicted and measured values were calculated for the calibration, validation and test set. The value of R lies in the range of 0 to 1, 1 means perfect fitting. R is given by,

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (\hat{c}_i - c_i)^2}{\sum_{i=1}^n (c_i - \bar{c})^2}} \quad (\text{Eq. 4.7})$$

Where \bar{c} is the mean of the reference measurement results for n samples in the calibration, validation and test sets.

The samples with maximum and minimum contents of puerarin, daidzin and total iso-flavonoids content were arranged in the calibration set while the remaining samples were split into three groups, calibration, validation and test sets, by Kennard-Stone (KS) algorithm. For the KS algorithm, the mean and standard deviation of the puerarin, daidzin and the total iso-flavonoids content in the three sets were maintained as close as possible, which ensure a similar prediction ability of the model over the entire calibration range. As seen from Table 4.8, the reference values of puerarin, daidzin and total iso-flavonoids content spanned over the entire range in both sets, therefore the distribution of the samples is even in the calibration, validation and test sets. The total iso-flavonoids content in here is defined as the total amounts of all the five iso-flavonoids standards that can be found in HPLC chromatogram.

	Samples	Puerarin (mg/g)	Mean	SD
Calibration set	49	0.64 – 71.62	24.00	14.65
Validation set	22	2.56 – 57.44	28.42	18.70
Test set	18	4.69 – 55.57	29.36	14.30
	Samples	Daidzin (mg/g)	Mean	SD
Calibration set	49	0.03 – 15.50	6.24	3.19
Validation set	22	0.47 – 12.18	5.52	3.62
Test set	18	1.13 – 11.77	6.62	3.13
	Samples	Total isoflavonoids (mg/g)	Mean	SD
Calibration set	49	0.84 – 83.72	31.46	17.15
Validation set	22	3.56 – 74.29	35.10	22.01
Test set	18	6.16 – 68.62	37.33	16.59

Table 4.8 Statistics of the isoflavonoids contents of the samples determined by the HPLC in the calibration, validation and test sets

PLS models for the quantitative analysis of puerarin, daidzin and total iso-flavonoids content was established in using the NIR spectra. Spectral pre-processing methods and the number of PLS factors are critical parameters for accurate models. The optimum number of factors is determined by the lowest PRESS value. A practical rule for deciding the number of PLS components to be retained is not to include the additional components unless that improve the PRESS by at least 2%²²⁵. Including more PLS factors in the model will better fit the calibration set, but rupture the prediction for other samples. This phenomenon is called ‘over-fitting’ of a model (shown in Fig 4.10). Specific information related to the calibration samples is included in the model, which may deteriorate the prediction results of the samples not in the training set.

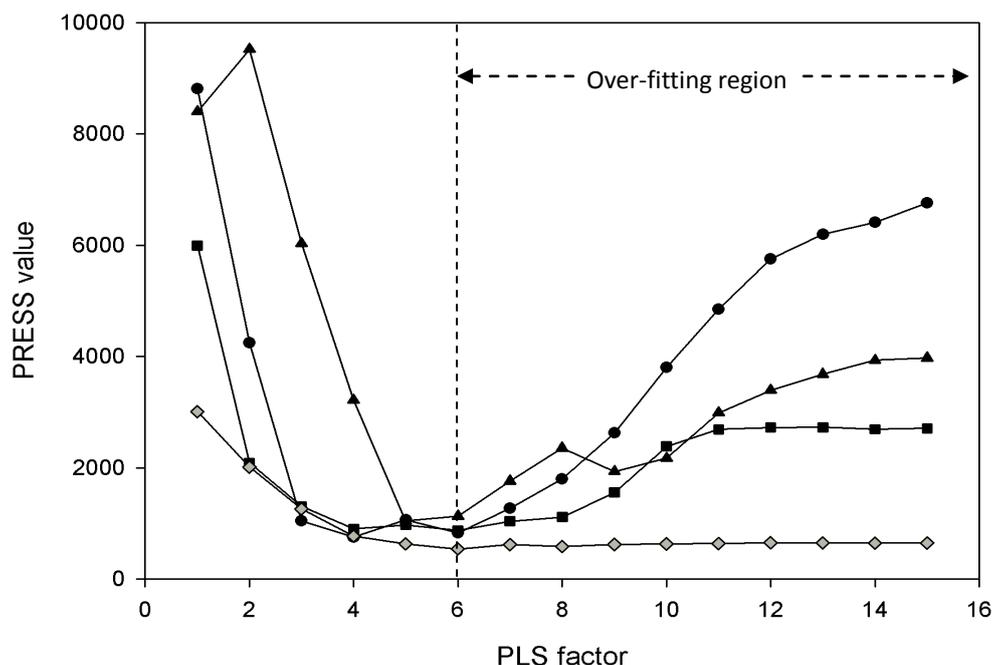


Fig. 4.10 A general PRESS against PLS factor graph showing the 'over-fitting' region

Puerarin model

Table 4.9 lists the RMSECV, RMSEV, RMSEP, BIAS, SEP and the correlation coefficient of the puerarin models developed by applying different pre-processing methods. In each pre-processing method only model of best performance with the lowest RMSECV value was shown. The model pretreated with SNV, detrend, second derivative (10,0) and S.G. (n=9,quintin) of whole NIR spectra (1100 – 2500nm) give the lowest RMSEV value among all other pre-processing methods (5 PLS factors). In fact, derivation of spectra with smoothing treatment enhances small differences between spectra and splits overlapping band. Also, MSC and detrend were used to correct the variations in light scattering due to different particle size distribution and packing density by adjusting the spectra based on intensities of spectra in regions carrying no specific chemical information¹⁴¹. This could be a reason explaining the good correlation results of puerarin content. Besides, as shown in Table 4.10, the model developed via second derivative pre-treatment of the entire spectral region gave

the lowest RMSEV value. However, it was not selected because of excessive number of PCA components used in the model, which is usually a signal of over-fitting.

Fig. 4.11 is a scatter plot showing the correlation between NIR predicted value and reference HPLC measurement of puerarin after treatment with SNV, detrend, second derivative and Savitsky-golay. Dot, circle and cross represent calibration, validation and test data points, respectively. The calibration, validation and test data have good correlation with the reference data and data point mainly fall on or close to the unity line. Puerarin content in the test set is predicted with a RMSEP value of 5.196. The correlation coefficients of this puerarin model for the calibration, validation and test sets are 0.977, 0.979 and 0.970, respectively.

Pretreatments	Factor	RMSECV	RMSEV	RMSEP	BIAS	SEP	R (calibration)	R (validation)	R (test)
No treatment	2	8.61	18.12	16.65	-1.26	16.60	0.805	0.479	0.384
2 nd derivative (10,0)	15	0.80	4.47	5.06	0.39	5.05	0.998	0.975	0.972
Detrend	6	5.37	6.93	7.26	-2.23	6.91	0.929	0.933	0.94
SNV, 2 nd derivative (10,0)	8	2.07	5.65	6.57	-0.24	6.57	0.990	0.954	0.944
MSC, 2 nd derivative (10,0)	8	2.06	5.60	6.53	-0.32	6.52	0.990	0.955	0.945
MSC, 2 nd derivative (10,0), S.G. (9,4)	5	3.10	4.00	5.25	-1.75	4.95	0.977	0.978	0.970
SNV, 2 nd derivative (10,0), S.G. (9,4)	5	3.11	3.96	5.22	-1.68	4.95	0.977	0.979	0.970
SNV, detrend, 2nd derivative (10,0), S.G. (9,4)	5	3.11	3.95	5.12	-1.65	4.93	0.977	0.979	0.970

Results from models with the best performance are marked in bold

Table 4.9 a) Performance of different PLS regression models for puerarin

Pretreatments	Factor	RMSECV	RMSEV	RMSEP	BIAS	SEP	R (calibration)	R (validation)	R (test)
No treatment	5	1.62	1.83	2.22	-0.92	2.02	0.871	0.826	0.887
MSC	3	1.85	1.94	2.31	-1.03	2.07	0.83	0.789	0.856
SNV	3	1.84	1.94	2.07	-0.87	2.25	0.831	0.79	0.851
2 nd derivative (10,0),	5	1.23	1.27	1.80	-0.65	1.68	0.928	0.916	0.911
3 rd derivative (10,0),	4	1.22	1.16	1.80	-0.56	1.71	0.929	0.931	0.909
2 nd derivative (10,0), S.G. (9,4)	9	0.42	1.16	1.51	0.62	1.38	0.992	0.929	0.937
3rd derivative (10,0), S.G. (7,3)	4	1.01	1.38	1.64	0.18	1.63	0.952	0.902	0.912
SNV, detrend, 2 nd derivative (10,0), S.G. (9,4)	4	1.22	1.28	1.77	-0.53	1.69	0.929	0.915	0.912

Results from models with the best performance are marked in bold

Table 4.9 b) Performance of different PLS regression models for daidzin

Pretreatments	Factor	RMSECV	RMSEV	RMSEP	BIAS	SEP	R (calibration)	R (validation)	R (test)
No treatment	3	9.38	21.48	20.25	-0.90	20.23	0.834	0.487	0.349
Detrend	8	4.47	6.63	9.49	-4.63	8.28	0.96	0.963	0.934
Detrend, 2 nd derivative (10,0)	7	2.63	6.53	7.51	-0.46	7.50	0.988	0.956	0.947
SNV, 2 nd derivative (10,0)	8	1.97	6.62	7.45	-0.82	7.41	0.993	0.955	0.948
MSC, 2 nd derivative (10,0)	8	1.96	6.61	7.47	-0.88	7.14	0.993	0.955	0.948
MSC, 2 nd derivative (10,0), S.G. (9,4)	8	1.96	6.61	7.47	-0.88	7.42	0.993	0.955	0.948
SNV, 2 nd derivative (10,0), S.G. (9,4)	5	3.28	4.90	6.46	-2.60	5.91	0.981	0.977	0.969
SNV, detrend, 2nd derivative (10,0), S.G. (9,4)	5	3.28	4.89	6.42	-2.56	5.89	0.981	0.977	0.969

Results from models with the best performance are marked in bold

MSC, Multiplicative Scatter Correction; SG, Savitsky-golay (data point, polynomial order); SNV, Standard Normal Variate transformation; derivative (segment, gap); RMSECV, Root Mean of Square Error of Cross Validation; RMSEV, Root Mean of Square Error of Validation; RMSEP, Root Mean of Square Error of Prediction; BIAS, Bias; SEP, Standard Error of Prediction; R, correlation coefficient.

Table 4.9 c) Performance of different PLS regression models for total isoflavonoids content

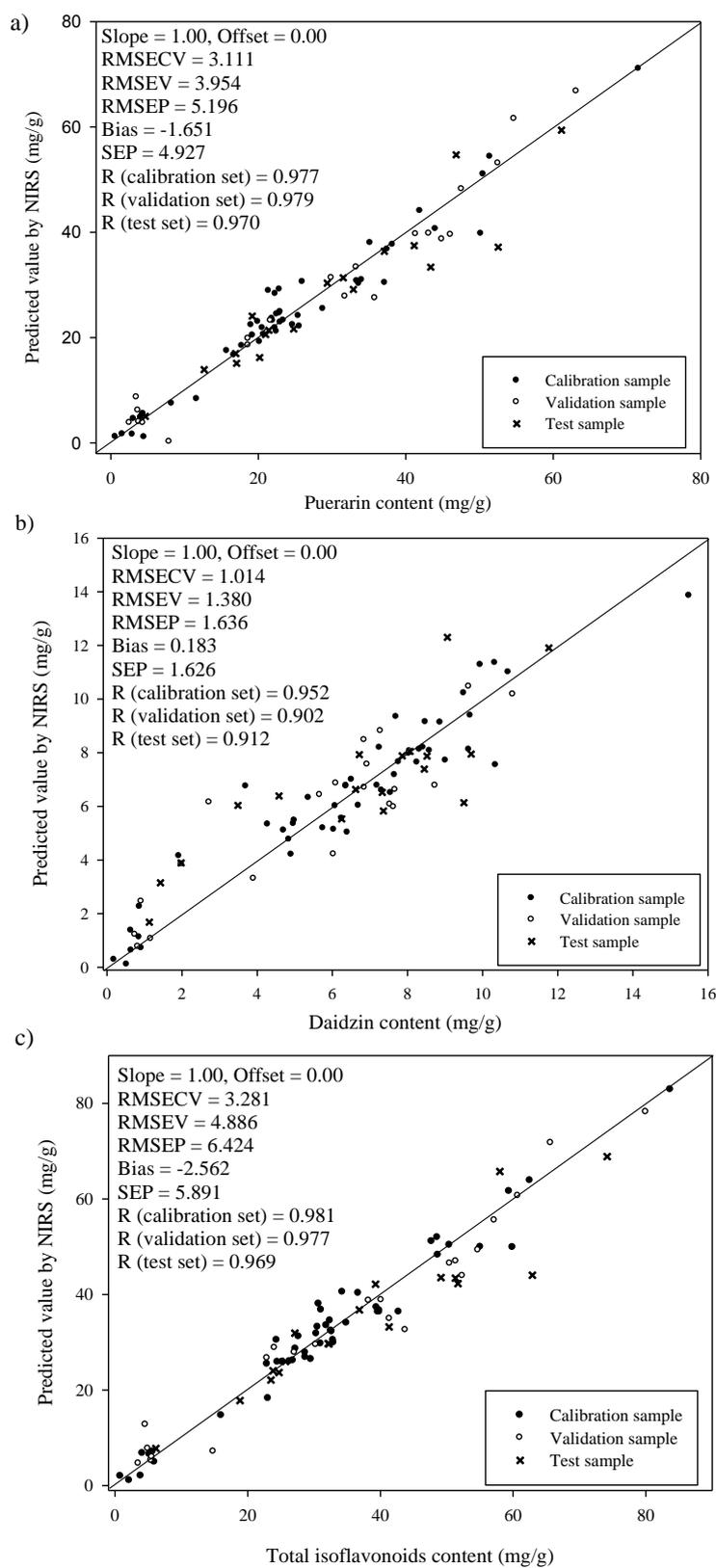


Fig. 4.11 NIRS predicted values against HPLC measurement for the content of a) puerarin, b) daidzin and c) total isoflavonoids in the *Puerariae Radix* samples ($n = 89$).

Daidzin model

Table 4.9 lists the RMSECV, RMSEV, RMSEP, BIAS, SEP and the correlation coefficient of the daidzin models developed by applying different pre-processing methods. For each pre-processing method, only the best model with the lowest RMSEV value was shown. The model pre-treated with third derivative (10,0) and SG (n=7,quartic) on the whole NIR spectra gave the lowest RMSEV value among all other pre-processing methods (4 PLS factors). Fig. 4.11 b) is a scatter plot depicting the correlation between NIR predicted value and HPLC reference measurement of daidzin model treating by the third derivative and Savitsky-Golay. Dot, circle and cross represent calibration, validation and test data, respectively. The predicted results had a good correlation with HPLC reference measurement in calibration, validation and test sets. Daidzin content of the test set is predicted with a RMSEP of 1.636. The correlation coefficients of this daidzin model for the calibration, validation and test set are 0.952, 0.902 and 0.912, respectively.

Total iso-flavonoids model

Table 4.9 lists the RMSECV, RMSEV, RMSEP, BIAS, SEP and the correlation coefficient of the total isoflavonoids models developed by applying different pre-processing methods. Again, only the best model with the lowest RMSEV value

for each pre-processing method was shown. The model pre-treated with SNV, detrend, second derivative (10,0) and SG (n=9,quintin) on the whole NIR spectra gave the lowest RMSEV value among all other pre-processing methods (5 PLS factors). Fig. 4.11 c) is a scatter plot depicting the correlation between NIR predicted value and the HPLC reference measurement of total isoflavonoids model developed by treating with SNV, detrend, second derivative and Savitsky-golay. Dot, circle and cross represent calibration, validation and test data, respectively. The results were great with good correlation of the predicted results to HPLC reference measurement in calibration, validation and test sets. Total iso-flavonoids content in the test set is predicted with RMSEP of value 6.424. The correlation coefficients of the total isoflavonoids model for the calibration, validation and test set are 0.981, 0.977 and 0.969, respectively.

4.5.7 Study of YG of different age

In this study, some samples of known age were collected from the same site (Table 4.1). Table 4.10 summarizes the variation of isoflavonoids content in YG from one to six years. Total isoflavonoid content was gradually increase during one to three years, and then significantly decreased to five years to reach its lowest level. For individual

isoflavonoids, content variations of puerarin and daidzein were in agreement with those of total isoflavonoid content while variations of daidzin, genistein and genistin exhibited a slightly different trend. However, these differences had little impact on the general variation of total isoflavonoids content since the amount of three isoflavonoids accounted for only a small fraction of the total isoflavonoids. Thus, the analytical results suggested that YG harvested in three year should be of best quality regarding isoflavonoids.

Year	Puerarin (mg/g)	Daidzin (mg/g)	Daidzein (mg/g)	Genistein (mg/g)	Genistin (mg/g)	Total isoflavonoids (mg/g)
1	20.83 ± 1.76	7.57 ± 0.77	0.99 ± 0.20	0.06 ± 0.03	0.21 ± 0.04	29.66 ± 2.37
2	28.14 ± 6.40	7.65 ± 1.86	1.06 ± 0.27	0.13 ± 0.10	0.22 ± 0.05	37.20 ± 6.34
3	33.35 ± 10.66	8.37 ± 3.09	1.47 ± 0.41	0.23 ± 0.13	0.41 ± 0.18	43.82 ± 13.28
4	17.96 ± 2.03	7.83 ± 1.13	0.93 ± 0.20	0.12 ± 0.03	0.19 ± 0.03	27.03 ± 2.74
5	16.58 ± 6.42	9.01 ± 1.95	0.76 ± 0.10	0.10 ± 0.07	0.21 ± 0.04	26.66 ± 6.49
> 5	29.17 ± 6.53	5.97 ± 1.25	0.95 ± 0.38	0.30 ± 0.16	0.41 ± 0.38	36.82 ± 4.67

Table 4.10 Variations in the content of individual and total isoflavonoids in YG

In order to evaluate the variation of YG, principal component analysis (PCA) was further performed based on the isoflavonoids contents of 36 samples from HPLC

profiles. In order to use the entire chromatographic profiles of YG for data analysis, it is necessary to correct the retention times of the common peaks identified in the previous section as the multivariate analysis is very sensitive to retention time shifts. In this study, the five isoflavonoids were used for Local least-square (LLS) alignment. Also, the variation in sample concentration might affect the multivariate analysis of the entire chromatographic profile so that normalization was carried out before analysis. Fig. 4.12a showed that year one, four and five YG samples were grouped together while the better quality samples including year two, three and more than year five YG samples scatter out (from the view of total isoflavonoids content $> 35\text{mg/g}$). In addition, centering the PCA scores of each year's sample, it was found that year three samples (age of good quality sample) could be separated from other cultivation years. At the same time, their corresponding NIR spectra were also evaluated by PCA, and found that the obtained result (shown in Fig. 4.12b) is similar to that of HPLC profile. The result suggested that NIRS PCA model may provide a faster way to estimate the cultivation year of YG sample. More samples are required for further substantiate the conclusion.

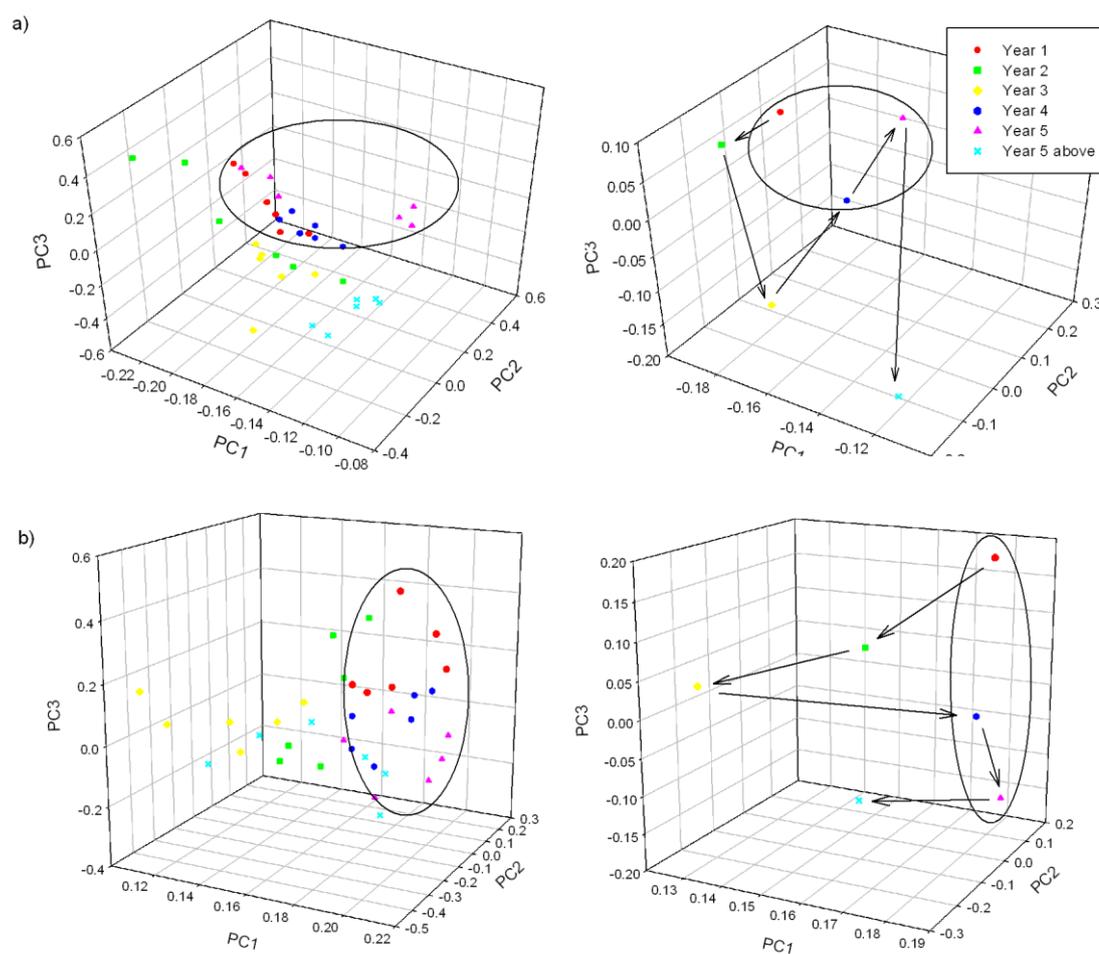


Fig. 4.12 The scatter plots obtained by PCA of the YG samples with different cultivation years by using a) HPLC data and b) NIR data

4.5.8 Implementation of NIRS method on quality control of CHM

In summary, data pre-processing is critical in obtaining the quantitative models of puerarin and total isoflavonoid content (see Table 4.9). In this work, second-order differentiation was used to enhance the differences between spectra and splits overlapping band. MSC and detrend were used to correct the variations in light

scattering due to different particle size distribution and packing density by adjusting the spectra based on intensities of spectra in regions carrying no specific chemical information¹⁴¹. In the case of daidzin, the RMSEP of models with data pre-processing is one forth smaller. Different quantitation models for other isoflavonoids such as genistin, genistein and daidzein were also established. However, the RMSEP are higher than 50% of the mean predicted contents of the isoflavonoids, significantly larger than the other models developed and reported in here. This suggested that the content prediction based on NIRS for components less than 1% is more difficult.

The spectral differences in NIRS show the variations among *Puerariae Radix* samples from different species. Development of the identification and quantification models in NIRS involves a lot of time, costs and efforts. Yet, NIRS provides the advantage of non-destructive, short measurement time when model is established. Even though physical interference may deteriorate the prediction results of the NIRS model, results comparable to that determined by HPLC can still be obtained when appropriate chemometrics treatments were employed in the model. Although the average error predicted from NIRS is much larger than those obtained from HPLC (nearly 4 times), it is still acceptable that the herbal sample nearly remain unchanged after the measurement. The present study also demonstrated that NIRS is applicable

in quality control of *Puerariae Radix* samples and the method can be adapted for online monitoring.

Development of correlation models and quantitative models is the most time demanding step in the analysis of CHM using NIRS. Taking *Puerariae Radix* as an example, the time required to develop the correlation models and quantitative models is about 60 working days. These included the time for adapting an HPLC condition to a chromatographic system, obtaining the NIR spectra, sample preparations and chromatographic experiments for the 89 samples and data analysis. However, for people or organizations accessing the quality of particular herbs or herbal products regularly, such as CHM manufacturers or traders, the time for model development is greatly reduced as the HPLC data or other reference data are already available. Assuming that a conventional quality control method is being performed regularly, the total time to build the models for NIRS analysis would be about two to three working days for acquiring the spectra and analyzing the data. In the case of *Puerariae Radix*, the time required for sample preparation and spectra measurement of one sample would be about 15 min compared to 3 hours for the HPLC method. The accuracy of NIRS is not good enough to replace the conventional HPLC method. The NIRS can still serve as a method to screen the samples before submission to HPLC analysis.

Thus the use of NIRS greatly reduced the number of HPLC analysis, without compromising the level of quality control.

4.6 Conclusion

The results demonstrated the feasibility of using NIRS for quality control of CHM in identification of species and quantification of major constituents with a short analysis time. Successful classification models between *Pueraria lobata* (YG) and *Pueraria thomsonii* (FG) and good correlation quantitative models of puerarin, daidzin and total isoflavonoids content in both species were obtained in our works. It can be concluded that major component(s) in CHM with content higher than 0.1% can be analyzed fast and simultaneously by NIRS coupled with PLS algorithm, and this real-time measurement will significantly improve the efficiency of quality control and assurance.

**Chapter 5: Variable Selection in
Near-Infrared Spectroscopy:
Analysis of Protoberberine Alkaloids
in Coptidis Rhizoma**

5.1 Introduction

In the previous chapter, Near-Infrared Spectroscopy (NIRS) had been applied successfully on the classification of *Puerariae Radix* and quantitation of puerarin, daidzin and total isoflavonoids in it. The conventional Partial least squares (PLS) regression method had been demonstrated to be useful in treating large data matrices, extracting relevant information from the spectra as well as producing reliable models. However, in most of the herbs, the complexity of the compositions may hinder the model development. Moreover, when the average content of the target marker in herb is less than 1%, the modeling becomes difficult. In these cases, the PLS regression result of the target marker would be more sensitive to noise and interference. The situation becomes even worse in the case of limited sample size and strong matrix effect.

Variable selection could be an important step in multivariate data analysis, particularly, when sample size is relatively small, number of variables is large and many of these variables containing redundant or noisy information. In these scenarios, a variable or feature selection procedure may be useful to avoid the over-fitting problem. Over-fitting occurred when the model picks up the idiosyncrasy of the data, the noise is also being fitted, and the model loses its

prediction ability for samples not in the training set.

Various variable selection (wavelength selection in our case) algorithms and criteria for use with multivariate calibration have been reported²²⁶⁻²⁴⁰. PLS regression is widely utilized in processing full spectra because of its ability to extract analyte information from the variance within the spectral data matrix. Wavelength selection methods have traditionally not been used with PLS regression models because of this ability to decompose the data matrix in a manner biased toward the isolation of analyte-dependent information. However, recent studies have indicated that the performance of PLS models can be improved through wavelength selection^{226, 236, 237, 240}. A mathematical justification of the theory supporting that wavelength selection can enhance the performance of PLS models was also reported recently²⁴⁰.

Several techniques like Uninformative Variable Elimination (UVE)²⁴¹, Support Vector Machine (SVM)^{242, 243}, Artificial Neural Networks (ANN)²⁴⁴, Simulated Annealing (SA)²⁴⁵ and Genetic Algorithms (GA)^{234, 246} can be applied to select the most informative variables. However, Genetic Algorithms (GA) has already been demonstrated its effectiveness in selecting wavelength region for further data analysis^{230, 247-251}. The performance on variable selection is very good for NIR data

sets²⁵²⁻²⁵⁴.

In order to verify the applicability and the performance of the GA in improving the NIRS models, *Coptidis Rhizoma* was studied. It is a commonly used medicinal herb in China for the treatment of dysentery, hypertension, inflammation, and liver diseases^{255, 256}. It is known as “Hunaglian” in Chinese. The reason of choosing this herb for the study is that it contains a class of protoberberine alkaloids, such as berberine, palmatine, coptisine and jatrorrhizine present in significant amount ranging between 0.1-7%. The protoberberine alkaloids also contribute to the biological activities of this herb²⁵⁷⁻²⁶⁰. The objective is to develop methods that could determine as many markers as possible even though the contents of some markers are less than 1%, as we concluded in previous chapter; it is the detection limit for quantitation of marker in CHM using NIR.

In this work, we aimed to establish fast and non-destructive analytical methods using NIRS for quantitative analysis of five alkaloids and quality control of *Coptidis Rhizoma*. Prior to NIRS modeling, a RRLC-DAD analysis was carried out in determining the five alkaloids contents in *Coptidis Rhizoma*. RRLC uses column of smaller particle size to enhance separation power so that the entire chromatographic

analysis of *Coptidis Rhizoma* could be obtained in a much shorter duration (~10 min). Validation parameters of the RRLC-DAD method such as sensitivity, linear range, precision and accuracy were determined. In addition, Taguchi's design was utilized to optimize the extraction conditions. Finally, all 47 *Coptidis Rhizoma* samples were used in model development for quantitation of the five markers berberine, epiberberine, palmatine, coptisine, and jatrorrhizine, as well as that of the total alkaloids by using GA-PLS and their models performance is also compared with PLS. A new algorithm GA-PLS is developed in this work which uses GA to select suitable wavelength regions which give best PLS models.

5.2 Background of *Coptidis Rhizoma*

Coptidis Rhizoma, known as Hunaglian in Chinese, is a commonly used medicinal herb in China for the treatment of dysentery, hypertension, inflammation, and liver diseases^{255, 256}. According to the latest version of Chinese Pharmacopoeia¹⁹⁶, *Coptidis Rhizoma* is the dried rhizome of *Coptis chinensis* Franch (WeiL 味連), *Coptis deltoidea* C.Y. Cheng et Hsiao (YaL 雅連) or *Coptis teeta* Wall (YunL 雲連). The contents of four markers, berberine, epiberberine, palmatine and coptisine are the quality parameter of *Coptidis Rhizoma*. This is more stringent than the previous requirement as the berberine content was the only quality parameter in Chinese

Pharmacopeia 2005²⁶¹.

Coptidis Rhizoma is mainly found in the temperate zone of the Northern Hemisphere where the Ranunculaceous plants generally grow. In China, it usually grows in the mountainous region of the southwest and south district²⁶². WeiL is mainly cultivated in Sichuan, Hubei, Hunan and Shanxi, YaL is produced in Emei and Hongya in Sichuan, and YunL is distributed in Yunnan²⁶³.

WeiL presents in cluster from which looks like a chicken claw with the 3-6 cm of length with 0.3-0.8 cm of diameter. It has rough, grayish yellow or yellowish brown surface with irregular node-like bulges, and the smooth part between the bulges is called 'Gou Qiao'. The upper part of the herb has brown scale leaves residues and the top part of the herb has remained shoots of leaf stems. Its cortex is red or brown, and the xylem is bright yellow or orange and arranges in radial shape, the pith sometimes is empty. YaL is usually in form of thin cylinder-liked single branch with the 4-8 cm of length with 0.5-1 cm of diameter. It has few shoot residue and longer 'Gou Qiao'. YunL is smaller in size. It bends appearing like a hook, and presents in individual form of branch²⁶¹.

5.2.1 Chemical studies

Coptidis Rhizoma contains rich protoberberine alkaloids components such as berberine, palmatine, coptisine, jatrorrhizine, worenine, epiberberine, berberastine and groenlandicine. Berberine is the main alkaloids present in Coptidis Rhizoma and its content should not be less than 5.5%²⁶¹.

Several methods have been reported for the determination of alkaloid contents in Coptidis Rhizoma extracts, which included colorimetry²⁶⁴, Thin Layer Chromatography (TLC)²⁶⁵, Capillary Electrophoresis^{264, 266, 267}, Micellar Electrokinetic Chromatography²⁶⁸, High Performance Liquid Chromatography (HPLC) with detection using diode array^{69, 269} and mass spectrometry²⁷⁰⁻²⁷³, and proton nuclear magnetic resonance (H-NMR) spectroscopy²⁷⁴.

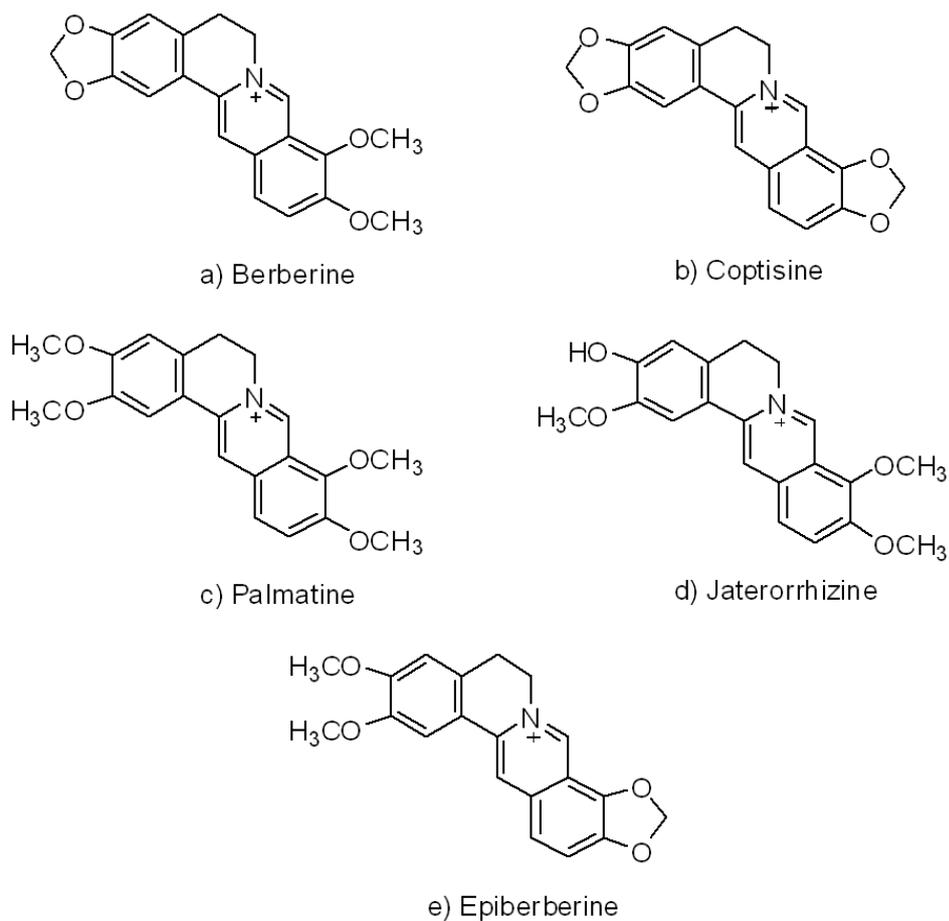


Fig. 5.1 The chemical structure of the major protoberberine alkaloids a) berberine, b) palmatine, c) coptisine, d) jatorrhizine, and e) epiberberine

5.2.2 Biological studies of alkaloids

For the biological function, the extract of *Coptidis Rhizoma* reduced the inflammation caused by TNF- α in dermatological conditions²⁷⁵. *Coptidis Rhizoma* extract can also lower that serum cholesterol level and liver cholesterol level²⁷⁶.

