This electronic thesis or dissertation has been downloaded from the King's Research Portal at https://kclpure.kcl.ac.uk/portal/



Approaches to Chemical- and Activity-based Standardization of Traditional Chinese Medicine

Zhao, Huiying

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 08. Oct. 2023

Approaches to Chemical- and Activity-based Standardization of Traditional Chinese Medicine

Huiying Zhao

A thesis presented for the degree of
Doctor of Philosophy

Institute of Pharmaceutical Science
School of Biomedical Sciences
King's College London
2014

Acknowledgement

First I would like to express my deepest appreciation to my supervisor Professor Peter J. Hylands. He has a great knowledge on traditional Chinese medicine and phytochemistry. He has been very supportive and understanding during my Ph.D study. Apart from supervising me in my research, his great interpersonal and management skills have also influenced me. I also thank the support from the Centre for Natural Medicines Research which he direct.

I would like to thank Professor Peter Jones and Professor Shanta Persaud from King's College Endocrinology group, for being enthusiastic and supportive during the collaboration, providing lab space and training. I would like to thank Altaf Romaiyan who had showed me cell culture experiment and insulin secretion test. I would also like to thank my colleges Shuang Song, Mustafa, Robert for being my company and helping out in the lab.

In University of East London, I thank Dr. Olivia Corcoran who had been providing lab space and actively collaborating with us in the beginning of the proteomic project. I give my sincerely thank to Dr. Jiayu Gao who had showed me how to do 2D gel electrophoresis, with who I co-authored a paper.

I appreciate Professor Peter Emery in the Department of Nutrition & Dietetics, KCL, who welcomed me to use DIGE equipments in his lab.

In Perkin Elmer I would like to thank Patrick Courtney, who is interested in applying ATR-IR on TCM and welcomed me to his lab. I thank Dr. Ben Perston for the kind help on using ATR-IR instrument and giving helpful comment on the application notes, as a result of our collaboration.

I thank master students Jin Xu, and Helen for preparing some ginseng extracts.

I thank Professor Burkhard Bechinger in University of Strasbourg for providing office space for me to finish my thesis.

I thank Royal science society for providing Oversea Student Scholarship and King's College London for providing Annual Funding Studentship throughout my Ph.D. study.

I would also like to show appreciation to the following people for providing useful discussions: Professor Robert Hider, Dr. Sukhi bansal, Dr. Paul Royall, Dr. David Mountford, Dr. Cristina Legido-Quigley, Dr. Xiaole Kong, and Dr. Rui Fang from Institute of Pharmaceutical Science, Andrew Western and Ricardo Sainz Fuertes from the Institute of Psychiatry.

I appreciate my friend Nan Li who have purchased ginseng samples from China for my research, for her being my friend and accompanying me whenever I am happy, crazy, or sad.

I thank my parents for their endless love and support during my Ph.D. Without whom, I would not pursue my study in the UK. I thank them for taking care of me in the UK when I was in severe illness, and for overcoming great difficulties without speaking English.

My biggest thank is to Dr. Louic Vermeer, who is my husband, friend and colleague, who gave me enormous support in life and work. He gave useful suggestions of my thesis, and often discussed with me about my research. He took care of me and being strong together with me.

I also want to thank my cat "Hulu". She brought me so much laughter during my thesis writing!

Abstract

There are difficulties in the standardization of traditional Chinese medicines (TCM) because of the complexity of TCM preparations. This leads to regulatory problems. In the present work, several Chinese medicinal plants were investigated including Danshen(*Salvia miltiorrhiza* Bunge), Sanqi(*Panax notoginseng* Burk.), compound Danshen dripping pill, and several species of the genus *Panax*.

Plant materials and commercial products containing these taxa were analysed by a range of techniques, including chemical- and activity-based standardization methods: the chemical techniques used include high performance liquid chromatographic (HPLC), nuclear magnetic resonance (NMR) spectroscopic and infrared (IR) spectroscopic methods, followed by chemometric analysis.

Different brands of Danshen finished products, Danshen plant root materials of different origin, and Sanqi plant root materials with different production dates were successfully discriminated from each other. Asian and American ginseng were successfully discriminated, and arginine was found for the first time as a main difference, together with other primary and secondary metabolites, using ^{1}H-NMR-PCA. For wild and cultivated American ginseng which had reputed activity in diabetes, an activity-based standardization method based on a cell viability test and proteomic analysis (using two-dimensional difference gel electrophoresis (2D-DIGE) analysis) of mouse insulinoma (MIN6) cells was performed. 83 proteins were found to be significantly up or down-regulated.

This work shows that HPLC, and NMR and IR spectroscopy coupled with PCA methods are applicable to standardization methodologies. The

advantages and disadvantages of HPLC, and NMR and IR spectroscopic technology are discussed and compared. The 2D-DIGE protein profile can be correlated with TCM treatments.

This work contributes to the current problem of regulatory control of TCM preparation as traditional medicines and may suggest a practical way forward.

Abbreviations

2D-GE = two dimensional gel electrophoresis
ASG = Asian ginseng
ATP = adenosine triphosphate
ATR = attenuated total reflectance
BSA = bovine serum albumin
CAG = cultivated American ginseng
CDDP = compound Danshen dripping pill
CP = Chinese Pharmacopoeia
DAD = photodiode array detection
DIGE = difference in gel electrophoresis
DMSO = dimethyl sulfoxide
DSS = 4,4-dimethyl-4-silapentane-1-sulfonic acid
DTT = dithiothreitol
EDTA = ethylenediaminetetraacetic acid
ELSD = evaporative light scattering detectors
FBS = fetal bovine serum
FDA = the Food and Drug Administration
FDR = false discovery rate
GC = gas chromatography
HCA = hierarchical cluster analysis
HPLC = high performance liquid chromatography
IEF = iso-electric focusing
IPG = immobilised pH gradient
IR = infrared spectroscopy
MDA = multivariate data analysis

MIN = mouse insulinoma
MS = mass spectrometry
MTT = 3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide
NL = non linear
NMR = nuclear magnetic resonance
PAT = process analytical technology
PBS = phosphate buffered saline
PCA = principal component analysis
PEG = polyethylene glycol
PLS-DA = partial least squares discriminant analysis
PTM = post-translational modification
QA = quality assurance
QC = quality control
RAPD = random amplification of polymorphic DNA
SAB = salvianolic acid B
SDS = sodium dodecyl sulfate
TCM = traditional Chinese medicine
TLC = thin - layer chromatography
TMS = tetramethylsilane
UV = ultraviolet
WAG = wild American ginseng

Contents

Abstract **3**

Abbreviations **5**

List of Figures **10**

List of Tables **10**

1 General introduction **17**

1.1 Traditional Chinese medicine 17

1.1.1 Current quality control methods for TCM and the problems 19

1.2 Chemical-based standardization of TCM 20

1.2.1 Chromatographic methods used in quality control of TCM 22

1.2.2 Spectroscopic methods used in quality control of TCM 24

1.2.3 Statistical methods used in the quality control of TCM 25

1.2.4 Metabolomic 26

1.2.5 Metabolomics quality control of TCM 27

1.3 Activity-based standardization 28

1.4 Structure of this thesis 29

1.5 Aim of study 30

2 Evaluation of QC methods **32**

2.1 Introduction 32

2.1.1 Danshen and Sanqi plant 34

2.1.2 Compound Danshen dripping pill 34
2.1.3 The general manufacturing process of dripping pills . . 37
2.1.4 Primary and secondary metabolites 38
2.1.5 Major compounds and their functions in compound Danshen dripping pill 40
2.1.6 Current quality standards and the limitations 41
2.2 Materials and methods 45
2.2.1 Materials 45
2.2.2 Sample preparation 47
2.2.3 HPLC analysis 49
2.2.4 NMR spectroscopy 49
2.2.5 IR spectroscopic analysis 50
2.2.6 Sensitivity test 51
2.2.7 Data pretreatment 52
2.2.8 Multivariate data analysis (MVA) 58
2.3 Results and discussion 60
2.3.1 Comparison of IR sampling methods 60
2.3.2 Crystallization affects the results of ATR-IR test . . . 62
2.3.3 Visual inspection of HPLC chromatogram, NMR and IR spectra 67
2.3.4 Multivariate data analysis 68
2.3.5 Sensitivity test 69
2.3.6 Discrimination of Danshen samples from different regions 73
2.3.7 Differentiation of Sanqi samples with different supply date 79
2.3.8 Differentiation of different batches of CDDP 85
2.3.9 Comparison of analytical approaches for raw material and finished products 87
2.3.10 Overall Comparison of HPLC, NMR and IR spectroscopy for Danshen and Sanqi product 89
2.4 Conclusion 90

3 Metabolomic quality control 93
3.1 Introduction . . . 93
3.1.1 Ginseng species . . . 93
3.1.2 History of use . . . 94
3.1.3 Morphology characters of ginseng samples . . . 95
3.1.4 Constituents of ginseng in general . . . 97
3.1.5 Current regulation methods . . . 98
3.1.6 Advanced quality control methods . . . 99
3.1.7 Aim of study . . . 101
3.2 Materials and methods . . . 101
3.2.1 *Panax* samples studied . . . 101
3.2.2 Chemicals used . . . 102
3.2.3 Ginseng roots extraction method . . . 102
3.2.4 NMR measurement . . . 104
3.2.5 Data analysis . . . 104
3.3 Results and discussion . . . 105
3.3.1 NMR spectra of 4 ginseng samples . . . 105
3.3.2 Assignment of peaks . . . 112
3.3.3 Differentiation of ginseng types based on their sugar contents . . . 114
3.3.4 Differentiation of *Panax* species following excision of the sugar region . . . 119
3.3.5 Difference among wild American ginseng samples . . . 132
3.3.6 Difference of a suspected counterfeit cultivated American ginseng sample number 20 . . . 133
3.4 Conclusion . . . 134

4 Proteomic quality control 137
4.1 Introduction . . . 137
4.1.1 Various "Omics" techniques used in TCM study . . . 138
4.1.2 Metabolomics and metabonomics . . . 138
4.1.3 Genomics . . . 139
4.1.4 Proteomics . . . 141

4.1.5 Diabetes and MIN6 cells . . . 148
4.1.6 Herbal medicine including American ginseng used in treatment of diabetes . . . 149
4.2 Methods . . . 149
4.2.1 Preparation of ginseng extract . . . 149
4.2.2 Cell culture . . . 150
4.2.3 Estimation of cell number . . . 153
4.2.4 CellTiter-Glo® Luminescent Cell Viability Assay . . . 153
4.2.5 Protein extraction and purification . . . 154
4.2.6 Protein quantification . . . 155
4.2.7 Experimental design . . . 155
4.2.8 CyDye labelling . . . 157
4.2.9 Two dimensional electrophoresis . . . 158
4.2.10 DIGE gel scanning . . . 160
4.2.11 Protein statistics . . . 160
4.2.12 Protein identification . . . 161
4.2.13 Data processing and database searching . . . 163
4.3 Results and discussion . . . 163
4.3.1 Viability test . . . 163
4.3.2 A brief analysis of the DIGE images . . . 166
4.3.3 Issue with protein post-translational modifications . . . 170
4.3.4 Expression change of protein M2659 as an example . . . 172
4.3.5 Effect of cultivated American ginseng on protein expression of MIN6 cells . . . 177
4.3.6 Effect of wild American ginseng on protein expression of MIN6 cells . . . 182
4.3.7 Change of protein expression between treatments by wild and cultivated American ginseng . . . 185
4.3.8 Protein identification . . . 188
4.4 Conclusion and future studies . . . 189

5 General conclusion 192

List of Figures

1.1 Yearly publications on quality assessment of herbal medicines using HPLC and NMR spectroscopic techniques. 21

2.1 Commercially available Danshen and Sanqi roots. 35
2.2 Commercially available compound Danshen products. 36
2.3 Structure of representative diterpenoid quinones in Danshen. . 39
2.4 Representative saponins from Sanqi 40
2.5 A map of Chinese provinces 47
2.6 Alignment of HPLC chromatograms of 5 batches of CDDP from 16 to 20 minutes. 53
2.7 Effect of normalization on ^{1}H-NMR spectra 5 batches of Danshen from 3.5 to 4.5 ppm. 56
2.8 IR spectra of different SAB-spiked Danshen extract 2,000 cm^{-1} - 650 cm^{-1}, with and without first derivative. 59
2.9 Top spectra: ATR-IR spectrum of extraction of CDDP using evaporation method at 1073 second. Bottom spectrum: ATR-IR spectrum of extraction of CDDP using evaporation method at 972 second 63
2.10 Total absorption of ATR-IR spectrum of 70% aqueous methanol extraction of CDDP over time. 64
2.11 CDDP measured using a microscope with a polarization mirror, temperature raising from 30.0°C to 70.0°C 66
2.12 Overlapped HPLC chromatograms of 5 Danshen no. 1-5. . . . 67
2.13 PCA scores plot of HPLC chromatogram dataset of 6 non-spiked CDDP sample 70

2.14 PCA scores plot of IR spectra of Danshen extract spiked by 1% to 20% SAB. 72
2.15 Scores plot PC1 versus PC2 obtained from HPLC fingerprints of Danshen from different regions. 74
2.16 Scores plot PC1 versus PC2 obtained from ATR-IR spectroscopic dataset of Danshen from different regions. 77
2.17 Scores plot PC1 versus PC2 was obtained from ATR-IR spectroscopic dataset of Sanqi supplied in different dates. 80
2.18 NMR spectra of 5 batches of Sanqi supplied in different dates. 82
2.19 Scores plot PC1 versus PC2 was obtained from NMR spectroscopic dataset of Sanqi supplied in different dates. 83
2.20 Scores plot PC1 versus PC2 obtained from HPLC dataset of CDDP produced by Tasly with different batch numbers. . . . 85
2.21 Scores plot PC1 versus PC2 obtained from ATR-IR spectroscopic dataset of CDDP produced in different years. 86
2.22 Correlated raw material-Danshen and intermediate material intermediate products A–E made from Danshen A–E and Sanqi A-E respectively 88

3.1 Representative photos of 6 ginseng samples. 96
3.2 A general structure of ginsenosides. 98
3.3 Representative NMR spectra of 4 ginseng types. 106
3.4 Structures of sucrose, glucose and xylose. 107
3.5 Structures of assigned amino acids in the ^{1}H-NMR spectra of ginseng extract 108
3.6 Structures of assigned organic acids in the ^{1}H-NMR spectra of ginseng extract 109
3.7 Assignment of primary metabolites in the ^{1}H-NMR spectrum of Asian ginseng sample no.19, from 0.9 to 3.3 ppm. 110
3.8 Assignment of primary metabolites in the ^{1}H-NMR spectrum of Asian ginseng sample no.19, from 4.2 to 8.5 ppm. 111

3.9 PCA scores plot of ^{1}H-NMR spectra of 4 ginseng types including wild and cultivated American ginseng, Asian ginseng, sanqi ginseng and sample 20. 115
3.10 PCA loadings plot of ^{1}H-NMR spectra of 4 ginseng samples. . 116
3.11 PCA scores plot of 4 ginseng samples and sample 20. 120
3.12 PCA scores plot of 3 *Panax* species. 121
3.13 PCA loadings plot of 3 *Panax* species correlating to the PCA scores plot in figure 3.12. 123
3.14 Enlarged L1 of PCA loadings plot of 3 *Panax* species focusing on the positive side of PC1. 124
3.15 Relative levels of selected metabolite (from figure 3.14) that differentiate between cultivated American ginseng (CAG), wild American ginseng (WAG) and Asian ginseng (ASG). 126
3.16 Enlarged L2 area of PCA loadings plot of 3 *Panax* species focusing on the positive side of PC2. 129
3.17 Enlarged L3 area of PCA loadings plot of 3 *Panax* species focusing on the negative side of PC1 and PC2. 130
3.18 Relative levels of selected metabolites (from figure 3.16 and figure 3.17) in cultivated American ginseng (CAG), wild American ginseng (WAG) and Asian ginseng (ASG). 131

4.1 Simplified schematic representation showing how to create a pooled standard in DIGE. 144
4.2 A simplified illustration of a haemocytometer grid and glass slides. 152
4.3 Cell titre viability test of MIN6 cells treated by a serial concentration of wild and cultivated American ginseng 164
4.4 Representative 2D-DIGE image of proteins extracted from MIN6 cells treated by American ginseng. 167
4.5 Representative 2D and 3D DIGE images of three adjacent spots 2611, 2601 and 2604 shift into each other after treated with American ginseng. 171

4.6 Representative 2D and 3D DIGE images of protein M1990 before and after treatment. 173
4.7 Expression change of protein M2659 treated by cultivated and wild American ginseng for 48 hours. 174
4.8 Plot of standard abundance of protein M2659 in CAG treated and control MIN6 cells. 176
4.9 The changed ratio of standard abundance of selected proteins extracted from control and CAG treated MIN6 cells. Top figure: An up-regulating trend is shown in all 24, 48 and 66 hours. Bottom figure: Where the ratio goes up above 1.2 in 48 and 66 hours treatment, but not showing a strong tendency of increasing in 24 hours. 178
4.10 The changed ratio of standard abundance of selected proteins extracted from control and CAG treated MIN6 cells.Top figure: where the protein up-regulated in 48 and 66 hours treatment. Bottom figure: protein up-regulated only after 66 hours treatment, but not after 24 or 66 hours. 180
4.11 The changed ratio of standard abundance of selected proteins extracted from control and CAG treated MIN6 cells. Top figure: Proteins were down regulated in one or more time points of 24, 48 and 66 hours. Bottom figure: Proteins showed different regulation trend in 24, 48 and 66 hours. 181
4.12 The changed ratio of standard abundance of selected proteins extracted from control and WAG treated MIN6 cells. The proteins in this group are up-regulated after treatment. . . . 183
4.13 The changed ratio of standard abundance of selected proteins extracted from control and WAG treated MIN6 cells. The proteins in this group are down-regulated after treatment. . . 184
4.14 Comparison of proteins treated by wild and cultivated American ginseng. 186

List of Tables

2.1 Details of herbal materials studied in this chapter. 46
2.2 Details of CDDP studied. 48
2.3 Comparison of the derivatized spectra of Sanqi plant root, ginsenoside Re and Rb_1. Unit in cm^{-1}. 81
2.4 Ability of distinguishing minor difference of TCM samples. . 89

3.1 Chemical structures of some major ginsenosides. 99
3.2 Ginseng studied in this chapter 103
3.3 Extra information for ginseng sample 1-6. 104
3.4 Assignment of sucrose, glucose and xylose- on ^{1}H-NMR spectrum of a representative Asian ginseng sample. 112
3.5 Assignment of 18 primary metabolites on ^{1}H-NMR spectrum of a representative Asian ginseng sample. 113
3.6 Summary of previous reports on the ^{1}H-NMR assignments of major ginsenosides including ginsenoside Rb_1, Rg_1, Re, Rd and notoginsenoside R1. 128
3.7 Summary of difference in metabolites between CAG, WAG and ASG. 135

4.1 Experimental design of CyDye labelling. 156
4.2 A summary of number of proteins in DIGE gel of CAG and WAG treated and control MIN6 cells 168
4.3 Comparison of proteins treated by wild and cultivated American ginseng. 187

4.4 MIN6 cell protein identified by 2D gel electrophoresis and ELSD-MS. 188

Chapter 1

General introduction

1.1 Traditional Chinese medicine

Traditional Chinese medicine (TCM) is mainly based on the *Yin-yang* and five element theory (metal, wood, water, fire and earth) in traditional Chinese philosophy. TCM include herbal medicines, animal products, minerals, and metals. A TCM preparation is often a mixture consisting of up to a dozen TCM acting as the "emperor", "minister", "assistant" and "guide" respectively. The "emperor" is the main medicine targeting as the disease. The "minister" medicine assists the "emperor" or to target at the accompanying symptom. The "assistant" may inhibit the toxic effect of the "emperor" and the "minister" and the "guide" leads all the TCM to the diseased region. In a TCM preparation, the "emperor" is always present, while the others are not always necessary. TCM practitioners prescribe an individualised mixture according to each patient's symptoms and physical condition. However, there are also famous combinations of TCM that have been repeatedly prescribed. Traditionally, the prescribed TCM is boiled by the patients and the soap is consumed. Currently, preprepared TCM are made into tablets, pills, granulations, and liquids for the convenience.

The use of TCM was first recorded from the years 2,698 - 2,596 BC by the supposed yellow emperor. The book *Neijing Suwen* or *Inner Canon: Basic Questions*, also known as the *Huangdi Neijing* (*Yellow Emperor's Inner*

Canon) was written in that period. In the *Han* Dynasty (202 BC - 220 AD), TCM flourished. The representative character is Zhang Zhongjing who wrote *Shang Han Za Bing Lun* which was composed of two parts "Treatise on Cold Damage" - focusing on febrile conditions attributed to "Cold", and the *Jingui yaolue* ("*Essentials of the Golden Cabinet*") - focusing on "miscellaneous illnesses". *Shang Han Za Bing Lun* is the first literature of TCM including theories, methods, treatments and medicines which has great significance in the history of TCM.

In the *Ming* Dynasty (1,368-1,644 AD), Li Shizhen wrote *Ben Cao Gang Mu* (Compendium of Materia Medica) which is regarded as the most complete and comprehensive medical book ever written on TCM. It lists all the plants, animals, minerals, and other objects which were believed then to have medicinal functions.

TCM has been widely used in China, Korea, Japan, *etc.* Coming to the present day, natural products have been a productive source of drug discovery. Over a 100 natural product derived medicine are undergoing clinical development, especially anti-cancer and anti-infective medicines[1]. Nowadays, TCM have been found to have effects on anaemia[2], dysmenorrhea[3], cardiovascular disease[4] *etc.*

TCM has drawn the world's attention. Around 40% Americans with chronic diseases have used complementary and alternative medicines[5], while TCM is the most commonly used. In the past five years, the Chinese government has invested 130 million US dollars for TCM research and development[6].

TCM has been extensively reviewed on the treatment of diabetes [7], hepatitis B virus infection [8]. TCM has been used to treat cancer widely in China [9].

Being widely used in the world, the quality, safety and efficacy of TCM became the biggest concern. In this chapter, current situation, various methods used in quality control of TCM are briefly reviewed. Chapters 2, 3 and 4 consist of more information in their introduction sections.

1.1.1 Current quality control methods for TCM and the problems

Currently there are some 3,000 Chinese practitioners in the UK. Although some parts of TCM have been found to be effective, it has had difficulties in being accepted by western countries. Standardization is a prerequisite for western clinical evaluation of the materials if they are to be successfully registered as medicinal products. There are several difficulties for the definition of Chinese Medicines. Firstly, an individual plant species contains many (active and inactive) substances and the medicinal remedies often contain many different plants used together. Purification of the various compounds turns out to be difficult and time consuming. The identification of all constituents in one single plant would be prohibitively expensive. Secondly, the same species of plant grown in different places[10] or harvested in different years [11] may have different constituents. Thirdly, the health care policy needs to be modified in order to bring TCM into western countries because herbals cannot be simply treated the same as chemical compounds. Conventional chemical drugs only consist one or a few chemicals, therefore quality control is straightforward. For botanical drugs, with more than hundreds of chemicals present, precise control of each chemical is impossible. Till now, the underlying ethnopharmacology mechanisms of none of the herbal medicine are fully explained. Therefore, it is not sufficient to say that any a few biomarkers could represent the true quality of any TCM. Finally, an effective method is required in order to analyse the numerous constituents. Due to all these difficulties, herbal materials are sometimes unidentified and almost always unstandardised[12].

Recently, modern technologies have been introduced in the quality control of TCM. In the 2005 Chinese Pharmacopoeia[13], 1,523 items are tested by TLC, 518 items are tested by high performance liquid chromatography (HPLC) and 47 items are tested by gas chromatography (GC). Most of the products have only one compound as quality control marker while in recent years some of them have two or more. However, the chemical markers are usually not the only active compounds in the product[14] .

NMR spectroscopy is not included in the standard tests in the Chinese Pharmacopoeia. However, it becomes increasingly interesting as a non-reductive method for quality control of TCM, and is often used together with statistical analysis[15][16].

Despite all the difficulties, some herbal products have been registered as drug. In 2006, Veregen® Ointment (`http://www.veregen.com/`) gained FDA approval (`http://www.accessdata.fda.gov/drugsatfda_docs/nda/2006/021902s000TOC.cfm`), as the first botanical drug. It is used to treat external genital warts. The active ingredient in Veregen® Ointment is Sinecatechins, a partially defined purified water extract from green tea leaves.

1.2 Chemical-based standardization of TCM

Chemical-based standardization, as is named by the author, means using various analytical approaches such as chromatography or spectroscopy to quantify, identify, and analyse the chemical composition of TCM, and therefore control the quality of TCM. Researchers try to understand the differences in compositions of TCM from different species, different regions[10], different parts of the plants[17], different years of harvesting[11], *etc.* Chemical-based standardization is the most common method used to control the quality of TCM. The interest in quality control of TCM has been growing quickly over the last few years.

Figure 1.1 was generated using the following search conditions on PubMed:

(authentication[Title/Abstract] OR quality control[Title/Abstract] OR metabolomics[Title/Abstract] OR fingerprinting[Title/Abstract]) AND (herbal medicine[Text word] OR Chinese medicine[Text word]) AND HPLC [Title/Abstract] AND ("20xx/01/01"[PDAT] : "20xx/12/31"[PDAT])

For the papers related to NMR or IR, "HPLC" was substituted by into "NMR" or "IR" in the searching terms. The years indicated in "20xx" need to be replaced accordingly. For IR, only 17 papers ever published and many of them came from a low quality journal. Therefore it was not presented in this figure.

The results shown were manually checked and most of them are related

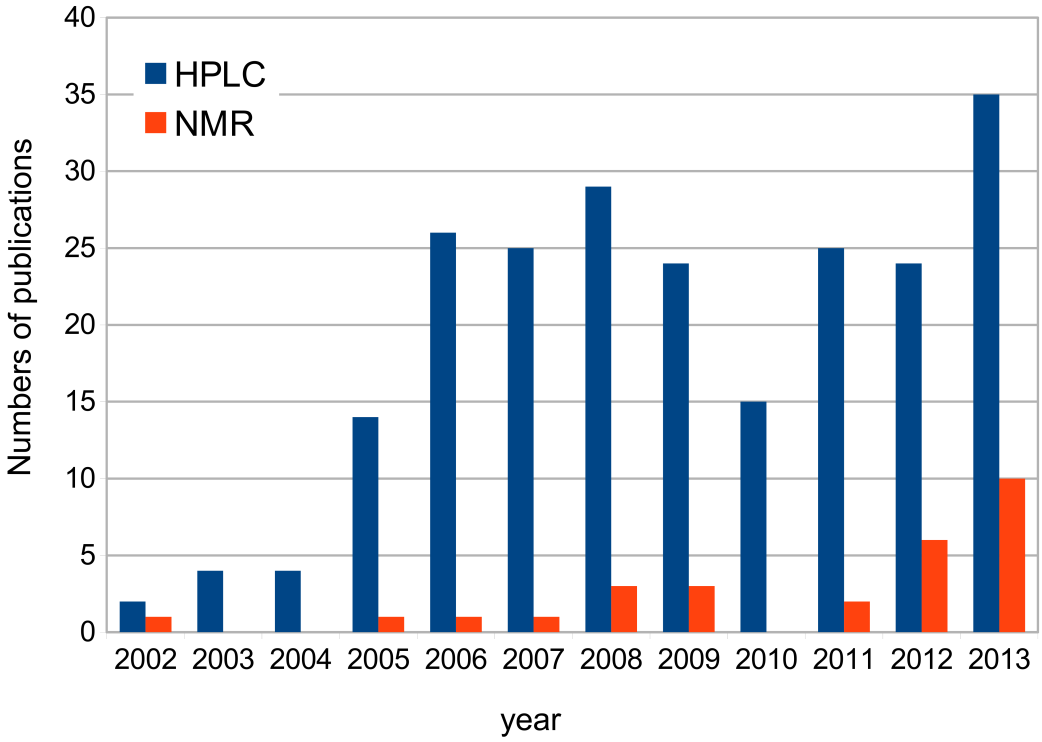


Figure 1.1: *Yearly publications on quality assessment of herbal medicines using HPLC and NMR spectroscopic techniques.*

to quality control of TCM. There are also other papers are related to quality control of TCM but not found using this term, because the key word they used may be"discrimination", "separation", or only used the Latin or Chinese name of a specific herb, without mentioning "herbal medicine" or "Chinese medicine". Therefore figure 1.1 is only a brief indication of the trend.

HPLC was used the most in quality control of TCM, because it is easy to use, and the separated known chemical compounds can be identified and quantified easily. Fewer papers used NMR, because chemical compounds are not separated in the spectrum, resulting the identification and quantification more difficult than HPLC. However, since 2006, the publications using HPLC did not increase much, but those using NMR showed an increasing trend. The number of publications using NMR on quality control of TCM increased 2-fold in 2012 compared to 2011. The reason may be that researchers realised that NMR, as a non-reductive method, can detect all the compounds with protons, therefore coincide with the "holistic" concept of the TCM.

1.2.1 Chromatographic methods used in quality control of TCM

The most popular chromatographic methods used in quality control of TCM include thin-layer chromatography, high performance liquid chromatography and gas chromatography. Chromatographic methods have different mechanisms (adsorption, partition, ion exchange *etc.*) to separate compounds from a mixture using different detection systems. Most common detectors include ultraviolet (UV), photodiode array detection (DAD), and evaporative light scattering detectors (ELSD) detectors. DAD detector can be used to measure several UV wavelengths at the same time, but can not detect non-UV-absorbing components. ELSD is good at detecting non-chromophoric compounds[14], but has a slightly low sensitivity to low molecular weight components due to their small size. However, none of the detectors could detect all the compounds. Therefore, chromatographic method is also referred to as a reductive method[15]. Despite of this, chromatographic methods are still the most widely used since it is easy to quantify and identify separated

compounds.

Thin-layer chromatography (TLC)

TLC is widely applied because it is easy, fast and cheap. Compounds without UV absorption can be detected by TLC by spraying with a visualization reagent. In a single TCL plate, more than 30 compounds can be detected. It is easy to identify the existence of the compounds in botanical extracts simply by comparing the retentions times of the spots with all the chromatographic conditions fixed. However, this method is not as powerful because sometimes some spots may overlap therefore could not be distinguished sufficiently[14].

TLC is a simple and rapid way to perform qualitative analysis and could provide a general idea of the compound contents although the basic TCL has lower accuracy and repeatability than other chromatographic methods such as HPLC and GC. Enhanced forms of basic TLC such as high performance thin layer chromatography (HPTLC) is used to gain higher resolution and accuracy. [18] TLC is a convenient method to perform a quick check to distinguish certain adulterants or fake TCM.

HPLC

HPLC is widely used in standardization of TCM for its powerful separation and quantitative capabilities. HPLC is easily attached to different detectors, such as UV, DAD and ELSD, and spectroscopy such as MS and NMR[19]. Mass spectrometry helps to analyse the structure of the compound as well as gaining information about the molecular weights.

In recent years, Jacob *et al.* have applied ultra-performance liquid chromatography (UPLC) combined with mass spectrometry (MS) on research of the chromatographic fingerprints of herbal medicines[20]. UPLC has a smaller particle size in the stationary phase than a normal HPLC, therefore band-broadening was decreased but resolution and sensitivity were increased.

Zhou *et al.* have developed an improved quality control method for finished Chinese medicine products using HPLC-DAD hyphenated technology [21]. Six hydrophilic and lipophilic components were compared between fin-

ished medicinal products produced by 8 different companies. The result indicated that the content of the six components vary considerably between different brands and tablets.

Gas chromatography(GC)

Many active compounds existing in the Chinese medicines are volatile. GC has high sensitivity on separating most of the volatile compounds. However, it has some disadvantages such as the inability of separating non-volatile or polar compounds.

Coupled with MS, both qualitative and quantitative information can be obtained. Two dimensional GC/MS has been applied on the research on *Pogostemon cablin* Benth.[22]. Volatile oils were successfully separated into more than 800 chromatographic peaks and 394 of these were identified.

1.2.2 Spectroscopic methods used in quality control of TCM

Nuclear magnetic resonance (NMR) spectroscopy

^{1}H-NMR is the most commonly used NMR spectroscopy in TCM, because hydrogen is widely present in most of organic compounds. ^{1}H-NMR needs less sample and less time to run than ^{13}C-NMR.

Contrary to HPLC, in ^{1}H-NMR spectra, one compound can result in multiple peaks, depending on the different magnetic properties of its protons. TCM extractions contain many compounds, which resulting overlapping peaks in NMR spectrum, making the quantitative analysis difficult. However, quantification of compounds is possible by using an internal standard in NMR spectroscopic analysis of multi-ingredient TCM extracts[23].

Due to its advantage of whole structure analysis, NMR spectroscopy can give details of the molecular composition. However, the extract of TCM is usually a mixture of compounds and it is difficult to obtain pure compound. This mixture can show thousands of peaks in the NMR spectrum. Therefore statistical methods become necessary to extract useful information from this

mixture of data. Combining NMR spectra with statistical methods such as PCA or PLS is becoming a trend [11][16]. Several batches of authentic TCM were analysed by PCA to construct a model, in order to predict the authenticity of an unknown samples.

Infrared spectroscopy

Infrared spectroscopy exploits the specific frequencies that a molecule absorbs according to the characteristic of its structure. A beam of infrared light is used to pass through the sample and the spectrum is recorded. Absorption occurs when the frequency of light matches to the vibrational frequency of covalently bonded atoms and groups. All organic compounds absorb infrared radiation, and their absorption spectra are characteristic to the wide variety of vibration motions of molecules. Reciprocal centimetre (cm^{-1}) is used as the frequency scale to show the number of wave cycles in one centimetre.

As a fast, cheap, non-destructive method, Fourier transform near infrared spectroscopy (FT-NIRS) coupled with PCA has been used to successfully distinguish Danshen roots grown in four different regions [24].

Two dimensional near infrared correlation spectroscopy (2D-NIR) was used to discriminate Chinese herbal medicine of different geographic regions. This method was found to be reliable and reproducible[25].

1.2.3 Statistical methods used in the quality control of TCM

Statistical analysis is crucial in studies of TCM since large amounts of data from the complex composition of TCM need to be analysed. Common statistical methods in the quality control of TCM include hierarchical cluster analysis, principal component analysis, and partial least squares analysis.

Hierarchical cluster analysis (HCA) is a method to find out the natural groupings of samples. In HCA, each sample is regarded as a group initially, then the two most similar samples are linked, then the next two, until all the samples are clustered in one group.

HCA was use to find out some natural groupings of TCM species. *Aconitum kusnezoffii* Rchb. was differentiated from other species using HPLC fingerprints with HCA[26]. In this case, the differences between species is large, some difference can even be visualised on the chromatogram. However, in the current study HCA was found difficult to distinguish very similar TCM samples, such as different species of ginseng, and the difference is not easy correlated to the plant metabolites. Therefore HCA is not used in this thesis.

PCA is an unsupervised method that reduces the dimensions of datasets by reducing many variables to a few components (factors) according to the relationships between those variables, summarizes the main differences between datasets, and explains the differences. Unsupervised means no class or difference of the data has been predetermined. PCA can reveal inter-sample and inter-variable relationships visually. Principal component analysis is becoming increasingly popular for the quality control of plants and TCM. It has been used, for example, on grape[27], beer[28] and TCM including Danshen injections[29] and ginseng[30][16]. PCA was coupled with HPLC, NMR and ATR-IR to study five species of TCM, and some industrial compound products in Chapter 2 and Chapter 3 of this thesis. Details of principal component analysis are introduced in chapter 2.

Partial least squares analysis is a supervised method which means that the groups are defined beforehand. The algorithm of PLS-DA is based on PCA. Partial least squares analysis is often used to link the complex chemical composition to a specific activity value, therefore predict the activity of a new sample. For example, PLS has beem used to correlate the NMR and HPLC fingerprints of *Scutellaria baicalensis* under different extraction methods to the results of MTT assay[31], which is a colorimetric assay for assessing cell viability.

1.2.4 Metabolomic

Metabonomics was derived from the application of NMR spectroscopy on biofluids from the mid - 1980s. Metabonomics was applied together with the use of pattern recognition and multivariate statistical analysis on com-

plex datasets. The term of metabolomics grew out of metabonomics around 2001 and was defined as a comprehensive and quantitative analysis of all metabolites. The two words metabonomics and metabolomics are often used interchangeably[32]. Currently metabolomic methods are applied in many fields such as human health care, plant science, animal health and model organisms for research[32]. Nicholson's group from Imperial College London has been trying to apply metabolomic method on developing personalized health care. Urinary metabolites from 4,630 individuals from the U.K., the U.S. and Asia were analysed. Blood pressure data are correlated with the urine NMR spectroscopic data to determine if any consistent metabolic differences exist between individuals with and without hypertension[33].

HPLC and NMR coupled with PCA were used on differentiating praziquantel tablet from different manufacturers using metabolomic method. Preservative was found in the products from one manufacturer and counterfeit products are successfully distinguished from the authentic ones[34].

Metabolomics analysis has been carried out on herbal products *Artemisia afra* Jacq. using LC-MS and NMR spectroscopic methods[35]. This work found that those claimed anti-malaria products under study contain *Artemisia afra* Jacq. instead of *Artemisia annua* L. However, the anti-malaria compound artemisinin only exist in *Artemisia annua* L. which means that those claimed anti-malaria products could be far too low to be effective.

These works indicated that metabolomics analysis using NMR spectroscopic method is a rapid and valuable tool in study of metabolite and quality control of medicines.

1.2.5 Metabolomics quality control of TCM

Conventional pharmaceutically acceptable methods for standardization and quality control have until recently relied on inappropriate reductive analyses - for example, of (often inactive) marker substances. Increasingly, metabonomic applications are finding a place in non-reductive analyses[29][30][16]. Multivariate data analysis (MDA) including principal component analysis, partial least squares discriminant analysis (PLS-DA), and partial least squares

are the tools used in metabolomics studies.

A number of analytical techniques are available for metabolite profiling, including the reductive (involving methods separating constituents such as gas/liquid chromatography and HPLC) and the non-reductive (eg, mass spectrometry and high field NMR). A number of publications [30][16][36][37][38] showed the feasibility of the use of high field NMR spectroscopic profiling coupled with multivariate analysis a viable approach for the standardization and quality control of complex herbal mixtures. This approach is particularly well suited to the problem because, contrary to the chromatographic methods, the use of high field NMR spectroscopy does not rely on separation of individual compounds. Such a method is useful because of the increasingly validated concept of synergy[39] implying that the control of one compound (active or not) is not sufficient when the clinically observed activity is likely to be due to the biological effects of more than one compound.

This thesis involves the application of these approaches to the difficult issue of the standardization and quality control of widely used mixtures of several different plant species in proprietary and prescription TCM preparations. This would be an important advance, allowing the possibility of their precise definition prior to clinical evaluation and eventual registration as a drug. This outcome, if successful, would be a major advance in the much needed control of the quality of TCM.

1.3 Activity-based standardization

Activity-based standardization on TCM means linking the activity of the TCM on a biological system (such as cells, animal, human) to its quality. If the standard of known activities of good quality TCM is established, the tested TCM which can not achieve the same activities can be regarded a "worse" quality sample. Ideally, the selected method to test the activity of TCM should be "holistic". Because a TCM contains many chemical compounds, the activities caused by those compounds are complex. Some potential "holistic" methods may be feasible for the activity-based quality control of TCM, including metabonomics, genomics and proteomics. The meanings

of these terms are explained and their usage discussed in the introduction of Chapter 4 of this thesis.

Not much research has been done on the activity based standardization of TCM. Although some links have been made between plant to its effect on a biological system, the purposes of these research are often not quality control of TCM, but focus on the activities of the plant or the difference of the biological samples used. For example, human urine was tested after ingestion of chamomile tea using ^{1}H-NMR metabonomics, and the changes of metabolites were discovered[40].

In chapter 4 of this thesis, proteomics was used to study the holistic effects of American ginseng extract on a pancreatic β cell line. Proteins perform many basic activities in human. As a holistic method, proteomics quantitatively monitor the changes of thousands of proteins at the same time. This holistic approach fits the "wholism" concept of TCM, and can potentially be the future trend to standardize herbal medicines.

1.4 Structure of this thesis

This thesis consists of three experimental chapters. Each chapter is self contained, including its own introduction, materials and methods, results and discussion, and conclusion.

The study of the second chapter was done in the beginning of this PhD project, as an initial study of metabolomic quality control of TCM, using various analytical methods including HPLC-DAD, ^{1}H-NMR and ATR-IR. A brief comparison of those methods was carried out, and the feasibility of those methods on different material types was discussed.

In the third chapter, the metabolomic quality control methods were explored in a deeper level based on a large sample size (37 batches of ginseng). ^{1}H-NMR with PCA was used to assess the quality of 4 types of ginseng, including cultivated American ginseng (15 batches), wild American ginseng (11 batches), Asian ginseng (5 batches), Sanqi ginseng (5 batches). The difference between them were compared and discussed. Chemical shifts in ^{1}H-NMR spectra which are responsible for the main differences were assigned

to the primary and secondary metabolites. A two-step exclusion analysis was used to separate different types of ginseng. Firstly the whole spectra were analysed, and the main differences between ginseng types were due to sugars, which are highly abundant in ginseng extract. Secondly the sugar regions were excluded, then the differences between ginseng types were mainly due to some less abundant primary and secondary metabolites.

In the fourth chapter, a proteomic approach was used to study the changes of the protein composition of MIN6 pancreatic β cells, before and after treatment with wild or cultivated American ginseng, during a period of 24, 48 and 66 hours. The proteins that showed a significant up or down regulation were presented, and the trend of changes over time was plotted. Comparison has also been made between the protein changes of MIN6 cells treated by wild and cultivated American ginseng. This work allows the effect of ginseng extracts on MIN6 cells whole proteome to be characterised, and significant differences before and after treatment may be used to assess the quality of ginseng. For example, some products might not cause the same changes in proteins levels as good quality ginseng.

1.5 Aim of study

The aim of the project is to study the chemical- and activity-based standardization methods of Traditional Chinese Medicine. The feasibility of chemical-based methods such as HPLC-DAD, ^{1}H-NMR and ATR-IR coupled with principal component analysis were used to distinguish the chemical differences of TCM between growing regions, harvesting dates, species, wild and cultivated, and between batches of the same brands using Danshen, Sanqi and ginseng products. Also, raw material (original herb) is used, as well as intermediate preparations, and final product (TCM) samples of a well characterized TCM (compound Danshen dripping pill, CDDP) to demonstrate the effectiveness of the process analytical technology (PAT)'s in a whole TCM production process. The data from well characterized TCM samples (CDDP) from a reputable company (Tasly) should provide meaningful results to show if the technology is truly a suitable PAT for TCM production

process quality control/quality assurance (QA/QC). The plan is to test the feasibility of using a metabolomics approach as a PAT for TCM QA/QC to provide production/product data from TCM preparations with most of its content intact, which may be more convincing to the regulation for satisfying the desire of knowing everything about medication before granting a licence for use as a drug. These tests investigate the suitability of the metabolomic method for following, constituent changes during a TCM production process and for monitoring the constituents consistency in different batches of the TCM production.

This study also aims to assess the feasibility of proteomics for activity-based standardization of TCM. Difference in gel electrophoresis(DIGE) was used to investigate its suitability to distinguish the differences of protein regulation changes in a pancreatic β cell line (MIN6) before and after American ginseng treatment. Also, the protein changes in MIN6 cells caused by cultivated and wild American ginseng treatments were compared, therefore finding the similarity and differences of the the activities (in this case, protein changes) of similar TCM.

Chapter 2

Evaluation of quality control methods including multivariate analysis of HPLC, IR and NMR spectroscopic data using various Danshen products

2.1 Introduction

There are difficulties in the standardization of traditional Chinese medicine (TCM) because of the complexity of TCM preparations. Traditionally, TCM preparations are combinations of several ingredients designed individually for patients. In recent years, however, popular TCM combinations are sold as pre-prepared products. Compound Danshen product is one of the most well known pre-prepared TCM products. Compound Danshen products usually refer to a combination of Danshen(*Salvia miltiorrhiza* Bunge) , Sanqi(*Panax notoginseng* Burk.) and borneol. Danshen is used as the main effective ingredient which is called the emperor, and Sanqi is used as the minister. Compound Danshen products have been shown to improve cardiovascular function and consist mainly of extract of Danshen and Sanqi [41]. Compound

Danshen products are produced in various formulations including tablets, pills, granules, injections, oral liquids, *etc.* Among these, compound Danshen dripping pill (CDDP) is the most popular formulation, with worldwide annual sales exceeding $148 million in 2009[42]. CDDP will soon begin Phase III clinical investigation[42].

The U.S. Food and Drug Administration (FDA) new drug registration process consists of three steps: phase I, II, and III. Phase I is a small scale investigation that principally includes safety, and dose finding. Phase II is a clinical trial that studies safety and efficacy on hundreds volunteers and usually takes up to 5 years. In Phase III investigation, the drug is studied on a large group of people of thousands, in order to confirm efficacy, monitor adverse effects, collect safety information and compare to other commonly used drugs[43]. The Phase III investigation may take another 5 years. Despite the efforts and money spent on Phase I and II, 10% to 30% drugs fail during Phase III [43].

If the Phase III clinical trial is successfully completed, a "New drug application" can be submitted to the FDA. At this stage, various factors will be evaluated, especially if the company has the capability of manufacturing the drug with a consistent quality. Finally, the drug can be marketed and prescribed.

Control of drug quality is thus crucial in the process of passing FDA-regulated clinical trials. For conventional chemical drugs which only consist of one or a few chemicals, quality control is straightforward. For botanical drugs, with more than hundreds of chemicals present, precise control of each chemical is impossible. In 2004, the FDA published guidelines that includes botanical drugs into "food supplement" category.

In June 2004, the FDA released "Guidance for Industry: Botanical Drug Products" that also emphasized the importance of an effective quality control method for herbal medicines[44].

However, quality control of CDDP is still largely based on evaluating the quantity of a few active compounds[45]. The existing quality control method is inadequate since the synergy effects of the multi-ingredient CDDP is ignored[46].

In the present study, Danshen, Sanqi, intermediate product and finished products including commercial CDDP and compound Danshen tablets have been studied using HPLC, NMR and IR spectroscopic methods with PCA and PLS-DA data analysis methods in an attempt to improve the process.

2.1.1 Danshen and Sanqi plant

Danshen, *Salvia miltiorrhiza* Bunge, also known as red sage, Chinese sage or *Tanshen*, is from the *Lamiaceae* family. Danshen was recorded in *Shennong's Chinese Materia Medica*, which is considered the earliest Chinese Pharmacopoeia, and a significant reference work for Chinese medicine[47]. Danshen root (Figure 2.1(a)) is the most common part of the plant to be used for medication. Danshen inhibits platelet aggregation[48][49], dilates coronary arteries[50] and relieves angina pectoris[51].

Sanqi(Figure 2.1(b)), *Panax notoginseng* Burk., also named as Tianqi, or Jinbuhuan, is grown widely in Yunnan province in China. The root of Sanqi is the most commonly used part in TCM. Sanqi has been used to treat coronary atherosclerotic heart disease[52], and improve blood circulation[53]. Interestingly, Sanqi has been shown to have both antiplatelet, anticoagulant[54][55], and anti-bleeding effects[56]. The mechanisms of this dual-direction effects of Sanqi is unclear. Sanqi has also been shown to have antihyperglycemic effects[57].

2.1.2 Compound Danshen dripping pill

A group of compound Danshen products are shown in figure 2.2. Among them, CDDP and industry imitated intermediate material were studied in this thesis. Compound Danshen tablets were studied in the author's MSc. thesis, therefore the data was not presented here.

Compound Danshen dripping pill, also referred to as Cardiotonic Pill, is composed of aqueous extracts of Danshen, Sanqi and synthesized borneol.

Borneol is a terpene and bicyclic organic compound. Borneol is an adjuvant and guiding component in Chinese medicine, which does not count as the main effective medicine in a compound prescription. Synthetic borneol



(a) Danshen roots



(b) Sanqi roots

Figure 2.1: *Commercially available Danshen and Sanqi roots. (a): Danshen roots, adapted from* `http://www.tcm100.com/Images£_£ZhongYaoXue/Danshen-YaoCai.jpg`. *(b): Sanqi roots, adapted from* `http://files.mainone.com/company/companyNews/2010£_£01/18131555543.gif`.



Figure 2.2: *Commercially available compound Danshen products including compound Danshen tablets (CDT) and compound Danshen dripping pill. CDT are produced by various companies including: GKH Natural Herbs (CDT-G), Hefei Shenlu Double-Crane pharmaceutical company (CDT-G), and Herbal Inn (CDT-H). IM: intermediate product of CDDP. Photo was taken by Huiying Zhao on 31 Oct. 2011.*

is much cheaper to produce than the natural ones. Sythetical borneol was shown to have the same pharmacodynamics effect as the natural borneol[58]. It was not known if synthesized or natural borneol was used in CDDP in the current market.

However, research shows that borneol has the effect of anti-inflammation, and protecting against cerebral ischemia[59]. Borneol has been shown to enhance intestinal absorption of salvianolic acids, including a few biological effective compounds in Danshen such as salvianolic acid B, rosmarinic acid and salvianolic acid A[60]. Borneol is usually analysed by gas chromatogra-

phy and is not studied in this chapter.

More than 707 companies produce compound Danshen products. Among them, 687 companies produce compound Danshen tablets but only one company produces compound Danshen dripping pill. This is a leading TCM pharmaceutical manufacturer in China, called Tianjin Tasly pharmaceutical company (Tasly). CDDP has been produced for 15 years since 1994 and has become the leading product of Tasly. The sales of CDDP exceeded 1 billion Chinese Yuan in 2007 which makes it one of the leading Chinese Medicine products.[42] In January 2010, CDDP became the first compound Chinese medicine product (produced from more than one plant) that has completed phase II investigation (FDA, NCT00797953). Till 2010, CDDP has already been marketed as a drug in Russia, South Korea, Mongolia, Singapore, Vietnam, and South Africa[61][62][63].

CDDP has been found to be more effective than isosorbide dinitrate, which is routinely used medicine, in treating angina pectoris[51]. 60 eligible randomized controlled trials (with 6931 participants involved) published between 1994 and 2009 were used, and CDDP was shown more effective.

2.1.3 The general manufacturing process of dripping pills

The production method of dripping pills was first established in Denmark and was developed rapidly in China for production of TCM. Water and organic solvents are commonly used as extraction methods in herbal medicine manufacture. The herbal extracts are mixed with polyethylene glycol (PEG) 4000 or 6000 or sometimes a mixture of these two as a matrix[64] . The volumetric ratio of the liquid extract to the matrix is usually around 1:1.5 to 1:2.5[65] . The mixture of plant extracts and matrix is heated to 70°C to 80°C and then dripped from a height of about 1 metre into a coolant, often liquid paraffin, thereby forming the dripping pills in the form of spherical solid masses of about 3 mm diameter. The size and the temperature of the dripping nozzle, the rate of dripping, and the distance from the dripping nozzle to the coolant are all controlled in order to achieve the optimal size

and shape of the dripping pills. Common coolants include liquid paraffin, dimethyl silicone oil and vegetable oil. Dimethyl silicone oil is commonly used among water soluble dripping pills since it has a smaller surface tension therefore shapes the pills well[64]. A mixture of coolants is often used[65].

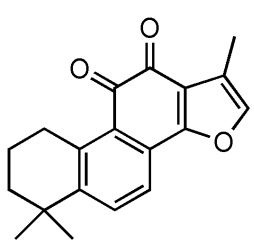
The manufacturing process for compound Danshen dripping pill

Danshen and Sanqi are refluxed with water. The water extract is decantated and concentrated by evaporation. Ethanol is added to the concentrated sample to cause sedimentation. The sediment is removed by decantation and the upper liquid is concentrated to give a sticky mass. Borneol is ground and mixed with the paste in molten PEG. The final mixture is dripped into paraffin to produce compound Danshen dripping pill, the dripping pills are then washed with water and packaged[45].

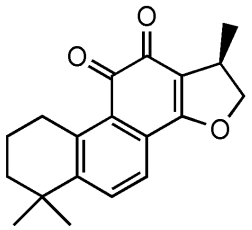
2.1.4 Primary and secondary metabolites

Primary and secondary metabolites are organic chemicals synthesised in plants. Primary metabolites serve the purpose for the plants' fundamental metabolic pathways involved in their growth and development. Primary metabolites include mainly carbohydrates, lipids, proteins and nucleic acids.

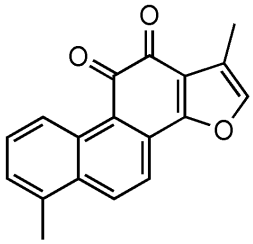
It was not until the late 20th century that secondary metabolites were clearly recognised to have important functions in plants. Secondary metabolites serve the purpose of signalling and defence systems in plants, such as pesticides, scent, colouring and flavouring. Secondary metabolites include alkaloids, phenolics and terpenoids. They usually present as unique combinations of specific chemicals in different plants, and serve specific purposes for that plant. Secondary metabolites have no nutritional value to humans, but may have proved biopharmacological effects, although most of the mechanisms remain unknown[66].



