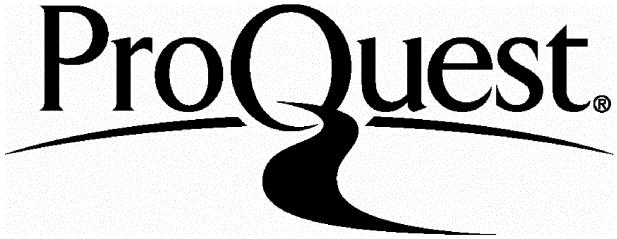
CHEMOMETRICS STUDY AND SPECTRAL DATABASE DEVELOPMENT OF PHARMACEUTICALS AND HERBAL PRODUCTS; MALAYSIAN PERSPECTIVE

Mazlina Mohd Said

The School of Pharmacy

University of London

A thesis submitted in partial fulfilment of the requirements of the University of London for the degree of Doctor of Philosophy.

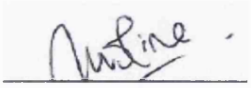


PLAGIARISM STATEMENT

This thesis describes research conducted at the School of Pharmacy, University of London between 5th November 2007 and 5th November 2010 under the supervision of Dr. Mire Zloh and Professor Simon Gibbons.

I certify that the research described is original and that any parts of the work that have been conducted by collaboration are clearly indicated. I also certify that I have written all text herein and have clearly indicated by suitable citation any part of this thesis that has already appeared in publication.

Mazlina Mohd Said



Signature

18.01.2012

Date

ABSTRACT

This research was initiated as part of the fight against the public health problems of rising counterfeit, substandard, and poor quality medicines and herbal products in the Malaysian market. A simple, quick and cost-effective drug screening procedure using an incremental near infra-red (NIR) spectral database of common medicines in combination with principal component analysis (PCA) was developed to facilitate drug analysis without depending on standard compounds or products from manufacturers. The novelty of the approach is demonstrated by this two-tier method which allowed application in product identification, drug quality study, herbal analysis, and the detection of counterfeit and adulterated medicines.

The NIR spectra database consisted of almost 4,000 spectra from 15 types of medicines and 3 types of herbal preparations, acquired and stored in the database throughout the study. The optimization procedure developed in this thesis on the database produced a search strategy using correlation and first derivative correlation algorithms on the full spectrum. The cut-off points of the hit quality index (HQI) were determined to classify the unknown sample in four categories; similar batch/match (classification type 1, <0.0001), same brand/different batch (classification type 2, <0.01), same type of medicine/different brand (classification type 3, <0.1) and different type of medicines (classification type 4, >0.2). The same set of spectra were analysed by multivariate methods, including PCA.

This two-tier screening approach proved successful when challenged firstly using simple compound drugs followed by complex mixtures of herbal preparations and then using alleged counterfeit and adulterated samples seized by the authority. The method developed has allowed samples to be identified without known background information which was difficult using other qualitative NIR techniques that required reference products for comparison. It is also allowed selection of suspected samples for further extensive analyses and in detection of adulterant contaminating herbal preparations.

The outcomes of this thesis should support further qualitative and quantitative researches on drugs and herbals products in Malaysian universities. The developed spectral database of drugs will providea valuable tool to assist in drug quality surveillance nation-wide.

ACKNOWLEDGEMENTS

I am greatly indebted to Dr. Mire Zloh for his willingness to supervise this research. His constant and expert supervision, his word of advice, patience and direction in guiding me through this study, and his invaluable help and support at all stages of the research remain unforgettable. Without his help, it would have been very hard for me to complete this work successfully. I cannot have asked for a more supportive supervision on this journey as well as I cannot thank him enough for all he has taught me and for the chance to work with him. His wisdom and guidance has enabled me to thrive as a student.

I am also very grateful to Professor Simon Gibbons for his willingness to co-supervise this research and his appreciable cooperation and support throughout the journey. Thank you also to Professor Tony Muffed, Dr. Robert Watt and Dr. Roger Jeep for their advice and support.

Special thanks are due to the Ministry of Higher Education Malaysia and my employer, *University Kebangsaan Malaysia* for their financial support.

I would also like to say my gratitude to several individuals; Dr. Abd. Razak Kasmuri-for giving me the early exposure in research work; Mr. Nazrul Izwan- for supplying me with the counterfeit and adulterated medicines from Malaysia; Mr. Emmanual Samuel, Dr. Kersti Karu , and Dr. Sybille Heidilberger - for answering all my questions in-and- out of the labs; Shilbe and Andreia- for making my life at SOP more enjoyable.

I am really grateful for the support from my parents/ in-laws. Life would be very difficult if you guys were not around. Thank you *mak dan abah* for the continuous support, words of wisdom, prayers, and most importantly, for always be there whenever I need you.

Finally, a very special thank you to my beloved husband, Zairul Effendy, and my beautiful daughter, Saffiya Hana, for all their love, sacrifice, patience, and support in giving me the motivation and ample time for me to complete this research.

LIST OF ABBREVIATIONS

ADR	Adverse drug reaction
API	Active pharmaceutical ingredient
APCI	Atmospheric pressure chemical ionisation
^{13}C NMR	Carbon NMR
DCA	Drug Control Authority
ESI	Electrospray ionisation
ERD	Entity relationship diagram
FDA	US Food and Drug Administration
FDIS	Fast drug identification system
^{1}H-NMR	Proton NMR
HPMC	Hydroxypropyl methylcellulose
HQI	Hit quality index
MALDI	Matrix assisted laser desorption ionization
MS	Mass spectrometry
MY	Malaysia
NIRS	Near-infrared spectroscopy
NMR	Nuclear magnetic resonance
NPCB	National Pharmaceutical Control Bureau
PC	Principal Component
PCA	Principal Component analysis
PCM	Paracetamol
PDE-5i	Phosphodiesterase-5 inhibitors
PLS	Partial least squares
PMP	Proprietary medicinal products
RDBMS	Relational database management system
RM	Ringgit Malaysia
SIMCA	Soft independent modelling of class analogy
SQL	Structure query language
TMS	Tetramethylsilane
TSP	Tetramethylsilyl propionate
UK	United Kingdom
WHO	World Health Organization

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background

The challenges faced by pharmaceutical enforcement bodies in the healthcare systems of developing countries include an influx of unregistered products, adulterated products, the adulteration of registered products, the adulteration of food and food-supplements, and counterfeit materials. Other problems faced by regulatory agencies are the results of smuggling, illegal entry, parallel importation, diversion, tampering, repackaging, relabeling, and the use of fake hologram meditags [1, 2, 3].

Among these challenges, the issue of counterfeit medicines is escalating across the globe [4]. The US Food and Drug Administration (FDA) estimates that counterfeit drugs make up more than 10% of the global medicines market [5], while the US Centre for Medicine in the Public Interest predicts that counterfeit drugs sales will reach USD 75 billion globally in 2010, an increase of more than 90% from 2005 [6].