Berberine is the marker required for *Coptidis Rhizoma* in the Chinese Pharmacopoeia 2005. It can inhibit the gastrointestinal infection caused by the toxic substances produced by *E. Coli*²⁷⁷. It can also dilate the blood vessels and inhibit the secretion of adrenal glands, and hence, lower the blood pressure. Besides, it can increase the production of bile acid and reduce the cholesterol level²⁷⁸. A study suggested that berberine, can reduce the blood glucose level. It has also been reported for inhibition of amoeba²⁷⁹.

Coptisine, one of the characteristic components found in *Coptidis Rhizoma*, can inhibit the vascular smooth muscle cells (VSMCs) proliferation and hence provide the protection from vascular disorder²⁸⁰. Jatrorrhizine has been shown as antibacterial, antifungal and parasite-fighting properties²⁸¹. Palmatine can inhibit the reverse transcriptase activity which makes it potentially effective against tumors. Besides, it was observed to inhibit the DNA synthesis against some bacteria, fungi and virus²⁸².

5.3 Chemometric techniques used

5.3.1 Generic Algorithm - Partial Least Square (GA-PLS)

Generic Algorithm, GA, is an efficient method for optimization on complex surfaces. It is based on the principle of genetics and natural selection. The technique combines the advantages of optimization algorithms based on random processes (stochastic methods) with those based on deterministic procedures. GAs are well known for their ability to escape from local optima and to find global optima in a relatively short period of time. The algorithm samples large number of points in the parameter space at each step. Several reviews of GAs and their application in chemistry-related problems are available ^{248, 249, 283, 284}. A brief description of the GA implementation used in this research is provided here.

A schematic of the operation of a simple GA is shown in Figure 5.2. Its implementation consists of four steps. A collection of values of the variables or parameters to be optimized is called a chromosome, and the variables themselves are called genes. In this work, the spectral points and the number of PLS factors used in the model are the genes. The initialization step creates an initial population of chromosomes. The initial population is either assigned manually or generated by

applying a mutation operator randomly to perturb an initial chromosome. The size of the population is an important parameter to control the size of the calculation and the efficiency of the optimization.

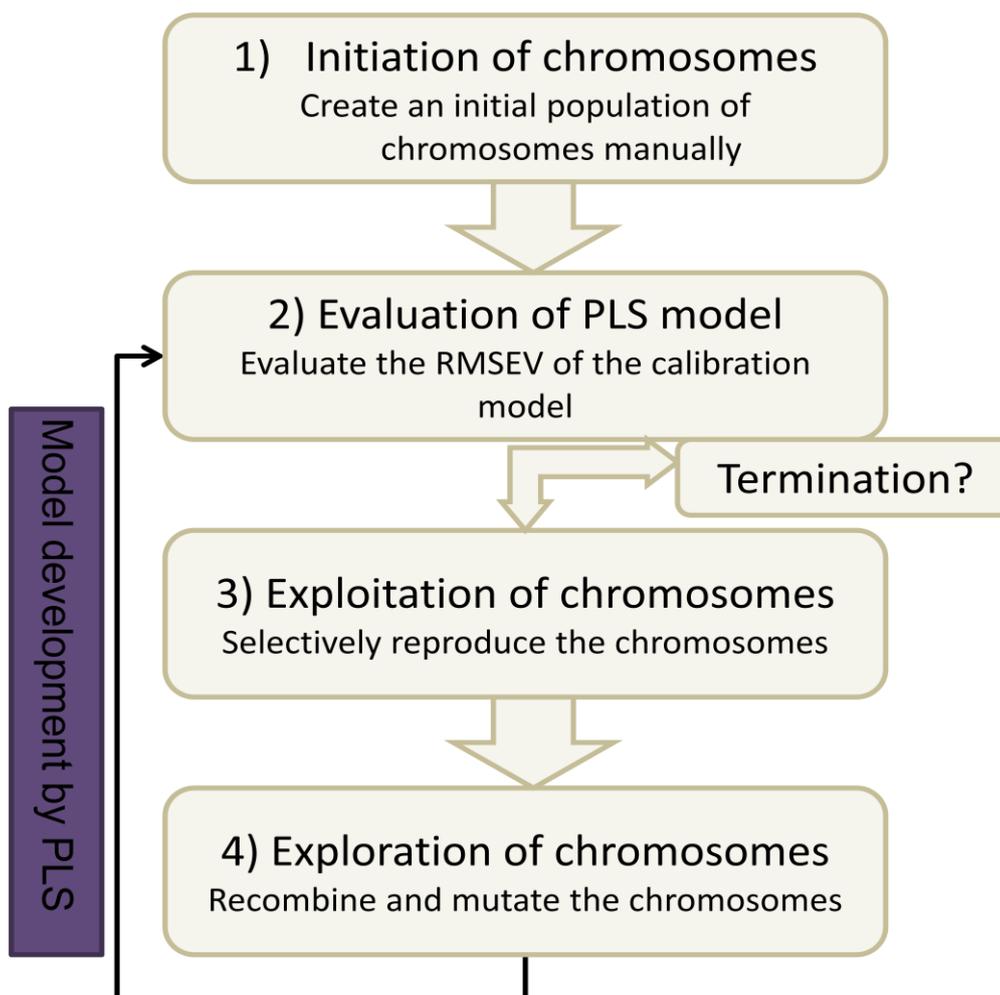


Fig. 5.2 Schematic diagram of a simple genetic algorithm

The second step is the evaluation step. The performance of the calibration model established from the spectral points and the number of PLS factors described by each chromosome are evaluated through an objective function, known as a fitness

function. The fitness value is a numerical measure of how well the computed calibration model performed in prediction. The fitness value is the criterion for guiding the GA to the global optimum. In this work, RMSEP (root mean square error of prediction) is used as the fitness value.

Each chromosome is composed of a set of genes, which are binary coded spectral variables (encoding). As the algorithm run, the genes are turned on or off (selected or not selected) and the chromosomes that produce models using PLS or MLR with lower values of RMSEP (root mean square error of prediction) for each generation are chosen to reproduce in the next generation.

In the exploitation step, the survived chromosomes are used to generate a new set of child chromosomes through the methods of recombination and mutation. The recombination approach that typically employed is termed single-point crossover. Given two selected parent chromosomes, a gene location on the chromosome is chosen randomly, and the values of all the genes up to that point are interchanged between the two parents to form two child chromosomes. Mutation is applied to the child chromosomes to alter the gene values on a gene-by-gene basis. The probability of mutation occur on a given gene is specified manually. Recombination and

mutation introduce diversity into the child chromosomes while preserving the information carried by the parents. The new population formed with the child chromosomes replacing the original and the chromosomes with higher fitness values in the new population are again selected to reproduce in the next generation through recombination and mutation. This procedure is an iterative evolutionary process seeking the chromosome with the highest fitness value. The crossover and mutation are introduced to overcome local optimization. This small perturbation is a good way to probe other parts of the gene keeping in a certain extent the adaptation of the considered individual.

The formation of each new population represents one iteration of the algorithm, which is termed as a generation. The algorithm terminates at certain generations or when a chromosome reached a user-specified level of fitness is found. The exploitation step removes chromosomes with poor fitness values, thus implementing a selection similar to that of evolution process.

5.3.2 Taguchi design

Taguchi-based design is a unique and powerful optimization technique that allows optimization with minimum number of experiments²⁸⁵. It is commonly used for designing and performing experiments to investigate processes where the output depends on many variables or factors without having to tediously carry out the experiments using all possible combinations of all variables. By identifying controllable variables and varying them in specific trials, variables with significant effect can be determined and optimized. Generally speaking, Taguchi's orthogonal array principle can be used to alter the variables or combination of variables to appear an equal number of times, and more than one variable is examined for each experiment. After the results are known, sum of the squares of different results are calculated and compared. This calculation finds the most significant improvement with the least variability. Therefore, Taguchi method can be used to reduce times of experiment, and produce sufficient information to determine the factors affecting the results.

5.4 Experimental details

5.4.1 Samples and reagents

In this study, 47 authenticated samples were collected from different provinces of China. All samples were authenticated by Dr Si-Bao Chen (State Key Laboratory of Chinese Medicine and Molecular Pharmacology, Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China). All samples were grinded into powder which is then passed through a 100 mesh (150 μ m) stainless sieve. The sieved powdered samples were stored in plastic containers at 20°C. Samples were freeze dried for two days before NIRS analysis.

No. of Samples	Sample Code	Latin Name	Origin
1	HL-01	<i>Coptis teeta</i> Wall. (YunL)	NanChengJiang, China
2	HL-02	<i>Coptis chinensis</i> Franch.(WeiL)	SiChuan, China
3	HL-03	<i>Coptis deltoidea</i> C.Y. Cheng et Hsiao (YaL)	YaAn, SiChuan, China
4-6	HL-04,07,08	<i>Coptis chinensis</i> Franch. (WeiL)	Unknown
7	HL-05	<i>Coptis chinensis</i> Franch. (WeiL)	ShiZhuYuan, SiChuan, China
8-18	HL-06,09,11-19	Unknown	Unkown
19-20	HL-10,30	<i>Coptis chinensis</i> Franch. (WeiL)	ErMei, SiChuan, China
21	HL-20	<i>Coptis chinensis</i> Franch. (WeiL)	Unknown
22-44	HL-21-23, 25-29, 31-32, 35-46	<i>Coptis chinensis</i> Franch. (WeiL)	ShiZhuYuan, SiChuan, China
44-46	HL-24,33-34	<i>Coptis chinensis</i> Franch. (WeiL)	Unknown
47	HL-47	Unknown	Unknown

Table 5.1 The List of the *Coptidis Rhizoma* samples collected

Berberine chloride and coptisine chloride were purchased from Wako while jatrorrhizine chloride and palmatine chloride were purchased from International Laboratory. Epiberberine was isolated in our laboratory. The purities of five standards are not less than 97% as determined by RRLC-DAD. HPLC grade acetonitrile and

methanol 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 from Milli-Q water system (Millipore Corp., Bedford, MA, USA) was used to prepare all buffers and sample solutions.

5.4.2 Rapid Resolution Liquid Chromatography (RRLC) analysis

A sample of 0.2g was weighted and sonicated with 50mL 0.5% (2M) acidified methanol for 30 min under 40°C. The mixture was filtered and the filtrate was concentrated with a rotary evaporator. Then, the concentrated filtrate was transferred to a volumetric flask and made up to 10mL solution. Finally, the sample solution was passed through a 0.45µm syringe filter before the chromatographic analysis.

Chromatographic analysis was carried out on a ZORBAX column (100mm× 4.6 mm, 1.8µm) at 40°C using an Agilent 1200 Series Rapid Resolution LC system (Agilent Technologies), equipped with a quaternary solvent delivery system, an auto sampler a thermostatic column compartment and UV detector. 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 gradient elution. The eluting program is shown as follow: initial 98% (A) at 0–1 min, 98% to 93% (A) at 1–2 min, 93% to 83% (A) at 2–6 min, 83% to 65% (A) at 6–10 min Re-equilibration duration was 3 min between individual runs. The flow rate was 1.5mLmin^{-1} and aliquots of $1\ \mu\text{L}$ were injected.

Standard solutions of the five alkaloids were prepared and diluted to appropriate concentrations to determine the calibration curves, limit of detection and quantification. A range of seven concentrations of the five alkaloids was analyzed by four replicate injections. For the chromatographic analysis of each alkaloid, the limits of detection (LOD) and quantification (LOQ) were determined at signal-to-noise ratio (S/N) of 3 and 10, respectively. The intra-day variability of each alkaloid was examined by five replicated injections in one day. Meanwhile, the recovery tests were performed to examine the accuracy of the analytical method. Accurate amounts of the alkaloids of different concentration levels (low, medium and high) were added to 0.2g pre-treated samples, followed by the extraction and analysis as described above.

5.4.3 Near-Infrared Spectroscopy (NIRS) and data analysis

All NIR spectra were recorded by a NIRSystems Model XDS spectrometer (Foss NIRSystems, Silver Spring, MD, USA) equipped with a quartz halogen lamp and a PbS detector. The spectra were collected in the reflectance mode with a ceramic reference standard over the spectral region of 1100 to 2500 nm with a resolution of 0.5 nm. The spectra were acquired with a circular sample cup with a quartz window. Each spectrum is an average of 32 scans. Each sample measurement was repeated three times after every 120° rotation of the cup, and the spectra obtained were averaged.

In this work, different data pre-processing methods were utilized to minimize interference effects like particle scattering and baseline shift, and to enhance the data quality for model establishment. These methods include mean centering (MC), auto-scaling, standard normal variate transformation (SNV), derivative, smoothing, multiplicative scatter correction (MSC) and de-trend. Good results have been obtained by using SNV, second derivative with Savitsky-Golay (SG), moving window average smoothing and MSC techniques.

All computer programs used in this study were coded in MATLAB 7.0 and all

computations were performed on a computer equipped with an Intel (R) E6300 CPU (1.86GHz) and 2 GB memories.

5.5 Results and discussion

5.5.1 Optimization of extraction condition using Taguchi experimental design

Taguchi is used to optimize the extraction condition for the chromatographic analysis. Four factors including ultrasonic duration, extraction temperature, amount of acid added and volume of the extraction solvent (Methanol) used are examined at three levels. The optimization considered the extraction efficiency of four alkaloids except epiberberine as it could not be found in one type of the *Coptidis Rhizoma* samples. The values of the extraction factors (A-D) and matrix (3^2) under the levels of consideration are listed in Tables 5.2 and 5.3, respectively.

Sample extraction factors	Levels used		
	Low	Medium	High
A: (ultrasonic duration, min)	15	30	45
B: (extraction temperature, °C)	30	40	50
C: (amount of 2M HCl acid added, v/v %)	0	0.5	1.0
D: (volume of the solvent used, mL)	30	40	50

Table 5.2 Sample extraction factors and variables for Taguchi's experimental design

(L-0 orthogonal array)

Experimental no.	Experimental order	A	B	C	D
1	2	15	30	0	30
2	9	15	40	0.5	40
3	7	15	50	1	50
4	3	30	30	0.5	50
5	4	30	40	1	30
6	1	30	50	0	40
7	5	45	30	1	40
8	6	45	40	0	50
9	8	45	50	0.5	30

Table 5.3 Randomized runs of Taguchi experimental design

The mean responses at different level for the four markers were calculated. In this work, nine experiments were done and the sequence of the experiments was randomized to avoid any subjective bias. From the data obtained, the interaction between these factors was found to be very small. The sum of squares for different factors were calculated and the results (Table 5.4) suggested that the most important factor contributing to the extraction efficiency of the four alkaloids markers was factor C (amount of acid added) followed by factor D (volume of the solvent used), factor A (ultrasonic duration) and B (extraction temperature) played a lesser role. Based on these, under the best condition, the optimal condition is the ultrasound extraction using 50mL of methanol with 0.25mL of 2M HCl at 40°C for 30 min.

Cospitine			
Factors	Degree of freedom	Sum of squares	Percentage contribution
Ultrasonic duration (A)	2	946.3	3.89%
Extraction temperature (B)	2	711.8	2.93%
Amount of acid added (C)	2	20860.7	85.74%
Volume of the solvent used (D)	2	1811.6	7.45%
Berberine			
Ultrasonic duration (A)	2	5981.1	7.19%
Extraction temperature (B)	2	3722.1	4.47%
Amount of acid added (C)	2	56904.8	68.42%
Volume of the solvent used (D)	2	16567.6	19.92%
Palmatine			
Ultrasonic duration (A)	2	349.4	8.98%
Extraction temperature (B)	2	122.4	3.15%
Amount of acid added (C)	2	2324.4	59.74%
Volume of the solvent used (D)	2	1094.6	28.13%
Jatrorrhizine			
Ultrasonic duration (A)	2	62.6	10.10%
Extraction temperature (B)	2	4.3	0.69%
Amount of acid added (C)	2	444.7	71.71%
Volume of the solvent used (D)	2	108.5	17.50%

Table 5.4 Results for experimental responses in the (3²) matrix

Finally, to validate the optimized conditions determined by Taguchi's design, the extraction efficiency for the four alkaloids was determined. Four successive extractions were carried out and the results are shown in Table 4. More than 98% alkaloids (98.2% for coptisine; 99.1% for jatrorrhizine; 98.7% for berberine; 98.1% for palmatine) in *Coptidis Rhizoma* are being extracted by the optimized extraction method for two times (Fig. 5.3).

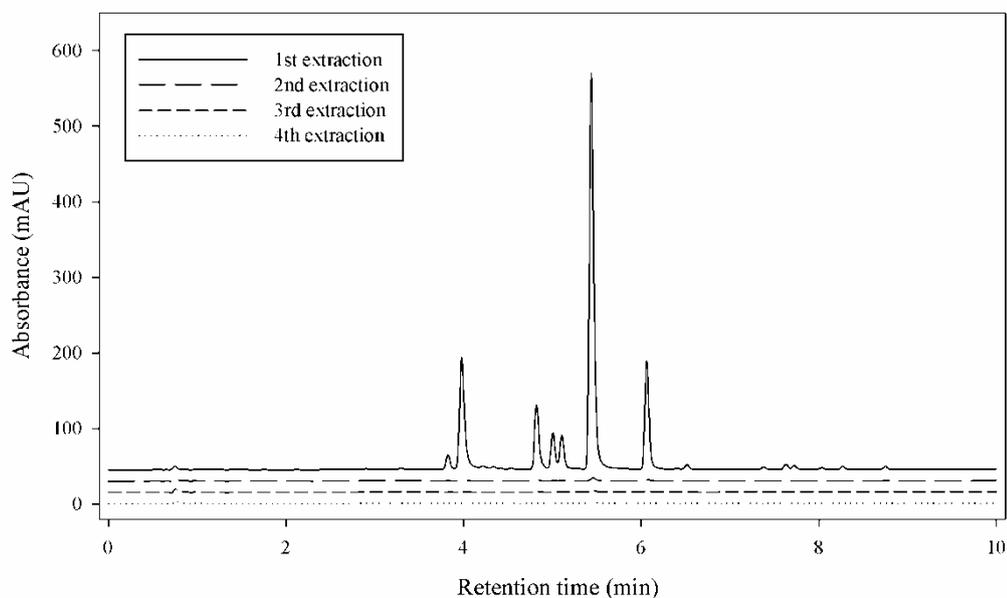


Fig. 5.3 The RRLC–DAD chromatogram of four successive extractions of *Coptidis Rhizoma* extract at 346 nm.

5.5.2 Chromatographic study of *Coptidis Rhizoma* by RRLC analysis

The chromatographic conditions were optimized in order to obtain chromatograms with a good resolution of the analyte peaks within a short period of time. Tetrabutylammonium chloride salt was added into acidified mobile phase so as to minimize the tailing effect during the separation process. In addition, further increase in column temperature resulted in incomplete separation between epiberberine and jatrorrhizine. Therefore, the column temperature was kept at 40°C. The best separation was achieved by adjusting the mobile phase compositions, flow-rates and gradient programs to those mentioned in section 5.4.2.

Fig. 5.4 shows the typical mean chromatograms of the five alkaloid components in the standard solution, extracts of YunL, WeiL and YaL. Under the experimental conditions mentioned above, peaks were identified by comparing the retention times and UV spectra with those of the standards. Retention time for standards, coptisine, epiberberine, jatrorrhizine, berberine and palmatine were 3.93, 4.80, 5.06, 5.40 and 6.03 min, respectively. It can be seen that the chromatograms of the three *Coptidis Rhizoma* species are similar to each other.

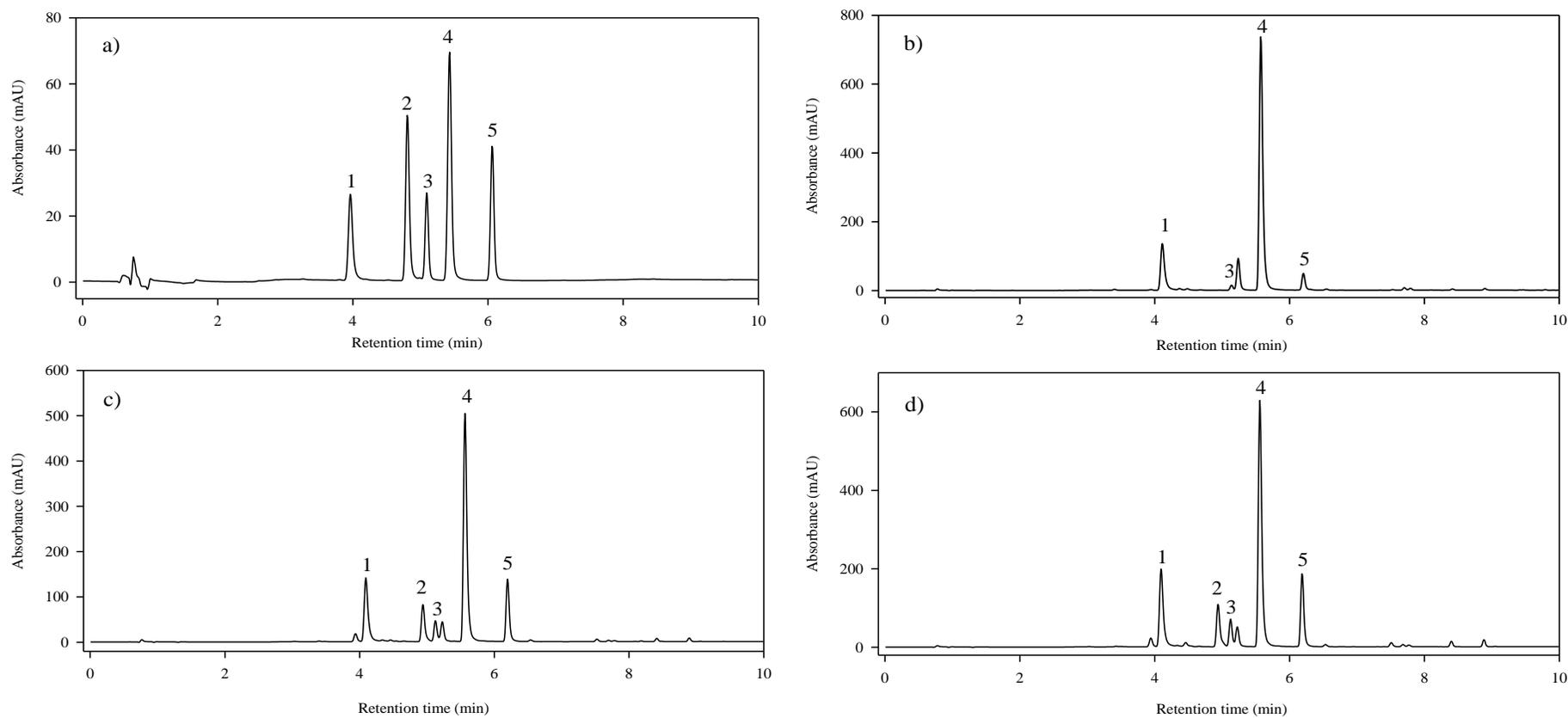


Fig. 5.4 RRLC/UV (346 nm) chromatograms of the five alkaloid components studied in a) standard solution, b) extract of *Coptis teeta* Wall (YunL), c) extract of *Coptis chinensis* Franch (WeiL) and d) extract of *Coptis deltoidea* C.Y. Cheng et Hsiao (YaL): 1) coptisine; 2) epiberberine; 3) jatrorrhizine; 4) berberine; 5) palmatine

5.5.3 Method validation

The RRLC method developed was validated according to the International Conference on Harmonization (ICH) guidelines and the results of linearity and precision obtained were tabulated in Table 5.5. All the calibration curves of the five standards exhibited good linearity ($r^2 > 0.999$) within their test ranges. The LOD and LOQ were in the range 0.12-0.20 ng and 0.41-0.66 ng, respectively. Intra-day variability was determined with both standard and sample extracts and almost all extracts were less than 1.5%. This indicated that the five analytes were stable. Also, the developed analytical method had good accuracy and the overall recoveries for all the four analytes concerned in the samples were summarized in table 5.6. All the results suggested that this assay showed good reproducibility. The amount of epiberberine isolated was not sufficient for the recovery studies, however as the structure is similar with other compounds studied; we believe that the reproducibility of epiberberine is also good.

The five alkaloids in 47 samples were quantitatively determined by the developed RRLC-DAD method and shown in Fig 5.5. The figure shows that there is a large variation in the content of the five alkaloids among the 47 samples. Results obtained

from this investigation are consistent with previous reports. This suggested that the samples of *Coptidis Rhizoma* collected for the study is a good representation of the general feature of *Coptidis Rhizoma*.

By using external standard method, the actual contents of the five alkaloids in the 47 samples were determined and their contents vary significantly (Table 5.3).

According to the standard recommended by the Pharmacopoeia of the People's Republic of China 2010, the content of berberine, epiberberine, palmatine and cosptisine should be more than 3.6%, 0.8%, 1.5% and 1.6% respectively. In the samples collected, 17 out of 47 *Coptidis Rhizoma* samples do not fulfill the limit and they are not suitable for medical use. Most of these samples cannot fulfill the requirement on the palmatine content. From this survey, there are large variations in the quality of samples, and so, the quality control for *Coptidis Rhizoma* is quiet important.

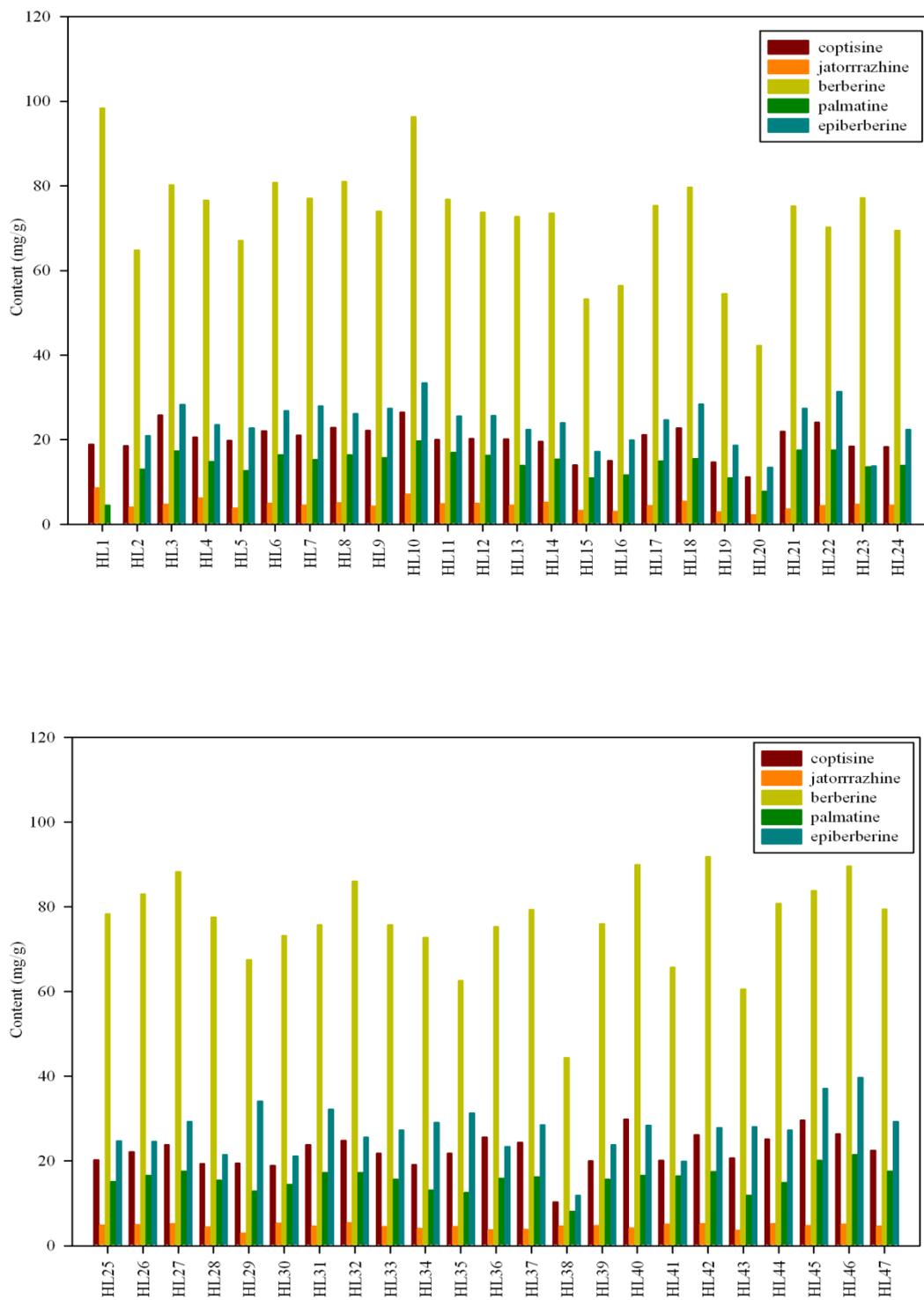


Fig. 5.5 The content of five alkaloids in 47 *Coptidis Rhizoma* samples

Component	r^2	Test range (ng)	LOQ (ng)	LOD (ng)	Intra-day variability (%)	Average content (mg/g)	Range	S.D.
Coptisine	0.9995	5 – 3000	0.66	0.20	0.63 ±0.52	21.20	10.33 – 29.86	4.04
Epiberberine	0.9993	13 – 1950	0.30	0.09	0.55 ±0.38	25.08	0.00 – 39.70	6.73
Jatrorrhizine	0.9992	2 – 1200	0.41	0.12	0.81 ±0.76	4.62	2.27 – 8.53	1.03
Berberine	0.9998	5 – 5000	0.53	0.16	0.62 ±0.44	74.58	42.28 – 98.34	11.76
Palmatine	1.000	5 – 3000	0.41	0.12	0.62 ±0.44	14.97	4.46 – 21.55	3.14

Table 5.5 RRLC method validation for the five alkaloids and results of the alkaloids contents determined by RRLC-DAD in 47 *Coptidis*

Rhizoma samples (n=4)

Analyte	Origin (mg/g)	Spiked (mg/g)	Found (mg/g)	Recovery (%)	Mean (%)	RSD (%)
Coptisine	20.78	7.22	28.15	102.2	101.1	2.54
		16.13	36.62	98.2		
		21.23	42.62	102.9		
Jatrorrhizine	4.72	2.23	6.80	113.7	111.9	1.60
		5.01	9.84	111.1		
		7.24	12.3	110.8		
Berberine	77.32	25.18	102.51	100.0	99.4	0.70
		75.56	151.86	98.66		
		110.82	187.76	99.66		
Palmatine	15.50	4.47	20.17	104.4	103.5	1.71
		13.41	29.11	101.6		
		22.34	38.87	104.61		

Table 5.6 Recoveries of the four alkaloids in Coptidis Rhizoma sample (n=4)

5.5.4 Quantitation NIR models development by GA-PLS

Raw NIR spectra of all the 47 samples are shown in Fig. 5.7a and the intensive bands in these spectra are from the first overtone of C-H stretching in the aromatic ring (1685nm), the first overtone of C-H asymmetry stretching in CH₃ (1715nm), combination band of O-H stretching and deformation vibration (2090 nm) followed by the combination band of C-H stretching and C-H bending in CH₃ (2285nm) ²⁷⁸.

Those vibrations mainly come from constituents such as alkaloids, polysaccharide and non-volatile acid. As shown in Fig. 5.6a, the raw spectra have baseline shift with broad bands. Therefore, these NIR spectra have to be preprocessed to remove baseline shift, noises and backgrounds before information extraction. Here, all the NIR data were treated by using a combination of several pre-processing tools including SNV, second derivative with the Savitsky-golay (SG) method, moving window average smoothing and MSC treatment. Then, the preprocessed spectra (Fig. 5.6b) were used in PLS regression. These mathematics tools were found to be effective in minimizing the noise, interferences of scattering and particle size and remove the baseline offset.

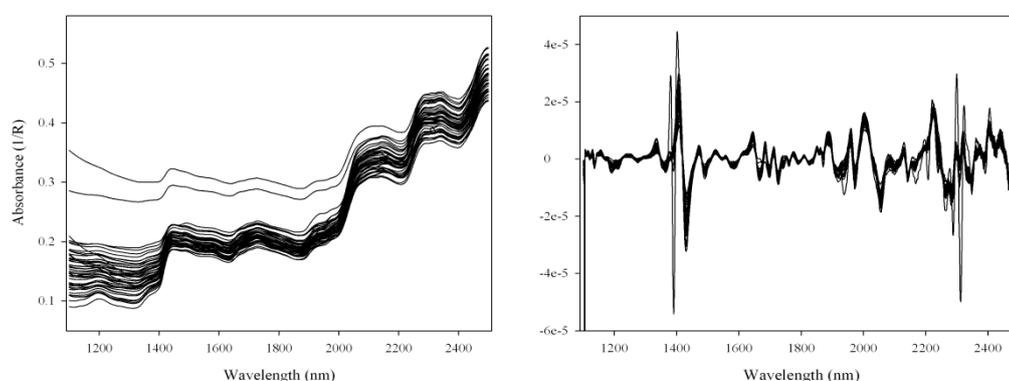


Fig. 5.6 47 NIR spectra of Coptidis Rhizoma obtained from, a) raw data, b) treatment of SNV, second derivative with Savitsky-golay (SG) method, moving window average smoothing and MSC

In this study, GA was firstly used to choose the related spectral data points for each alkaloid with the parameters shown in Table 5.7, which is the same as the reference²⁵³. In order to reduce the autocorrelation of spectral variables and increase the robustness of the model built, the number of NIR data points have been linearly interpolated to 220 (data point or gene) for each sample. Through this method, the well-defined spectral regions can be selected with good explicable features. The prediction ability is significantly improved in contrast to the full-spectrum model.

In our implementation of the GA, the values to be optimized were which of the individual spectral points being used as input variables in the PLS calculation and the number of the resulting PLS factors used to construct the calibration model. Unlike other variable selection methods which select scattered spectral points, GA-PLS usually extracts contiguous wavelength regions which can provide more useful information for further interpretation and examination. This means that the spectra in the selected region would be more relevant to the target components selected for modeling. Despite its simplicity and efficiency, due to its stochastic nature, the selection of variables by GA may not be reproducible. It is also a major disadvantage using GA for spectral selection during model development. Thus, a strategy based on the selection frequency of spectral variables (wavelength) was

proposed to build robust models in this study. The samples were split randomly into 3 equal groups, two groups were combined to form the training sets, and GA-PLS was applied on a training set to find the optimized spectrum variables. The remaining group then was used as validation set to evaluate the PLS models. The number of PLS factors used in the model was determined by minimizing the values of RMSEV. The performance of the final model was evaluated by RMSECV, RMSEV, BIAS, SEV and R. Then, the GA-PLS and modeling process were repeated until all groups were utilized as test set once. The spectral variables with selection frequency higher than one time in the 3-fold cross-validation step, and were then used to build the final calibration model. This guarantees the maximum robustness and prediction ability of the final calibration model. Again, a three-fold cross-validation was used to establish the final PLS calibration model and the GA-PLS process repeated 100 times for each analyte and the total alkaloid content. Details of the GA algorithm can be found in our previous works^{96,96,98}. In this work, total alkaloids content is defined as sum of the amounts of all the alkaloids standards that can be found in the RRLC-DAD chromatogram, i.e. the sum of the contents of the five standards (berberine, coptisine, palmatine, epiberberine, jatrorrhizine). Quantitation models for berberine, coptisine, palmatine, epiberberine, jatrorrhizine and total alkaloids content were then established from the NIR spectra.