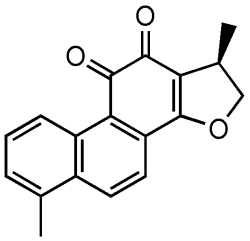
tanshinone IIA



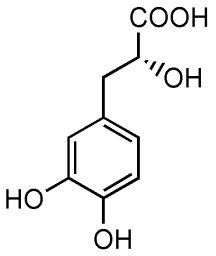
cryptotanshinone



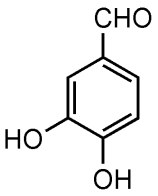
tanshinone I



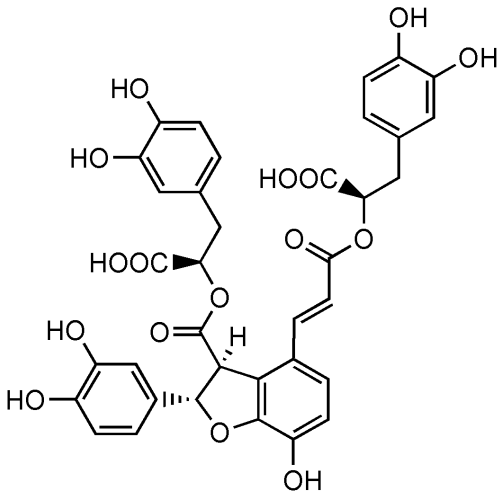
15,16-dihydrotanshinone I



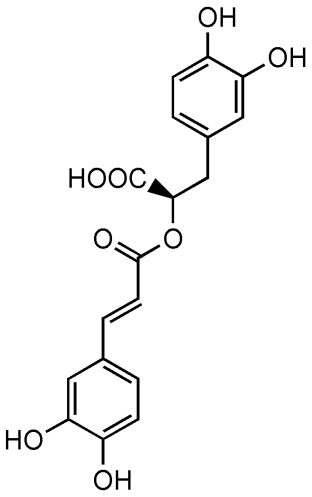
salvianolic acid A



protocatechuic aldehyde

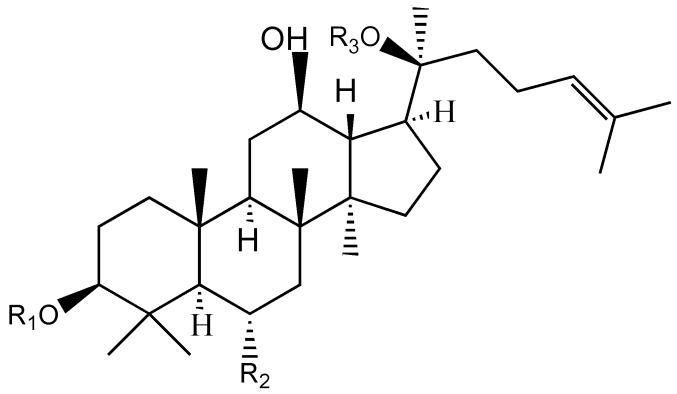


salvianolic acid B



rosmarinic acid

Figure 2.3: *Structure of representative diterpenoid quinones in Danshen including: tanshinone IIA, cryptotanshinone, tanshinone I and 15,16-dihydrotanshinone, representative phenolic compounds including: salvianolic acid A, protocatechuic aldehyde, salvianolic acid and rosmarinic acid.*



	R_1	R_2	R_3
notoginsenoside R_1	-H	-O-glc(2→1)xyl	-glc
ginsenoside Rg_1	-H	-O-glc	-glc
ginsenoside Rb_1	-glc(2→1)glc	-H	-glc(6→1)glc
ginsenoside Rd	-glc(2→1)glc	-H	-glc

Figure 2.4: *Representative saponins from Sanqi*

2.1.5 Major compounds and their functions in compound Danshen dripping pill

There are 4 major saponins, 4 major phenolic acids and 4 major diterpenoid quinones in CDDP. The 4 major phenolic acids from Danshen include salvianolic acid A (*Danshensu*, 3-(3,4-dihydroxyphenyl)-2-hydroxylpropionic acid), rosmarinic acid, protocatechuic aldehyde and salvianolic acid B (Figure 2.3). The 4 major diterpenoid quinones that come from Danshen include 15,16-dihydrotanshinone, cryptotanshinone, tanshinone I and tanshinone IIA (Figure 2.3). All of them are bioactive compounds[67][21]. The 4 major saponins come from Sanqi include notoginsenoside R1, ginsenoside Rg_1, ginsenoside Rb_1 and ginsenoside Rd(Figure 2.4).

Salvianolic acid A has vasodilatation ability[68]. It was shown that salvianolic acid A inhibits platelet aggregation, improves micro circulation, protects the myocardium from reperfusion injury of the ischemic heart, and has antitumer activity[69]. The mechanisms maybe related to inhibition of

Ca^{2+}aggregation in cardiac muscle cells[70], inhibiting angiogenesis and tumor cell invasion[69], *etc.*

Salvianolic acid B has been shown to have free radical-scavening activity [70]. It has been shown that salvianolic acid B has a higher antioxidant activity than salvianolic acid A and vitamin C. The mechanisms of salvianolic B and salvianolic acid A's antioxidant activities were not fully understood but are believed to be related to the phenol group[70].

The cardioprotective effects[71] of tanshinone IIA include platelet aggregation, anti-oxidant and anti-inflammatory actions[72]. It has been shown to significantly reduce mycardial infarct size[72].

Notoginsenoside R_1 is the main compound in Sanqi that has cardiovascular effect [73] .Notoginsenoside R1 may protect the progression of atherosclerosis[74].

The biomarkers mentioned above were the most studied ones in Danshen and Sanqi products but the precise mechanisms of the biomarkers are not fully understood. It is even more complex when many compounds work together to achieve a reinforced effect than when single compounds are used separately[75].

2.1.6 Current quality standards and the limitations

Quality control of traditional Chinese medicine is important because the same plant species produced in different places can have different chemical constituents, therefore leading to different effects[10]. Adulterants are usually cheaper and less effective than the authentic plant, or not effective at all. For example, *Arctium lappa* L., *Ipomoea batatas* Lam. (sweet potato) and *Salvia przewalskii* Maxim. are common adulterants of Danshen[76][10]. *Ipomoea batatas* Lam. can be coloured to look very similar to Danshen.

In the past, organoleptic tests such as appearance, taste, smell and feel were the main quality control methods for TCM. These tests still play an important part of quality evaluation in China[45]. However, organoleptic tests are not sufficient since it is impossible for them to be standardised.

In the Chinese Pharmacopoeia 2010 (CP 2010), the following factors are

mentioned for TCM products: description (appearance), microscopic and chemical identification, quantitative chemical assay, heavy metals and harmful elements, microbial limits, and animal safety test. HPLC is used for quantitative analysis for Danshen and Sanqi biomarkers.

The standards for Danshen, Sanqi and CDDP in Chinese Pharmacopeia are used by the Tianjin Tasly Parmaceutical Company as its quality control method. In Chinese Pharmacopoeia 2005, three compounds had been selected to be quantitative biomarkers for Danshen. Tanshinone IIA should not be less than 0.20%, salvianolic acid B should not be less than 3.0 % and salvianolic acid A should not be less than 1.2%; for Sanqi, ginsenoside Rg_1, ginsenoside Re, ginsenoside Rb_1 and notogingsenoside R_1 should not be less than 6.0% in total[13].

In the CP 2010, salvianolic acid A in Danshen and ginsenoside Re in Sanqi were not listed in their quantitative analysis(CP2010, volume 1. Sanqi, page 11; Danshen, page 70). Only two compounds were selected to be quantitative biomarkers for Danshen. Tanshinone IIA should not be less than 0.20% and salvianolic acid B should not be less than 3.0%. For Sanqi, ginsenoside Rg_1, ginsenoside Rb_1 and notogingsenoside R_1 should not be less than 5.0% in total[45].

A limitation of Chinese Pharmacopoeia is that only the lowest value of chemical markers was defined, not the highest. However, an overdose of effective chemicals might cause side effects[77]. The lowest threshold of chemical markers defined by the CP 2010 might only work using exactly the same experimental parameters. For example, according to the CP 2010, salvianolic acid B should not be less than 3.0% and tanshinone IIA should not be less than 0.20% in Danshen. In recent studies by Zhong *et al.*, 74 Danshen samples were collected from different cultivation locations and provincial hospitals in China, and were simultaneously quantified using a UPLC method[10]. Comparing to the CP, only 11 out of 74 samples contained level of tanshinone IIA higher than 0.20%, and 17 out of 74 samples contained level of salvianolic acid B higher than 3.0%. It appears that only 3 out of 74 samples meet both requirement by CP 2010. This might have been caused by the difference of extraction methods and experimental conditions. In Chinese Pharmacopoeia

2010, 75% methanol was used to extract salvianolic acid B. Zhong *et al.* used methanol instead. Salvianolic acid B dissolves better in 75% methanol than in methanol. This might be the reason that a lower content of salvianolic acid B is obtained in the majority of samples by Zhong *et al.* than in CP 2010. For tanshinone IIA, methanol was used in both Zhong *et al.*'s research and the CP 2010, but the extraction methods are different. Dried Danshen powder was sonicated for 45 min in Zhong *et al.*'s study, but was refluxed for 1 hour in the CP 2010. This might means that the mentioned reflux method is more efficient in extracting tanshinone IIA than sonication method.

Research on simultaneous quantification analysis of Danshen marker compounds by Li *et al.* [78] showed similar results to Zhong *et al.*[10]. The concentration of tanshinone IIA is 1.58% on average (with a standard deviation of 0.15, samples with highest and lowest value were not included since they were far off the average) of Danshen from 7 geographical locations. Ethanol was used as the extraction solvent. Therefore, the concentration of chemical markers were not comparable to those obtained using methanol. There are also many other studies on quantification of chemical markers of Danshen using various extraction solvents and processes[79], in order to determine the quantities of biomarkers simultaneously. None of the results obtained can be compared to the requirements of the CP 2010, or to the research of other groups, unless the same extraction method was used.

For the quantitative analysis of tanshinone IIA, UPLC (detected at 280 nm) was used by Zhong and Li[10], and HPLC (detected at 270 nm) was used in CP 2010. However, this should not be a problem causing the concentration of tanshinone IIA different, since pure reference compound was used for calibration in both cases.

Adulterants are not always without bioavailability. The Danshen adulterant *Salvia przewalskii* Maxim. has a unique diterpene quinone called przewaquinone, which has been shown to be effective against vascular disorders[80] . *Salvia przewalskii* Maxim has a higher content of rosmarinic acid and a lower content of salvianolic acid B than Danshen. This makes it easy to be identified from Danshen, using HPLC. Quantitative control of reference chemicals helps to distinguish some adulterants and poor quality

herbs, which have a different content of marker chemicals from the desired herb. Many studies have suggested quantifying chemical markers as a quality control method for Danshen. For compound Danshen preparations, Zhou *et al.*[21] have selected 6 markers and Wei *et al.* [67] have selected 12 compounds as quality control markers. The content of each compound was measured by HPLC and the overall stability variations and the overall recoveries were determined. By comparing an unknown batch of product with the known ones, the quality of the unknown product could be evaluated. If the product has a higher content of the selected biomarkers than the required amount, it is usually considered a qualified product. However, this herb might have very distinct concentrations of other chemicals other than the selected biomarkers, therefore it might also have a very different biological effect. For example, CP 2010 only requires content of salvianolic acid B (SAB) 3.0% and tashinone IIA 0.2% in Danshen. However, other compounds such as rosmarinic acid also have reported effects on human[81] .

Therefore, a chemical-based non-selective method has the advantage in quality control of TCM. As many chemicals as possible can be considered in a non-selective method, which combines the whole chromatography or spectroscopic fingerprints with statistical evaluation methods. All the chemicals contributed to the signals in the fingerprints were considered in this quality control method. Non-selective quality control method fits the "synergy" concept of TCM, and is considered the future of modernization of TCM[82].

Various chromatographic techniques were used to access the quality of TCM[83]. Ultra-performance liquid chromatography (UPLC) was used with rPCA (robust principal component analysis) to distinguish commercially available extract granules of Danshen[84]. rPCA is very similar to PCA, but use the median instead of the mean of data for the calculation, therefore minimising the impact of outliers[85].

Fourier transform near infrared spectroscopy was used with PCA and PLS to distinguish geographical regions of Danshen and its antioxidant activity was correlated to different regions[24]. Direct analysis in real time mass spectrometry (DART-MS) was used with multivariate data analysis to distinguish Danshen injections and analytical markers were found[29].

In this chapter, HPLC-DAD (photodiode array detection), ^{1}H-NMR and ATR(attenuated total reflectance)-IR coupled with PCA were used to study the quality of Danshen, sanqi, intermediate product and finished commercial product (CDDP). Two ATR-IR methods including extraction evaporation and solid sample method were evaluated using CDDP. Danshen grown from different regions and Sanqi supplied in different dates were analysed. CDDP produced from different years were compared.

2.2 Materials and methods

2.2.1 Materials

5 batches of Danshen (numbered 1-5) produced in different regions, Sanqi plant root material supplied in different dates (numbered 1-5) were supplied by Tianjin Tasly company.

5 batches of Danshen (numbered A-E, provided by Tasly to Zhejiang University, China), Sanqi (numbered A-E, produced by Hangzhou Chinese medicine factory) and correlated intermediate products alphabetised from A to E were provided by Professor Yongping Yu in Zhejiang University. The intermediate product was prepared by Yu's group using the same process as Tasly (Details described in section 2.2.2). The origins of plant materials are listed in table 2.1. A map of China is shown in figure 2.5 for better understanding of the origins of Danshen and Sanqi.

9 batches of CDDP produced by Tasly in 2007 and 2009 were listed in table 2.2.

Voucher specimens of tested samples were deposited in the pharmacognosy library in Institute of Pharmaceutical Science, King's College London. Methanol, acetonitrile, water were all of HPLC grade and purchased from Fisher Scientific (Fisher Scientific Ltd., Pittsburgh, PA), salvianolic acid B, tanshinone IIA, salvianolic acid A, ginsenoside Rg_1, Rb_1 were purchased from Carbosynth (Carbosynth Ltd, Berkshire, UK). Notoginsenoside R_1 was purchased from The National Institutes for Food and Drug Control, Beijing. Protocatechuic aldehyde (D108405, 97%), deuterium oxide (D_2O, 661643,

Table 2.1: *Details of herbal materials studied in this chapter. Each batch has a unique warehouse number (not provided for reasons of confidentiality) and batch number. Danshen 1–5 were provided by Tasly and used for region-to-region study. Sanqi 1–5 were provided by Tasly and used for date-to-date study. Danshen A–E, Sanqi A–E and intermediate products A–E were used for products in industrial process study. Abbreviations: NOM: nature of material. P: ground root powder, R: dried and trimmed root, S: sticky syrup. Studies: which studies have been performed in this chapter. R: region-to-region study, D: date-to-date study. I: products in industrial process study. IM: intermediate product. pt.: planting base. HZ: Hangzhou pharmaceutical factory. ZU: Zhejiang University.*

Name	NOM	Grown in	Supply date	Supplied by	Studies
Danshen 1	P	Shaanxi pt.	20/01/2009	Tasly	R
Danshen 2	P	Shaanxi pt.	09/03/2009	Tasly	R
Danshen 3	P	Shandong pt.	04/03/2009	Tasly	R
Danshen 4	P	Shandong pt.	04/03/2009	Tasly	R
Danshen 5	P	Shandong pt.	04/03/2009	Tasly	R
Sanqi 1	P	Yunnan pt.	28/10/2008	Tasly	D
Sanqi 2	P	Yunnan pt.	20/10/2008	Tasly	D
Sanqi 3	P	Yunnan pt.	25/10/2008	Tasly	D
Sanqi 4	P	Yunnan pt.	15/11/2008	Tasly	D
Sanqi 5	P	Yunnan pt.	03/03/2009	Tasly	D
Danshen A	R	Tasly pt.	08/2009	Tasly	I
Danshen B	R	Tasly pt.	08/2009	Tasly	I
Danshen C	R	Tasly pt.	08/2009	Tasly	I
Danshen D	R	Tasly pt.	08/2009	Tasly	I
Danshen E	R	Tasly pt.	08/2009	Tasly	I
Sanqi A	R	Yunnan Prov.	08/2009	HZ	I
Sanqi B	R	Yunnan Prov.	08/2009	HZ	I
Sanqi C	R	Yunnan Prov.	08/2009	HZ	I
Sanqi D	R	Yunnan Prov.	08/2009	HZ	I
Sanqi E	R	Yunnan Prov.	08/2009	HZ	I
IM A	S	N/A	08/2009	ZU	I
IM B	S	N/A	08/2009	ZU	I
IM C	S	N/A	08/2009	ZU	I
IM D	S	N/A	08/2009	ZU	I
IM E	S	N/A	08/2009	ZU	I



Figure 2.5: *A map of Chinese provinces. Underlined provinces are related to the origin of plant used in this chapter. Shandong and Shaanxi provinces produces Danshen. Yunnan province produce Sanqi. Adapted from* `http://www.sacu.org/`.

%99.99 atom % D, contains 1% DSS-d_6) was from Sigma (Sigma Co., St Louis, MO, USA). 5 mm NMR tubes (Wilmad 507-PP) and methanol-d_4 (CD_3OD; 99.8% purity) were purchased from Goss Scientific (Goss Scientific Instruments, Great Baddow, Essex, UK).

2.2.2 Sample preparation

Danshen and Sanqi dried root material were ground using a coffee grinder and filtered through a 350 μm sieve. CDDPs were briefly ground in a mortar, in order to break and remove the coating. The coating was formed when the

Table 2.2: *Details of CDDP studied. CDDP 1–9 are compound Danshen dripping pills produced by Tasly in 2009 and 2007. B.B.: best before.*

Product	Batch no.	Production date	B.B. date	Manufacturer
CDDP 1	090312	20/03/2009	02/2013	Tasly
CDDP 2	090310	28/03/2009	02/2013	Tasly
CDDP 3	090311	04/03/2009	02/2013	Tasly
CDDP 4	090313	04/03/2009	02/2013	Tasly
CDDP 5	090314	04/03/2009	02/2013	Tasly
CDDP 6	20070909	22/08/2007	07/2011	Tasly
CDDP 7	20070403	23/03/2007	02/2011	Tasly
CDDP 8	20070404	23/03/2007	02/2011	Tasly
CDDP 9	20070308	28/02/2007	01/2011	Tasly

liquid plant extract was dripped into the coolant, usually liquid paraffin (see section 2.1.3).

0.5 g of the samples was extracted with 25 ml 70% methanol in a volumetric flask, by sonicating for 15 minutes at 30°C followed by shaking manually and sonicating for a further 15 min. After cooling to room temperature, evaporated solvent was replaced with 70% methanol and the volume of the solvent was made up to 25 ml. The supernatant was filtered with a 0.2 μm syringe filter and stored at 4°C for future use.

Preparation method for intermediate product

150 g Danshen and 50 g Sanqi plant material of each batch were refluxed with 2000 ml water (pH was adjusted to 7-7.5 using NH_3) for 2 hours and then filtered. The residue was taken and refluxed by 2000 ml water (pH was adjusted to 7-7.5 using NH_3) for 2 hours. The liquid extractions were combined and concentrated to 1:3 to 1:5 (extract: water). 95% ethanol was added to the concentrated liquid extraction and the final solvent was made to contain 80% ethanol by volume. The solvent was stirred and left overnight and then filtered and concentrated until no more solvent could be evaporated.

2.2.3 HPLC analysis

HPLC column: Zorbax SB-C18 150mmX2.1mm,5μm, flow rate 0.2ml/min, Agilent, UK HPLC system: Hewlett-Packard, 1050 Liquid chromatography.

HPLC mobile phase solvent was filtered through a 0.2 μm membrane filter. The mobile phase was composed of (A) aqueous trifluoroacetic acid (0.1%, v/v) and (B) acetonitrile, over a gradient elution of 0-90% B at 0-40 minutes, 90% B from 40 to 45 minutes, followed by 90-0% B from 46 to 50 minutes, and 0% B from 51 to 53 minutes.

Chromatograms were produced by detection at 205, 210, 230, 254 and 281 nm. 281 nm was chosen as the detection wavelength for Danshen and CDDP; 205 nm was chosen for detection of Sanqi plants.

Chromatographic data of each run were saved as *txt.* files and subjected to data analysis.

2.2.4 NMR spectroscopy

20 ml of the supernatant obtained in section 2.2.2 was evaporated by rotary evaporation under reduced pressure at 40°C for 20 minutes. About 7 ml solvent was left after evaporation. This solvent was removed by lyophilization.

D_2O(D_2O, 661643, %99.99 atom % D, contains 1% DSS(4,4-dimethyl-4-silapentane-1-sulfonic acid)-d_6), [D_4]-MeOD (99.8%) with 0.05% (v/v) tetramethylsilane (TMS, Cambridge Isotope Laboratories Inc., Hook, Hampshire, UK) and DMSO were attempted to reconstitute the lyophilised samples. Both D_2O and CD_3OD were found feasible to dissolve Danshen and CDDP extract, but D_2O is better at dissolving Sanqi extract.

Lyophilized samples (10 mg) were dissolved in 1 ml deuterated solvent. One dimensional ^{1}H NMR spectra were measured at a temperature of 300K on a Bruker DRX 400 MHz Spectrometer (Bruker, Coventry, UK) equipped with a 5 mm triple resonance inverse detection (TXI) probe. Spectra were the result of the summation of 16 free induction decays (FIDs), with data collected into 65k data points and a sweep width of 20 ppm. Acquisition parameters were 0.126 Hz/point, pulse width was 14 μs 90° and relaxation delay was 1.0 s. Prior to Fourier transformation, an exponential line broad-

ening equivalent to 0.3 Hz was applied to the FIDs. The signal intensities for all samples were referenced to TMS at 0.00 ppm. Spectroscopic data for each run were saved, then transformed to *txt.* data using MestReNova® (MNova) software version 5.2.3(Mestrelab research S.L. Santiago de Compostela, Spain).

2.2.5 IR spectroscopic analysis

IR spectroscopic analysis was performed using FT-IR Spectrometers (Perkin Elmer, Massachusetts) with a Universal ATR Polarization Accessory. Powdered and liquid extraction samples prepared in section 2.2.2 were used in this test. Diamond was used on the sampling plate as the detection surface.

Danshen and Sanqi powder can be measured directly on top of the diamond with pressure applied. This is called the solid sample method.

Liquid extraction could not be tested directly on the unheated diamond plate, because solvents such as water and methanol gives overwhelming signals than plant metabolites. However, the sampling plate can be heated by a temperature controller, so that solvent could be removed by evaporation. This is called the evaporation method.

These two preparation methods were investigated on all the materials. Each sample was prepared in triplicate before analysis.

Solid sample method

For the plant root material, 2 mg of powder was enough to cover the crystal on the ATR sampling plate. For the dripping pills: each pill was cut through the middle and half of the pill was placed on the crystal with the cut side down. The biomarkers were so tested as fine powders. Pressure was applied on the sample until the first safe lock was clicked. The spectra were then obtained using 8 scans of each sample. The background signal of the empty sampling plate was recorded every three sample measurements and subtracted from the sample spectra.

Evaporation method

0.5g of sample was extracted with 25 ml 70% methanol under sonication for 30 minutes at 30°C[63]. The solvent was filtered though a 0.2 μm syringe filter. 10 μl of the solvent was added in 5 steps on the ATR-IR sampling plate which was maintained at 40°C. After the signal stabilized the spectra were obtained using 8 scans for each sample. The background signal of the empty sampling plate was recorded for each sample and subtracted from the sample spectra.

2.2.6 Sensitivity test

HPLC

Protocatechuic aldehyde, a representative phenolic compound in CDDP was used in this sensitivity test. The highest peak in CDDP chromatogram detected at 281 nm was caused by protocatechuic aldehyde. Protocatechuic aldehyde is also the cheapest representative chemical biomarker of CDDP. Standard solutions contain 10, 20, 40, 80, 120 and 160 μg/ml protocatechuic aldehyde were subjected to HPLC analysis using method described in section 2.2.3. The peak at 12.82 minutes was caused by protocatechuic aldehyde. The concentration of protocatechuic aldehyde in CDDP extract was calculated to be 41.19 μg/ml, using the standard curve. 10, 20, 50, 100, 200, 300 and 500 μg protocatechuic aldehyde was added into every 25 ml CDDP 70% methanol extraction, in order to achieve 1%, 2%, 5%, 10%, 20%, 30% and 50% protocatechuic aldehyde-spiked CDDP solutions.

NMR spectroscopy

The protocatechuic aldehyde-spiked Danshen solutions obtained in HPLC sensitivity test in the previous section were used in this NMR spectroscopy sensitivity test. 1 ml of each obtained solution was evaporated into dryness, and dissolved in 0.6 ml of deuterated methanol. This solution was analysed by NMR spectroscopy using the same parameter described in section 2.2.4.

IR spectroscopy

Salvianolic acid B, a representative phenolic compound from Danshen, was used in this sensitivity test. 0.5 g of Danshen powder was extracted by 25 ml 70% methanol under sonication for 30 minutes at 30 °C. The solvent was filtered though a 0.45 μm syringe filters. The dried residual is 400 mg therefore 100 mg of the Danshen was extracted into the 70% methanol solvent. To the extract, different percentages of SAB in terms of the dry extract (100 mg) were used to spike the Danshen: 1%, 3%, 5%, 7.5%, 10%, 15%, and 20% (0.04, 0.12, 0.16, 0.20, 0.32, 0.4, 0.6 and 0.8 mg SAB was added to 1 ml of DS extraction respectively). Extraction evaporation method was used in this study.

2.2.7 Data pretreatment

The raw dataset exported from chromatograms and spectra are usually not appropriate to perform PCA directly. For chromatographic data, alignment is necessary because the minor shifts of peaks are inevitable due to the intrinsic limitations, such as minor changes in mobile phase compositions, degradation of the stationary phase, and problems related to the detectors[86].

Alignment was proved to be helpful, in a study, after alignment of the HPLC chromatograms of 79 batches of herbal medicine, the generated PCA model can explain 35.1% of the total variance, instead of 22.4% before alignment[86].

For NMR spectra, bucketing is usually necessary to reduce the large dataset into a smaller one. Usually the bucket size range from 0.002 to 0.04 ppm. Reduced dataset requires less time to calculate, and also act as a solution to the problem of the minor shifted peaks in NMR spectra.

Normalization processes have been applied on all the data obtained by all methods, so removing concentration effects.

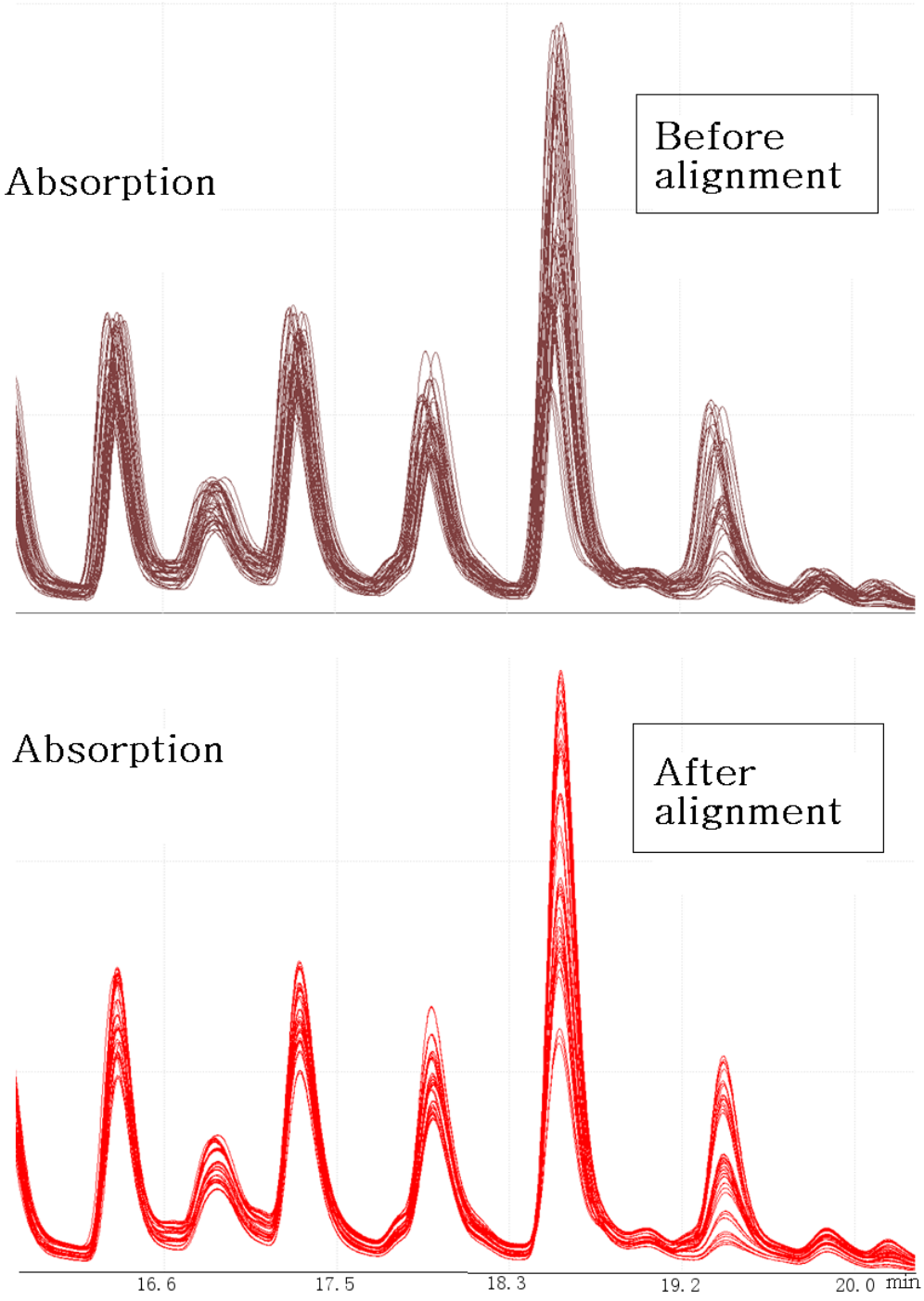


Figure 2.6: *HPLC chromatograms of 5 batches of CDDP from 16 to 20 minutes. Top figure: HPLC before alignment. Bottom figure: HPLC after alignment using Lineup Software.*

Peak alignment

Alignment is an essential part of multivariate data analysis such as PCA. The retention times of constituents in identical material often shift slightly because of time, temperature, gas accumulated in HPLC system, slight variance of composition of mobile phase, *etc.* Chromatograms without alignment are not suitable for PCA since the retention times of the same compound may be slightly different, which effects the correlation between PCA scores plot and loadings plot.

When performing alignment, all the data files are aligned to one target file. Target file is the most reliable fingerprinting signal obtained that all the other files are aligned according to it. In this study, the target file was selected by Pirouette® using outlier diagnostics. The sample with the lowest Mahalanobis distance was selected as the target file. A lower Mahalanobis distance means closer to the average of all the signals. In this case, sample 3-1-1 was selected as the target file because it met the above criteria. Figure 2.6 is an example of HPLC chromatogram of 5 batches of CDDP before and after alignment. The retention times of the largest peak in the chromatograms of 5 batches of CDDP were in a range from 18.548 to 18.598 minutes before alignment. After alignment, this range has become 18.577 to 18.580 minutes. The difference between retention times of the same compound has been changed from 0.05 minute to 0.003 minute before and after alignment which made the variability more than 30 times less. This test indicates that Lineup Software (Informetrix Inc., Woodinville, WA) is suitable for alignment of HPLC chromatograms before PCA test.

For NMR spectra, if a large bucketing size is used, such as 0.04 ppm, alignment is usually not necessary. Because bucketing itself acts as an alignment function. However, if a smaller bucketing size such as 0.002 ppm is used, alignment can help to improve the data quality.

The HPLC chromatograms and NMR spectra were aligned using Lineup™ software. A few parameters needed to be set before alignment. Slack parameters (Warp) is usually set to be 2 to 4. The segment size was best set as the quantity of data points to form a peak. For example, if on average a peak

contains 30 data points in a chromatogram, the segment size will be set as 30. The sides of the fingerprints are sometimes not aligned so extended baselines (usually called paddings) are added to both sides of the profiles. For example, many zeros was added to the sides of chromatograms. However, this was found not necessary in this study as the peaks in both chromatograms and spectra are not near the edge. "Shifts" value means how many data points are desired in order to shift the fingerprint globally. If the shift between target file and samples file is more than three peaks width, the alignment results will not be satisfactory. Therefore "shifts" value needs to be set up. In this study, the alignment parameters were set up as follows: for HPLC, warp, 3; slack parameters, 100; shift, 0; for NMR spectra, warp, 3; slack parameters, 20; shift, 6.

Normalization

There are a few ways to do normalisation. The normalization algorithm of Mnova software is simple: the selected peak (usually the highest one) is set to 100 and each data point in the spectrum is expressed in comparison to it. The result using Pirouette® was better since it is based on a more powerful algorithm known as vector length normalization (equation 2.1). Each variable is divided by the length of the sample vector f_i. The symbol m* indicates included variables, i is the number of sample, j is the variable number. In the case of NMR, number of sample means the number of spectrum, variable number means the chemical shifts number.

$$f_i = \left(\sum_{i}^{m^*} x_{ij}^2 \right)^{1/2} \tag{2.1}$$

Figure 2.7 shows the difference before and after normalisation of ^{1}H-NMR spectra 5 batches of Danshen. Before normalisation, some concentration difference can be seen in the intensities. After normalisation, the peaks are more overlapping each other.

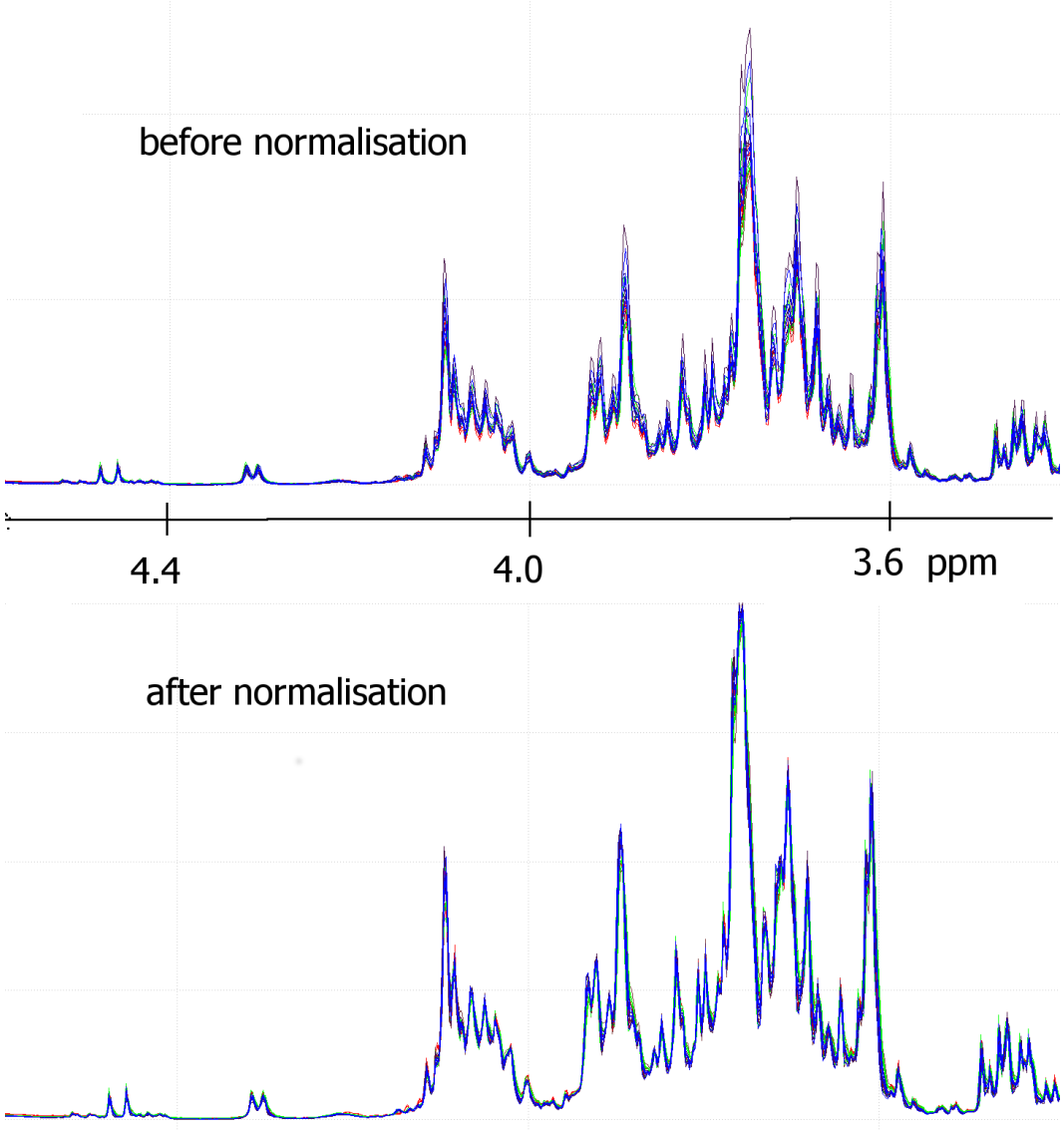


Figure 2.7: *Effect of normalization on ^{1}H-NMR spectra 5 batches of Danshen from 3.5 to 4.5 ppm. The Y-axis is the intensity in arbitrary units, the number was not shown.*

Scaling

The high intensity peaks in NMR spectra of TCM extracts were usually caused by primary metabolites. Due to the low content of secondary metabolites in herbal medicines, the NMR spectroscopic peak of secondary metabolite is usually low. However, secondary metabolites have important pharmacological effect despite their low content[87][88][89]. Therefore, the difference of content of secondary metabolites should be taken account in the analysis of TCM. If the original NMR spectroscopic data (without scaling) were put into analysis directly, the analytical results would be dominated by the larger peaks, which came from primary metabolites. To solve this problem, data scaling method is used to give equal or increased importance of smaller peaks.

Different scaling methods can be used prior PCA. The most popular ones for NMR spectroscopic data are autoscaling equation 2.2 and pareto scaling equation 2.3 [90].

auto scaling:

$$x_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \tag{2.2}$$

pareto scaling:

$$x_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{s_j}} \tag{2.3}$$

In the above equations, x_{ij} is the value of the sample located in row i and column j. $\bar{x}_j$ means the average value of samples in column j, in this case, the average value of responses with the same chemical shift in different samples. s_j means the standard deviation of samples in column j.

For autoscaling, all the data points in NMR spectra are given the same importance. Noise is often introduced in this way, because baseline data points are also considered to be important to the results using autoscaling. Pareto scaling is very similar to autoscaling, however the difference between each data point and the average data point was divided by the square root of standard deviation instead of the standard deviation. This is a compromise approach between non-scaling and autoscaling. Small peaks in NMR spectra are given greater importance than the original data, yet the baseline noise

are minimalised in the later PCA. Pareto scaling was used for NMR data throughout studies in this thesis.

Baseline related issues

For ATR-IR spectra, first derivative was performed for baseline correction and normalisation was performed for intensity correction. Figure 2.8 shows the quality of data before and after derivatisation. Before derivatisation, the baseline near 2,000 cm^{-1} is not aligned while after derivatisation, all the baselines are almost overlapped, therefore only the real difference will be analysed.

First derivatives of the HPLC chromatograms of CDDP were also tried. Derivatisation measures the slope of the signals at each point therefore the integrity will not be lost, but may permit great discrimination. However, noise of the original signal might be increased by derivatisation. After the first derivative, batch 1 and 3 were separated better but batches 1, 2, 4, and 5 were not separated as clearly as the one before derivatisation. Therefore derivatives are also not recommended unless the baseline is particularly bad.

Baseline exclusion was tried on ^{1}H-NMR spectra of 5 batches of Danshen. Baseline noise was deleted and the data was subjected to PCA. The scores plot of this PCA was compared with the previous PCA and no difference was found in the scores plot. This indicates that the PCA of the NMR spectrum was not affected by the noise in the baseline because it is too small.

2.2.8 Multivariate data analysis (MVA)

Principal components analysis (PCA) was performed using Pirouette® 4.0 chemometrics modelling software (Infometrix, USA).

HPLC chromatograms comprising 2,100 discrete regions by data acquisition every second from 0.00 to 45.00 min. Alignment for HPLC chromatograms used LineUp software 3.0 (Infometrix, USA). The resulting data were formatted as an ASCII-text file and exported into Microsoft Excel 2007 using Totalchrom Convert 6.31(PerkinElmer, USA). The region concerning the internal standard which had been added to all samples between 3.06 and

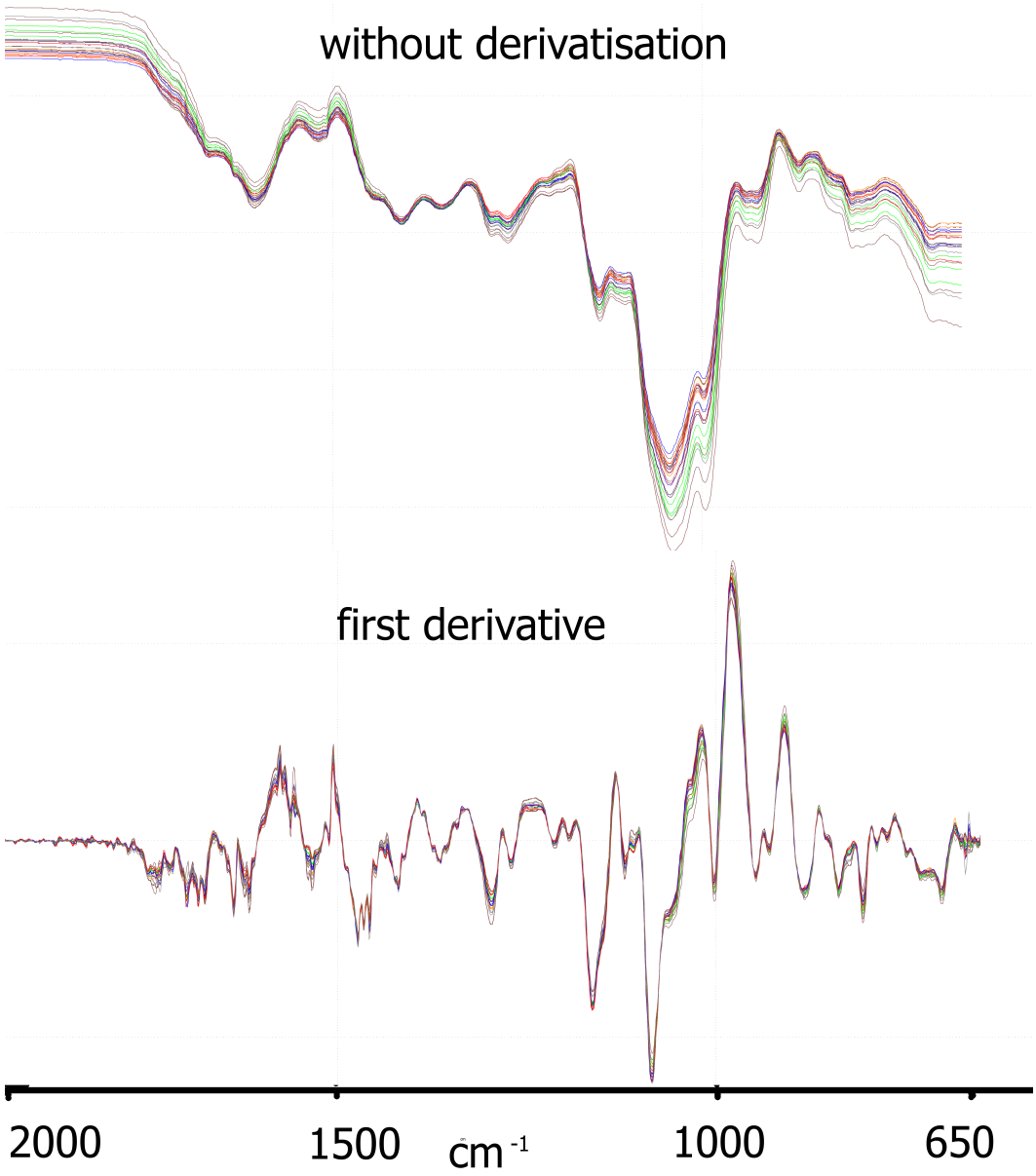


Figure 2.8: *IR spectra of different SAB-spiked Danshen extract 2,000 cm^{-1} - 650 cm^{-1}. Upper figure: the IR spectra after normalisation without derivatisation. Lower image: the IR spectra after normalisation with first derivative.*

4.35 min was eliminated. The remaining 2,050 integral regions were retained for further analysis. Normalization and mean-centering without any scaling was applied prior to PCA and PLS analysis.

NMR spectra were reduced to 4,250 discrete chemical shift regions by digitisation to produce a series of sequentially integrated regions, δ0.002 wide, between δ0.5 and 9, using MestReC 1.1 (Mestrenova, Spain). The resulting data were exported into Pirouette® 4.0 and selected regions removed around the residual solvents including water (δ4.566-5.078), methanol (δ 3.244-3.338), and also the lipid region (δ1.192-1.454, only for CDDP). The remaining 3813 integral regions were retained for further analysis. Normalization of NMR spectra was carried out using Pirouette® 4.0 (Infometrix, USA). Normalization and pareto scaling was applied prior to PCA.

IR spectra from 665-2000 cm^{-1} was analysed. 1^{st} derivatisation, normalization and pareto scaling was applied prior to PCA.

2.3 Results and discussion

All of the materials listed in table 2.1 and table 2.2 were tested using HPLC, NMR and ATR-IR spectroscopy. Two sample preparation methods for ATR-IR spectroscopic analysis were tested and compared. Danshen grown in Shandong and Shaanxi province in China were compared. Sanqi plants supplied on different dates from one Tasly's planting base to Tasly company were studied, because the production time might also affect the quality of Sanqi. CDDP produced by Tasly with different batch numbers and production dates were studied, because the quality of each batch of CDDP might be affected by the plant materials, operations, and reproducibility of manufacture machines. Correlations between plant materials, intermediate materials and finished products were studied.

2.3.1 Comparison of IR sampling methods

For the IR spectroscopy, different sampling methods have been tested including the solid sample method and extraction evaporation method. The

extraction evaporation method has a few disadvantages: 1. It is time consuming. It took 5 minutes for the liquid extraction to evaporate into dryness. The solid sample method only needed 30 seconds. 2. It might have the risk of compound reactions using the hot plate, although the hot plate was only set at 40°C. Samples could be measured at room temperature using solid sample methods, which is lower than 40°C, and less likely to cause any reactions. 3. For TCM products that decrystallise at high temperature, the solid sample method is preferred, because low resolution spectra were obtained from decrystallised samples. Without the effects of crystallisation, the resolution of IR spectra obtained from extraction evaporation method is the same as solid sample method. 4. Solid sample method is less selective than extraction evaporation, because no extraction was performed, no compound was lost. In extraction evaporation method, some compound may be lost using specific solvents, because no solvents could extract all compounds from a plant.

The extraction evaporation method has some advantages over the solid sample method on sensitivity test. In the sensitivity test, different percentages of SAB were added to the Danshen plant. There are two ways of doing this: 1. Powdered SAB standard compound could be added to the Danshen powder in percentage, followed by analysing with solid sample method. 2. SAB standard was dissolved in Danshen liquid extraction at a known percentage, followed by extraction evaporation method. For method 1 using the solid sample method, ground plant root powder was filtered through a 320 μm sieve. However the measuring crystal is only 2 mm in diameter, and the evanescent wave only protrudes the sample for 0.5 - 5 μm (FT-IR Spectroscopy Attenuated Total Reflectance, Technical note, Perkin Elmer, Massachusetts). If the SAB powder was not perfectly suspended in plant root powder, a representative spectrum could not be obtained. Therefore extract evaporation method might give more reliable spectra compared to the solid sample method.

The method chosen depended on experimental needs. Solid sample method has advantages over extraction evaporation method on all the samples studied in this chapter, except in the sensitivity test.

2.3.2 Crystallization affects the results of ATR-IR test

It is difficult to obtain good resolution spectra of liquid extraction of CDDP using the evaporation method described in section 2.2.5. In order to obtain well resolved spectra using ATR-IR, the sample has to be crystallized. Sometimes unsharp spectra were obtained (Figure 2.9 "State 2") when measuring CDDP solvents and this was suspected to be a problem related to crystallisation. In order to investigate this problem, a measurement of total absorption over time of CDDP extract using different temperatures was carried out(Figure 2.10).

The CDDP liquid extraction spectrum was found to be temperature sensitive using the evaporation method. Figure 2.10 shows the total absorption of CDDP (integration of absorption between $4000\,\mathrm{cm}^{-1}$ and $650\,\mathrm{cm}^{-1}$ over time). The temperature of the sampling plate was maintained at 40°C before the sample was loaded. At the 7^{th} second, $5\,\mu$l of CDDP liquid extraction was added to the crystal. The total absorption increased rapidly because the liquid solvents contain methanol and water which have strong absorptions. The absorbance between 3100-3500 cm^{-1} is caused by -OH absorption of water. The absorbance between 2800-3000 cm^{-1} is contributed by methanol (refer to figure 2.9). At the 34^{th} second, methanol had evaporated so the concentration of water in the thin layer on top of the crystal had become higher, therefore the total absorption increased. ATR-IR spectroscopy only measures a thin layer of samples on top of the sampling plate, no matter how much sample was loaded on the sampling plate. After 150 seconds, all the water had evaporated, leading to a decrease of the total absorption. The crystallization process of the sample started from the 220^{th} second to the 280^{th} second increasing the total absorption from the 0.047 to 0.054 (termed "state 2"). This may be because the extraction of CDDP crystallized after being dried continuously at 40°C. This might be caused by the presence of polyethylene glycol (PEG) 4000 or 6000 in CDDP. By raising the temperature to 50°C at the 577^{th} second, the total absorption became lower, and the material dropped to state 2. This might be because crystallisation of CDDP extract was disturbed by the high temperature. The total absorption

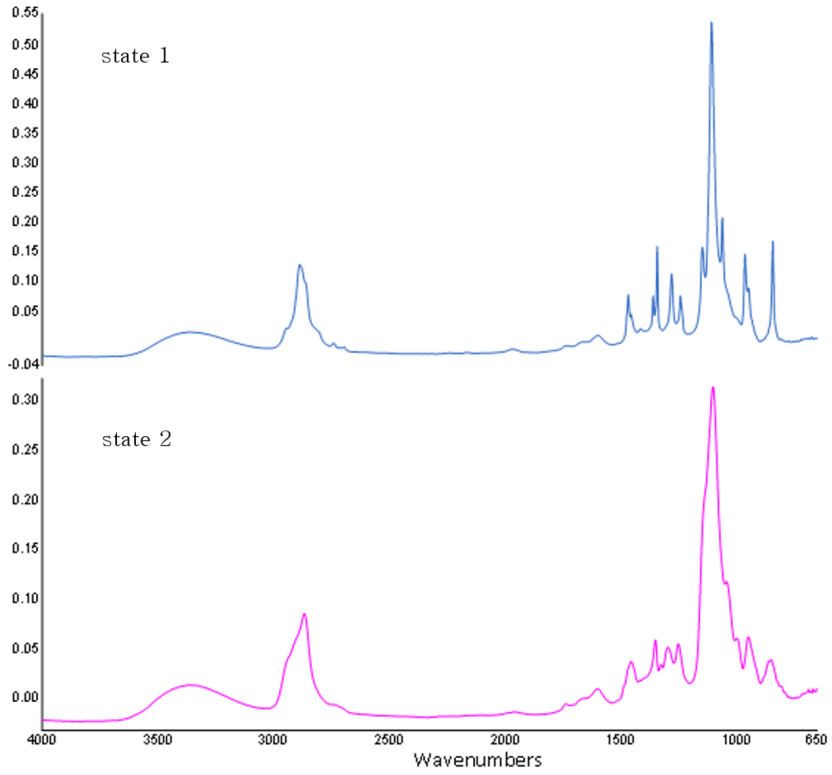


Figure 2.9: *Top spectra: ATR-IR spectrum of extraction of CDDP using evaporation method at 1073 second of figure 2.10. This is a representative spectrum of which so called "State 1" by the author. In state 1, CDDP extract was crystallised, and the IR spectrum shows sharp peaks. Bottom spectrum: ATR-IR spectrum of extraction of CDDP using evaporation method at 972 second of figure 2.10. This is a representative spectrum of which so called "State 2" by the author. In state 2, CDDP extract was not crystallised, and the IR spectrum shows broader peaks. Y-axis: absorption.*

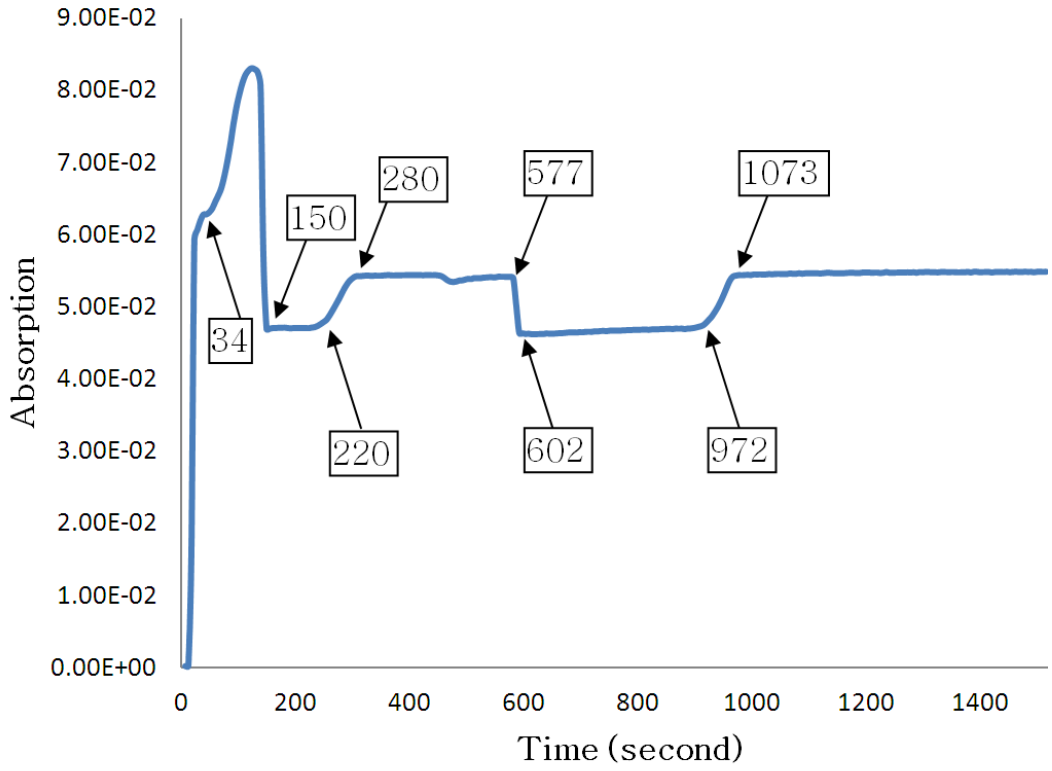


Figure 2.10: *Total absorption of 70% aqueous methanol extraction of CDDP over time. Total absorption is the integral of absorptions of a single IR spectrum over the range 650-4000 cm^{-1}. This curve monitored the changes of the total absorption changes by adding samples and adjusting the temperatures of the sampling plate. The following actions were carried out by the author at these time points. 0 second: the temperature of the sampling plate was maintained at 40° C, 7^{th} second: 5 ul of 70% aqueous methanol extraction of CDDP was added to the crystal, 577^{th} second : the temperature of the sampling plate was raised to 50° C, 972^{th} second: the temperature of the sampling plate was reduced to 40° C.*

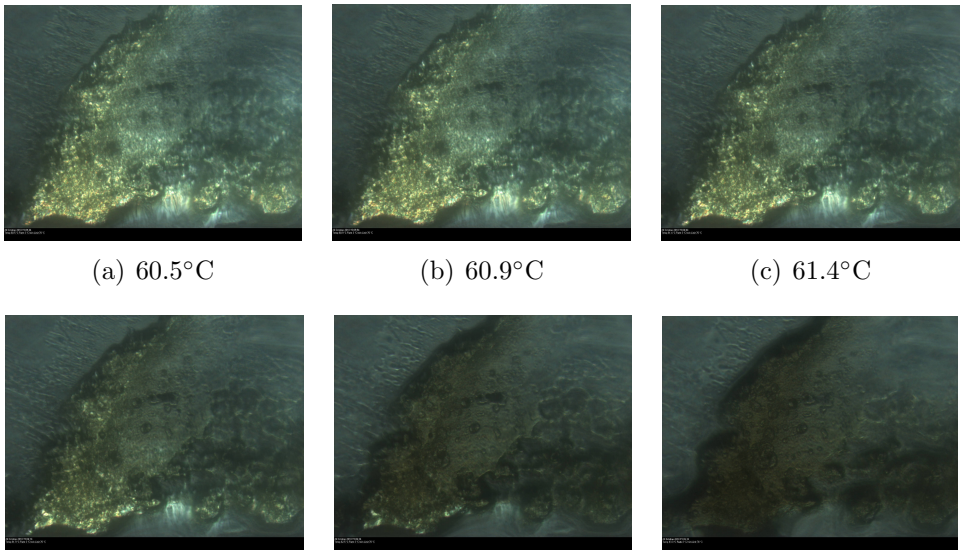
remained at state 2 until the temperature was dropped to the 40°C at the 972^{nd} second, and the spectrum switched to "state 1". It could be that the extract of CDDP recrystallised at 40°C. Crystallised sample have less bond movement freedom therefore has a lower absorption in IR spectra. In figure 2.10, from the 150 second, high resolution spectra were obtained from the 280^{th} seconds to the 577^{th} seconds and from the 1073^{rd} seconds to the end. Low resolution spectra were obtained from the 150^{th} seconds to the 220^{th} seconds and from the 602^{nd} seconds to the 972^{nd} second(see figure 2.9).