Counterfeit medicines, by definition, are medicines that are deliberately and fraudulently mislabelled with respect to identity and/or source. This can apply to both branded and generic products. Counterfeit products can include products with the right or wrong ingredients, the wrong dosage or absence of active ingredients or with the fake packaging [7]. A substandard drug falls into one class of counterfeit medicines, and is described as a preparation that has insufficient active ingredients compared with those purported [7]. Furthermore, adulterated medicines are products that consist in whole or in part of any filthy, putrid, or decomposed substance [8].

Pharmaceutical enforcement agencies in each country aim to verify the quality of pharmaceutical products on the market and to ensure the products received by consumers are safe and effective. However, this is not a simple task. The influx of medicines from different sources into healthcare systems of developing countries presents a challenge in monitoring their origins and, more importantly, their quality. This situation is further compounded by the surge of generic products in the market [9] and the increased interest in purchasing pharmaceuticals through Internet sales. More than 50% of the medicines purchased over the Internet are from illegal sites that did not display their physical addresses and have been found to be counterfeited [7].

The Malaysian pharmaceutical industry is an example of the problems that a developing country may experience. The Pharmaceutical Service Division of the Ministry of Health Malaysia (DCA) is the organisation that is responsible for carrying out enforcement activities to ensure the safety, quality, and efficacy of pharmaceuticals, health care, and personal care products that are marketed in Malaysia. One of the agency's ways to achieve these objectives is by monitoring the quality of registered products on the market. Over time, raids have been carried out on individuals, companies, and businesses following inputs from the intelligence-gathering unit, diversion control unit, precursor unit, and other sources [10]. Thousands of products that cost billions of Malaysian ringgit (RM) have been seized each year by the enforcement division in their reactive measures to overcome these drug quality issues [11, 12].

This study was initiated by the Faculty of Pharmacy at the National University of Malaysia as part of the fight against the public health problems of rising counterfeit, substandard, and poor quality medicines and herbal products. A simple and quick drug screening procedure using a spectral database of medicines has been developed for the identification and classification of drugs, particularly samples with poor labelling or from unknown origins, without relying on reference standards for comparisons.

This spectral database should assist the DCA in conducting regular post-marketing surveillance or periodical investigations on over-the-counter products. It is also suggested to be a cost-effective method for screening mass products in

the market and selecting appropriate samples that require more complicated and expensive analysis.

1.1.1 Adulterated and counterfeit medicines – a Malaysian review

The issue of drug counterfeiting is a growing concern in Asia. The corporate affairs director of Pfizer Malaysia Narinder Kaur said that counterfeit drugs of Aricept, Celebrex, Diflucan, Feldene, Lipitor, Norvasc, Ponstant, Zoloft, and Viagra have been found in Asia [13]. In Malaysia, attention is being particularly paid to unregistered and adulterated products, a classification of counterfeit medicines that can jeopardise the safety of consumers.

It is difficult to estimate the depth of these problems because of under-reported cases and the difficulty in detecting, investigating, and quantifying counterfeit medicines. However, the Ministry of Health has revealed that the value of counterfeit medicines confiscated in the Malaysian market in 2007 amounted to RM 35.8 million compared with RM 25.9 million in 2004 [11]. This information came as a shock to many Malaysians as there has never been a serious warning on particular products, product recalls, or reported cases in order to attract public awareness.

In 1997, the Ministry of Health established that the pervasiveness of counterfeit drugs was approximately 5.8% [14]. A survey in 2005 by the Pharmaceutical Association of Malaysia (PhaMA) discovered that 14 out of 285 (nearly 5%) samples of three prescription medicines that were purchased from 196 pharmacies and clinics in six states were counterfeit [15]. In 2008, Mazlan Ismail, senior assistant director of enforcement at the DCA, said that counterfeit medicines accounted for 5.28% of over-the-counter products [11]. These estimates indicate that the prevalence of counterfeit medicines has been consistent in Malaysia for over a decade.

There is limited public accessible information about the spread of counterfeit medicines in Malaysia. PhaMA, consisting of pharmaceutical companies in Malaysia, documented the occurrence of counterfeit medicines from 1998 to 2002 on analgesics and eye drop products (Table 1.1). PhaMA has also noted other counterfeited products that has been reported, including Betnovate N

Cream (Sabah), Lomac (Kuala Lumpur), Ventolin inhalers (Sabah and Perak), and Zantac (Perak) [15].

Drug counterfeiting is a serious offence. A life is at stake when consumers take medication that is not what it is supposed to be. Weakness in the regulatory framework is believed to be one of the reasons behind the existence of counterfeit medicines in Malaysia. Table 1.2 shows a summary of drug counterfeiting cases taken to court [15].

These figures may not give a true reflection of the problem in Malaysia as the data were only based on reported cases and a survey in 2005 in selected regions. There could be many other products out there that are counterfeited and many patients could have become victims of these onslaughts.

In addition to this consistent problem, the adulteration of herbal medicines seems to be on the increase. From 2006 to 2008, 17 products for men's health, six products for weight loss, five cough medicines, and four products for joint pain were found to be adulterated with synthetic drugs to increase their pharmacological effects [16]. This could be an exceptionally dangerous problem as herbal products are highly complex and conventional single chemical entity–herbal drug combination pharmacology is poorly understood.

Therefore, a comprehensive nationwide survey needs to be conducted to get an accurate picture of this problem in Malaysia, thus making necessary interventions possible. Identifying the actual components of a counterfeit medicine is also crucial, especially in tracking illegal manufacturers.

Table 1.1 Cases of counterfeit medicines documented by PhaMA in 1998–2002 [15].

States	1998		1999		2000		2001		Until Sept. 2002	
	Analgesics	Eye drops	Analgesics	Eye drops	Analgesics	Eye drops	Analgesics	Eye drops	Analgesics	Eye drops
Kedah	1	-	1	1	-	1	2	2	-	-
Penang	5	-	-	-	-	-	-	-	-	-
Perak	-	2	3	5	-	1	-	-	-	-
KL	2	-	4	7	-	-	-	-	-	-
Selangor	-	-	3	4	1	-	-	4	-	4
Malacca	1	1	1	-	-	-	-	-	-	3
Johor	1	1	1	3	1	2	1	4	1	4
Pahang	-	-	1	1	-	-	-	-	-	-
Kelantan	-	3	3	6	6	5	-	-	1	-
T'ganu	-	-	-	1	-	1	-	-	-	-
Sarawak	1	-	-	-	-	-	-	-	-	-
Sabah	1	-	1	1	-	1	-	-	2	3
Total	**12**	**7**	**18**	**29**	**8**	**11**	**3**	**10**	**4**	**14**

Table 1.2 Drug counterfeiting cases taken to court and the penalties assigned [15].