Population size: 30 chromosomes
Average length of initial chromosome: 5
Maximum length of the chromosome: 30
Probability of mutation: 1%
Maximum number of PLS components: 15
Number of runs of variable selection: 100
Window size for smoothing: 3

Table 5.7 Parameters used GA

The scatter plots in Fig. 5.7 shows the correlation between the NIR predicted values and the reference values obtained by RRLC-DAD measurement. The dot and cross symbol represent the calibration and validation data points, respectively. It can be seen from these plots that the data in the calibration and test sets are highly correlated with the reference data.

The SEV of the quantitation models of berberine, coptisine, palmatine, epiberberine, jatrorrhizine and total alkaloids content are 4.069, 1.235, 0.982, 2.420, 0.514 and 4.980, respectively, with the relative error of these models as 5.46%, 5.83%, 6.60%, 9.65%, 11.13%, 3.55%, respectively. The R values of the test sets of the berberine,

coptisine, palmatine, epiberberine, and jatrorrhizine and total alkaloids content models are 0.937, 0.951, 0.948, 0.931, 0.862, and 0.974, respectively. This indicates that the predicted values from NIRS are very close with those determined value by RRLC-DAD. For the six NIRS models, only the palmatine and jatrorrhizine models involved more than 5 PLS factors. As for the jatrorrhizine, the average content is close to 0.4%, which is approaching or below the limit of NIRS in analyzing the herbal extracts²⁷⁷. This explains why more PLS factors are required in establishing the models.

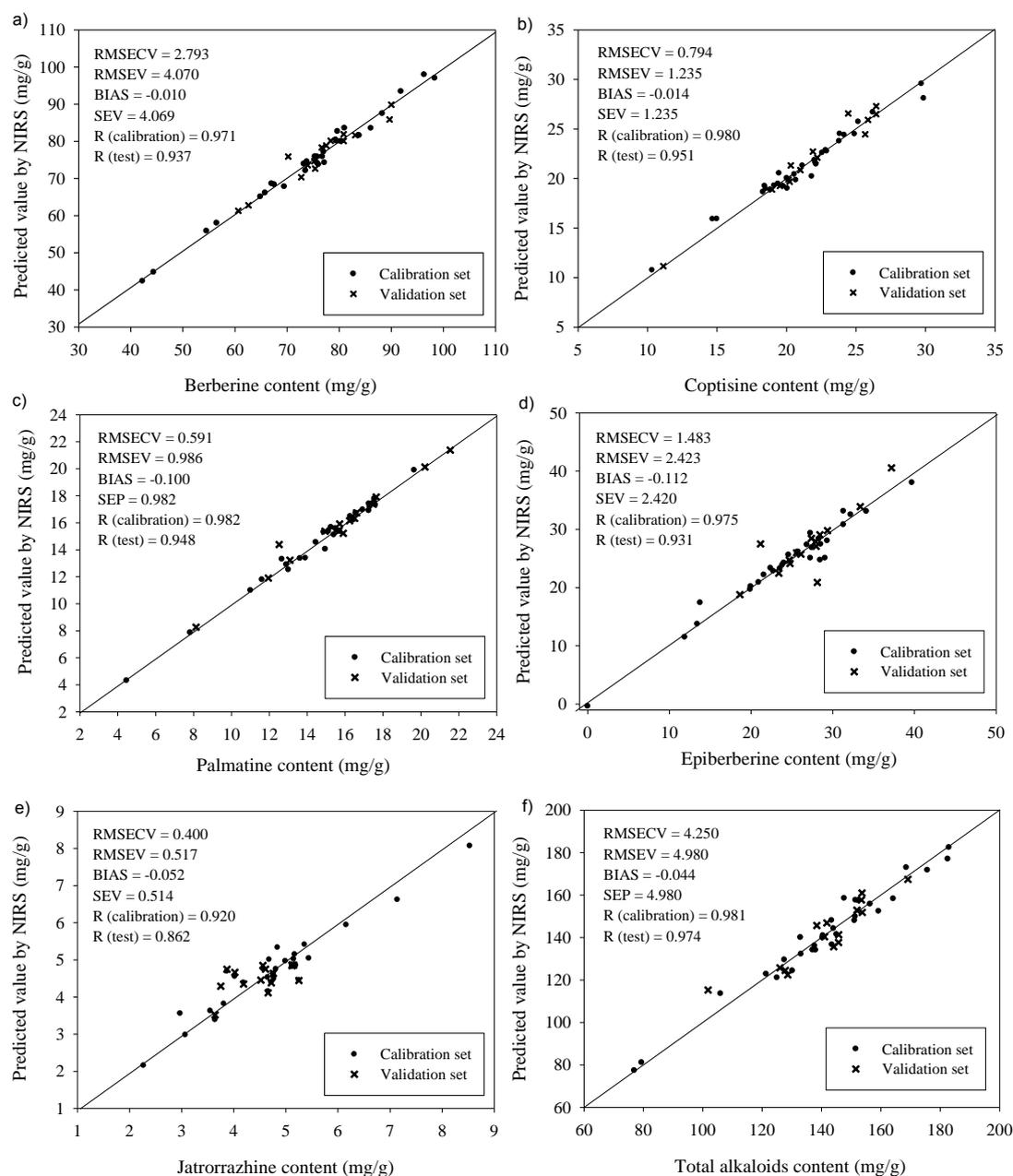


Fig. 5.7 Prediction values against RRLC-DAD measured content of a) berberine, b) coptisine, c) palmatine, d) epiberberine, e) jatrorrhizine and f) total alkaloids content in all *Coptidis Rhizoma* samples (n = 47)

5.5.5 Comparison of the NIR models developed by PLS and GA-PLS

Table 5.8 is the summary of NIRS quantitation models developed using PLS and GA-PLS. For the data set using the ordinary PLS regression method, the RMSEV of the quantitation models of berberine, coptisine, palmatine, epiberberine, jatrorrhizine and total alkaloids content are 5.227, 1.434, 0.173, 3.297, 0.674 and 4.980, respectively, with the relative error of these models as 7.01%, 6.76%, 7.84%, 13.15%, 14.59%, 4.54%, respectively. According to the results obtained, it is obvious that the average relative error for the individual alkaloid as well as total alkaloids using GA-PLS modeling is smaller than that those using ordinary PLS.

Marker	PLS		GA-PLS	
	RMSECV	RMSEV	RMSECV	RMSEV
Berberine	3.879	5.227	2.793	4.070
Coptisine	1.361	1.434	0.794	1.235
Palmatine	1.160	1.173	0.591	0.986
Epiberberine	2.132	3.297	1.483	2.423
Jatrorrhizine	0.437	0.674	0.400	0.517
Total alkaloids Content	6.286	6.458	4.250	4.980

Table 5.8 Performance of the models developed by PLS and GA-PLS

To investigate the detail of GA-PLS modeling, the variables used for GA-PLS models were examined. 35, 36, 36, 50, 32 and 49 variables were finally selected as wavelength regions, shown in Fig. 5.8, through the selection of 100 GA-PLS runs, and were used to build up the quantitation models with PLS regression for berberine, coptisine, palmatine, epiberberine, jatrorrhizine and total alkaloids content, respectively. As shown in Table 5.8, not only the prediction ability was slightly better than the ordinary PLS, the variables involved in model had been a significant reduction as shown in Fig. 5.8. Variables carrying useless information were eliminated by GA.

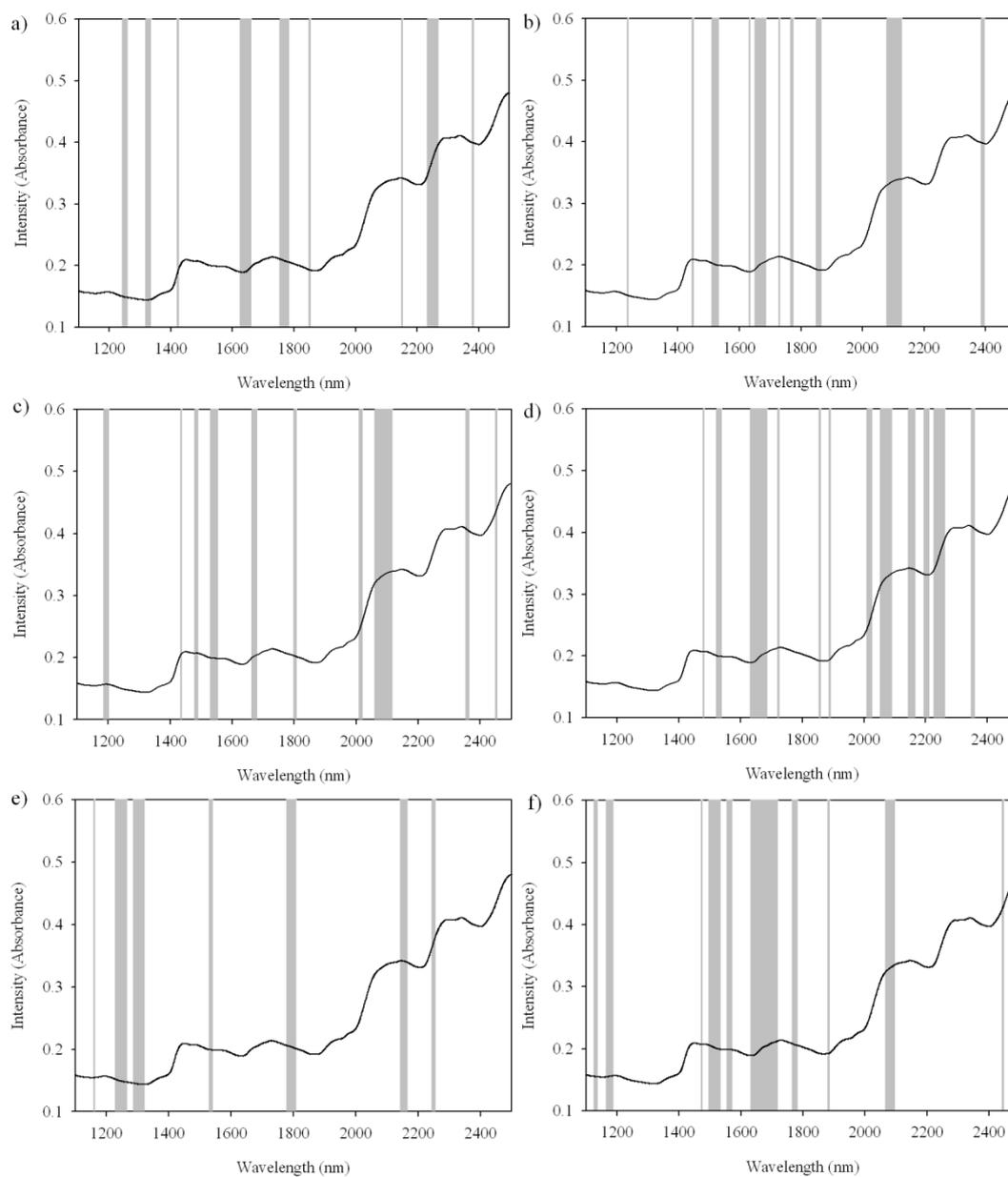


Fig. 5.8 Overview of the spectral variables selected by GA a) berberine, b) coptisine, c) palmatine, d) epiberberine, e) jatrorrhizine and f) total alkaloids content overlaid with the average NIR spectrum of *Coptidis Rhizoma*

In the algorithm, variables are removed, that might decrease the variances in the data set. The robustness of the calibration model should be improved. However, considering fewer test samples in the model means fewer parameters and potentially more bias. It might possibly have a lower prediction error than for the full-spectrum model. It should be noted that variable selection itself is an additive source of variance that may contribute to the prediction error. In the case considered here, a full cross-validation was then performed and the spectral variables with selection frequency higher than one time in the 3-fold cross-validation were chosen to build the final calibration model. This guarantees the prediction ability of the final calibration model, even though the population size is limited and the content of target could be as low as 0.1%.

Compared to the ordinary PLS regression, GA-PLS only focuses on relevant (selected) wavelength regions. Thus, we can expect that predictions are be less impacted by a change on the spectra caused by the variation of environmental factors such as temperature change and humidity. Since the use of genetic algorithms reduces the relative error slightly (only 30-50 out of 220 wavelength regions have a coefficient), we can suppose that the model robustness will be more stable over time if fewer wavelength regions are used in modeling.

5.5.6 Validation of quantitation models developed by GA-PLS

In our work, all the proposed NIRS models were validated in accordance with the ICH, EMEA and PASG guideline through assessing their specificity, linearity, accuracy and precision²⁸⁶⁻²⁸⁹.

In the NIRS assay, the specificity is defined as the ability to measure the alkaloids contents in a specific matrix – the Coptidis Rhizoma herbal matrix. Therefore, a spectral library allowing the identification of Coptidis Rhizoma samples was built in this work. The important wavelength regions including 1503 – 1541nm (first overtone of N-H), 1560 – 1580nm (first overtone of intra-molecular O-H), 1637 – 1727nm (first overtone of C=C, first overtone of C-H in aromatic ring & symmetrical first overtone of –CH₃) and 1772 – 1791nm (asymmetrical first overtone of –CH₃)²⁹⁰ had significant loading in our PLS model for the total alkaloids content, and these region were selected and used in model development. The specificity of the Coptidis Rhizoma library is assessed by external samples which have not been used in the modeling. In this case, Phellodendri Cortex, known as Huangbo in Chinese, samples were served as external samples in the selectivity assessment. From our previous study⁹⁸, Phellodendri Cortex contains about 6-8% alkaloids components, including berberine, palmatine and jatrorrhizine. We believed

that Phellodendri Cortex would be suitable for this purpose. NIR spectra of 30 Phellodendri Cortex collected for the study are shown Fig. 5.9a. It is quite clear that the difference in the spectra of Phellodendri Cortex and Coptidis Rhizoma (mean NIR spectra, shown in Fig. 5.9b) is not significant.

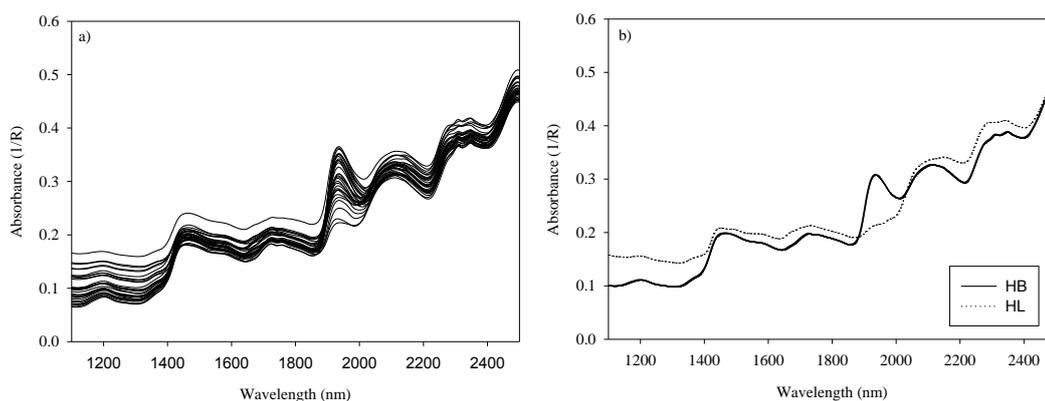


Fig. 5.9 Raw NIR spectra of a) 30 Phellodendri Cortex samples, b) mean Phellodendri Cortex and Coptidis Rhizoma

To examine the specificity of the Coptidis Rhizoma library, the Mahalanobis distance was used as a measure of the similarity between the individual spectrum (test sample) and the calibration spectra (mean spectrum). The Mahalanobis distance for each sample spectrum is calculated after the PCA is performed. Table 5.9 shows that all the Phellodendri Cortex spectra analyzed have their Mahalanobis distances, which were much higher than the threshold defined by the method used in Coptidis Rhizoma (1.416), and all the Phellodendri Cortex samples, were considered as the

outliers. On the other hand, all *Coptidis Rhizoma* spectra have their Mahalanobis distance below the method threshold, showing that the method is specific to *Coptidis Rhizoma*.

Samples	Spectrum mahalanobis distance	Mahalanobis distance threshold	Outlier
<i>Coptidis Rhizoma</i>	0.765 ± 0.472 (n=47)	1.416	No
<i>Phellodendri Cortex</i>	3.177 ± 0.702 (n=30)		Yes

Table 5.9 Method selectivity for the *Coptidis Rhizoma* and *Phellodendri Cortex* samples

The linearity of an analytical method relates to the range in which the test results are directly proportional to the concentration of analytes in samples. In multivariate calibration methods, the calibration model is established by using the concentrations of the sample extract rather than that of the standard solutions as the effect of the sample matrix cannot be removed explicitly. Thus, the linearity is limited by that of the reference method. In the ideal case, the linearity of the NIRS is the same as that of the reference method if the regression line with a unity slope and a zero intercept, i.e. the NIRS model provided the same results as the reference method for a sample

set spanning the same concentration range. A range of $\pm 20\%$ around the nominal value for each analyte in all the developed NIRS models was used in linearity assessment. This led to all the six NIRS models met the criteria over the concentration ranges examined (see Table 5.10).

Parameter	Component	Berberine	Coptisine	Palmatine	Epiberberine	Jatrorrhizine	Total alkaloids content
Linearity	Range (mg/g)	42.28 – 98.34	10.33 – 29.86	4.46 – 21.55	0.00 – 39.70	2.27 – 8.53	76.94 – 182.93
	Equation	$y=0.923x+5.65$	$y=0.964x+0.74$	$y=0.976x+0.46$	$y=0.827x+4.23$	$y=0.929x+0.28$	$y=0.918x+11.15$
	Intercept	0.948	0.124	0.092	0.738	0.069	1.820
	Slope	0.025	0.011	0.012	0.056	0.029	0.025
	r	0.937	0.951	0.948	0.931	0.886	0.974
Accuracy	Average Difference (mg/g)	0.095	0.802	-0.094	0.120	0.052	0.043
	SD	4.113	1.248	0.992	2.447	0.520	5.034
	t	0.16	0.08	0.64	0.33	0.68	0.06
	t (Table)	2.01					
	Precision	RSD (%)	3.37	3.46	3.94	5.21	6.94

Table 5.10 Results of the validation of all the NIRS models

The accuracy of an analytical method is evaluated by how close of the test results obtained by that method to the true values and the accuracy of the method should be established across its range. To establish the accuracy of the NIRS models, a paired *t*-test was performed between the reference method and NIRS model to check whether the results between them were significantly different or not. In our study, all samples in each model were executed in the paired *t*-test. Before the paired *t*-test, an F-test ($\alpha = 0.05$) was made, to verify whether any significance differences between the values obtained from two methods being used, for all samples. As all the developed models present statistically similar variances, the paired *t*-test was performed. The acceptance criteria is defined by the formula: $|t_{\text{exp}}| \leq |t_{\text{tab}}|$, and the results for a t_{tab} value of 2.01 for $P = 0.05$. The results obtained can be found in Table 6. $|t_{\text{exp}}| \leq t_{\text{tab}}$ with being obtained for all the samples in all the NIRS models, this clearly showed that the accuracy of all our NIRS models are comparable to the reference method.

Precision is the degree of agreement among individual test results when the method is applied repeatedly to a homogeneous sample. It is usually measured by the repeatability. In this work, we did it by analyzing a single sample three times on the same day in this work. Results for these calculations are listed in Table 5.5. The highest relative standard deviation (RSD) for jatrorrhizine was 6.94%, and but still lower than the widely accepted tolerance of 10% for this type of non-destructive determination.

5.6 Conclusion

The NIRS methods developed in this work allows simultaneous quantification of five individual alkaloids and total alkaloids content of the *Coptidis Rhizoma* samples accurately, precisely and expeditiously with minimal sample treatment. These promising results were also validated by specificity, linearity, accuracy and precision under the ICH, EMEA and PASG guideline. The prediction ability GA-PLS was better than the ordinary PLS. The number of variables considered in model decreased significantly by using GA-PLS. Some variables carrying useless information were eliminated by GA. The result obtained using GA-PLS, suggested that markers with low content in CHM could be predicted by NIRS models. The detection limit of NIRS was estimated to be 0.1%. We believe that the GA-PLS technique can be used to significantly enhance the applicability of NIRS in quality control and assurance of CHM.

**Chapter 6: Modeling the Biological
Activity of CHM: Predicting the
Anti-Oxidant Effect of *Ganoderma***

6.1 Introduction

As herbal supplements became more and more popular, people are more concerned about the quality of herbs. Conventional analytical approach only looks at the contents of one or a few selected compounds. However, the content of the marker components cannot fully represent the consistency of chemical composition in the herb, not to say to reflect their biological activities. Correlation of biological effects of CHM to its chemical compounds is not an easy task since the effects could be contributed by a group or class of compounds, while conventional analytical methods only examine a few of them. Previous work suggested that NIRS is capable of predicting content of some compounds in the herb. Here, we examine the possibility of predicting the biological activity of an herb using NIRS. This made the quality control scheme more relevant to the final goal of assuring the efficacy of the herbs or herbal products.

In this chapter, the GA-PLS technique was applied to establish correlation models of NIR spectra of *Ganoderma* with the triterpenoids content and the anti-oxidant effect of the sample. It is a famous CHM for its nourishing actions. The challenges in this study are the complex composition of *Ganoderma*, low content of marker components and small number of samples collected (n=33). *Ganoderma* contains hundreds triterpenoids which makes separation of all triterpenoid very difficult, and the contents of some triterpenoids are very low (0.02%), which is far below the detection limit of NIRS technique. For the same reasons, the reference method for the identification and quantification of the target triterpenoids of *Ganoderma* was Liquid chromatography coupled with mass spectrometry (LC/MS), so that the separation efficiency and the accuracy were monitored by checking the purity and the mass to charge ratio (m/z) of the corresponding peak of interest. The correlation of the antioxidant activity of

Ganoderma with triterpenoids content as well as with the NIR spectra was studied.

6.2 Background of *Ganoderma*

According to the Chinese Pharmacopoeia 2010, the species used for *Ganoderma* are *Ganoderma lucidum* (Leyss. Ex Fr.) Karst. or *Ganoderma sinense* Zhao, Xu et Zhang, the former is known as red *Ganoderma* and the latter is known as purple or black *Ganoderma*. Both forms are generally sold dried and whole²⁹¹.

Ganoderma is can be found naturally in temperate and subtropical zones. It is a fungal plant grows on dead or dying tree in areas that are lack of sunshine, fertile and high humidity. Nowadays, the supplies of *Ganoderma* are mainly from cultivation in Jiangxi and Zhejiang and other provinces of China.

Ganoderma is sweet and bitter in taste, neutral in nature. The tradition function of *Ganoderma* is to replenish Qi, to ease the mind, to relieve cough and asthma, and it is recommended for dizziness, insomnia, palpitation, and shortness of breath²⁹². Nowadays, *Ganoderma* are even more widely used. They are used in treating stress, allergies, cancers, hypertension, heart diseases, high cholesterol (LDL), diabetes, headache, stomachache, arthritis, Chronic Fatigue Syndrome (CFS) and hepatitis.

It has been a long history that Asians believed in the miracle health effects of *Ganoderma*. More recent reports on its consumption, mostly from Asia with some from North America and Europe also, support some of the ancient health claims. Quite a number of studies concerning the chemical constituents of *Ganoderma* were reported. Nowadays, a variety of commercial *Ganoderma* products are available in various forms, such as powders, dietary supplements, and tea. They are produced from

different parts of the mushroom, including mycelia, fruit bodies, and spores.

The fruiting body of *G. lucidum* is annual, persisting many months, growth form highly variable, specimens from North America typically large and shelf-like while those from the tropics and wild are usually smaller with cap and stalk. Intermediate forms exist, including the rare antlered form. The cap is circular to semicircular, fan- or kidney-shaped, 2-20cm board, 4-8cm thick. Upper surface smooth or with concentric ripples, corky, tough, with or without varnished appearance; dark red to reddish-brown or reddish-black in center, and yellowish toward the margin, edge is white when actively growing, flesh yellowish-brown to dark brown. The cap of fresh *G. lucidum* is soft, moist, and leathery, corky and tough when it is old. The cap becomes woody when dry. The pores are presented in one layer of spore-producing tubes, each tube 0.02-0.2 cm long, spores minute, dense, whitish when fresh, bruising brown. The stalk is dorsally or laterally attached, 3-14cm long and 0.5-4cm thick, may be enlarged at base, same color and appearance as cap surface.

The fruiting body of *G. sinense* is annual, persisting many months; stipitate. The cap is circular to semicircular, fan- or kidney-shaped, 2.5-9.5 cm board, 0.4-1.2 cm thick; upper surface smooth or with concentric ripples, corky, tough, with or without varnished appearance; purplish-black to black, margin narrow, edge pale brown when actively growing, flesh yellowish brown to dark brown. The cap of fresh *G. sinense* is soft, moist, and leathery, corky and tough when it is old. The cap becomes woody when dry. The pores are also presented in one layer of spore-producing tubes, each tube 0.3-1cm long, spores 5-6 per mm, off-white or light to dark brown. The stalk is dorsally or laterally attached, 7-19 cm long, 0.5-1 cm thick, often twisted, may be

enlarged at base, same color and appearance as cap surface.

Other species like *G. tsugae* Murrill and *G. applanatum* sometimes are being traded as counterfeit of *Ganoderma*. The cap surface of *G. tsugae* is quite similar to that of *G. lucidum*; however, its flesh is white, usually has a stalk and confined to conifers. *G. applanatum* is easily distinguished from *G. lucidum*. The fruiting body is shelf like, without a stalk, and is fan-shaped or semicircular in outline. The cap is 5-75 cm broad and 2-20 cm thick, with a hard woody surface that is gray to brown and is not varnished.

6.3 Chemical and biological studies of *Ganoderma*

6.3.1 Chemical studies

Ganoderma contains 1.8% ash, 26-28% carbohydrate, 3-5% crude fat, 59% crude fiber, and 7-8% protein²⁹³. Polysaccharides and triterpenoids are the two major active constituents of *Ganoderma*. Polysaccharides (β -Glucan) and triterpenoids in *Ganoderma* and other species have been investigated thoroughly. One of the characteristic of the fruiting body of *Ganoderma* is its bitterness taste. The components responsible for the bitterness is the triterpenoids which can also serve as a marker for pharmacological evaluation²⁹⁴.

In this study, we focused on the triterpenoids which are the bitter components of *Ganoderma* with well known pharmacological activities²⁹⁵. Triterpenes are naturally occurring compounds which has a basic skeleton of C30 (see Fig.6.1). Since the first isolation of two bitter triterpenoids, ganoderic acid A and B, from the dried epidermis of *Ganoderma* in 1982²⁹⁶, more than 130 oxygenated triterpenoids (mostly lanostane-type triterpenes) have subsequently been isolated from the fruiting bodies, spores, mycelia and culture media of *Ganoderma*^{295, 297}.

The predominant triterpenoids are ganoderic acid A-Z. Other triterpenoids include ganoderal A and B, ganoderenic acid A-D, and ganoderic acids, etc²⁹¹. The contents of triterpenoids vary significantly in different parts of the fruiting body. Quantitatively, the cap is the richest source of triterpenoids, followed by the stem, and then the spores. The outer side of the outer layer of the cap yields a higher concentration of triterpenoids than the other parts of the cap. Four ganoderic acids, ganoderic acid A,

C2, F and H (their chemical structures are shown Fig. 6.1) in *Ganoderma*, were commercially available and served as marker compounds in this study.

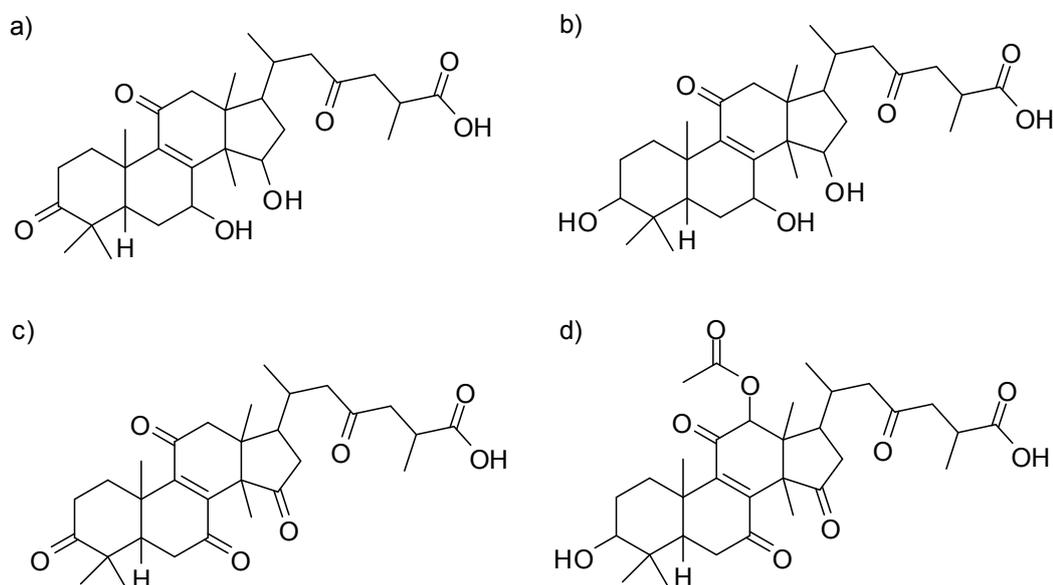


Fig. 6.1 Structures of a) ganoderic acid A b) ganoderic acid C2 c) ganoderic acid F and d) ganoderic acid H

Several analytical methods including TLC^{298, 299}, HPLC-DAD³⁰⁰⁻³⁰² and HPLC-MS^{303, 304} have been reported for the analysis of the triterpenoids profile of *Ganoderma*. The TLC methods for differentiation of *Ganoderma* from several other fungal crude drugs, such as hoelen, omphalia and polyporus, has been reported by Kohda et al.²⁹⁸. It was found that ganoderic acids B and C are unique constituents of *Ganoderma*, and these substances showed characteristic bright red spots on a TLC plate after spraying with H₂SO₄ and followed by heating. Similarly, distinctive TLC patterns of various triterpenes obtained from the extracts of the fruiting bodies of *G. lucidum* and *G. tsugae* have also been reported by other researchers²⁹⁹.

Chyr and Shiao³⁰¹ reported a method using HPLC-DAD for the classification of *G.*

lucidum and related species by comparing the content and the pattern of their triterpenoid profiles. Su et al.³⁰² reported a simple and reliable HPLC method for the identification of different members of *Ganoderma*. In this study, the 64 samples collected could be divided into 18 groups based on characteristics of triterpenoids profile shown in the HPLC chromatograms, these were in good agreement with the classification based on morphological examination. To speed up the chemical analysis, Sye³⁰⁰ investigated the influence of the extraction using a small Sep-Pak C18 cartridge on the HPLC pattern of the triterpene profile of *Ganoderma*. Results showed that similar triterpenoids pattern comparing to the conventional extraction method can be obtained.

LC-MS is increasingly employed in analyzing Chinese medicine, especially those with complex chemical constituents. Thus, it has been used in characterization of the complex triterpenoids in *Ganoderma*. A total of 32 triterpenoids, including six new ones, were identified or tentatively characterized by LC/MS in negative mode. The fragmentations pathways were reported for the first time, and were implemented for the analysis of triterpenoids in *G. lucidum*³⁰⁴. A total of 31 triterpenoids were identified or tentatively characterized from rat bile after oral administration triterpenoids-enriched extract by HPLC-DAD-ESI-MS and LC-ESI-IT-TOF/MS³⁰⁵. Two triterpenoids ganoderic acid T (GA-T) and ganoderic acid Me (GA-Me), from triterpenoids-enriched extracts of *G. lucidum* mycelia were identified by LC/MS³⁰³.

6.3.2 Biological studies of *Ganoderma*

In general, triterpenes have been reported to possess significant bioactivities, such as anti-oxidation ³⁰⁶, hepatoprotection ^{307 308 309}, anti-allergy ³¹⁰, anti-hypertension ³¹¹, anti-tumor, anti-bacterial ³¹², anti-inflammatory effects ³¹³, as well as inhibiting platelet aggregation ³¹⁴, due to the inhibition of enzymes such as β - galactosidase, angiotension converting enzyme, cholesterol synthase, etc. The following summarize the major biological activities studied.

Lowering of blood sugar and pressure

Extract of *G. lucidum* has also been found to be effective in reducing the blood glucose level after two months of treatment ³¹⁵. Ganoderan B was considered to enhance glucose utilization because it increased the plasma insulin level in normal and glucose loaded mice, but did not affect the insulin binding to isolated adipocytes ³¹⁶. The hypoglycemic activity of *G. lucidum* is thus due to an increase of the plasma insulin level and an acceleration of glucose metabolism occurring not only in the peripheral tissues but also in the liver. *G. lucidum* is also effective in lowering hypertensive blood pressure. This is due to the presence of lanostane derivatives, such as ganoderic acids B, D, F, H, K, S and Y, which exhibits hypertensive activities ³¹⁷.

Anti-bacterial and Anti-viral activities

Anti-bacterial activity has been observed against Gram-positive bacteria from the basidiocarp extracts of *G. lucidum* ³¹⁸. Another study reported that seven Indonesian *Ganoderma* species inhibited the growth of *Bacillus subtilis* ³¹⁹. Yoon et al. ³²⁰

investigated the additive effect on the activity of an aqueous extract of *G. lucidum* with four known antibiotics and observed that the anti-bacterial activity increased. There are relatively few studies on extracts from the liquid cultivated mycelium. Russell and Paterson³²¹ noted that methanolic extracts of the mycelia and culture extracts of *G. resinaceum* and *G. lucidum* inhibited *B. subtilis*. *G. resinaceum* also inhibited *Staphylococcus aureus*. Low molecular weight aqueous fractions of *G. lucidum* extract strongly inhibited multiplication of HIV-1³²².

Anti-cancer effect

Polysaccharides isolated from *G. lucidum* was found to inhibit the growth of sarcoma S180 tumor in mice³²³. The protein-bound polysaccharides showed inhibition on the neutrophil apoptosis³²⁴ and phagocytosis³²⁵, and activated the immune response by stimulating macrophages and T-lymphocytes³²⁶. Lucidimol A, B, and ganoderiol F are found to be cytotoxic on sarcoma and lung carcinoma cells³²⁷⁻³²⁹. *G. lucidum* extract showed inhibition on cell adhesion and cell migration of NF-kB cell, which conceivable that it inhibit the breast and prostate cancer^{330, 331}. And it also prevents apoptosis of Erk1 cell and PKC cell, which is pheochromacytoma and hematoma cell^{332, 333}.

The effect of *Ganoderma* upon cancer cell could be summarized as below, the mushroom inhibits (a) proliferation and invasive behavior of breast and prostate cancer cells; (b) growth and induces apoptosis of breast and prostate cancer cells; (c) growth of hematoma cells; and (d) secretion of vascular endothelial growth factor suppressing angiogenesis and transforming growth factor from prostate cancer cells

^{334, 335}. In addition, it induces apoptosis of colon cancer cells ³³⁴. Stanley et al. ³³⁶ demonstrated that *G. lucidum* induces apoptosis, inhibits cell proliferation, and suppresses cell migration of human prostate cancer cells. However, the molecular mechanism(s) has not been fully elucidated.

Furthermore, an extract from *G. lucidum* was screened by Muller et al. ³³⁷ for anti-proliferative activity using a human cancer cell lines. The results indicated that *G. lucidum* had a profound activity against leukemia (HL-60, U937, K562, THP-1, NB4), lymphoma ((B- and T-ALL) and multiple myeloma cells (RPMI8226, ARH77, U266, NCI-H929) and may be a novel adjunctive therapy for the treatment of hematologic malignancies.