Crystallization study of CDDP using hot stage microscopy

In order to further evaluate the crystallisation of CDDP, a polarization study was performed using a Leitz Dialux 22EB polarized light microscope (Leitz, Wetzlar, FRG) with a Linkam TP 92 hot stage control unit. A camera was attached to this microscopy and software called Linksys 32 could capture pictures and videos automatically. Natural lights vibrates in random directions. In the contrary, polarized light vibrates in only one direction. In a polarized light microscope, the test sample is placed in between a polarizer and an analyzer, which are perpendicular to each other. When no sample or an uncrystallised sample is being tested, the light polarized by the polarizer is blocked by the analyzer, therefore it appears dark. When a crystallized sample is being tested, the light is rotated by the sample and can get through the analyser, therefore,it appears shiny. Less than 0.5 mg CDDP was put on the sampling plate. This is because heat transfer for big piece of sample is slow, so long time is needed to be crystallise/decrystallise.

Figure 2.10 shows the process of decrystallisation of CDDP with increased temperature. The temperature was increased from 30°C to 70°C in the speed of 3°C per minute. At 61.4°C, the CDDP sample was still shiny under the microscope, which means it was still crystallised. Between 61.4°C and 62.4°C, the CDDP sample appeared darker and darker, until 63.0°C, the sample appeared completely dark, which means that the sample decrystallised completely.

The cause of decrystallisation of CDDP in high temperature might be due



(a) 60.5°C (b) 60.9°C (c) 61.4°C

(d) 61.9°C (e) 62.4°C (f) 63.0°C

Figure 2.11: *0.05 mg CDDP was measured on a hot plate under a microscope with a polarization mirror, temperature raising from 30.0° C to 70.0° C in the speed of 3° C per minute. Decrystallization observed from 60.5° C to 63.0° C. (a) 60.5° C, (b) 60.9° C, (c) 61.4° C, (d) 61.9° C, (e) 62.4° C and (f) 63.0° C*

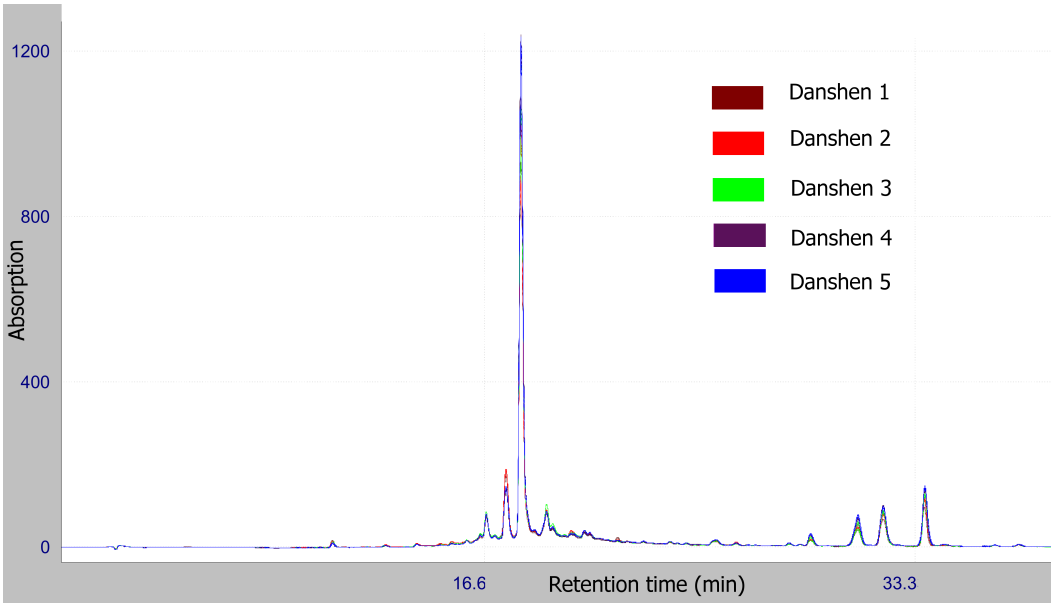


Figure 2.12: *Overlapped HPLC chromatograms from 0-40 min of 5 Danshen no. 1-5. Danshen 1 and 2 were produced in Shaanxi; Danshen 3, 4 and 5 were produced in Shandong.*

to the presence of PEG. Decrystallisation was only found to occur in CDDP but not in Danshen or Sanqi plant extraction. The only extra ingredient of in CDDP than Danshen and Sanqi are borneol and PEG. Borneol is highly volatile and evaporated easily in high temperature. PEG is easy to dissolve in water. Therefore, the 70% aqueous methanol extraction of CDDP contains PEG. So only PEG might be the reason for decrystallisation of CDDP. This was also found in previous research[91].

2.3.3 Visual inspection of HPLC chromatogram, NMR and IR spectra

70% methanol extract of 5 batches of Danshen roots grown in different regions (2 batches grown in Shaanxi, 3 batches grown in Shandong supplied by Tianjin Tasly pharmaceutical Co. Ltd.) were analysed by HPLC to evaluate the effect of growing environment. Figure 2.12 is a collection of 5 batches Danshen HPLC chromatograms. All the chromatograms looks almost iden-

tical except some small differences. For example, Danshen 1 and 2 (Shaanxi) have a higher signal at 17.5 min than Danshen 3, 4 and 5 (Shandong).

NMR spectra of 5 batches of Sanqi supplied by Tasly planting based on different dates looks almost identical (Figure 2.18). It is also difficult using IR spectra of to distinguish the highly similar TCM samples without PCA.

Therefore, multivariate data analysis was used to analyse the fingerprints of HPLC, IR and NMR spectroscopy.

2.3.4 Multivariate data analysis

PCA is an unsupervised method to reduce the dimensionality of the variables in sets of data. For example, in the HPLC chromatogram dataset, each retention time is one variable. In this study, each HPLC chromatogram was recorded over 53 minutes and the signal was detected every 0.0016 minutes, thus results in (53/0.0016=) 31,800 data point in each chromatogram dataset, therefore each HPLC chromatogram dataset has 31,800 variables. Each variable represents the absorbance peak height at that specific retention time. All the chromatograms studied in this chapter used identical chromatography condition, therefore all the chromatograms can share the same retention times, thus all the chromatograms can form one large dataset with 31,800 variables. PCA reduces and summarise these 31,800 variables into a few principal components.

The first PC counts for the most variance among all the PCs. Usually, the first two or three PCs are chosen to show the scores plot. Use a 2 PCs scores plot as a example, PC 1 and PC 2 are the axis of the plot. On this plot, observations are plotted according to their scores. In the case of HPLC, each chromatogram is one observation. The distance between samples is related to how different they are. The further away two samples are from each other, the more different they are. (0, 0) means the average of all the chromatograms. If a sample is very far away from the centre of the coordinate, this sample is very different from the average of all the samples.

The loadings are the coefficients multiplied by the variables. The projection of variables on the PC axis means how much this variable is contributing

to this PC. Loadings plots can be used to detect which compound contribute the most to the differentiation of the observations.

For NMR and IR spectroscopic analysis, the theories of PCA is the same, but the variables means chemical shifts (ppm) and wave numbers (cm^{-1}) respectively in this case.

2.3.5 Sensitivity test

The sensitivity test acts as an initial evaluation of different fingerprinting methods. A series of percentages of a standard pure chemical were use to manually spike Danshen or CDDP extractions, thus create a series of very similar samples, but with known differences. This allows to evaluate if PCA coupled fingerprinting methods are sensitive enough to pick up the differences, and to tell what is the difference.

HPLC

Protocatechuic aldehyde, a representative phenolic compound in CDDP was used in this sensitivity test as described in section 2.2.6.

The PCA scores plot (Figure 2.13) shows 6 replicates of CDDP extract without spiking (brown) and the same CDDP extract spiked by 1% - 30% of protocatechuic aldehyde.

The non-spiked CDDP samples are located on the left, and the most spiked CDDP samples are located on right. In between them are the other percentage spiked CDDP extract arranged from low to high concentration from left to right. The non-spiked samples are located further away from the >5% spiked samples than from the <5% spiked samples, which means that the CDDP samples contain 5% extra protocatechuic aldehyde is very different from CDDP solutions. The non-spiked CDDP samples overlapps with 1% spiked samples, but not mixed with 2% spiked samples. Therefore CDDP samples which contain 2% extra protocatechuic aldehyde could be differentiated from non-spiked samples. This means PCA-HPLC method is sensitive enough to distinguish 2% and above difference of protocatechuic aldehyde among CDDP samples. Figure 2.13 also indicates that the content

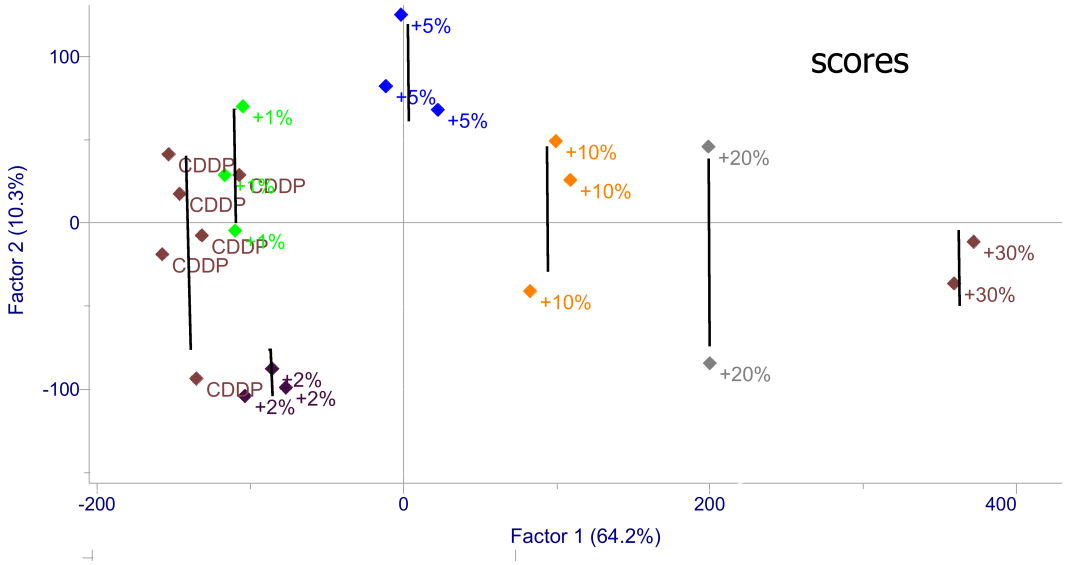


Figure 2.13: *PCA scores plot of HPLC chromatogram dataset of 6 non-spiked CDDP sample (indicated by brown squares. These samples were circled in the plot.) and 1% (green), 2% (purple), 5% (blue), 10% (orange), 20%(grey), and 30% (dark red) protocatechuic aldehyde-spiked CDDP samples.*

difference of protocatechuic aldehyde in CDDP is less than 2% which means that the HPLC analysis was sensitive and the batch difference of CDDP is small.

^{1}H-NMR spectroscopy

0%, 1%, 2%, 5%, 10%, 20%, 30% and 50% protocatechuic aldehyde-spiked CDDP solutions were used in NMR sensitivity test. These different percentages spiked samples could not be distinguished by PCA scores plot of the NMR datasets. The ^{1}H-NMR assignment for pure protocatechuic aldehyde is as follows: 2.24 (1, 3H), 6.83 (5, 1H, *dd*, J=8.44 Hz, J=1.28 Hz), 7.59 (6, 1H, *dd*, J=8.44 Hz), J=1.73 Hz), 7.35 (7, 1H, *dd*, J=1.73 Hz, J=1.28 Hz). These peaks could not be identified from the spiked CDDP NMR spectra. These peaks were either too low to be detected, or only contributed to a small percentage of a peak that overlapped by the response of other compounds.

The content difference of 0% and 50% spiked CDDP samples was only 20.5 μg protocatechuic aldehyde. This small amount is difficult to be detected. (`http://www.chem.wisc.edu/~cic/nmr/Guides/Other/sensitivity-NMR.pdf`). No difference could be seen between 0%, 1%, 2%, 5%, 10%, 20%, 30% and 50% protocatechuic aldehyde-spiked CDDP solutions by the NMR-PCA method in the present study.

However, this does not mean that NMR-PCA methods could not be used to distinguish minor differences between plants. Previous studies showed that ^{1}H-NMR-PCA could be used to distinguish small differences between herbs, especially primary metabolites[30][16].

IR spectroscopy

Salvianolic acid B was used to spike the Danshen liquid extraction to produce 1%, 3%, 5%, 7.5%, 10%, 15%, and 20% of dry extract of Danshen solution (w/w). It is worth noticing that the percentage of SAB does not mean the percentage of SAB in terms of the SAB content of Danshen, but the total dry extract of the Danshen. This experiment was designed like it is because ATR-IR was not expected to be as sensitive as HPLC, and a very small

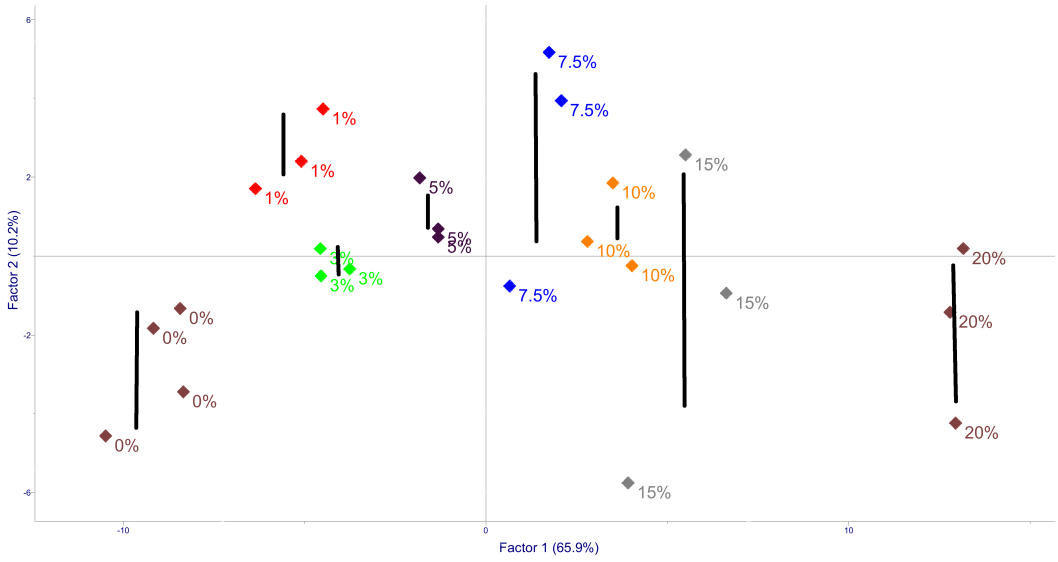


Figure 2.14: *PCA scores plot of IR spectra of Danshen extract spiked by 1% (red), 3% (green), 5%(purple), 7.5%(blue), 10% (orange), 15% (grey), and 20% (dark red) SAB. 0% (brown) means no spiking. From the highest to the lowest percentages SAB-spiked sample groups were separated along factor 1. Black lines that run through each sample group have been manually drawn for easy visualization of locations.*

amount of SAB will probably not be detected.

Figure 2.14 is the PCA scores plot of Danshen extract spiked by 1% to 20% SAB. This PCA scores plot consists of 2 factors: factor 1 accounts for 65.9% of the total variance, and factor 2 accounts for 10.2%. From the highest to the lowest percentages SAB-spiked sample groups were separated along factor 1. This means factor 1 is the main difference among those sample groups, and it is the common difference between any groups.

There are also some differences within groups. Samples within the same group spread along factor 2.

By inspecting the corresponding loadings plot, factor 1 is mainly positively contributed by 1050, 1142, and 1165 cm^{-1}, and negatively contributed by 1537, 1615, 1300, and 1121 cm^{-1}. These loadings should be related to the addition of SAB. However, it is difficult to assign or interpret the data because first derivatized spectra show the slope of each datapoint forming the peak, instead of directly measuring the height of the peak. A spectrum represent a mixture of chemical compound is even more difficult to interpret.

1% SAB-spiked samples are separated from 0% spiked samples indicating that ATR-IR combined with PCA can distinguish 1% of SAB in terms of dry extract of Danshen from the non-spiked Danshen.

2.3.6 Discrimination of Danshen samples from different regions

In this section, three methods were used to distinguish Danshen roots produced from different planting regions. In table 2.1, Danshen 1 and Danshen 2 were produced in Tasly's Shaanxi planting base. Danshen 3, Danshen 4, and Danshen 5 were produced in Tasly's Shandong planting base. CDDP produced by Tasly mainly used Danshen from these two planting bases. It is important that these two planting bases produce almost the same quality of Danshen, otherwise the standard of CDDP could not be set up.

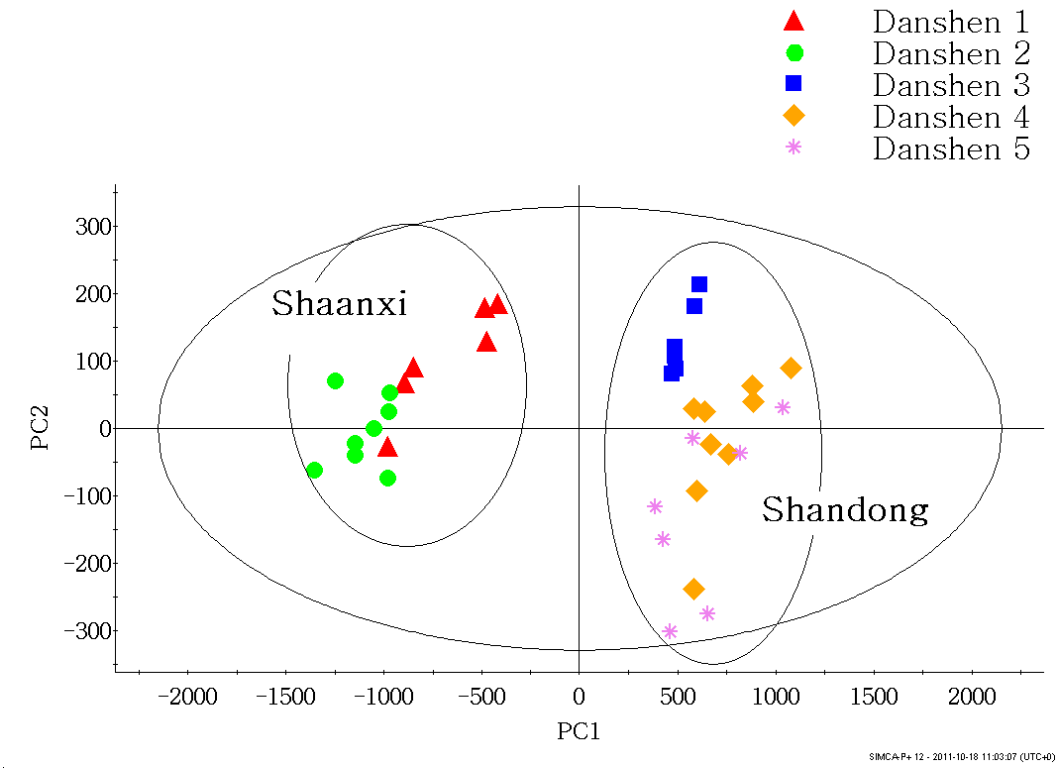


Figure 2.15: *Scores plot PC1 versus PC2 obtained from HPLC fingerprints of Danshen from different regions. Left hand side of zero coordinate are Danshen grown in Shaanxi province: Red triangle, Danshen 1; green dot, Danshen 2. Right hand side of zero coordinate are Danshen grown in Shandong province: blue square, Danshen 3; orange diamond, Danshen 4; purple star, Danshen 5.*

HPLC method

Danshen plants with batch numbers 1 and 2 were grown in Shaanxi province and batch numbers 3, 4, and 5 were grown in Shandong province, China. This region-to-region difference is shown in HPLC-PCA study (Figure 2.15) .There are 9 technical replicates of each batch of Danshen, because the extraction process was repeated 3 times, and each extract was injected 3 times into the HPLC.

PCA scores plots consist of many principal components (PCs). Only the first two main PCs were shown in this case. PC1 and PC2 counts for 83.7% of the total variance, therefore it is a good indication of the differences between samples. PC1 is the summary of HPLC retention times, at which makes the biggest variance between all the HPLC datasets. PC2 is the summary of HPLC retention times, at which makes the second biggest variance between all the HPLC datasets. The retention times belong to PC1 and PC2 were not related to each other, which mean that retention times contributing to PC1 and PC2 should not belong to the same compound.

In figure 2.15, Danshen 1 and 2 were produced in Shaanxi province. These samples are located on the left of the PCA plot with a negative value of PC1, which means they have a negative contribution of PC1. PC1 is a summary of many variables, in this case retention times. Therefore Danshen produced in Shaanxi province had lower peaks at the retention times positively contributed to PC1.

The points for Danshen produced in Shandong province were located on the positive side of PC1. This means that Danshen produced in Shandong have higher HPLC peaks at those retention times which contribute positively to PC1. Higher HPLC peak at a certain retention time and the same experimental condition, means a higher concentration of a certain compound. Therefore Danshen grown in Shandong has a higher concentration of the compounds positively contributed to PC1 than Danshen grown in Shaanxi, and Danshen grown in Shaanxi have a higher concentration of the compounds negatively contributed to PC1. The retention times correlated to those compounds could be found out by comparing the PCA scores plot to the PCA

loadings plot.

PCA loadings plot also consists of many PCs. Loadings means the variables, in this case retention times. The loadings located in the positive side of PC1 contribute positively to PC1, and the loadings located in the negative side of PC1 contributed negatively to PC1. The same case is for PC2. The loadings plots could be compared to the scores plot. If the variable in a loadings plot was found at the same location as a sample in the scores plot, that sample has a high value at that variable, in this case mean the sample has a high concentration of the compound with a retention time at that variable.

By comparing the loadings plot to the original chromatogram, it was found that the peaks with retention times 33.7, 32.12, 31.13, 29.30, 25.63, 18.63, 18.07, 16.44 min contribute positively to PC1, and peaks at 21.83, 17.50, 12.83 and 10.70 min contribute negatively to PC1. Since Danshen grown in Shandong located on the positive side of PC1, they have a higher content of compounds with retention times on 33.7, 32.12, 31.13, 29.30, 25.63, 18.63, 18.07, 16.44 min. Conversely, Danshen grown in Shaanxi have a higher content of compounds with retention times on 21.83, 17.50, 12.83 and 10.70 min.

Most of the peaks contributing to PC1 were identified using HPLC standard addition method. A small amount of Danshen standard reference was used to spike the Danshen root extraction. The samples before and after spiked extraction were subjected to HPLC analysis. The peak area difference of the standard reference was calculated. If the difference was equal to the amount of spiking, then the identity of the peak was confirmed.

The peaks were identified as following: 10.70 min, salvianolic acid A; 12.83 min, protocatechuic aldehyde; 18.07 min, salvianolic acid B (SAB).

Therefore, Danshen grown in Shandong have a higher concentration of SAB but Danshen gown in Shaanxi have a higher content of salvianolic acid A and protocatechuic aldehyde. As a preliminary study to evaluate methods, not all the peaks were identified.

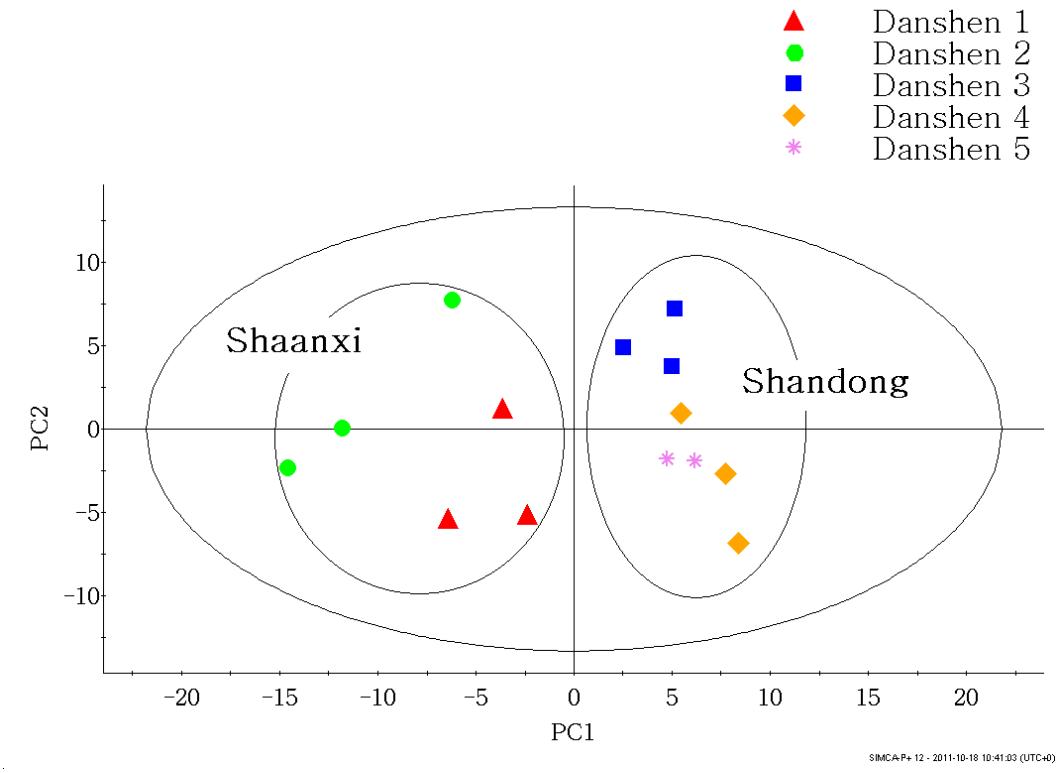


Figure 2.16: *Scores plot PC1 versus PC2 obtained from ATR-IR spectroscopic dataset of Danshen from different regions. Danshen was in the form of powder and measured with the solid sample method. Left hand side of zero coordinate are Danshen grown in Shaanxi province: Red triangle: Danshen 1, green dot: Danshen 2, Right hand side of zero coordinate are Danshen grown in Shandong province: blue square: Danshen 3, orange diamond: Danshen 4, purple star: Danshen 5. Details of the samples were listed in Table 2.1.*

ATR-IR method

Difference between different geographical Danshen could also be detected by ATR-IR -PCA. According to the PCA on the Danshen root powder (Figure 2.16), Danshen grown in Shaanxi and Shandong can be distinguished by PC1. PCA on the Danshen extract using the evaporation method shows the same differentiation as the analysis using Danshen powder with the solid sample method. This indicates that both powdered and extract forms provide good differentiation of Danshen samples of different origins.

By analysing the corresponding loadings plot and observing the reconstructed spectra after first derivatisation and normalisation, the differences between Danshen grown in these two plantations were found to be as follows:

Danshen batch 1 and 2 grown in Shaanxi have higher responses at 2353, 2348, 2305, 2025, 2091, 2085, 2082, 2062, 2026, 2019, 1977, 1972, 1966, 1960, 1956, 1952, 1569, 1558, 1551, 1360, 1115, 1060 and 945 cm^{-1}.

Danshen batch 3, 4 and 5 grown in Shandong have higher responses at 1760-1720, 1630-1610 (1624, 1616), 1513, 1015, 1002, 979, 829, 788 and 675 cm^{-1}.

It is difficult to assign the above bands to a standard compound due to the many overlapping peaks. The spectra of standard reference compounds, such as SAB, consist of mainly sharp peaks and the spectra of Danshen or its extraction consist of mainly broad peaks. This is because Danshen/ its extraction is a mixture of hundreds of compounds and many of their IR absorptions overlap. Therefore the identification of a single biomarker from the spectra of the plant is difficult. However, in regions where it happens to have no overlapping, the identity of some biomarkers could be detected from the spectra of Danshen plants. The derivatized and normalized IR band between 1760-1720 cm^{-1} of salvianolic acid B and Danshen powder show the same shape. This means that IR band (1760-1720 cm^{-1}) in the spectrum of Danshen roots is possibly directly due to absorption by salvianolic acid B. Because 1760-1720 cm^{-1} is higher in Danshen grown in Shandong, it may mean that those Danshen has a higher content of SAB. If true, this coincides with the result shown by HPLC analysis.

NMR spectroscopic method

Good resolution NMR spectra of 5 batches of Danshen products were obtained. However, Danshen grown in Shandong and Shaanxi province could not be distinguished by NMR-PCA method. All the spectra obtained looked almost identical. This is because the concentration of the secondary metabolite is very low, therefore a slight difference cannot be detected.

2.3.7 Differentiation of Sanqi samples with different supply date

ATR-IR method

Sanqi plant roots supplied in different months from the plantation base of Tasly can be differentiated by PCA (Figure 2.17). Sanqi 5 is located on the left side of the PCA map, and has the latest supply date as March 2009. Sanqi 4 (supplied in November 2008) is located in the middle of the map and Sanqi 1 to 3 (supplied in October 2008) are located on the right of the map. A few IR bands have been found to be responsible for the differences and those bands are compared with Sanqi biomarkers ginsenoside Re and Rb_1 in table 2.4. Ginsenoside Re and Rb_1 are two main secondary metabolites in Sanqi.

Table 2.4 shows the derivatized wavelengths of the difference between Sanqi plant roots are close to the derivatized wavelengths ginsenoside Re and Rb_1. The first derivative plots the value of the slope of each data in the IR spectra. It is not directly related to the intensity. However, if one peak is high in the original spectra, it has a higher slope and the lower peaks with the same peak width. This becomes very complex when the testing material is TCM, which contains many chemicals. However, the correlation of differences between Sanqi to ginsenosides might indicate that, the content of ginsenoside Re and Rb_1 are likely to contribute to the date-to-date differences between 5 batches of Sanqi plants. There might be other biomarkers of Sanqi also contributing to the date-to-date difference of Sanqi samples which are not tested.

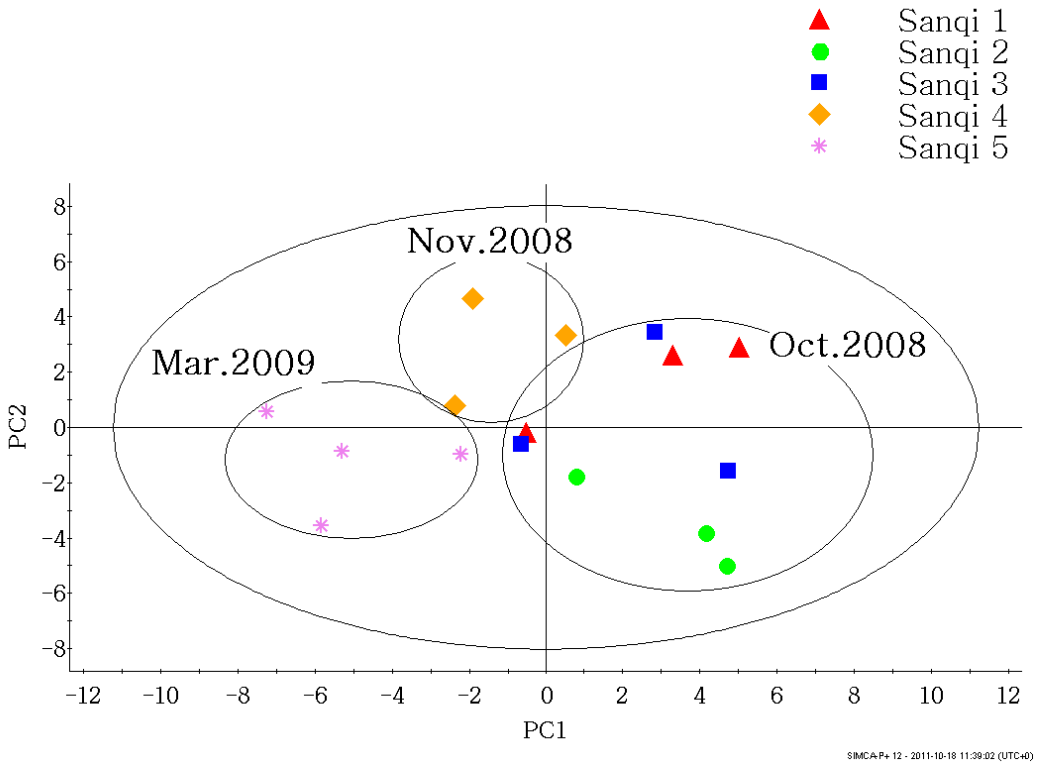


Figure 2.17: *Scores plot PC1 versus PC2 was obtained from ATR-IR spectroscopic dataset of Sanqi supplied in different dates. Sanqi 1,2 and 3 were supplied in October 2008, Sanqi 4 was supplied in November 2008, and Sanqi 5 was supplied in March 2009. All of them were grown in the same planting base in Yunnan. Red triangle: Sanqi 1 supplied on 28/10/2008, green dot: Sanqi 2 supplied on 20/10/2008, blue box: Sanqi 3 supplied on 25/10/2008, orange diamond: Sanqi 4 supplied on 15/11/2008, purple star: supplied on 03/03/2009. Details of the production dates are listed in Table 2.1.*

Table 2.3: *Comparison of the derivatized spectra of Sanqi plant root, ginsenoside Re and Rb_1. Unit in cm^{-1}.*

Difference between Sanqi samples	ginsenoside Re	ginsenoside Rb_1
1751	1752	-
1713	1715	1717
1692	1752	1690
1672	-	1670
1652	1653	-
1649	1648	1648
1460	1464	-
1436	1432	1438
1407	1407	-
1391	1392	1394
1386	1387	-
1340	1342	1338

NMR spectroscopic method

Figure 2.18 are 5 batches of Sanqi supplied in different dates from Tasly's planting base. No differences can be spotted by the naked eye.

Figure 2.19 is the PCA scores plot of 5 the NMR spectra of 5 batches of Sanqi. From right to left, the supply dates arranged as Oct. 2008, Nov.2008, Mar. 2009. This sequence is the same as the IR-PCA shown in figure 2.17 from left to right.

In PCA scores plot, the importance of the negative and positive side of a principal component is equal. Using difference PCA software, the scores plot obtained can be mirrored. Therefore, it does not matter if the sequence of samples if from one side or the other, as long as the order is the same.

Therefore, Sanqi samples were arranged similarly in NMR and IR -PCA scores plot. In NMR-PCA, Sanqi with different dates are more separated from each other than in IR-PCA. This may mean NMR is better at finding the differences from Sanqi metabolites.

According to the loadings plot, Sanqi supplied in Mar. 2009 have a higher response at 5.29, 3.7 and 4.0 ppm than other batches. Sanqi supplied in Oct. 2008 have a higher response at 0.81, 1.52, 3.12, 3.29, 3.61 and 3.65 ppm. As a

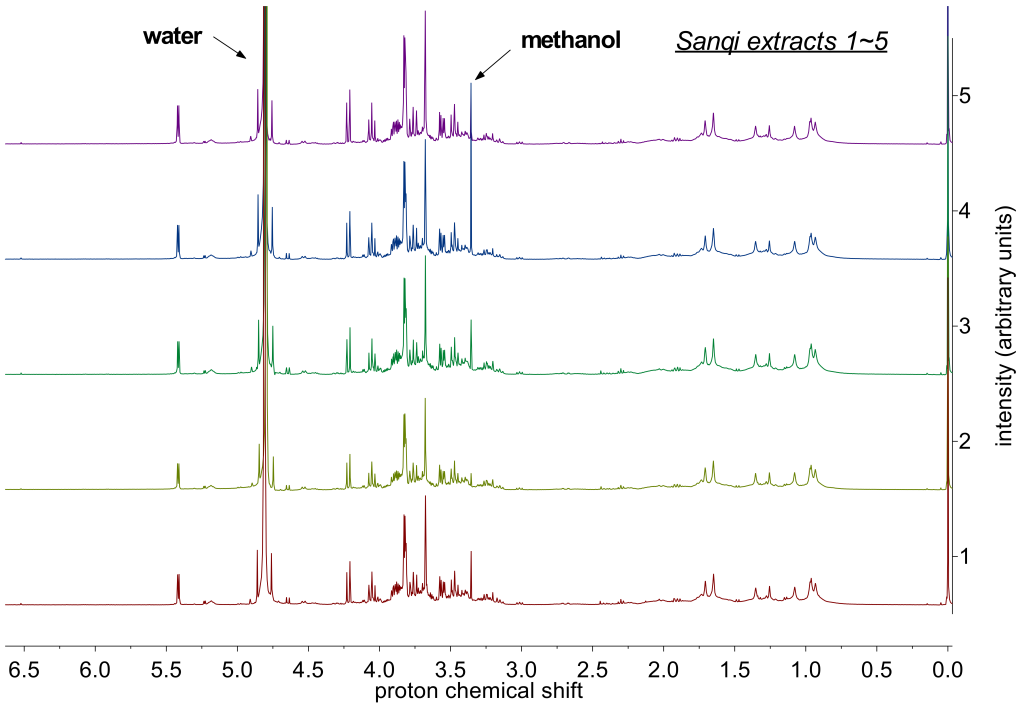


Figure 2.18: *NMR spectra of 5 batches of Sanqi supplied in different dates. Sanqi 1,2 and 3 were supplied in October 2008, Sanqi 4 was supplied in November 2008, and Sanqi 5 was supplied in March 2009. All of them were grown in the same planting base in Yunnan. 0–10 ppm. 1: Sanqi 1, 2: Sanqi 2, 3, Sanqi 3, 4: Sanqi 4 and 5: Sanqi 5*

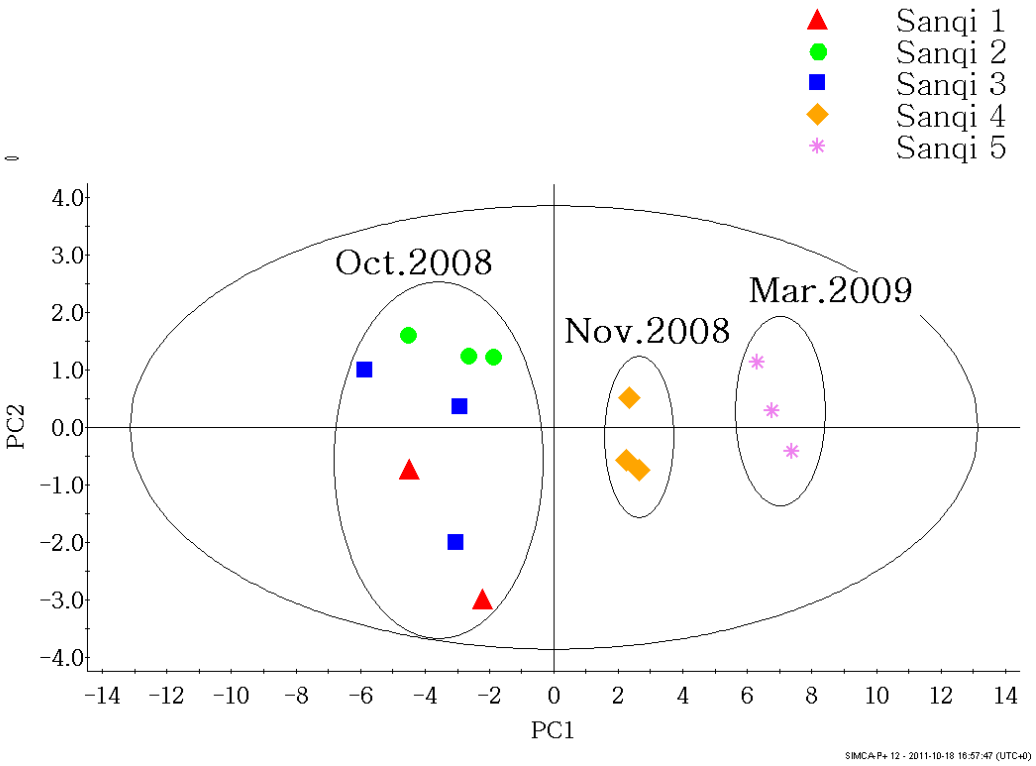


Figure 2.19: *Scores plot PC1 versus PC2 was obtained from NMR spectroscopic dataset of Sanqi supplied in different dates. Sanqi 1, 2 and 3 were supplied in October 2008, Sanqi 4 was supplied in November 2008, and Sanqi 5 was supplied in March 2009. All of them were grown in the same planting base in Yunnan. Red triangle: Sanqi 1 supplied on 28/10/2008, green dot: Sanqi 2 supplied on 20/10/2008, blue box: Sanqi 3 supplied on 25/10/2008, orange diamond: Sanqi 4 supplied on 15/11/2008, purple star: supplied on 03/03/2009. Details of the production dates in Table 2.1.*

brief study of different spectroscopic methods, these peaks are not identified.

However, Sanqi roots supplied in the earliest date(Oct. 2008) are located on one side, those supplied in the latest date (Mar. 2009) are located on the other side, and those supplied in the time between (Nov.2008) are located in the middle in the PCA scores plot, indicates the composition of Sanqi is changing with time. Sanqi is usually harvested after the third year since it was planted. It usually flowers in August, and fruits in September, matures in October or November. The supply date is probably quickly after the harvesting of Sanqi. The Sanqi supplied in Oct. and Nov. 2008 are probably from the same year harvesting. The Sanqi supplied in Mar. 2009 might be a late harvested batch, or have been stored in the warehouse for a while. If this is the case, the separation of the spectra of Sanqi supplied at different times indicates a difference of metabolite composition of Sanqi harvested in different times of the year or stored for different lengths of time. Storage factors such as temperature and moisture affects the shelf-life of herbal medicines[92]. However, the exact storage condition of Sanqi from he company is unknown. In future work, the chemical shifts that contribute to the difference can be assigned to Sanqi metabolites. In chapter 3, the author demonstrated that ^{1}H-NMR is an effective method to identify ginseng metabolites. This provides useful information of quality control of raw material for TCM companies.

HPLC method

Sanqi is difficult to be analysed by PCA using HPLC fingerprint since the main UV absorption of Sanqi extract is at 205 nm which is very close to the detection limit of the DAD detector. The baseline of the HPLC chromatogram leans down with the increase of composition of acetonitrile in the HPLC mobile phase. The baseline is related to the pressure of the system, and the pressure of the system is related to the viscosity of mobile phase. Acetonitrile has a low viscosity. Therefore, an increase of acetonitrile in mobile phase results in a decreased system pressure, thus leading to a decreased baseline. It is very difficult to use PCA to handle the tilted chromatogram,

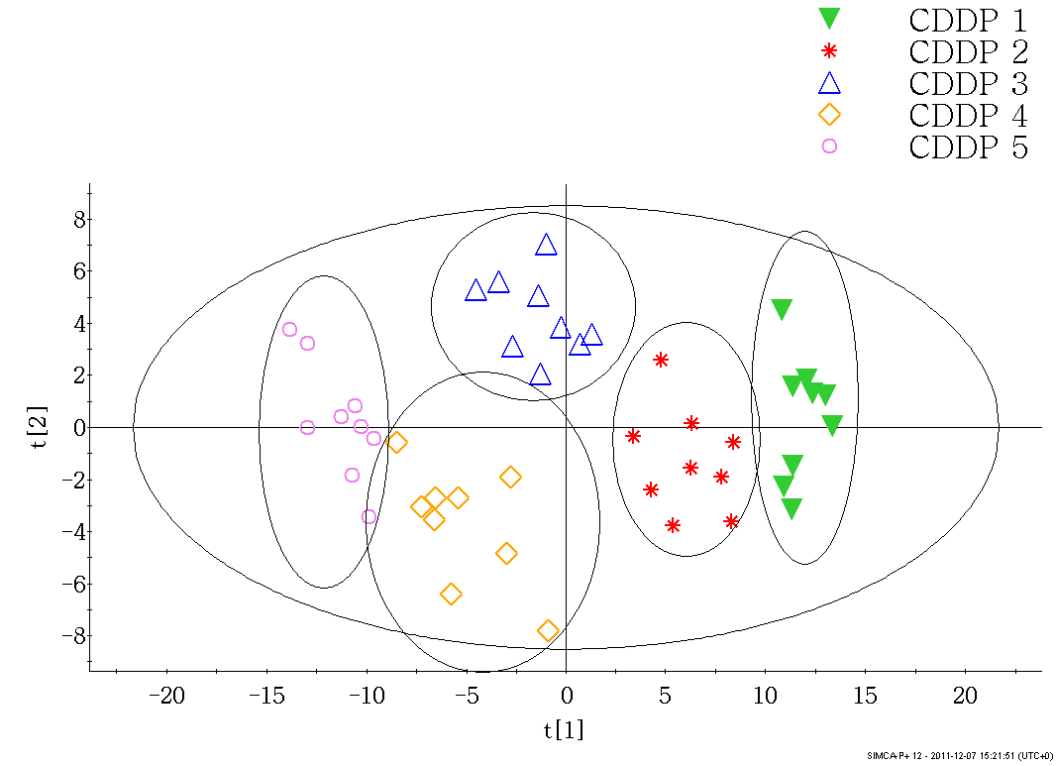


Figure 2.20: *Scores plot PC1(t[1]) versus PC2(t[2]) obtained from HPLC dataset of CDDP produced by Tasly with different batch numbers. Batch number: CDDP 1: 090312, CDDP 2: 090310, CDDP 3: 090311, CDDP 4: 090313, CDDP 5: 090314. Details of the production dates in Table 2.2.*

even if baseline subtraction was performed. Maybe it is possible to adjust the baseline manually before PCA, but it is time consuming. ELSD was demonstrated in previous research [93] that it is feasible to detect Sanqi metabolites but it is not available in our lab. Therefore HPLC was not used to study Sanqi.

2.3.8 Differentiation of different batches of CDDP

HPLC method

5 batches of CDDP were produced in one week by Tasly. Figure 2.21 is the PCA scores plot. However, still differences in composition between these

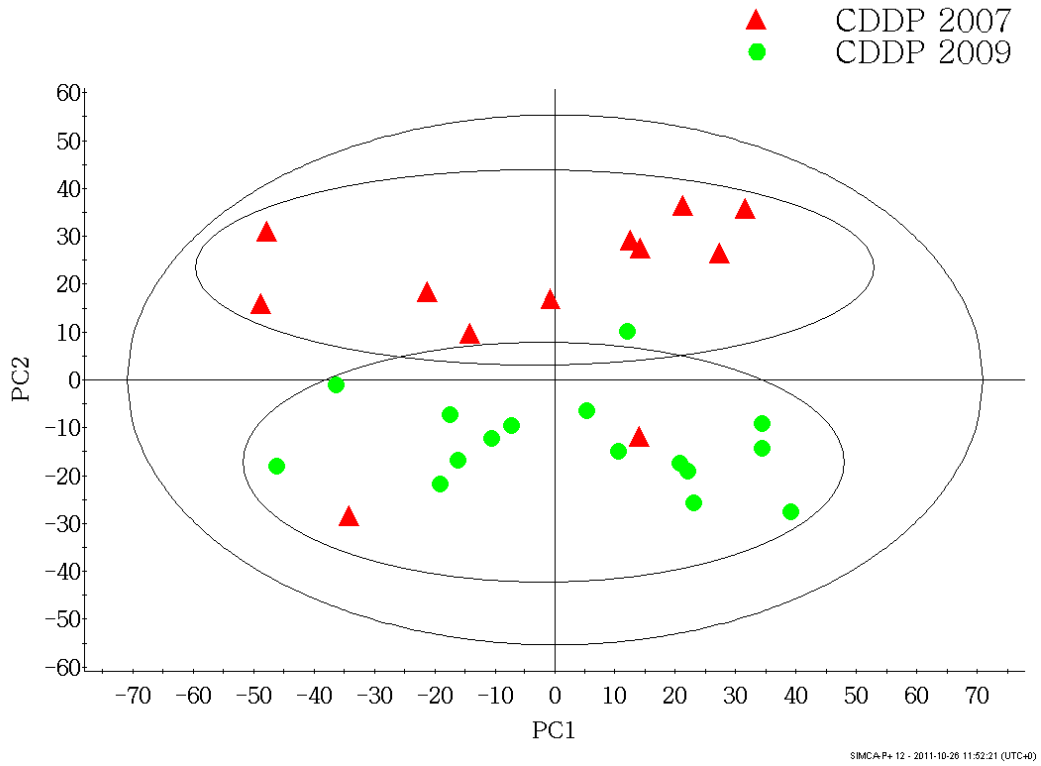


Figure 2.21: *Scores plot PC1 versus PC2 obtained from ATR-IR spectroscopic dataset of CDDP produced in different years. Red triangle: CDDP produced in 2009 including CDDP 1 – 5, green dot: CDDP produced in 2007 including CDDP 6 – 9. Details of the production dates in Table 2.2*

batches are evident. PC1 is positively contributed by 18.65 min which is due to SAB, negatively contributed by 19.48 min. PC1 is positively contributed by 12.75 min which is due to protocatechuic aldehyde. Therefore, CDDP batch 3 has a higher content of protocatechuic aldehyde than the rest of bathes. Batch 4 and 5 have a higher content of SAB. Batch 1 and 2 have a higher content of a compound with a retention time of 19.48 min.