Year	Place	Crime	Penalty
1997	Ipoh	A sundry shop owner having in possession and supplying for sale 140 counterfeit Panadol (paracetamol, GSK) tablets.	RM 5,000 fine and 6 months' imprisonment. Reduced to RM 10,000 without jail after appeal.
1999	Not stated	A manufacturer producing counterfeit Panadol and having in possession about RM 500,000 worth of goods and equipment.	RM 15,000 fine.
1999	Not stated	Three men were acquitted on a charge of manufacturing 47,300 tablets of counterfeit Panadol.	Prosecution has failed as there was insufficient evidence.
2000	Kota Kinabalu	MDTCA seized 3 and 6 units of Ventolin inhalers in two different pharmacies.	RM 5,000 fine for both cases.
2003	Johor	Three pharmacies in Johor Bahru were found to sell counterfeit products imported from Singapore.	Case not concluded.

1.2 Near-infrared (NIR) spectroscopy for drug analysis

Near infrared spectroscopy (NIRS) has become a valuable analytical tool in pharmaceutical analysis. Using NIRS, drug analysis has become easier and faster as generally sample preparation is not required. The ability to analyse samples intact is another advantage of the technique whereby samples can be retained for further analysis or as evidence for forensic purposes.

The NIRS was shown to be useful for identification of pharmaceutical substances and dosage forms based on the chemical information from active ingredient or the excipient independently and the physical properties of the sample [17]. This information-rich spectrum provides a unique fingerprint for each product which is very advantageous particularly for the identification and detection of counterfeit and substandard medicines.

The most common technique in counterfeit drug detection by NIRS is through the comparison of suspected tablets with the spectra of authentic tablets [18, 19, 20, 21]. This analysis requires the original products from manufacturer to obtain reference spectra. However, many manufacturers, especially in developing countries, refuse to co-operate in drug quality studies leading to problems in acquiring the reference samples.

An alternative approach for drug identification by NIR is to determine the active pharmaceutical ingredients (APIs) [22] or the sources of the tablets [23]. In both techniques, the applications of multivariate analysis are essential to process the highly overlapped and broad spectra produced by the instrument to get the required information.

A "fast drug identification system" (FDIS) was developed in China for the quick inspection of medicines sold in rural areas in an attempt to fight against fake and substandard medicines. Mobile vehicles were fitted with NIR pre-screening systems [24, 25] and fast chemical identification equipment [26, 27]. The application of NIRS analysis in these systems is based on the universal qualitative [24, 25] and quantitative models [28, 29]. A universal model is a calibration model that contains products from different manufacturers. Such models can be used to identify the sample by comparing its spectrum with the other spectrain

the model using partial least squares (PLS). However, many different factors need to be considered, such as the variation in product formulation and the types or amount of excipients. This highlights the need to have a sufficient number of samples that will cover these variations in the model [29].

Besides qualitative analysis, NIR spectroscopy is used for quantitative assaying of intact tablets as the most obvious way to detect substandard preparations [30, 31]. These methods are based on developing calibration curves of APIs [32]. In all these studies, information about the tablets analysed were available and the analyses were designed based on this knowledge.

The combination of NIRS and chemometric analyses such as PCA, PLS and SIMCA, has proven to be a simple, fast and robust drug analysis procedure. However, application of these methods mostly depends on availability of reference compounds or products for comparison with the new spectra.

However, one of the real problems occurring in developing countries is a consequence of medicines being sold or dispensed without their original packaging; tablets are either wrapped in a piece of paper or given in an envelope or a plastic bag, where the information about the tablets may be given verbally or just handwritten on the container [33]. Poor and improper labelling has made identification of medicines to be almost impossible using the NIRS techniques mentioned earlier. Without the information about the tablet content, it is not possible to choose either a reference sample for comparison or property of the tablet to be used for development of the calibration curve.

In this scenario, the questions to be answered by analytical methods are more complex: has the correct medicine been dispensed? What is the composition of the dispensed formulation? Is the dispensed medicine counterfeit?

1.2.1 The use of NIR spectral library in pharmaceutical industry

NIR spectroscopy has gained wide acceptance for various pharmaceutical applications [34, 35]. One major current application of this technique is for Process Analytical Technology (PAT). PAT can be defined as "a system for designing, analysing, and controlling manufacturing through timely

measurements (i.e. during processing) of critical quality and performance attributes of raw and in process materials with the goal of ensuring final product quality" [36]. Figure 1.1 shows the PAT opportunities at typical pharmaceutical manufacturing operations.

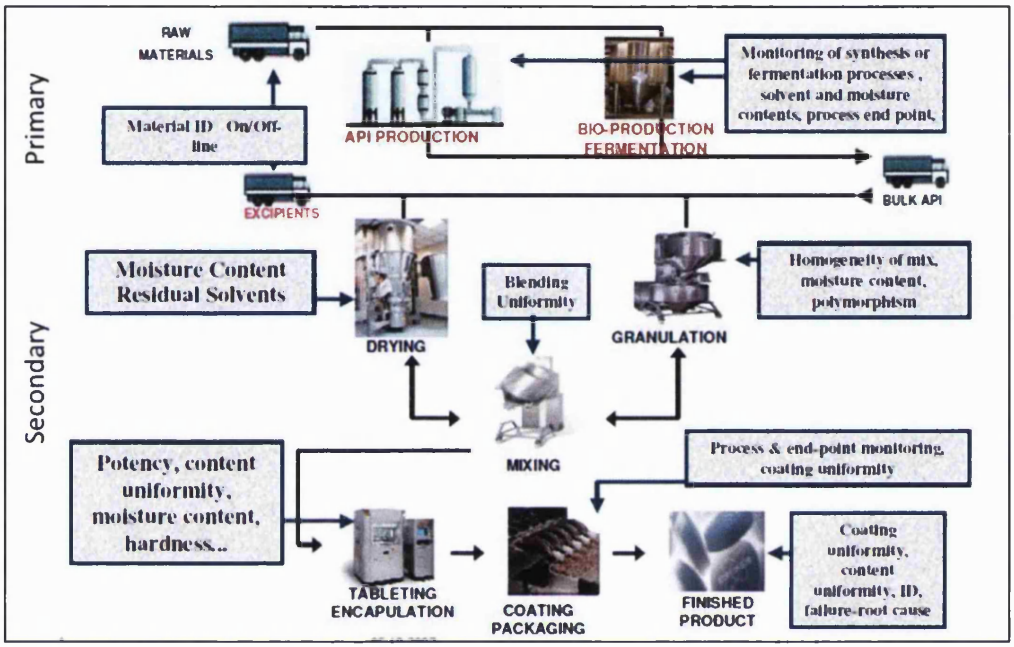


Figure 1.1 PAT opportunities in typical pharmaceutical manufacturing operations [adapted from: Li H, ref 37].