Immune-regulating effect

G. lucidum may have an immune-modulating effect in patients with advanced colorectal cancer ³³⁸. Further studies are needed to explore the benefits and safety to cancer patients. Pero et al. ³³⁹ report on a combination of mushroom extracts (*Cordyceps sinensis*, *Grifola blazei*, *Gr. frondosa*, *Trametes versicolor* and *Ganoderma lucidum*) into a formulation designed to optimize different modes of immunostimulatory action, and yet that would avoid metabolic anti-oxidant competition. The efficacy of the result is less than expected. However, despite this, the data were taken as strong evidence that the combination gave additive or synergistic effects to improve health. Kuo et al. ³⁴⁰ obtained supportive evidence for the effect of dried mycelium of *G. lucidum* on the enhancement of innate immune response. Lin et al. ³⁴¹ investigated the effects of two *G. tsugae* supplement products and the results suggest that supplementation diets might alleviate bronchoalveolar inflammation via

decreasing the infiltration of inflammatory cells and the secretion of inflammatory mediators into the lungs and airways.

6.4 Experimental methods and materials

6.4.1 Samples and reagents

Twenty *G. lucidum* and eleven *G. sinense* samples were collected from different provinces of China, mainly from Sichuan, Zhejiang, Fujian. Eighteen *G. lucidum* samples were purchased from local market in Hong Kong. All samples were authenticated by Dr Dawn Au (Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hung Hom, Hong Kong). The species and the geographic origin were tabulated in Table 6.1.

All *Ganoderma* samples were grinded into powder, and then passed through a 100 mesh (150 μ m) stainless sieve. The sieved powdered samples were stored in plastic containers at 20°C for subsequent studies.

Ganoderic acid A, C1, H and F with purity above 90% were purchased from Chromadex (USA). HPLC grade acetonitrile, methanol and acetic acid were purchased from Tedia (USA). The Chloroform used for sample preparation was in AR grade, purchased from Tedia (USA) also. Double deionized water was prepared by a Milli-Q water-purification system (Millipore, MA, USA).

No. of Samples	Sample Code	Latin Name	Origin
1-6	C1 - 6	<i>Ganoderma lucidum</i>	Sichuan, China
7-9	C7, 16, 19	<i>Ganoderma lucidum</i>	Zhejiang, China
10-11	C8, 15	<i>Ganoderma lucidum</i>	Guangxi, China
12	C9	<i>Ganoderma lucidum</i>	Hunan, China
13	C10	<i>Ganoderma lucidum</i>	Shanxi, China
14	C11	<i>Ganoderma lucidum</i>	Guizhou, China
15	C12	<i>Ganoderma lucidum</i>	Anhui, China
16	C13	<i>Ganoderma lucidum</i>	Jiling, China
17	C14	<i>Ganoderma lucidum</i>	Shandong, China
18-20	C17-18, 20	<i>Ganoderma lucidum</i>	Fujian, China
21-38	L1 - 18	<i>Ganoderma lucidum</i>	Local Market
39-41	P1-3	<i>Ganoderma sinense</i>	Hunan, China
42-44	P4-5, 11	<i>Ganoderma sinense</i>	Zhejiang, China
45	P6	<i>Ganoderma sinense</i>	Sichuan, China
46	P7	<i>Ganoderma sinense</i>	Guangxi, China
47	P8	<i>Ganoderma sinense</i>	Fujian, China
48	P9	<i>Ganoderma sinense</i>	Shandong, China
49	P10	<i>Ganoderma sinense</i>	Jiangxi, China

Table 6.1 The list of the *Ganoderma* samples collected

6.4.2 High Performance Liquid Chromatography coupled with Mass Spectrometry (LC/MS) analysis

0.5g sample was accurately weighed, and sonicated with 40mL chloroform in 45°C for 20 min. This extraction was repeated in twice. Then, the mixture was combined, filtered and evaporated to dryness with a rotary evaporator. Afterward, the dry residue

was dissolved in 5mL methanol and filtrated through a 0.45 μ m syringe filter.

Chromatographic analysis was carried out on a Thermo ODS hypersil column (250mm \times 4.6mm,5 μ m) at 25 $^{\circ}$ C using an Agilent 1100 liquid chromatography system, equipped with a quaternary solvent deliver system, an auto-sampler and a DAD system. The detection wavelength was 254 nm, and the UV spectra were recorded in the range of 190-400nm. The gradient elution of the mobile phase consisting of (A) acetonitrile and (B) 0.1% (v/v) formic acid is as follows: 25–30% (A) at 0–20 min; 30% (A) at 20–30 min; 30 to 40% (A) at 30–40 min; 40% (A) at 40–55min; 40 to 55% (A) at 55–60 min and 55% to 70% (A) at 60–70 min. 10 min re-equilibrium was allowed between injections. The flow rate was 1.0mL/min and aliquots of 10 μ L were injected into the HPLC.

An Agilent MSD Trap VL module mass spectrometer was connected to the Agilent 1100 HPLC instrument via an ESI interface. The LC effluent was introduced into the ESI source in a post-column splitting of 5:1. Ultra high purity helium (He) was used as the collision gas while high-purity nitrogen (N²) was used as the nebulizing gas. All the parameters were optimized by the instrument's automatic tuning procedure. The optimized parameters in the negative ion mode were as follow: ion spray voltage, 4.0 kV; the heated capillary temperature, 270 $^{\circ}$ C; sheath gas, 30L min⁻¹. Data acquisition was performed in the full scan mode form m/z 100 to 1500 for MS and with an accumulation time of 200 ms and 7 microscans was averaged per recorded scan.

6.4.3 The Ferric Reducing Antioxidant Power (FRAP) assay

FRAP assay measures the ferric reducing ability of antioxidant because the redox reaction would be involved in the reaction with ROS. Thus, the reducing capacity could reflect the antioxidant activity of the samples. In this assay, the ability of the sample in reducing ferric tripyridyltriazine complex (Fe^{3+} -TPTZ) into ferrous tripyridyltriazine complex (Fe^{2+} -TPTZ) at low pH is measured at 593nm³⁴². Fe^{2+} -TPTZ absorbs most at 593nm and the production of this complex is detected. The recorded absorbance of the ferrous ion is multiplied by the conversion factor. This conversion factor was found from the absorbance of 1000 μM of standard ferrous ion and its concentration. Thus, the recorded absorbance is converted into FRAP value which serves as the indicator of the strength of the antioxidant activity and is related to the concentration of Fe (II) ion complex formed.

It was performed according to the procedure described by Benzie and Strain³⁴². COBAS FARA II spectrofluorometric centrifugal analyzer (Roche) was utilized. Ascorbic acid standards were tested as control. 300 μL of FRAP reagent, the mixture of 20mL of acetate buffer solution, 2mL of Fe(III) solution and 2mL of TPTZ solution, was freshly prepared. The reagent blank reading was measured at 593nm. 10 μL of sample was added. The change of absorbance between final reading selected and the reagent blank reading was calculated for each sample after 4 minutes. The change of absorbance of Fe(II) standard solution was measured in parallel. The antioxidant activity was expressed in terms of FRAP values which is arbitrarily defined as the reduction of 1mole of Fe(III) ion to Fe(II) ion.

6.4.4 Near-Infrared Spectroscopic (NIRS) analysis

All NIR spectra were recorded by a NIRSystems Model XDS spectrometer (Foss NIRSystems, Silver Spring, MD, USA) equipped with a quartz halogen lamp and PbS detector. NIR spectra of the all samples were collected in the reflectance mode over the spectral region 1100 to 2500 nm with 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). 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. Each sample measurement was repeated in three times after every rotation of the cup for 120° , and the corresponding spectra were averaged.

Several data preprocessing methods, including derivation, multiplicative scatter correction (MSC), Savitsky-golay (SG), standard normal variate transformation (SNV) and detrend have been applied to all NIR spectra to minimize the interferences such as scattering effect and baseline shift, as well as to enhance the spectral differences and to smooth the spectra etc. All data analysis was performed using a VISION (3.0.1.0), FOSS Nirsystems, MATLAB (version 6.5) and Sigma Plot (version 7.0).

6.5 Results and discussion

6.5.1 Optimization of extraction condition

To extract the triterpenoids from *Ganoderma* sample, methanol^{343, 344} and chloroform^{303, 304, 345-348} are usually used as the extraction solvent. In order to find the best extraction solvent, 0.5 g of *Ganoderma* sample (C1) was sonicated with 40 mL methanol and chloroform for 60 min in twice. After evaporating the solvent, the final extractive was dissolved in 5 mL methanol. The UV chromatograms of sample C1 using the two different solvents were shown below in Fig. 6.2. The pattern of the chromatograms is similar to each other, but the entire chromatogram of the methanol extract (Fig. 6.3b) has lower response comparing to chloroform extract (Fig. 5.3a). Also, the number of peaks detected in chloroform extract is larger than that of methanol extract. Furthermore, chloroform is able to extract 5%-15% more of the ganoderic acids than the methanol as shown in the Table 6.3. Thus, chloroform was selected as the extraction solvent.

	Ganoderic acid C2 (mAU)	Ganoderic acid A (mAU)	Ganoderic acid H (mAU)	Ganoderic acid F (mAU)
Chloroform	153.1	575.8	126.1	296.1
Methanol	158.2	539	120	258.4

Table 6.2 The comparison of the peak area of the ganoderic acid C2, A, H and F in chloroform and methanol extracts

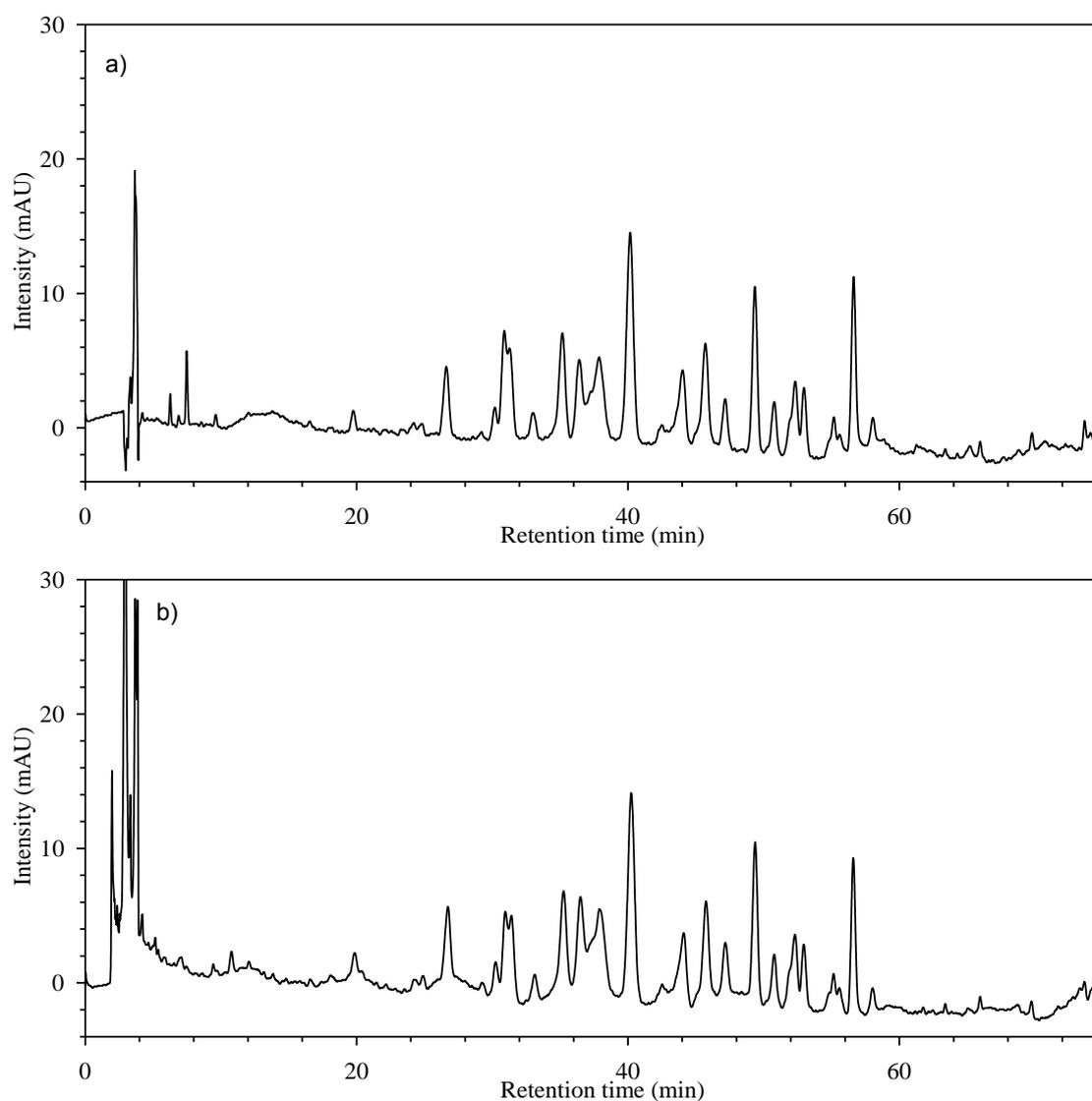


Fig. 6.2 The HPLC-DAD chromatogram of *Ganoderma* (C1 sample) extracted with a) chloroform, and b) methanol

After selecting the extraction solvent, different extraction methods including sonication and heat-reflux were compared. The results indicated that both methods got similar extraction efficiencies for the four triterpenoid markers. Sonication was finally selected because it is simpler to operate. Different extraction duration, 20 min, 60 min and 120 min were also compared. The results suggested no significant difference in extraction efficiency of 20min, 60 min and 120 min extraction and the chromatograms were similar to each other. That explains why most previous studies^{304, 345, 348} used 20

min for extraction as well.

Finally, a series of experiments were conducted to determine the most suitable volume of the solvent used and the number of successive extractions being carried out. The extraction efficiencies of the two major triterpenoids ganoderic acid A and F in *Ganoderma* samples were the criteria for optimization. From Table 6.4, 10 mL chloroform could not extract all the ganoderic acid A and F in 3 successive extractions. 20 mL chloroform was sufficient to extract them all in 3 successive extractions, while 40 mL chloroform could achieve the same purpose in 2 successive extractions. Thus, sonification with 40 mL chloroform for 20 min twice was the optimal extracting condition for triterpenoids components in *Ganoderma* samples.

Volume of Chloroform (mL)	1 st successive extraction		2 nd successive extraction		3 rd successive extraction	
	GA* A (mAU)	GA F (mAU)	GA A (mAU)	GA F (mAU)	GA A (mAU)	GA F (mAU)
10	288.6	156.1	126.6	66.27	71.5	32.4
20	433.2	231.9	116.5	61.8	32	15
40	495.8	271.3	70.1	48.3	0	0

*GA: Ganoderic acid

Table 6.3 The peak area of ganoderic acid A and F of the sample using different solvent volume in different successive extraction cycles

6.5.2 Optimization of chromatographic conditions

Triterpenoids have good absorption in about 240-260 nm, 254 nm was selected as that is the absorption maxima. For the chromatographic conditions, the influence of the stationary phase and mobile phase were firstly investigated. Triterpenoids are acidic in nature and small amount of formic acid was added to the mobile phase to reduce the ionization of these compounds. Different elution conditions with methanol-water, acetonitrile-water and different concentration of formic acid in water were examined to obtain the most suitable mobile phase. The best resolution was obtained by a combination of the acetonitrile-water/formic acid (0.1% v/v). The mobile phase used in the study differed from the earlier systems reported in the literature by replacing acetic acid with formic acid to make the mobile phase more suitable for MS study³⁴⁶.

Our studies suggested that the separation is rather sensitive to the stationary phase used. Three HPLC columns, Agilent, Sorbax C18 (250mm X 4.6mm i.d. 5 μ m) and Thermo, Hypersil ODS C18 (250mm X 4.6mm i.d. 5 μ m) and Thermo Hypersil Gold C18 (250mm X 4.6mm i.d. 5 μ m) were used to separate the triterpenoids components in the same *Ganoderma* sample and their corresponding HPLC-DAD chromatograms were shown in Fig. 6.3.

The chromatograms of the sample using Agilent Sorbax C18 column is quite similar to that produced by Thermo Hypersil ODS C18 column. Agilent Sorbax C18 column results in chromatograms of better peak shape and higher resolution, while the Thermo Hypersil ODS C18 is capable to resolve more peaks, particularly; the four markers used were well resolved. The markers ganoderic acid A and H were co-eluted as a single peak by using Agilent Sorbax C18 column while they are completely

resolved in Thermo Hypersil ODS C18. To our best knowledge, there is no paper reporting the separation of ganoderic acid A and H, even though ganoderic acid A is the major triterpenoid in *Ganoderma*. Therefore, Thermo Hypersil ODS C18 was selected for subsequent HPLC-DAD-MS studies.

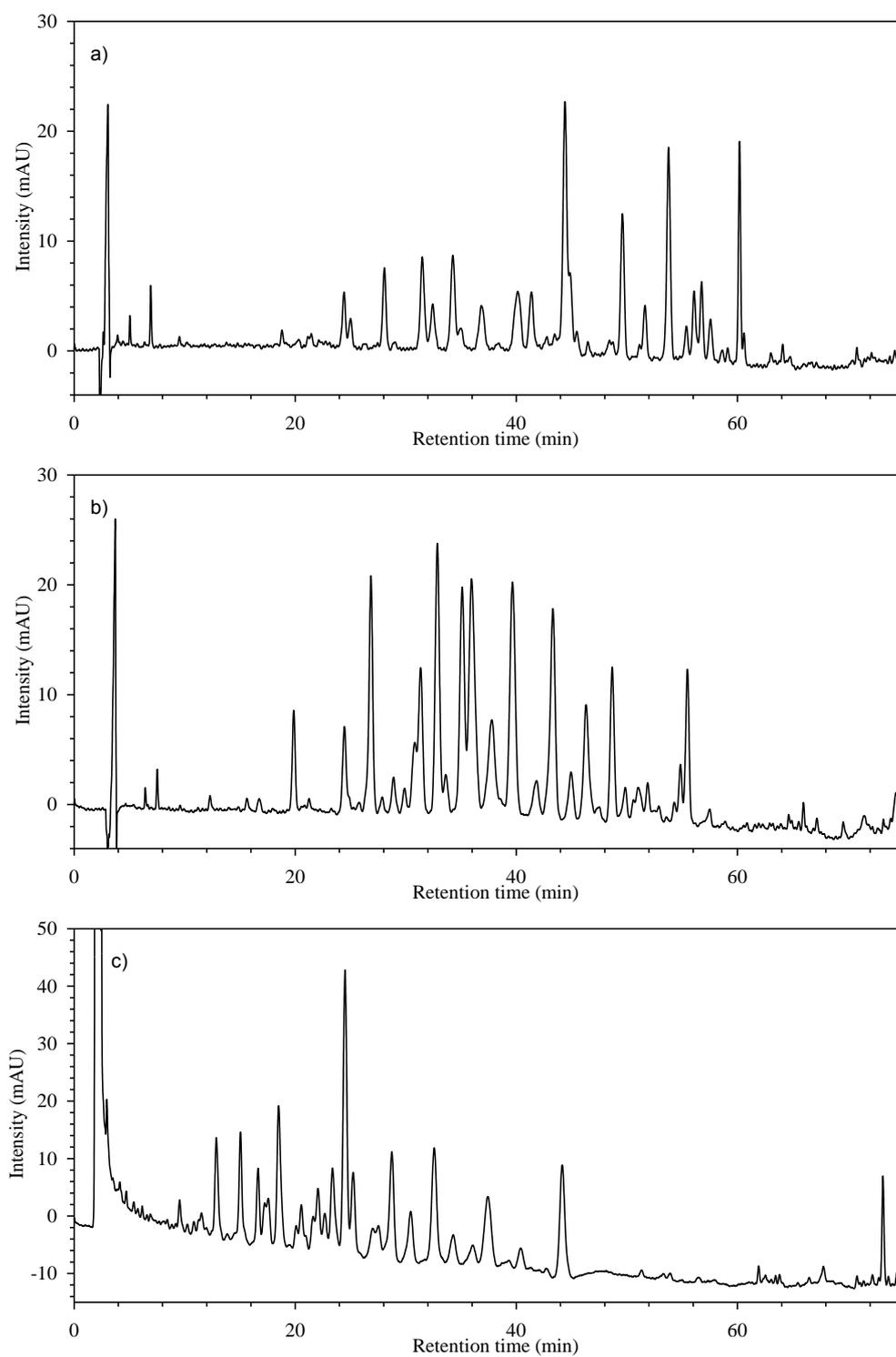


Fig. 6.3 The HPLC chromatogram of *Ganoderma* (C1 sample) using different columns. a) Agilent, Sorbax C18, b) Thermo, Hypersil ODS C18 and c) Thermo Hypersil Gold C18

6.5.3 Quantitative analysis of triterpenoids in *Ganoderma*

Fig. 6.4 shows typical HPLC-UV chromatograms of the standard solution containing four standards mixture of ganoderic acid C2, A, H and F, the extract of *G. lucidum* and the extract of *G. sinense* (type I) and the extract of *G. sinense* (type II), under the optimized chromatographic conditions. The presence of the four triterpenoids was confirmed by comparing the retention time, UV spectrum as well as the parent ions of the corresponding peak with those of the standards. The retention time of ganoderic acid C2, A, H and F are 26.4, 39.7, 43.3 and 56.0 min respectively. The respective parent ions of ganoderic acid C2, A, H and F are 500.1, 498.9, 573.0 and 570.1 m/z.

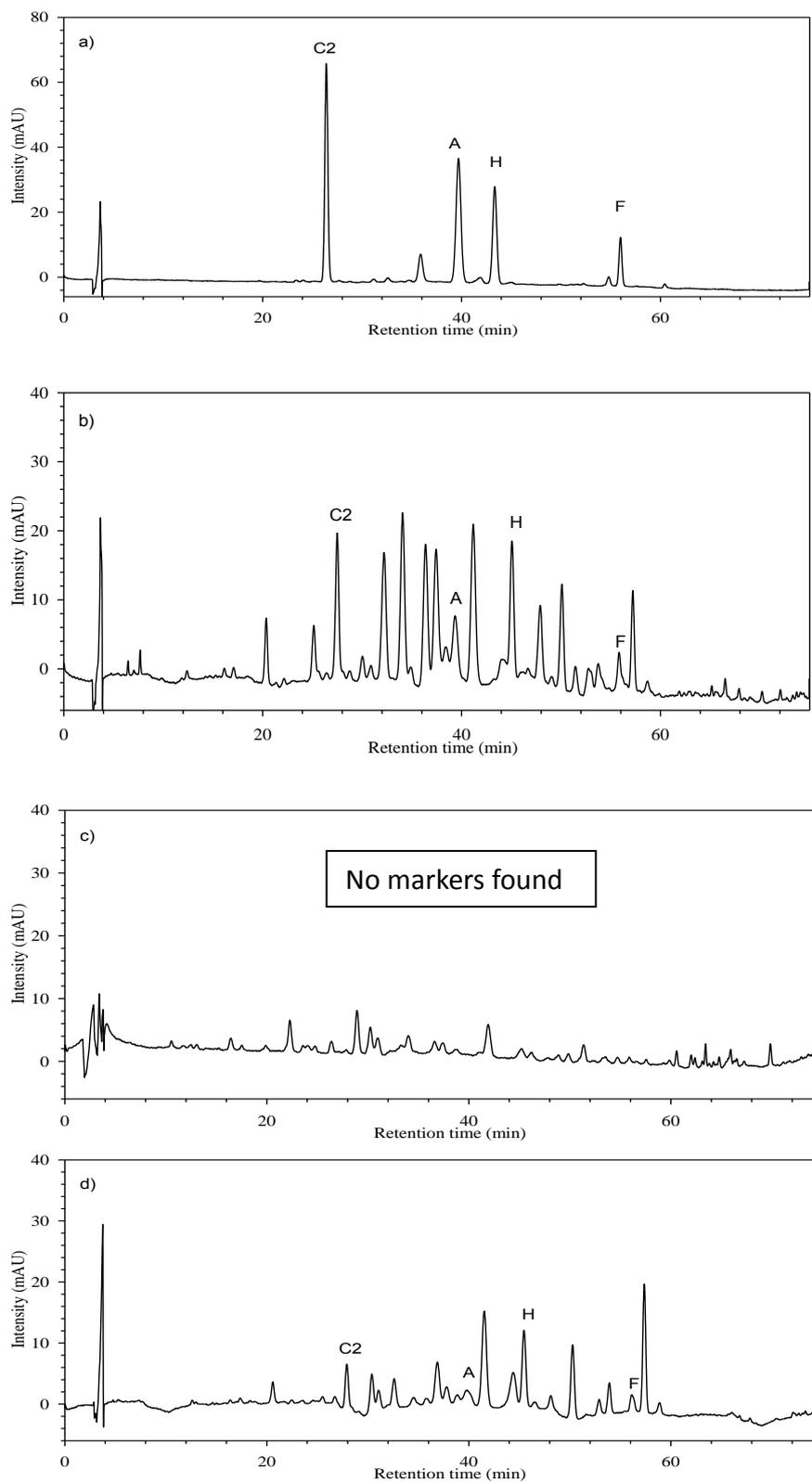


Fig. 6.4 a) The HPLC-UV typical chromatograms of the standards solution containing four standards mixture of ganoderic Acid C2, A, H and F, b) the extract of *G. lucidum*, c) the extract of *G. sinense* (type I) and d) the extract of *G. sinense* (type II)

To determine the actual contents of the four triterpenoids present in the *Ganoderma* samples, the calibrations of four triterpenoids were determined. Standard solutions of four triterpenoids were prepared and diluted to appropriate concentrations. A range of six concentrations of the four triterpenoids were analyzed by four replicate injections. The calibration curves were constructed by plotting the peak areas versus the actual amount (ng) of the triterpenoids. The contents in the samples were calculated from the linear equations of the calibration curves as shown in Table 6.5. All the calibration curves of the four standards showed good linearity ($r^2 > 0.999$) within the test ranges.

Components	Calibration curve*	r^2	Test range (ppm)
Ganoderic acid C2	$y = 4.90x - 4.38$	1.0000	10 – 300
Ganoderic acid A	$y = 4.50x - 9.66$	0.9998	10 – 300
Ganoderic acid H	$y = 2.94x - 6.51$	0.9997	10 – 300
Ganoderic acid F	$y = 3.15x - 2.20$	0.9995	5 – 200

Table 6.4 Calibration curves for four triterpenoids in *Ganoderma* where x is the content (ppm) of the standard solution and y is the peak area (mAU)

Four triterpenoids in the forty-nine samples of *Ganoderma* were quantitatively determined by the developed HPLC-DAD method and the results were shown in Table in 6.6. It was found that sixteen samples did not contain any of these four chemical markers. Upon closer inspection, five samples were found to be extracted previously while another five samples of *G. sinense* (type I) did not contained any of these four markers. The remaining six samples were further authenticated as other *Ganoderma* species, *G. duropora*, which has been reported that no triterpenoids are present. Thus, only 33 samples were further analyzed by NIRS.

	GA* C2 (mg/g)	RSD (%)	GA A (mg/g)	RSD (%)	GA H (mg/g)	RSD (%)	GA F (mg/g)	RSD (%)
C1	1.053	0.66	1.670	5.18	2.141	5.32	1.035	1.60
C2	0.422	1.30	0.833	4.97	1.282	9.79	1.498	1.91
C3	0.293	3.48	0.490	3.27	0.443	3.47	0.251	3.50
C5	0.119	5.68	0.642	2.39	N.D		0.216	3.21
C6	0.530	4.77	0.517	2.33	0.366	3.66	0.240	2.58
C7	1.828	1.57	1.415	3.72	0.741	1.48	0.538	5.88
C8	2.026	6.82	2.101	5.93	1.509	6.14	0.751	0.87
C10	1.116	1.83	1.233	2.30	0.661	1.97	0.644	1.65
C11	1.198	2.09	1.462	4.44	1.174	4.40	0.459	5.52
C12	0.223	5.35	1.097	7.02	0.977	4.20	0.734	4.23
C13	0.123	4.38	0.463	5.70	0.281	3.11	0.505	4.63
C14	0.261	5.76	1.234	3.13	0.309	4.16	0.871	0.79
C16	0.374	3.47	0.671	2.96	1.442	1.44	0.606	4.86
C17	0.341	2.87	1.248	2.50	0.829	3.64	1.082	1.08
C18	0.267	6.46	0.847	5.07	1.205	2.31	0.595	2.87
C19	0.223	9.01	0.934	4.20	1.160	4.91	0.606	4.30
C20	0.348	2.42	0.983	2.69	2.092	2.43	0.874	2.30
P3	0.282	3.46	0.909	2.53	0.779	4.92	1.113	6.17

P7	0.376	6.69	1.151	1.01	1.122	3.25	1.440	0.80
P8	0.477	3.69	0.609	7.01	0.653	3.91	0.572	1.70
P9	0.373	6.69	1.139	1.01	1.111	3.26	1.426	0.82
L1	0.348	0.01	1.705	2.26	0.754	8.95	1.096	4.47
L2	1.763	1.80	6.602	3.88	1.789	9.64	1.818	2.62
L5	N.D.		0.159	7.70	0.103	5.31	0.149	9.73
L6	0.282	3.00	0.159	0.50	0.724	2.15	0.323	9.11
L7	0.100	5.37	0.644	1.38	0.411	6.52	0.424	3.90
L8	0.586	1.27	1.427	6.25	0.159	4.81	0.294	4.76
L9	0.137	1.36	0.645	5.85	0.350	7.82	0.398	8.83
L11	0.269	4.18	1.002	1.46	0.489	4.07	0.894	7.92
L12	0.091	1.25	1.385	2.82	0.174	4.20	0.532	4.02
L13	0.443	4.87	1.327	1.59	0.818	4.20	1.105	3.35
L14	0.428	2.22	0.445	7.25	0.720	2.07	0.323	4.60
L18	0.282	0.64	2.222	1.36	0.159	5.91	0.323	2.50

*GA: Ganoderic acid

Table 6.5 The content of the four triterpenoids in the 33 *Ganoderma* samples

The average content of the four markers in the *Ganoderma* samples were tabulated in Table 6.7. It was found that the content of the markers in the local and mainland samples varied significantly. For example, the standard deviation of ganoderic acid C2 is close to the average content. In *G. sinense* samples, two types of samples were identified. Type I samples contained no ganoderic acids, which is consistent with literatures while type II samples have a similar pattern as *G. lucidum*, which contained all four ganoderic acids³⁴⁷⁻³⁴⁹. Comparing the marker contents of *G. lucidum* and *G. sinense* (type II), the triterpenoids contents in *G. lucidum* is larger than that in *G. sinense* (type II).

As shown in the Fig. 6.5, ganodeic acid A was found to be the most abundant marker in most of *Ganoderma* samples, followed by ganoderic acid H, F and C2. Comparing local and mainland samples, the ganoderic acid contents of samples collected in local market were generally lower than that in mainland. This may due to the prolonged storage of these samples.

Samples	Content (mg/g)				
	Ganoderic acid C2	Ganoderic acid A	Ganoderic acid H	Ganoderic acid F	Total triterpenoids
<i>Ganoderma lucidum</i> (China) n=17	0.653 ± 0.589	1.089 ± 0.443	1.086 ± 0.580	0.712 ± 0.399	3.108 ± 1.164
<i>Ganoderma sinense</i> (type I) n=5	ND	ND	ND	ND	ND
<i>Ganoderma sinense</i> (type II) n=4	0.380 ± 0.076	0.918 ± 0.226	0.892 ± 0.213	1.096 ± 0.372	2.063 ± 1.051
<i>Ganoderma lucidum</i> (Local) n=12	0.430 ± 0.468	1.477 ± 1.734	0.554 ± 0.469	0.640 ± 0.491	3.065 ± 2.972
All samples n=33	0.542 ± 0.514	1.209 ± 1.087	0.862 ± 0.555	0.732 ± 0.442	2.965 ± 1.983

Table 6.6 The average contents of the ganoderic acid markers in *Ganoderma* samples obtained using HPLC-DAD

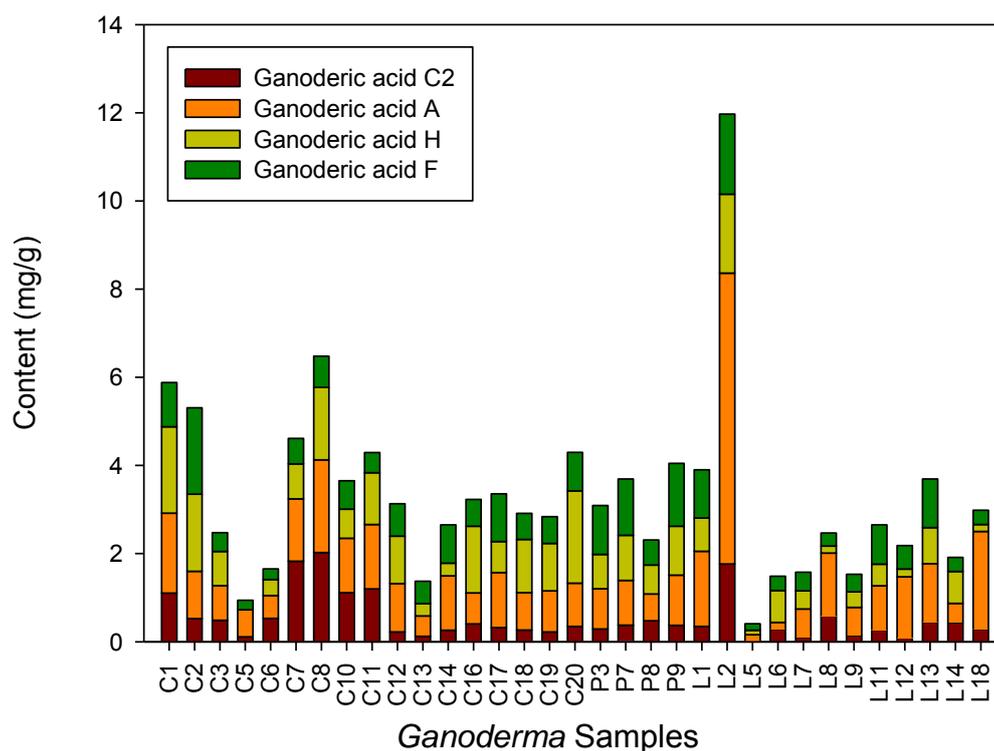


Fig. 6.5 Histogram of the four triterpenoids contents in *Ganoderma* samples obtained using HPLC-DAD.

6.5.4 Identification of triterpenoids in *Ganoderma* by LC/MS

According to literature, the characteristics of UV spectra of the triterpenoids have absorption maxima at 240–260 nm³³¹⁻³³⁴. Since the UV spectra of triterpenoids are similar to each other and the commercially availability of triterpenoids are limited, LC-MS was employed to study the chemical constituents of *Ganoderma* samples. In our work, electro spray ionization (ESI) in both negative and positive modes were used and compared for the detection of triterpenoid. Negative mode ESI was found to be more sensitive for detecting triterpenoid. Therefore, all MS data was obtained in ESI negative mode.

Fig. 6.6 shows the TIC typical chromatograms of the *G. lucidum*, *G. sinense* (type I) and *G. sinense* (type II) extracts obtained using the optimized chromatographic and mass spectroscopic conditions. A total of 19 triterpenoids (shown in Table 6.8a) were characterized from the chloroform extract of *G. lucidum*, which were tentatively identified by comparing their retention times and mass spectra with those of reference standards as well as ESI-MS spectra reported previously in the literatures^{304, 346, 349}. The MS data of the types of *G. sinense* in the obtained chromatograms are also very different from each other. By further investigating the MS data, 3 and 9 triterpenoids (shown in Table 6.8b & c) were tentatively identified from the chloroform extract of *G. sinense* type I and *G. sinense* type II, respectively.