IR spectroscopy

5 batches of CDDP produced in 2009 and 4 batches produced in 2007 can be distinguished by IR-PCA method. Figure 2.21 is the PCA scores plot of

9 batches of CDDP. PC2 is the main component in the PCA to distinguish CDDP produced in different years. The CDDP produced in 2007 contributed mainly positively to PC2 and the CDDP produced in 2009 contributed mainly negatively to PC2. The following wave numbers were found to contribute the main differences between CDDP that were produced in different years: 751, 1713, 1692, 1672, 1652, 1649, 1460, 1436, 1407, 1391, 1386, 1340 cm^{-1}. These wave numbers coincide with the wave numbers that distinguish Sanqi with different supply dates. These could mean the difference of CDDPs produced in 2007 and 2009 mainly due to the different of the Sanqi ingredient. It is possibly because they have changed the production methods or sites of the Sanqi product in the years between 2007 and 2009, although the exact information was not provided by the company. This could be a potential method to trace the quality from finished products to the raw plant.

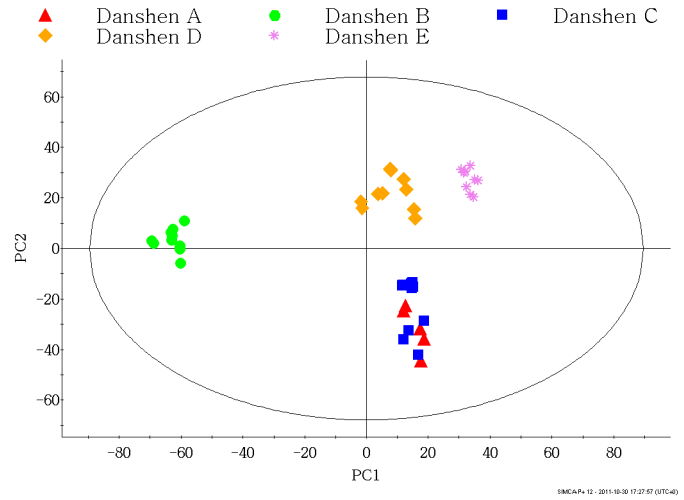
A 2D IR might be performed in order to correlate these wave numbers to compounds.

NMR spectroscopy

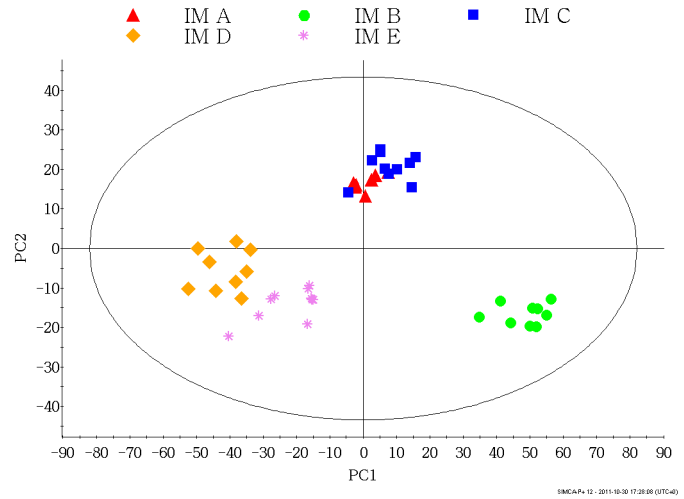
The NMR-PCA of 5 batches of CDDP produced in 2009 showed no difference between them. This might be because ^{1}H-NMR mainly detects primary metabolites (more explanation in section 2.1.4) which are largely present in the extract of herbal remedies, and the primary metabolites present in the CDDPs are quite similar. This may be also because NMR is intrinsically less sensitive than HPLC.

2.3.9 Comparison of analytical approaches for raw material and finished products

The weight to weight concentrations of salvianolic acid A and protocatechuic aldehyde is much higher in CDDP than in Danshen root. In this study 70% aqueous methanol is used. In the industrial precess, water is used. Therefore the polar compounds such as salvianolic acid A and protocatechuic aldehyde are better extracted in industry than in this study.



(a) PLS-DA scores plot of HPLC dataset of 5 Danshen A–E



(b) PLS-DA scores plot of HPLC dataset of 5 intermediate products A–E

Figure 2.22: *Correlated raw material-Danshen and intermediate material intermediate products A–E made from Danshen A–E and Sanqi A-E respectively.*

Figure 2.22 shows correlated Danshen and the intermediate product. Intermediate product A (IM A) was made from Danshen A and Sanqi. In the upper scores plot, Danshen batched A and C are overlapping, D and E are next to each other, B is on its own. The same grouping trend is shown in the corresponding intermediate product. The intermediate products were prepared following the same method as in the industry. This indicates that the difference between raw materials can possibly be carried to the intermediate product, and likely to the final product. Therefore, the control of the raw material is very important.

2.3.10 Overall Comparison of HPLC, NMR and IR spectroscopy for Danshen and Sanqi product

Table 2.4: *Ability of distinguishing minor difference of TCM samples. Abbreviations: l/o: lyophilization/evaporation. Correlation: Correlation from raw to intermediate products. * 4 hours is based on individual sample lyophilising time in small quantities. In reality samples subjected to NMR spectroscopic analysis were lyophilysed for 24 hours with many batches together.*

Products	HPLC	NMR	ATR-IR
Danshen	Good	Difficult	Difficult
Sanqi	Difficult	Good	Good
CDDP	Good	Good	Good
Correlation to metabolites	Good	difficult	N/A
Method Evaluation	Discrete	Non-discrete	Non-discrete
Peak Assignment	Easy	Difficult	Very difficult
Sample recycle	l/o	l/o	direct use
Sample preparation time	50 min	4 hours*	10 min
Measuring time	60 min	5 min	2 min
Sensitivity	Good	Ok	OK

HPLC and NMR, ATR-IR coupled with PCA have the ability to distinguish between minor differences of most Danshen and Sanqi products that are studied in this chapter. HPLC could distinguish the differences between region-to-region difference of Danshen, different batches of CDDP produced

in the same year, and correlate from raw material to intermediate material. However, Sanqi with different supply date could not be distinguished by HPLC-DAD. Ginsenosides mainly absorb at 205 nm which makes the HPLC analysis very difficult.[94] Other detectors such as Evaporative Light Scattering Detector (ELSD) might be used to detect ginsenosides[94].

NMR spectroscopic analysis could distinguish differences between Sanqi with different supply dates. Although good resolution spectra were obtained for all the samples studied, Danshen region to region and CDDP batch-to-batch differences are not able to be seen by NMR-PCA method.

ATR-IR spectroscopy is faster than HPLC and NMR spectroscopy. ATR-IR does not require complex sample preprocessing procedures. Ground plant powder could be measured directly. 12 minutes or less is usually enough for one sample using ATR-IR spectroscopy, from product to spectrum. TCM products need to be extracted in solvents before HPLC and NMR spectroscopic analysis. 50 minutes was needed to prepare the extraction. In practice, 6 or 9 samples were usually handled as one group. For the measurement time, at least 45 minutes is required for a HPLC chromatogram of most TCM products, and 5 minutes is usually enough to obtain a NMR spectrum with good resolution.

It is easy to correlate signals to compounds in HPLC because standard compounds can be used to match the HPLC peaks. ^{1}H-NMR spectroscopic is non-selective in plant study since proton exists in most of plant metabolites.

Because the secondary metabolites are less abundant than the primary metabolites in plants, and NMR spectroscopy is intrinsically insensitive, it is not usually used to study the secondary metabolites, but the primary metabolites.

2.4 Conclusion

In this chapter various TCM materials from raw to finished products have been analysed using HPLC-DAD, ATR-IR and ^{1}H-NMR spectroscopy with PCA data analysis.

A sensitivity test was carried out on all the fingerprinting methods. Iden-

tical protocatechuic aldehyde-spiked Danshen extractions were used for NMR and HPLC-PCA study. HPLC was found to be more sensitive than NMR, and can distinguish samples with a >5% protocatechuic aldehyde difference.

Sample preparation and measuring method for each technique were optimised. A comparison of solid sample method and extraction evaporation method of ATR-IR was made. The solid sample method was chosen as the more easy and reliable one. The ATR-IR evaporation method is not recommended because: 1. It is time consuming. 2. It does not give a better resolution for the spectra than the solid sample method, 3. There is a risk of minor component losses during extraction.

Danshen from different regions can be distinguished by HPLC, ATR-IR, but not NMR spectroscopy. This is the first time that ATR-IR is reported to discriminate geographical different plant to the author's best knowledge.

Sanqi with different supply date but from the same region can be distinguished by ATR-IR and NMR spectroscopy, but not HPLC. CDDP with different batch numbers can be distinguished by HPLC, but not ATR-IR and NMR spectroscopy. CDDP produced in different years (2007 and 2009) can be distinguished by ATR-IR spectroscopy. This was not tested using other techniques. Different brands of compound Danshen products were successfully distinguished by PCA by both HPLC and NMR methods. This was not tested using ATR-IR spectroscopy. Intermediate products were not tested by ATR-IR due to their hard and sticky nature.

Differences between batches can be distinguished in raw plant material, intermediate products and finished commercial products. Correlated batches cluster similarly in Danshen raw and intermediate products. Geographical different Danshen can be distinguished.

HPLC and NMR, ATR-IR spectroscopy can be used as complementary methods depending on the nature of the compounds concerned. The advantages and disadvantages of each method has been discussed. HPLC was found to be able to distinguish small differences between samples. As a powerful separation technique, the identification of chemicals is relatively easy. As a brief study, pure chemical standards were used to determine some HPLC peaks. To verify the identification, HPLC is often coupled with MS

and NMR. In the TCM industry, HPLC provides a quick and easy solution for profiling TCM metabolites, allows the user to obtain a chromatogram which is easy to interpret. However, as a discrete method, HPLC can not detect all the metabolites. In the future world of quality control of TCM, which emphasise the synergy effect of all the metabolites, HPLC might lose its primary position.

^{1}H-NMR was found to be a powerful non-discrete method, to reveal all the protons present in a mixture of TCM extract. This fits the synergy concept of TCM, and might be a potential method for quality control of TCM. The assignments of NMR peaks is not straight forward, but are possible with the assistance of 2D-NMR such as correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY) *etc.* ^{1}H-NMR is fast and reliable and can be used as a high-throughput methods for pharmacies and industries to perform TCM quality check.

ATR-IR is not used as much as the above methods in the quality control of TCM, because the spectra are almost impossible to interpret. However, this study demonstrates its ability to discriminate very similar TCM samples. ATR-IR is small and portable, which is an advantage over HPLC and ^{1}H-NMR. It is also solvent free that TCM powder can be measured directly. With these benefits, ATR-IR can be carried to TCM planting field, to perform a fast and high-throughput checking.

PCA is proved to be a powerful multivariate statistical analysis, which reveals differences between similar samples that otherwise difficult to be determined. It was found especially useful when coupled with spectroscopic methods. PCA also helps to visualise the differences of HPLC chromatograms of very similar TCM samples[83].

Chapter 3

Metabolomic quality control of commercial ginseng products using nuclear magnetic resonance spectroscopy

3.1 Introduction

In this chapter, ^{1}H-NMR metabolomics was used to catalogue variations in the spectra of 45 batches of commercially available ginseng. Ginseng samples used in this study were purchased from five different regions in China and include 4 different species, and possibly some counterfeit products.

This chapter starts by introducing some background knowledge about ginseng, including its history, use, morphology and chemical composition. This is followed by an introduction of the existing quality control methods of ginseng, and the method used in the current thesis.

3.1.1 Ginseng species

The term "ginseng" is usually used to indicate the root of *Panax* species. The genus *Panax* consists of 12 species including Asian ginseng (*Panax ginseng* C. A. Meyer), American ginseng ((*Panax quinquefolius* L.) and Sanqi (*Panax*

notoginseng). Asian and American ginseng are the two *Panax* species mainly used in TCM. *Panax* belongs to the family Araliaceae. The genus *Panax* is a relic of the Tertiary Period dating back to 0.3 billion years ago. The root is the part used in traditional medicine.

3.1.2 History of use

The use of Asian ginseng dates back 4,000 years. Its first use was probably mystical, because the shape of ginseng looks human. It was first recorded in "Shennong's Chinese Materia Medica" (*Shen Nong Ben Cao Jing*), considered the oldest book on traditional Chinese medicine, compiled approximately during the Latter *Han* Dynasty (ca. 25-220 A.D.). [95] Asian ginseng was recorded with other 119 herbs, as "the first treatise", which were used for their "stimulating properties", and stated to be harmless to humans. The function of Asian ginseng was recorded as "supplement for five main internal organs" and "mental-tranquillization"[47].

Till the 16th century, Asian ginseng gained great popularity. In *Ben Cao Gang Mu* ("Compendium of Materia Medica"; ca.1593), regarded as the most complete and comprehensive medical book written in the history of TCM, the use of Asian ginseng in treating 31 medical conditions including sweats, weakness, and nausea, was described comprehensively in respect of dose, combination with other TCM, and preparation methods[96]. Asian ginseng was almost seen as an omnipotent herb. Due to its popularity, although strictly controlled by several Emperors, the wild Asian ginseng became almost extinct in China.

This passion for Asian ginseng from the Chinese influenced French missionaries in the 17th century. A similar looking ginseng was soon found by them in north America, which was used by the native Americans as a healthy food. This was later known as American ginseng. It was first documented in "New compilation of Materia Medica" (*Bencao Congxin*) by Yiluo Wu in 1757[97]. American ginseng was found to have a similar effect to Asian ginseng. In fact, the chemical composition of American ginseng is similar to Asian ginseng[98].

American ginseng quickly gained popularity in China. In the late 19^{th} century, about 282 tons of wild American ginseng was imported to China[99]. As might have been expected, wild American ginseng decreased rapidly in north America as well.

Due to the complex conditions for its successful cultivation, it was not until 1885 that George Stanton, who is often called today "The father of American Ginseng", successfully planted 150 acres of American ginseng. Since then, American ginseng is successfully cultivated in the US, Canada, Australia and China.

Ginseng has the effect of decreasing fatigue. Possible mechanisms of action of ginseng is linked to antioxidant[100], anti-inflammatory, anti-tumour[101] and immunostimulatory activities[17]. American ginseng stimulates insulin production and has an anti-hyperglycaemic effect[17].

3.1.3 Morphology characters of ginseng samples

The shape of different ginseng samples is shown in figure 3.1. WAG32 is an example of wild American ginseng with an age older than 15 years according to the retailer. CAG 2 and CAG 5 are examples of cylindrical and conical shaped cultivated American ginseng with an age of around 4-5 years. ASG is an example of Asian ginseng. The shape of Asian ginseng is long and slim with an untrimmed beard. American ginseng is usually trimmed. The shape of the cylinder shaped American ginseng is very similar to Asian ginseng.

There are minor differences between wild and cultivated American ginseng samples. The head of the wild American ginseng is smaller compared to cultivated American ginseng. There are circles that develop from the head and all the way down the body of the roots. The colour of wild American ginseng is usually dark whereas the colour of cultivated American ginseng is lighter and the skin is more smooth and with even scars.

Sample 20 (figure 3.1) was purchased from a less renowned small ginseng store on the famous Chinese medicine retailing area called "Shanghuan", where it was claimed to be cultivated American ginseng grown in Wisconsin, USA, and it was supplied with so claimed certificate of authenticity. The gin-

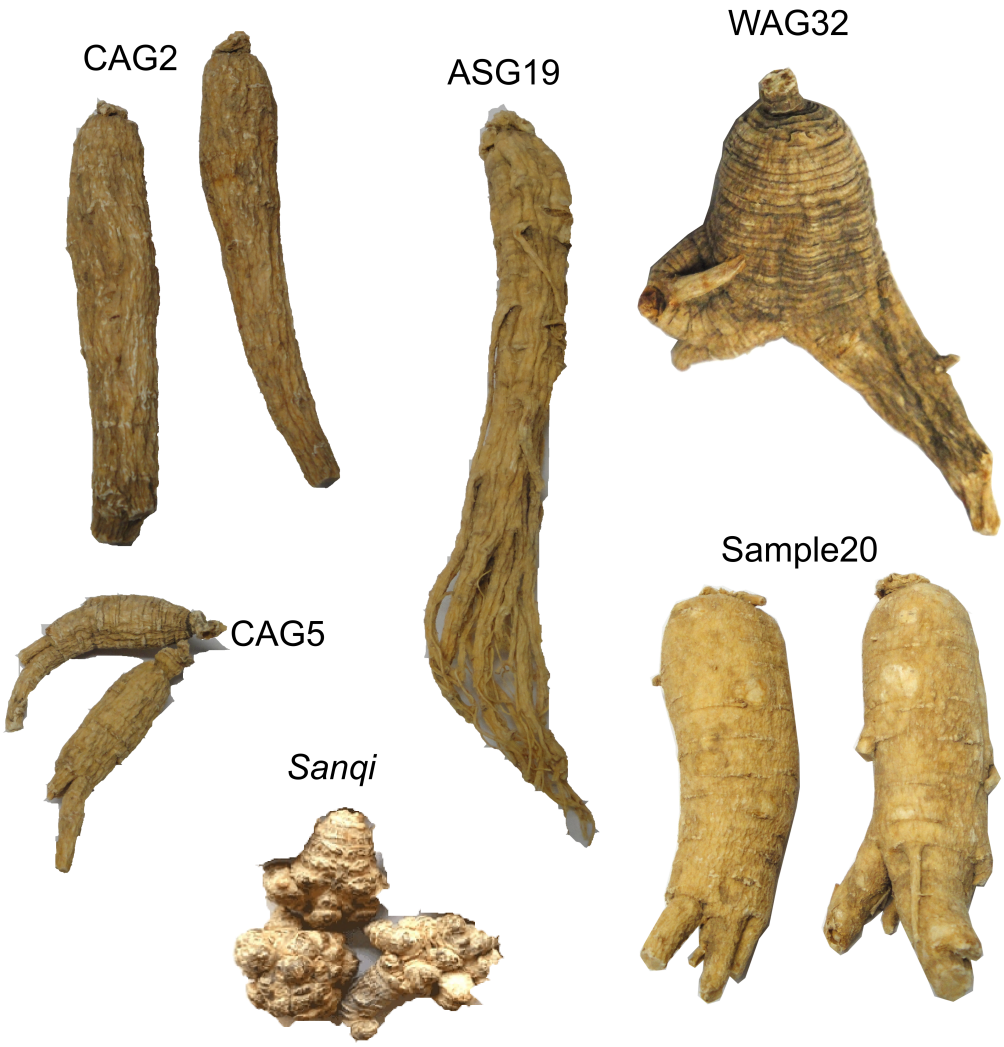


Figure 3.1: *Representative photos of 6 ginseng samples. The numbers in the names indicate the experimental numbers used in the present study in table 3.2. WAG: Wild American ginseng root. CAG2: an example of long shaped cultivated American ginseng root. CAG5: an example of short shaped cultivated American ginseng root. ASG: Asian ginseng root. Sample20: a claimed cultivated American ginseng batch but with distinct morphological characters. Photo taken by the author.*

seng roots were stored in large baskets and trimmed by staff in the store. The claimed American ginseng roots are about 5 cm long and 3 cm in diameter. As seen in figure 3.1, it is much bigger, less wrinkled, lighter in colour and cheaper compared to American ginseng samples purchased from well reputed stores. This sample was suspected to be a counterfeit product and this was confirmed in this study (in section 3.3.6).

In the markets, there is a huge price variation among the different grades of American ginseng. The price is generally in the trend as follows: wild American ginseng > cultivated American ginseng > Asian ginseng > Sanqi. In China, wild American ginseng can be as expensive as £20 per gram while the cheaper cultivated American ginseng can be £1 per gram. Cheaper ginseng was sometimes used to adulterate or sold as a higher price ginseng, especially when the samples were sliced, it is more difficult to identify its morphological characters. The huge profit margin of wild American ginseng drives shops to adulterate cultivated American ginseng using Asian ginseng or other plants since their shapes appear very similar to inexperienced customers. The look of Sanqi is very different from American and Asian ginseng, however, the composition and medical use are quite similar[55]. Sanqi was sometimes used to adulterate American and Asian ginseng.

3.1.4 Constituents of ginseng in general

The constituents of ginseng root include amino acids, fatty acids, carbohydrates, alkaloids, triterpenoid saponins, polysaccharides, sesquiterpenes, polyacetylene, peptidoglycans, minor elements, vitamins, and phenolic compounds[102][11].

Most attention has been focused on a group of triterpenoid saponins, the ginsenosides, which have been associated with significant pharmacological effects, including antioxidant[87], anti-diabetic[88] and anti-tumor effects[89]. However, the exact therapeutic mechanisms were not clear.

More than 30 ginsenosides have been identified so far. The most common ones are listed in figure 3.2 and table 3.1. Ginsenosides are divided into 2 groups according to their aglycone moieties: the protopanaxadiol group:

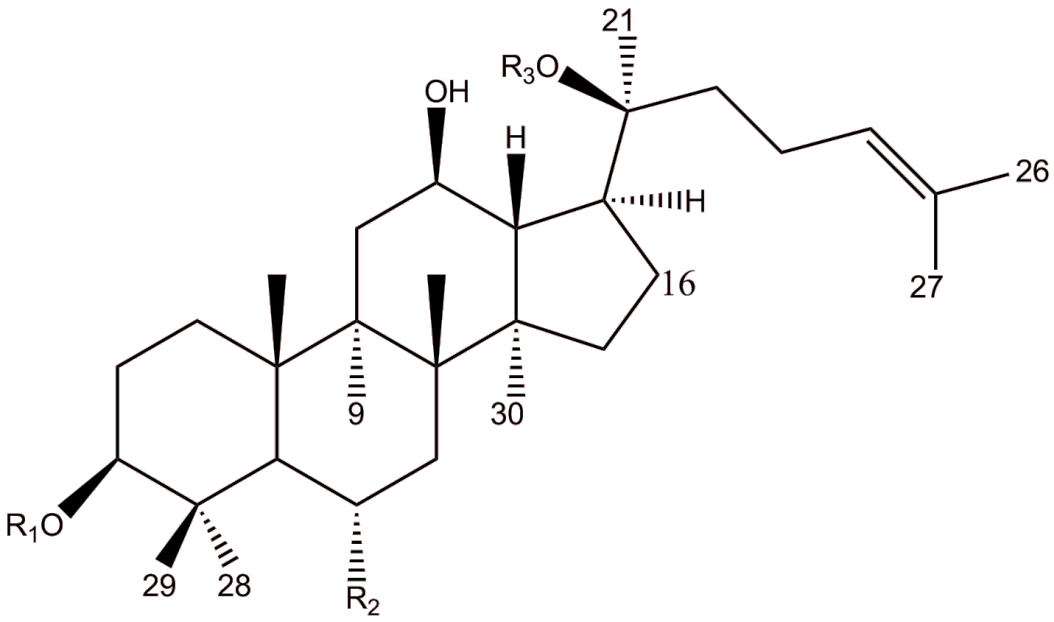


Figure 3.2: *A general structure of ginsenosides. Values for R_1, R_2 and R_3 of major ginsenosides are listed in table 3.1*

Rb_1, Rb_2, Rc, Rd, and the protopanaxatriol group: for example Re, Rf, Rg_1, Rg_2)[103]. A general structure of ginsenoside is illustrates in figure 3.2. The most abundant saponins in ginseng are Rb_1, Rb_2, Rc, Rd, Re and Rg_1.

Other studies showed that ginseng polysaccharides have various important biological activities [17] such as antioxidant [100], anti-tumor, anti-adhesive, immunomodulating[101] and hypoglycemic activities.

3.1.5 Current regulation methods

Many ginseng products in the market are in the form of slices or powder in capsules, which makes identification using morphological characters impossible. Even when the ginseng is not sliced or ground, the identification might be difficult since some of the cheap products (lower age, cultivated, fake) have a similar look to the expensive ones. The identification using morphological characters requires considerable expertise in the field and it is not always practical. Adulteration and faulty products could cause unexpected

Table 3.1: *Chemical structures of some major ginsenosides referring to figure 3.2. Glc = glucose, Ara(p) = alpha-L arabinopyranose, Ara(f) = alpha-L arabinofuranose. Rha = rhamnose.*

ginsenoside	R_1	R_2	R_3
protopanaxadiols			
Rb_1	-Glc^2-Glc	-H	-Glc^6-Glc
Rb_2	-Glc^2-Glc	-H	-Glc^6-Ara(p)
Rc	-Glc^2-Glc	-H	-Glc^6-Ara(f)
Rd	-Glc^2-Glc	-H	-Glc
protopanaxatriols			
Re	-H	-Glc^2-Rha	-Glc
Rf	-H	-Glc^2-Glc	-H
Rg_1	-H	-Glc	-Glc
Rg_2	-H	-Glc^2-Rha	-H

side effects. Therefore the need of developing an easy and effective quality control method is urgent.

In order to regulate the market, the Chinese Pharmacopoeia defines the morphological characters and a few simple tests including identification and quantification of a few secondary metabolites using TLC and HPLC[45]. The total amount of ginsenoside Rg_1, Re and Rb_1 should not be less than 0.5% of the dried Asian ginseng (Rg_1 and Re should be higher than 0.3% in total, Rb_1 should be higher than 0.2%). The analogous requirement is 2.0% for American ginseng. There is no quantification requirement for other potentially effective metabolites.

The test methods by Chinese pharmacopoeia are not adequate to distinguish all the counterfeit products[104].

3.1.6 Advanced quality control methods

Because of the inadequacy of the Chinese Pharmacopoeia, some reports have suggested using advanced quality assessment method on ginseng products including HPLC, NMR and IR [103][16][11][105].

HPLC is the most widely used and popular method due to its simplicity, sensitivity and accuracy. The resulting data is easy to be interpreted since

one compounds present in the form of one peak. The detection power of HPLC was made stronger with various detectors such as UV, DAD, and ELSD[103].

HPLC with different detectors such as UV[106] and ELSD[103] have been used in these methods to simultaneously quantify a dozen of ginsenosides. 19 ginsenosides were quantified in black ginseng developed from Asian ginseng using HPLC-ELSD[107].

However, there are limitations of these methods. Firstly, HPLC methods are selective. Saponins mostly absorb at 203 nm, which is close to the detection limit of DAD detectors. Many solvents and even oxygen also absorb at this wavelength thus increase the baseline noise and lower the sensitivity of detection. The baseline is poor at 203 nm especially when using a gradient elution method. The baseline is related to the pressure of the system, and the pressure of the system is related to the viscosity of mobile phase. Acetonitrile has a low viscosity. Therefore, an increase of acetonitrile in mobile phase results in a decreased system pressure, thus leading to a decreased baseline.

Secondly, it is not sufficient to quantify only ginsenosides because they are not the only active compounds existing in ginseng. However, despite the limitations of HPLC, it is still the most used method in quality control of TCM using selected biomarkers.

NMR spectroscopy became increasingly popular in quality assessment of herbal medicines[15][16]. Most of the reports on quality assessment of TCM by ^{1}H NMR metabolite fingerprinting analysed the primary metabolites, because primary metabolites are more abundant in plants than secondary metabolites, and NMR spectroscopy intrinsically lacks high sensitivity.

PCA or PLS methods have often been used to analyse datasets of NMR spectra. Because the NMR spectra of plant extracts are very complex with many overlapping peaks. PCA is thus useful to summarise and find the differences between groups. Therefore it is often used with NMR spectroscopy on quality assessment of similar herbal medicines[15][16]..

There are several reports on the metabolite compositions of various ginseng types. Lee *et al.* analysed the primary metabolite composition of Asian

ginseng samples aged 2 to 6 years using NMR-PCA method.[16]. However, most of the reports were not representative since only a small number of samples from each species were analysed[38][16].

3.1.7 Aim of study

The aim of study in this chapter is 1. to investigate the suitability of ^{1}H-NMR-PCA on distinguishing Asian ginseng, cultivated and wild American ginseng. 2. to investigate the differences in primary and secondary metabolites between Asian ginseng, cultivated and wild American ginseng. 3. to study the differences in between wild American ginseng samples. 4. to distinguish a possibly fake CAG product sample 20.

3.2 Materials and methods

3.2.1 *Panax* samples studied

37 batches of crude commercial ginseng products were purchased from five different sources in China including HongKong, Beijing, Taiyuan, Hangzhou and Tangshan. These products include 15 batches of cultivated American ginseng, 11 batches of wild American ginseng, 5 batches of Asian ginseng (all cultivated), 5 batchs of sanqi and 1 batch claimed to be cultivated American ginseng, with a clear morphological difference from other cultivated American ginseng samples. A typical commercial ginseng product from each group is shown in figure 3.1. Except for the wild American ginseng, all the other samples were cultivated. A detailed sample list is specified in table 3.2.

The grading criteria of ginseng is objective depending on different companies. Usually the older ginseng, the better. However, exact criteria in companies is unknown.

As 90% of wild American ginseng samples from the US was exported to HongKong, and the samples used in this research were purchased from the most prestige TCM companies in HongKong, together with the knowledge of the author on identified the appearance of wild American ginseng, it is very

likely that the wild American ginseng samples used in this study are genuine as stated.

Among them, sample 18 was purchased from a less noted ginseng store, the average size of this batch was around 2-3 cm, the morphology features mostly matched the features of the wild American ginseng, although some of them look like cultivated American ginseng. Sample 20 was purchased from a store in "Shanghuan" in Hong Kong, the morphological features is distinctively different from all the other ginseng purchased (figure 3.1).

3.2.2 Chemicals used

Ginsenoside Re, Rg_1, Rb_1, Rg_2, Rc, Rd were purchased from Carbosynth (Berkshire, UK).

All the sugar samples including sucrose, glucose, lactose, fructose and deuterium oxide (D_2O, 661643, %99.99 atom % D, contains 1% DSS-d_6) were purchased from Sigma (Sigma Co., St Louis, MO, USA). Other deuterated solvents and 5 mm NMR tubes (Wilmad 507-PP) were purchased from Goss Scientific (Goss Scientific Instruments, Great Baddow, Essex, UK).

3.2.3 Ginseng roots extraction method

Ginseng dried root material was ground using a coffee grinder and filtered through a 350 μm sieve. Large pieces which did not go through were re-ground, passed though the same sieve and mixed with the rest of the powder. 2 g of the powder was extracted with 100 ml 70% methanol in a volumetric flask, using a bath sonicator for 15 minutes at 30°C followed by shaking manually and sonicating for a further 15 min. After cooling to room temperature, evaporated solvent was replaced with 70% methanol and the volume of the solvent was made up to 100 ml. The supernatant was filtered with a 0.2 μm syringe filter.

The filtered solvent was evaporated under reduced pressure for 30 minutes to remove most of the methanol. The residual solution was frozen in dry ice then lyophilized for 24 hours.

Table 3.2: *Details of 37 batches of ginseng studied in this chapter. P.F.: purchased from. COP:country of production. CAG: cultivated American ginseng. ASG: Asian ginseng. WAG: wild American ginseng. TFH: Tung Fong Hung Ltd, HongKong. EYS: Eu Yan Sang, HongKong. BJ: Beijing. TRT: Tong Ren Tang Group Co., Ltd. Anhui WS: Anhui Wan Sheng pharmaceutics. FHCT: Fang Hui Chun Tang, Hangzhou £/37.8 g: Price per 37.8 g (Chinese weight unit "Liang") in British Pounds. Cy.: cylindrical shaped. Co.: conical shaped. P. no.: production number. Blank area means information not available.*

No	Type	P.F.	COP	Grade	£/37.8 g	P. no.
1	CAG	TRT,BJ	Ca.	Co.,5	9.1	20090547
2	CAG	TRT,BJ	Ca.	Cy., 1	7.5	20090448
4	CAG	TRT,BJ	Ca.	Co.	8.7	20100101
5	CAG	TRT,BJ	BJ	Sliced	8.7	100107
6	CAG	TRT,BJ	BJ	Co.	11.2	706069
8	CAG	TFH	US		20	142418
9	CAG	TFH	US		16	141617
10	CAG	TFH	US		12	140414
11	CAG	TFH	US		9	142714
12	CAG	TFH	US		6	142614
13	CAG	TFH	US		2.5	140718
28	CAG	TRT,LD	US			
15	CAG	NPH	US		16	40603
16	CAG	NPH	US		4.8	41101
21	CAG	FHCT				
19	ASG	TRT,BJ	China		1.9	
22	ASG	TRT,SX	China		9.3	101201
23	ASG	TRT,JL	China		2.1	263274
24	ASG	Anhui WS	China			
25	ASG	Shanxi PS				
7	WAG	TFH	US		159	
14	WAG	NPH	US		130	40101
17	WAG	EYS	US		120	11115
18	WAG	TakTai	US			
29	WAG	TFH	US	4	132	149924
30	WAG	TFH	US	3	199	138910
31	WAG	TFH	US	1	373	139824
32	WAG	TFH	US	1	398	
33	WAG	EYS	US	Cy.	95	11115
34	WAG	EYS	US	4	252	11050
35	WAG	EYS	US	1	410	11155
20	unknown	HongKong				
41-45	Sanqi	Yunnan province				08/2009

Table 3.3: *Extra information for ginseng sample 1-6. TRT: Tong Ren Tang Group Co., Ltd. JL: Jilin Ginseng Co., Ltd. GS: Ginseng Co., Ltd. BXLB: Benxi Longbao group gingseng Co., Ltd. BJTH: Huairou, Beijing Tianhui ginseng industrial Co., Ltd. BJCG: Beijing Chongguang medical Co., Ltd.*

No	Type	Production Company	Manufacture date	Expiry date
1	CAG	TRT, JL	20090511	20120510
2	CAG	TRT, GS	20090415	20120414
4	CAG	BXLB	20100102	20120101
5	CAG	BJTH	20100112	201212
6	CAG	BJCG	20080817	20130817

3.2.4 NMR measurement

D_2O, [D_4]-MeOD, and DMSO were attempted to reconstitute the lyophilised samples, and D_2O was found the best solvent. Concentrations of 5, 10, 15, and 30 mg/ml were tested and the best spectra were found to be produced at a concentration of 10 mg/ml. The spectra of 15, and 30 mg/ml contain broad peaks, which were probably due to undissolved particles. The spectra of 5 mg/ml were not as clear as those of 10 mg/ml.

One dimensional ^{1}H NMR spectra were measured at a temperature of 300K on a Bruker DRX 400 MHz Spectrometer (Bruker, Coventry, UK) equipped with a 5 mm triple resonance inverse detection (TXI) probe. Spectra were the result of the summation of 16 free induction decays (FIDs), with data collected into 65k data points and a sweep width of 20 ppm. Acquisition parameters were 0.126 Hz/point, pulse width was 17.857 kHz 90 and relaxation delay was 1.0 s. Prior to Fourier transformation, an exponential line broadening equivalent to 0.3 Hz was applied to the FIDs. The signal intensities for all samples were referenced to DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) at 0.00 ppm.

3.2.5 Data analysis

Data pretreatment

The NMR spectra of all the ginseng samples were bucketed in 0.002 ppm, instead of 0.04 ppm which was used in most of the previous reports using

NMR-PCA method to do quality assessment of herbal medicine[16]. The small bucket size allows details to be revealed and provides more information of small peaks. Bucketing of spectra was performed using MestReNova® (MNova) software version 5.2.3. The NMR spectra were aligned using LineUp software (Infometrix), and treated with vector length normalization (refer to section 2.2.7) using Pirouette® software (Infometrix).

The water peak at 5.07-4.6 ppm, DSS peak at 0-0.07 ppm and methanol peaks at 3.3-3.4 ppm were excluded before normalisation (section 2.2.7), pareto scaling(section 2.2.7), and PCA.

PCA

PCA is a multivariate data analysis method to summarise massive variables in a dataset into a few correlated variables. In this study each bucketed ppm value is considered to be one variable. There are 2,270 variables in each spectrum. In the results and discussion section, some selected spectroscopic regions were deleted during the analysis as the study progressed and became more focused.

PCA, normalization and pareto scaling were performed using Pirouette® software version 4.5. The terms "PC 1" and "factor 1" are used interchangeably.

3.3 Results and discussion

3.3.1 NMR spectra of 4 ginseng samples

Typical NMR spectra for ginseng samples including Sanqi root, cultivated American ginseng root, Asian ginseng root and wild American ginseng root at 0.5-6.5 ppm are shown in figure 3.3.

There are a few differences that can be observed by eye without statistical analysis. Apart from the water and methanol peaks which were excluded during analysis, the most noticeable responses in these chemical shifts are due to sugar, the peaks of which occur in the regions of 3.5-4.5 ppm and 5.0-5.5 ppm. Among these, sucrose largely dominates in this 70% aqueous

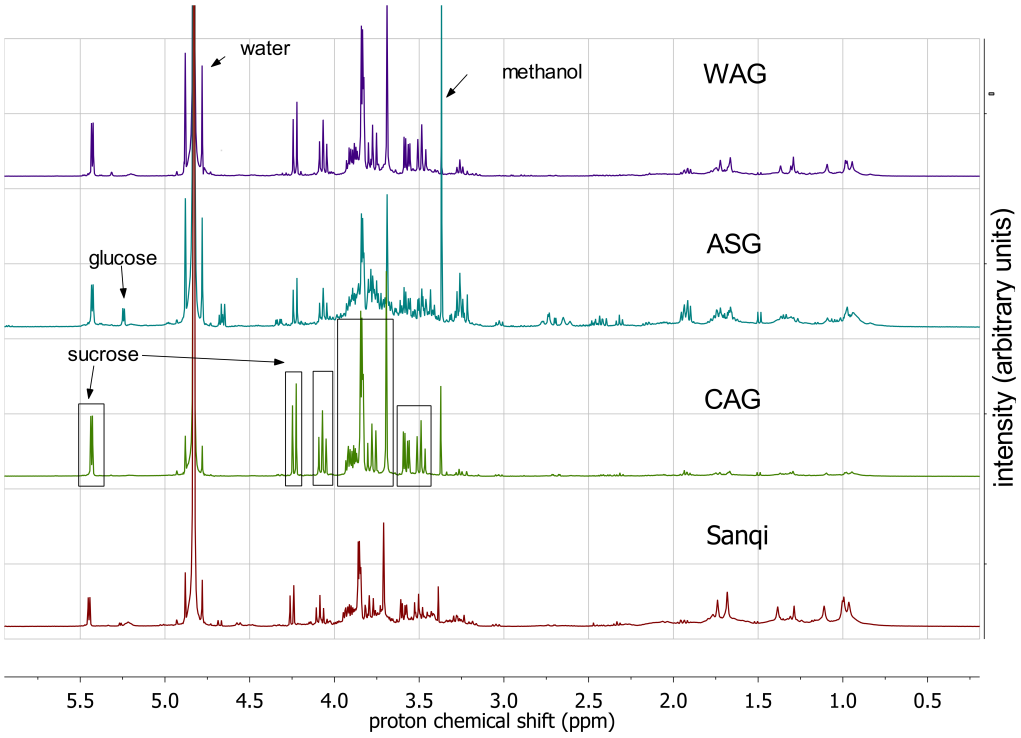


Figure 3.3: *Representative NMR spectra of extractions of 4 ginseng types. 0.5-6.0 ppm parts were shown as where the most metabolites presents. CAG: cultivated American ginseng batch 4 . ASG: Asian ginseng batch 24. WAG: wild American ginseng batch 35. Peaks marked in boxes in CAG belong to sucrose. Detailed assignment of sucrose is given in table 3.4.*

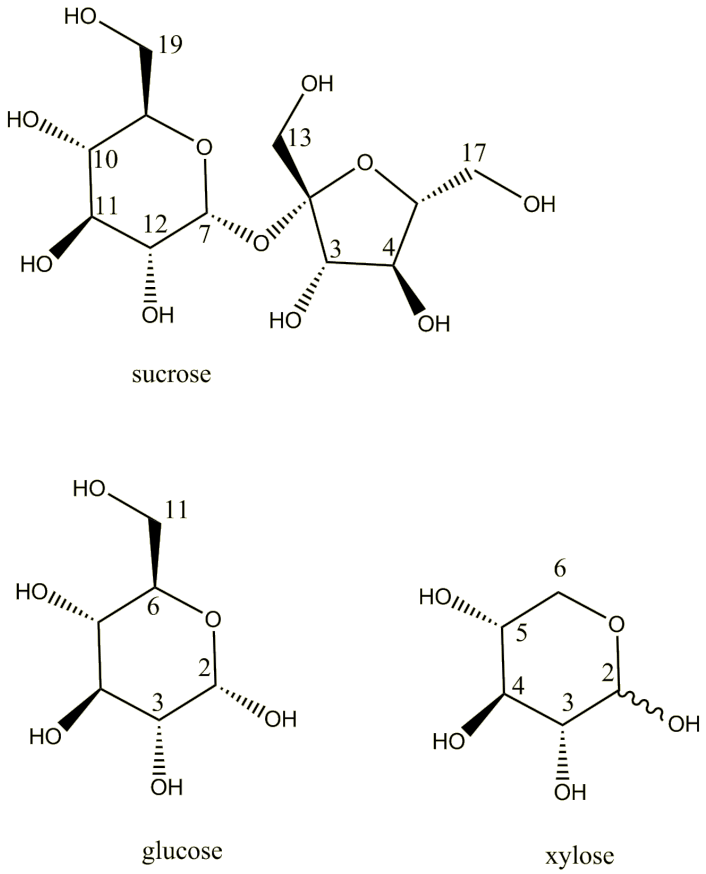


Figure 3.4: *Structures of sucrose, glucose and xylose. The* ^{1}*H-NMR assignments are listed in table 3.4.*

methanol extract of ginseng samples. The sucrose regions were marked using boxes in figure 3.3. Sucrose is the most abundant metabolite in the extract of all the ginseng samples studied in this chapter, except wild American ginseng sample 32 and 29, and Asian ginseng 24, where glucose is the most abundant metabolite. This is further discussed in the following subsection. Apart from the sugar regions, there are less abundant peaks from 1.0 to 3.5 ppm and from 4.0-8.5 ppm. These were mainly due to primary metabolites and these are assigned in the next section.

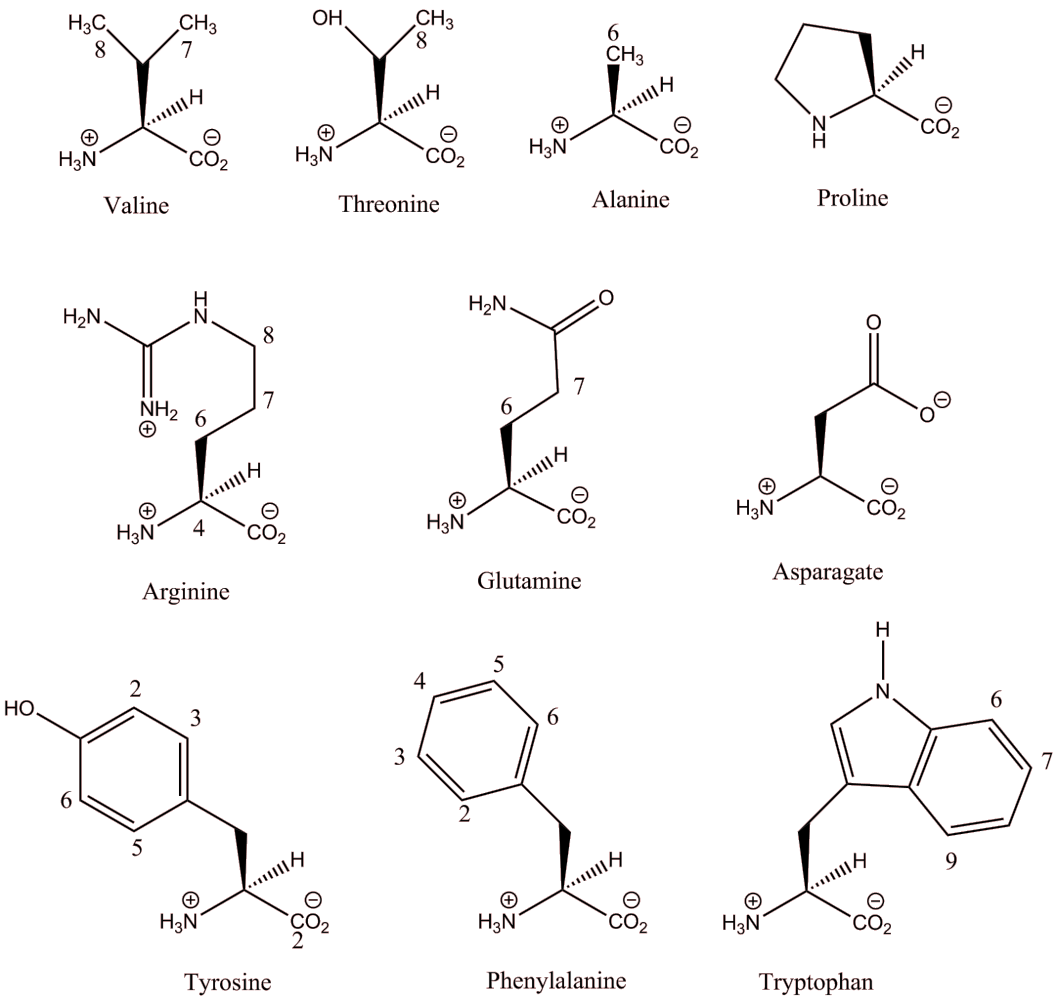


Figure 3.5: *Structures of assigned amino acids in the* ^{1}*H-NMR spectra of ginseng extract. The* ^{1}*H-NMR assignments are listed in table 3.5. The protons were numbered according to the Human Metabolome Database* `http://www.hmdb.ca/`.

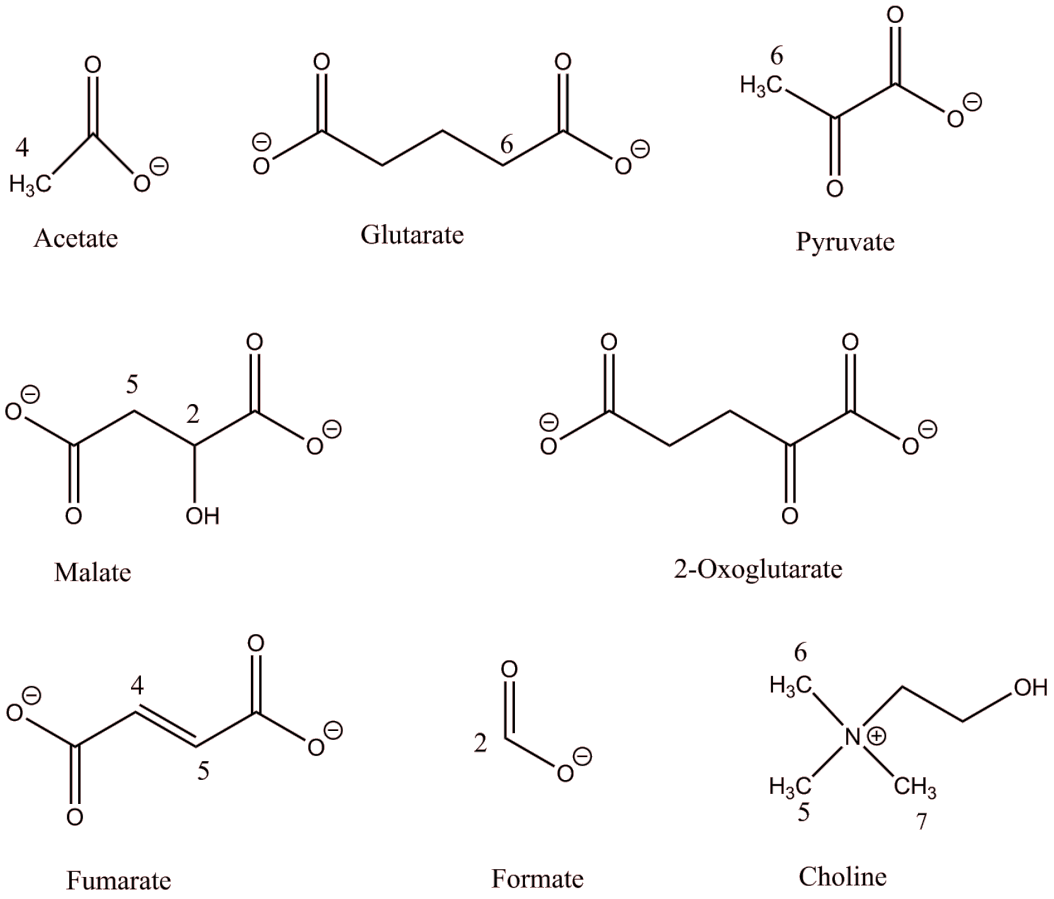


Figure 3.6: *Structures of assigned organic acids in the ^{1}H-NMR spectra of ginseng extract. The ^{1}H-NMR assignments are listed in table 3.5. The protons were numbered according to the Human Metabolome Database* `http://www.hmdb.ca/`.

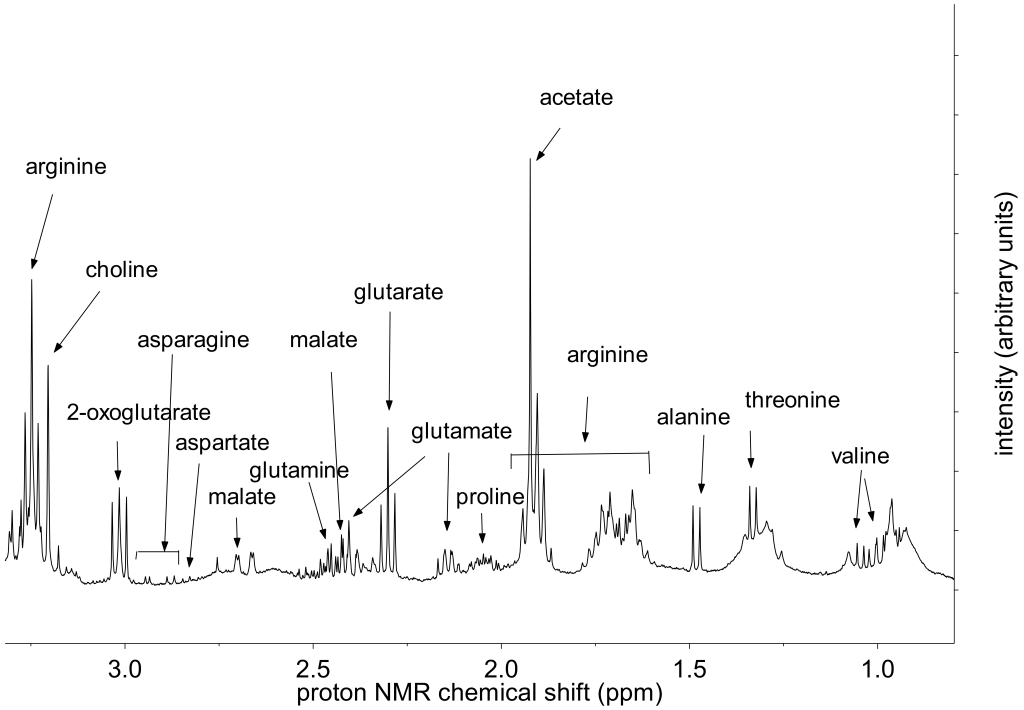


Figure 3.7: *Assignment of primary metabolites in the ^{1}H-NMR spectrum of Asian ginseng sample no.19, from 0.9 to 3.3 ppm. The sample was extracted using 70% aqueous methanol and dissolved in D_2O. Detailed assignments of primary metabolites in this region are listed in table 3.4.*

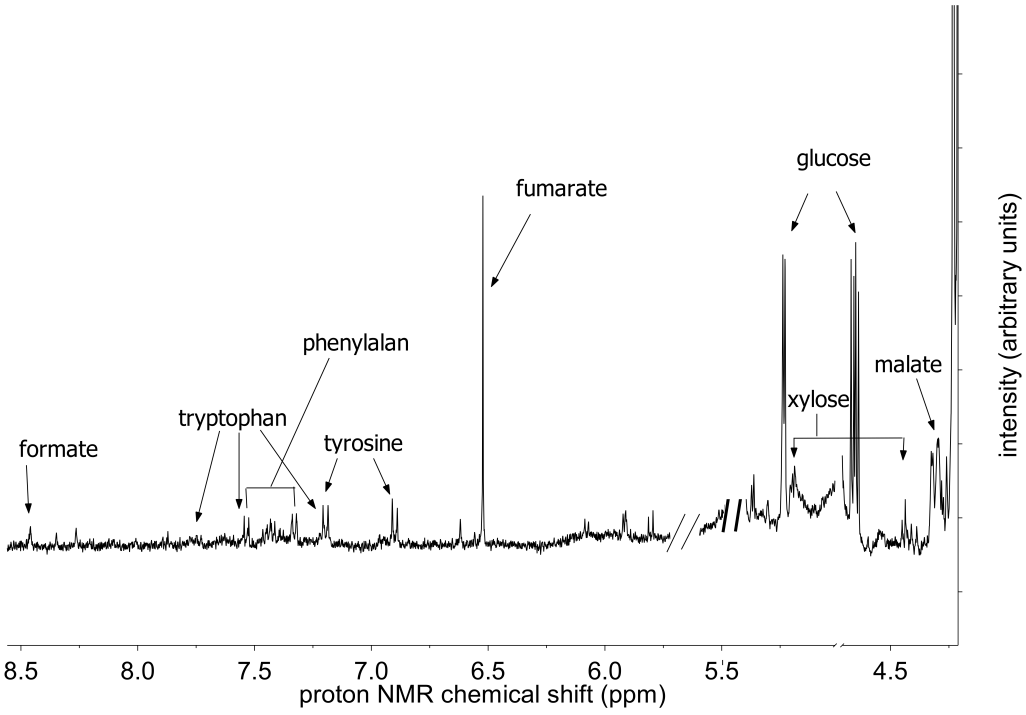


Figure 3.8: *Assignment of primary metabolites in the ^{1}H-NMR spectrum of Asian ginseng sample no.19, from 4.2 to 8.5 ppm. The sample was extracted using 70% aqueous methanol and dissolved in D_2O. The water peak (4.7 to 5.1 ppm) was deleted from the spectrum. Detailed assignments of primary metabolites in this region are listed in table 3.5.*

Table 3.4: *Assignment of sucrose, glucose and xylose- on ^{1}H-NMR spectrum of a representative Asian ginseng sample (No. 19), extracted using 70% aqueous methanol and dissolved in D_2O. s: singlet. d: doublet. dd: double doublets. t: triplet. J: J-coupling constants. Assignment of H see figure 3.4.*

Metabolite	^{1}H	Assignment
Sucrose	3.47(t,J=9.6)	H-10
	3.56(dd,J=10.0, 3.9)	H-12
	3.66(s)	H-13
	3.76(t,J=9.5)	H-11
	3.79-3.91(m)	H-17,19
	4.05(t,J=8.6)	H-4
	4.22(d,J=8.8)	H-3
	5.42(d,J=3.9)	H-7
Glucose	3.45-3.48(m)	H-6
	3.56(dd,J=9.9, 3.8)	H-3
	3.72-3.91(m)	H-11,6
	5.22(d,J=3.8)	H-2
Xylose	3.38(t), J=9.4)	H-4
	3.534(d, J=8.1)	H-2
	5.18(m), J=3.9)	H-2

3.3.2 Assignment of peaks

The assignments were performed by comparison to reference spectra obtained using standard compounds, assignments in previous publications[11][108][109], Human Metabolome Database (HMDB,`http://www.hmdb.ca/`), and the Biological Magnetic Resonance Data Bank (`http://www.bmrb.wisc.edu/metabolomics/`). J coupling constants were matches with references during assignment.