The ease of sampling, rapid analysis and the possibility of analysing a sample intact or in its packaging are some of the factors that led to the increasing use of NIR spectroscopy and NIR libraries in PAT, replacing conventional chromatography methods like high performance liquid chromatography (HPLC) and thin-layer chromatography (TLC) which are more costly and time consuming. One of the main roles of NIR in PAT is to ensure the right materials are used in every process. The procedure in qualitative analysis by NIR is illustrated in Figure 1.2.

In general, there are two steps involved in the identification process by NIR. Firstly, by constructing a 'spectral library' which consist of numerous individual spectra that are averaged to form a representative mean library spectrum. Once the library is validated, the spectrum of a test sample is acquired and compared

with those in the 'spectral library' on the basis of mathematical criteria for spectral similarity.

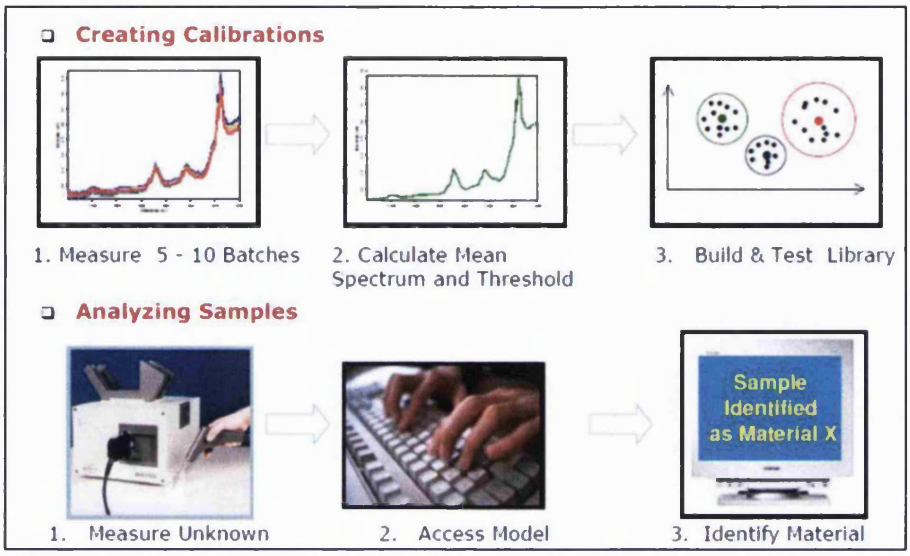


Figure 1.2 Qualitative analysis of sample by NIR. The calibration process involved spectra acquisition, inclusion of mean spectrum in the library followed by determination of the threshold and validation of the library. To analyse new sample, the spectrum is acquired and compared to the other spectra from the pre-built library [adapted from: Li H, ref. 37].

A common approach used for expressing similarity is the correlation coefficient, which is defined as the inverse cosine of the angle between the multi-wavelength vector representing the mean library and test spectra. If the similarity coefficient exceeds a pre-set threshold, then the two spectra compared are considered similar and the samples can be assumed to be similar. There are several guidelines and suggestions on how to construct and interpret NIR libraries for the pharmaceutical industry [38, 39, 40].

Different terms were used to depict the correlation coefficient strategy. For example, van der Vlies *et al.* has used the term spectral match value (SMV) for the identification of different types of cellulose [41] and ampicillin trihydrate [42]. Blanco *et al.* has demonstrated the use of discriminative correlation coefficient known as match index (MI) for the identification of pharmaceutical preparation including active compounds, excipients, vitamins and amino acids [43].

Although the correlation value represents similarity between the spectrum of the unknown sample and the average library spectrum, the correlation value does not have direct, physical meaning based on principles. Because of this, a pass/fail threshold limit should be determined empirically during method development and validation. This is also known as the 'qualification' method which involves expressing similarity based on distance terms in order to determine whether a sample falls within the normal variability range.

John and Pixley have described a statistical data driven decision making process to aid in the selection of correlation threshold value for the identification of pharmaceutical products [44]. The selection of the threshold value was based on the relationship between positive and negative control correlation values in a single product library generated from spectra taken across four different potencies and each with unique coating compositions.

The most common mathematical application in the 'qualification analysis' is by using the Mahalanobis distance. Dreassi *et al.* used this approach to monitor the quality of antibiotics during production, where the samples were characterized at different stage of the process [45] and distinguished from other products produced in the same production area [46].

Alternative to direct spectral computation is the use of PCA, where the correlation and distances are calculated from scores in the space bound by the principal component. This method has been used to identify and classify variety of substances using general identification library [47, 48], or single product library [49, 50, 51]. Shah and Gemperline used Mahalanobis distance in PC space to qualify different batch of Acivel PH101 microcrystalline cellulose [52]. It was assumed that a sample was qualified when the probability level for a X2 distribution fell in the range 1.0-0.05.

Qualitative analysis using standard algorithms and setting up acceptable threshold for qualification is a well know and commonly used method in the industry for products identification and classification.

1.3 Spectral database

A spectral database can be defined as a collection of chemical representations of products or compounds that can be used to identify other similar substances. There are over 10 million chemical compounds known to the scientific world today. However, most spectral databases cover as few as 100,000 compounds [53]. Most of them are also specialised in one particular technique [54] or within a very specific subject area [55]. A comprehensive list of spectroscopic databases can be viewed elsewhere [53, 56, 57]. However, this is not an exhaustive list as database development is a continuous process and keeps on expanding.

Building a spectral database is tedious, time consuming and costly in terms of obtaining samples and maintaining the laboratory infrastructure throughout the whole process. The data for the databases can also be collected from different labs or even from the literature [58]. This is a quicker and cheaper procedure to expand the database but with uncertain quality. In this work, the quality of the data for the database was ensured by using standard procedures during the acquisition of spectral data for each of the methods used and by maintaining standard instrument specifications throughout the analysis.

The utilisation of the spectral database for the identification of a compound or a product is based on algorithms developed to match an unknown spectrum to the reference spectra in the database. Criteria or statistical thresholds need to be defined to positively identify a sample. When an unknown sample spectrum in the database meets these criteria, then there will be confidence that the sample has been correctly identified.

1.3.1. Full spectrum search

Several research groups in the 1970/80s developed the full spectral search technique derived from pattern recognition methods [59, 60, 61]. In general, the unknown spectrum and each reference spectrum are represented as points in a multidimensional space, where each dimension corresponds to a particular wavenumber location in the spectrum. The similarity of the unknown spectrum to each reference spectrum is then computed as the 'distance' between the two

points in the 'hyperspace'. This is the basis of the nearest-neighbour technique, which assigns features to an unknown compound based on the features contained in the compounds whose spectra are 'closest' to it. In a full spectrum search, the spectra are often 'de-resolved' from the complete library spectrum, which allows for faster searching, but retains enough fine features of the peak position, intensity, and bandwidth [62].