By comparing the chemical profiles of UV and MS chromatograms of *G. lucidum* and *G. sinense*, they were different from each other and our observations are consistent with previous literatures³⁴⁶⁻³⁴⁹. Since the triterpenoids are the bioactive components in *Ganoderma* and their profiles and content varied significantly in different species of *Ganoderma*, the biological activities may also very different from each other. More works have to be done to clarify the health benefits of these medicinal mushrooms to human.

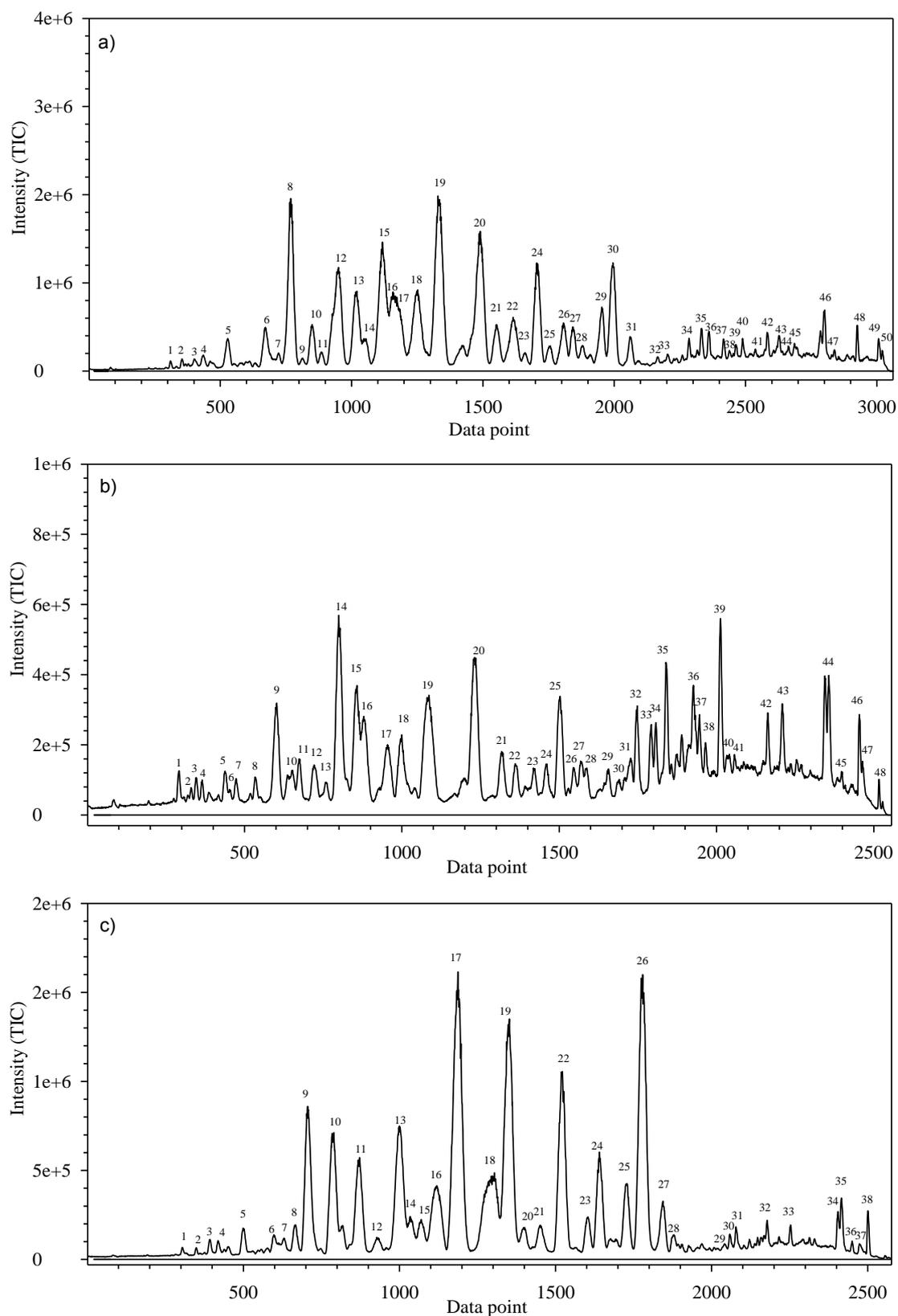


Fig. 6.6 The typical TIC chromatograms of the a) extract of *G. lucidum*, b) extract of *G. sinense* (type I) and c) extract of *G. sinense* (type II), under the optimized chromatographic and mass spectroscopic conditions.

a) *Ganoderma lucidum*

No	RT	Assigned identity	[M-H] ⁻	[2M-H] ⁻	MS ²	No	RT	Assigned identity	[M-H] ⁻	[2M-H] ⁻	MS ²
1	12.8	Unknown	533	1067		26	52.7	Lucidenic acid D	513	1027	513
2	14.5	12-Hydroxyganoderic acid C2	533		533	27	53.5	Ganoderic acid F	511	1023	511
3	16.4	Unknown	527			28	54.1	Unknown	641		
4	17.3	Unknown	529			29	55.8	Ganolucidic acid D	499	999	499
5	20.6	Unknown	531	1063		30	56.9	12-acetoxylanoderic acid F	569	1139	551
6	25.4	Lucidenic acid E	515	1031	515	31	58.6	Unknown	513	1027	
7	26.5	Unknown	571			32	61.9	Unknown	509		
8	27.7	Ganoderic acid C2	517	1035	517	33	63	Unknown	507		
9	28.8	unknown	529			34	65.1	Unknown	555		
10	30	unknown	511			35	66.3	Unknown	485		
11	31	Ganoderic acid C6	529	1059	511	36	66.9	Unknown	595		
12	32.4	Ganoderic acid G	531	1063	513	37	68.4	Unknown	483		
13	34.2	Ganoderenic acid B	513	1027	495	38	68.9	Unknown	501		
14	35	unknown	495			39	69.5	Unknown	483		
15	36.4	Ganoderic acid B	515	1031	497	40	70.2	Unknown	315		
16	37.4	Ganoderic acid AM1	513	1027	513	41	71.4	Unknown	483		
16	37.4	Ganoderenic acid K	571		553	42	73.5	Unknown	485		
17	38.3	Unknown	515	1031		43	74.4	Unknown	481		
18	39.4	Ganoderic acid K	573	1147	555	44	74.9	Unknown	297		

19	41.1	Ganoderic acid A	515	1031	515	45	77.4	Unknown	467	
20	45	Ganoderic acid H	571	1143	553	46	77.7	Unknown	467	
21	46.2	Lucidenic acid A	457	915	457	47	78.7	Unknown	469	
22	47.9	Unknown	511	1023		48	81	Unknown	279	
23	48.9	Ganoderenic acid D	511	1023	493	49	83.1	Unknown	281	
24	50	Ganoderic acid D	513	1027	495	50	83.6	Unknown	299	
25	51.2	Unknown	569	1139	551					
b) <i>Ganoderma sinense</i> (Type I)										
1	11.9	Unknown	473			25	53.2	Lucidenic acid D	513	1027
2	13.4	Unknown	473			26	54.6	Unknown	455	
3	14.1	Unknown	473			27	55.3	Unknown	511	
4	14.7	Unknown	473			28	55.9	Unknown	443	887
5	17.7	Unknown	455			29	58.3	Unknown	501	1003
6	17.1	Unknown	471			30	59.5	Unknown	339	
7	18.9	Unknown	471			31	60.7	Unknown	497	
8	23.8	Unknown	457			32	61.2	Unknown	557	
9	24.9	Unknown	473			33	62.6	Unknown	499	
10	25.4	Unknown	473			34	63	Unknown	555	
11	26.1	Unknown	461			35	64	Unknown	499	999
12	27.8	Unknown	457			36	64.8	Unknown	495	
13	29	Unknown	515			37	65.3	Unknown	555	
14	30.4	Unknown	457			38	66.2	Unknown	497	995

15	32	Unknown	457			39	66.6	Unknown	553		
16	32.6	Unknown	517			40	67.2	Unknown	495		
17	35	Ganoderenic acid K	571			41	68.4	Unknown	483	967	
18	36.3	Unknown	455			42	72.5	Unknown	295		
19	39	Unknown	515			43	73.8	Unknown	485	971	
20	44	Unknown	455			44	77.6	Unknown	467		
21	47	Ganoderenic acid D	511			45	77.9	Unknown	467		
22	48.5	Unknown	453			46	81	Unknown	279		
23	50.4	Unknown	569			47	83.3	Unknown	281		
24	51.9	Unknown	455			48	83.8	Unknown	299		
c) <i>Ganoderma sinense</i> (Type II)											
1	12.5	Unknown	533	1067		20	46.3	Unknown	553		
2	14.4	12-Hydroxyganoderic acid C2	533		533	21	47.9	Ganoderenic acid D	511	1023	493
3	16.1	Unknown	527			22	50.1	Ganoderic acid D	513	1027	495
4	17.1	Unknown	529			23	52.6	Unknown	553		509
5	20.1	Unknown	531	1063		24	53.7	Ganoderic acid F	511	1023	
6	23.9	Unknown	529			25	56.2	Ganolucidic acid D	567	999	499
7	25.1	Unknown	515			26	57.5	12-acetoxylanoderic acid F	551	1139	
8	26.4	Unknown	531			27	59.3	Unknown	513	1027	
9	27.5	Ganoderic acid C2	517	1035	517	28	60.3	Unknown	339		

10	30	Unknown	511			29	67.2	Unknown	551	485
11	31.9	Unknown	513			30	67.8	Unknown	595	483
12	33.8	Unknown	571	1027		31	69.3	Unknown	483	
13	36	Unknown	497	1031		32	71.1	Unknown	315	
14	37.5	Unknown	527			33	73.5	Unknown	295	
15	38.6	Unknown	555		513	34	78.6	Unknown	467	
16	39.4	Unknown	529			35	78.8	Unknown	467	
17	40.9	Ganoderic acid A	515	1031	515	36	80	Unknown	469	
18	43.9	unknown	569		511	37	80.9	Unknown	467	
19	45.0	Ganoderic acid H	571	1143	553	38	82	Unknown	279	

Table 6.7 Identification of triterpenic acids from the chloroform extract of a) *G. lucidum*, b) *G. sinense* (Type I) and c) *G. sinense* (Type II)

6.5.5 Antioxidant activity of *Ganoderma* samples

The total antioxidant capacities of the *Ganoderma* extracts collected for this work were determined by the FRAP assay, a simple speedy, inexpensive, and reproducible method, to the assay the antioxidant powers of botanicals^{350, 351}. In brief, 300 uL FRAP reagent was warmed to 37°C, $A_{593\text{nm}}$ of reagent blank was measured. 10uL sample was added, absorbance was measured at every 15s until 4 min at 593nm. The antioxidant activity of the samples were evaluated from the change of the absorbance at 593nm of the FRAP reagent at 4 min after the addition of the sample extracts. The increase in the absorbance is due to the reduction of the ferric tripyridyltriazine complex to the ferrous form. The absorbance of the ferrous standard solution without the introduction of the sample extract was measured as control.

The FRAP value of all methanolic *Ganoderma* extracts were shown in Fig.6.7 and Table 6.9. The antioxidant activities varied significantly among 33 samples (range from 88.3 to 747.7). The methanolic extracts were used to avoid the influence of polysarrcharide in the *Ganoderma* samples. All FRAP experiments were performed in duplicate and the RSD of the FRAP value of almost all *Ganoderma* extract are less than 5%. The average antioxidant activities of *G. lucidum* (mainland), *G. sinense* (type II) and *G. lucidum* (local) are $211.1 \pm 84.4 \mu\text{mol/L}$ (n=17), $223.7 \pm 67.8 \mu\text{mol/L}$ (n=4) and $244.4 \pm 174.7 \mu\text{mol/L}$ (n=12) respectively. There is no significant difference observed among them.

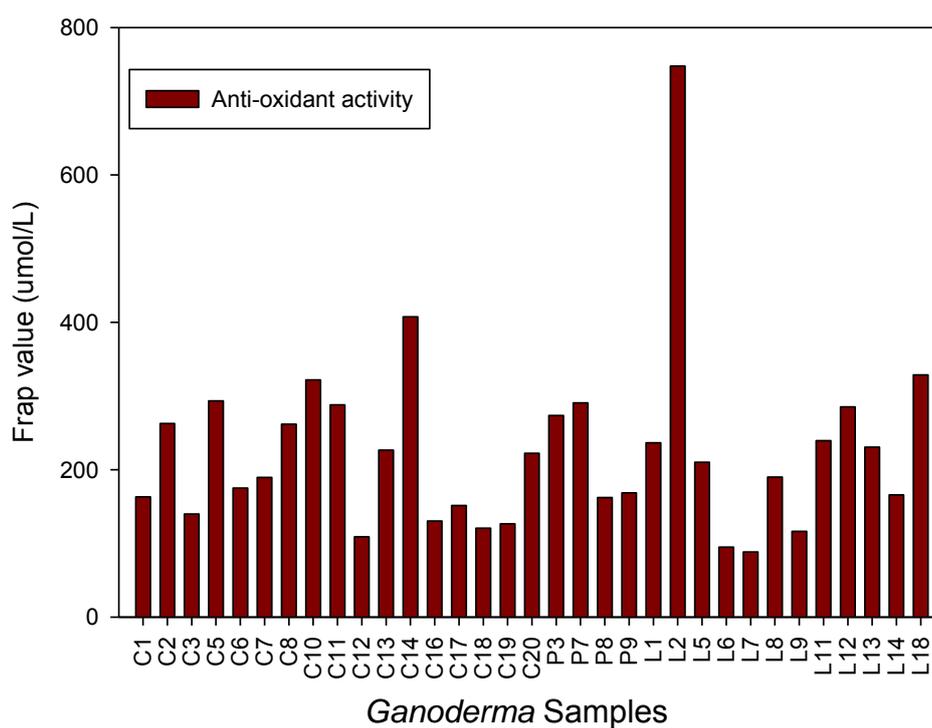


Fig. 6.7 Results of antioxidant activities in *Ganoderma* samples determined by FRAP assay

Sample	FRAP value (µmol/L)	Sample	FRAP (µmol/L)
C1	163.19 ± 7.98	P3	273.47 ± 3.13
C2	262.66 ± 1.42	P7	290.69 ± 5.99
C3	139.84 ± 2.58	P8	162.20 ± 17.83
C5	293.45 ± 8.03	P9	168.57 ± 1.26
C6	175.03 ± 0.68	L1	236.52 ± 4.70
C7	189.67 ± 0.19	L2	747.73 ± 6.69
C8	261.76 ± 28.8	L5	210.03 ± 0.03
C10	321.94 ± 9.95	L6	94.95 ± 9.17
C11	287.93 ± 0.44	L7	88.32 ± 6.31
C12	108.96 ± 2.27	L8	189.97 ± 13.26
C13	226.52 ± 3.01	L9	116.30 ± 8.32
C14	407.34 ± 1.88	L11	239.49 ± 0.82
C16	130.34 ± 3.55	L12	285.38 ± 1.57
C17	151.30 ± 5.24	L13	230.56 ± 9.53
C18	120.78 ± 4.60	L14	165.93 ± 1.74
C19	126.58 ± 2.97	L18	328.71 ± 6.30
C20	222.21 ± 5.05		

Table 6.8 The results of the antioxidant activity of 33 *Ganoderma* samples determined by FRAP assay

The correlations between the FRAP value of *Ganoderma* methanolic extracts and the total triterpenoids contents were shown in Fig. 6.8 and the Pearson correlation coefficients is 0.6874. Moderate correlation was observed between sum of 4 triterpenoids and the antioxidant activity. The possible reason is that the four triterpenoids may only take minor role in contributing the antioxidant effect, while besides triterpenoids; there are other compounds which may contribute to the antioxidant activity of *Ganoderma*. Phenolic compounds are possible candidates that contribute to the antioxidant activities. A recent paper reported³⁵² that methanolic extract of *G. lucidum* contained phenolic compounds, which are two phenolic acids, p-hydroxybenzoic and p-coumaric acids, and one related compound, cinnamic acid, were observed by HPLC-DAD-MS.

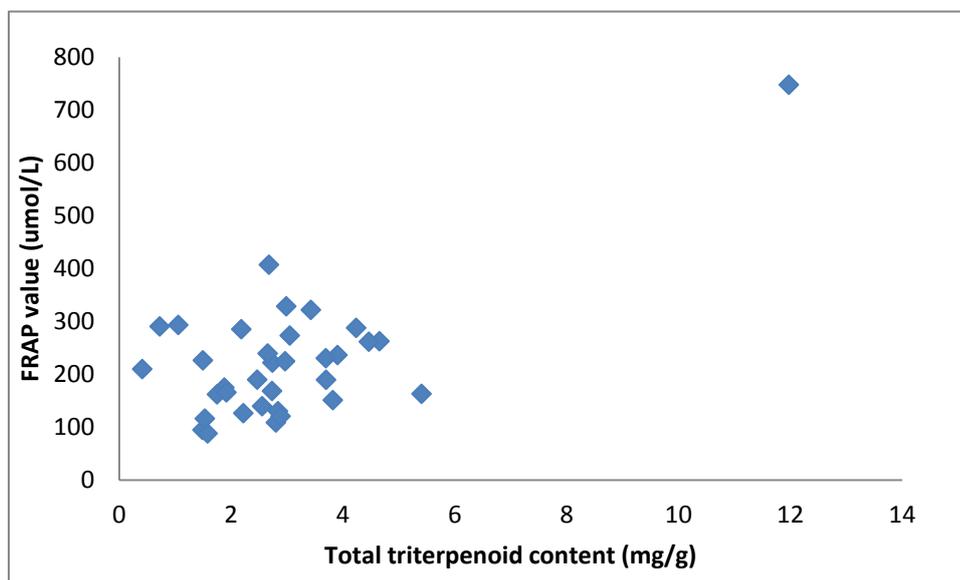


Fig. 6.8 A plot of total triterpenoid content against the antioxidant activity (FRAP value) of the 33 *Ganoderma* samples

6.5.6 Calibration model for triterpenoids content using NIRS

The contents of triterpenoids are much lower than 1%. Thus, GA-PLS developed in this work was used to select relevant spectral region for the models. The parameters shown in the Table 6.10 were used for the selection. All NIR spectra data were utilized as a gene. As described in previous chapter, the samples were split randomly into three equal groups and a full cross-validation was then performed and spectral regions were chosen according to selection frequency. The spectral variables with selection frequency equal to or higher than 1 time in the 3-fold cross-validation step, were used to build the final calibration model. The performance of the final model was evaluated by RMSECV, RMSEV, BIAS, SEV and R.

Population size: 30 chromosomes
Average length of initial chromosome: 20
Maximum length of the chromosome: 120
Probability of mutation: 1%
Maximum number of PLS components: 15
Number of runs of variable selection: 100
Window size for smoothing: 3

Table 6.9 The parameters of the GA-PLS used

Typical NIR spectra of *Ganoderma* samples are shown in Fig. 6.12. The NIR spectra of different types of *Ganoderma* are similar to each other. The intensive bands in these spectra are from the first overtone O-H stretching (1450nm), first overtone of C-H stretching in CH₃ (1730nm), combination band of O-H stretching and HOH deformation vibration (2100 nm) followed by the combination band of C-H stretching and C-H bending in CH₃ (2315nm) ³³⁷. Those vibrations mainly come from constituents such as phenolic compounds, triterpenoids, and polysaccharides.

As shown in Fig. 6.9, the raw spectra have baseline shift with broad bands. Therefore, these NIR spectra have to be preprocessed first to remove baseline shift, noise and backgrounds before data analysis. Here, all the 33 NIR spectral data were treated by using a combination of pre-processing tools including SNV, second derivative with the Savitsky-golay (SG) method, moving window average smoothing and MSC treatment. Then, the preprocessed spectra were used in PLS regression.

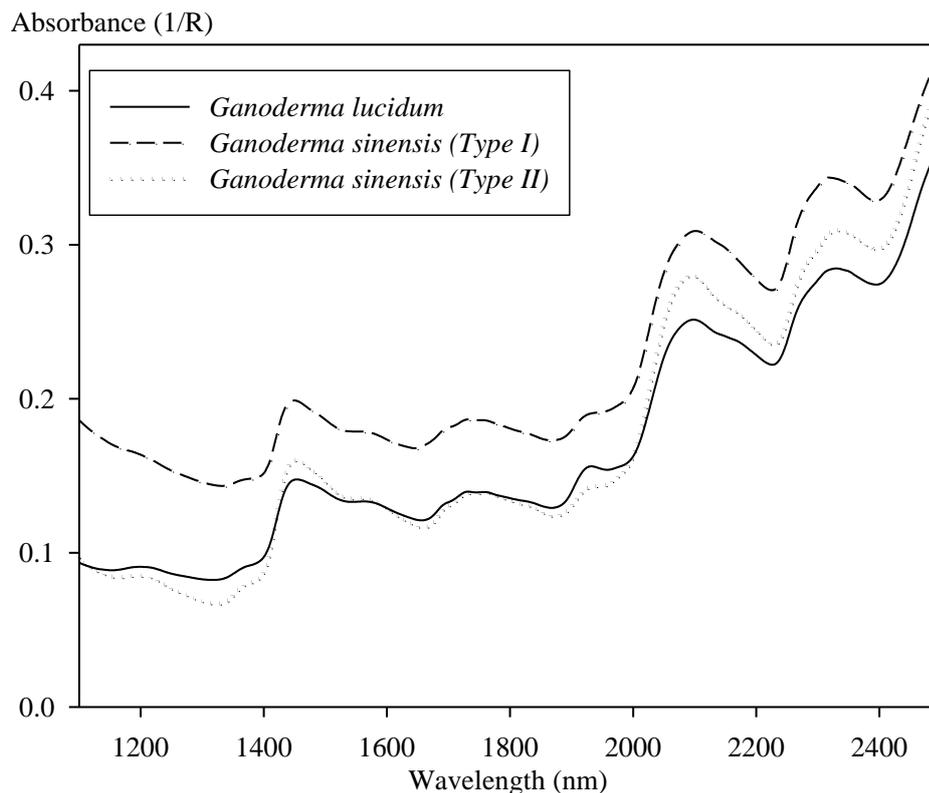


Fig. 6.9 Typical NIR spectra of a) *G. lucidum* (—), b) *G. sinense* (Type I) (- - -) and c) *G. sinense* (Type II) (.....)

In this work, the total triterpenoid content are described as the sum of the contents of the four standards (Ganoderic acid A, C₂, F and H). Quantitation models for ganoderic acid A, C₂, F, H and total triterpenoids content were then established from the NIR spectra.

Fig. 6.10 shows the correlation between the NIR predicted values and the reference values obtained by RRLC-DAD measurement. The dot and cross symbol represent the calibration and test data points, respectively. It can be seen from these plots that the

data in the calibration and test sets are highly correlated with the reference data. For the validation set, the SEV of the quantitation models of ganoderic acid A, C₂, F, H and total triterpenoids content being 0.232, 0.091, 0.151, 0.228 and 0.601, respectively, and the relative error of those models are 19.18%, 16.78%, 20.63%, 26.45% and 17.96%, respectively. The R values of the test sets of the ganoderic acid A, C₂, F, H and total triterpenoids content models are 0.976, 0.984, 0.937, 0.908, and 0.949, respectively. This indicates that the predicted values from NIRS are consistent with those determined by HPLC-DAD.

To investigate the detail of GA-PLS modeling, the variables used for GA-PLS modeling were examined. 70, 173, 171, 109, and 197 variables were finally selected as wavelength regions, after 100 iterations of genetic algorithms, and were used to build the final quantitation PLS regression models for ganoderic acid A, C₂, F, H and total triterpenoids content, respectively. Comparing to the detection limits of previous NIRS quantitation models, isoflavonoids content of *Puerariae Radix* and alkaloids content of *Coptidis Rhizoma*, these *Ganoderma* triterpenoids models were demonstrated to give reliable prediction for content as low as 0.02% of the content analyzed by using GA-PLS.

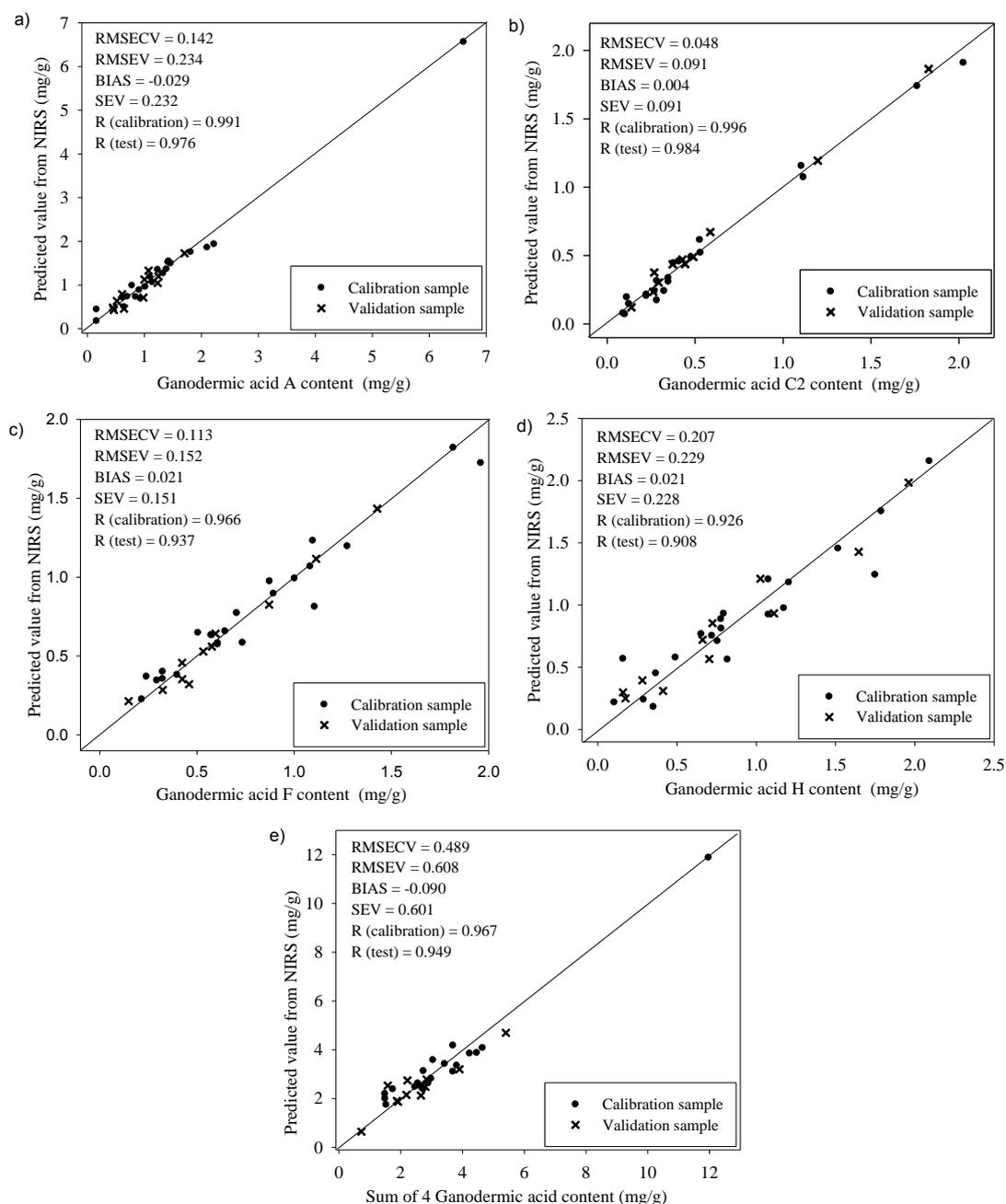


Fig. 6.10 Correlation plot of NIRS predicted value against HPLC-DAD measured content of a) ganoderic acid A, b) ganoderic acid C₂, c) ganoderic acid F, d) ganoderic acid H and e) total triterpenoids content (sum of four ganoderic acids) in *Ganoderma* samples (n = 33)

6.5.7 Calibration model for anti-oxidant activity

Fig. 6.11 shows the correlation between the NIR predicted values and the experimental FRAP values. It can be seen from this plot that the data in the calibration and validation sets are highly correlated with the antioxidant data. For the validation set, the SEV of the quantitation models of FRAP value being 42.15, with the relative error of those models are 18.72%. The R values of the test sets of the FRAP value models are 0.936.

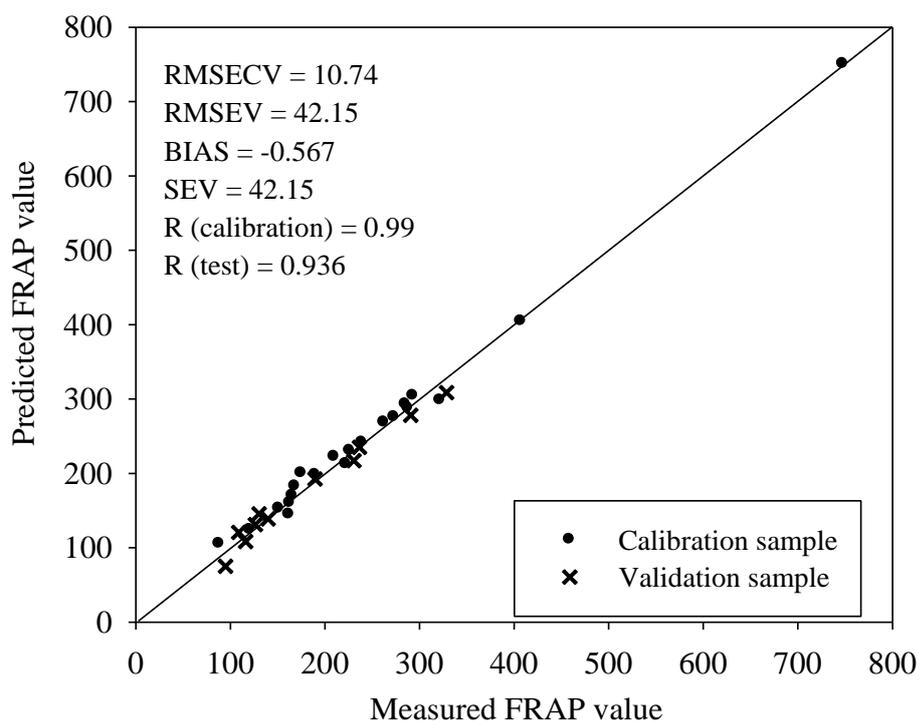


Fig. 6.11 Predicted activity by NIRS against antioxidant activity measured by FRAP assay for *Ganoderma* samples (n = 33)

To investigate the detail of GA-PLS modeling, the variables used for GA-PLS modeling were examined. 252 variables (less than 10% variables) were selected after 100 iterations of genetic algorithms to build the final antioxidant activity models with PLS regression. These variables are indicated in Fig. 6.12. This is the first time that the NIRS technique is employed to build the correlation model to predict biological activities of herbal extract. The strong correlation suggested that the NIRS technique can be used to estimate biological activity as well as chemical composition of the herbs and products reliably. This could be an important improvement to the quality control as the biological effect of herbs and herbal products is being monitored. Modeling other biological activities may be more difficult which may required more sophisticated algorithm and some of the biological effects may not be linear as well. Yet, this is certainly worth the efforts for more investigations and examinations.

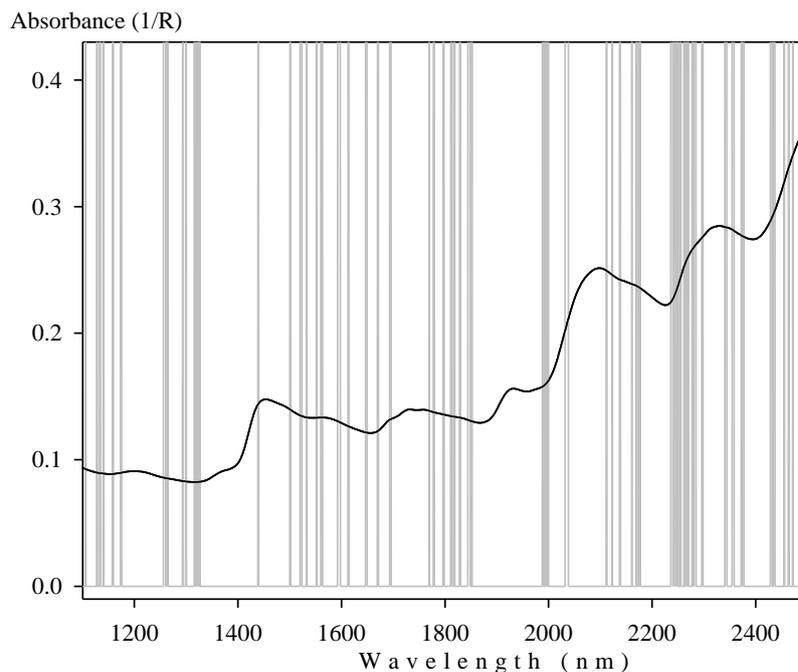


Fig. 6.12 Overview of the final spectral variables selected by GA for FRAP model

6.6 Conclusion

The results successfully demonstrated the feasibility of using NIRS for quality control of *Ganoderma* in quantification of constituents as low as 0.02% and antioxidant activities of the methanolic extract simultaneously, with the use of genetic algorithm for variable selection before the PLS regression. Moderate correlation was also found between the total triterpenoids content of *Ganoderma* and their antioxidant activities measured by FRAP method. This technique, NIRS coupled with the GA-PLS is capable of building reliable model predicting the antioxidant activities of *Ganoderma*. From this point of view, the NIRS could be an important analytical technique for determination of contents and biological activities of CHM.

Chapter 7: Conclusion

Conventional approach on the quality control of CHM relies on chromatographic techniques, which are usually unable to give instantaneous assessment on the quality of the CHM being examined. Moreover, current practice of detecting one or two marker compounds as a mean to identify the herbs is not sufficient to give a holistic quality assessment. Spectroscopic technique is one way to develop a rapid and holistic characterization of the chemical composition of the herb. Among different spectroscopic techniques, Near-Infrared Spectroscopy (NIRS) has the advantage of being rapid and non-destructive. The aim of this work is to examine the feasibility of NIRS for analyzing secondary metabolites which usually present at level of below 1% to several % in herbs, and to develop systematic experimental and data analysis procedures for developing robust analytical methods. In this work, NIRS was used to evaluate the quality of three CHM, *Puerariae Radix*, *Coptidis Rhizoma* and *Ganoderma*, in terms of differentiation of the species, prediction of markers content and biological effect.

Puerariae Radix was studied first because of relative high content of the markers. In this work, NIRS has been demonstrated to be capable of predicting major marker contents which are higher than 1%, and identifying the species of multi-species CHM. The successful classification model developed by SIMCA demonstrated a clear

discrimination between the two species of *Puerariae Radix* without any mistake. The results may be associated with the large difference in the isoflavonoid contents of these two species. The strong correlation of the contents of two isoflavonoids markers puerarin (0.97), daidzin (0.912) and the total isoflavonoids (0.969) with the NIR spectra of the testing samples suggested that CHM with components of content higher than 1% could be determined accurately and rapidly by NIR quantitation models. In order to avoid the bias on sample selection in the model development, all samples were divided into three groups by the Kennard-Stone algorithm. An independent testing set was used to validate the model developed, which make the models more robust.