The assignments are listed in table 3.4 and 3.5, and illustrated in figure 3.7 and figure 3.8. Peaks in 1.0-3.5 ppm and 4.5-8.5 ppm were assigned to 20 primary metabolites, including 2 sugars (structures shown in figure 3.4),10 amino acids (structures shown in figure 3.5), and 7 organic acids and 1 vitamin (structures shown in figure 3.6).

A first impression of an American ginseng spectrum is that it was dominated by sucrose signals. This was confirmed by adding 3 mg of sucrose into 10 mg of Asian ginseng extract.

Table 3.5: *Assignment of 18 primary metabolites on ^{1}H-NMR spectrum of a representative Asian ginseng sample (No. 19), extracted using 70% aqueous methanol and dissolved in D_2O. The assignment of protons are marked on the molecules in figure 3.5 and figure 3.6. The chemical shift values of some metabolites were not assigned to a specific proton. s: singlet. d: doublet. dd: double doublets. t: triplet. J: J-coupling constants (Hz).*

No	Metabolite	δ ^{1}H (p.p.m.)	Assignment
Amino acids			
	Valine	0.98(d,J=7.0)	H-8
		1.02(d,J=7.0)	H-7
	Threonine	1.34(d,J=6.5)	H-8
	Alanine	1.46(d,J=6.5)	H-6
	Arginine	1.61-1.76(m)	H-7
		1.87-1.96(m)	H-6
		3.26(t,J=6.9)	H-8
		3.768(t,J=6.11)	H-4
	Proline	1.99-2.09(m)	H-3,4
	Glutamate	2.10-2.17(m)	H-6
		2.34-2.37(m)	H-7
	Aspartate	2.82(dd,J=17.1, 3.7)	H-6
	Tyrosine	6.89-6.91(m)	H-2,6
		7.19-7.20(m)	H-3,5
	Tryptophan	7.21-7.25(m)	H-9
		7.54(d,J=8.0)	H-6
		7.74(d,J=7.9)	H-7
	Phenylalanine	7.34(d,J=7.6)	H-2,6
		7.34-7.40(m)	H-4
		7.41-7.44(m)	H-5,3
Organic acids			
	Acetate	1.90(s)	H-4
	Glutarate	2.27(t, J=7.46)	H-4,6
	Pyruvate	2.38(s)	H-6
	Malate	2.40(dd,J=15.5, 9.9)	H-5
		2.74(dd,J=15.5, 3.1)	H-5
		4.31(dd,J=9.9, 3.1)	H-2
	2-Oxoglutarate	3.02(t,J=7.4)	H-5
	Fumarate	6.51(s)	H-4,5
	Formate	8.42(s)	H-2
Vitamins			
	Choline	3.18(s)	H-5,6,7

3.3.3 Differentiation of ginseng types based on their sugar contents

Figure 3.9 is the PCA scores plot of ^{1}H-NMR spectra of 4 ginseng samples, including wild and cultivated American ginseng root, Asian ginseng, sanqi ginseng and sample 20, which is a different looking root but claimed to be cultivated American ginseng. The differentiation of sample 20 is further discussed in the end of this chapter, in section 3.3.6. This PCA scores plot is formed by factor 1 and factor 2, which count for 52.5% and 16.0% of the total variance respectively. The third principal component factor 3 counts for 5.3% of the total variance. Each factor is perpendicular to all other factors. These 3 factors explain in total of 73.5% of the total variance of the samples.

The cultivated American ginseng samples are located on the top left, Asian ginseng samples are located mainly on the left side below cultivated American ginseng samples, except sample 24 are located on the right side, sanqi samples are located on the top right, and wild American ginseng samples are located mainly in between Asian and cultivated American ginseng samples, except sample 29 and 32 which are located on the bottom right of the scores plot(figure 3.9) .

Cultivated American ginseng, Asian ginseng and sanqi are well separated in the PCA scores plot, but wild American ginseng samples overlap with all the other samples. In order to further analyse the difference of those samples, the corresponding PCA loadings plots is analysed. Figure 3.10 is the loadings plot correlated to the PCA scores plot(figure 3.9). On the top left of the PCA loadings plot, the loadings are assigned to sucrose chemical shifts. On the bottom right of the PCA loadings plot, the loadings are assigned to glucose chemical shifts. The loadings directions of sucrose and glucose are opposite each other in the loadings plot. This means that if more sucrose is present in ginseng, then less glucose is present. It may be because that sucrose broke down into glucose in plant's metabolism[110]. Previous report by Lee *et al.* also found that cultivated American ginseng has a higher content of sucrose than ASG, but Asian ginseng has a higher content of glucose than CAG[16], based on three ginseng samples. Yang *et*

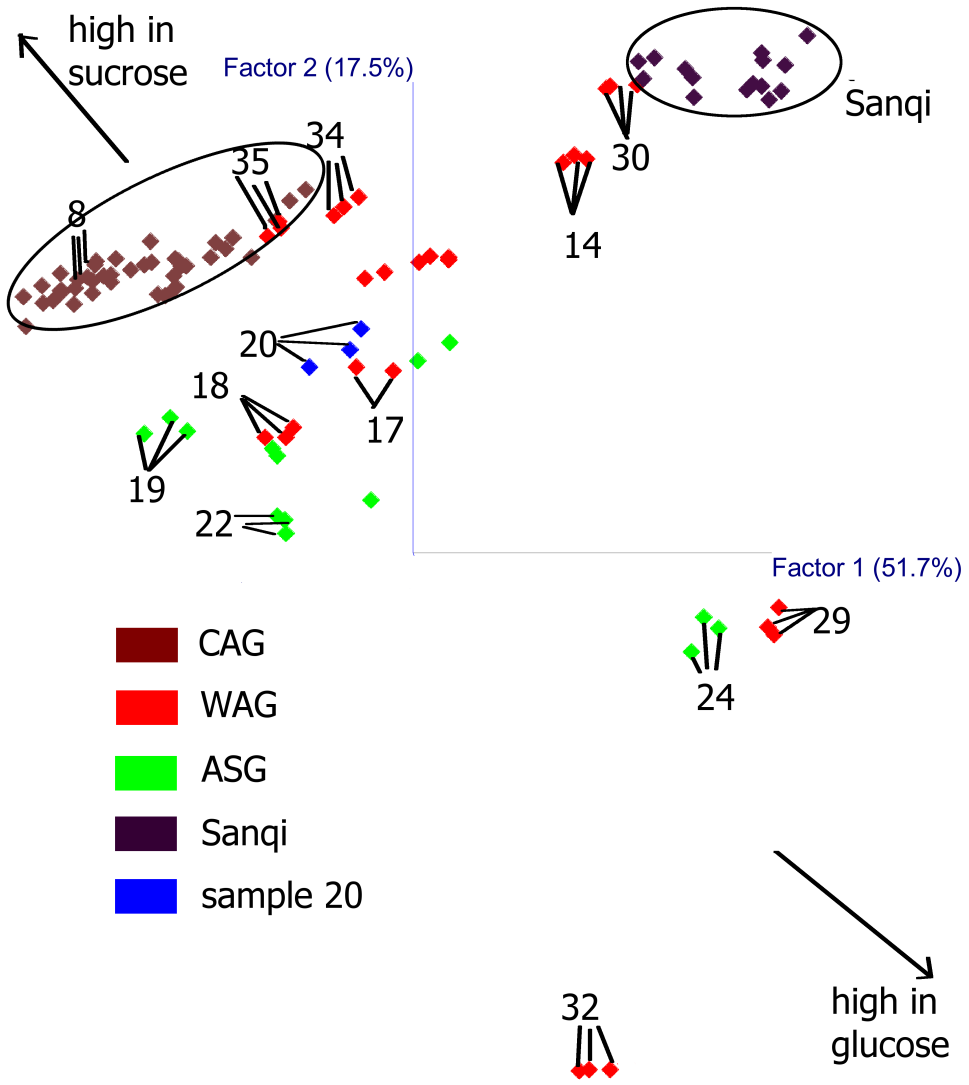


Figure 3.9: *PCA scores plot of* ^{1}*H-NMR spectra of 4 ginseng types including wild and cultivated American ginseng, Asian ginseng, sanqi ginseng and sample 20. Each colour represents a different group of samples. Abbreviations: ASG: Asian ginseng root (green spot). WAG: wild American ginseng root (red spot). CAG: cultivated American ginseng root (brown spot). Sanqi root(Purple spot). Sample 20: claimed to be CAG (Blue spot).*

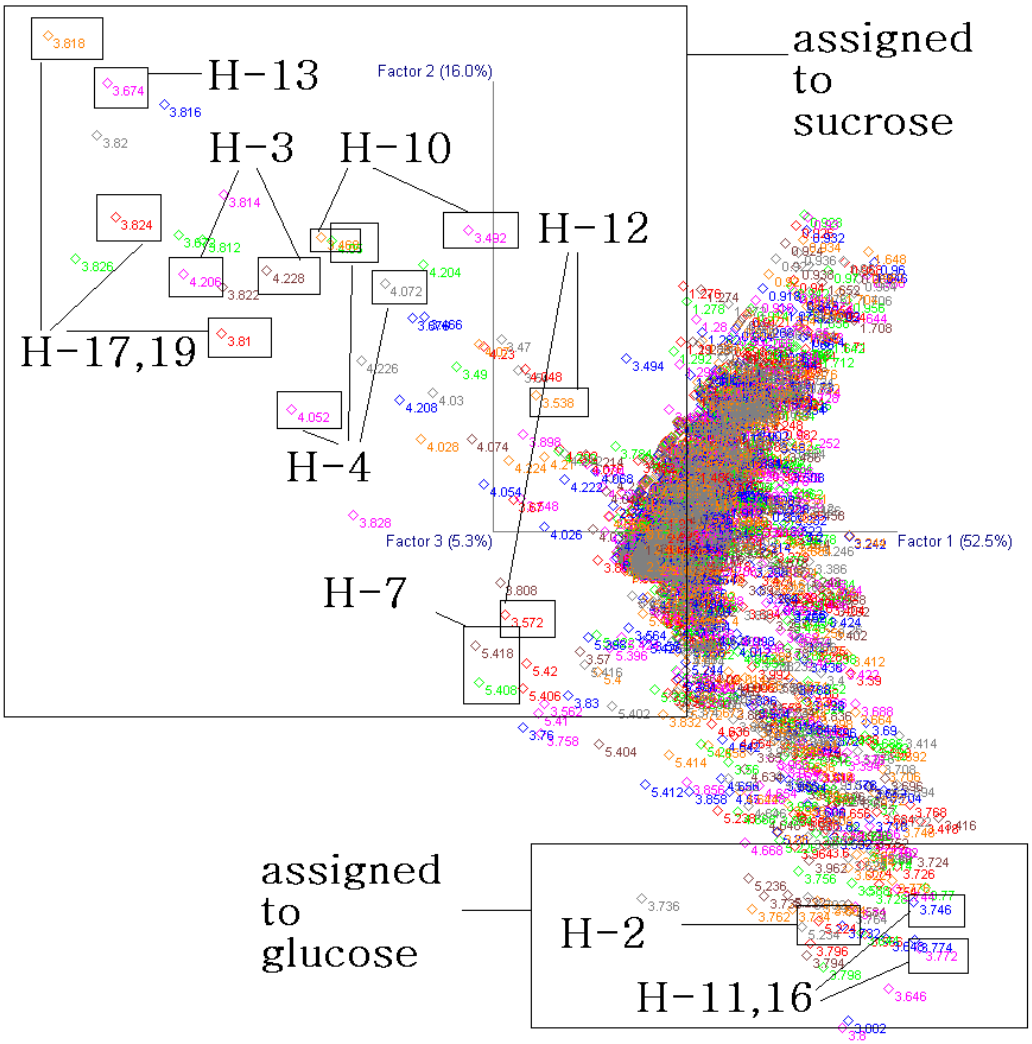


Figure 3.10: *PCA loadings plot of* ^{1}*H-NMR spectra of 4 ginseng samples. This correlates to the PCA scores plot of figure 3.9. Colours of the loadings are irrelevant to the analysis. On the top left side, the loadings were assigned to the main sucrose* ^{1}*H-NMR chemical shifts. On the bottom right side, the loadings were assigned to glucose* ^{1}*H-NMR chemical shifts. The details of the assignment are listed in table 3.4.*

al. reported the metabolite differences between cultivated American ginseng with different ages, sucrose was found in a higher level in age 4, 5 and 6 year old than 2, 3 year old ginseng. The levels of glucose are lower in age 1, 4, 5 and 6 year old than the 3 year old ones [11].

In ginsenosides, there are glucose attached to it (see table 3.1). However, the glucose signal was not contributed by the the sugar substituents in ginsenosides[30]. This was confirmed by Kang *et al.* using diffusion-ordered spectroscopy analysis[30]. The current study also shed a light on this, that many Asian ginseng samples are high in glucose, but have a similar level of ginsenosides comparing to CAG samples.

Along the direction of sucrose, which is to the top left of the PCA scores plot, the furthest top left group is cultivated American ginseng and wild American ginseng sample 34 and 35. This means these samples contain the highest percentage of sucrose, and lowest percentage of glucose among all the samples studied. The NMR spectra of ginseng samples suggest that only a small amount of glucose is present in cultivated American ginseng samples and wild American ginseng sample 34 and 35, some glucose is present in Asian ginseng samples. Conversely, more glucose but less sucrose are present in sample 24 (ASG), 29 and 32 (WAG).

Apart from the variables contributed by sucrose and glucose, most of the other variables are located close to the centre of the PCA loadings plot. This means that sucrose and glucose contents represent the largest difference among the four types of ginseng samples. The relative sucrose content among cultivated American ginseng samples is very similar, since the shape of its group stretched out almost perpendicular to the direction of sucrose loadings, as do the sanqi samples. The relative sucrose/glucose content of Asian ginseng and wild American ginseng varies from almost none to a high content within the group. This may be related to the growth condition of individual batch of samples. This is especially true for wild American ginseng, because their growth conditions are almost isolated. The growth conditions of cultivated American ginseng are controlled to meet the desired way, therefore it is understandable that it forms a tighter group in the PCA scores plot showing good similarity between the samples.

Details of the assignment of sucrose peaks are listed in table 3.4. 3.81, 3.818, 3.824 ppm peaks were assigned to H-17,19 of sucrose. These are the larger peaks of sucrose, therefore are weighted more in PCA, and were located further more from the centre in the PCA loadings plot. 5.408 ppm and 5.418 ppm peaks are assigned to H-7 of sucrose. These are smaller peaks in sucrose, therefore they were counted of less importance in PCA, so these loadings are located closer to the centre of the loadings plot. The rest of the sucrose peaks are located in the top right area between the centre and 3.818 ppm in the PCA loadings plot. This is because pareto scaling was used in this PCA: not all the variables were not given the same importance, therefore higher peaks counted more in the analysis. This is discussed in the method section.

The sucrose/glucose content may be used as a rough indicator to differentiate cultivated American ginseng and Asian ginseng. The sucrose/glucose content within 15 batches of cultivated American ginseng is very similar to each other , but not very similar within the 5 batches of Asian ginseng studied. However, the sucrose/glucose content in the 5 batches of Asian ginseng studied are all lower than that of the cultivated American ginseng. 8 out of 10 of the wild American ginseng studied in this chapter also have a high content of sucrose and a low content of glucose, but the other 2 wild American ginseng have shown a similar content of sucrose/glucose with the cultivated American ginseng. Therefore sucrose/glucose content can not be used to distinguish cultivated and wild American ginseng, or Asian and wild American ginseng.

Xylose is more abundant in Sanqi and wild American ginseng batch 14 and 30. However, levels of detected xylose are very low compared to other metabolites in ginseng samples; therefore it is not a main difference in PCA.

The sugar region shows the most intense signal in the NMR spectra of *Panax* species, and often hides signals caused by other compounds in the same region. Therefore, in some research on the quality assessment of Dangshen, the sugar was removed by water elution using a C-18 cartridge[23]. However, research indicates that polysaccharides in *Panax* species have various activities including anti-fatigue[111][17], and immunomodulating[101].

Therefore in the present study, sugar was not removed during preparation of samples for NMR spectroscopic study. It is also because this work is aiming to find an easy solution in quality control of ginseng samples with minimal preparation and modification steps. Although sugar was included during extraction process, its NMR spectroscopic region could be carefully excluded electronically to serve the purpose to highlight the differences caused by other compounds. In order to further study the differences between the *Panax* groups, the sucrose and glucose regions were excluded in future analysis.

3.3.4 Differentiation of *Panax* species following excision of the sugar region

For the analysis below, sugar regions from 3.33-4.94 ppm, 5.20-5.26 ppm, and 5.35-5.51 ppm were excluded before PCA.

In the PCA scores plot shown in figure 3.11, cultivated American ginseng, Asian ginseng, and Sanqi were well separated from each other and formed a tight plot within each group. Asian ginseng products represented by green spots are located in the top left, wild American ginseng samples (red spots) are located in the middle, cultivated American ginseng (brown spots) are located in the bottom left, and *Sanqi* samples (dark purple spots) are located in the bottom right of the PCA scores plot. The NMR spectra of wild American ginseng are spread out over a large region in the scores plot, indicating they are different from each other. This is possibly because the purchased wild American ginseng roots have different ages, growth conditions, gene type *etc.* The size of wild American ginseng is usually an indicator of the age, the price is often proportional with the size.

Even if the sugar region was deleted, the separation between 4 kinds of ginseng remains good. This means that sugars are not the only metabolites distinguishing ginseng samples.

Because Sanqi is distinctly different from Asian and American ginseng and highly influence the analysis, it was excluded in the next part of study and the differences of the remaining three ginseng types were further explored.

PCA scores plot in figure 3.12 is composed of PC1 and PC2 which count

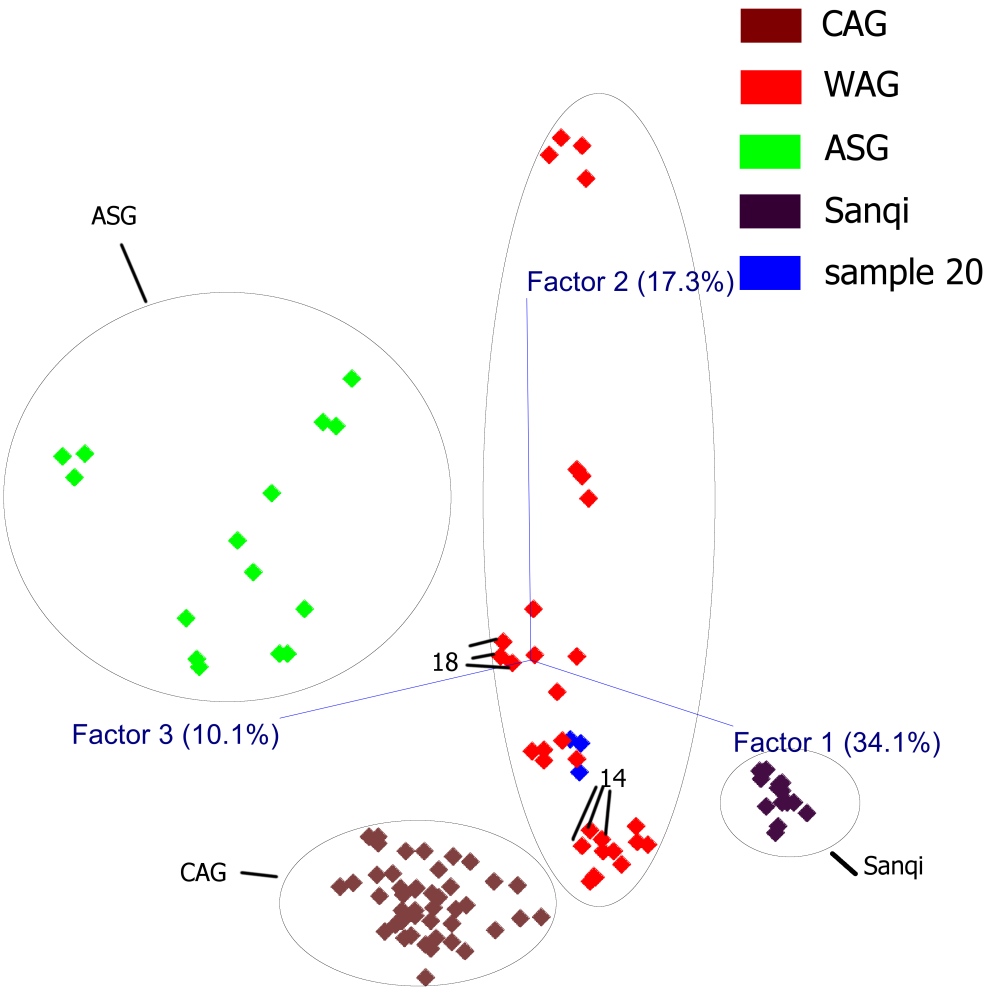


Figure 3.11: *PCA scores plot of 4 ginseng samples and sample 20. Each colour represents a different group of samples. The number written behind the coloured spots represents different batches of samples, which is correlated to table 3.2. "a", "b" or "c" after the batch numbers are repeated extracts of the same batch of sample. Dark purple spot: Sanqi root. Brown spot: Cultivated American ginseng root. Green spot: Asian ginseng root. Red spot: Wild American ginseng root. Blue spot: Sample 20.*

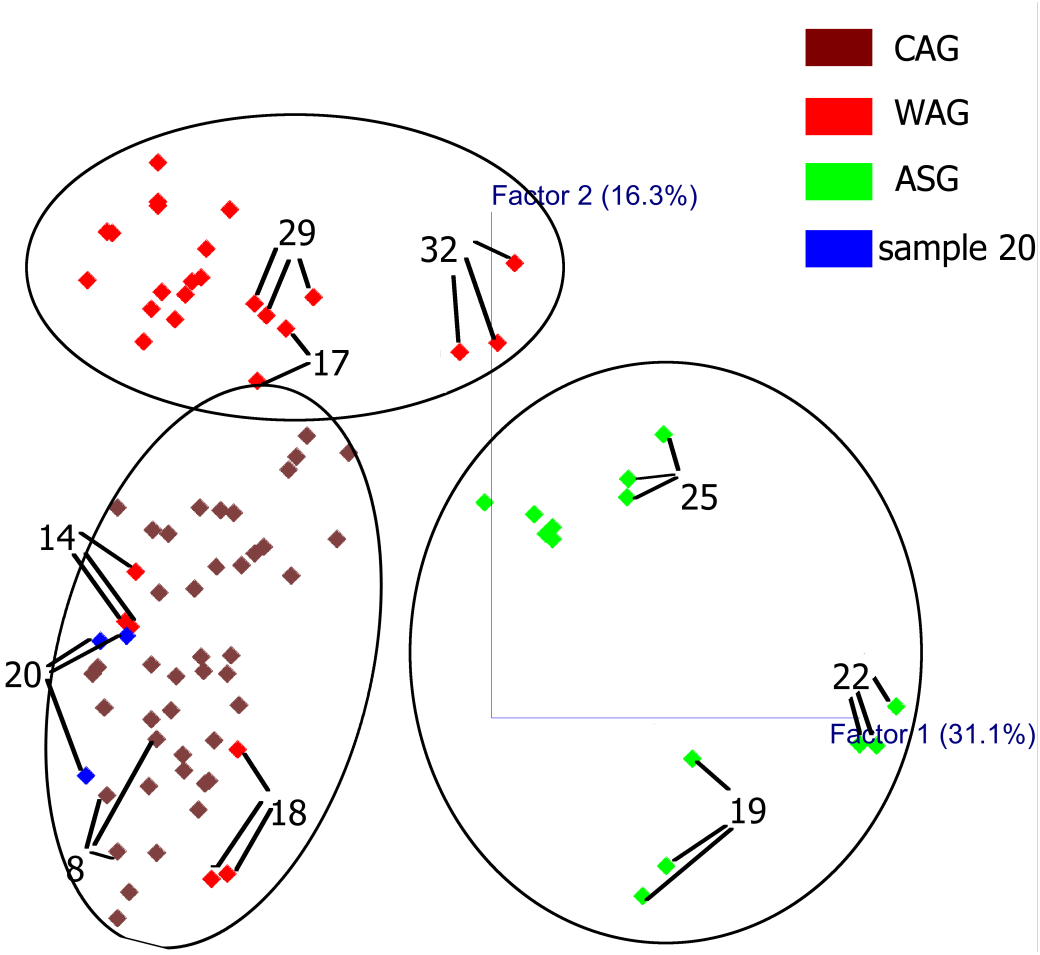


Figure 3.12: *PCA scores plot of 3* Panax *species. Each colour represents a different group of samples. The number written behind the coloured spots represents different batches of samples, which is correlated to table 3.2. Purple spot: Sanqi root. Brown spot: Cultivated American ginseng root. Green spot: Asian ginseng root. Red spot: Wild American ginseng root. Blue spot: Sample 20.*

for 31.1% and 16.3% of the total variance. American ginseng were mainly located on the left side of the PCA scores plot, except wild American ginseng sample 32. Cultivated American ginseng samples are all located on the left side of PC1 axis. Asian ginseng is located on the right side. This means that cultivated American ginseng and Asian ginseng were clearly separated by PC1 alone. Wild American ginseng samples 18 and 14 are grouped closely to the cultivated American ginseng samples. However, in the three dimensional view shown in figure 3.11, samples 14 and 18 are separated from CAG. This means that WAG samples 18 and 14 are different from the CAG samples.

As mentioned before, sample 18 was purchased from a less noted store and some of the ginseng roots in this sample looks like cultivated American ginseng. This may suggest wild American ginseng sample 18 was adulterated by cultivated American ginseng. This could also because the size of the root of sample 18 is relatively small comparing with other wild ginseng samples. This indicates a younger age. Therefore it might have a different composition of metabolites according to its different location in the scores plot.

Figure 3.13 is the corresponding PCA loadings plot of the PCA scores plot shown in figure 3.12. The variables in the loadings plot is packed therefore they were shown in coloured spots for easier visualisation. Areas marked L1 (figure 3.14), L2 (figure 3.16) and L3 (figure 3.17) are in the same location in the PCA loadings plot as ASG, WAG and CAG in the PCA scores plot. This means ASG has a higher relative content of the metabolite correspond to the variables indicated in L1. Variables with a high value on a principal component contribute more to that principal component. For example, variable 3.244 ppm is located on the most right side of the loadings plot, therefore it has the highest value on PC1. Also PC1 is the major difference among all the PCs, thus the variable 3.244 ppm contributes the most to the PCA scores separation. Therefore the metabolite with a chemical shift at 3.244 ppm is most responsible for the separation along PC1, in this case, the difference between Asian ginseng and American ginseng.

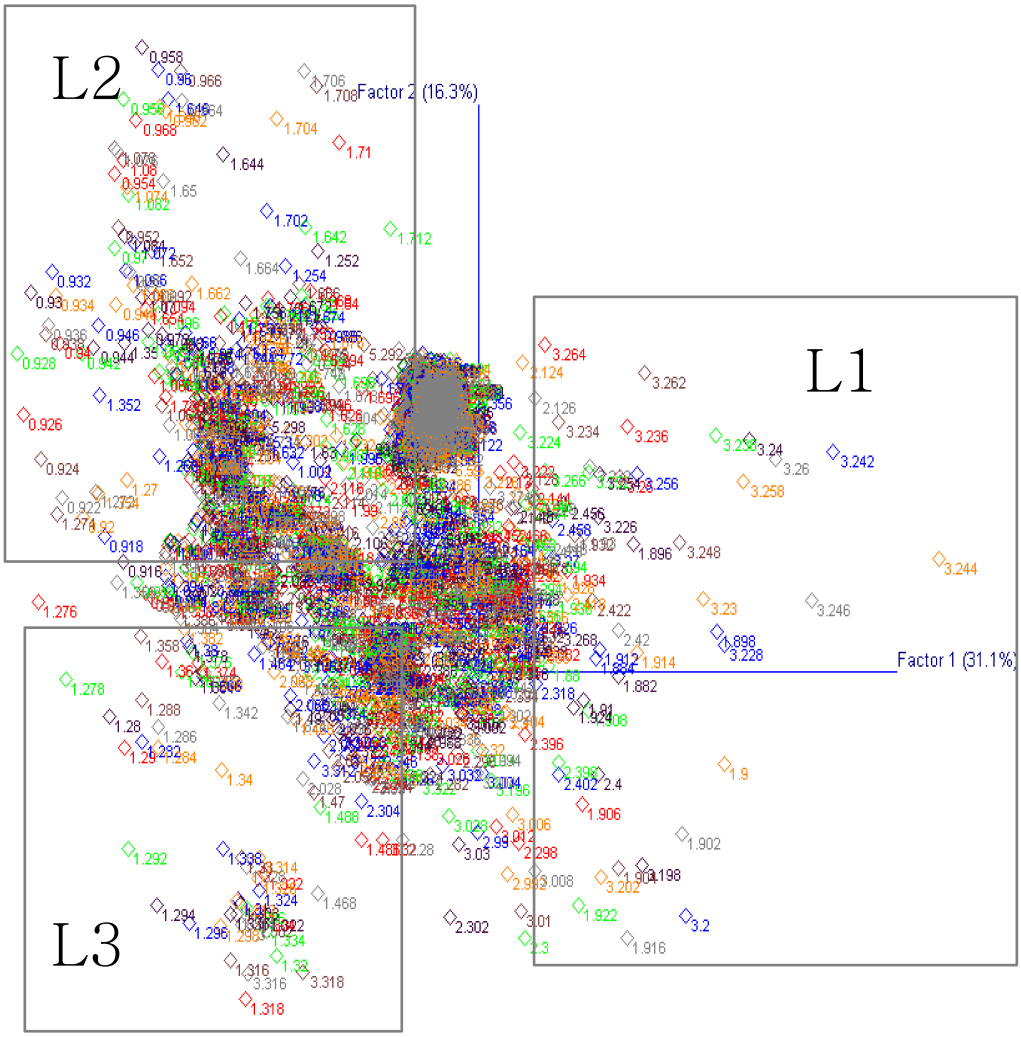


Figure 3.13: *PCA loadings plot of 3* Panax *species correlating to the PCA scores plot in figure 3.12. Each spot represents a different variable, in this case, ppm value. The zoomed in figure of L1, L2 and L3 are shown in figure 3.14, figure 3.16 and figure 3.17.*

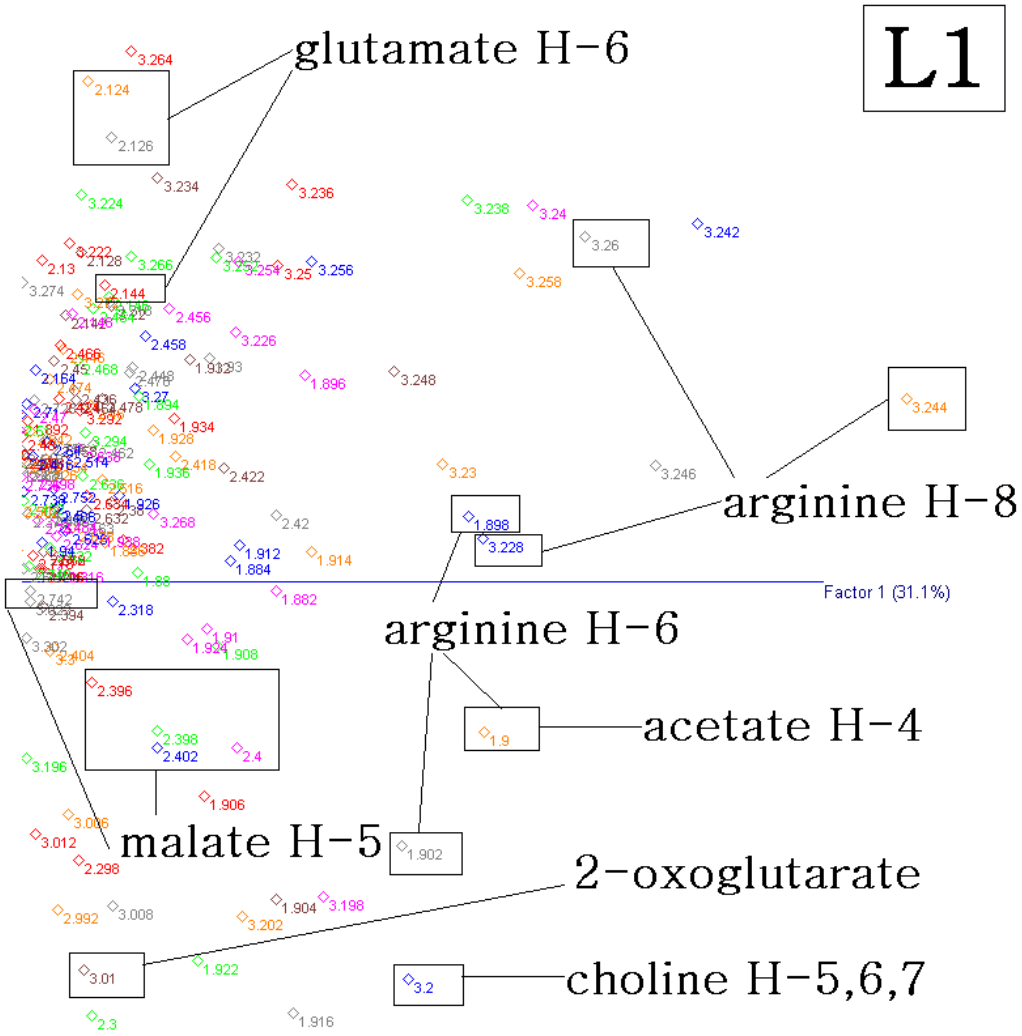


Figure 3.14: *Enlarged L1 of PCA loadings plot of 3* Panax *species focusing on the positive side of PC1. The colour of the loadings are irrelevant to the analysis. The loadings that are highly contributing to PC1 are highlighted in boxes and assigned to metabolites. 3.228, 3.244, 3.262 ppm: arginine H-8; 1.898, 1.90, 1.902 ppm: arginine H-6; 1.90 ppm: acetate H-4;3.2 ppm: choline H-5,6,7; 3.01 ppm: 2-oxoglutarate; 2.396, 2.398, 2.4, 2.402, 2.742 ppm: malate H-5; 2.124, 2.126, 2.144 ppm: glutamate H-6;*

Differentiation of Asian ginseng and American ginseng

In order to closely study the difference between all the ginseng samples, the loadings areas L1, L2 and L3 were enlarged in figure 3.14, figure 3.16 and figure 3.17 and the variables are assigned to ginseng root metabolites.

The relative intensities of the NMR spectral peaks on the selected variables were calculated and plotted in figure 3.15. In order to eliminate the potential influence from sample concentration and quantity, normalization was performed on the whole dataset before obtaining the intensity values on the interested chemical shifts. Standard errors were used to plot the error bars. t-test with 95% confidence level was used to compare samples. The average value for cultivated American ginseng samples was calculated base on 15 batches. The values for Wild American ginseng and Asian ginseng samples were based on 11 and 5 batches respectively.

By inspecting the original spectra, it was found that fumarate, tyrosine, phenylalanine, and tryptophan are clearly visible in all the Asian ginseng samples, but not in American ginseng samples. The responses of these peaks are very low, around a few times higher than the baseline noise. Therefore these peaks were not picked up by PCA loadings. However these peaks can be checked manually to assist the separation of Asian and American ginseng.

In the loadings figure 3.14, apart from the highlighted variables, there are other variables next to them but with some 0.002-0.006 ppm difference. These usually belong to the same peak. Because the bucket size used in this study is 0.002 ppm, it is normal that one NMR spectroscopic peak covers several variables. 3.228, 3.244, 3.262 ppm were assigned to arginine H-8. 1.898, 1.90 and 1.902 ppm were assigned to arginine H-6. The fact that levels of arginine are around 2-fold in ASG than in CAG or WAG at all the chemical shift related to arginine indicates a real difference (figure 3.15).

The level of arginine is also higher in WAG than in CAG at most of the assigned chemical shifts, including 3.26, 3.244, 3.768 ppm. However, at 3.288 and 1.898 ppm the difference is not significance. For 3.288 ppm, it may be because interference of small peaks presented at 3.218 ppm. Nonetheless, the ratio of arginine abundance between CAG, WAG and ASG are well sug-

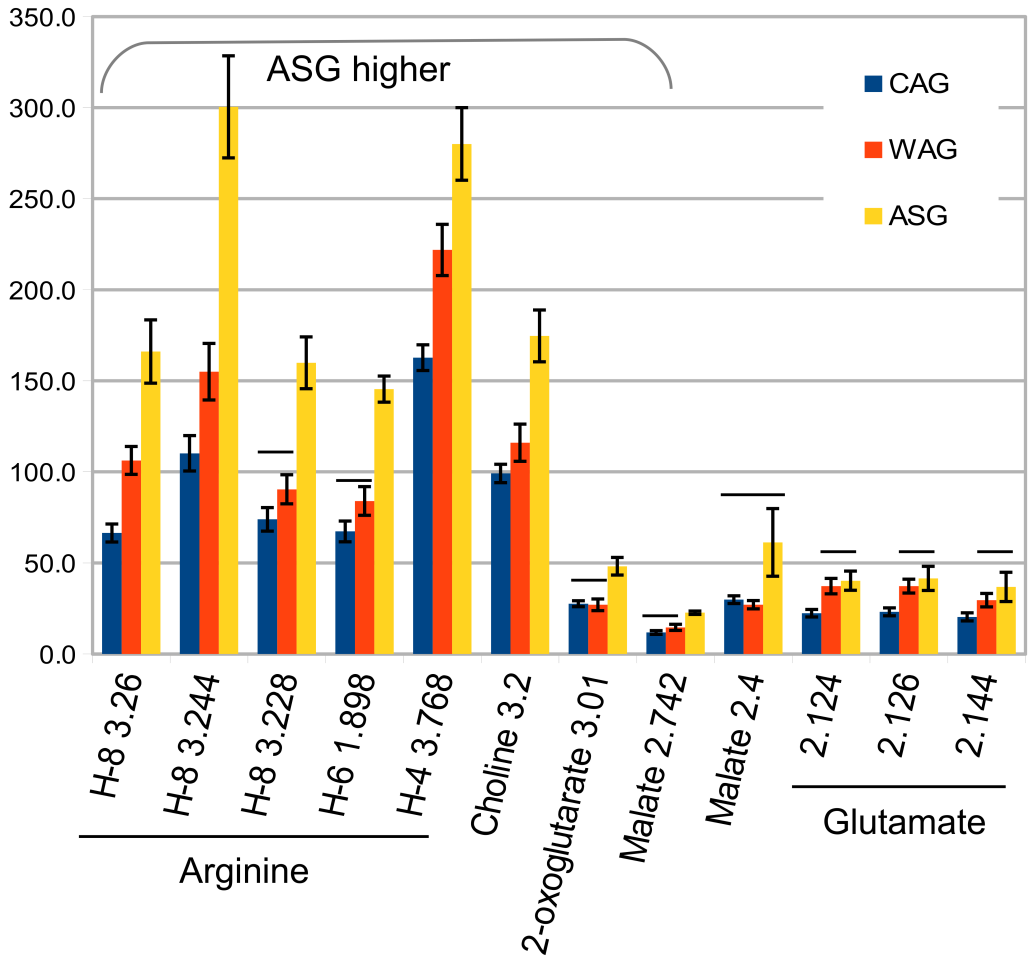


Figure 3.15: *Relative levels of selected metabolite (from figure 3.14) that differentiate between cultivated American ginseng (CAG), wild American ginseng (WAG) and Asian ginseng (ASG). The Y axis is the relative intensity of the NMR spectra on selected chemical shifts. The X axis are the metabolite names assigned to the chemical shift values in buckets with ppm as unit. Error bars represent standard errors based on average values. A t-test with $p<0.05$ was used to indicate a real difference between samples. Within the same chemical shift group (3 samples), if a short "-" is seen on top of a sample, it means that this sample is not significantly different from others which also have a "-" on them, but a sample with a "-" is significantly different from a sample without a "-" within the same group.*

gested by the H-8 and H-6 peak present at 3.26, 3.244 and 3.768 ppm, as no overlapping and interfering peaks were found in this regions.

1.90 ppm assigned to single peak caused by acetate H-4, which overlaps with arginine H-6, but the fact that the difference between ASG and WAG or CAG thoughout the chemical shifts of arginine is the same, indicate that the levels of acetate is the same in all the ginseng. Therefore arginine is the metabolite that contributes the most difference between the Asian ginseng and the American ginseng apart from the sugars.

Apart from arginine, there are other variables also contribute to the separation of the Asian ginseng from the American ginseng. These are 3.2 ppm assigned to choline H-5,6,7, and 3.01 ppm assigned to 2-oxoglutarate. These metabolites contribute the most to PC1. Since Asian ginseng samples are located on the positive side of PC1, it indicates a higher content of arginine, choline and 2-oxoglutarate in Asian ginseng samples. The bar plot in figure 3.15 confirmed this.

There are other peaks contribute moderately to factor 1, these are 2.40 and 2.74 ppm which were assigned to malate H-5, 2.124, 2.126 and 2.144 ppm assigned to glutamate H-6. These chemical shift values are located slightly to the right of PC2 axis. By examining the original spectra, it is confirmed that glutamate is higher in ASG and WAG than in CAG. The levels of malate is higher in Asian than American ginseng samples, as the assignment to 2.742 shows. However, at 2.4 ppm the error bar for ASG is very high. By inspecting the original spectra, it was found that many small peaks are overlapping around 2.4 ppm, this might explain the large variance.

To summarise, the metabolite composition of Asian ginseng is distinctively different from American ginseng samples. Arginine, and choline are much higher, 2-oxoglutarate and malate are moderately higher in Asian ginseng samples than American ginseng samples. Levels of glutamate and arginine are higher in WAG than in CAG.

Table 3.6: *Summary of previous report[108][109] on the ^{1}H-NMR assignments of major ginsenosides including ginsenoside Rb_1, Rg_1, Re, Rd and notoginsenoside R1. The underlined ginsenosides are assigned in figure 3.16.*

Hydrogen	Rb_1	Rg_1	Re	Rg_2	Rb_2	Rc	Rd	R1
H-9	1.33	1.49	1.51	1.53	1.35	1.33	1.95	1.45
H-16β	1.82	1.26	1.24	1.29	1.36	1.33	1.35	1.26
H-16α	2.18	1.72	1.74	1.77	1.82	1.81	1.76	1.73
H-26	1.59	1.59	1.60	1.67	1.65	1.59	1.60	1.59
H-27	1.64	1.59	1.57	1.63	1.62	1.64	1.60	1.59
H-28	1.25	2.03	2.07	2.10	1.26	1.25	1.27	2.04
H-29	1.08	1.57	1.34	1.35	1.07	1.07	1.09	1.44
H-30	0.94	0.79	0.94	0.95	0.94	0.92	0.95	0.78

Differentiation of wild and cultivated American ginseng

Loadings area L2 is enlarged in figure 3.16. The variables were assigned to methyl groups in ginsenosides which are secondary metabolite instead of primary metabolite. A typical structure of ginsenoside is shown in figure 3.2. The ^{1}H-NMR assignments are shown in table 3.6.

Ginsenosides are assigned to the variables in the loadings figure 3.16 according to previous reports[108][109] and ^{1}H-NMR spectra of ginsenoside Re, Rg_1 and Rb_1. It was found that the NMR spectra of various ginsenosides are very similar to each other, especially the chemical shifts of methyl groups. The chemical shifts shown in figure 3.16 can be assigned to the underlined ginsenosides shown in table 3.6. Because L2 locates in the same area in the loadings plot (figure 3.12) as WAG in the scores plot (figure 3.13), it indicates WAG has a higher intensity of methyl peaks, and means a probably higher content of ginsenosides in wild than cultivated American ginseng. However, previous reports were not able to show a difference of total ginsenosides content between wild and cultivated American ginseng[112][113].

Figure 3.17 is the enlarged left bottom side of the PCA loadings plot. 1.486 and 1.468 ppm were assigned to alanine H-6. 1.334 and 1.318 ppm were assigned to threonine H-8.

A quantitative analysis of the levels of metabolite that assigned to the left side of the PCA loadings plot (figure 3.13) including area L2 (figure 3.16)

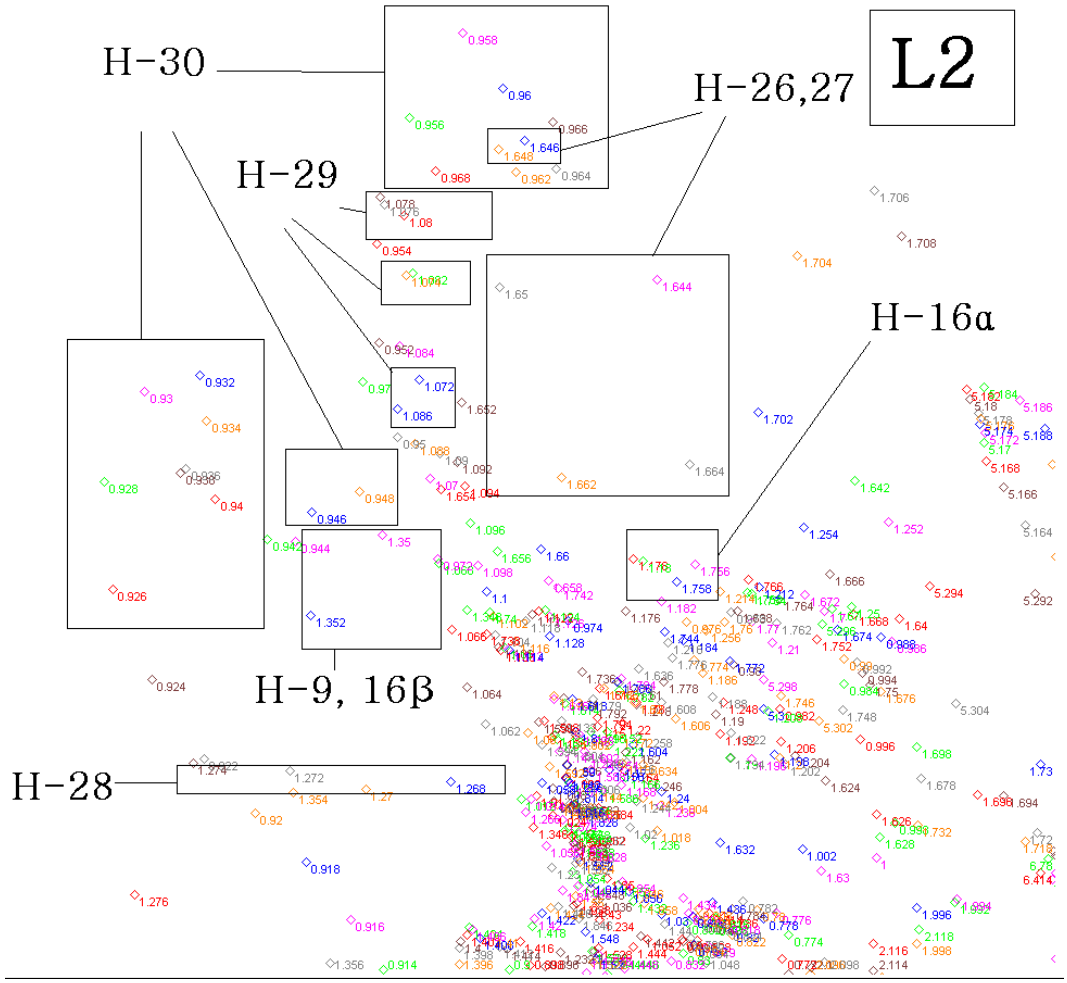


Figure 3.16: *Enlarged L2 area of PCA loadings plot of 3* Panax *species focusing on the positive side of PC2. The colour of the loadings are irrelevant to the analysis. The loadings that are highly contributed to PC2 are highlighted in boxes and assigned to the general structure of ginsenosides, referring to table 3.6.*

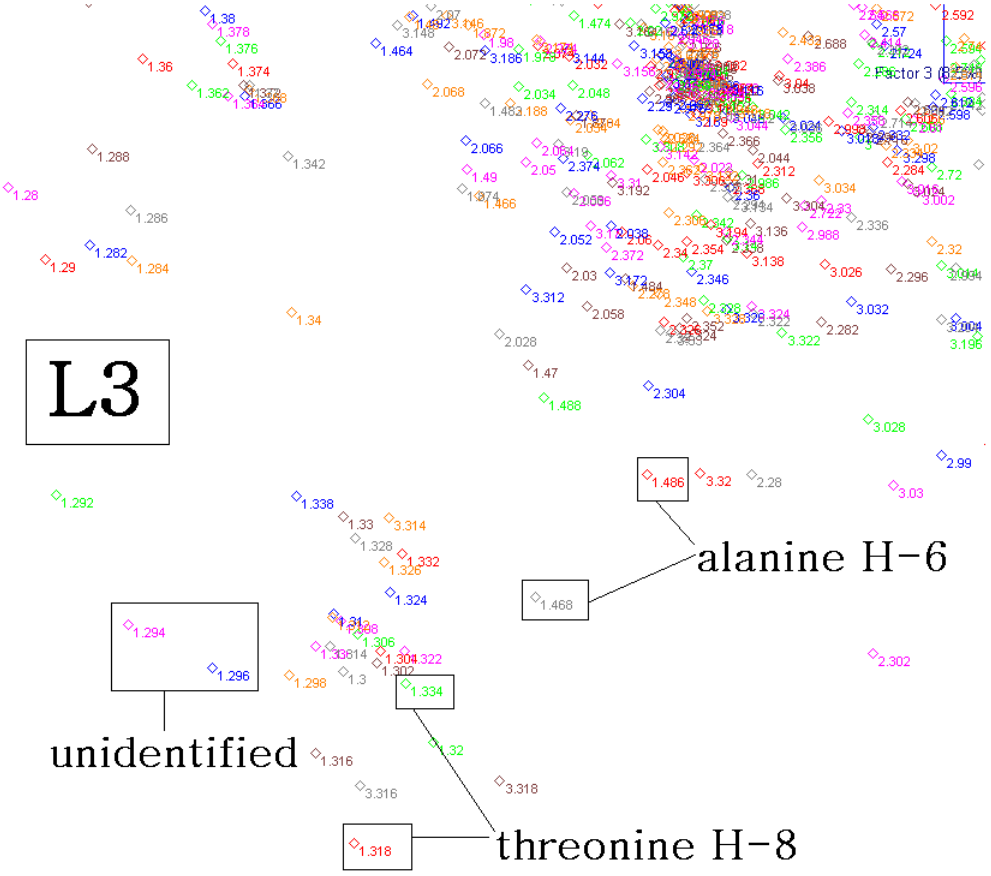


Figure 3.17: *Enlarged L3 area of PCA loadings plot of 3* Panax *species focusing on the negative side of PC1 and PC2. The colour of the loadings are irrelevant to the analysis. The loadings that are contributing to the negative side of PC1 and PC2 are highlighted in boxes and assigned to metabolites as follows: 1.486, 1.468 ppm alanine H-6; 1.334, 1.318 ppm threonine H-8. Peaks at 1.296 and 1.298 ppm are probably due to ginsenosides H-16.*

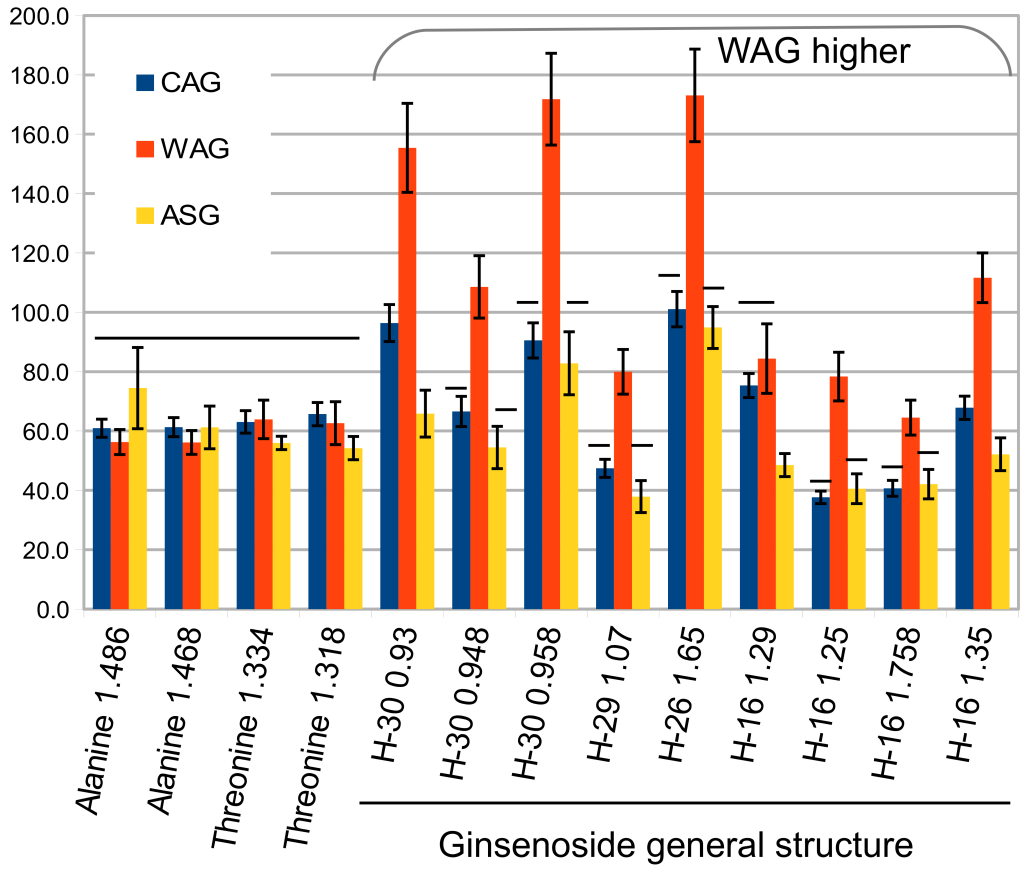


Figure 3.18: *Relative levels of selected metabolites (from figure 3.16 and figure 3.17) in cultivated American ginseng (CAG), wild American ginseng (WAG) and Asian ginseng (ASG). GSS: General ginsenoside structure referring to figure 3.2. The Y axis is the relative intensity of the NMR spectra on selected chemical shifts. The X axis are the metabolites names assigned to the chemical shift values in buckets with ppm as unit. Error bars represent standard errors. A* t*-test with $p<0.05$ was used to indicate a real difference between samples. Within the same chemical shift group (3 samples), if a short "-" is seen on top of a sample, it means that this sample is not significantly different from others which also have a "-" on them, but a sample with a "-" is significantly different from a sample without a"-" within the same group.*

and L3 (figure 3.17) were plotted in figure 3.18.