Spectral ID software provides seven different algorithms for full spectrum searches: Euclidian distance, absolute value, first derivative absolute value, least squares, first derivative least squares, correlation, and first derivative correlation. The equations used to calculate the differences and similarities between the unknown and known spectra for each algorithm are described below [59, 62, 63]. The hit quality index (HQI) is calculated where A_i is the absorbance values of the i^{th} point in the unknown spectrum, L_i is the corresponding point in the reference spectrum being searched (in the database), and n is the number of wavelengths in each vector.

1.3.1.1 Euclidean Distance

The Euclidean distance of an unknown spectrum to a database entry is effectively a normalised least squares dot product on the unknown:

$$HQI_{ed} = 1 - \left[\frac{\sqrt{\sum_{i=1}^{n} A_i L_i}}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} L_i^2}} \right]$$

The best match (library=unknown spectrum) is given at a distance of 0 and the worst given at a distance of 1.41421 according to the values returned by the full spectrum search of the many commercial search programs that use the GIFTS baseline data reduction (as described in section 1.6.1.3, page 70) and the Euclidean Distance search algorithm [64]. This algorithm is the industrial standard and the most common algorithm used in performing library data searches [64].

1.3.1.2 Absolute Value

The Absolute Value HQI of the unknown from a library entry is calculated as the sum of the absolute difference vector:

$$HQI_{av} = \frac{\sum_{i=1}^{n}|L_i - A_i|}{n}$$

Each library or unknown data point is 0 (best match) to 1 (worst match). This algorithm is suitable for cases where there are additional large spikes in the unknown but provides poor discrimination and typically returns many library entries with identical or similar HQI values [65].

1.3.1.3 Least Squares

The Least Squares HQI of the unknown from the library entry is calculated as the sum of the squares of the vector difference:

$$HQI_{ls} = \frac{\sum_{i=1}^{n}(L_i - A_i)^2}{n}$$

Similar to the absolute value algorithm, each library or unknown data point is 0 to 1. Therefore, the best match is 0 and the worst is 1.

1.3.1.4 First derivatives absolute value/least square

Both algorithms are similar to their non-derivative counterparts except that the derivative of the unknown and library entry are calculated before the HQI is derived. This derivative is simply the difference between each set of points in the spectrum:

$$HQI_{1d.av} = \frac{\sum_{i=1}^{n}[(L_i - L_{i-1}) - (A_i - A_{i-1})]}{n}$$

$$HQI_{1d.ls} = \frac{\sum_{i=1}^{n}[(L_i - L_{i-1})^2 - (A_i - A_{i-1})^2]}{n}$$

Using derivatives gives the advantage of removing most of the errors arising from the varying baselines. These are especially useful when the baseline error cannot be removed by the standard GIFTS algorithm.

1.3.1.5 Correlation

The correlation algorithm is similar to the Euclidean Distance algorithm with one distinction; both the unknown and the library data are centred on their respective means before the vector dot products are calculated:

$$HQI_c = 1 - \frac{(L_m \times A_m)^2}{(L_m \times L_m)(A_m \times A_m)}$$

where the vectors are defined as:

$$L_m = L - \frac{\sum_{i=1}^{n} L_i}{n} \qquad A_m = A - \frac{\sum_{i=1}^{n} A_i}{n}$$

The advantage of using this algorithm is that both the known and unknown spectra have the same mean values when the signal-to-noise ratio and baselines in the library and unknown spectra are similar.

1.3.1.6 First derivative correlation

This method is the same as the correlation algorithm, except that the first derivatives of both the unknown and library spectra are used to calculate the HQI. The first derivative is taken by subtracting previous points during mean centring calculations and removing the non-linear effects of the baseline that cannot be compensated by baseline pre-processing or correlation algorithms.

$$L_m = L - \frac{\sum_{i=1}^{n} Li - L_{i-1}}{n} \qquad A_m = A - \frac{\sum_{i=1}^{n} A_i - Ai - 1}{n}$$

1.3.2 Peak search

The peak search systems are conducted using the peak and an intensity level [66]. Different methods include incorporating the molecular formula and chemical structure information in the location of peak maxima to the nearest tenth micrometre[67] and using information theory to improve the selection of spectral ranges whereby the peak intensity information is used as encoded data [68].

In this software system, each library entry includes a peak table containing up to 127 peaks, whereby each peak is identified by its X axis value and intensity between 0 and 9. The largest peak in the entry is always normalised to an intensity of 9 and the other peaks are scaled relative to the largest [69]. The two main problems of this search procedure are the lack of intensity value and the selection of predefined windows [70]. Some parameters to consider in conducting a peak search include the search type (forward or reverse), pick picker (standard or double sided), level, and sensitivity.

1.3.2.1 Forward and reverse search

The forward and reverse search type determines how the unknown peaks compare with the peak tables in selected libraries. The forward peak search takes each peak in the unknown spectrum and compares it to the reference spectra in the database, while in a reverse peak search, as its name indicates each peak in a reference spectrum peak table is compared with the peaks of the unknown spectra. When an unknown spectrum is loaded into the database, the peaks are extracted using the same algorithm used to extract the reference spectra. A score is then calculated according to the weighting described below [62].

For the forward peak search, the absolute minimum (of X distance) between the unknown spectrum and each of the reference spectra is calculated by subtracting the X value of the first unknown peak from the X values for each peak of the reference spectra. A score of this unknown peak is assigned based on the values in Table 1.3 (A).

Next, the absolute value (of amplitude difference) is calculated by subtracting the hit amplitude of the unknown peak (values 1 to 9) from the value of the nearest

library peak. A score of this unknown peak is assigned based on the values in Table 1.3 (B).

The absolute value (of amplitude difference) times 10 is subtracted from the score value for the peak (of X distance) determined in Table 1.3 (A) and the result is added to the total score for the current library entry. If the result is less than 0, no hit value is added to the total. This process is repeated for each peak in the unknown, and then the total is normalised by dividing by the number of peaks in the unknown to get a hit score between 0 and 100.

Table 1.3 (A) The score value representative of the absolute peaks in X distance and (B) the score value representative of the absolute peak in Y distance. The absolute peak was obtained by subtracting the hit amplitude of the unknown peak from the corresponding peak in the reference.

A

Absolute peak (X distance)	**Score for peak**
0 or 1	100
2	80
3	40
4	20
5	10
6	8
7	4
8	2
9	1
Greater than 10	0

B

Absolute peak (amplitude difference)	**Subtracted from peak score**
0	0
1	10
2	20
3	30
4	40
5	50
6	60
7	70
8	80

1.4 Sampling theory

1.4.1 Sampling techniques

Sampling comprises the operations designed to select a portion of a pharmaceutical product for a defined purpose [71]. The choice of sampling plan should always take into consideration the specific objectives of the sampling and the risks and consequences associated with inherent decision errors. In this research, the appropriate sampling procedure was chosen to obtain a general idea about the quality of medicines in selected areas.