However, CHM with marker contents higher than 1% are not common; most of them contain chemical components with lower contents. To extend the application of NIRS in chemical analysis and quality control of CHM, a variable selection procedure based on Genetic Algorithm (GA) has been developed for building quantitation models using NIR in the study of *Coptidis Rhizoma*. Compare to ordinary PLS, the GA-PLS greatly reduced the Square Error of Validation (SEV) of the prediction models by 14-27%. The sensitivity of the NIRS analysis was enhanced tremendously; detection of minor components in CHM with content higher than 0.1% is capable with the use

of GA-PLS in quantitation model development. The results proved that GA is useful in mining appropriate wavelength regions with relevant information, thus, building robust quantitative models for minor components in CHM even when the sample size is small.

Finally, we have also demonstrated that NIRS could be used to predict biological activities of the CHM samples. The anti-oxidant activity of *Ganoderma* was found highly correlated ($R=0.936$) with the NIR spectra with 18.72% relative error of prediction. The results suggested that NIRS is a better technique for correlation of biological activities of CHM than other approaches like correlating the antioxidant activity with marker contents (0.687). Since NIRS examines the CHM as a whole, the principle of NIRS is in line with the characteristics of CHM of multi-target and synergistic action due to different chemical components present. Biological activities may have complex mechanisms which involve different compounds, which identities have not revealed yet. Therefore, it is better to mining the relevant information corresponding to different chemical compounds from NIR spectra, as spectra of the herbs were taken contribution from all compounds. Furthermore, the correlation coefficient of ganoderic acid A (0.976), C2 (0.984), F (0.937), H (0.908), and total triterpenoids (0.949) suggested that the GA-PLS algorithm is capable of correlating

marker of content lower than 0.05% with the NIR spectra of the CHM. Additionally, the carefully selected chromatographic condition in this study results in clear separation of four ganoderic acid A, C2, H and F. This is the first chromatographic condition that fully resolved the ganoderic acid A and H while were usually co-eluted in previous reports.

This study clearly demonstrated that NIR spectroscopy with suitable pre-processing and variable selection methods could be used for identification of the CHM, quantitative analysis of the selected chemical compounds in the CHM as well as evaluation of related biological effects. The GA-PLS algorithm developed widens the scope of the application of NIRS in quality control of CHM as the sensitivity was tremendously enhanced by ten times. As a conclusion, NIR spectroscopy could be an effective tool for rapid quality control of CHM.

The NIR spectroscopy requires no or little sample preparation and is non-destructive.

The measurement is fast and can be performed in less than a second, which make it an ideal tool for on-line applications. Several constituents of the sample can be measured at the same time. However, there are some drawbacks hindering the development of NIRS for quality control of CHM.

The NIR signal is non-specific, so information has to be retrieved by using proper chemometric techniques. However, the use of chemometric models is not a trial task³⁵³ and this is one of the barriers prohibiting NIR spectroscopy in the analysis of CHM. A professional personnels are usually needed to develop chemometric models from the NIR data. Most importantly, unlike western drugs or food, the chemical constituents in CHM are very low in content (<1%), and the matrix effect contributing from other constituents in samples makes it even more difficult to develop reliable prediction models. The systematic approach presented in this study for model development, using NIR spectroscopy combined with GA-PLS data analysis can enhance the sensitivity, robustness and repeatability of the models developed, and open up the possibility of on-line quality control of CHM in manufacturing process. The developed techniques could be transferred into a user-friendly software package. After a proper training, all chemistry graduates should be capable to operate the NIR device and develop simple prediction models.

Collection of sufficient samples for model development is the major difficulties we encountered in this study. The techniques certainly receive much more attention if more industrial collaboration can be established to apply the technique to more samples. Samples used in our study were collected from different geographical origins,

the NIR technique will perform even in a particular manufacturer as they are sourcing materials from selected suppliers only.

NIR method is certainly a valuable tool in CHM manufacturing to impose consistency of products for the advancement of the TCM practice. The procedures developed in this work provide a simple and systematic approach for model development. With more extensive studies, and implementation of user friendly data analysis software, NIR spectroscopy will certainly play a key role in the CHM manufacturing processes and quality control.

References

1. Hua, R.; Sun, S. Q.; Zhou, Q.; Noda, I.; Wang, B. Q., Discrimination of Fritillary according to geographical origin with Fourier transform infrared spectroscopy and two-dimensional correlation IR spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis* **2003**, *33* (2), 199-209.
2. Chan, K.; Lee, H., *The way forward for chinese medicine*. Harwood Academic Publishers, The Taylor and Francis Group 2002.
3. Chan, K., Chinese medicinal materials and their interface with Western medical concepts. *Journal of Ethnopharmacology* **2005**, *96* (1-2), 1-18.
4. Hosbach, I.; Neeb, G.; Hager, S.; Kirchhoff, S.; Kirschbaum, B., In defence of traditional Chinese herbal medicine. *Anaesthesia* **2003**, *58* (3), 282-283.
5. Ko, R. J., A U.S. perspective on the adverse reactions from traditional Chinese medicines. *Journal of the Chinese Medical Association* **2004**, *67* (3), 109-16.
6. Leung, A. Y., Scientific studies and reports in the herbal literature: What are we studying and reporting? *HerbalGram* **2000**, *48*, 62-63.
7. Li, H. B.; Jiang, Y.; Chen, F., Separation methods used for *Scutellaria baicalensis* active components. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2004**, *812* (1-2), 277-290.
8. Tsao, R.; Deng, Z. Y., Separation procedures for naturally occurring antioxidant phytochemicals. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2004**, *812* (1-2), 85-99.
9. Deng, C. H.; Liu, N.; Gao, M. X.; Zhang, X. M., Recent developments in sample preparation techniques for chromatography analysis of traditional Chinese medicines. *Journal of Chromatography A* **2007**, *1153* (1-2), 90-96.
10. Drasar, P.; Moravcova, J., Recent advances in analysis of Chinese medical plants and traditional medicines. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2004**, *812* (1-2), 3-21.
11. Li, P.; Qi, L. W.; Liu, E. H.; Zhou, J. L.; Wen, X. D., Analysis of Chinese herbal medicines with holistic approaches and integrated evaluation models. *Trac-Trends in*

Analytical Chemistry **2008**, 27 (1), 66-77.

12. Li, S. P.; Zhao, J.; Yang, B., Strategies for quality control of Chinese medicines. *Journal of pharmaceutical and biomedical analysis* **2011**, 55 (4), 802-809.

13. Huie, C. W., A review of modern sample-preparation techniques for the extraction and analysis of medicinal plants. *Analytical and Bioanalytical Chemistry* **2002**, 373 (1-2), 23-30.

14. Tang, F.; Zhang, Q. L.; Nie, Z.; Chen, B.; Yao, S. Z., Sample preparation for analyzing traditional Chinese medicines. *Trac-Trends in Analytical Chemistry* **2009**, 28 (11), 1253-1262.

15. Li, H.; Chen, B.; Yao, S. Z., Application of ultrasonic technique for extracting chlorogenic acid from *Eucommia ulmoides* Oliv. (*E. ulmoides*). *Ultrasonics Sonochemistry* **2005**, 12 (4), 295-300.

16. Lee, M. H.; Lin, C. C., Comparison of techniques for extraction of isoflavones from the root of *Radix Puerariae*: Ultrasonic and pressurized solvent extractions. *Food Chemistry* **2007**, 105 (1), 223-228.

17. Ferhat, M. A.; Tigrine-Kordjani, N.; Chemat, S.; Meklati, B. Y.; Chemat, F., Rapid extraction of volatile compounds using a new simultaneous microwave distillation: Solvent extraction device. *Chromatographia* **2007**, 65 (3-4), 217-222.

18. Tigrine-Kordjani, N.; Chemat, F.; Meklati, B. Y.; Tuduri, L.; Giraudel, J. L.; Montury, M., Relative characterization of rosemary samples according to their geographical origins using microwave-accelerated distillation, solid-phase microextraction and Kohonen self-organizing maps. *Analytical and Bioanalytical Chemistry* **2007**, 389 (2), 631-641.

19. Zhang, T. M.; Liang, Y. Z.; Ding, F.; Yang, Q.; Zhu, Y. X., Simultaneous determination of six trace elements in plant medicines by derivative adsorption waves in conjunction with a microwave technique. *Instrumentation Science & Technology* **2007**, 35 (1), 95-111.

20. Chen, X. J.; Guo, B. L.; Li, S. P.; Zhang, Q. W.; Tu, P. F.; Wang, Y. T., Simultaneous determination of 15 flavonoids in *Epimedium* using pressurized liquid

extraction and high-performance liquid chromatography. *Journal of Chromatography A* **2007**, *1163* (1-2), 96-104.

21. Chen, X. J.; Ji, H.; Zhang, Q. W.; Tu, P. F.; Wang, Y. T.; Guo, B. L.; Li, S. P., A rapid method for simultaneous determination of 15 flavonoids in Epimedium using pressurized liquid extraction and ultra-performance liquid chromatography. *Journal of Pharmaceutical and Biomedical Analysis* **2008**, *46* (2), 226-235.

22. Qin, N. Y.; Yang, F. Q.; Wang, Y. T.; Li, S. P., Quantitative determination of eight components in rhizome (Jianghuang) and tuberous root (Yujin) of *Curcuma longa* using pressurized liquid extraction and gas chromatography-mass spectrometry. *Journal of Pharmaceutical and Biomedical Analysis* **2007**, *43* (2), 486-492.

23. Qian, Z. M.; Lu, J.; Gao, Q. P.; Li, S. P., Rapid method for simultaneous determination of flavonoid, saponins and polyacetylenes in *Folium Ginseng* and *Radix Ginseng* by pressurized liquid extraction and high-performance liquid chromatography coupled with diode array detection and mass spectrometry. *Journal of Chromatography A* **2009**, *1216* (18), 3825-3830.

24. Clarkson, C.; Staerk, D.; Hansen, S. H.; Jaroszewski, J. W., Hyphenation of solid-phase extraction with liquid chromatography and nuclear magnetic resonance: Application of HPLC-DAD-SPE-NMR to identification of constituents of *Kanahia laniflora*. *Analytical Chemistry* **2005**, *77* (11), 3547-3553.

25. Kim, M. R.; Abd El-Aty, A. M.; Choi, J. H.; Lee, K. B.; Shim, J. H., Identification of volatile components in *Angelica* species using supercritical-CO₂ fluid extraction and solid phase microextraction coupled to gas chromatography-mass spectrometry. *Biomedical Chromatography* **2006**, *20* (11), 1267-73.

26. Wood, J. A.; Bernards, M. A.; Wan, W. K.; Charpentier, P. A., Extraction of ginsenosides from North American ginseng using modified supercritical carbon dioxide. *Journal of Supercritical Fluids* **2006**, *39* (1), 40-47.

27. Liu, B.; Chang, Y.; Jiang, H.; Shen, B., Extraction of paeonol from Jisheng Shenqi Wan using supercritical fluid extraction. *Biomedical Chromatography* **2007**, *21* (1), 79-83.

28. Yuan, J. B.; Liu, Q.; Wei, G. B.; Tang, F.; Ding, L.; Yao, S. Z., Characterization

and determination of six aristolochic acids and three aristololactams in medicinal plants and their preparations by high-performance liquid chromatography-photodiode array detection/electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **2007**, *21* (14), 2332-2342.

29. Xie, J. J.; Lu, J.; Qian, Z. M.; Yu, Y.; Duan, J. A.; Li, S. P., Optimization and comparison of five methods for extraction of coniferyl ferulate from *Angelica sinensis*. *Molecules* **2009**, *14* (1), 555-565.

30. Nortier, J. L.; Vanherweghem, J. L., Renal interstitial fibrosis and urothelial carcinoma associated with the use of a Chinese herb (*Aristolochia fangchi*). *Toxicology* **2002**, *181*, 577-580.

31. Ernst, E., Toxic heavy metals and undeclared drugs in Asian herbal medicines. *Trends in Pharmacological Sciences* **2002**, *23* (3), 136-139.

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

33. Chen, S. B.; Liu, H. P.; Tian, R. T.; Yang, D. J.; Chen, S. L.; Xu, H. X.; Chan, A. S. C.; Xie, P. S., High-performance thin-layer chromatographic fingerprints of isoflavonoids for distinguishing between *Radix Puerariae lobate* and *Radix Puerariae thomsonii*. *Journal of Chromatography A* **2006**, *1121* (1), 114-119.

34. Di, X.; Chan, K. V. K. C.; Leung, H. W.; 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), 85-95.

35. Cao, J.; Qi, M. L.; Fang, L. H.; Zhou, S.; Fu, R. N.; Zhang, P. P., Solid-phase microextraction-gas chromatographic-mass spectrometric analysis of volatile compounds from *Curcuma wenyujin*. *Journal of Pharmaceutical and Biomedical Analysis* **2006**, *40* (3), 552-558.

36. Deng, C. H.; Xu, X. Q.; Yao, N.; Li, N.; Zhang, X. M., Rapid determination of essential oil compounds in *Artemisia Selengensis Turcz* by gas chromatography-mass

spectrometry with microwave distillation and simultaneous solid-phase microextraction. *Analytica Chimica Acta* **2006**, 556 (2), 289-294.

37. Dalluge, J.; Beens, J.; Brinkman, U. A. T., Comprehensive two-dimensional gas chromatography: a powerful and versatile analytical tool. *Journal of Chromatography A* **2003**, 1000 (1-2), 69-108.

38. Qiu, Y. Q.; Lu, X.; Pang, T.; Zhu, S. K.; Kong, H. W.; Xu, G. W., Study of traditional Chinese medicine volatile oils from different geographical origins by comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry (GC x GC-TOFMS) in combination with multivariate analysis. *Journal of Pharmaceutical and Biomedical Analysis* **2007**, 43 (5), 1721-1727.

39. Adahchour, M.; Beens, J.; Vreuls, R. J. J.; Brinkman, U. A. T., Recent developments in comprehensive two-dimensional gas chromatography (GC X GC) - IV. Further applications, conclusions and perspectives. *Trac-Trends in Analytical Chemistry* **2006**, 25 (8), 821-840.

40. Jiang, Y.; Li, P.; Li, S. P.; Wang, Y. T.; Tu, P., Optimization of pressurized liquid extraction of five major flavanoids from *Lysimachia clethroides*. *Journal of Pharmaceutical and Biomedical Analysis* **2007**, 43 (1), 341-345.

41. Jiang, Y.; Li, S. P.; Chang, H. T.; Wang, Y. T.; Tu, P. F., Pressurized liquid extraction followed by high-performance liquid chromatography for determination of seven active compounds in *Cortex Dictamni*. *Journal of Chromatography A* **2006**, 1108 (2), 268-272.

42. Qi, L. W.; Yu, Q. T.; Li, P.; Li, S. L.; Wang, Y. X.; Sheng, L. H.; Yi, L., Quality evaluation of *Radix Astragali* through a simultaneous determination of six major active isoflavonoids and four main saponins by high-performance liquid chromatography coupled with diode array and evaporative light scattering detectors. *Journal of Chromatography A* **2006**, 1134 (1-2), 162-169.

43. Yan, S. K.; Luo, G. A.; Wang, Y. M.; Cheng, Y. Y., Simultaneous determination of nine components in Qingkailing injection by HPLC/ELSD/DAD and its application to the quality control. *Journal of Pharmaceutical and Biomedical Analysis* **2006**, 40 (4), 889-895.

44. Zhao, J.; Li, S. P.; Yang, F. Q.; Li, P.; Wang, Y. T., Simultaneous determination of saponins and fatty acids in *Ziziphus jujuba* (Suanzaoren) by high performance liquid chromatography-evaporative light scattering detection and pressurized liquid extraction. *Journal of Chromatography A* **2006**, *1108* (2), 188-194.
45. Ma, L. J.; Zhang, X. Z.; Guo, H.; Gan, Y. R., Determination of four water-soluble compounds in *Salvia miltiorrhiza* Bunge by high-performance liquid chromatography with a coulometric electrode array system. *Journal of Chromatography B* **2006**, *833* (2), 260-263.
46. Ren, M. T.; Chen, J.; Song, Y.; Sheng, L. S.; Li, P.; Qi, L. W., Identification and quantification of 32 bioactive compounds in *Lonicera* species by high performance liquid chromatography coupled with time-of-flight mass spectrometry. *Journal of Pharmaceutical and Biomedical Analysis* **2008**, *48* (5), 1351-1360.
47. Liang, X. M.; Jin, Y.; Wang, Y. P.; Jin, G. W.; Fu, Q.; Xiao, Y. S., Qualitative and quantitative analysis in quality control of traditional Chinese medicines. *Journal of Chromatography A* **2009**, *1216* (11), 2033-2044.
48. Li, W. K.; Fitzloff, J. F., A validated method for quantitative determination of saponins in notoginseng (*Panax notoginseng*) using high-performance liquid chromatography with evaporative light-scattering detection. *Journal of Pharmacy and Pharmacology* **2001**, *53* (12), 1637-1643.
49. Chai, X. Y.; Li, S. L.; Li, P., Quality evaluation of *Flos Lonicerae* through a simultaneous determination of seven saponins by HPLC with ELSD. *Journal of Chromatography A* **2005**, *1070* (1-2), 43-48.
50. de Villiers, A.; Gorecki, T.; Lynen, F.; Szucs, R.; Sandra, P., Improving the universal response of evaporative light scattering detection by mobile phase compensation. *Journal of Chromatography A* **2007**, *1161* (1-2), 183-191.
51. Luo, X. J.; Bi, K. S.; Zhou, S. Y.; Wei, Q. H.; Zhang, R. H., Determination of Danshensu, a major active compound of *Radix Salvia miltiorrhiza* in dog plasma by HPLC with fluorescence detection. *Biomedical Chromatography* **2001**, *15* (8), 493-496.
52. Chan, W.; Lee, K. C.; Liu, N.; Cai, Z. W., A sensitivity enhanced

high-performance liquid chromatography fluorescence method for the detection of nephrotoxic and carcinogenic aristolochic acid in herbal medicines. *Journal of Chromatography A* **2007**, *1164* (1-2), 113-119.

53. Huang, L. F.; Liang, Y. Z.; Guo, F. Q.; Zhou, Z. F.; Cheng, B. M., Simultaneous separation and determination of active components in *Cordyceps sinensis* and *Cordyceps militaris* by LC/ESI-MS. *Journal of Pharmaceutical and Biomedical Analysis* **2003**, *33* (5), 1155-1162.

54. Zhang, H. J.; Shen, P.; Cheng, Y. Y., Identification and determination of the major constituents in traditional Chinese medicine Si-Wu-Tang by HPLC coupled with DAD and ESI-MS. *Journal of Pharmaceutical and Biomedical Analysis* **2004**, *34* (3), 705-713.

55. Xiao, H. B.; Krucker, M.; Albert, K.; Liang, X. M., Determination and identification of isoflavonoids in *Radix astragali* by matrix solid-phase dispersion extraction and high-performance liquid chromatography with photodiode array and mass spectrometric detection. *Journal of Chromatography A* **2004**, *1032* (1-2), 117-124.

56. Lai, C. M.; Li, S. P.; Yu, H.; Wan, J. B.; Kan, K. W.; Wang, Y. T., Rapid HPLC-ESI-MS/MS for qualitative and quantitative analysis of saponins in "XUESETONG" injection. *Journal of Pharmaceutical and Biomedical Analysis* **2006**, *40* (3), 669-678.

57. Lu, Y. H.; Zhang, C. W.; Bucheli, P.; Wei, D. Z., Citrus flavonoids in fruit and traditional Chinese medicinal food ingredients in China. *Plant Foods for Human Nutrition* **2006**, *61* (2), 57-65.

58. Li, Y. H.; Jiang, S. Y.; Guan, Y. L.; Liu, X.; Zhou, Y.; Li, L. M.; Huang, S. X.; Sun, H. D.; Peng, S. L.; Zhou, Y., Quantitative determination of the chemical profile of the plant material "Qiang-huo" by LC-ESI-MS-MS. *Chromatographia* **2006**, *64* (7-8), 405-411.

59. Ye, M.; Guo, H.; Guo, H. Z.; Han, J.; Guo, D., Simultaneous determination of cytotoxic bufadienolides in the Chinese medicine ChanSu by high-performance liquid chromatography coupled with photodiode array and mass spectrometry detections. *Journal of Chromatography B* **2006**, *838* (2), 86-95.

60. Yi, T.; Leung, K. S. Y.; Lu, G. H.; Zhang, H.; Chan, K., Identification and determination of the major constituents in traditional Chinese medicinal plant *Polygonum multiflorum* Thunb by HPLC coupled with PAD and ESI/MS. *Phytochemical Analysis* **2007**, *18* (3), 181-187.
61. Shi, Q. R.; Yan, S. K.; Liang, M. J.; Yang, Y.; Wang, Y.; Zhang, W. D., Simultaneous determination of eight components in *Radix Tinosporae* by high-performance liquid chromatography coupled with diode array detector and electrospray tandem mass spectrometry. *Journal of Pharmaceutical and Biomedical Analysis* **2007**, *43* (3), 994-999.
62. Qian, Z. M.; Li, H. J.; Li, P.; Ren, M. T.; Tang, D., Simultaneous qualification and quantification of thirteen bioactive compounds in *Flos Lonicerae* by high-performance liquid chromatography with diode array detector and mass spectrometry. *Chemical & Pharmaceutical Bulletin* **2007**, *55* (7), 1073-1076.
63. Ding, B.; Zhou, T. T.; Fan, G. R.; Hong, Z. Y.; Wu, Y. T., Qualitative and quantitative determination of ten alkaloids in traditional Chinese medicine *Corydalis yanhusuo* WT Wang by LC-MS/MS and LC-DAD. *Journal of Pharmaceutical and Biomedical Analysis* **2007**, *45* (2), 219-226.
64. Yi, T.; Leung, K. S. Y.; Lu, G. H.; Zhang, H., Comparative analysis of *Ligusticum chuanxiong* and related umbelliferous medicinal plants by high performance liquid chromatography-electrospray ionization mass spectrometry. *Planta Medica* **2007**, *73* (4), 392-398.
65. Huang, L. Y.; Chen, T. W.; Ye, Z.; Chen, G. N., Use of liquid chromatography-atmospheric pressure chemical ionization-ion trap mass spectrometry for identification of oleanolic acid and ursolic acid in *Anoectochilus roxburghii* (wall.) Lindl. *Journal of Mass Spectrometry* **2007**, *42* (7), 910-917.
66. Tolonen, A.; Hohtola, A.; Jalonen, J., Comparison of electrospray ionization and atmospheric pressure chemical ionization techniques in the analysis of the main constituents from *Rhodiola rosea* extracts by liquid chromatography/mass spectrometry. *Journal of Mass Spectrometry* **2003**, *38* (8), 845-853.
67. Chen, J.; Song, Y.; Li, P., Capillary high-performance liquid chromatography with mass spectrometry for simultaneous determination of major flavonoids, iridoid

glucosides and saponins in Flos Lonicerae. *Journal of Chromatography A* **2007**, *1157* (1-2), 217-226.

68. Zheng, X. T.; Shi, P. Y.; Cheng, Y. Y.; Qu, H. B., Rapid analysis of a Chinese herbal prescription by liquid chromatography-time-of-flight tandem mass spectrometry. *Journal of Chromatography A* **2008**, *1206* (2), 140-146.

69. Chen, J. H.; Wang, F. M.; Liu, J.; Lee, F. S. C.; Wang, X. R.; Yang, H. H., Analysis of alkaloids in *Coptis chinensis* Franch by accelerated solvent extraction combined with ultra performance liquid chromatographic analysis with photodiode array and tandem mass spectrometry detections. *Analytica Chimica Acta* **2008**, *613* (2), 184-195.

70. Guan, J.; Lai, C. M.; Li, S. P., A rapid method for the simultaneous determination of 11 saponins in *Panax notoginseng* using ultra performance liquid chromatography. *Journal of Pharmaceutical and Biomedical Analysis* **2007**, *44* (4), 996-1000.

71. Liu, Y. M.; Xue, X. Y.; Guo, Z. M.; Xu, Q.; Zhang, F. F.; Liang, X. M., Novel two-dimensional reversed-phase liquid chromatography/hydrophilic interaction chromatography, an excellent orthogonal system for practical analysis. *Journal of Chromatography A* **2008**, *1208* (1-2), 133-140.

72. Cogne, A. L.; Queiroz, E. F.; Wolfender, J. L.; Marston, A.; Mavi, S.; Hostettmann, K., On-line identification of unstable catalpol derivatives from *Jamesbrittenia fodina* by LC-MS and LC-NMR. *Phytochemical Analysis* **2003**, *14* (2), 67-73.

73. de Rijke, E.; de Kanter, F.; Ariese, F.; Brinkman, U. A. T.; Gooijer, C., Liquid chromatography coupled to nuclear magnetic resonance spectroscopy for the identification of isoflavone glucoside malonates in *T-pratense* L. leaves. *Journal of Separation Science* **2004**, *27* (13), 1061-1070.

74. Seger, C.; Godejohann, M.; Tseng, L. H.; Spraul, M.; Girtler, A.; Sturm, S.; Stuppner, H., LC-DAD-MS/SPE-NMR hyphenation. A tool for the analysis of pharmaceutically used plant extracts: Identification of isobaric iridoid glycoside regioisomers from *Harpagophytum procumbens*. *Analytical Chemistry* **2005**, *77* (3), 878-885.

75. Cogne, A. L.; Queiroz, E. F.; Marston, A.; Wolfender, J. L.; Mavi, S.; Hostettmann, K., On-line identification of unstable iridoids from *Jamesbrittenia fodina* by HPLC-MS and HPLC-NMR. *Phytochemical Analysis* **2005**, *16* (6), 429-439.
76. Exarchou, V.; Fiamegos, Y. C.; van Beek, T. A.; Nanos, C.; Vervoort, J., Hyphenated chromatographic techniques for the rapid screening and identification of antioxidants in methanolic extracts of pharmaceutically used plants. *Journal of Chromatography A* **2006**, *1112* (1-2), 293-302.
77. Liu, J. J.; Li, S. P.; 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), 344-349.
78. Ye, N. S.; Zhu, R. H.; Gu, X. X.; Zou, H., Determination of scopolamine, atropine and anisodamine in *Flos daturae* by capillary electrophoresis. *Biomedical Chromatography* **2001**, *15* (8), 509-512.
79. Gao, Y.; Xiang, Q.; Xu, Y. H.; Tian, Y. L.; Wang, E. K., The use of CE-electrochemiluminescence with ionic liquid for the determination of bioactive constituents in Chinese traditional medicine. *Electrophoresis* **2006**, *27* (23), 4842-4848.
80. Yu, L. J.; Xu, Y.; Feng, H. T.; Li, S. F. Y., Separation and determination of toxic pyrrolizidine alkaloids in traditional Chinese herbal medicines by micellar electrokinetic chromatography with organic modifier. *Electrophoresis* **2005**, *26* (17), 3397-3404.
81. Qi, S. D.; Ding, L.; Tian, K.; Chen, X. G.; Hu, Z. D., Novel and simple nonaqueous capillary electrophoresis separation and determination bioactive triterpenes in Chinese herbs. *Journal of Pharmaceutical and Biomedical Analysis* **2006**, *40* (1), 35-41.
82. Xu, C. H.; Sun, S. Q.; Guo, C. Q.; Zhou, Q.; Tao, J. X.; Noda, I., Multi-steps infrared macro-fingerprint analysis for thermal processing of *Fructus viticis*. *Vibrational Spectroscopy* **2006**, *41* (1), 118-125.

83. Liu, H. X.; Sun, S. Q.; Lv, G. H.; Liang, X. Y., Discrimination of extracted lipophilic constituents of Angelica with multi-steps infrared macro-fingerprint method. *Vibrational Spectroscopy* **2006**, *40* (2), 202-208.
84. Wu, Y. W.; Sun, S. Q.; Zhou, Q.; Tao, J. X.; Noda, I., Volatility-dependent 2D IR correlation analysis of traditional Chinese medicine 'Red Flower Oil' preparation from different manufacturers. *Journal of Molecular Structure* **2008**, *882* (1-3), 107-115.
85. Pei, L. K.; Sun, S. Q.; Guo, B. L.; Huang, W. H.; Xiao, P. G., Fast quality control of Herba Epimedii by using Fourier transform infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2008**, *70* (2), 258-264.
86. Liu, H. X.; Sun, S. Q.; Lv, G. H.; Chan, K. K., Study on Angelica and its different extracts by Fourier transform infrared spectroscopy and two-dimensional correlation IR spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2006**, *64* (2), 321-326.
87. Liu, H. X.; Zhou, Q.; Sun, S. Q.; Bao, H. J., Discrimination of different Chrysanthemums with Fourier transform infrared spectroscopy. *Journal of Molecular Structure* **2008**, *883-884*, 38-47.
88. Zhou, Y. Q.; Yu, H.; Zhang, Y. L.; Sun, S. Q.; Chen, S. L.; Zhao, R. H.; Zhou, Q.; Noda, I., Evaluation on intrinsic quality of licorice influenced by environmental factors by using FTIR combined with 2D-IR correlation spectroscopy. *Journal of Molecular Structure* **2010**, *974*, 127-131.
89. Liu, X. H.; Xu, C. H.; Sun, S. Q.; Huang, J.; Zhang, K.; Li, G. Y.; Zhu, Y.; Zhou, Q.; Zhang, Z. C.; Wang, J. H., Discrimination of different genuine Danshen and their extracts by Fourier transform infrared spectroscopy combined with two-dimensional correlation infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2012**, *97*, 290-296.
90. Li, D.; Jin, Z.; Zhou, Q.; Chen, J.; Lei, Y.; Sun, S., Discrimination of five species of Fritillaria and its extracts by FT-IR and 2D-IR. *Journal of Molecular Structure* **974** (1??), 68-72.
91. Xu, B.; Zhang, G.; Sun, S.; Xu, C.; Chen, J.; Tu, Y.; Zhou, Q.; Cui, M.; Wang, J.; Wen, C., Rapid discrimination of three kinds of Radix Puerariae and their extracts by

Fourier transform infrared spectroscopy and two-dimensional correlation infrared spectroscopy. *Journal of Molecular Structure* 1018 (0), 88-95.

92. Liu, Y.; Zhang, G.-J.; Sun, S.-Q.; Noda, I., Study on similar traditional Chinese medicines Cornu Cervi Pantotrichum, Cornu Cervi and Cornu Cervi Degelatinatum by FT-IR and 2D-IR correlation spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis* 52 (4), 631-635.

93. Zhou, Q.; Sun, S. Q.; Zuo, L., Study on traditional Chinese medicine ‘Qing Kai Ling’ injections from different manufactures by 2D IR correlation. *Vibrational Spectroscopy* 2004, 36, 207-212.

94. Chen, X.; Liu, X.; Sheng, D.; Huang, D.; Li, W.; Wang, X., Distinction of broken cellular wall *Ganoderma lucidum* spores and *G. lucidum* spores using FTIR microspectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 97 (0), 667-672.

95. Lu, J.; Xiang, B. R.; Liu, H.; Xiang, S. Y.; Xie, S. F.; Deng, H. S., Application of two-dimensional near-infrared correlation spectroscopy to the discrimination of Chinese herbal medicine of different geographic regions. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* 2008, 69 (2), 580-586.

96. Lau, C. C.; Chan, C. O.; Chau, F. T.; Mok, D. K. W., Rapid analysis of *Radix puerariae* by near-infrared spectroscopy. *Journal of Chromatography A* 2009, 1216 (11), 2130-2135.

97. Li, W.; Xing, L.; Fang, L.; Wang, J.; Qu, H., Application of near infrared spectroscopy for rapid analysis of intermediates of Tanreqing injection. *Journal of Pharmaceutical and Biomedical Analysis* 2010, 53 (3), 350-358.

98. Chan, C. O.; Chu, C. C.; Mok, D. K. W.; Chau, F. T., Analysis of berberine and total alkaloid content in *Cortex Phellodendri* by near infrared spectroscopy (NIRS) compared with high-performance liquid chromatography coupled with ultra-visible spectrometric detection. *Analytica Chimica Acta* 2007, 592 (2), 121-131.

99. Luo, X.; Yu, X.; Wu, X.; Cheng, Y.; Qu, H., Rapid determination of *Paeoniae Radix* using near infrared spectroscopy. *Microchemical Journal* 2008, 90 (1), 8-12.

100. Lin, H.; Zhao, J.; Chen, Q.; Zhou, F.; Sun, L., Discrimination of Radix Pseudostellariae according to geographical origins using NIR spectroscopy and support vector data description. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2011**, *79* (5), 1381-1385.
101. Li, W.; Xing, L.; Cai, Y.; Qu, H., Classification and quantification analysis of Radix scutellariae from different origins with near infrared diffuse reflection spectroscopy. *Vibrational Spectroscopy* **2011**, *55* (1), 58-64.
102. Li, W.; Cheng, Z.; Wang, Y.; Qu, H., Quality control of Lonicerae Japonicae Flos using near infrared spectroscopy and chemometrics. *Journal of Pharmaceutical and Biomedical Analysis* **2013**, *72* (0), 33-39.
103. Lai, Y.; Ni, Y.; Kokot, S., Discrimination of Rhizoma Corydalis from two sources by near-infrared spectroscopy supported by the wavelet transform and least-squares support vector machine methods. *Vibrational Spectroscopy* **2011**, *56* (2), 154-160.
104. Xiong, H.; Gong, X.; Qu, H., Monitoring batch-to-batch reproducibility of liquid-liquid extraction process using in-line near-infrared spectroscopy combined with multivariate analysis. *Journal of Pharmaceutical and Biomedical Analysis* **2012**, *70* (0), 178-187.
105. Wu, Y.; Jin, Y.; Li, Y.; Sun, D.; Liu, X.; Chen, Y., NIR spectroscopy as a process analytical technology (PAT) tool for on-line and real-time monitoring of an extraction process. *Vibrational Spectroscopy* **2012**, *58* (0), 109-118.
106. Wu, Z.; Tao, O.; Dai, X.; Du, M.; Shi, X.; Qiao, Y., Monitoring of a pharmaceutical blending process using near infrared chemical imaging. *Vibrational Spectroscopy* **2012**, *63* (0), 371-379.
107. Krishnan, P.; Kruger, N. J.; Ratcliffe, R. G., Metabolite fingerprinting and profiling in plants using NMR. *Journal of Experimental Botany* **2005**, *56* (410), 255-65.
108. Ward, J. L.; Baker, J. M.; Beale, M. H., Recent applications of NMR spectroscopy in plant metabolomics. *The FEBS Journal* **2007**, *274* (5), 1126-31.
109. Holzgrabe, U.; Deubner, R.; Schollmayer, C.; Waibel, B., Quantitative NMR

spectroscopy--applications in drug analysis. *Journal of Pharmaceutical and Biomedical Analysis* **2005**, *38* (5), 806-12.

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

111. Lee, S. I.; Kwon, H. J.; Lee, Y. M.; Lee, J. H.; Hong, S. P., Simultaneous analysis method for polar and non-polar ginsenosides in red ginseng by reversed-phase HPLC-PAD. *Journal of pharmaceutical and biomedical analysis* **2012**, *60*, 80-85.

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

113. Tistaert, C.; Dejaegher, B.; Vander Heyden, Y., Chromatographic separation techniques and data handling methods for herbal fingerprints: A review. *Analytica Chimica Acta* **2011**, *690* (2), 148-161.

114. Cai, M.; Zhou, Y.; Gesang, S. L.; Bianba, C. R.; Ding, L. S., Chemical fingerprint analysis of rhizomes of *Gymnadenia conopsea* by HPLC-DAD-MSn. *Journal of Chromatography B* **2006**, *844* (2), 301-307.