The chemical shifts of the proton on the same location in the general structure are sometimes the same, but sometimes different between different types of ginsenosides. Take H-16β for example, in Rg_1, Re, and R1, its chemical shift is around 1.25 ppm, but in Rb_2, Rc and Rd, its chemical shift is around 1.35 ppm. This difference may be due to the different aglycone moieties. Rg_1 and Re are the protopanaxadiols, but Rb_2, Rc and Rd are the protopanaxiatriols. However, the aglycone moiety is not solely responsible for the difference. Rb_1 also belongs to the protopanaxiatriol group, but its H-16β chemical shift is 1.82 ppm, which is very different from the other ginsenosides in this group, which have chemical shifts around 1.35 ppm.

Apart from this, peaks at the same chemical shift may be caused by protons located on different place in the GSS. For example, 1.25 ppm can be caused by H-28 in Rb_1, Rb_2, Rc and Rd, or by H-16β in Rg_1, Re, and R1.

For many reasons discussed above, it is difficult to attribute the selected chemical shifts to particular ginsenosides, but to the ginsenoside general structure. It was found that for almost all the selected chemical shifts analysed in figure 3.18, their intensities in WAG is almost twice as high as in CAG and ASG, except at 1.29 ppm, where CAG is at the same level as WAG. This may be because 1.29 ppm is only contributed by Rg_2 (GSS-16β).

The suggested chemical shifts may be used as important markers to identify the wild American ginseng from the cultivated ones.

At 5.29 ppm there is almost no peaks in Asian ginseng samples, but there is a single peak in all the cultivated American ginseng samples and most of the wild American ginseng samples. At 1.29 ppm, the peak intensities in WAG and CAG are almost twice as high as in ASG. However these peaks remain unidentified.

3.3.5 Difference among wild American ginseng samples

Wild American ginseng sample 18, 29 and 32 are distinct from both wild and cultivated American and Asian ginseng samples, and from each other. Peaks

at (3.936, 3.962, 3.988)t, 3.646(s), 3.622(s), and (3.584, 3.594)d are high in WAG 18, 29 and 32. However, these peaks are unidentified, but are likely due to polysaccharides.

There are four peaks between 1.27-1.34 ppm present obviously in American ginseng, moderately in Asian ginseng, not detected from sample 32. The intensities of these peaks in other WAG including 18 and 29 vary.

Sucrose peaks in sample 32 and 29 are higher than other WAG, but for glucose peaks, it is the other way round. In almost all cases, either glucose or sucrose responses are high, but not both. It might be that sucrose degraded in to glucose in the metabolism [110].

Several reports suggested there are two chemotypes of wild American ginseng: the "low Rg_1/high Re" type and the "high Rg_1/low Re" type[114][115]. This can potentially be one of the reason that the wild American ginseng studied in this chapter are so different from each other. However, this was not confirmed in this study because the ^{1}H-NMR spectra of these two ginsenosides are very similar.

It is difficult to correlate the age and growth condition of wild American ginseng to their compositions of metabolites, because suppliers of wild American ginseng do not wish to reveal the location of it. A common way of detecting the age of ginseng is usually by counting the number of the bud scars on the neck of the root. However, it is often trimmed when seen in the market. Research of wild American ginseng is very limited due to the limited availability.

3.3.6 Difference of a suspected counterfeit cultivated American ginseng sample number 20

In the PCA of 4 different kinds of root samples, as shown in figure 3.11, sample 20 was outside the group of cultivated American ginseng, this confirms the suspicion that this sample is probably not from a batch of genuine cultivated American ginseng. However, sample 20 seems to be close to most of the wild American ginseng samples. Since the variations between wild American ginseng themselves were quite large, it is not adequate to say that

sample 20 is wild American ginseng, and clearly, it is not, because there is no appearance features of sample 20 that fits those of wild American ginseng.

After exclusion of the sugar region, in the PCA (see scores plot figure 3.12) of 31 batches of cultivated American ginseng, wild American ginseng and Asian ginseng, the scores of sample 20 appeared close to the cultivated American ginseng in the PCA scores plot, but further away from wild American ginseng.

By closely investigating the overlapping spectra of 4 kinds of ginseng samples, it is found that at 3.213 ppm, there is a doublet peak in sample 20, which does not occur in any of the cultivated American ginseng samples.

At 2.40, 2.74 and 4.31 ppm, doublet-doublet peaks were present in almost all the ginseng samples, except for sample 20. These peaks were confirmed to be malate.

At 4.65 ppm a a double-doublet peak was present in all the wild American ginseng samples and sample 20, but not in cultivated American ginseng samples. All these evidences show sample 20 is neither wild nor cultivated American ginseng, nor Asian ginseng.

Although PCA did not pick up sample 20 in the first stage, excluding the sugar revealed that sample 20 is suspect. Based on ^{1}H-NMR spectra, it is almost certain that sample 20 is a different species other than those ginseng studied in this chapter. Malate and 3.21, 4.65 ppm can be used to distinguish sample 20 from commercially valuable ginseng, and potentially be usefully at identifying fake ginseng with non-commercial value.

3.4 Conclusion

For the first time robust and distinctive difference of Asian ginseng from American ginseng were found and the key metabolites responsible were identified as sucrose, glucose, arginine, and choline, based on the data of 15 batches of CAG, 11 batches of WAG and 5 batches of ASG from Chinese medicine stores all over China. A previous report quantifying the metabolite content did not report any difference of arginine and choline between cultivated American and Asian ginseng[16]. In that report only 2 geographically

Table 3.7: *Summary of difference in metabolites between CAG, WAG and ASG. The underlined metabolites present a higher content in the underlined ginseng type in each comparison.*

Comparison	metabolites that are different
CAG and ASG	sucrose, glucose, arginine, choline, 2-oxoglutarate, malate (2.742 ppm), glutamate, GSS-16, 28 (1.35, 1.29 ppm)
WAG and ASG	arginine, choline, 2-oxoglutarate, malate (2.742 ppm), glutamate, GSS-30, 29, 28, 26, 16
CAG and WAG	(sucrose and glucose: for some of the WAG) glutamate, GSS-30, 29, 28, 26, 16

different Asian ginseng batches and 1 American ginseng batch was studied. Although 5 biological replicates were used, the fact that they were grown in the same field and the same year results in the metabolite compositions being very similar. Therefore the result might not be a good representative.

To the author's best knowledge, it is the first time that ^{1}H-NMR-PCA method is used to analyse primary metabolites of wild American ginseng samples, although there are several reports on major ginsenosides contents of wild and cultivated American ginseng using HPLC[113][116][112].

This study used a large sample size of commercially available ginseng samples collected from 15 pharmacy stores over 5 provinces in China. It is a good representation of the metabolite composition of wild, cultivated American ginseng and Asian ginseng. The relative content of metabolites were quantified and compared, and summarised in table 3.7.

It was found that there is a large variance within the WAG and ASG samples themselves. The variance between CAG samples is much smaller comparing to those. Differences between wild American ginseng were discussed and analysed. A suspected fake American ginseng sample (Sample 20) was successfully distinguished from the genuine samples.

Novelty in the present study used a multi-step region-excision PCA on analysing the metabolite composition of ginseng extract. This model excises the most abundant metabolites of ginseng extract from the NMR spectra before PCA, and correlated the differences between ginseng types to the less

abundant metabolites.

Chapter 4

Proteomic quality control of American ginseng on MIN6 pancreatic β cells using DIGE

4.1 Introduction

In the previous chapters, the quality of traditional Chinese medicines was assessed using chromatography and spectroscopy of complex herbal extracts by principal component analysis. Clear separations between very similar TCM were achieved, including separation of Asian and American ginseng (Chapter 3), and separation of Danshen grown in different locations (Chapter 2). This was carried out using principal component analysis on a holistic profile of chemical constituents, obtained by either NMR, HPLC or IR. This chemical based quality control method is usually applied using a number of TCM samples from trusted sources and excluding the outliers using statistical analysis on the profiles of chemical constituents, then comparing unknown plant samples with the known ones.

However, there are always some undetectable compounds present due to the limitations of detection methods (discussed in section 3.1.6), and those compounds might also contribute to the effects of TCM. If the effects of TCM can be measured and controlled directly, the control of chemical composition

might be skipped. This is the purpose of the study in this chapter.

In order to control the complex effects of the multiple ingredient TCM, a high-throughput, information-rich biological assay is needed. In the next section, Various high-throughput, information-rich "Omics" techniques are compared and discussed, based on their potential application on quality control of TCM. Among those, proteomics was chosen as a potential method.

Various concentrations of wild and cultivated American ginseng were tested on pancreatic MIN6 cells. The highest concentrations without causing >5% cell death were selected, and used to treat MIN6 cells over a period of 24, 48 and 66 hours. The cellular proteins were analysed using the difference in gel electrophoresis (DIGE), and the up/down regulated proteins before and after treatment were selected and compared between wild and cultivated American ginseng treatment.

4.1.1 Various "Omics" techniques used in TCM study

In recent years, information rich, high through put "Omics" methods are becoming increasingly popular, including metabolomics, genomics, transcripnomics, and proteomics. The "Omics" techniques are favourable for TCM studies is because their holistic nature coincides with TCM's "synergy" principle[117].

Those "omics" methods can be used to describe the TCM's own biological features; or used as a systems biology approach to evaluate the biological features of the *in vitro/in vivo* samples treated by the TCM.

Many of the previous published reports using "omics" focus on biological features such as the metabolites[15], DNA sequence [118–120], proteomic profile [121] of the TCM itself, instead of linking it to its effects on another biological system, such as cells or animals. Only a fraction of the reports using the "omics" methods are using the systems biology approach[117].

4.1.2 Metabolomics and metabonomics

Metabolomics and metabonomics are often used interchangeably, however, strictly speaking there are some subtle differences in terms of their differ-

ent usages. Metabolomics is usually about the analytical description of the metabolite profile of a biological system, but metabonomics focuses more on the effects of stimuli on global dynamic metabolites [122]. In terms of their usage on TCM, metabolomics can be used to focus on the metabolites of a TCM, as discussed in chapter 2 and 3; metabonomics may be used to relate the quality of TCM to its biological effect. Usually, metabonomics studies on TCM are done by collecting urine or blood samples of human or animal test subjects treated by a TCM, thereby linking the changes in their metabolite profiles to its effects [123].

A urinary metabonomic method based on ultra-performance liquid chromatography coupled with mass spectrometry (UPLC/MS) was used to study the therapeutic effects of Rhizoma Drynariae on the "Kidney-Yang Deficiency syndrome" [123]. Some significantly changed metabolites were found to be related to the disturbance in energy metabolism, amino acid metabolism and gut microflora, which were related to the syndrome. However, this research is only focusing on the effect of the TCM, but not discussing the quality which might lead to different effect.

To the author's best knowledge, there is no report using metabonomics on evaluating the quality of TCM by linking the TCM quality to the metabolite changes in the treated and control animal/cell model. Animal metabolites are usually primary metabolites, the numbers are less than the rich and distinctive secondary metabolites existing in plants. There might be some animal metabolites that specifically response to the TCM treatment, and make the quality control process possible. Since some research has successfully shown that the metabolite profiles of animal models change distinctly before and after TCM treatment [123], it might also be possible to see some difference of the profiles of the metabolite treated by different qualities of TCM.

4.1.3 Genomics

DNA sequencing

Various DNA methods including direct DNA sequencing, Random Amplification of Polymorphic DNA (RAPD) [118], DNA bar coding [119] and DNA

microchip electrophoresis were used to distinguish similar species of TCM including Asian and American ginseng samples [120].

Asian and American ginseng were successfully distinguished by direct DNA sequencing [118]. Within the Asian ginseng species, there are different processing methods used, including steaming and drying. However, the DNA methods can not distinguish these differences. Since these two types of ginseng have some distinctly different chemicals due to different processing methods, it is important to be able to see the difference. The ^{1}H-NMR-PCA method has been used successfully on this [16].

RAPD had also been used in this research, but only parts of the samples gave good quality fingerprints [118]. The RAPD method is criticised by its potential to amplify the DNA of contaminations such as bacteria and fungus. Besides, RAPD only works well with reasonably preserved DNA, so it is not feasible for many commercially available dried or ground herbal medicines with highly degraded DNAs [118].

However, the composition of DNA is not affected by the growth environment, the age and different parts of the plant, but the chemical compositions are affected [11][124]. Because the biological effects of the plant in the human body are mainly related to the chemical composition of the plants, it is not adequate to only authenticate the DNA.

To summarise the above discussions, DNA methods are good to distinguish genetically different plants, for example, different species, which are used as adulterants of the authentic ones, but not be able to detect the various acquired features which are also very important for the chemical constituents, which might lead to different effects. Therefore the DNA method is not a true indication of the quality of TCM.

Transcriptomics

Transcriptomics is the study of the entire collection of RNAs (transcriptome) encoded by the genome of a specific cell or organism at a specific time or under a specific set of conditions. Since the genome is almost identical in every cell within human or other animals, it is interesting to know which gene

is turned on (transcribed into RNA molecules) or off in the cells or tissues in an organism.

Some researchers used transcriptomics on the study of the activities of TCM. A well established TCM formula "Huang Qin Tang" consists of four herbs which were found to reduce chemotherapy(CPT-11)-induced gastrointestinal toxicity in a phase I/II clinical trial [125]. The expression levels of four important marker genes related to cell recovery were monitored in the control and treated experiments, and "Huang Qin Tang" was found to be able to recover the gene expression levels of Lgr5 and Olfm4, which was decreased by CPT-11 [125].

This demonstrates a potential possibility of using gene expression microarray analysis for activity-based quality control of TCM. Although it is not "high-throughput", in that they have used only four marker genes suggested by previous research to investigate the activity of TCM, it is partially "systematic" by using a herbal mixture instead of isolated compounds. The ideal systems biology quality control method for TCM should be "system" to "system", that is, from the "system" of complex herbal compounds mixture, to the "system" of the TCM treated biological sample, focusing on either RNA, small molecule metabolite or proteins in cell, animal, or human. It will be even better to unite all of the information mentioned together, which is impossible with the current ability.

4.1.4 Proteomics

The original definition of "PROTEOME" analysis means "The analysis of the entire PROTEin complement expressed by a genOME, or by a cell or tissue type." [126] The idea of studying the complete complement of proteins in a cell was made possible by two dimensional gel electrophoresis (2D-GE), which had already been used in the 1970s to separate proteins [127]. 2D-GE offers a high resolution of protein separation, that up to thousands of proteins can be separated in one gel.

Various proteomic methods

2D-GE separates proteins according to their isoelectric points in the first dimension and molecular sizes in the second dimension, which are independent of each other. These two parameters can separate several thousands of proteins in a gel. The separated proteins can be preserved in the gel matrix until further identification using MS.

Apart from 2D-GE techniques, gel-free methods including iTRAQ (isobaric tag for relative and absolute quantitation) and iCATs (isotope-coded affinity tags) also became popular [128][129][130]. These methods use isotope tags to label proteins, then separate them using multiple liquid chromatography, followed by MS identification. The gel-free proteomic methods eliminate the complex 2D-GE experimental process, but are not be able to visualise a large part of proteome and analyse complex protein lysates as easily. Currently no report was found using gel-free proteomic methods on quality control of TCM.

Conventional two dimensional electrophoresis gel and its application on plant proteomic analysis

Only a few preliminary reports have discussed the feasibility of using proteomics on quality control of TCM. Most of the reports use proteomics to study the proteome of TCM, but not TCM-treated calls or animals. For example, silver staining 2D-GE was used to distinguish Asian and American ginseng by profiling the proteome of their cultured cells.

The proteome profile of four parts of Asian ginseng including the main roots, lateral roots, rhizome head and ginseng skin were analysed and 9 common protein spots were found throughout the gels. Although hundreds of proteins were shown in the gels obtained, cross-gel matching is poor and the the clarity of the image is not good enough to draw any firm conclusion on the difference of the proteome of different parts of the plant [131].

Coomassie Brilliant Blue staining has been applied to the proteomic analysis of rice seedling infection [132]. The changes in the proteomics profile of rice plants before and after *Sinorhizobium meliloti* 1021 infection was stud-

ied. It was found that photosynthesis related proteins were up-regulated only in leaf sheath and leaf, while the up-regulated proteins in roots were solely defense related [132].

Two-fold changes of protein abundance are regarded as significant using Coomassie Brilliant Blue. However, this ratio of changes can be reduced to 1.2 with the use of the modern DIGE method, since difference of protein changes less than 10% can be detected with over 95% confidence [133].

Difference in gel electrophoresis(DIGE)

Comparing to the conventional 1-sample-per-gel 2D-GE techniques including silver staining or Coomassie Brilliant Blue methods, the 3-sample-per gel DIGE method is a great improvement allowing internal standard, control and treated samples to be loaded on the same gel, therefore reducing the number of technical replicates needed.

One of the drawbacks of using silver staining 2D-GE technique is that not only biological replicates, but also technical replicates should be used for each sample. "Biological replicates" means different batches of the prepared protein samples, in order to eliminate the protein preparation/sample variance; "technical replicates" means to run a few (at least 3) gels for each batch of proteins, in order to eliminate the many possibilities causing variance during gel preparing/running/staining practice. It is a great amount of work to run enough gels to meet the statistical significance requirement. Therefore many papers published using conventional staining methods including silver or Coomassie Brilliant Blue, have difficulties to quantify the proteins and to tell the differences between gels.

As indicated in figure 4.1, 1/3 of control and 1/3 of treated samples were taken and combined to make the internal standard sample, which was labelled with Cy2. Test samples (Control and treated) are labelled with Cy3 and Cy5. The three labelled samples are then mixed together and loaded on an immobilised pH gradient (IPG) strip, which is the first dimensional separation according to the isoelectric points of the proteins. Then the IPG strip (1D) is placed on top of an SDS gel for the second dimension separation

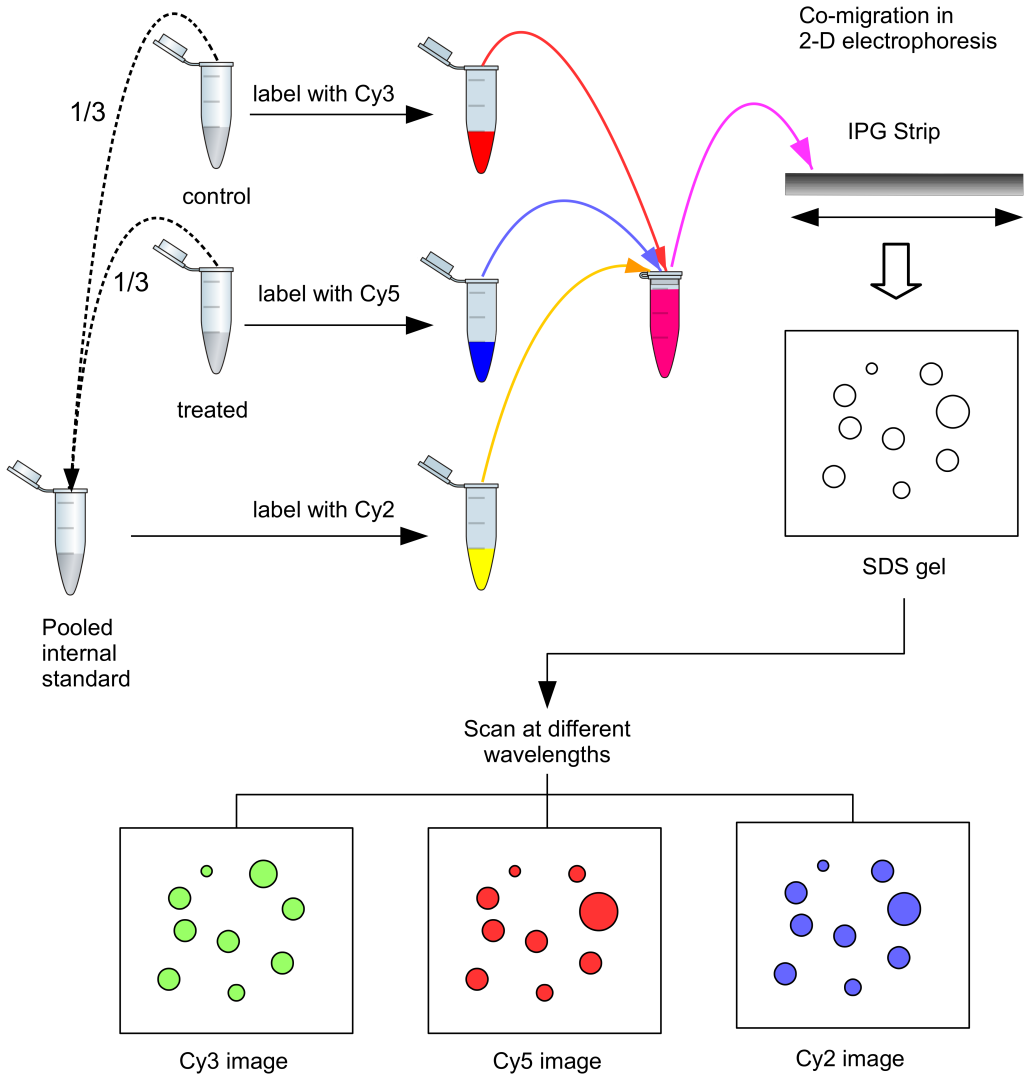


Figure 4.1: *Simplified schematic representation showing the creation of a pooled standard in DIGE. The sample proteins are labelled with Cy3 and Cy5, the internal standard with Cy2. Details of the dyes see section 4.2.8. Adapted from "Proteomics in Practice" by Westermeieret al. (2008) [133].*

according to the molecular weight of the proteins.

After the second dimension is finished, the gel is then scanned at different wavelengths to obtain three separate images representing the protein profiles of the three samples. Because the three protein samples are run in the same gel, they have the same experimental conditions, and therefore a largely improved reproducibility and accuracy of 2D-GE technique. The excitation range of Cy3, Cy5 and Cy2 are not overlapping with each other, therefore there is no risk of interfering during scanning.

Comparing to the medium sensitive Coomassie Brilliant Blue staining method, with a sensitivity of 20 ng of BSA, DIGE fluorescent labelling has a high sensitivity of 25 pg. Although this sensitivity is similar to silver staining which is around 1-60 ng, the dynamic range of DIGE fluorescent labelling ($< 10^5$) is much higher that that of the silver staining($< 10^2$) method (Ettan DIGE System, User Manual, GE Healthcare). Therefore in terms of precise, quantitative, and sensitive analysis, DIGE is much preferable over Coomassie brilliant Blue and silver staining. The invention of DIGE greatly advanced the ability of 2D-GE and remains unrivalled among the various proteomic techniques [134].

Limitations

Although DIGE is already a great improvement than the traditional one protein per gel staining method, it still has some shortcomings:

1. The following proteins cannot be analysed: proteins with PI lower than 3 or higher than 11; similarly, proteins having a molecular weight lower than 8 kDa or higher than 250 kDa.

2. The experimental time is long. It usually takes 4 days from labelling the protein to obtain a DIGE image. This excludes the time spent on cell culturing, protein extraction and purification, and MS identification, which could easily be a few weeks.

3. Hundred of experimental steps were included in the protocol. Therefore technical skill and experience is required. Attention should be paid to every detail. For example, the sides of DIGE image can be blank if excessive

water was present in the paper pads on the side of the IEF gel during 1D. This can sometimes be seen even in published results.

4. There is a limitation of the sample numbers that can be analysed in one batch. The capacities of both 1D IEF and 2D electrophoresis equipments only allow 12 samples to be loaded.

Cross lab comparison has always been a difficulty in a 2D-GE. During the experiment, about 20 self-mixed solvents are used and some solvents could be stored in the fridge or at room temperature for a few weeks to a few months. Various storage times across labs might result in slight difference in the results. The self-made polyacrylamide gel might also be a main source of variance. Different labs usually have different optimised recipes for making the gels. The slight difference in gel composition might also affect the results. Besides, among the many steps involved in preparing the DIGE experiment, for many of them there are several ways of doing it. Usually, different labs have their own preferences depends on the protein samples they are analysing.

Some internet-based program might assist the comparison between the gels obtained and similar ones existed in online databases[135], including SIB (Swiss Institute of Bioinformatics), by warping and matching the 2D gel images automatically and manually . However, even if the same cell type was used, due to the many factors that might affect the result - including cell treatment, protein extraction method and pH range selected, the comparison might become very difficult. In the current study, only one paper showing a DIGE image of MIN6 cell was found, however this paper was using pH 4-7 as opposed to the current thesis using pH 3-10[136]. This image is not included in the SIB. The most similar cell listed in SIB to the current study is a mouse pancreatic islet cell 2D page map. Although MIN6 cell count for 65-80% of the cells in mouse pancreatic islet, it is still difficult to correlate without much information of MS identified proteins. However if the DIGE parameters are standardised, cross lab comparison should be possible.

The accuracy and reproducibility of the method is important. In a previous report, a preliminary comparison of results across 5 labs and 10 practitioners were carried out, using 2 similar test samples, with each participant using the same protocol and need to report the 200 most different proteins

between these two samples (Voshol, H., Oral presentation, Reproducibility of 2D gel-based proteomics experiments Workshop for Proteomics Validation, HUPO 2008, 7th Annual World Conference., 2008). The results showed that the cross lab reproducibility is good (>84 similarity) and the results of with-lab but different operator is excellent(>90 similarity). Most of the variation is believed to come from the sample preparation procedure.

Application of DIGE on plant proteomic analysis

DIGE has been used to distinguish very similar plants. Strawberries grown in different conditions were distinguished by DIGE according to their different proteomes [121]. Grape berries during post harvest withering were also analysed using DIGE [137]. 90 proteins were found to express differently during berry ripening/withering and 72 were identified using MS/MS. These papers demonstrate the possibilities of using DIGE to distinguish different TCM. However, this method still focus on studying plant proteins, but not link the quality to its activity on cells.

DIGE has been used to study the activities herbal medicine on cells. Conventional and herbal antidepressant treatments were tested on H22 neuronal cells, and similar changes in protein expression levels were found [138]. This demonstrates the possibility of analysing the proteome of cells treated by TCM.

Although the purposes of none of the papers mentioned were to use DIGE to profile the proteome of the cells treated by different TCM, therefore to control the quality of TCM, it indicate the possibility of doing it.

Proteins participate in the most essential life activities, and also reflect the effect of a medicine. It will be highly beneficial if the quality of TCM could be linked to its effect in a cell.

In this chapter, DIGE is used to study the feasibility of activity-based QC of TCM using 2D-GE. Because ginseng products were thoroughly studied in Chapter 3, they was selected as the ideal TCM to study their biological effects on cells. Besides, there some known effect of ginseng on diabetes, which is further explained in section 4.1.6. Therefore came the idea using one stable

cell line related to diabetes (pancreatic β cell line MIN6, which is further explained in the next section 4.1.5) as the biological system. The changes of proteome in the cell before and after treatments by American ginseng was studied, to explore the feasibility of quality control of TCM using proteomics.

4.1.5 Diabetes and MIN6 cells

There are three main types of diabetes: Type 1, Type 2, and gestational diabetes. Type 1 diabetes, sometimes called insulin-dependent diabetes, is considered to be an autoimmune disease. In this type of diabetes, pancreatic β cells lost the function of producing insulin. Type 2 diabetes, sometimes called non-insulin dependent diabetes, is the most common type which accounts for at least 90% diabetic patients. Type 2 diabetes is related to insulin resistance or relative insulin deficiency, usually together with pancreatic β cell failure.

Pancreatic β cells are closely related to diabetes, due to their function of storing and releasing insulin. They make up of 65-80% of the pancreatic islet.

MIN6 cells are insulinoma pancreatic β cells derived from transgenic mouse developed by Miyazaki[139]. The MIN6 cell line retains the ability to respond to glucose and other secretagogues comparable with cultured normal mouse islet cells and has been used extensively for insulin secretion studies [140].

In order to apply proteomics on quality control of TCM, a stable protein composition is preferred. The reason that MIN6 cells are used instead of pancreatic islet tissue is because, the latter is made up of at least 5 types of cells, resulting a complex and unstable mixture of proteins that is difficult to quantify.

4.1.6 Herbal medicine including American ginseng used in treatment of diabetes

Traditional Chinese medicine has been used to treat diabetes mellitus for a long history. Li *et al.* have summarised 86 traditional Chinese medicines which have demonstrated experimental or clinical effectiveness, including Asian and American ginseng, *Radix Astragali seu Hedysari*, *Rhizoma Anemarrhenae*[7].

Besides TCM, there are other natural products also reported to have anti-diabetic effects. The Indian herbal medicine *Gemnema sylvestre* was reported to stimulate insulin secretion and increase intake of glucose in both *in-vivo* (Wistar rats) and *in-vitro* (β cells in Langerhans islets) studies[141]. The effect is considered to be contributed by a group of triterpenoid saponins, known as *Gemnema* acids [142].

Among those medicines, American ginseng has been found to increase ATP and insulin production/secretion, and reduce apoptosis [143]. American ginseng down regulates uncoupling protein-2, which is a critical protein in the process of insulin secretion and cell survival [143]. More research is needed to fully understand the mechanisms of the effects of American ginseng on pancreatic β cells.

Wild (WAG) and cultivated American ginseng (CAG) studied in the previous chapter were used to treat MIN6 cells. The changes of the protein profiles of the cells were compared and analysed between treated and untreated cells, and between different treatment conditions (CAG and WAG over 24, 48, and 66 hours), thereby attempting to link the activity of American ginseng to its quality.

4.2 Methods

4.2.1 Preparation of ginseng extract

Wild American ginseng batch 17 and cultivated American ginseng batch 8 studied in chapter 3 were used in this study following the same extraction

method. These two batches are available in a larger quantity and reflect a representative character of wild and cultivated American ginseng. Ginseng dried root material was ground using a coffee grinder and filtered through a 350 μm sieve. 2.00 g of the samples was extracted with 100 ml 70% methanol in a volumetric flask, by sonicating for 15 minutes at 30°C followed by shaking manually and sonicating for a further 15 min. After cooling to room temperature, evaporated solvent was replaced with 70% methanol and the volume of the solvent was made up to 100 ml. The supernatant was filtered with a 0.2 μm syringe filter.

The filtered solvent was under rotary evaporation for 30 minutes to evaporate methanol. The remaining of the solvent was frozen using dry ice then lyophilized for 24 hours.

4.2.2 Cell culture

Maintenance of MIN6 cells

MIN6 cell lines were kindly provided by Professor Shanta Persaud and Professor Peter Jones (Diabetes & Nutritional Sciences Division, School of Medicine, King's College London).

MIN6 cells were grown in cell culture medium DMEM (D5796, Sigma) containing 10% fetal bovine serum FBS, 1% glutamine and 2% penicillin-streptomycin at 30°C in an atmosphere of 95% air/5% CO_2. The cell culture flask is negatively charged for cells to attach. Cell culture medium was changed every 3 or 4 days. When changing medium for cells in a T75 flask, 6 ml room temperature sterile phosphate buffered saline (PBS) was used to wash the cells briefly followed by adding 10 ml 37°C cell culture medium. Passages 27 to 37 of MIN6 cells were used in this chapter.

Subculturing

MIN6 cells were subcultured at around 70% confluence. The cells were briefly washed with 6 ml PBS, then trypsinized using 3 ml trypsin/EDTA solution in a 37°C cell culture incubator. When cells start to detach from the bottom,

which usually takes 2-3 minutes, the flask was knocked firmly against the palm of the hand to detach the cells completely. 7 ml of cell culture medium was added to the flask. An autopipette was used to pipette the medium up and down until the cells were separated into singlet or doublet. This is important because if the cells are aggregated, they will grow in groups in new flasks. This will result in uneven growth rate, which means that cells within the aggregation group will grow faster and start to die while other parts of the flask are still empty. The combined 10 ml medium was centrifuged at 1,000 rpm (200g) for 3 min. The supernatant was discarded by aspiration. The cells were then suspended by 1 ml cell culture medium and seeded in 4 flasks.

American ginseng-treated MIN6 cells

The seeded cells grow into 70% confluence in about 7 days. Various concentrations of American ginseng were used to treat MIN6 cells for different time lengths in cell viability test. When speaking of "treated" and the corresponding "control" cells, it means that the same passage of MIN6 cells were used in these two groups cultured in the same incubator.

In order to obtain a sufficient amount of protein for DIGE experiment, American ginseng extracts were added to the cell flasks in different time points to make sure the cells will be at 70% confluency when harvest. For example, in order to treat the cells for 24 hours, the American ginseng extract were added in day 6, to make sure at the 7^{th} day (when the cells are harvested), the cell confluency is at 70%. By the same logic, in order to obtain 48 hours treated cell, ginseng extract need to be added on the 5^{th} day, and for the 66 hours treated cells on the 4^{th} day.

0.125 mg/ml wild American ginseng and 0.25 mg/ml cultivated American ginseng were prepared by dissolving the American ginseng extract in cell culture medium directly. These ginseng concentrations were selected using cell viability test, which are further explained in section 4.2.4 and section 4.3.1. The solution was filtered through a sterile 0.2 μm syringe filter to exclude bacteria and particles.

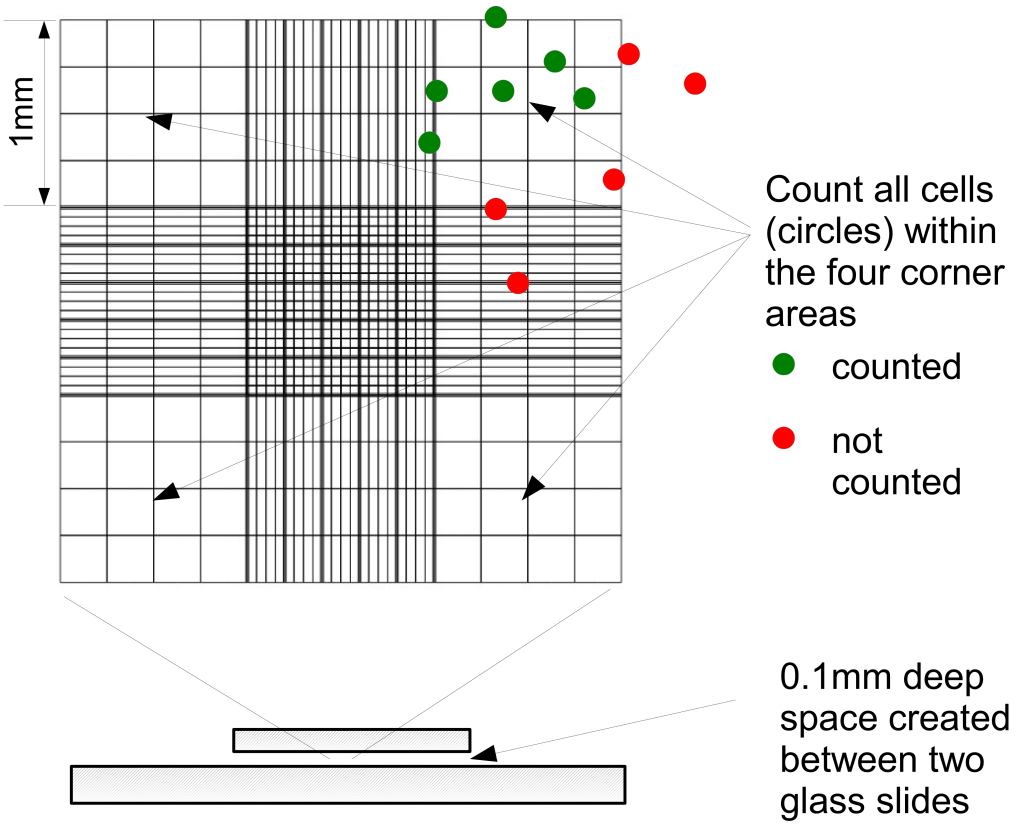


Figure 4.2: *A simplified illustration of a haemocytometer grid and glass slides. Cells that are within or touch the left or top boundary of each primary squares are counted (shown in green), and the cells that are outside or touch the right or bottom are not counted (shown in red).*

4.2.3 Estimation of cell number

For some experiments such as the viability test in section 4.2.4, a defined number of cells are required. A Neubauer haemocytometer (figure 4.2) was used to accurately determine the numbers of cells in a sample. It consists of a heavy glass slide and a cover slip, together creating a 0.1 mm depth space. A 10 μl pipette was used to transfer cell solution into this narrow space due to capillary action.

Under the microscope, it can be observed that the haemocytometer grid on the thick glass slide consists of 9 main squares. Each square is 1 mm wide. The volume of the space under the four main squares can thus be calculated: $1\,\text{mm}^2 \times 0.1\,\text{mm} \times 4 = 0.4\,\text{mm}^3$.

The cells in the 4 main squares on the four corners were used in cell counting. To further define what means "in" the square, an illustration of cell counting is shown in figure 4.2. Cells that are within or touching the left or top boundary of the primary squares are counted(shown in green), and the cells that are outside or touching the right or bottom of the primary squares are not counted (shown in red). Using the information of the cell number and the volume ($0.4\,\text{mm}^3$), the cell concentration of a tested sample can be calculated. Therefore if a certain number of cells are required, the required volume of the sample can be calculated.

4.2.4 CellTiter-Glo® Luminescent Cell Viability Assay

In order to select the highest possible concentration of American ginseng to treat MIN6 cells without causing >% cell death, a cell viability test was performed. CellTiter-Glo® Luminescent Cell Viability Assay (G7570, Promega, Madison, WI, USA) was selected due to its simplicity and precision.

Approximately 10,000 cells were seeded in each well on a 8×12 well plate, except the 26 wells on the edge of the plates, because the medium in the wells on the edge might evaporate more than those in the middle. Therefore only 6×10 wells in the centre of the plate were used. After 24 hours, A series concentration of WAG and CAG were added to each well.

Each sample was prepared with 6 technical replicates and 3 biological replicates. After another 24 hours, the CellTiter-Glo® Luminescent Cell Viability Assay was added to each well. The plate was scanned by a luminescent scanner. The reagent causes cell lysis and by reacting with ATP, it generates a luminescent signal. This signal is directly proportional to the quantity of ATP, thus proportional to metabolically active cell numbers. By calculating the ratio between absorption values of treated and control MIN6 cells, the cell viability percentage was obtained. The procedure of this experiment was following a standard protocol provided by the company(Technical Bulletin, CellTiter-Glo® Luminescent Cell Viability Assay.`http://www.promega.co.uk/resources/protocols/technical-bulletins/0/celltiter-glo-luminescent-cell-viability-assay-pro-\tocol/`).

4.2.5 Protein extraction and purification

Treated and control MIN6 cells were trypsinized from the T75 flasks and washed with PBS solution. This suspension was centrifuged at 13,000 rpm, 0°C for 10 min. The supernatant was discarded and the cell pellet was lysed in a lysis buffer containing 30 mM Tris, 7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 5 μM EDTA and 10 μl/ml Protease *Arrest*™ (G-Biosciences, St Louis, MO. USA). Protease *Arrest*™ is suitable to protect protein by inhibiting protease activities.

This cell mixture was subjected to two quick freeze-thaw cycles with liquid nitrogen and subsequent thawing at room temperature. The protein mixture was centrifuged at 13,000 rpm, 0°C for 20 min. The supernatant was collected and subjected to protein purification using Perfect-FOCUS™ (82021-274, G-Biosciences, St. Louis, MO. USA). This step is to get rid of agents including ionic detergents, salts, lipids, charged polysaccharides, peptides, nucleic acids, enzyme substrates, *etc.*) These agents might change the net protein charge, therefore interfere with the later iso-electric focusing (IEF) experiment(Technical Bulletin, Perfect-FOCUS™, G-Biosciences). The purification process also has the function to concentrate protein samples, since

there will be a minimal concentration requirement (1-5 mg/ml) in the later fluorescence labelling process. The purification procedure involves 11 steps and was performed according to the manufacturer's instruction. Purified protein was dissolved in 30-50 μl lysis buffer for further experiment.

4.2.6 Protein quantification

Protein quantification was performed using Bradford Reagent (B6916, Sigma). Bradford Reagent contains Brilliant Blue G in phosphoric acid and methanol. Protein bonded with Brilliant Blue G shifts its absorption from 465 nm to 595 nm. The quantity of absorption is directly proportional to the quantity of protein present in the linear range from 0.1-1.4 mg/ml.

The "96 Well Plate Assay Protocol" indicated in the technical bulletin of Bradford Reagent was used in this study. A series of protein standards (blank, 0.1, 0.2, 0.4, 0.8, 1.0, 1.2, 1.4 mg/ml) were prepared in a 1/3 diluted lysis buffer (30 mM Tris, 7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 5 μM EDTA and 10 μl/ml Protease Arrest) using bovine serum albumin (BSA). 5 μl of BSA standards and test samples were added in triplicate to separate wells. Bradford reagent stored in the fridge was gently shaken and brought to room temperature before use. 250 μl of Bradford reagent was added to each well that contains protein sample or standard. The 96-well plate was gently shaken for 30 seconds and left at room temperature. After 20 minutes, absorbance was measured using a plate reader (EL 311SX, BIOTEX, UK). A standard curve was obtained based on the series of standard BSA protein concentration and used to calculate the protein concentration of test samples.

4.2.7 Experimental design

70% aqueous methanol extract of wild and cultivated American ginseng were used to treat MIN6 cells for 24, 48 and 66 hours respectively. The control and treated cells were harvested and lyophilized, and their proteins were extracted, purified and quantified according to method described in section 4.2.5. The proteins obtained were labelled according to table 4.1. Details of the dyes are described in section 4.2.8 . Cy2 was used to label

Table 4.1: *Experimental design of CyDye labelling. 22 gels including 44 samples were ran in two batches. Cy3 and Cy5 were used to label test samples, and Cy2 is used to label the internal standard. Pooled sample is made by taking 1/3 of each test sample and mixed together in each experimental batch. Gel 1-12 include 13 batches of cultivated American ginseng-treated cells: 24 hours × 3, 48 hours × 5, and 66 hours × 4. Gel 13-22 include 12 batches of wild American ginseng- treated samples: 24 hours × 5, 48 hours × 3, and 66 hours × 4. Abbreviation: AG, American ginseng*

Gel no.	Cy2	Cy3	Cy5
	Internal standard	Test samples	Test samples
1	Pool	Cultivated AG 24h	Control
2	Pool	Cultivated AG 24h	Cultivated AG 66h
3	Pool	Cultivated AG 48h	Control
4	Pool	Cultivated AG 48h	Cultivated AG 24h
5	Pool	Cultivated AG 66h	Control
6	Pool	Cultivated AG 66h	Control
7	Pool	Cultivated AG 66h	Control
8	Pool	Control	Cultivated AG 48h
9	Pool	Control	Wild AG 24h
10	Pool	Control	Cultivated AG 48h
11	Pool	Control	Cultivated AG 66h
12	Pool	Control	Cultivated AG 48h
13	Pool	Wild AG 24h	Control
14	Pool	Wild AG 48h	Control
15	Pool	Wild AG 48h	Wild AG 66h
16	Pool	Wild AG 66h	Control
17	Pool	Wild AG 66h	Wild AG 24h
18	Pool	Wild AG 24h	Control
19	Pool	Control	Wild AG 24h
20	Pool	Control	Wild AG 48h
21	Pool	Control	Wild AG 66h
22	Pool	Control	Wild AG 24h

the pooled protein samples, which was used as an internal standard across gels for normalisation. Pooled protein was made by taking 25μg from each test sample and mixed together, and 50μg was taken from the mixture as an internal standard used in each gel.

The protein samples were labelled with either Cy3 and Cy5. Protein replicates were avoided to be labelled all by the same dye, in order to avoid bias. For example, there are four biological replicates for "Wild AG 66h" in table 4.1, two of them were labelled with Cy3 in gel 16 and 17, the other two were labelled with Cy5 in gel 15 and 21.

4.2.8 CyDye labelling

CyDye DIGE Fluro minimal dyes are cyanine dyes including Cy3, Cy5 and Cy2, with an NHS ester group which can bond with lysine residues in proteins. This does not significantly alter the pI of the protein because a dye carries a "+1" charge to replace a "+1" charge in the lysine residue. CyDye DIGE Fluor saturation dyes are available for cysteine labelling, but lysine labelling is preferable due to the simpler procedure, and the fact that lysine is one of the most frequent present amino acids in proteins. It is rare that a protein does not have a lysine residue.

The ratio that 50μg protein was labelled by 400 pmol CyDye is to ensure that only 3-5% of protein will receive a tag, and the rest of the proteins will remain unlabelled. Using this ratio, most of the labelled proteins have only one (or two sometimes) lysine residue labelled.

50 μg protein solution was adjusted to pH 8.5 by adding approximately 0.1 μl diluted sodium hydroxide (100 mM). 50 μg protein was labelled with 400 pmol CyDye Fluors. The labelling reaction was performed in an ice bucket in the dark for 30 minutes. Excessive dye was quenched with 1 μl 10 mM lysine in the dark on ice for 10 min. A detailed protocol of the labelling reaction is available on the product booklet and GE Healthcare website (`http://www.gelifesciences.com/webapp/wcs/stores/servlet/product\ById/zh/GELifeSciences-fr/25800861`).

Some important details need to be noted during the experiment. 1μl of

the CyDye working solution which contains 400 pmol CyDye DIGE Fluor was used to label 50 μg of protein. Extra attention should be paid on measuring the volume of 1 μl. An 2μl air displacement auto-pipette (Gilson Pipetman P2 Classic Pipettes, 0.2 to 2UL) and polypropylene pipette tips were used in this experiment. Normally water is used to calibrate the auto pipette, using the weight and the density to calculate the corresponding volume. However, in this case it was found not appropriate to use water to calibrate 1 μl, because more than 1 μl, actually around 1.2 μl CyDye working solution was taken using this water calibrated to 1 μl pipette. 12 μl CyDye working solution was finished after withdrawal 10 times. In order to obtain the correct volume of CyDye working solution, the pipette was adjusted to be able to separate 12 μl CyDye working solution into 12 parts. In this experiment, the pipette volume mark was found to be 0.85 μl to measure 1 μl working solution. The reason might be the density of DMF which was used to dissolve the Cydyes is 944 kg/m^3, lower than that of water which is 1,000 kg/m^3. For air displacement pipettes, a larger dose of liquid is aspirated for lower density solutions. This is caused by the flexible dead air volume along with the earth gravity (Good Laboratory Pipetting Guide, Thermo Scientific Finnpipette Finntip, hermo Scientific). The purpose of noticing this is to make sure the expensive CyDye DIGE Fluor is not over used than necessary. Although 400 pmol CyDye is suggested to label 50μg of protein, it has been reported that 200 pmol CyDye should be enough [144].

4.2.9 Two dimensional electrophoresis

1D-isoelectric focusing (IEF)

Immobiline DryStrip was used in 1D-IEF. This drystrip comes with different pH ranges. When rehydrated, it reswells and absorbs proteins. During the IEF, proteins migrate into positions according to their isoelectric points. pH3-10NL(NL: non linear) was chosen for this experiment, because pH3-10 covers a reasonably large amount of proteins yet gives relatively good separation. Non-linear means the pH range of the 1D strip is not linearly distributed. Instead, the part where most proteins were concentrated which

is between pH5-7 is enlarged comparing to the pH3-5 and 7-10 parts. This allows a better separation.

50μg pooled internal standard protein, 50μg test sample labelled with Cy3, and 50μg test sample labelled with Cy5 (according to table 4.2) were combined and topped up to 450μl using rehydration buffer(30 mM Tris, 7M urea, 2M Thiourea, 4% (w/v) CHAPS, 0.5% IPG buffer, 18 mM DTT and 0.002% bromophenol blue). This buffer was placed in a reswelling tray together with the Immobilized dryStrips pH3-10NL (24 cm) over night (>10 hours), for the proteins to diffuse into the gels in the strips.

The Immobiline DryStrip was rinsed briefly by spraying water using a 1 ml auto-pipette tip and wiped dry. IEF was performed using an Ettan IPGphor II (GE Healthcare) following a standard protocol: 500 V (1 h, step-n-hold), 1000 V (8 h, gradient), 8000 V (3 h, gradient), 8000 V (3.45 h, step-n-hold), 10,000 V (3 h, gradient) and 10,000 V (2.45 h, step-n-hold).

After the IEF, 1D strips were equilibrated using equilibration buffer containing 50 mM TrisHCl (pH 8.8), 6 M urea, 30% (v/v) glycerol, 2% (w/v) SDS, and 0.002% (w/v) bromophenol blue. The equilibration process involves two steps. In the first step, 1D strips were incubated using the equilibration buffer with 1% DTT for 12 minutes. In the second step, 1D strips were incubated using the equilibration buffer with 4% iodoacetamide for another 12 min.

2D-GE

1D strips were rinsed by cathode buffer containing 0.25 M Tris, 1.92 M glysine and 1% (w/v) SDS using an auto-pipette, then carefully sealed into a 18x24 cm 12.5% ployacrylamide gel (28-9374-51, GE Healthcare) with agarose solution (gel buffer with 0.5% agarose and 0.002% bromophenol blue). The second dimension of electrophoresis was carried out as follows: Step A, 1W/gel for 1 hour; Step B, 1.5 W/gel for 15-16 hours until the bromephenol blue lines ran to the bottom of the gels (2-D Electrophoresis: Principles and Methods, Handbook, GE Heathcare).

4.2.10 DIGE gel scanning

DIGE gels sealed in low fluorescent glass plates were wiped clean using Aldrich-Crew® lint free wipers (Z236810, Sigma). This was used to avoid dust showing up in the scanned images. An Ettan DIGE scanner (GE Healthcare, Chalfont St Giles, England) was used to scan the gels. The scanner is equipped with using three lasers with wavelength at 532 nm, 633 nm, and 488 nm. The image scanning process was performed according to the manufacturer's instruction. (Ettan™ DIGE Imager, User Manual, GE Healthcare)

Exposure and emission settings of each dye were set according to the protocol. The pixel size was set to 100 μm which is required for the subsequent analysis using DeCyder™ software. A test scan was carried out on a small area of a random gel in order to optimise the exposure settings. The area between pH 5-7 is usually where the most abundant proteins are present, thus a good indication of the whole gel. If this area is not saturated, then it is likely that no other experimental areas (excluding the edges) is saturated. The exposure time should be as long as necessary yet short enough to avoid any saturation. The maximum pixel value of the scanned image is ideally between 30,000 and 55,000, and must be below 65,000. The same settings should be used across all the gels belong to the same experiment. Following these criteria, exposure settings used through out the current study were: CY2 at 0.6 exposure, CY3 at 0.12 exposure, CY5 at 0.2 exposure.

The edges of the images were excluded during the analysis because they were usually saturated with large amount of proteins outside the detecting range.

4.2.11 Protein statistics

The scanned and cropped gel images were imported into DeCyder™ 2D V. 7.0 software(GE Healthcare, Chalfont St Giles, England), which was designed for the analysis of DIGE . It can normalise the gels according to the internal standards and align the protein spots across images to an automatically selected "master" image.