Random, convenience, and lot quality assurance sampling (LQAS) are three types of sampling procedures that are commonly used in drug quality surveillance/post-marketing monitoring. Table 1.4 summarises their advantages and disadvantages. All sampling techniques, except convenience sampling, require a sampling frame from which the sample can be drawn. It is important to design a sampling frame that represents the whole population, as this will enable the results to be generalised throughout. However, this requires detailed sampling locations (for example the location of each pharmacy outlet in the sampling area) and this information may be difficult to obtain.

Owing to this reason, many previous surveys on drug quality have been conducted using the convenience method [72, 73, 74] compared with random sampling [75, 76, 77] or LQAS [78, 79]. Convenience sampling, although naturally prone to biases, can provide an initial indication of a problem. If required, this could open up subsequent detailed surveys [80].

1.4.2 Sampling personnel

Who should perform the sample depends largely on the regulatory status of the medicines and on whether or not the drugs sellers know the quality of drugs that they are selling. Open sampling is one method whereby the drugs seller is introduced to the sampler and the purpose of the survey is explained. However, this will only be beneficial if sellers are concerned about the quality of products they are selling. Previous reports have showed varying levels of knowledge among drugs sellers regarding the quality of their wares [72]. Another sample

collection method is using 'mystery shoppers', whereby the identity of the collector and the sampling objective are undisclosed.

1.4.3 Ethical and legal aspects of sampling

Some ethical issues must be considered during sampling for drug quality assessment; (i) informed consent from the seller is not always necessary but they have the ethical responsibility to confirm or refute the suspicion (if it arises) as soon as possible, (ii) it is potentially difficult for the researcher to buy all the stock of an essential medicine when there are only a few suppliers, (iii) depending on the local pharmaceutical law, sampling where prescriptions are required may be more difficult, and (iv) medicine sampling by academics is unlikely to be used in a court of law as evidence as it is not normally collected using legally robust 'chain of evidence' procedures.

Table 1.4 Different type of sampling techniques commonly used in drug quality surveillance.

Type of sampling	Description	Advantages	Disadvantages
Convenience survey	-No specific guidance -Choice of outlet depending on the collector - Findings could highlight the initial indication of a problem	-Simplest method -Minimal cost -Time consuming	-Prone to bias -Results are specific to sampling area and cannot being generalised
Random sampling	-More objective method -Requires proper planning		-Large sample size -Labour costs
LQAS	-Pre-assessment to provide binary result before formal random sampling to determine if prevalence exceeds the acceptable threshold	-Smaller sample size -Economical methods	-Does not estimate an exact prevalence

1.5 Spectroscopic analysis

1.5.1 NIR

1.5.1.1 Principles

Harmonic oscillator

Figure 1.3 shows the simplest model of a diatomic molecule, which is connected with a rigid bond that can only rotate about its centre of gravity. The forces of attraction and repulsion both occur between two bonding atoms. Repulsive forces exist between the positively charged nuclei of the two atoms and between the two negatively charged electron clouds. Attraction forces exist between the positively charged nuclei of one atom and the negatively charged electron cloud of the other. When the total energy of the system is at its minimum, the two atoms positioned at the mean inter-nuclear distance and the forces are balanced.

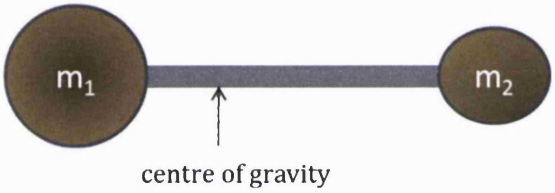


Figure 1.3 Illustration of a pair of rigid diatomic molecules.

For a molecule to vibrate, the bond must be elastic. Therefore, energy is required to overcome the forces of attraction and repulsion existing between the two atoms. The stretching and squeezing of the bond can be approximated to a single mass (equal to the reduced mass) attached to a spring with one end fixed.

The reduced mass of the system, μ, is defined as:

$$\mu = \frac{(m_1 m_2)}{(m_1 + m_2)}$$

where m_1 and m_2 are the masses of the two atoms.

Simple harmonic oscillators

The disturbance of mass along the axis of a stretching and squeezing spring is described by Hooke's law [69]:

$$F = -k\,(r - r_{eq})$$

where$F$ is the restoring force, k is the force constant of the bond, r is the inter-nuclear distance, and r_{eq} is the equilibrium inter-nuclear distance. The potential energy, E, of this system is a parabolic function of the inter-nuclear distance, given by:

$$E = \frac{1}{2}\, k\,(r - r_{eq})$$

For a diatomic molecule, the vibrating mechanical model can be described by the classical vibrational frequency, v_f:

$$v_f = \frac{1}{2}\pi\sqrt{\frac{k}{\mu}}$$

The harmonic potential energy curve is shown in Figure 1.4.

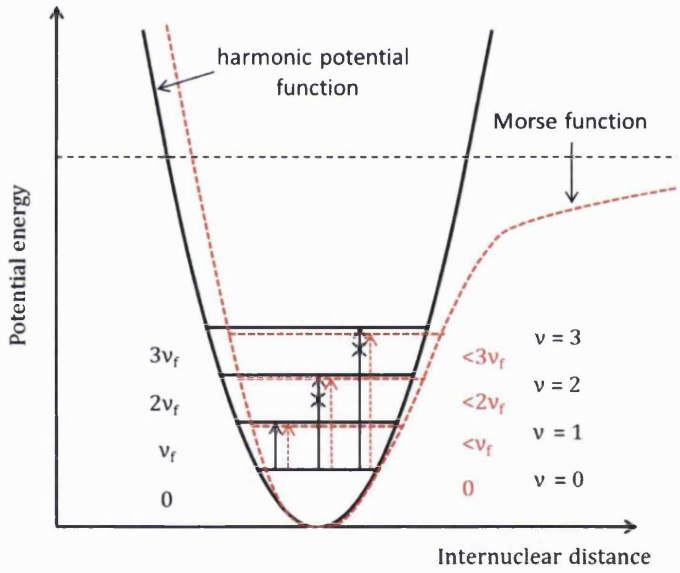


Figure 1.4 Harmonic (Hooke's law) and anharmonic (Morse function) potential functions for a diatomic oscillator [adapted from: Ciurczak, ref. 81].

The theory of quantum mechanics predicts that the vibrational energy for a harmonic oscillator can only have certain discrete values, which are equally spaced:

$$E_v = \left(v + \frac{1}{2}\right) h v_f$$

whereE_v is the value of the energy level, h is Planck's constant, v_f is the classical vibrational frequency, and $v$is the vibrational quantum number, which can only take integer values and change by ± 1.