115. Li, Y.; Hu, Z.; He, L., An approach to develop binary chromatographic fingerprints of the total alkaloids from *Caulophyllum robustum* by high performance liquid chromatography/diode array detector and gas chromatography/mass spectrometry. *Journal of Pharmaceutical and Biomedical Analysis* **2007**, *43* (5), 1667-72.

116. van Nederkassel, A. M.; Vijverman, V.; Massart, D. L.; Heyden, Y. V., Development of a *Ginkgo biloba* fingerprint chromatogram with UV and evaporative light scattering detection and optimization of the evaporative light scattering detector operating conditions. *Journal of Chromatography A* **2005**, *1085* (2), 230-239.

117. Yan, S. K.; Xin, W. F.; Luo, G. A.; Wang, Y. M.; Cheng, Y. Y., An approach to develop two-dimensional fingerprint for the quality control of Qingkailing injection by high-performance liquid chromatography with diode array detection. *Journal of Chromatography A* **2005**, *1090* (1-2), 90-7.

118. Liang, Y. Z.; Kvalheim, O. M.; Keller, H. R.; Massart, D. L.; Kiechle, P.; Erni, F., Heuristic Evolving Latent Projections - Resolving 2-Way Multicomponent Data .2. Detection and Resolution of Minor Constituents. *Analytical Chemistry* **1992**, *64* (8), 946-953.
119. Kvalheim, O. M.; 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), 936-946.
120. Dai, D. Q.; Shih, T. M.; Chau, F. T., Polynomial preserving algorithm for digital image interpolation. *Signal Processing* **1998**, *67* (1), 109-121.
121. Sun, L. X.; Ning, L. L.; Bi, K. S.; Lao, X., Quality evaluation of indigowoad root and leaf by chemical pattern recognition. *Journal of Chinese Medicinal Materials* **2000**, *23*, 609-613.
122. Norris, K. H., History of NIR. *Journal of Near Infrared Spectroscopy* **1996**, *4* (1), 31-37.
123. Roggo, Y.; Duponchel, L.; Huvenne, J. P., Quality evaluation of sugar beet (*Beta vulgaris*) by near-infrared spectroscopy. *Journal of Agricultural and Food Chemistry* **2004**, *52* (5), 1055-61.
124. Williams, P.; Norris, N., *Near-Infrared Technology in the Agricultural and Food Industries*. USA, 1987.
125. Larrechi, M. S.; Callao, M. P., Strategy for introducing NIR spectroscopy and multivariate calibration techniques in industry. *TrAC Trends in Analytical Chemistry* **2003**, *22* (9), 634-640.
126. Blanco, M.; Maspocho, S.; Villarroya, I.; Peralta, X.; González, J. M.; Torres, J., Geographical Origin Classification of Petroleum Crudes from Near-Infrared Spectra of Bitumens. *Applied Spectroscopy* **2001**, *55* (7), 834-839.
127. Siesler, H. W.; Ozaki, Y.; Kawata, S.; Heise, H. M., *Near-Infrared Spectroscopy, Principles Instruments Applications*. Wiley-VCH: Weinheim, 2002.
128. Burns, D. A.; Ciurczak, E. W., *Handbook of Near-Infrared Analysis, Revised and*

Expanded. 2nd ed ed.; Marcel Dekker, Inc: New York, Basel, Hong Kong, 2001.

129. Bokobza, L., Near infrared spectroscopy. *Journal of Near Infrared Spectroscopy* **1998**, *6* (1), 3-17.

130. Breiman, L.; Friedman, R.; Olsen, R.; Stone, C., *Classification and Regression Trees*. Wadsworth: Pacific Grove, CA, 1984.

131. Massart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; Jong, S. D.; Lewi, P. J.; Smeyers-Verbeke, J., *Handbook of Chemometrics and Qualimetrics: Part A*. Elsevier: Amsterdam, 1997.

132. Vandeginste, B. G. M.; Massart, D. L.; Buydens, L. M. C.; Jong, S. D.; Lewi, P. J.; Smeyers-Verbeke, J., *Handbook of Chemometrics and Qualimetrics: Part B*. Elsevier: Amsterdam, 1998.

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

134. Kortüm, G., *Reflectance spectroscopy : Principles, Methods, Applications*. Springer: New York 1969.

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

136. Wu, W.; Walczak, B.; Massart, D. L.; Prebble, K. A.; Last, I. R., Spectral Transformation and Wavelength Selection in near-Infrared Spectra Classification. *Anal Chim Acta* **1995**, *315* (3), 243-255.

137. Salamin, P. A.; Cornelis, Y.; Bartels, H., Identification of Chemical-Substances by Their near-Infrared Spectra. *Chemometr Intell Lab* **1988**, *3* (4), 329-333.

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

139. Barnes, R. J.; Dhanoa, M. S.; Lister, S. J., Letter: Correction to the description of

Standard Normal Variate (SNV) and De-Trend (DT) transformations in Practical Spectroscopy with Applications in Food and Beverage Analysis—2nd Edition. *Journal of Near Infrared Spectroscopy* **1993**, *1* (3), 185-186.

140. Dhanoa, M. S.; Lister, S. J.; Sanderson, R.; 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* (1), 43-47.

141. Candolfi, A.; De Maesschalck, R.; Jouan-Rimbaud, D.; Hailey, P. A.; 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), 115-132.

142. Barnes, R. J.; Dhanoa, M. S.; Lister, S. J., Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy* **1989**, *43* (5), 772-777.

143. Dhanoa, M. S.; Lister, S. J.; Barnes, R. J., On the scales associated with near-infrared reflectance difference spectra. *Applied Spectroscopy* **1995**, *49* (6), 765-772.

144. Savitzky, A.; Golay, M. J. E., Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* **1964**, *36* (8), 1627-1639.

145. González-Arjona, D.; González-Gallero, V.; Pablos, F.; Gustavo González, A., Authentication and differentiation of Irish whiskeys by higher-alcohol congener analysis. *Analytica Chimica Acta* **1999**, *381* (2-3), 257-264.

146. Simon, L.; Nazmul Karim, M., Probabilistic neural networks using Bayesian decision strategies and a modified Gompertz model for growth phase classification in the batch culture of *Bacillus subtilis*. *Biochemical Engineering Journal* **2001**, *7* (1), 41-48.

147. Marini, F.; Balestrieri, F.; Bucci, R.; Magrì, A. D.; Magrì, A. L.; Marini, D., Supervised pattern recognition to authenticate Italian extra virgin olive oil varieties. *Chemometrics and Intelligent Laboratory Systems* **2004**, *73* (1), 85-93.

148. Massart, D. L.; Vandeginste, B. G. M.; Deming, S. M.; Michotte, Y.; Kaufmann, L., *Chemometrics: A textbook*. Elsevier.: Amsterdam., 2003.
149. Martens, H.; Martens, M., *Multivariate Analysis of Quality An Introduction*. John Wiley & Sons: Chichester, 2001.
150. Roggo, Y.; Roeseler, C.; Ulmschneider, M., Near infrared spectroscopy for qualitative comparison of pharmaceutical batches. *Journal of Pharmaceutical and Biomedical Analysis* **2004**, *36* (4), 777-786.
151. EMEA/CVMP/961/01,2003. *The European Agency for the Evaluation of Medicinal Products* **2003**.
152. Daszykowski, M.; Walczak, B.; Massart, D. L., Looking for natural patterns in data: Part 1. Density-based approach. *Chemometrics and Intelligent Laboratory Systems* **2001**, *56* (2), 83-92.
153. Daszykowski, M.; Walczak, B.; Massart, D. L., Representative subset selection. *Analytica Chimica Acta* **2002**, *468* (1), 91-103.
154. Lopes, J. A.; Costa, P. F.; Alves, T. P.; Menezes, J. C., Chemometrics in bioprocess engineering: process analytical technology (PAT) applications. *Chemometrics and Intelligent Laboratory Systems* **2004**, *74* (2), 269-275.
155. Danzer, K.; Hobert, H.; Fischbacher, C.; Jagemann, K. U., *Chemometrik, Grundlagen und Anwendungen*. Springer: Berlin, 2011.
156. Mark, H. L.; Tunnell, D., Qualitative near-infrared reflectance analysis using Mahalanobis distances. *Analytical Chemistry* **1985**, *57* (7), 1449-1456.
157. Coomans, D.; Jonckheer, M.; Massart, D. L.; Broeckaert, I.; Blockx, P., The application of linear discriminant analysis in the diagnosis of thyroid diseases. *Analytica Chimica Acta* **1978**, *103* (4), 409-415.
158. Sharaf, M.; Illman, D.; Kowalski, B., *Chemometrics*. Wiley/Interscience: New York, 1986.
159. Cover, T. M.; Hart, P. E., Nearest Neighbor Pattern Classification. *IEEE*

Transactions on Information Theory **1967**, 13 (1), 21-27.

160. Derde, M. P.; Buydens, L.; Guns, C.; Massart, D. L.; Hopke, P. K., Comparison of rule-building expert systems with pattern recognition for the classification of analytical data. *Analytical chemistry* **1987**, 59 (14), 1868–1871.

161. Coomans, D.; Massart, D. L., Alternative k-nearest neighbour rules in supervised pattern recognition : Part 2. Probabilistic classification on the basis of the kNN method modified for direct density estimation. *Analytica Chimica Acta* **1982**, 138, 153-165.

162. Wold, S., Pattern-recognition by means of disjoint principal components models. *Pattern Recognition* **1976**, 8 (3), 127-139.

163. Shaffer, R. E.; Rose-Pehrsson, S. L.; McGill, R. A., A comparison study of chemical sensor array pattern recognition algorithms. *Analytica Chimica Acta* **1999**, 384 (3), 305-317.

164. Blanco, M.; Romero, M. A.; Alcala, M., Strategies for constructing the calibration set for a near infrared spectroscopic quantitation method. *Talanta* **2004**, 64 (3), 597-602.

165. Blanco, M.; Coello, J.; Iturriaga, H.; MasPOCH, S.; delaPezuela, C., Strategies for constructing the calibration set in the determination of active principles in pharmaceuticals by near infrared diffuse reflectance spectrometry. *Analyst* **1997**, 122 (8), 761-765.

166. Kalivas, J. H., Basis sets for multivariate regression. *Analytica Chimica Acta* **2001**, 428 (1), 31-40.

167. Galvao, R. K. H.; Araujo, M. C. U.; Jose, G. E.; Pontes, M. J. C.; Silva, E. C.; Saldanha, T. C. B., A method for calibration and validation subset partitioning. *Talanta* **2005**, 67 (4), 736-740.

168. Filho, H. A. D.; Galvão, R. K. H.; Araújo, M. C. U.; da Silva, E. C.; Saldanha, T. C. B.; José, G. E.; Pasquini, C.; Milton Raimundo Jr, I.; Rohwedder, J. J. R., A strategy for selecting calibration samples for multivariate modelling. *Chemometrics and Intelligent Laboratory Systems* **2004**, 72 (1), 83-91.

169. Wu, W.; Walczak, B.; Massart, D. L.; Heuerding, S.; Erni, F.; Last, I. R.; Prebble, K. A., Artificial neural networks in classification of NIR spectral data: Design of the training set. *Chemometrics and Intelligent Laboratory Systems* **1996**, *33* (1), 35-46.
170. Martens, H.; Naes, T., *Multivariate Calibration*. JohnWiley&Sons: Chichester, 1996.
171. Naes, T.; Isaksson, T.; Fearn, T.; Davies, T., *Multivariate Calibration and Classification*. NIR Publications: Chichester, UK, 2002.
172. Adams, M. J., *Chemometrics in Analytical Spectroscopy*. Royal Society of Chemistry: Cambridge, UK, 1995.
173. Dou, Y.; Sun, Y.; Ren, Y. Q.; Ju, P.; Ren, Y. L., Simultaneous non-destructive determination of two components of combined paracetamol and amantadine hydrochloride in tablets and powder by NIR spectroscopy and artificial neural networks. *Journal of Pharmaceutical and Biomedical Analysis* **2005**, *37* (3), 543-549.
174. Chen, Y. X.; Thosar, S. S.; Forbess, R. A.; Kemper, M. S.; Rubinovitz, R. L.; Shukla, A. J., Prediction of drug content and hardness of intact tablets using artificial neural network and near-infrared spectroscopy. *Drug Development and Industrial Pharmacy* **2001**, *27* (7), 623-631.
175. Findlay, W. P.; Peck, G. R.; Morris, K. R., Determination of fluidized bed granulation end point using near-infrared spectroscopy and phenomenological analysis. *Journal of Pharmaceutical Sciences* **2005**, *94* (3), 604-612.
176. Zhou, X. J.; Hines, P.; Borer, M. W., Moisture determination in hygroscopic drug substances by near infrared spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis* **1998**, *17* (2), 219-225.
177. Rantanen, J.; Antikainen, O.; Mannermaa, J. P.; Yliruusi, J., Use of the near-infrared reflectance method for measurement of moisture content during granulation. *Pharmaceutical Development and Technology* **2000**, *5* (2), 209-217.
178. Buice, R. G.; Gold, T. B.; Lodder, R. A.; Digenis, G. A., Determination of moisture in intact gelatin capsules by near-infrared spectrophotometry. *Pharmaceutical Research* **1995**, *12* (1), 161-163.

179. Berntsson, O.; Zackrisson, G.; Ostling, G., Determination of moisture in hard gelatin capsules using near-infrared spectroscopy: Applications to at-line process control of pharmaceuticals. *Journal of Pharmaceutical and Biomedical Analysis* **1997**, *15* (7), 895-900.
180. Broad, N. W.; Jee, R. D.; Moffat, A. C.; Eaves, M. J.; Mann, W. C.; Dziki, W., Non-invasive determination of ethanol, propylene glycol and water in a multi-component pharmaceutical oral liquid by direct measurement through amber plastic bottles using Fourier transform near-infrared spectroscopy. *Analyst* **2000**, *125* (11), 2054-2058.
181. Boyer, C.; Gaudin, K.; Kauss, T.; Gaubert, A.; Boudis, A.; Verschelden, J.; Franc, M.; Roussille, J.; Boucher, J.; Olliaro, P.; White, N. J.; Millet, P.; Dubost, J. P., Development of NIRS method for quality control of drug combination artesunate-azithromycin for the treatment of severe malaria. *Journal of Pharmaceutical and Biomedical Analysis* **2012**, *67-68*, 10-15.
182. Dou, Y.; Qu, N.; Wang, B.; Chi, Y. Z.; Ren, Y. L., Simultaneous determination of two active components in compound aspirin tablets using principal component artificial neural networks (PC-ANNs) on NIR spectroscopy. *European Federation for Pharmaceutical Sciences* **2007**, *32* (3), 193-199.
183. Blanco, M.; Alcalá, M.; Bautista, M., Pharmaceutical gel analysis by NIR spectroscopy. Determination of the active principle and low concentration of preservatives. *European Federation for Pharmaceutical Sciences* **2008**, *33* (4-5), 409-414.
184. Roggo, Y.; Jent, N.; Edmond, A.; Chalus, P.; Ulmschneider, M., Characterizing process effects on pharmaceutical solid forms using near-infrared spectroscopy and infrared imaging. *European Journal of Pharmaceutics and Biopharmaceutics* **2005**, *61* (1-2), 100-110.
185. Blanco, M.; Coello, J.; Iturriaga, H.; Maspocho, S.; Serrano, D., Near-infrared analytical control of pharmaceuticals. A single calibration model from mixed phase to coated tablets. *Analyst* **1998**, *123* (11), 2307-2312.
186. Han, S. M.; Faulkner, P. G., Determination of SE 216469-S during tablet production using near-infrared reflectance spectroscopy. *Journal of Pharmaceutical*

and Biomedical Analysis **1996**, *14* (12), 1681-1689.

187. Chen, Y. X.; 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), 491-496.

188. Ren, G. X.; 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), 2771-2775.

189. Rager, I.; Roos, G.; Schmidt, P. C.; 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), 439-446.

190. Li, W.; Cheng, Z.; Wang, Y.; Qu, H., A study on the use of near-infrared spectroscopy for the rapid quantification of major compounds in Tanreqing injection. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2013**, *101* (0), 1-7.

191. Jiang, Y.; David, B.; Tu, P. F.; Barbin, Y., Recent analytical approaches in quality control of traditional Chinese medicines-A review. *Analytica Chimica Acta* **2010**, *657* (1), 9-18.

192. Luypaert, J.; Zhang, M. H.; Massart, D. L., Feasibility study for the use of near infrared spectroscopy in the qualitative and quantitative analysis of green tea, *Camellia sinensis* (L.). *Analytica Chimica Acta* **2003**, *478* (2), 303-312.

193. Woo, Y. A.; Kim, H. J.; Ze, K. R.; 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), 955-9.

194. Laasonen, M.; Harmia-Pulkkinen, T.; Simard, C. L.; Michiels, E.; Rasanen, M.; Vuorela, H., Fast identification of *Echinacea purpurea* dried roots using near-infrared spectroscopy. *Analytical chemistry* **2002**, *74* (11), 2493-9.

195. Chen, Q. S.; Zhao, J. W.; Zhang, H. D.; Wang, X. Y., Feasibility study on qualitative and quantitative analysis in tea by near infrared spectroscopy with

multivariate calibration. *Analytica Chimica Acta* **2006**, 572 (1), 77-84.

196. *The State Pharmacopoeia Commission of People's Republic of China, Pharmacopoeia of the People's Republic of China*. China Medical Science and Technology Press: Beijing, 2010; Vol. 1, p 285.

197. Wangner, H.; Bauer, R.; Xiao, P. G.; Chen, J. M.; Michler, H.; Bacher, S., *Chinese Drug Monographs and Analysis*. Verlag fur GanzheitlicheMedizin, 2003; Vol. 4.

198. Cervellati, R.; Renzulli, C.; Guerra, M. C.; Speroni, E., Evaluation of antioxidant activity of some natural polyphenolic compounds using the Briggs-Rauscher reaction method. *Journal of agricultural and food chemistry* **2002**, 50 (26), 7504-9.

199. Ding, L.; Yang, Y.; Han, S. Puerarin injection for perfusion for treating heart disease. 1998.

200. Lukas, S. E.; Penetar, D.; Berko, J.; Vicens, L.; Palmer, C.; Mallya, G.; Macklin, E. A.; Lee, D. Y. W., An extract of the Chinese herbal root Kudzu reduces alcohol drinking by heavy drinkers in a naturalistic setting. *Alcoholism, Clinical and Experimental Research* **2005**, 29 (5), 756-762.

201. Keung, W. M., Anti-dipsotropic isoflavones: The potential therapeutic agents for alcohol dependence. *Medicinal Research Reviews* **2003**, 23 (6), 669-696.

202. Jiang, R. W.; Lau, K. M.; Lam, H. M.; Yam, W. S.; Leung, L. K.; Choi, K. L.; Waye, M. M. Y.; Mak, T. C. W.; Woo, K. S.; Fung, K. P., A comparative study on aqueous root extracts of *Pueraria thomsonii* and *Pueraria lobata* by antioxidant assay and HPLC fingerprint analysis. *Journal of Ethnopharmacology* **2005**, 96 (1-2), 133-138.

203. *The State Pharmacopoeia Commission of People's Republic of China, Pharmacopoeia of the 313 People's Republic of China*. Chemical Industry Press: Beijing, 2005.

204. Cherdshewasart, W.; Subtang, S.; Dahlan, W., Major isoflavonoid contents of the phytoestrogen rich-herb *Pueraria mirifica* in comparison with *Pueraria lobata*. *Journal of Pharmaceutical and Biomedical Analysis* **2007**, 43 (2), 428-434.

205. Chen, S. L.; Chen, S. B.; Yang, D. J.; Chan, A. S. C.; Xu, H. X., HPLC fingerprint analyses of Radix Puerariae. *Chinese Traditional and Herbal Drugs* **2003**, *23*, 661-663.
206. Fang, C. B.; Wan, X. C.; Jiang, C. J.; Tan, H. R.; Hu, Y. H.; Cao, H. Q., Comparison of HPTLC, HPLC, and HPCE for fingerprinting of Pueraria radix. *Journal of Planar Chromatography - Modern Tlc* **2006**, *19* (111), 348-354.
207. Keung, W. M.; Vallee, B. L., *Phytochemistry* **1997**, *47* (4), 499-506.
208. Lin, R. C.; Guthrie, S.; Xie, C. Y.; Mai, K.; Lee, D. Y.; Lumeng, L.; Li, T. K., Isoflavonoid compounds extracted from Pueraria lobata suppress alcohol preference in a pharmacogenetic rat model of alcoholism. *Alcoholism, clinical and experimental research* **1996**, *20* (4), 659-63.
209. Kang, K. A.; Chae, S.; Koh, Y. S.; Kim, J. S.; Lee, J. H.; You, H. J.; Hyun, J. W., Protective effect of Puerariae radix on oxidative stress induced by hydrogen peroxide and streptozotocin. *Biological & Pharmaceutical Bulletin* **2005**, *28* (7), 1154-1160.
210. Xiong, F. L.; Sun, X. H.; Lu, G.; Yang, X. L.; Xu, H. B., Puerarin protects rat pancreatic islets from damage by hydrogen peroxide. *European Journal of Pharmacology* **2006**, *529* (1-3), 1-7.
211. Guerra, M. C.; Speroni, E.; Broccoli, M.; Cangini, M.; Pasini, P.; Minghetti, A.; Crespi-Perellino, N.; Mirasoli, M.; Cantelli-Forti, G.; Paolini, M., Comparison between Chinese medical herb Pueraria lobata crude extract and its main isoflavone puerarin - Antioxidant properties and effects on rat liver CYP-catalysed drug metabolism. *Life Sciences* **2000**, *67* (24), 2997-3006.
212. Meng, F.; Liu, R.; Bai, H.; Liu, B. W.; Liu, Y., Inhibitory effect of quercetin, rutin and puerarin on HDL oxidation induced by Cu²⁺. *Sichuan Da Xue Xue Bao Yi Xue Ban* **2004**, *35* (6), 836-838.
213. Wang, J. F.; Guo, Y. X.; Niu, J. Z.; Liu, J.; Wang, L. Q.; Li, P. H., Effects of Radix Puerariae flavones on liver lipid metabolism in ovariectomized rats. *World Journal of Gastroenterology* **2004**, *10* (13), 1967-1970.
214. Zhang, Y.; Chen, J.; Zhang, C.; Wu, W.; Liang, X., Analysis of the estrogenic

components in kudzu root by bioassay and high performance liquid chromatography. *The Journal of steroid biochemistry and molecular biology* **2005**, 94 (4), 375-81.

215. Jeon, G. C.; Park, M. S.; Yoon, D. Y.; Shin, C. H.; Sin, H. S.; Um, S. J., Antitumor activity of spinasterol isolated from Pueraria roots. *Experimental and molecular medicine* **2005**, 37 (2), 111-20.

216. Yu, Z. L.; Li, W. J., Induction of apoptosis by puerarin in colon cancer HT-29 cells. *Cancer Letters* **2006**, 238 (1), 53-60.

217. Yan, B.; Wang, D. Y.; Xing, D. M.; Ding, Y.; Wang, R. F.; Lei, F.; Du, L. J., The antidepressant effect of ethanol extract of radix puerariae in mice exposed to cerebral ischemia reperfusion. *Pharmacology, Biochemistry, and Behavior* **2004**, 78 (2), 319-25.

218. Zhao, Y.; Du, G. Y.; Cui, H. F.; Cao, C. Y.; Wang, X. R.; Zhang, C. Y., Experimental study of protective effect of pueraria compound on the cerebral ischemic injury. *Zhongguo Zhong Yao Za Zhi* **2005**, 30 (7), 548-51.

219. Ren, X. Q.; Li, J. S.; Feng, Y. M.; Lu, Y. Q., Neuro-protective effect of naomaitong to brain damage after focal cerebral ischemia reperfusion (I/R) in the aged rats. *Zhongguo Zhong Yao Za Zhi* **2004**, 29 (1), 66-70.

220. Yan, B.; Xing, D.; Ding, Y.; Tao, J.; Du, L. J., HPLC method for the determination and pharmacokinetic studies on puerarin in cerebral ischemia reperfusion rat plasma after intravenous administration of puerariae radix isoflavone. *Journal of Pharmaceutical and Biomedical Analysis* **2005**, 37 (2), 297-301.

221. Yu, B. S.; Yan, X. P.; Zhen, G. B.; Rao, Y. P., RP-HPLC determination of puerarin in Chinese traditional medicinal preparations containing pueraria. *Journal of Pharmaceutical and Biomedical Analysis* **2002**, 30 (3), 843-849.

222. Wang, C. Y.; Huang, H. Y.; Kuo, K. L.; Hsieh, Y. Z., Analysis of Puerariae radix and its medicinal preparations by capillary electrophoresis. *Journal of Chromatography A* **1998**, 802 (1), 225-231.

223. Chen, G.; Zhang, J. X.; Ye, J. N., Determination of puerarin, daidzein and rutin in Pueraria lobata (Wild.) Ohwi by capillary electrophoresis with electrochemical

detection. *Journal of Chromatography A* **2001**, 923 (1-2), 255-262.

224. de Groot, P. J.; Postma, G. J.; Melssen, W. J.; Buydens, L. M. C., Selecting a representative training set for the classification of demolition waste using remote NIR sensing. *Analytica Chimica Acta* **1999**, 392 (1), 67-75.

225. Haaland, D. M.; Thomas, E. V., Partial Least-Squares Methods for Spectral Analyses .1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information. *Analytical Chemistry* **1988**, 60 (11), 1193-1202.

226. Bangalore, A. S.; Shaffer, R. E.; Small, G. W.; Arnold, M. A., Genetic algorithm-based method for selecting wavelengths and model size for use with partial least-squares regression: application to near-infrared spectroscopy. *Analytical chemistry* **1996**, 68 (23), 4200-12.

227. Brown, P. J., Wavelength Selection in Multicomponent near-Infrared Calibration. *Journal of Chemometrics* **1992**, 6 (3), 151-161.

228. Brown, P. J., Graphics for Linearity and Selectivity and Prediction Diagnostics for Multicomponent Spectra. *Journal of Chemometrics* **1993**, 7 (4), 255-265.

229. Horchner, U.; Kalivas, J. H., Further Investigation on a Comparative-Study of Simulated Annealing and Genetic Algorithm for Wavelength Selection. *Analytica Chimica Acta* **1995**, 311 (1), 1-13.

230. Jouanrimbaud, D.; Massart, D. L.; Leardi, R.; Denoord, O. E., Genetic Algorithms as a Tool for Wavelength Selection in Multivariate Calibration. *Analytical Chemistry* **1995**, 67 (23), 4295-4301.

231. Jouanrimbaud, D.; Walczak, B.; Massart, D. L.; Last, I. R.; Prebble, K. A., Comparison of Multivariate Methods Based on Latent Vectors and Methods Based on Wavelength Selection for the Analysis of near-Infrared Spectroscopic Data. *Analytica Chimica Acta* **1995**, 304 (3), 285-295.

232. Kalivas, J. H.; Roberts, N.; Sutter, J. M., GLOBAL OPTIMIZATION BY SIMULATED ANNEALING WITH WAVELENGTH SELECTION FOR ULTRAVIOLET VISIBLE SPECTROPHOTOMETRY. *Analytical Chemistry* **1989**, 61 (18), 2024-2030.

233. Liang, Y. Z.; Xie, Y. L.; Yu, R. Q., Accuracy criteria and optimal wavelength selection for multicomponent spectrophotometric determinations. *Analytica Chimica Acta* **1989**, 222 (1), 347-357.
234. Lucasius, C. B.; Beckers, M. L. M.; Kateman, G., Genetic Algorithms in Wavelength Selection - a Comparative-Study. *Analytica Chimica Acta* **1994**, 286 (2), 135-153.
235. Lucasius, C. B.; Kateman, G., Genetic Algorithms for Large-Scale Optimization in Chemometrics - an Application. *Trac-Trends in Analytical Chemistry* **1991**, 10 (8), 254-261.
236. McShane, M. J.; Cote, G. L.; Spiegelman, C., Variable selection in multivariate calibration of a spectroscopic glucose sensor. *Applied Spectroscopy* **1997**, 51 (10), 1559-1564.
237. Navarro-Villoslada, F.; Pérez-Arribas, L. V.; León-González, M. E.; Polo-Díez, L. M., Selection of calibration mixtures and wavelengths for different multivariate calibration methods. *Analytica Chimica Acta* **1995**, 313 (1-2), 93-101.
238. Salamin, P. A.; Bartels, H.; Forster, P., A Wavelength and Optical-Path Length Selection Procedure for Spectroscopic Multicomponent Analysis. *Chemometrics and Intelligent Laboratory Systems* **1991**, 11 (1), 57-62.
239. Sasaki, K.; Kawata, S.; Minami, S., Optimal Wavelength Selection for Quantitative-Analysis. *Applied Spectroscopy* **1986**, 40 (2), 185-190.
240. Spiegelman, C. H.; McShane, M. J.; Goetz, M. J.; Motamedi, M.; Yue, Q. L.; Cote, G. L., Theoretical justification of wavelength selection in PLS calibration development of a new algorithm. *Analytical Chemistry* **1998**, 70 (1), 35-44.
241. Centner, V.; Massart, D. L.; deNoord, O. E.; deJong, S.; Vandeginste, B. M.; Sterna, C., Elimination of uninformative variables for multivariate calibration. *Analytical Chemistry* **1996**, 68 (21), 3851-3858.
242. Cortes, C.; Vapnik, V., Support-Vector Networks. *Machine Learning* **1995**, 20 (3), 273-297.

243. Burges, C. J. C., A tutorial on Support Vector Machines for pattern recognition. *Data Mining and Knowledge Discovery* **1998**, 2 (2), 121-167.
244. Boger, Z., Selection of quasi-optimal inputs in chemometrics modeling by artificial neural network analysis. *Analytica Chimica Acta* **2003**, 490 (1-2), 31-40.
245. Swierenga, H.; de Groot, P. J.; de Weijer, A. P.; Derksen, M. W. J.; Buydens, L. M. C., Improvement of PLS model transferability by robust wavelength selection. *Chemometrics and Intelligent Laboratory Systems* **1998**, 41 (2), 237-248.
246. Polgar, O.; Fried, M.; Lohner, T.; Barsony, I., Comparison of algorithms used for evaluation of ellipsometric measurements - Random search, genetic algorithms, simulated annealing and hill climbing graph-searches. *Surface Science* **2000**, 457 (1-2), 157-177.
247. Leardi, R., Genetic algorithms in feature selection. In *Genetic Algorithms in Molecular Modeling*, Devillers, J., Ed. Academic Press: 1996; p 67.
248. Lucasius, C. B.; Kateman, G., Understanding and Using Genetic Algorithms .1. Concepts, Properties and Context. *Chemometrics and Intelligent Laboratory Systems* **1993**, 19 (1), 1-33.
249. Hibbert, D. B., Genetic Algorithms in Chemistry. *Chemometrics and Intelligent Laboratory Systems* **1993**, 19 (3), 277-293.
250. Leardi, R.; Boggia, R.; Terrile, M., Genetic Algorithms as a Strategy for Feature-Selection. *Journal of Chemometrics* **1992**, 6 (5), 267-281.
251. Leardi, R., Application of a Genetic Algorithm to Feature-Selection under Full Validation Conditions and to Outlier Detection. *Journal of Chemometrics* **1994**, 8 (1), 65-79.
252. Leardi, R.; Seasholtz, M. B.; Pell, R. J., Variable selection for multivariate calibration using a genetic algorithm: prediction of additive concentrations in polymer films from Fourier transform-infrared spectral data. *Analytica Chimica Acta* **2002**, 461 (2), 189-200.
253. Leardi, R., Application of genetic algorithm-PLS for feature selection in spectral

data sets. *Journal of Chemometrics* **2000**, *14* (5-6), 643-655.

254. Ghasemi, J.; Niazi, A.; Leardi, R., Genetic-algorithm-based wavelength selection in multicomponent spectrophotometric determination by PLS: application on copper and zinc mixture. *Talanta* **2003**, *59* (2), 311-317.

255. Bova, S.; Padrini, R.; Goldman, W. F.; Berman, D. M.; Cargnelli, G., On the Mechanism of Vasodilating Action of Berberine - Possible Role of Inositol Lipid Signaling System. *Journal of Pharmacology and Experimental Therapeutics* **1992**, *261* (1), 318-323.

256. Chiou, W. F.; Yen, M. H.; Chen, C. F., Mechanism of Vasodilatory Effect of Berberine in Rat Mesenteric-Artery. *European Journal of Pharmacology* **1991**, *204* (1), 35-40.

257. Yokozawa, T.; Satoh, A.; Cho, E. J.; Kashiwada, Y.; Ikeshiro, Y., Protective role of Coptidis Rhizoma alkaloids against peroxy-nitrite-induced damage to renal tubular epithelial cells. *Journal of Pharmacy and Pharmacology* **2005**, *57* (3), 367-374.

258. Ro, J. S.; Lee, S. S.; Lee, K. S.; Lee, M. K., Inhibition of type A monoamine oxidase by coptisine in mouse brain. *Life Sciences* **2001**, *70* (6), 639-645.

259. Janbaz, K. H.; Gilani, A. H., Studies on preventive and curative effects of berberine on chemical-induced hepatotoxicity in rodents. *Fitoterapia* **2000**, *71* (1), 25-33.

260. Yokozawa, T.; Ishida, A.; Kashiwada, Y.; Cho, E. J.; Kim, H. Y.; Ikeshiro, Y., Coptidis Rhizoma: protective effects against peroxy-nitrite-induced oxidative damage and elucidation of its active components. *Journal of Pharmacy and Pharmacology* **2004**, *56* (4), 547-556.

261. *China Pharmacopoeia Committie, Chinese Pharmacopoeia book 1*. Chinese Industry Press: Beijing, 2005.

262. 周榮漢, *中藥資源學*. 中國醫藥科技出版社: 北京, 1993.

263. 潘綱, *易混淆的中藥鑒別*. 江蘇科學技術出版社: 南京 1987.

264. Gao, W. H.; Lin, S. Y.; Jia, L.; Gu, X. K.; Chen, X. G.; 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), 92-97.
265. Ge, Z. C.; Zhou, J. M., Determination of berberine, palmatine and jatrorrhizine in coptis and its preparations by micellar thin layer, chromatographic scanning method. *Chinese Journal of Analytical Chemistry* **2004**, *32* (1), 99-101.
266. Sun, S. W.; Tseng, H. M., Sensitivity improvement on detection of Coptidis alkaloids by sweeping in capillary electrophoresis. *Journal of Pharmaceutical and Biomedical Analysis* **2005**, *37* (1), 39-45.
267. Chen, Y. R.; Wen, K. C.; 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), 273-280.
268. Chang, L. C.; Sun, S. W., Micellar electrokinetic chromatography for separation of a mixture of coptis alkaloids, scute flavonoids, and rhubarb anthraquinones and bianthrone. *Journal of Pharmaceutical and Biomedical Analysis* **2006**, *40* (1), 62-67.
269. Kong, W. J.; Zhao, Y. L.; Xiao, X. H.; Jin, C.; Li, Z. L., Quantitative and chemical fingerprint analysis for quality control of Rhizoma Coptidis chinensis based on UPLC-PAD combined with chemometrics methods. *Phytomedicine* **2009**, *16* (10), 950-959.
270. Ren, L. L.; Xue, X. Y.; Zhang, F. F.; Xu, Q.; Liang, X. M., High performance liquid chromatography-mass spectrometry analysis of protoberberine alkaloids in medicine herbs. *Journal of Separation Science* **2007**, *30* (6), 833-842.
271. Chu, C. Y.; Sheu, S. J., Liquid chromatographic separation of the alkaloids in coptis-evodia herb couple. *Journal of Chromatography A* **1996**, *756* (1-2), 137-144.
272. Luo, X. B.; Chen, B.; Yao, S. Z., Simultaneous analysis of protoberberine, indolequinoline and quinolone alkaloids in coptis-evodia herb couple and the Chinese herbal preparations by high-performance liquid chromatography-electrospray mass spectrometry. *Talanta* **2005**, *66* (1), 103-110.