4.2.12 Protein identification

After statistical analysis of DIGE gels, interesting proteins are often selected for further understandings of pathways and mechanisms. This is done by running a separate preparative 2D gel which contains around 300-450μg proteins for identification purposes. The preparative gel is then subjected to silver staining and matched to the analytical gels. The selected proteins are cut out and subjected to MS/MS analysis. In this thesis, only a few proteins were identified as a preliminary test. Protein identification was carried out by the proteomic service in University of Cambridge.

Silver staining

Preparative 2D gels were prepared for protein identification. CyDye labelling was used for analytical gels and silver staining was used for preparative gels. 450 μg protein was dissolved in 450 μl rehydration buffer and loaded on a 1D Immobiline DryStrip. The experimental conditions of the IEF and 2D-GE remain the same as descried in section 4.2.9. After the second dimension, the gel was removed carefully from the glass plates, and fixed in 45%(v/v) methanol with 1%(v/v) acetic acid overnight. The gel was then washed with Milli-Q water to regain shape overnight. The fixed gel was sensitised with 0.02% (w/v) sodium thiosulfate for 1 minutes, then rinsed with Milli-Q water for 3 times, each time 20 seconds. The gel was stained using 0.2% silver nitrate and 0.02% formaldehyde for 20 minutes, then rinsed with Milli-Q water for 2 times, each time 20 seconds. The stained gel was developed using 3% sodium carbonate, 0.0005% (w/v) sodium thiosulfate and 0.05% formaldehyde for approximately 3 minutes until the desired colour level of the spots was shown, then termination solution which is 5% acetic acid was poured in the gel tank from a corner. The gel was then scanned by an Epson Expression 10000 XL scanner, and a tiff file was obtained. This image was matched with analytical gels, and selected proteins during analysis were cut out and subjected to in-gel digestion and further MS identification.

In gel digestion

Desired spots were cut off using a 10 ml pipette tip in a laminar hood, and placed into a 96-well plate. Briefly, the gel bands were destained, reduced using DTT and alkylated using iodoacetamide and subjected to enzymatic digestion with trypsin overnight at 37°C. After digestion, 10μl of supernatant was pipetted into a sample vial and loaded onto an auto sampler for automated LC-MS/MS analysis.

Orbitrap LC MS/MS

All LC-MS/MS experiments were performed using an Eksigent NanoLC-1D Plus (Eksigent Technologies, Dublin, CA) HPLC system and an LTQ Orbitrap Velos mass spectrometer (ThermoFisher, Waltham, MA). Separation of peptides was performed by reverse-phase chromatography with a flow rate of 300 nL/min and an LC-Packings (Dionex, Sunnyvale, CA) PepMap 100 column (C18, 75 μM i.d. x 150 mm, 3 μM particle size). Peptides were loaded onto a precolumn (Dionex Acclaim PepMap 100 C18, 5 μM particle size, 100 A, 300 μM i.d x 5 mm) from the autosampler with 0.1% formic acid for 5 minutes at a flow rate of 10 μL/min. After this period, the valve was switched to allow elution of peptides from the precolumn onto the analytical column. Solvent A was water + 0.1% formic acid and solvent B was acetonitrile + 0.1% formic acid. The gradient employed was 5-50% B in 45 minutes.

The LC eluent was sprayed into the mass spectrometer by means of a New Objective nanospray source. All m/z (mass number/charge number) values of eluting ions were measured in an Orbitrap Velos mass analyser, set at a resolution to 30,000. Data dependent scans (Top 20) were employed to automatically isolate and generate fragment ions by collision-induced dissociation in the linear ion trap, resulting in the generation of MS/MS spectra. Ions with charge states of 2^+ and above were selected for fragmentation.

4.2.13 Data processing and database searching

Post-run, the data was processed using Protein Discoverer (version 1.2, ThermoFisher). Briefly, all MS/MS data were converted to mgf (text) files. These files were then submitted to the Mascot search algorithm (Matrix Science, London UK) and searched against Uniprot mouse database, using a fixed modification of carbamidomethyl and variable modifications of oxidation.

4.3 Results and discussion

4.3.1 Viability test

Figure 4.3 shows the result of the cell titre viability test (section 4.2.4) of MIN6 cells treated by a series of concentrations of 1, 0.5, 0.25, 0.125 and 0.063 mg/ml of wild American ginseng (WAG), and 0.5, 0.25, 0.125 and 0.063 mg/ml cultivated American ginseng (CAG) during 24, 48 and 66 hours. These values of all the treated cells were normalised to the corresponding control cells. Figure 4.3 shows that a high American ginseng concentration and a long treatment time are the two key factors causing cell death.

For MIN6 cells treated with 1 mg/ml WAG, at 24 hours 19% of the cells were dead comparing to the controls. At 48 hours, 93% of the cells were dead and in 66 hours, 97%. This death rate is much higher than the 0.5 mg/ml WAG treated cells. It could be that the high concentration of WAG promoted necrosis of the MIN6 cells. Typical morphological changes for necrosis was visible under microscope: cells appeared swollen up like a balloon, with the cell nucleus clearly visible [145].

For cells treated with 0.5 mg/ml wild American ginseng extract, 124% of cells at 24 hours were alive, this means the cell number was increased by 24% compared to the untreated controls. However, after 48 hours the cell viability level dropped to 48%, and after 66 hours it dropped to 25%. This means that 0.5 mg/ml wild American ginseng extract promoted cell growth in the first 24 hours, then promoted cell death.

The same case happened in 0.5 mg/ml cultivated American ginseng-treated cells, higher than 60% cell deaths occurred in 48 and 66 hours treatment. The

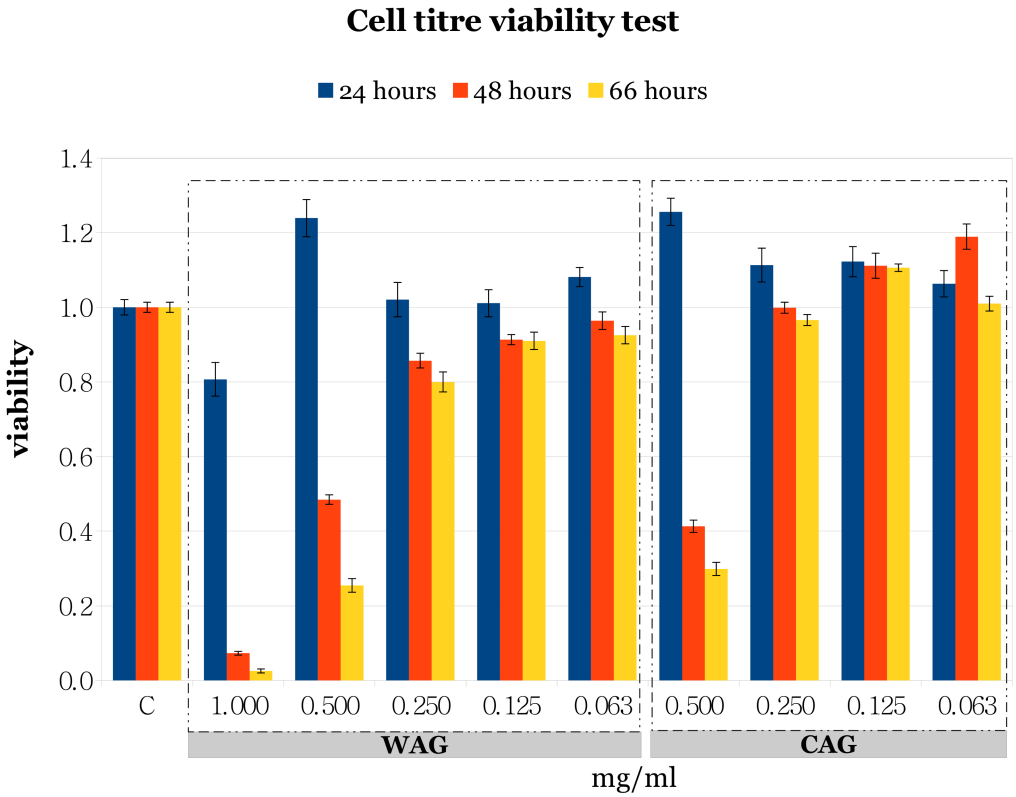


Figure 4.3: *Cell titre viability test of MIN6 cells treated by a serial concentration of 1, 0.5, 0.25, 0.125 and 0.063 mg/ml of wild American ginseng (WAG), and 0.5, 0.25, 0.125 and 0.063 mg/ml cultivated American ginseng (CAG) during 24, 48, 66 hours. Control cells are untreated MIN6 cells. The y-axis is the viability level where "1" means 100% alive. Error bars are the standard deviations based on 6 replicates of each sample.*

cell death caused by high concentration wild/cultivated American ginseng may be related to the ginsenosides. Rb1 and Rg1 are two of the main ginsenosides in American ginseng extract [146], and have been found to enhance MIN6 cell viability throughout a four hours treatment compared to control. This might explain the increased cell growth that occurred in this experiment. On the other hand, ginsenoside Rb1 and Rg1 are saponins, which may be detrimental to membrane integrity and cell viability because of their detergent properties [140]. This might explain cell death under treatment with high concentrations.

The viability of MIN6 cells treated with 0.25 mg/ml WAG maintained the same as the control after a 24 hours treatment, but was only around 80% after 48 and 66 hours treatment time. Therefore, about 20% cell dead between 24 hours and 66 hours treated by 0.25 mg/ml wild American ginseng.

Comparing to WAG, In the case of 0.25 mg/ml, the viability level of CAG treated MIN6 cells were 110%, 100% and 98% respectively, higher than the WAG treated cells which were 102%, 86%($p < 0.05$), and 80% ($p < 0.05$).

When treated by lower than 0.25 mg/ml CAG, MIN6 cells have a viability value higher than 95% after 24, 48, and 66 hours respectively. However for wild American ginseng, the higher than 95 % viability level only occurred when using a concentration below 0.125 mg/ml after 24, 48, and 66 hours respectively. Therefore wild American ginseng is "stronger" than cultivated American ginseng when using the concentration between 0.125-0.25 mg/ml on MIN6 cells. This finding coincide with the metabolic research in chapter 3 of this thesis. Wild American ginseng batch 17 and cultivated American ginseng batch 8 from chapter 3 were used in this viability study. Wild American ginseng was shown to have more glucose but less sucrose than cultivated American ginseng as shown by ^{1}H-NMR, refer to section 3.3.3, and possibly more ginsenosides as well, refer to section 3.3.4. MIN6 cells response to glucose by secreting insulin. However, considering the very small quantity of WAG added, the glucose content is unlikely to make a difference to MIN6 cells[136]. The exact interaction between ginsenosides and MIN6 is unknown. There are also many primary metabolites present in ginseng extract, and they may stimulate cell growth or induce cell death in a different

composition. More research is needed to explain the exact mechanism.

The highest possible concentrations that caused less than 10% cell death were chosen for further experiments. Therefore, 0.125 mg/ml for wild American ginseng and 0.25 mg/ml for cultivated American ginseng were chosen for treatment concentrations.

4.3.2 A brief analysis of the DIGE images

Figure 4.4 is a typical image of a DIGE gel of MIN6 cells, treated by cultivated American ginseng. Each spot represents one protein. Horizontally the proteins are between pH 3-10 NL (non-linear, with a smaller slope between pH 5-7). Vertically the molecular weights are logarithmically arranged from 12 to 120 kDa. This image was set to black and white and the contrast and brightness levels were manually adjusted to provide the best possible visualisation. When a higher contrast and a higher brightness settings were used, the smaller protein spots tend to be invisible, and with the opposite settings the bigger protein spots tend to join with the adjacent spots and form a back smear. This only changes the visualisation and does not affect the stored image because the data of protein spots are stored as volume numbers, which are the absorption value of each spot. It is not possible to analyse the complex DIGE images by visualisation. The function of the visualisable image is to manually check if the alignment has been done correctly by the software and spots were correctly selected. Sometimes when two spots are almost overlapping each other, the software might consider them as the same spot. This only happens in rare cases, that it can only be confirmed to be one or two peaks by comparing it with the same spot in contain replicates of that protein sample. In this case, adjacent proteins can be manually split or merged as needed.

The scanned images of DIGE protein profile (a sample is shown in figure 4.4) were subjected to the batch analysis using DeCyder™ 2-D Differential Analysis Software v7.0. The edges of the DIGE image were cropped before analysis, because they are highly saturated with proteins that do not belong to the analytical range (pH 3-10, molecular weight between 12-120 kDa),

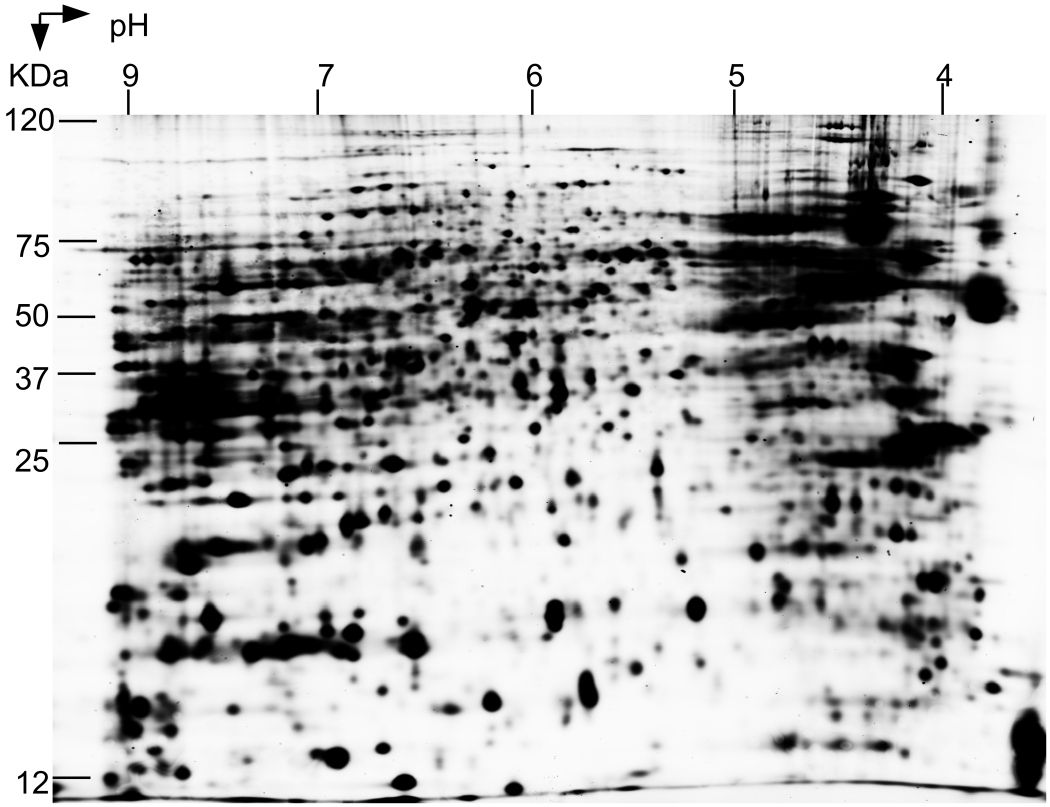


Figure 4.4: *Representative 2D-DIGE image of proteins extracted from MIN6 cells treated by American ginseng. This specific image is MIN6 treated with cultivated American ginseng for 66 hours labelled with Cy5. Horizontally the proteins are between pH 3-10 (non-linear, with a smaller slop between pH 5-7, referencing to the pH slope graph offered by GE). Vertically the molecular weights are logarithmically arranged from 12-120 kDa referencing to a molecular weight marker. This image was set to black and white and the contrast and brightness level were manually adjusted to provide the best possible visualisation.*

thus will affect the normalisation process during the calculation.

Each of the gels were aligned to the master gel, which was automatically selected by the software. The same proteins might not show up in exactly the same place in different images, because of the unavoidable distortion, curving of the gels. This can happen if there is a bubble trapped in between the 1D IPG strip and the 2D gel, after the 2D gel is developed, the whole vertical line under that bubble will be shifted up a little, therefore cause a curve horizontally. It is difficult to avoid this kind of curve during 2D protein development, because the top of the 2D gel is not usually perfectly flat, therefore is not perfectly contacted with the 1D strip. Thus it is important to perform alignment. The same protein across gels were automatically aligned according to their relative positions. The protein spots labelled by Cy3 and Cy5 were normalised to the corresponding internal standard protein labelled by Cy2, within the same gel, in order to eliminate gel-to-gel variance.

Table 4.2: *A summary of number of proteins in DIGE gel of CAG and WAG treated and control MIN6 cells using DeCyder software. total: total number of protein detected. up-reg: number of up regulated protein.*

Treatment	time	total	up-reg.	down-reg.
CAG	24h	3563	1705	1579
	48h	3563	1382	1553
	66h	3563	1600	1574
WAG	48h	3504	1559	1539
	66h	3504	1519	1432

As indicated in table 4.2, there are 3,563 proteins detected in the cultivated American ginseng-treated and control MIN6 cells, and 3,504 proteins in the WAG treated and control MIN6 cells.

The data for treatment with WAG for 24 hours was not available, because less than 75 μg protein was collected after the cell lysis process, which is not enough to perform DIGE (minimum 75 μg).

The average of protein abundance of one treated protein among replicate gels is divided by the average of protein abundance of that protein in control gels. Thus a ratio of treated to control was obtained. As a brief calculation,

without using t-test, and any ratio above 1.01 is considered up-regulated and any below -1.01 were considered down-regulated, there are 1,705 protein up-regulated, and 1,579 proteins down-regulated out of 3,563 proteins in 24 hours treatment with CAG. This table allows an overview of the data obtained. However, in order to understand the changes of protein expression levels before and after treatment, the data were analysed using student t-test. The ratios of the abundance of the proteins in the treated MIN6 to the controlled MIN6 cells were calculated. The t-test P value < 0.05 and a ratio $>$ 1.2-fold and <-1.2-fold were used as a selection criteria. Besides, the protein need to appear in more than 90% of all the gels.

When analysing thousands of data points (protein expression levels), p=0.05 means for every 1000 proteins there are 50 "statistically significant" proteins can not be ruled out by the null hypothesis, which means that the expression changes of these 50 proteins happened by randomness. One way to solve this problem is to correct false discovery rate (FDR) on p values. Correction methods include the Benjamini-Hochberg and the Bonferroni corrections[147]. The basic theory to apply these corrections is to multiply the p values by 3,000 (the number of protein analysed. However, this method is often found too stringent in DIGE analysis. When applied, often few proteins were found significant, because of the inherent variations in 2D-GE method, and usually not many replicates are available due to the expenses. Therefore few publications reported the FDR corrected p values.

To reduce the effect of false discovery, all the proteins that meet the initial selection criteria were manually checked for validity, in order to exclude spots due to dust (appeared as a very sharp peak), very tiny or overlapped proteins, which are difficult to calculate and be distinguished from the neighbouring proteins.

When a wide pH range (3-10) was used, the proteins are very crowded in the top part of the gels and in the anode area, resulting the edge of the proteins difficult to define. Sometimes it is difficult to know if the proteins adjacent to each other are different proteins or the post-translational modifications (PTMs) of the same protein (further discussed in section 4.3.3). When proteins are too crowded, it is also difficult to perform alignment on

them, because their relative position are difficult to define. Proteins present in small amount are especially affected because they can easily hide in the shoulder of a large protein spot.

A narrow pH range such as 4-7, or even one unit pH range such as 4-5, 5-6, can be used to achieve a high resolution, but less proteins can be included. Therefore this is a trade-off.

However, as this chapter is only an initial study on American ginseng-treated MIN6, a wider range of proteins were desired to be included. As a matter of fact, some proteins have been found to be responsive to the treatment, such as protein 2659 (figure 4.7), with a theoretical pI=8.15 which would otherwise not show up in the narrow range pH 4-7 gel.

4.3.3 Issue with protein post-translational modifications

The effect of single or multiple post-translational modifications (PTMs) can cause one protein to exhibit slightly different isoelectric points and molecular weights on a 2D gel[148]. Therefore, multiple adjacent spots maybe the same protein with different PTMs. This may be confirmed with protein identification using mass spectrometry in future studies.

If the adjacent spots are all increasing (or decreasing), this means that either several proteins are up (or down)-regulated, or PTMs of the same protein are increasing(or decreasing), thus only one protein is up(or down)-regulated. However, if the adjacent proteins show different regulation trends, for example one increases but the one next to it decreases, it might be different proteins with different regulation trends, or PTMs of the same protein shift into each other. In the later case, the expression level of the protein might not change. Figure 4.5 and figure 4.6 are two examples.

Adjacent spots show different regulation trends

Figure 4.5 shows an examples that one spot was up-regulated while its neighbouring spot was down regulated before and after treatment. If these spots

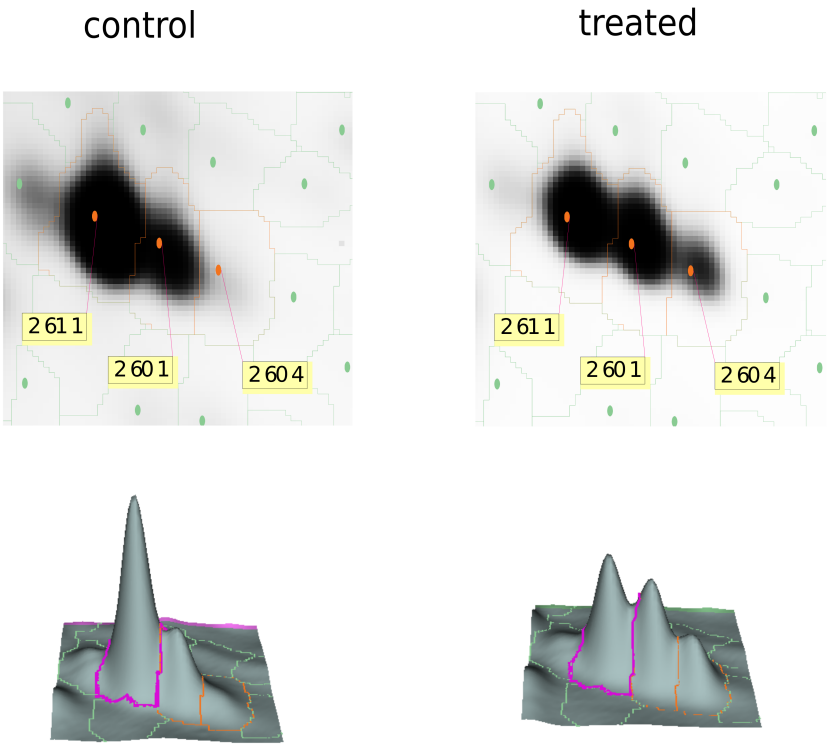


Figure 4.5: *Representative 2D and 3D DIGE images of three adjacent spots 2611, 2601 and 2604 shift into each other after treated with American ginseng. This specific image is protein extracted from MIN6 control and treated with cultivated American ginseng for 48 hours.*

were measured as a group, their expression levels remain the same. These spots are likely due to PTMs.

If this was not considered during the analysis, it is likely to appear as a false discovery. For example, spot 2611 was decreased by 60% in 48 hours after treatment, with a *t*-test p value 0.017. It seems to be a significant change. However, its neighbouring spot 2604 increased by 44% (p=0.23). With a close inspection, it was found that even in replicate gels, spot 2611 and its neighbouring spots move into each other. After adding up all three spots, it was found that the group expression decreased by only 12% before and after treatment. The three spots are likely to be different PTMs of one single protein. If that is the case, the addition of American ginseng extract only promoted the post-translational modification of that protein. Again this is something that can be confirmed by MS.

Adjacent spots did not change

As indicated in figure 4.6, spot M1990 is down regulated by 46% with a very small p value of 0.00033. However the amount of this spot is very low and it is located in between two proteins with high intensity. These two proteins did not change after treatment (with a T/C ratio=1.01). It is likely that M1990 is not an individual protein, but a PTM form of the neighbouring proteins. This is why manual inspection was performed for all the proteins that passed the initial selection criteria.

4.3.4 Expression change of protein M2659 as an example

In figure 4.7, protein M2659 was selected as an example showing protein expression changes with wild and cultivated American ginseng treatment for 48 hours. Protein M2659 has a master number 2642 in wild American ginseng-treated MIN6 cells, and master number 2659 in cultivated American ginseng. Master number is generated by the DeCyder software according to the location for each spot detected, started from number 1. Because proteins extracted from WAG and CAG treated cells were analysed in two

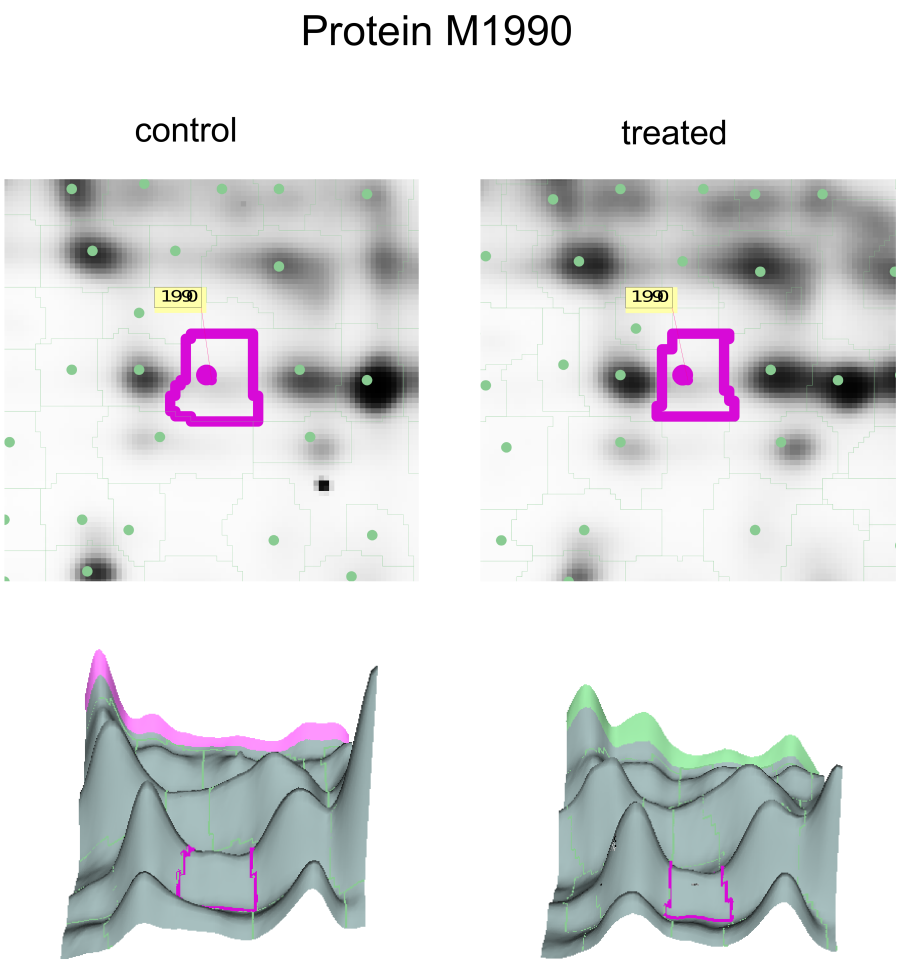


Figure 4.6: *Representative 2D and 3D DIGE images of protein M1990 before and after treatment. This specific image is protein extracted from MIN6 control and treated with cultivated American ginseng for 48 hours.*

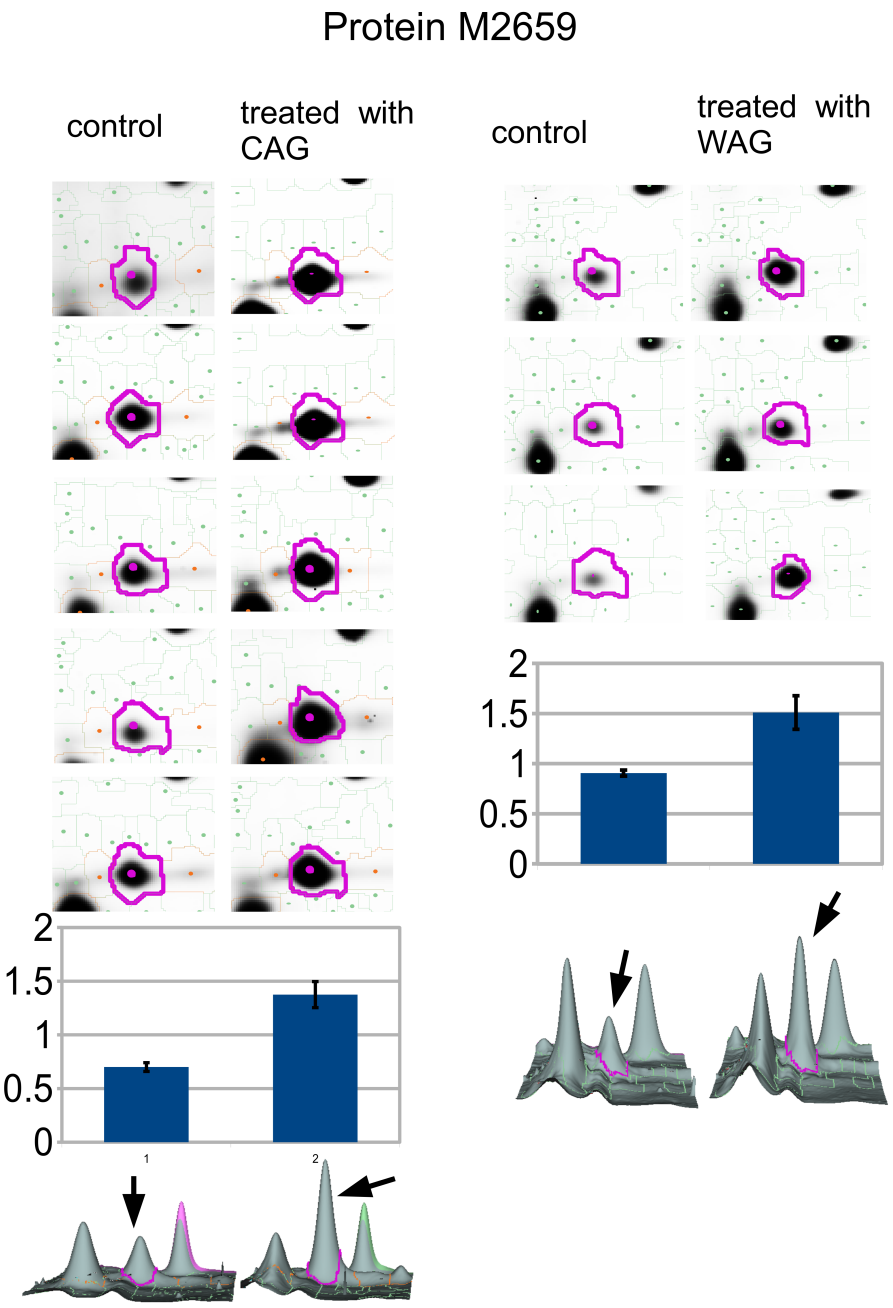


Figure 4.7: *Expression change of protein M2659 treated by cultivated (CAG:left) and wild (WAG: right) American ginseng for 48 hours. Top: View of protein spots in 2D gel. Protein M2659 were circled in purple lines. 5 replicates of MIN6 treated with cultivated American ginseng and 3 replicates with wild American ginseng were shown. Middle: The average standard abundance of replicates with standard error bars. t-test $p<0.05$. Bottom: 3D view of protein M2659, arrow pointed.*

batches, the master numbers for the same protein are different in each batch. However a manual correlation can be made according to the relative positions of the proteins. That is the reason in this study each protein has two master numbers when mentioning the WAG and CAG comparison.

5 replicates of protein profiles of control and treated MIN6 cells with cultivated American ginseng were available, and 3 replicates for wild American ginseng were available. Highlighted spots represent protein M2659 in replicate DIGE gels. An up-regulated 1.96-fold (t-test p=0.00028) change was observed for cultivated American ginseng and an up-regulated 1.73-fold (t-test p=0.0027) change was observed for wild American ginseng in 48 hours.

This result means that M2659 responded to both cultivated and wild American ginseng treatment, but more strongly (error bars of the ratios do not overlap) to 1 mg/ml cultivated than 0.5 mg/ml wild American ginseng treatment.

In order to further understand how does protein M2659 respond to cultivated American ginseng ginseng treatment over time, a plot consisting of the standard abundance of M2659 at 24, 48, and 66 hours is shown in figure 4.8.

Figure 4.8 consists of 3 sets of data, each representing the standard abundance of protein M2659 in control and treated MIN6 cells in 24, 48 and 66 hours respectively. 1.77, 1.96 and 1.46 fold increased protein expression were found at 24, 48 and 66 hours. Among those, the 2D views of 48 hours control and treated proteins were shown on the left hand side of figure 4.7.

The ratios of treated to control proteins are plotted in the later discussions in the form of three columns, instead of plotting each of the replicate proteins. If only the ratio is plotted, the error bar of the ratios must also be plotted, based on the error bar of the control and treated protein abundance. The next section explains this.

Calculation of protein expression changes

The ratio of protein standard abundance of treated cells to the corresponding control cells were calculated and unpaired t test was used, because each sample come from a different flask. The standard error of the ratio is calculated

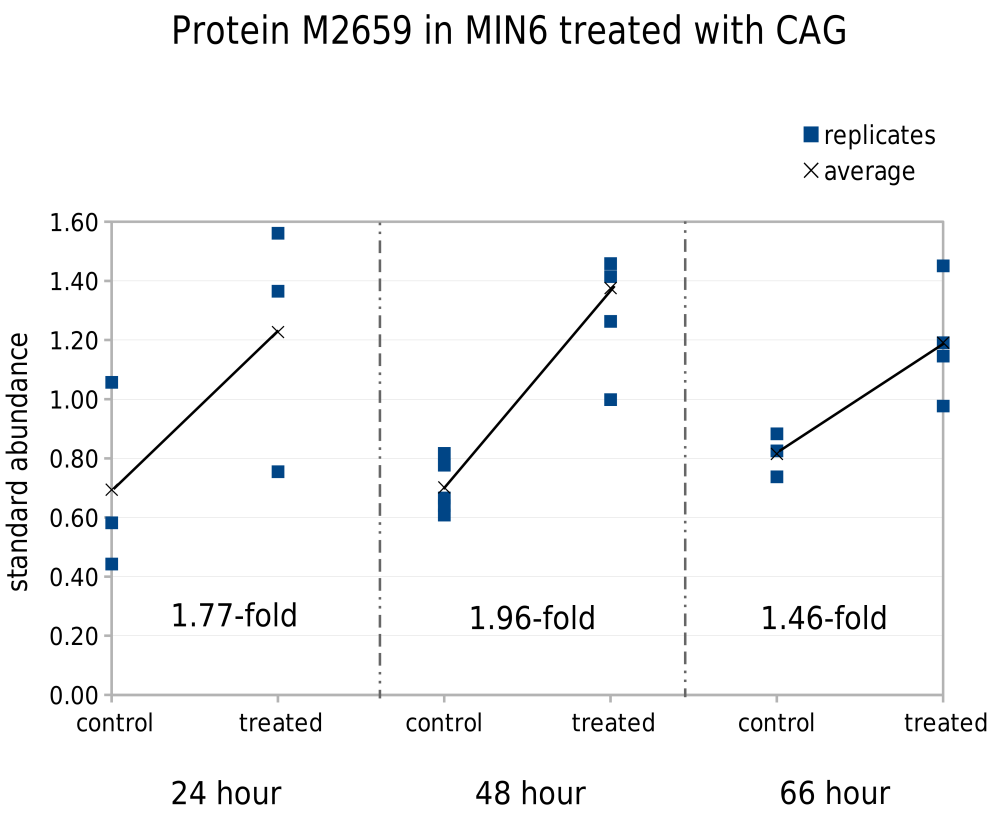


Figure 4.8: *Plot of standard abundance of protein M2659 between CAG treated and control MIN6 in a period 24 hours (3 replicates for both control and treated MIN6 cells), 48 (5 replicates for both control and treated) and 66 hours (3 replicates for control, 4 for treated) treatment.*

as in equation 4.1

$$St.err.Ratio = \sqrt{\left(\frac{st.err.T}{Aver.T}\right)^2 + \left(\frac{st.err.C}{Aver.C}\right)^2} \tag{4.1}$$

The abbreviations are: st.err.: standard error. Ratio: Ratio of protein standard abundance of treated to control cells. Aver.: Average. T: protein standard abundance of treated cells. C: protein standard abundance of control cells.

4.3.5 Effect of cultivated American ginseng on protein expression of MIN6 cells

In order to study the protein expression changes of MIN6 cells caused by the extract of American ginseng, some proteins were selected based on the criteria that on one or more treatment times, the expression of the protein was up- or down-regulated after treatment more than 1.2-fold, with a t-test p value less than 0.05, and the protein must appear in more than 33 of the 36 gels. After the initial selection, 250 proteins out of 3563 meet the criteria.

The selected protein were then manually checked, and any fault detections such as half proteins on the edge of the gel, dusts and PTMs were excluded. So 73 out of the 250 proteins were finally selected and analysed. Among those, 48 proteins were significantly up-regulated, and 35 proteins were significantly down regulated.

Based on the trend of expression changes with time, the selected proteins were shown in 6 graphs in figure 4.9, 4.10 and 4.11. The changed ratio of standard abundance of selected proteins extracted from control and treated MIN6 cells were plotted.

A box drawn around some proteins using dotted lines means the included proteins are adjacent and showed similar trend of regulation after treatment, and they are potential PTMs of the same protein.

In the top graph of figure 4.9, proteins are up-regulated in all three time points. Some proteins were up-regulated significantly over 50% in 48 or/and 66 hours, including 2344, 2659, 2642, 2652, 2110. This significant large

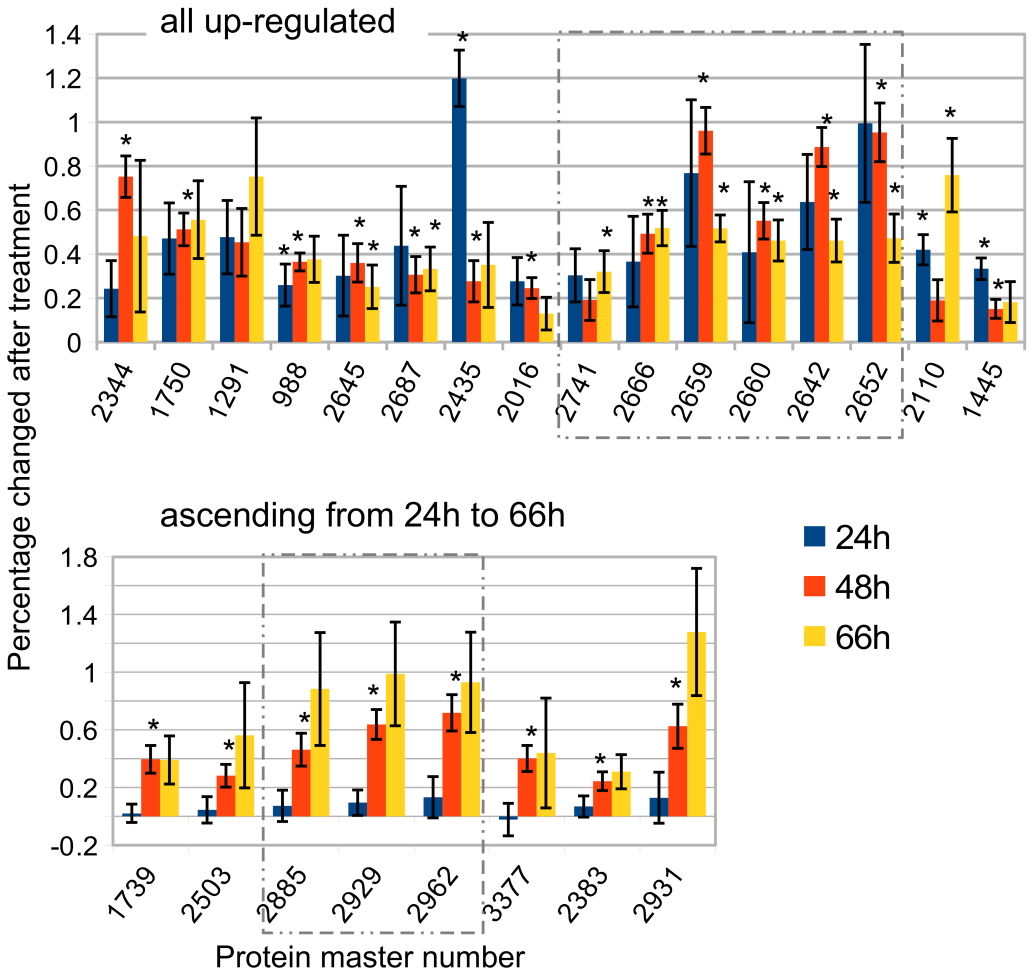


Figure 4.9: *The changed ratio of standard abundance of selected proteins extracted from control and treated MIN6 cells. MIN6 cells were treated with CAG in a period 24, 48, 66 hours. Top figure: An up-regulating trend is shown in all 24, 48 and 66 hours. Bottom figure: Where the ratio goes up above 1.2 in 48 and 66 hours treatment, but not showing a strong tendency of increasing in 24 hours. The figures shown here are with a general trend of up regulating and reaching its peak at 66 hours. The X-axis indicates the master number of the proteins. The Y-axis is ratio of standard abundance of protein between treated and control samples. Error bars are calculated according to equation 4.1 based on 3 control and 3 treated samples for 24 hours, 5 control and 5 treated samples for 48 hours, 4 control and 3 treated samples for 66 hours. * sign on top of the bar means that the p value from the t-test for that data is less than 0.05, and the ratio between control and treated is <-1.2 and >1.2. A box drawn around some proteins using dotted lines means the included proteins are adjacent and showed similar trend of regulation after treatment, and they are potential PTMs of the same protein.*

amount of changes indicate some proteins are especially affected by CAG treatment. Identification of those proteins might assist further understanding of the mechanism of how American ginseng affects pancreatic β cells.

For protein 2435, in 24 hours the expression increased more than 120% (p=0.0037), and in 48 and 66 hours it increased only 28% (p=0.28) and 35% (p=0.16). This could be that the protein responded immediately when contacted the American ginseng, then it slowed down as time passes.

For protein 2110, it increased 76% (p=0.0024) in 66 hours, 42% (p=0.0076) and 19% (p=0.067) in 24 and 48 hours respectively. The irregular changes of this protein is unclear but may be related to the cell growth cycle.

There are some proteins very close to each other and form a tight group. For example 2741, 2666, 2659, 2660, 2642, and 2652, these 6 proteins are grouped together in the same area with a theoretical molecular weight of 22,600 kDa and pH 7.4-8.3.

A group of up regulated proteins are shown in the bottom of figure 4.9. The general regulation trend for this group of proteins after treatment is increasing from 24 hours and reached their highest levels in 66 hours. Proteins 2885, 2929, and 2962 are adjacent spots and the same regulation trends were observed between them.

In the top graph in figure 4.10, the expression levels of proteins treated for 24 and 48 hours were increased significantly, but in 66 hours there is no significant change between control and treated proteins. Proteins 805, 798, 791, 877, 863, and 853 are adjacent and a similar regulation trend was observed between them.

In the bottom graph in figure 4.10, only proteins treated for 48 hours were significantly up regulated, but proteins treated for 24 and 66 hours were not changed significantly after treatment. Among them, proteins 1779, 1785, 1799 and 1811 are adjacent. They were all increased over 30% after treatment in 48 hours. For these proteins, at 66 hours treatment their expression levels were decreased. The error bars for these proteins are large, this could mean that the protein expression level started to drop down around 66 hours.

In the top graph in figure 4.11, down-regulated proteins are listed. Protein 2862 and 3172 were significantly decreased by 96% and 111% after 33 hours

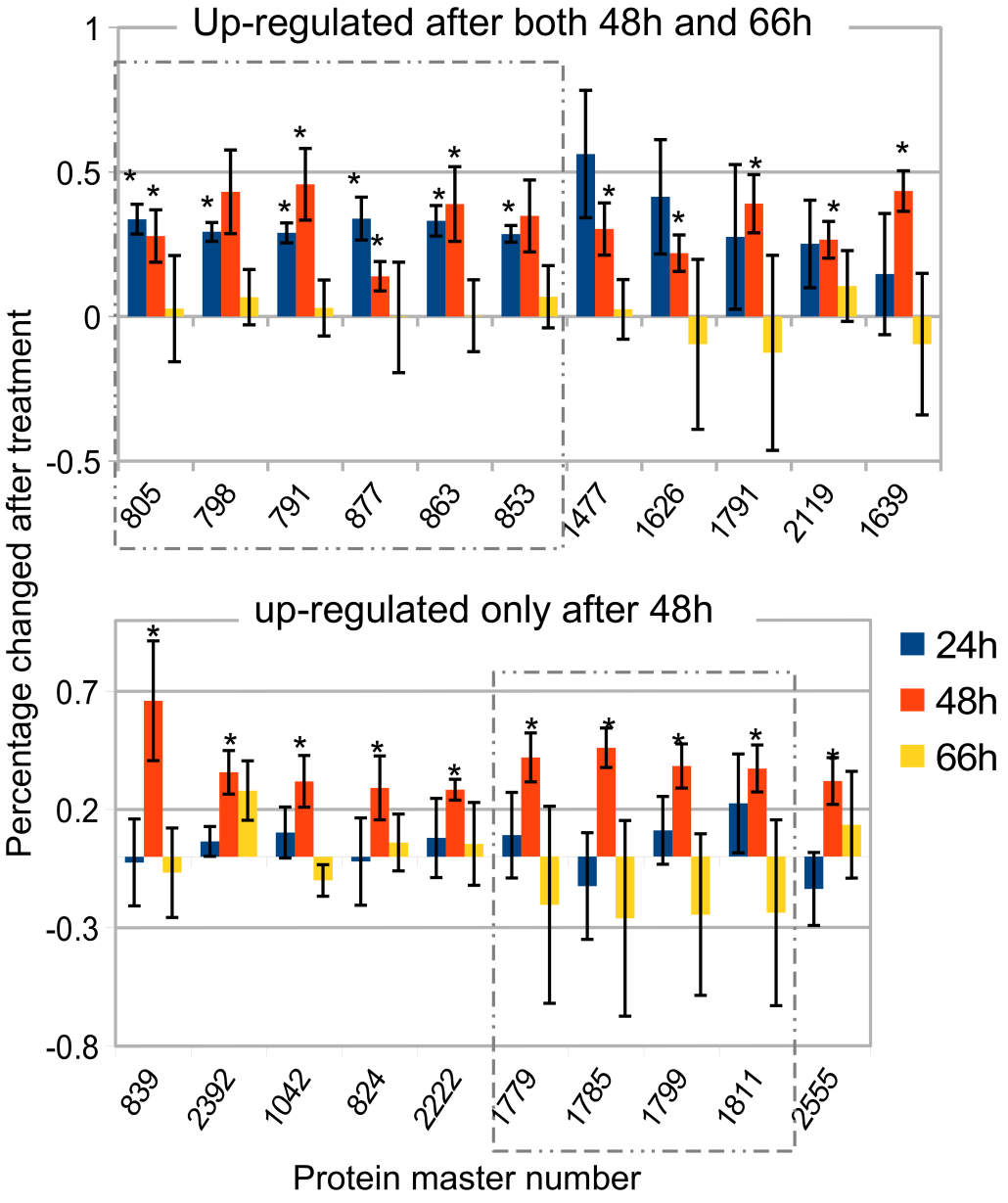


Figure 4.10: *The changed ratio of standard abundance of selected proteins extracted from control and treated MIN6 cells. MIN6 cells were treated with CAG in a period 24, 48, 66 hours. Top figure: where the protein up-regulated in 48 and 66 hours treatment. However did not change in 66 hours. Bottom figure: protein up-regulated only after 66 hours treatment, but not after 24 or 66 hours. The X-axis indicates the master number of the proteins. * Specifications are the same as for figure 4.9.*

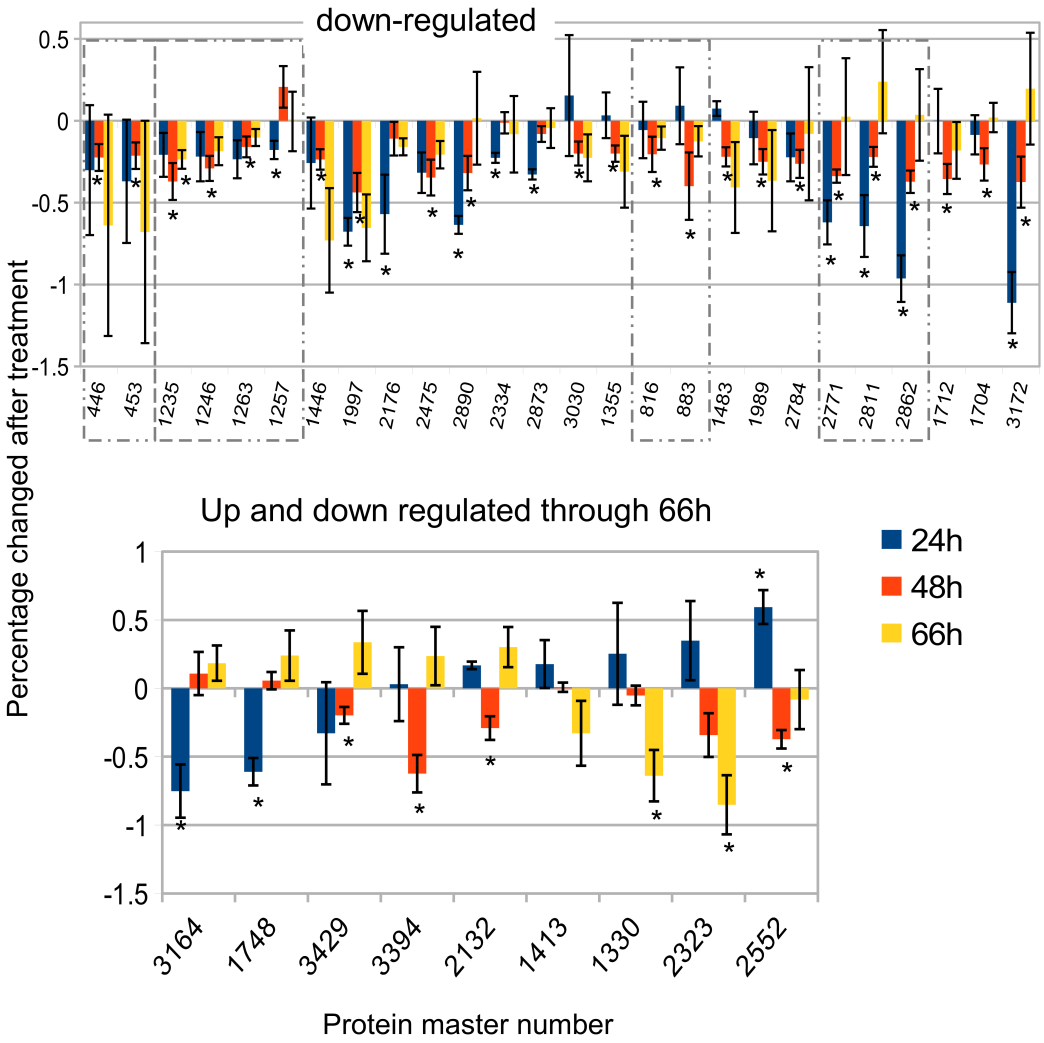


Figure 4.11: *The changed ratio of standard abundance of selected proteins extracted from control and treated MIN6 cells. MIN6 cells were treated with CAG in a period 24, 48, 66 hours. Top figure: Proteins were down regulated in one or more time points of 24, 48 and 66 hours. Bottom figure: Proteins showed different regulation trend in 24, 48 and 66 hours. The X-axis indicates the master number of the proteins. The Y-axis is ratio of standard abundance of protein between treated and control samples. Error bars are calculated according to equation 4.1 based on 3 control and 3 treated samples for 24 hours, 5 control and 5 treated samples for 48 hours, 4 control and 3 treated samples for 66 hours. * sign on top of the bar means that the p value from the t-test for that data is less than 0.05, and the ratio between control and treated is <-1.2 and >1.2. A box drawn around some proteins using dotted lines means the included proteins are adjacent and showed similar trend of regulation after treatment, and they are potential PTMs of the same protein.*

.

treatment. Proteins 446, 453 belong to one group, 1235, 1246 1263 and 1257 belong to one group, 816 and 883 belong to one group. 2771 2811 and 2862 belong to one group.

In the bottom graph in figure 4.11, most of the proteins significantly decreased in one time point, either 24, 48 or 66 hours. For example, protein 3164 decreased significantly by 75% (p=0.049) in 24 hours and protein 3394 decreased significantly by 62% (p=0.0086) in 48 hours. However, the proteins in this group also have one or more time points down regulated, for example, protein 2552 is significantly up-regulated in 24 hours treatment, but significantly decreased in 48 hours. Not all of the up-regulation trends shown in this graph are with great significance, but it shows that those proteins responded to the stimuli differently over time.

4.3.6 Effect of wild American ginseng on protein expression of MIN6 cells

The data in this section for wild American ginseng underwent identical analysis as for cultivated American ginseng. Selected proteins are presented in figure 4.12 and figure 4.13. The first figure shows up-regulated proteins after treatment, and the second figure shows the down-regulated proteins.