The energy is at the minimum energy state when v=o, given by:

$$E_v = \frac{hv}{2}$$

At room temperature, most molecules exist in the v=0 vibrational state. Thus, the transition v=0 →v=1 is the most common; this is referred to as the fundamental transition that normally occurs in the mid-IR region.

Anharmonic oscillators

Real molecules, however, are not harmonic oscillators and they do not obey Hooke's law exactly. When a bond is stretched, extension is easier than is squeezing and it is subjected to a maximum point where the bond breaks. Morse [82] described the potential energy curve illustrating these properties, given by:

$$V = D_e\left(1 - e^{\beta(r_{eq}-r)}\right)^2$$

where β is a constant and D_e is the dissociation energy (Figure 1.2).

The energy levels for the Morse function are no longer equally spaced according to the Schrödinger wave equation:

$$E_v \approx \left(v + \frac{1}{f2}hf\right) - \left(v + \frac{1}{2}\right)^2 hfx$$

where*f*is the equilibrium oscillation frequency and x is the anharmonicity constant.

The selection rules for the anharmonic oscillator are the same as those for the harmonic oscillator but with the possibility of transition between energy levels that differ by two or more vibrational quantum numbers ($\Delta v = \pm 1, \pm 2, \pm 3 \ldots\ldots$).

The absorbance at approximately f is the fundamental absorption and this generally occurs at wavenumbers less than 4,000 cm^{-1}(termed the mid-IR region). Transitions that occur in the NIR region have an absorbance of approximately 2f or 3f, and these are termed the first and second overtones,

respectively. In addition, combination bands may arise within this region when two or more vibrations interact:

$v = 0 \rightarrow v = 1, \Delta v = \pm 1 \quad \Delta E = hv\,(1 - 2x)$ Fundamental (most intense)

$v = 0 \rightarrow v = 2, \Delta v = \pm 2 \quad \Delta E = 2hv\,(1 - 3x)\, 1^{st}$ overtone (small intensity)

$v = 0 \rightarrow v = 3, \Delta v = \pm 3 \quad \Delta E = 3hv(1 - 4x)\; 2^{nd}$ overtone (min. intensity)

Modes of vibration

Vibrations can be described as the bending or stretching of bonds. Bending is caused by a change in the bond angle (Figure 1.5 a, b, c, and d), while stretching is a change in the diatomic distance along the axis of the bond (Figure 1.5 e and f).

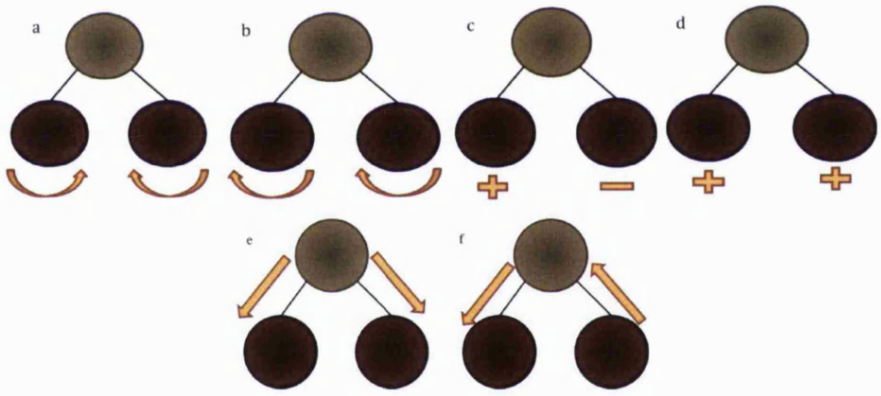


Figure 1.5 Modes of vibrations for a non-linear tri-atomic molecule: a) symmetrical in plane deformation (scissoring), b) asymmetrical in-plane deformation (rocking), c) asymmetric out-of-plane deformation (twisting), d) symmetric out-of-plane deformation (wagging), e) symmetrical stretching, and f) asymmetrical stretching [adapted from: Barton, ref. 83].

Atoms move at a similar frequency as the radiation absorbed but at a different amplitude depending on the change in dipole moment associated with the vibration. For example, a lighter molecule will vibrate with larger amplitude compared with heavier ones.

The first set of spectra was recorded by Coblents [84] in 1905 and confirmed by Ellis [85] in 1922. Further research on the band assignment and spectral interpretation led to the compilation of a common reference table (illustrated in Figure 1.6) [86].

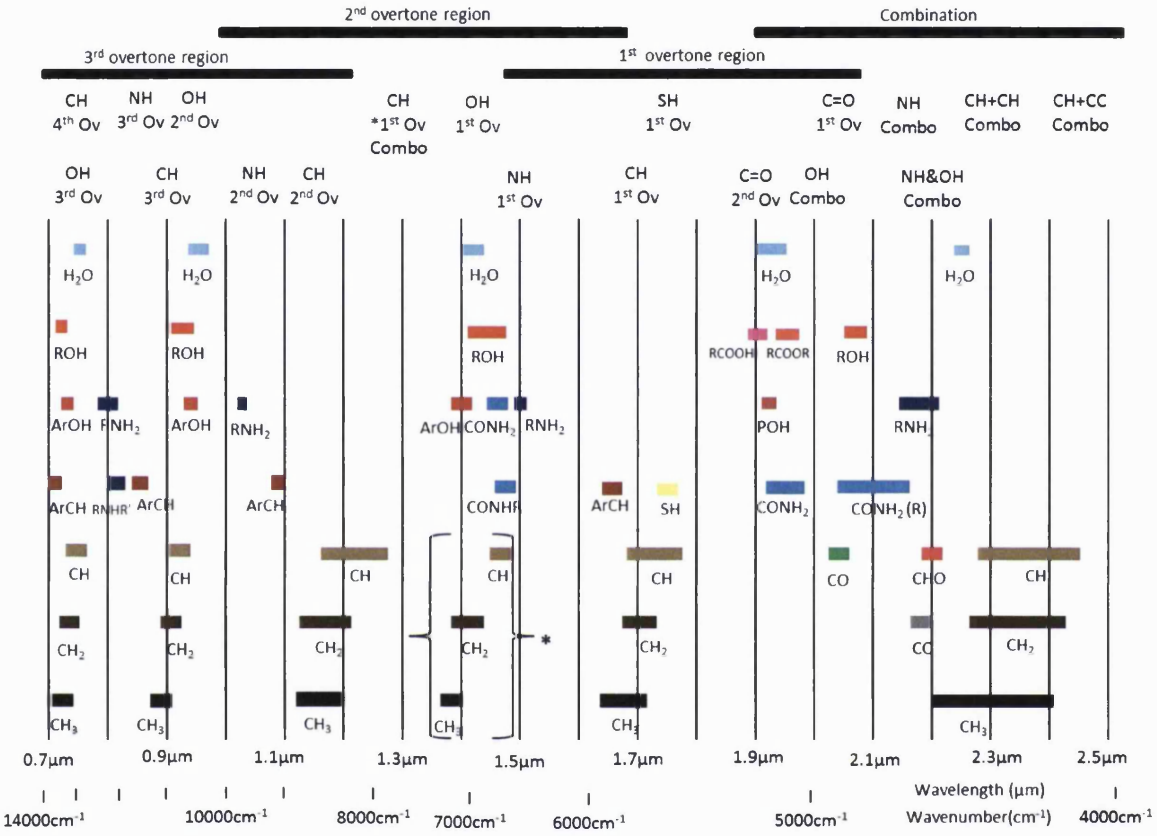


Figure 1.6 Common absorption of the band assignments in the NIR region [adapted from: FOSS, ref. 86].