273. Wang, D. W.; Liu, Z. Q.; Guo, M. Q.; Liu, S. Y., Structural elucidation and identification of alkaloids in Rhizoma Coptidis by electrospray ionization tandem mass spectrometry. *Journal of Mass Spectrometry* **2004**, *39* (11), 1356-1365.
274. Li, C. Y.; Tsai, S. I.; Damu, A. G.; Wu, T. S., A rapid and simple determination of protoberberine alkaloids in Rhizoma Coptidis by ¹H NMR and its application for quality control of commercial prescriptions. *Journal of Pharmaceutical and Biomedical Analysis* **2009**, *49* (5), 1272-1276.
275. Enk, R.; Eehalt, R.; Graham, J. E.; Bierhaus, A.; Remppis, A.; Greten, H. J., Differential effect of Rhizoma coptidis and its main alkaloid compound berberine on TNF-alpha induced NFkappaB translocation in human keratinocytes. *Journal of ethnopharmacology* **2007**, *109* (1), 170-175.
276. Yokozawa, T.; Ishida, A.; Cho, E. J.; Nakagawa, T., The effects of Coptidis Rhizoma extract on a hypercholesterolemic animal model. *Phytomedicine* **2003**, *10* (1), 17-22.
277. Wang, X. Y.; Shi, Y. L.; Zeng, Y. J., *Chinese Journal of New Drugs* **2003**, *12* (7).
278. Chen, J.; Chen, T., *Chinese Medical Herbology and Pharmacology*. 2004.
279. 徐國鈞, *中國藥材學*. 中國醫藥科技出版社: 北京, 1996.
280. Tanabe, H.; Suzuki, H.; Mizukami, H.; Inoue, M., Double blockade of cell cycle progression by coptisine in vascular smooth muscle cells. *Biochemical Pharmacology* **2005**, *70* (8), 1176-1184.
281. Li, Y.; He, W. Y.; Liu, J. Q.; Sheng, F. L.; Hu, Z. D.; Chen, X. G., Binding of the bioactive component Jatrorrhizine to human serum albumin. *Biochimica Et Biophysica Acta-General Subjects* **2005**, *1722* (1), 15-21.
282. Chang, Y. L.; Usami, S.; Hsieh, M. T.; Jiang, M. J., Effects of palmatine on isometric force and intracellular calcium levels of arterial smooth muscle. *Life Sciences* **1999**, *64* (8), 597-606.
283. Goldberg, D. E., *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley: New York, 1989.

284. Lucasius, C. B.; Kateman, G., Understanding and Using Genetic Algorithms .2. Representation, Configuration and Hybridization. *Chemometrics and Intelligent Laboratory Systems* **1994**, 25 (2), 99-145.
285. Yang, L. J., Plasma surface hardening of ASSAB 760 steel specimens with Taguchi optimisation of the processing parameters. *Journal of Materials Processing Technology* **2001**, 113 (1-3), 521-526.
286. *International Conference on Harmonization (ICH) of Technical Requirements for the Registration of Pharmaceuticals for Human Use, Validation of analytical procedures, ICH-Q2A, Geneva.* 1996.
287. *Note for guidance on the use of Near-infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations, The European Agency for the Evaluation of Medicinal Products.* February 2003.
288. *PASG Guidelines for the development and Validation of Near-Infrared (NIR) Spectroscopy methods*
PASG NIR Subgroup: 2001.
289. International Conference on Harmonization (ICH) of Technical Requirements for the Registration of Pharmaceuticals for Human Use, Validation of analytical procedures, ICH-Q2B. Geneva, 1996.
290. Workman, J.; Weyer, L., *Practical guide to interpretive near-infrared spectroscopy.* CRC Press: 2007.
291. Upton, R., *Reishi mushroom : Ganoderma lucidum : standards of analysis, quality control, and therapeutics.* American Herbal Pharmacopoeia: Santa Cruz, Calif, 2000.
292. *The state pharmacopoeia commission of people's republic of China, pharmacopoeia of the 313 People's Republic of China, Vol. 1.* Chemical Industry Press: Beijing, 2005.
293. Mau, J. L.; Lin, H. C.; Chen, C. C., Non-volatile components of several medicinal mushrooms. *Food Research International* **2001**, 34 (6), 521-526.

294. Shiao, M. S.; Lee, K. R.; Lin, L. J.; Wang, C. T., Natural products and biological activities of the Chinese medical fungus *Ganoderma Lucidum*. *American Chemical Society Symposium Series* **1994**, 547, 342-354.
295. Kim, H. W.; Kim, B. K., Biomedicinal triterpenoids of *Ganoderma Lucidum* (Curt.: Fr.) P. Karst (Aphyllophoromy- cetideae). *International Journal of Medicinal Mushrooms* **1999**, 1, 121-138.
296. Kubota, T.; Asaka, Y.; Miura, I.; Mori, H., Structures of Ganoderic Acid A and B, Two New Lanostane Type Bitter Triterpenes from *Ganoderma Lucidum* (FR.) KARST. *Helvetica Chimica Acta* **1982**, 65 (2), 611-619.
297. Chairul, S. M.; Hayashi, Y., Lanostanoid triterpenes from *Ganoderma applanatum*. *Phytochemistry* **1994**, 35 (5), 1305-1308.
298. Kohda, H.; Tokumoto, W.; Sakamoto, K.; Fjii, M.; Hirai, Y.; Yamasaki, K.; Komoda, Y.; Nakamura, H.; Ishihara, S.; Uchida, M., The Biologically Active Constituents of *Ganoderma lucidum* (FR.) KARST. Histamine Release-Inhibitory Triterpenes. *Chemical & Pharmaceutical Bulletin* **1985**, 33 (4), 1367-1374.
299. Nishitoba, T.; Sato, H.; Shirasu, S.; Sakamura, S., Evidence on the strainspecific terpenoid pattern of *Ganoderma lucidum*. *Agricultural and biological chemistry* **1986**, 50, 2151-2154.
300. Sye, W. F., Improvement method of extraction and High-Performance Liquid Chromatographic separation of ganoderic acids form *Ganoderma Lucidum* *Journal of the Chinese Chemical Society* **1991**, 38, 179-182.
301. Chyr, R.; Shiao, M. S., Liquid chromatographic characterization of the triterpenoid patterns in *Ganoderma Lucidum* and related species. *Journal of chromatography* **1991**, 542, 327-336.
302. Su, C. H.; Yang, Y. Z.; Ho, H. O.; Hu, C. H.; Sheu, M. T., High-performance liquid chromatographic analysis for the characterization of triterpenoids from *Ganoderma*. *Journal of Chromatographic Science* **2001**, 39 (3), 93-100.
303. Tang, W.; Gu, T.; Zhong, J. J., Separation of targeted ganoderic acids from *Ganoderma lucidum* by reversed phase liquid chromatography with ultraviolet and

mass spectrometry detections. *Biochemical Engineering Journal* **2006**, *32* (3), 205-210.

304. Yang, M.; Wang, X.; Guan, S.; Xia, J.; Sun, J.; Guo, H.; Guo, D. A., Analysis of triterpenoids in *Ganoderma Lucidum* using Liquid Chromatography Coupled with electrospray ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry* **2007**, *18* (5), 927-939.

305. Guo, X. Y.; Han, J.; Ye, M.; Ma, X. C.; Shen, X.; Xue, B. B.; Che, Q. M., Identification of major compounds in rat bile after oral administration of total triterpenoids of *Ganoderma lucidum* by high-performance liquid chromatography with electrospray ionization tandem mass spectrometry. *Journal of pharmaceutical and biomedical analysis* **2012**, *63*, 29-39.

306. Ahmad, S., *Oxidative stress and anti-oxidant defences in biology*. Chapman and Hall: New York, 1995.

307. Kim, D. H.; Shim, S. B.; Kim, N. J.; Jang, I. S., Beta-glucuronidase-inhibitory activity and hepatoprotective effect of *Ganoderma lucidum*. *Biological & Pharmaceutical Bulletin* **1999**, *22* (2), 162-4.

308. Shieh, Y. H.; Liu, C. F.; Huang, Y. K.; Yang, J. Y.; Wu, I. L.; Lin, C. H.; Li, S. C., Evaluation of the hepatic and renal-protective effects of *Ganoderma lucidum* in mice. *The American Journal of Chinese Medicine* **2001**, *29* (3-4), 501-507.

309. Lin, J. M.; Lin, C. C.; Chen, M. F.; Ujiie, T.; Takada, A., Radical scavenger and antihepatotoxic activity of *Ganoderma formosanum*, *Ganoderma lucidum* and *Ganoderma neo-japonicum*. *Journal of Ethnopharmacology* **1995**, *47* (1), 33-41.

310. Tasaka, K.; Mio, M.; Isushi, K.; Akagi, M.; Mikino, T., Anti-allergic constituents in the culture medium of *Ganoderma Lucidum* (II). *Agents Actions* **1998**, *23*, 157-160.

311. Kabir, Y.; Kimura, S.; Tamura, T., Dietary effect of *Ganoderma lucidum* mushroom on blood pressure and lipid levels in spontaneously hypertensive rats (SHR). *Journal of nutritional science and vitaminology* **1988**, *34* (4), 433-8.

312. Gao, Y.; Zhou, S.; Huang, M.; Xu, A., Antibacterial and Antiviral Value of the Genus *Ganoderma* P. Karst. Species (Aphyllophoromycetideae): A Review.

International Journal of Medicinal Mushrooms **2003**, *5*, 235-246.

313. Kendrick, B., *The Fifth Kingdom*. Mycologue Publication: Waterloo, 1985.

314. Su, C. Y.; Shiao, M. S.; Wang, C. T., Differential effects of ganodermic acid S on the thromboxane A₂-signaling pathways in human platelets. *Biochemical Pharmacology* **1999**, *58* (4), 587-595.

315. Teow, S. S., The effective application of Ganoderma nutraceuticals. In *Recent Progress in Ganoderma lucidum Research*, Kim, B. K.; Moon, C. K.; Kim, T. S., Eds. The Pharmaceutical Society of Korea: Seoul, 1997; pp 21-39.

316. Hikino, H.; Ishiyama, M.; Suzuki, Y.; Konno, C., Mechanisms of hypoglycemic activity of ganoderan B: a glycan of Ganoderma lucidum fruit bodies. *Planta medica* **1989**, *55* (5), 423-428.

317. Morigiwa, A.; Kitabatake, K.; Fujimoto, Y.; Ikekawa, N., Angiotensin converting enzyme-inhibitory triterpenes from Ganoderma lucidum. *Chemical & pharmaceutical bulletin* **1986**, *34* (7), 3025-3028.

318. Kim, B. K.; Cho, H. Y.; Kim, J. S.; Kim, H. W.; Choi, E. C., Studies on constituents of higher fungi of Korea (LXVIII). Antitumor components of the cultured mycelia of Ganoderma lucidum. *Korea Journal of Pharmacology* **1993**, *24*, 203-212.

319. Li, W.-J.; Nie, S.-P.; Liu, X.-Z.; Zhang, H.; Yang, Y.; Yu, Q.; Xie, M.-Y., Antimicrobial properties, antioxidant activity and cytotoxicity of ethanol-soluble acidic components from Ganoderma atrum. *Food and Chemical Toxicology* **50** (3??), 689-694.

320. Yoon, S. Y.; Eo, S. K.; Kim, Y. S.; Lee, C. K.; Han, S. S., Antimicrobial activity of Ganoderma Lucidum extract alone and in combination with some antibiotics. *Archives of Pharmacal Research* **1994**, *17* (6), 438-442.

321. Russell, R.; M. Paterson, Ganoderma - A therapeutic fungal biofactory. *Phytochemistry* **2006**, *67* (18), 1985-2001.

322. Kim, H. W.; Shim, M. J.; Choi, E. C.; Kim, B. K., Inhibition of cytopathic effect of human immunodeficiency virus-1 by water-soluble extract of Ganoderma Lucidum.

Archives of Pharmacal Research **1997**, *20*, 425-431.

323. Sone, Y.; Okuda, R.; Wada, N.; Kishida, E.; Misaki, A., Structure and antitumor activities of polysaccharides isolated from fruiting body and the growing culture of mycelium of *Ganoderma lucidum*. *Agricultural and Biological Chemistry* **1985**, *49*, 2641-2653.

324. Hsu, M. J.; Lee, S. S.; Lin, W. W., Polysaccharide purified from *Ganoderma lucidum* inhibits spontaneous and Fas-mediated apoptosis in human neutrophils through activation of the phosphatidylinositol 3 kinase/Akt pathways. *Journal of leukocyte biology* **2002**, *72* (1), 207-216.

325. Hsu, M. J.; Lee, S. S.; Lee, S. T.; Lin, W. W., Signaling mechanisms of enhanced neutrophil phagocytosis and chemotaxis by the polysaccharide purified from *Ganoderma lucidum*. *British journal of pharmacology* **2003**, *139* (2), 289-98.

326. Wang, S. Y.; Hsu, M. L.; Hsu, H. C.; Tzeng, C. H.; Lee, S. S.; Shiao, M. S.; Ho, C. K., The antitumor effect of *Ganoderma lucidum* is mediated by cytokines released from activated macrophages and T lymphocytes. *International Journal of Cancer* **1997**, *70* (6), 699-705.

327. Min, B. S.; Gao, J. J.; Nakamura, N.; Hattori, M., Triterpenes from the spores of *Ganoderma lucidum* and their cytotoxicity against Meth-A and LLC tumor cells. *Chemical & Pharmaceutical Bulletin* **2000**, *48* (7), 1026-1033.

328. Min, B. S.; Nakamura, N.; Miyashiro, H.; Bae, K. W.; Hattori, M., Triterpenes from the spores of *Ganoderma lucidum* and their inhibitory activity against HIV-1 protease. *Chemical & Pharmaceutical Bulletin* **1998**, *46* (10), 1607-1612.

329. El-Mekkawy, S.; Meselhy, M. R.; Nakamura, N.; Tezuka, Y.; Hattori, M.; Kakiuchi, N.; Shimotohno, K.; Kawahata, T.; Otake, T., Anti-HIV-1 and anti-HIV-1-protease substances from *Ganoderma Lucidum*. *Phytochemistry* **1998**, *49* (6), 1651-1657.

330. Sliva, D.; Labarrere, C.; Slivova, V.; Sedlak, M.; Lloyd, F. P., Jr.; Ho, N. W., *Ganoderma lucidum* suppresses motility of highly invasive breast and prostate cancer cells. *Biochemical and Biophysical Research Communications* **2002**, *298* (4), 603-612.

331. Sliva, D.; Sedlak, M.; Slivova, V.; Valachovicova, T.; Lloyd, F. P., Jr.; Ho, N. W., Biologic activity of spores and dried powder from *Ganoderma lucidum* for the inhibition of highly invasive human breast and prostate cancer cells. *Journal of Alternative and Complementary Medicine (New York, N.Y.)* **2003**, *9* (4), 491-497.
332. Cheung, W. M.; Hui, W. S.; Chu, P. W.; Chiu, S. W.; Ip, N. Y., *Ganoderma* extract activates MAP kinases and induces the neuronal differentiation of rat pheochromocytoma PC12 cells. *FEBS letters* **2000**, *486* (3), 291-6.
333. Lin, S. B.; Li, C. H.; Lee, S. S.; Kan, L. S., Triterpene-enriched extracts from *Ganoderma lucidum* inhibit growth of hepatoma cells via suppressing protein kinase C, activating mitogen-activated protein kinases and G2-phase cell cycle arrest. *Life sciences* **2003**, *72* (21), 2381-2390.
334. Sliva, D., *Ganoderma lucidum* in cancer research. *Leukemia research* **2006**, *30* (7), 767-768.
335. Aydemir, G., Research on nutrition and cancer: The importance of the standardized dietary assessments. *Asian Pacific Journal of Cancer Prevention* **2002**, *3* (2), 177-180.
336. Stanley, G.; Harvey, K.; Slivova, V.; Jiang, J.; Sliva, D., *Ganoderma lucidum* suppresses angiogenesis through the inhibition of secretion of VEGF and TGF-beta1 from prostate cancer cells. *Biochemical and Biophysical Research Communications* **2005**, *330* (1), 46-52.
337. Muller, C. I.; Kumagai, T.; O'Kelly, J.; Seeram, N. P.; Heber, D.; Koeffler, H. P., *Ganoderma lucidum* causes apoptosis in leukemia, lymphoma and multiple myeloma cells. *Leukemia research* **2006**, *30* (7), 841-848.
338. Chen, X.; Hu, Z. P.; Yang, X. X.; Huang, M.; Gao, Y.; Tang, W.; Chan, S. Y.; Dai, X.; Ye, J.; Ho, P. C.; Duan, W.; Yang, H. Y.; Zhu, Y. Z.; Zhou, S. F., Monitoring of immune responses to a herbal immuno-modulator in patients with advanced colorectal cancer. *International Immunopharmacology* **2006**, *6* (3), 499-508.
339. Pero, R. W.; Amiri, A.; Sheng, Y.; Welther, M.; Rich, M., Formulation and in vitro/in vivo evaluation of combining DNA repair and immune enhancing nutritional supplements. *Phytomedicine* **2005**, *12* (4), 255-263.

340. Kuo, M. C.; Weng, C. Y.; Ha, C. L.; Wu, M. J., Ganoderma lucidum mycelia enhance innate immunity by activating NF-kappaB. *Journal of Ethnopharmacology* **2006**, *103* (2), 217-22.
341. Lin, J. Y.; Chen, M. L.; Chiang, B. L.; Lin, B. F., Ganoderma tsugae supplementation alleviates bronchoalveolar inflammation in an airway sensitization and challenge mouse model. *International Immunopharmacology* **2006**, *6* (2), 241-51.
342. Benzie, I. F.; Strain, J. J., The ferric reducing ability of plasma (FRAP) as a measure of "antioxidant power": the FRAP assay. *Analytical biochemistry* **1996**, *239* (1), 70-76.
343. Lin, L. J.; Shiao, M. S., Separation of oxygenated triterpenoids from Ganoderma lucidum by high-performance liquid chromatography. *Journal of Chromatography A* **1987**, *410*, 195-200.
344. Chen, Y.; Zhu, S. B.; Xie, M. Y.; Nie, S. P.; Liu, W.; Li, C.; Gong, X. F.; Wang, Y. X., Quality control and original discrimination of Ganoderma lucidum based on high-performance liquid chromatographic fingerprints and combined chemometrics methods. *Analytica Chimica Acta* **2008**, *623* (2), 146-156.
345. Keypour, S.; Rafati, H.; Riahi, H.; Mirzajani, F.; Moradali, M. F., Qualitative analysis of ganoderic acids in Ganoderma lucidum from Iran and China by RP-HPLC and electrospray ionisation-mass spectrometry (ESI-MS). *Food Chemistry* **2010**, *119* (4), 1704-1708.
346. Chen, Y.; Yan, Y.; Xie, M. Y.; Nie, S. P.; Liu, W.; Gong, X. F.; Wang, Y. X., Development of a chromatographic fingerprint for the chloroform extracts of Ganoderma lucidum by HPLC and LC-MS. *Journal of pharmaceutical and biomedical analysis* **2008**, *47* (3), 469-477.
347. Wang, X. M.; Yang, M.; Guan, S. H.; Liu, R. X.; Xia, J. M.; Bi, K. S.; Guo, D. A., Quantitative determination of six major triterpenoids in Ganoderma lucidum and related species by high performance liquid chromatography. *Journal of Pharmaceutical and Biomedical Analysis* **2006**, *41* (3), 838-844.
348. Liu, Y.; Qiu, F.; Di, X., Sensitive and selective liquid chromatography-tandem mass spectrometry method for the determination of five ganoderic acids in

Ganoderma lucidum and its related species. *Journal of Pharmaceutical and Biomedical Analysis* **2011**, 54 (4), 717-721.

349. Da, J.; Wu, W. Y.; Hou, J. J.; Long, H. L.; Yao, S.; Yang, Z.; Cai, L. Y.; Yang, M.; Jiang, B. H.; Liu, X.; Cheng, C. R.; Li, Y. F.; Guo, D. A., Comparison of two officinal Chinese pharmacopoeia species of Ganoderma based on chemical research with multiple technologies and chemometrics analysis. *Journal of chromatography A* **2012**, 1222, 59-70.

350. Prior, R. L.; Wu, X.; Schaich, K., Standardized methods for the determination of antioxidant capacity and phenolics in foods and dietary supplements. *Journal of Agricultural and Food Chemistry* **2005**, 53 (10), 4290-4302.

351. Pulido, R.; Bravo, L.; Saura-Calixto, F., Antioxidant activity of dietary polyphenols as determined by a modified ferric reducing/antioxidant power assay. *Journal of Agricultural and Food Chemistry* **2000**, 48 (8), 3396-3402.

352. Smina, T. P.; Mathew, J.; Janardhanan, K. K.; Devasagayam, T. P., Antioxidant activity and toxicity profile of total triterpenes isolated from Ganoderma lucidum (Fr.) P. Karst occurring in South India. *Environmental Toxicology and Pharmacology* **2011**, 32 (3), 438-446.

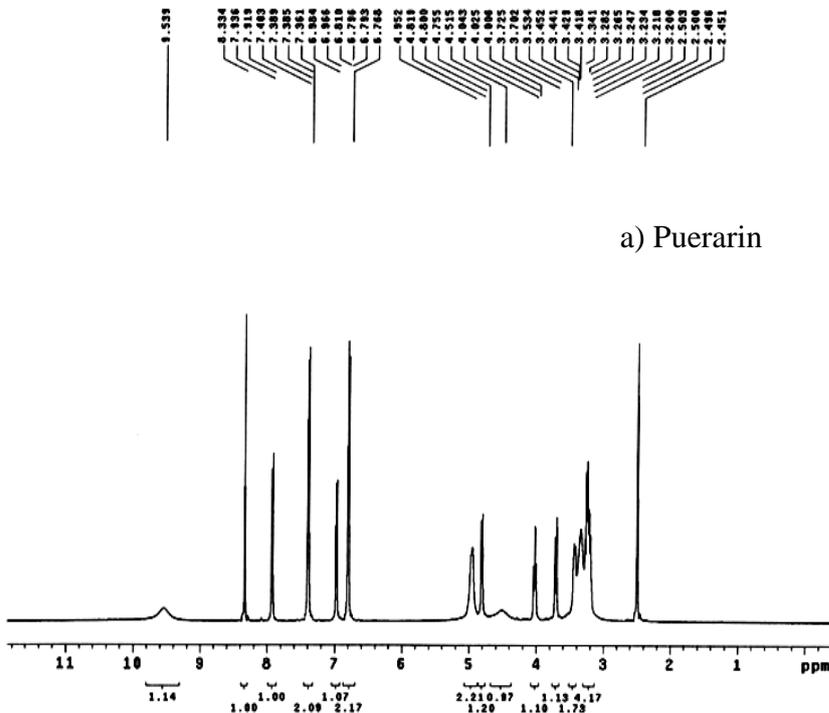
353. Jamrogiewicz, M., Application of the near-infrared spectroscopy in the pharmaceutical technology. *Journal of Pharmaceutical and Biomedical Analysis* **2012**, 66 (0), 1-10.

Appendices

Appendix 4.1 NMR spectrum of a) puerarin, b) daidzin, c) daidzein, d) genistin and d) genistein

```

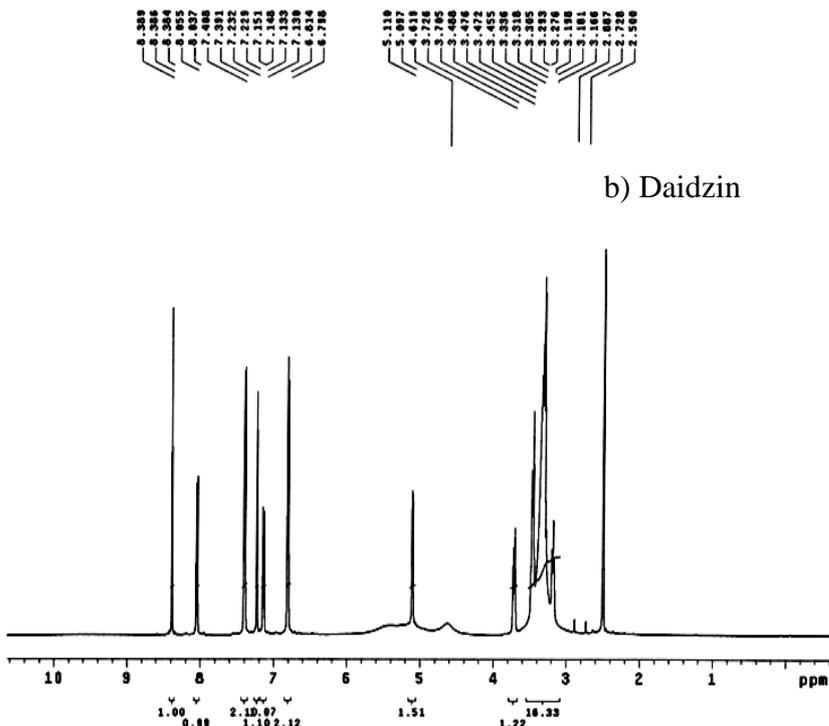
Puerarin
exp1 Proton
SAMPLE SPECIAL
date Oct 24 2008 temp not used
solvent dmsc gain not used
file exp spin not used
ACQUISITION hst 0.008
sw 10494.6 pw90 8.500
at 2.049 alfa 8.000
np 42888
fb 8000 11 FLAGS n
bs 32 1n n
ss 2 dp y
d1 1.000 hs nm
nt 8
ct 8 1b PROCESSING 0.20
      8 fn 65536
tn H1 DISPLAY
sfrq 498.745 sp -231.2
tof 249.7 wp 8182.3
tpwr 57 rfl 3730.2
pw 4.250 rfp 1249.4
DECOUPLER rfp -175.3
dn C13 lp -3.2
dof 0 PLOT
dm nmn wc 180
dms c sc 0
dpwr 50 vs 173
dmf 13766 th 2
      al cdc ph
  
```



a) Puerarin

```

Daidzin
exp1 Proton
SAMPLE SPECIAL
date Oct 24 2008 temp not used
solvent dmsc gain not used
file exp spin not used
ACQUISITION hst 0.008
sw 10494.6 pw90 8.500
at 2.049 alfa 8.000
np 42888
fb 8000 11 FLAGS n
bs 32 1n n
ss 2 dp y
d1 1.000 hs nm
nt 8
ct 8 1b PROCESSING 0.20
      8 fn 65536
tn H1 DISPLAY
sfrq 498.745 sp -328.6
tof 249.7 wp 8048.3
tpwr 57 rfl 3730.8
pw 4.250 rfp 1249.4
DECOUPLER rfp -175.1
dn C13 lp -12.8
dof 0 PLOT
dm nmn wc 180
dms c sc 0
dpwr 50 vs 197
dmf 13766 th 2
      al cdc ph
  
```

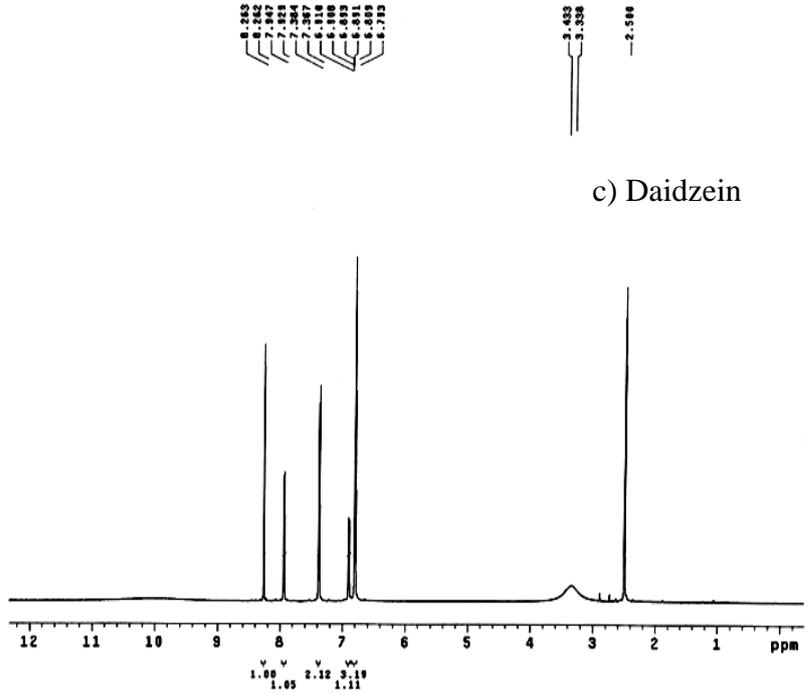


b) Daidzin

```

Daidzein
expl Proton

SAMPLE SPECIAL
date Oct 24 2008 temp not used
solvent dms0 gain not used
file exp spin 20
ACQUISITION hst 0.000
sv 10494.6 pwr0 8.500
at 2.049 a1fa 8.000
np 42888 FLAGS
fb 8000 i1 n
bs 32 in n
ss 2 dp y
dl 1.000 hs nn
nt 8 PROCESSING
ct 8 lb 0.20
TRANSMITTER 8 fn 85536
tn H1 DISPLAY
sfrq 489.745 sp -189.8
tof 249.7 wp 8354.8
tpr 57 rfl 3737.9
pw 4.250 rfp 1249.4
DECOUPLER rp 189.2
dn C13 lp -4.0
dof 0 PLOT
dm nnn wc 189
dm c sc 8
dpwr 50 ve 53
dwt 13788 th cdc ph 3
    
```

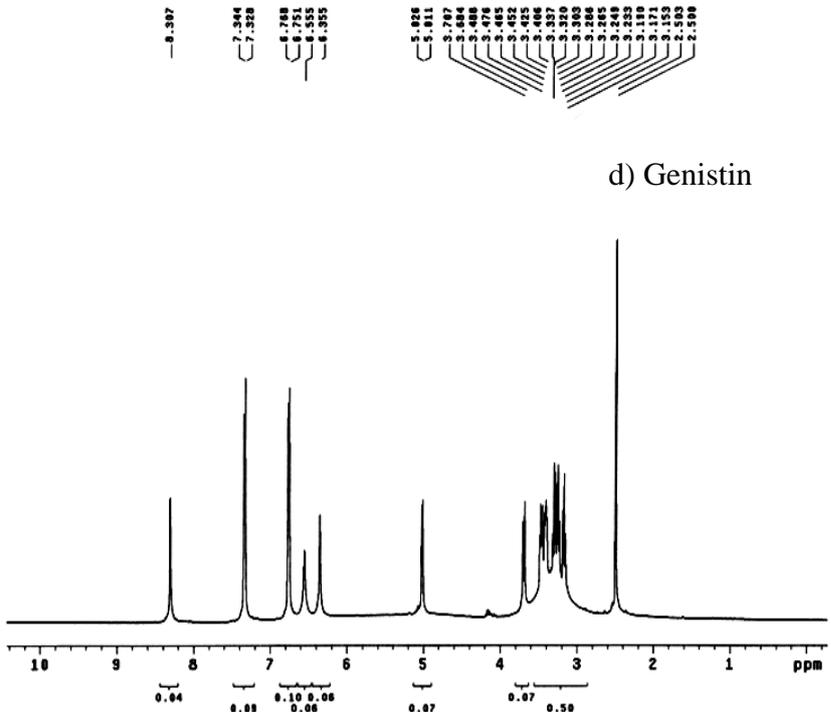


c) Daidzein

```

Genistein
expl Proton

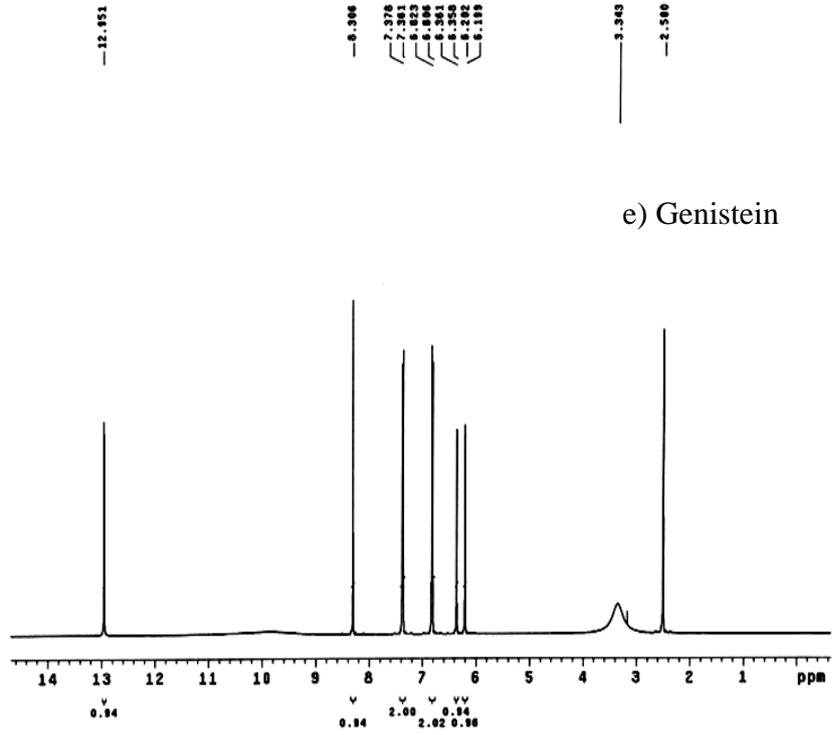
SAMPLE SPECIAL
date Oct 24 2008 temp not used
solvent dms0 gain not used
file exp spin 20
ACQUISITION hst 0.000
sv 10494.6 pwr0 8.500
at 2.049 a1fa 8.000
np 42888 FLAGS
fb 8000 i1 n
bs 32 in n
ss 2 dp y
dl 1.000 hs nn
nt 8 PROCESSING
ct 8 lb 0.20
TRANSMITTER 8 fn 85536
tn H1 DISPLAY
sfrq 489.745 sp -136.7
tof 249.7 wp 5353.9
tpr 57 rfl 3738.1
pw 4.250 rfp 1248.4
DECOUPLER rp 189.8
dn C13 lp -2.0
dof 0 PLOT
dm nnn wc 189
dm c sc 8
dpwr 50 ve 215
dwt 13788 th cdc ph 7
    
```



d) Genistein

```

Genistein
expl Proton
SAMPLE SPECIAL
date Oct 24 2008 temp not used
solvent dms0 gain not used
file exp spin 20
ACQUISITION hst 0.000
sw 10494.8 pu90 0.500
et 2.048 alfa 0.000
np 42000
fb 6000 l1 n
bs 32 ln n
ss 2 dp y
dl 1.000 hs nn
nt 8
ct 8 PROCESSING
TRANSMITTER H1 fb 0.20
fn 65536
tn H1 DISPLAY
efrq 499.745 sp -320.6
tof 249.7 wp 7870.7
tpr 57 rf1 3797.9
pw 4.250 rfp 1249.4
DECOUPLER rp 176.3
dn C13 lp -10.0
dof 0 PLOT
dm nnn wc 100
dms c sc 0
dpr 50 vs 50
dnt 13768 th 5
df at cdc ph
    
```



e) Genistein