On the left of figure 4.12, proteins 2389, 808, 456, 461, 752, 2046 and 1732, have a higher ratio of treated to control protein expression level in 66 hours than in 48 hours. However, on the right part of the figure, an opposite trend is shown in protein 250, 2321, and 235, that the changed protein expression ratio of treated to control is higher in 48 hours than in 66 hours. This may indicate that for the proteins on the left part of the figure, a longer treatment time induces a larger increase of protein level.

Protein 461 and 752 increased more than 70% after treatment for 66 hours; protein 2589 and 2642 increased more than 70% after treatment for 48 hours.

In figure 4.13, some down-regulated proteins are summarised. Protein 3439 was down-regulated by 200% after 48 hours treatment, and after 66 hours, it only went down by 20%. Therefore the most changes occurred

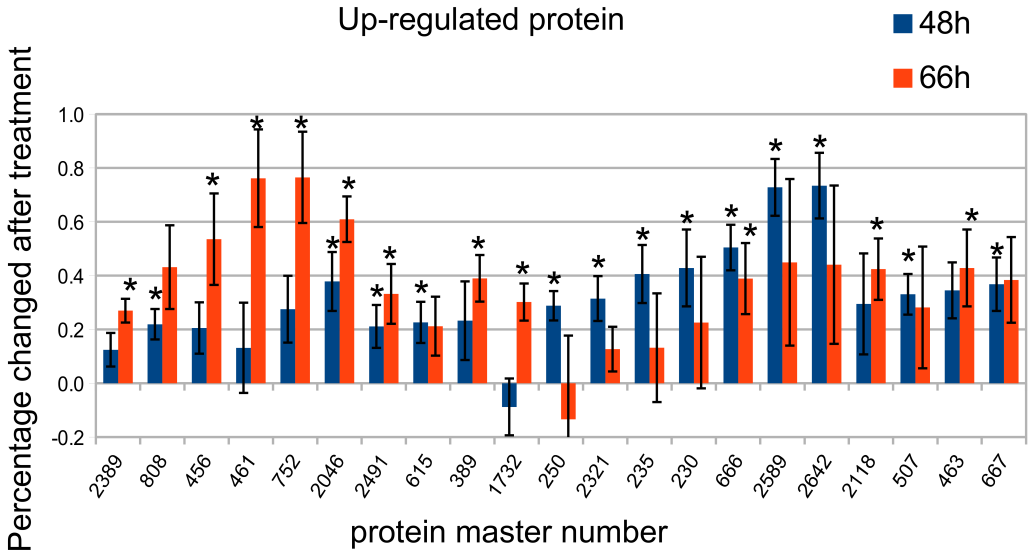


Figure 4.12: *The changed ratio of standard abundance of selected proteins extracted from control and treated MIN6 cells. MIN6 cells were treated with wild American ginseng in a period of 48 and 66 hours. The proteins in this group are up-regulated after treatment. The X-axis indicates the master number of the proteins. The Y-axis is ratio of standard abundance of protein between treated and control samples. Error bars are calculated according to equation 4.1 based on 5 control and 3 treated samples for 48 hours, 4 control and 4 treated samples for 66 hours.* sign on top of the bar means that the p value from the t-test for that data is less than 0.05, and the ratio between control and treated is <-1.2 and >1.2.*

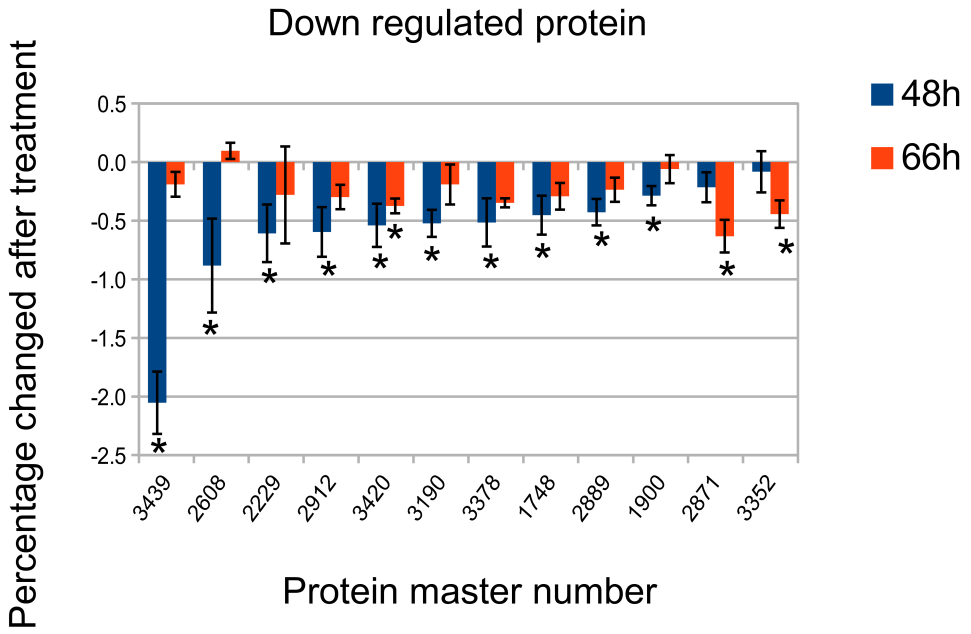


Figure 4.13: *The changed ratio of standard abundance of selected proteins extracted from control and treated MIN6 cells. MIN6 cells were treated with wild American ginseng in a period of 48 and 66 hours. The X-axis indicates the master number of the proteins. The proteins in this group are down-regulated after treatment. The Y-axis is ratio of standard abundance of protein between treated and control samples. Error bars are calculated according to equation 4.1 based on 5 control and 3 treated samples for 48 hours, 4 control and 4 treated samples for 66 hours. * sign on top of the bar means that the p value from the t-test for that data is less than 0.05, and the ratio between control and treated is <-1.2 and >1.2.*

around 48 hours. However, the same protein did not change after treatment with cultivated American ginseng in any of the treated time length. Proteins like this might be used to distinguish treatment with wild and cultivated American ginseng.

The mechanisms of the protein changes are not clear unless further identification is carried out.

4.3.7 Change of protein expression between treatments by wild and cultivated American ginseng

Two batches of DIGE gels with proteins extracted from MIN6 treated by wild and cultivated American ginseng were analysed separately, and the selected proteins were matched and compared in table 4.3, and plotted in figure 4.14 (re-arranged).

Most of the proteins showed similar regulation trend treated with both cultivated and wild American ginseng. However, the amount of changes of expression levels is sometimes different.

Some proteins shows a similar regulation trend between wild and cultivated American ginseng-treated MIN6 cells, including protein 2659 in cultivated American ginseng, up-regulated by 1.96 in 48 hours, with a t-test p value = 0.00028. When treated by wild American ginseng (master number 2642), upregulated by 1.73 in 48 hours, with a p value 0.0027. Protein 2652 in cultivated American ginseng, up-regulated by 1.95 (p=0.00078) in 48 hours, when treated by in wild American ginseng (protein master number 2589), up-regulated by 1.73 (p=0.0035) in 48 hours. In both cases proteins in MIN6 cells treated with 1mg/ml cultivated American ginseng were up-regulated more than those treated with 0.5mg/ml wild American ginseng. This could mean that cultivated American ginseng has a stronger effect with the maximum concentration that does not cause cell death on these proteins. Although a higher concentration (0.25 mg/ml) of CAG was used than WAG (0.125 mg/ml), these concentrations are the highest concentrations that does not cause more than 5% cell death. The consistent changes of levels of proteins in MIN6 indicated, that there are similar responses of MIN6 cells toward

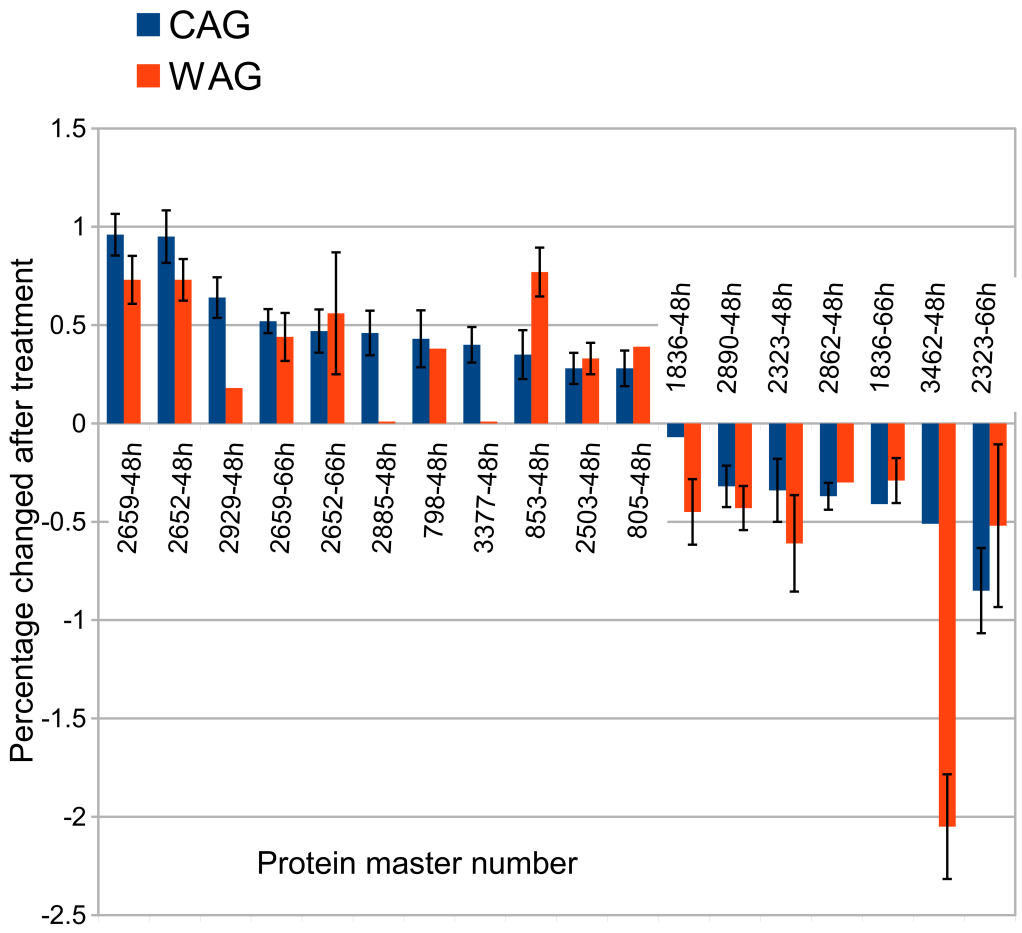


Figure 4.14: *Comparison of proteins treated by wild and cultivated American ginseng. This bar plot presents the same data in table 4.3. Error bars are calculated. Data bars without error bars ploted are not significant.*

Table 4.3: *Comparison of proteins treated by wild and cultivated American ginseng.* [1] *Protein master number, outside the bracket is the master number for cultivated American ginseng MIN6, inside the bracket is for wild American ginseng.* [2] *Theoretical molecular weight and protein isoelectric point.* [3] *Treatment time.*[4] *Ratio/t-test p value for CAG treated.* [5] *Ratio/t-test p value for WAG treated.*

Protein[1]	Th. MW/PI[2]	time [3]	r/p CAG[4]	r/p WAG[5]
2659(2642)	22738/8.15	48h	0.00028/1.96	0.0027 /1.73
		66h	0.0017/1.52	0.32/1.44
2652(2589)	22290/7.62	48h	0.00078/1.95	0.0035 /1.73
		66h	0.026/1.47	0.31/1.56
2890(2889)	18323/5.28	48h	0.026 /-1.32	0.04/ -1.43
3462(3439)	10366/8.37	48h	0.038/-1.51	0.038/-3.05
2885(2848)	18453/6.13	48h	0.011/1.46	0.86 / 1.01
2503(2491)	25802/5.10	66h	0.012/1.28	0.05 /1.33
2862(2865)	18824/4.78	48h	0.0016/-1.37	0.057 / -1.3
3377(3334)	11305/7.16	48h	0.005/1.40	0.083/1.01
2929(2893)	17810/6.13	48h	0.0011/1.64	0.18/1.18
798(667)	70232/4.72	48h	0.063/1.43	0.1/1.38
805(666)	69834/4.80	48h	0.041/1.28	0.054/1.39
853(752)	68459/4.62	48h	0.07/1.35	0.017/1.77
1836(1748)	41822/4.46	48h	0.53/-1.07	0.044/-1.45
		66h	0.041/-1.41	0.086/ -1.29
2323(2229)	30037/7.51	48h	0.069 /-1.34	0.027 / -1.61
		66h	0.044/ -1.85	0.31/-1.52

different concentration/types of American ginseng treatment.

There are also proteins that respond differently to these different treatments. For protein 2929(2893), treatment with CAG for 48 hours caused an increase of 1.64-fold (p=0.0011), whereas treated with WAG the expression level did not change.

Protein 798, 805, and 853 are adjacent proteins, and all showed an up-regulating trend. Although the p value for 798 is larger than 0.05, there is still confidence that the up regulation trend is real.

Protein 1836(1748) were down-regulated by -1.45-fold with WAG, but did not change with CAG in 48 hours. However, in 66 hours this protein down

regulated -1.41-fold in CAG, but only -1.29-fold in WAG. This means that this protein started to decrease earlier in WAG treated MIN6. This same case is shown in protein 2323(2229).

Because the data for 24h WAG treated MIN6 are not available, and less replicates samples were used for WAG (3 replicates) than CAG (5 replicates), standard error of protein data in CAG study is smaller than in WAG study. Therefore more proteins in CAG study were selected as statistically significant. Some proteins in WAG study may be ruled out with a larger p value, purely because of technical reasons, for example, they are located on a perturbation area of the 2D-Gel. However, despite of the fact that less replicates were used, there are still these proteins listed in table 4.3 that showed similar protein changes in MIN6 cells between CAG and WAG study.

Because the DIGE experiment were performed in different batches for WAG and CAG treated cells, the results indicate that some protein responses of treatment with American ginseng are reproducible. This means DIGE can be a potential tool to assess the protein profiles of MIN6 cells before and after treated with American ginseng. If the difference of protein profiles treated by different TCM can be standardised, DIGE can be used as a reliable tool to perform quality control of TCM.

4.3.8 Protein identification

Table 4.4: *MIN6 cell proteins identified by 2D gel electrophoresis and ELSD-MS. PAN: primary (citable) accession number. M.W.: molecular weight. Oat: Ornithine aminotransferase, mitochondrial*

Protein name	PAN	M.W.	location
Oat	P04182	48723	mitochondrion matrix

In order to demonstrate the feasibility of orbitrap LC MS/MS for protein identification of 2D-GE, one protein listed in table 4.4 was identified. Ornithine aminotransferase, mitochondrial (Oat) (p1446) was down-regulated by 24% (p=0.0089) at 48 hours treated by CAG. It is not possible to discuss the mechanisms of American ginseng on MIN6 cells with few proteins

identified.

In the future studies, it is suggested that all the selected proteins in section 4.3.5, section 4.3.6 and section 4.3.7 should be identified, and their pathways will be analysed. It is possible that proteins shows similar regulation trend after treatment belongs to the same pathway.

Although most of the selected proteins are not identified, their theoretical pIs and molecular weights are known because the pH scale of the 1D IPG strip and the range of the molecular weights of the second dimension are known.

4.4 Conclusion and future studies

This chapter explored the possibility of connecting ginseng quality to the proteomic profile of the MIN6 cells treated by ginsengs. A MIN6 cell viability test was carried out before the proteomic study. When ginseng extracts were higher than 0.5 mg/ml in cell culturing medium, both cultivated and wild ginseng caused more than 50% MIN6 cell death after 48 hours. Wild American ginseng was found to cause more cell death than cultivated American ginseng at 0.25 and 0.125 mg/ml, in 48 and 66 hours treatment. This might indicate a more toxic effect of wild than cultivated American ginseng. 0.25 mg/ml for cultivated American ginseng and 0.125 mg/ml wild American ginseng were selected to treat MIN6 because those concentrations caused less than 5% cell death.

In the case that several proteins are next to each other, there may be chance of a false discovery of protein change. Some potential PTMs of proteins and a protein shoulder were illustrated. This information can contribute and be analysed together with the future MS identifications.

Wild and cultivated American ginseng were used to treat MIN6 cells for 24, 48 and 66 hours, and DIGE analysis was performed on MIN6 cells before and after treatment.

This method demonstrated changes more than 20% in certain proteins in one or more of the selected time points. The protein expression level increase or decrease in response to the treatment. Some proteins were up- or

down-regulated in all of the three treatment times, in various degrees. Some proteins increased or decreased more in one time point than the other time points.

Interesting proteins were selected based on the criteria that a ratio of > 1.2 or <-1.2 of proteins abundance between control and treated was shown, besides the *t*-test p value should be lower than 0.05, and the protein spots should show up in more than 90% of all the gels analysed. 48 proteins out of 3563 detected spots have demonstrated significant up-regulation, and 35 proteins have demonstrated significant down-regulation in at least one treatment times in cultivated American ginseng. For wild American ginseng, the numbers are 21 for up-regulation and 12 for down-regulation out of 3504. The number of proteins which meet the selection criteria may increase if more replicates were available, because the p values will be smaller accordingly.

It is possible that the regulation trends of proteins are related to protein pathways, but it is not possible to understand the meaning and the relationship between changes of proteins and their correlations to the chemical composition of TCM used without further study.

Apart from this, the biological samples can also be variable from each other, because they are unique biological individuals. Even for the most stable cell lines, different passages might have slight different compositions. Previous research showed the difference of protein profiles between high and low passage proteins (passage 19 and 40) in MIN6 cells[136]. In the current study, cell passages close to each other (33, 35, and 37) were used in order to make sure the result was not affected by the passage differences. The same passage of cells was used for each control and the correspondent treated experiment.

The effect of American ginseng has long been known as gentle, these might leading to a mild effect toward the protein changes of MIN6 cells. If a TCM with proven strong effect are selected to treat a cell line, it might be easier to see bigger difference before and after treatment.

It is also recommended that for a preliminary experiment, less treatment conditions including different time points, should be used but increase the biological replicates, in order to obtain a good statistical result. For other

techniques such as ^{1}H-NMR, HPLC, it is possible to use many technical and biological replicates, where using a large sample size is cheap and easy. However this is not possible for the current DIGE analysis, due to the high price and many experimental steps involved. It is recommended that 24 samples are loaded on 12 gels as one experiment, this could include 4 groups, for example 2 control and 2 treated groups, with 6 replicates each. This will allow a good statistic result to be obtained. In this study, 5 batches of MIN6 cells treated by cultivated American ginseng for 48 hours were available for analysis, and a very good reproducibility and a small error bar was obtained.

This is a preliminary study innovatively using DIGE technique on the systematic quality control of herbal medicines, attempting to connect the activity of TCM, in this case the whole protein profile to the whole plant extract. In this study the protein expression level of MIN6 changed by treating with CAG and WAG, besides there are some similar and different changes caused by CAG and WAG. Statistical analysis shows that these changes are significant and stable, therefore it is possible to correlate these changes to the quality of the American ginseng.

MS identifications of all the selected proteins in this chapter will help to understand the relationship of changes of protein expression level to the different types or concentrations of American ginseng. Some pathways might be found to be related to disease control. If this can be linked successfully, the quality of TCM can be linked to the protein changes directly, thus using the "effect" to control the quality, and skip the complex intermediate steps which might or might not relevant to the eventual purpose of TCM.

Chapter 5

General conclusion

This thesis explored the feasibility of chemical- and acitivity-based quality control methods on TCMs including Danshen and Sanqi roots, with their compound products, and three ginseng species.

Chemical-based methods including chromatography using HPLC-DAD and spectroscopy including ^{1}H-NMR and ATR-IR were used with PCA analysis. The advantages and disadvantages of each method was discussed. Each method was found to be able to distinguish between some very similar TCMs, depending on the type. Among them, ATR-IR spectrum is fast and easy to do a quick screening, but was found difficult to be correlated to TCM metabolites.

In the second chapter, various Danshen industrial and commercial products were screened using different methods. Danshen grown in different regions, Sanqi supplied at different dates, and compound Danshen dripping pill (CDDP) produced in different years were found to be different using HPLC-DAD, ^{1}H-NMR and ATR-IR hyphenated with PCA. As an initial study some metabolites were identified.

This study shows TCM grown in different regions, and supplied at different times (possibly different harvesting times) have different metabolite compositions. Therefore, it is necessary to set quality standards in planting farms and perform routine checks. On-site labs may be set up to minimise the quality change during harvesting, drying and transporting.

In the third chapter, a large sample of 4 types of commercial ginseng was studied using ^{1}H-NMR-PCA. The differences were identified to be primary and secondary metabolites. American ginseng and Asian ginseng were successfully distinguished from each other. Primary metabolites including arginine and choline were found for the first time to be potential candidates to distinguish the two species. This has provided a potential usage of these metabolites to identify adulterant from American ginseng. Wild American ginseng samples with different prices and suppliers were found to be very different from each other, and some were found similar to the cultivated ones potentially due to adulteration or low age. However, the majority of wild American ginseng showed potentially a higher content of ginsenosides.

Chapters 2 and 3 show chemical-based TCM standardisation methods are fast, quantitative, and comprehensive. However, it is not known if it is effective. Chemical-based methods focus on the chemical composition of TCM, it can tell how different and what the differences are in terms of chemical metabolites. However, what does those complex chemical mixtures mean to human diseases, is unknown. Therefore, it is very difficult to say which chemical composition is "good" and which is "bad". Unless the composition of TCM is correlated to its efficacy, it is not possible to judge whether the quality is good.

Some studies have been carried out using TCM on specific diseases[2][3][4]. However, the effect of TCM might not only be related to one disease category according to the western medicine. In the concept of TCM practice, different diseases in western medicine can belong to one *Zheng*, which means the underlying principles of diseases according to the TCM philosophy. Therefore it is not adequate to test TCM on only one type of disease. A non-discrete method that can reveal the holistic activity of TCM is needed. The "-omics" techniques profiles a whole biological system from different points of view, whether it is transcriptome or proteome. Using such a method, a holistic view can be achieved, and details of this view may be explored if needed.

Therefore, a proteomic approach was chosen to evaluate the quality of TCM. In the fourth chapter, concentrations of 0.25 mg/ml cultivated American ginseng and 0.125 mg/ml wild American ginseng extract were selected as

the highest concentrations without causing more than 5% MIN6 cell death. The fact that the selected concentrations were different suggested that wild and cultivated American ginseng are different. As it was discussed, the difference might be due to the higher content of total ginsenosides in wild American ginseng. A cell viability test may be used to test the potential toxicity of TCM before it is used.

After the appropriate concentrations of ginseng extracts were selected, they were used to treat MIN6 cells, and the proteins in MIN6 cells were extracted, purified and subjected to DIGE analysis. The advantages of DIGE are that it is quantitative, precise, and reproducible, than conventional 2D-gel staining methods such as silver or Coomassie blue. The use of three Cy-Dyes with no interaction to each other allows three proteins samples to loaded on the same gel, with one sample being the internal standard. The introduction of internal standard allows the comparison between multiple gels, which was not possible before the invention of DIGE.

The results of the proteomic chapter showed DIGE is sensitive enough to identify protein changes before and after ginseng treatment, and between wild and cultivated ginseng treatment. The protein changes after treatment were plotted at 24, 48, and 66 hours, thus a trend can be observed. After automatic and manual selection, 48 and 35 protein spots out of 3563 were significantly up or down-regulated by more than 20% after treated with cultivated American ginseng. Among them, 20 proteins were changed by more than 50%. The cells treated with wild American ginseng were analysed by the same way. 33 protein spots of of 3504 were found to be significantly increased or decreased by more than 20%. 10 proteins among the selected ones were found to be up/down-regulated similarly between wild and cultivated American ginseng, but the rest of the selected proteins either changed with wild, or with cultivated ginseng extract.

The differences of protein changes of MIN6 cells caused by cultivated and wild American ginseng extracts can possibly be used to distinguish ginseng types and control the quality. Although currently it is too expensive and time consuming to use DIGE as a routine quality control method, it shows the possibility, and the price, and efficiency may be improved in the future

with the advancement of techniques.

This is the first time that DIGE was used as a holistic activity-based method to assess the quality of TCM. The significant protein changes after treatment and between treatments indicated that protein changes are related to the addition of American ginseng, and to the types of ginseng extracts added. These changes were also found to be reliable. For future studies, it is suggested that the changed proteins be identified, thus a better understanding of the relationship between protein pathways and ginseng metabolites can be achieved.

The DIGE activity-based method used in this study can be used to study other TCMs, preferably with a cell line which it has some known effects on. With the invention and improvement of modern techniques, it is possible that DIGE will work faster, cheaper and more precise.

This thesis evaluated many different techniques, and showed DIGE to be possible as an potential TCM quality control method. This is a small step achieving the ultimate goal of controlling the quality of a very complex mixture of nature products. It is to be hoped that this work may contribute to the eventual licensing of TCM preparation as traditional medicines.

Bibliography

[1] A. L. Harvey. "Natural products in drug discovery." *Drug Discov. Today* **13** (2008), 894–901.

[2] X. Yang et al. "Effect of Danggui Buxue Tang on immune-mediated aplastic anemia bone marrow proliferation mice." *Phytomedicine* (2013). Article In Press.

[3] I. L. I. Hook. "Danggui to *Angelica sinensis* root: Lost in translation? Are potential benefits of a TCM lost to European women? A review." *J. Ethnopharmacol.* (2013). Article In Press.

[4] L. Li and X. Liang. "Fufang danshen pill prevent atherosclerosis by decreasing plasma lipid and modifying vascular endothelial function in rabbits". *J. Mol. Cell. Cardiol.* **44** (2008), 747–748.

[5] J. K. Rao et al. "Use of complementary therapies for arthritis among patients of rheumatologists." *Ann. Intern. Med.* **131** (1999), 409–416.

[6] J. Qiu. "China plans to modernize traditional medicine." *Nature* **446** (2007), 590–591.

[7] W. L. Li et al. "Natural medicines used in the traditional Chinese medical system for therapy of diabetes mellitus." *J. Ethnopharmacol.* **92** (2004), 1–21.

[8] F. H. Qi et al. "Traditional Chinese medicine and related active compounds: A review of their role on hepatitis B virus infection." *Drug Discov. Ther.* **7** (2013), 212–224.

[9] X. Li et al. "Traditional Chinese medicine in cancer care: a review of controlled clinical studies published in Chinese." *PLoS One* **8** (2013), e60338.

[10] G.-X. Zhong et al. "Chemical Characteristics of *Salvia miltiorrhiza* (Danshen) Collected from Different Locations in China". *J. Agric. Food Chem.* **57** (2009), 6879–6887.

[11] S.-O. Yang et al. "NMR-based metabolic profiling and differentiation of ginseng roots according to cultivation ages". *J. Pharm. Biomed. Anal.* **58** (2012), 19–26.

[12] M. A. Doel and J. Segrott. "Materializing complementary and alternative medicine: aromatherapy, chiropractic, and Chinese herbal medicine in the UK". *Geoforum* **35** (2004), 727–738.

[13] C. P. Commission, ed. *The Pharmacopoeia of the People's Republic of China*. Chemical Industry Press, 2005.

[14] Y.-Z. Liang, P. Xie, and K. Chan. "Quality control of herbal medicines". *J. Chromatogr. B* **812** (2004), 53–70.

[15] E. Holmes et al. "The assessment of plant metabolite profiles by NMR-based methodologies." *Planta Med.* **72** (2006), 771–785.

[16] E.-J. Lee et al. "Quality Assessment of Ginseng by ^{1}H NMR Metabolite Fingerprinting and Profiling Analysis". *J. Agric. Food Chem.* **57** (2009), 7513–7522.

[17] Y. Sun. "Structure and biological activities of the polysaccharides from the leaves, roots and fruits of *Panax ginseng* C.A. Meyer: An overview". *Carbohydr. Polym.* **85** (2011), 490–499.

[18] S. Nyiredy. "Progress in forced-flow planar chromatography." *J. Chromatogr. A* **1000** (2003), 985–999.

[19] X.-H. Fan et al. "Multiple chromatographic fingerprinting and its application to the quality control of herbal medicines". *Anal. Chim. Acta* **555** (2006), 217–224.

[20] S. S. Jacob, N. W. Smith, and C. Legido-Quigley. "Assessment of Chinese medicinal herb metabolite profiles by UPLC-MS-based methodology for the detection of aristolochic acids". *J. Sep. Sci.* **30** (2007), 1200–1206.

[21] L. Zhou, M. Chow, and Z. Zuo. "Improved quality control method for Danshen products - Consideration of both hydrophilic and lipophilic active components". *J. Pharm. Biomed. Anal.* **41** (2006), 744–750.

[22] J. Wu et al. "Application of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry in the analysis of volatile oil of traditional Chinese medicines". *J. Chromatogr. A* **1034** (2004), 199–205.

[23] C.-Y. Li et al. "Quality assessment of *Radix Codonopsis* by quantitative nuclear magnetic resonance". *J. Chromatogr. A* **1216** (2009), 2124–2129.

[24] X. Duan et al. "Rapid discrimination of geographical origin and evaluation of antioxidant activity of *Salvia miltiorrhiza* var. alba by Fourier transform near infrared spectroscopy." *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **122C** (2013), 751–757.

[25] J. Lu et al. "Application of two-dimensional near-infrared correlation spectroscopy to the discrimination of Chinese herbal medicine of different geographic regions". *Spectrochim. Acta Mol. Biomol. Spectros.* **69** (2008), 580–586.

[26] Y.-Y. Zhao et al. "An expeditious HPLC method to distinguish *Aconitum kusnezoffii* from related species". *Fitoterapia* **80** (2009), 333–338.

[27] G. E. Pereira et al. "^{1}H NMR and chemometrics to characterize mature grape berries in four wine-growing areas in Bordeaux, France." *J. Agric. Food Chem.* **53** (2005), 6382–6389.

[28] I. Duarte et al. "High-resolution nuclear magnetic resonance spectroscopy and multivariate analysis for the characterization of beer." *J. Agric. Food Chem.* **50** (2002), 2475–2481.

[29] S. Zeng et al. "Direct analysis in real time mass spectrometry and multivariate data analysis: a novel approach to rapid identification of analytical markers for quality control of traditional Chinese medicine preparation." *Anal. Chim. Acta* **733** (2012), 38–47.

[30] J. Kang et al. "NMR-based metabolomics approach for the differentiation of ginseng (*Panax ginseng*) roots from different origins." *Arch. Pharm. Res.* **31** (2008), 330–336.

[31] J. Gao et al. "Secondary metabolite mapping identifies *Scutellaria* inhibitors of human lung cancer cells". *J. Pharm. Biomed. Anal.* **53** (2009), 723–728.

[32] J. K. Nicholson, E. Holmes, and J. C. Lindon. *The Handbook of Metabonomics and Metabolomics*. Ed. by J. C. Lindon, J. K. Nicholson, and E. Holmes. Amsterdam: Elsevier Science B.V., 2007. Chap. 1, 1–33.

[33] E. Holmes et al. "Human metabolic phenotype diversity and its association with diet and blood pressure". *Nature* **453** (2008), 396–400.

[34] J. Li et al. "A high-performance liquid chromatography and nuclear magnetic resonance spectroscopy-based analysis of commercially available praziquantel tablets". *J. Pharm. Biomed. Anal.* **45** (2007), 263–267.

[35] F. V. der Kooy, R. Verpoorte, and J. J. M. Meyer. "Metabolomic quality control of claimed anti-malarial *Artemisia afra* herbal remedy and *A. afra* and *A. annua* plant extracts". *S. Afr. J. Bot.* **74** (2008), 186–189.

[36] J. E. Rodrigues and A. M. Gil. "NMR methods for beer characterization and quality control." *Magn. Reson. Chem.* **49** (2011), S37–S45.

[37] N. J. C. Bailey et al. "Multi-component metabolic classification of commercial feverfew preparations via high-field ^{1}H-NMR spectroscopy and chemometrics." *Planta Med.* **68** (2002), 734–738.

[38] P. H. Chan et al. "Metabonomic analysis of water extracts from Chinese and American ginsengs by ^{1}H nuclear magnetic resonance: identification of chemical profile for quality control." *Chin. Med.* **7** (2012), 25.

[39] E. M. Williamson. "Synergy and other interactions in phytomedicines." *Phytomedicine* **8** (2001), 401–409.

[40] Y. Wang et al. "A metabonomic strategy for the detection of the metabolic effects of chamomile (*Matricaria recutita* L.) ingestion." *J. Agric. Food Chem.* **53** (2005), 191–196.

[41] Y. Chu et al. "The effect of Compound Danshen Dripping Pills, a Chinese herb medicine, on the pharmacokinetics and pharmacodynamics of warfarin in rats." *J. Ethnopharmacol.* **137** (2011), 1457–1461.

[42] "American botanical council, Chinese Herbal Medicine Clears US FDA Phase II Trials". *HerbalEGram* **7** (2010).

[43] "Guidance for Industry: INDs for Phase 2 and Phase 3 Studies Chemistry, Manufacturing, and Controls Information. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research. `http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070567.pdf`" (2003).

[44] "Guidance for Industry Botanical Drug Products, U.S. Food and Drug Administration, `http://www.fda.gov/downloads/Drugs/Guidance\ComplianceRegulatoryInformation/Guidances/ucm070491.pdf`" (2004).

[45] C. P. Commission, ed. *The Pharmacopoeia of the People's Republic of China*. Chemical Industry Press, 2010.

[46] P. Li et al. "Analysis of Chinese herbal medicines with holistic approaches and integrated evaluation models". *TrAC, Trends Anal. Chem.* **27** (2008), 66–77.

[47] *Shen Nong Ben Cao Jing (Shennong's Chinese Materia Medica), volume 1*. approximately ca. 25–220 ad.

[48] J.-Q. Liu et al. "Effects of tanshinone IIA, a major component of *Salvia miltiorrhiza*, on platelet aggregation in healthy newborn piglets". *J. Ethnopharmacol.* **137** (2011), 44–49.

[49] Y. P. Wu et al. "Salvianolic Acid B inhibits platelet adhesion under conditions of flow by a mechanism involving the collagen receptor $\alpha 2\beta 1$". *Thromb. Res.* **123** (2008), 298–305.

[50] F. F. Y. Lam et al. "Relaxant effects of danshen aqueous extract and its constituent danshensu on rat coronary artery are mediated by inhibition of calcium channels". *Vascul. Pharmacol.* **46** (2007), 271–277.

[51] Y. Jia et al. "Is danshen (*Salvia miltiorrhiza*) dripping pill more effective than isosorbide dinitrate in treating angina pectoris? A systematic review of randomized controlled trials." *Int. J. Cardiol.* **157** (2012), 330–340.

[52] Y. Jia et al. "*Panax notoginseng* saponins decrease cholesterol ester via up-regulating ATP-binding cassette transporter A1 in foam cells". *J. Ethnopharmacol.* **132** (2010), 297–302.

[53] Q. F. Xu, X. L. Fang, and D. F. Chen. "Pharmacokinetics and bioavailability of ginsenoside Rb1 and Rg1 from *Panax notoginseng* in rats". *J. Ethnopharmacol.* **84** (2003), 187–192.

[54] Y. Yao et al. "Proteomic analysis of differential protein expression in rat platelets treated with notoginsengnosides". *Phytomedicine* **15** (2008), 800–807.

[55] A.-J. Lau et al. "Antiplatelet and anticoagulant effects of *Panax notoginseng*: Comparison of raw and steamed *Panax notoginseng* with *Panax ginseng* and *Panax quinquefolium*". *J. Ethnopharmacol.* **125** (2009), 380–386.

[56] S. Wei, L. Meng, and R. Huang. "Effect of *Panax notoginseng* saponins on serum neuronal specific enolase and rehabilitation in patients with cerebral hemorrhage". *Zhongguo Zhong Xi Yi Jie He Za Zhi* **27(2)** (2007), 159–62.

[57] C.-Y. Yang et al. "Anti-diabetic effects of *Panax notoginseng* saponins and its major anti-hyperglycemic components". *J. Ethnopharmacol.* **130** (2010), 231–236.

[58] Y. Wang, X. Gao, and B. Zhang. "Study on the effect of crude previous termborneolnext term in Fufang Dansen pills and the connection between the amount of previous termborneol and pharmacodynamics". *Chin. Tradit. Herbal Drugs* **35** (2004), 672–673.

[59] R. Liu et al. "Protection by borneol on cortical neurons against oxygen-glucose deprivation/reperfusion: involvement of anti-oxidation and anti-inflammation through nuclear transcription factor kappaB signaling pathway." *Neuroscience* **176** (2011), 408–419.

[60] X.-J. Lai et al. "Comparative pharmacokinetic and bioavailability studies of three salvianolic acids after the administration of Salviae miltiorrhizae alone or with synthetical borneol in rats". *Fitoterapia* **82** (2011), 883–888.

[61] "Phase II Multi-Center Study of T89 to Treat Chronic Stable Angina (T89 phase 2). `http://clinicaltrials.gov/ct2/show/NCT00797953`" (2010).

[62] L. Zhou, Z. Zuo, and M. S. S. Chow. "Danshen: an overview of its chemistry, pharmacology, pharmacokinetics, and clinical use." *J. Clin. Pharmacol.* **45** (2005), 1345–1359.

[63] H.-T. Cheng et al. "Simultaneous quantification of selected compounds from Salvia herbs by HPLC method and their application". *Food Chem.* **130** (2012), 1031–1035.

[64] Z. Wang et al. "Study on optimum preparation procedure for Danggui Danshen dripping pill". *Chin. Trad. Pat. Med.* **28** (2006), 3.

[65] R. Zhu and S. Li. "Study on moulding techniques of Shuxin dripping pill". *Chin. Trad. Pat. Med.* **24** (2002), 3.

[66] Y. Tu. "The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine." *Nat. Med.* **17** (2011), 1217–1220.

[67] Y.-J. Wei et al. "Improved quality control method for Fufang Danshen preparations through simultaneous determination of phenolic acids, saponins and diterpenoid quinones by HPLC coupled with diode array and evaporative light scattering detectors". *J. Pharm. Biomed. Anal.* **45** (2007), 775–784.

[68] X. Zheng et al. "Determination of the main bioactive metabolites of Radix *Salvia miltiorrhizae* in Compound Danshen Dripping Pills and the tissue distribution of Danshensu in rabbit by SPE-HPLC-MS". *J. Sep. Sci.* **30** (2007), 851–857.

[69] L.-J. Zhang et al. "Danshensu has anti-tumor activity in B16F10 melanoma by inhibiting angiogenesis and tumor cell invasion." *Eur. J. Pharmacol.* **643** (2010), 195–201.

[70] G.-R. Zhao et al. "Characterization of the radical scavenging and antioxidant activities of danshensu and salvianolic acid B". *Food Chem. Toxicol.* **46** (2008), 73–81.

[71] J. Fu et al. "Tanshinone IIA protects cardiac myocytes against oxidative stress-triggered damage and apoptosis." *Eur. J. Pharmacol.* **568** (2007), 213–221.

[72] S. Gao et al. "Cardiovascular actions and therapeutic potential of tanshinone IIA." *Atherosclerosis* **220** (2012), 3–10.

[73] H.-S. Zhang and S.-Q. Wang. "Notoginsenoside R1 from *Panax notoginseng* inhibits TNF-induced PAI-1 production in human aortic smooth muscle cells". *Vascul. Pharmacol.* **44** (2006), 224–230.

[74] J. Yi et al. "Atherosclerosis lesion is accelerated by persistent systemic inflammation but attenuated by saponins from *Panax notoginseng* in rabbits". *J. Med. Coll. PLA* **23** (2008), 38–44.

[75] W. Jia et al. "The rediscovery of ancient Chinese herbal formulas." *Phytother. Res.* **18** (2004), 681–686.

[76] H. Li. "The Physical chemical test of Danshen and the mimic product - *Arctium lappa* L." *Zhong Yao Cai* **22** (1999), 22–25.

[77] Z.-P. Zeng and J.-G. Jiang. "Analysis of the adverse reactions induced by natural product-derived drugs." *Br. J. Pharmacol.* **159** (2010), 1374–1391.

[78] P. Li et al. "Optimizing ultraperformance liquid chromatographic analysis of 10 diterpenoid compounds in *Salvia miltiorrhiza* using central composite design." *J. Agric. Food Chem.* **56** (2008), 1164–1171.

[79] X. Liu, J. Zhang, and X. Chen. "Separation and determination of three water-soluble compounds in *Salvia miltiorrhiza* Bunge and two related traditional medicinal preparations by flow injection-capillary electrophoresis". *J. Chromatogr. B* **852** (2007), 325–332.

[80] A. K. S. Wan et al. "Vascular Effects of Different Lipophilic Components of Danshen, a Traditional Chinese Medicine, in the Isolated Porcine Coronary Artery". *J. Nat. Prod.* **71** (2008), 1825–1828.

[81] E.-J. Yang et al. "Barrier protective effects of rosmarinic acid on HMGB1-induced inflammatory responses *in vitro* and *in vivo*." *J. Cell. Physiol.* **228** (2013), 975–982.

[82] P.-S. Xie and A. Y. Leung. "Understanding the traditional aspect of Chinese medicine in order to achieve meaningful quality control of Chinese materia medica". *J. Chromatogr. A* **1216** (2009), 1933–1940.

[83] C. Tistaert, B. Dejaegher, and Y. Vander Heyden. "Chromatographic separation techniques and data handling methods for herbal fingerprints: a review." *Anal. Chim. Acta* **690** (2011), 148–161.

[84] J.-Z. Song et al. "A novel approach to rapidly explore analytical markers for quality control of Radix Salviae Miltiorrhizae extract granules by robust principal component analysis with ultra-high performance liquid chromatography-ultraviolet-quadrupole time-of-flight mass spectrometry." *J. Pharm. Biomed. Anal.* **53** (2010), 279–286.

[85] M. Daszykowski et al. "Robust statistics in data analysis- A review: Basic concepts". *Chemometrics and Intelligent Laboratory Systems* **85** (2007), 203–219.

[86] F. Gong et al. "Correction of retention time shifts for chromatographic fingerprints of herbal medicines". *J. Chromatogr. A* **1029** (2004), 173–183.

[87] C. L. Saw et al. "Pharmacodynamics of ginsenosides: antioxidant activities, activation of Nrf2, and potential synergistic effects of combinations." *Chem. Res. Toxicol.* **25** (2012), 1574–1580.

[88] Y. Xiong et al. "Antiobesity and antihyperglycemic effects of ginsenoside Rb1 in rats." *Diabetes* **59** (2010), 2505–2512.

[89] Y. S. Keum et al. "Antioxidant and anti-tumor promoting activities of the methanol extract of heat-processed ginseng." *Cancer Lett.* **150** (2000), 41–48.

[90] R. A. van den Berg et al. "Centering, scaling, and transformations: improving the biological information content of metabolomics data." *BMC Genomics* **7** (2006), 142.

[91] Y. Hu et al. "Crystallization and phase separation in blends of high stereoregular poly(lactide) with poly(ethylene glycol)". *Polymer* **44** (2003), 5681–5689.

[92] L. Thakur et al. "Novel approaches for stability improvement in natural medicines." *Pharmacogn. Rev.* **5** (2011), 48–54.

[93] J. B. Wan et al. "Chemical characteristics for different parts of *Panax notoginseng* using pressurized liquid extraction and HPLC-ELSD." *J. Pharm. Biomed. Anal.* **41** (2006), 1596–1601.

[94] N. Fuzzati. "Analysis methods of ginsenosides." *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* **812** (2004), 119–133.

[95] J. Needham, G. Lu, and H. Huang. *Science and Civilisation in China, Volume 6, Biology and Biological Technology, Part 1: Botany.* Cambridge University Press, 1986.

[96] L. Shizhen. *Ben Cao Gang Mu, (Compendium of Materia Medica).* ca. 1593.

[97] W. Yiluo. *New compilation of Materia Medica (Bencao Congxin).* 1757.

[98] L. Li et al. "Rapid qualitative and quantitative analyses of Asian ginseng in adulterated American ginseng preparations by UPLC/Q-TOF-MS". *J. Pharm. Biomed. Anal.* **52** (2010), 66–72.

[99] L. Williams. "Ginseng". *Economic Botany* **2** (1957), 347.

[100] D. Luo and B. Fang. "Structural identification of ginseng polysaccharides and testing of their antioxidant activities". *Carbohydrate Polymers* **72** (2008), 376–381.

[101] K. M. Park et al. "Nitric oxide is involved in the immunomodulating activities of acidic polysaccharide from *Panax ginseng*." *Planta Med.* **67** (2001), 122–126.

[102] Y.-Z. Xiang et al. "A comparison of the ancient use of ginseng in traditional Chinese medicine with modern pharmacological experiments and clinical trials." *Phytother. Res.* **22** (2008), 851–858.

[103] S. N. Kim et al. "Simultaneous quantification of 14 ginsenosides in *Panax ginseng* C.A. Meyer (Korean red ginseng) by HPLC-ELSD and its application to quality control". *J. Pharm. Biomed. Anal.* **45** (2007), 164–170.

[104] K. Chan, K. S.-Y. Leung, and S. S. Zhao. "Harmonization of monographic standards is needed to ensure the quality of Chinese medicinal materials." *Chin. Med.* **4** (2009), 18.

[105] K. Y.-L. Yap, S. Y. Chan, and C. S. Lim. "Authentication of traditional Chinese medicine using infrared spectroscopy distinguishing between ginseng and its morphological fakes." *J. Biomed. Sci.* **14** (2007), 265–273.

[106] R. M. Corbit et al. "Simplified Extraction of Ginsenosides from American Ginseng (Panax quinquefolius L.) for High-Performance Liquid Chromatography Ultraviolet Analysis". *J. Agric. Food Chem.* **53** (2005), 9867–9873.

[107] B.-S. Sun et al. "Simultaneous quantification of 19 ginsenosides in black ginseng developed from *Panax ginseng* by HPLC-ELSD." *J. Pharm. Biomed. Anal.* **50** (2009), 15–22.

[108] R. Teng et al. "Complete assignments of ^{1}H and ^{13}C NMR resonances of oleanolic acid, 18α-oleanolic acid, ursolic acid and their 11-oxo derivatives". *Magn. Reson.* **40** (2002), 483–488.

[109] J.-G. Cho et al. "Physicochemical Characterization and NMR Assignments of Ginsenosides Rb1, Rb2, Rc, and Rd Isolated from *Panax ginseng*". *J. Ginseng Res.* **34** (2010), 113–121.

[110] K. Koch. "Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development." *Curr. Opin. Plant Biol.* **7** (2004), 235–246.

[111] J. Wang et al. "Anti-fatigue activity of the water-soluble polysaccharides isolated from *Panax ginseng* C. A. Meyer". *J. Ethnopharmacol.* **130** (2010), 421–423.

[112] V. A. Assinewe et al. "Phytochemistry of Wild Populations of *Panax quinquefolius* L. (North American Ginseng)". *J. Agric. Food Chem.* **51** (2003), 4549–4553.

[113] E. M. Schlag and M. S. McIntosh. "Ginsenoside content and variation among and within American ginseng (*Panax quinquefolius* L.) populations". *Phytochemistry* **67** (2006), 1510–1519.

[114] J.-R. Wang et al. "Quantitative comparison of ginsenosides and polyacetylenes in wild and cultivated American ginseng." *Chem. Biodivers.* **7** (2010), 975–983.

[115] E. M. Schlag and M. S. McIntosh. "The relationship between genetic and chemotypic diversity in American ginseng (*Panax quinquefolius* L.)." *Phytochemistry* **93** (2013), 96–104.

[116] W. Lim, K. W. Mudge, and F. Vermeylen. "Effects of Population, Age, and Cultivation Methods on Ginsenoside Content of Wild American Ginseng (*Panax quinquefolium*)". *J. Agric. Food Chem.* **53** (2005), 8498–8505.

[117] A. Buriani et al. "Omic techniques in systems biology approaches to traditional Chinese medicine research: Present and future". *J. Ethnopharmacol.* **140** (2012), 535–544.

[118] J. J. Mihalov, A. D. Marderosian, and J. C. Pierce. "DNA Identification of Commercial Ginseng Samples". *J. Agric. Food Chem.* **48** (2000), 3744–3752.

[119] Y. Zuo et al. "DNA barcoding of Panax species." *Planta Med.* **77** (2011), 182–187.

[120] G. Heubl. "New aspects of DNA-based authentication of Chinese medicinal plants by molecular biological techniques." *Planta Med.* **76** (2010), 1963–1974.

[121] R. Alm et al. "Proteomic variation is as large within as between strawberry varieties". *J. Proteome Res.* **6** (2007), 3011–3020.

[122] J. K. Nicholson and J. C. Lindon. "Systems biology: Metabonomics." *Nature* **455** (2008), 1054–1056.

[123] X. Lu et al. "Metabonomic study on 'Kidney-Yang Deficiency syndrome' and intervention effects of *Rhizoma Drynariae* extracts in rats using ultra performance liquid chromatography coupled with mass spectrometry." *Talanta* **83** (2011), 700–708.

[124] W. Shi et al. "Investigation of ginsenosides in different parts and ages of Panax ginseng". *Food Chem.* **102** (2007), 664–668.

[125] W. Lam et al. "The four-herb Chinese medicine PHY906 reduces chemotherapy-induced gastrointestinal toxicity." *Sci. Transl. Med.* **2** (2010), 45ps41.

[126] V. Wasinger et al. "Progress with gene-product mapping of the Mollicutes Mycoplasma genitalium". *Electrophoresis* **16** (1995), 1090–1094.

[127] P. James. "Protein identification in the post-genome era: the rapid rise of proteomics." *Q. Rev. Biophys.* **30** (1997), 279–331.

[128] Y. Shiio and R. Aebersold. "Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry." *Nat. Protoc.* **1** (2006), 139–145.

[129] S. P. Gygi et al. "Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags." *J. Proteome Res.* **1** (2002), 47–54.

[130] R. Kandpal, B. Saviola, and J. Felton. "The era of "omics" unlimited." *Biotechniques* **46** (2009), 351–2, 354–5.

[131] J. H.-K. Lum et al. "Proteome of Oriental ginseng *Panax ginseng* C. A. Meyer and the potential to use it as an identification tool." *Proteomics* **2** (2002), 1123–1130.

[132] F. Chi et al. "Proteomic analysis of rice seedlings infected by *Sinorhizobium meliloti* 1021." *Proteomics* **10** (2010), 1861–1874.

[133] R. Westermeier, T. Naven, and H.-R. Hopker. *Proteomics in Practice. A Guide to Successesful Experimental Design Second Edition.* WILEY-VCH Verlag GmbH & Co.KGaA, 2008, 134.

[134] R. Marouga, S. David, and E. Hawkins. "The development of the DIGE system: 2D fluorescence difference gel analysis technology." *Anal. Bioanal. Chem.* **382** (2005), 669–678.

[135] P. F. Lemkin. "Comparing two-dimensional electrophoretic gel images across the Internet." *Electrophoresis* **18** (1997), 461–470.

[136] P. Dowling et al. "Proteomic screening of glucose-responsive and glucose non-responsive MIN-6 beta cells reveals differential expression of proteins involved in protein folding, secretion and oxidative stress." *Proteomics* **6** (2006), 6578–6587.

[137] M. Di Carli et al. "Two-dimensional differential in gel electrophoresis (2D-DIGE) analysis of grape berry proteome during postharvest withering." *J. Proteome Res.* **10** (2011), 429–446.

[138] K. Pennington et al. "A proteomic investigation of similarities between conventional and herbal antidepressant treatments". *J. Psychopharmacol.* **23** (2009), 520–530.

[139] M. Jun-Ichi et al. "Establishment of a Pancreatic β Cell Line That Retains Glucose-Inducible Insulin Secretion: Special Reference to Expression of Glucose Transporter Isoforms". *Endocrinology* **127** (1990), 126–132.

[140] A. Al-Romaiyan et al. "A novel *Gymnema sylvestre* extract stimulates insulin secretion from human islets *in vivo* and *in vitro*." *Phytother. Res.* **24** (2010), 1370–1376.

[141] A. B. A. Ahmed, A. Rao, and M. Rao. "*In vitro* callus and *in vivo* leaf extract of *Gymnema sylvestre* stimulate β-cells regeneration and anti-diabetic activity in Wistar rats". *Phytomedicine* **17** (2010), 1033–1039.

[142] N. P. Sahu et al. "Triterpenoid saponins from *Gymnema sylvestre*". *Phytochemistry* **41** (1996), 1181–1185.

[143] J. Z. Luo and L. Luo. "American Ginseng Stimulates Insulin Production and Prevents Apoptosis through Regulation of Uncoupling Protein-2 in Cultured β Cells". *Evid. Based Complement. Alternat. Med.* **3** (2006), 365–372.

[144] J. D. Rasinger. "Ph.D. Thesis: Proteomic evaluation of the neurotoxic potential of model environmental contaminants commonly found in farmed salmon." (2011), 112.

[145] G. Majno and I. Joris. "Apoptosis, oncosis, and necrosis. An overview of cell death." *Am. J. Pathol.* **146** (1995), 3–15.

[146] S. Park et al. "Ginsenosides Rb1 and Rg1 suppress triglyceride accumulation in 3T3-L1 adipocytes and enhance beta-cell insulin secretion and viability in MIN6 cells via PKA-dependent pathways." *Biosci. Biotechnol. Biochem.* **72** (2008), 2815–2823.

[147] J. H. McDonald. *Handbook of Biological Statistics, 2nd ed.* Sparky House Publishing, Baltimore, Maryland, 2009, 233–236.

[148] B. D. Halligan et al. "ProMoST (Protein Modification Screening Tool): a web-based tool for mapping protein modifications on two-dimensional gels." *Nucleic Acids Res.* **32** (2004), W638–W644.