1.5.1.2 Spectra acquisition

Four main techniques are commonly used to measure the interaction of radiation with matter: absorption, reflection, transmission, and transflection (Figure 1.7). In this work, only the diffused reflection and transflection techniques were used and thus these will be discussed further here. Reflection and transmission methods are most common for solid analysis, while transflection and transmission methods are more suitable for liquid analysis.

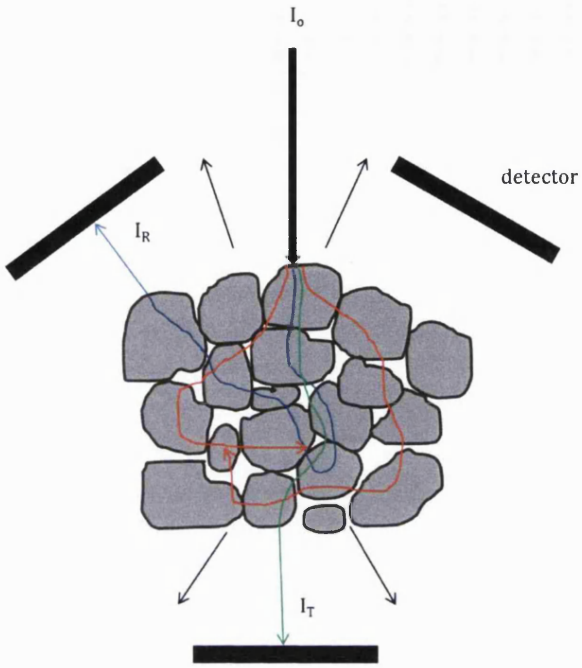


Figure 1.7 Four possible paths of radiation: incident (black), reflected (blue), absorbed (red), and transmitted (green) [adapted from: FOSS, ref. 86].

Diffused reflectance

Diffused reflection is a process where the radiation is scattered, transmitted, and absorbed by interactions within the sample. This produces a unique characterisation of the sample. The Kulbeka-Munk is the most common function used to calculate diffused reflectance.

$$f(R) = \frac{(1-R)^2}{2R} = \frac{k}{s}$$

where$k$ is the absorption coefficient and s is the scatter coefficient. This model assumes that the sample is illuminated with monochromatic radiation, the distribution of scattered radiation is isotropic, the particles in the sample layer are randomly distributed, the particles are much smaller than the thickness of the sample layer but bigger than the wavelength of irradiation, and edge effects are eliminated.

Transflectance

In a transflectance measurement, radiation passes through the sample to a reflector and back (Figure 1.8), the optical path length being approximately twice

the distance from the point of entry of the radiation to the reflector. Typically, the reflector is made from stainless steel, gold plate, PTFE, or another inert reflecting surface.

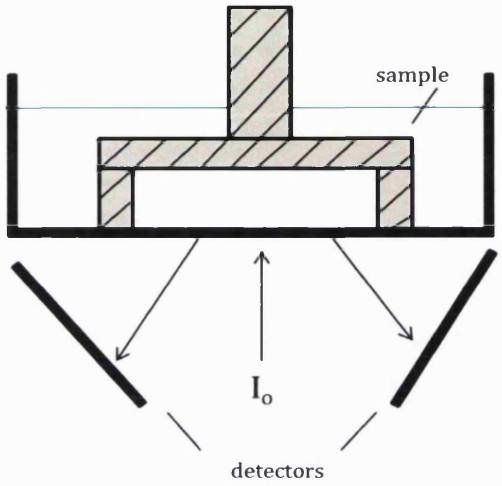


Figure 1.8 Schematic optical diagram for transflectance measurement using a sample cup and reflector [adapted from: Yoon et al., ref. 87].

1.5.1.3 Fundamentals of NIR instrumentation

In this work, the FOSS NIRSystem 6500 was used for NIR reflectance spectrum acquisition over the wavelength range of 1100 nm to 2498 nm. In general, the NIR instrument consists of the source, wavelength selector, sample stage, and detector (Figure 1.9).

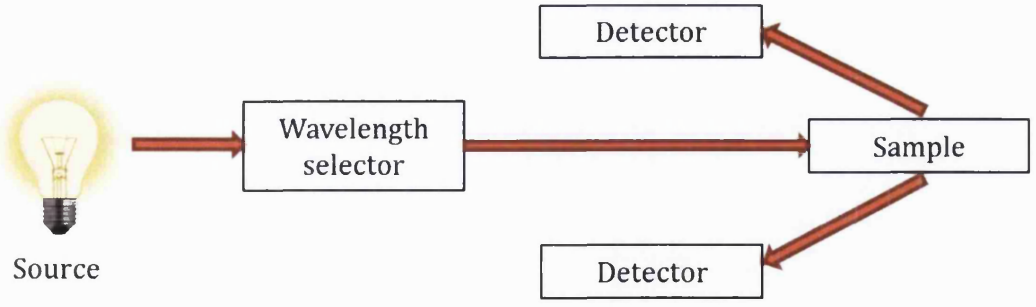


Figure 1.9 Schematic layout of the NIR reflectance instrument [adapted from: FOSS, ref. 86].

There are two types of sources available: thermal and non-thermal. The most common, which is also used in the FOSS NIRSystem 6500, is a tungsten halogen lamp. Non-thermal sources include lasers, discharge lamps, and light-emitting diodes. These types of sources in general are more compact and have a narrower

emitting wavelength range, which makes them compatible with specific detectors. The thermal source, however, is inexpensive compared with the non-thermal one.

The wavelength selector is the compartment that defines each instrument. Three of the most common types are dispersive, interferometer, and non-thermal (the FOSS NIRSystem 6500 is a dispersive instrument). Dispersive instruments generally use gratings to produce angular dispersion. Figure 1.10 illustrates the use of a reflective monochromator (grating) in a FOSS NIR instrument.

The sample stage of the instrument is an enclosed compartment attached to the spectrometers. It consists of an aperture for sample position with an adjustable pin for positioning the sample centrally.

Finally, the most common type of detector used is the lead sulphite (PbS) detector, a photoconductive mode (another type is the photovoltaic mode), with high sensitivity over the typical NIR range of 1000 to 2500 nm.

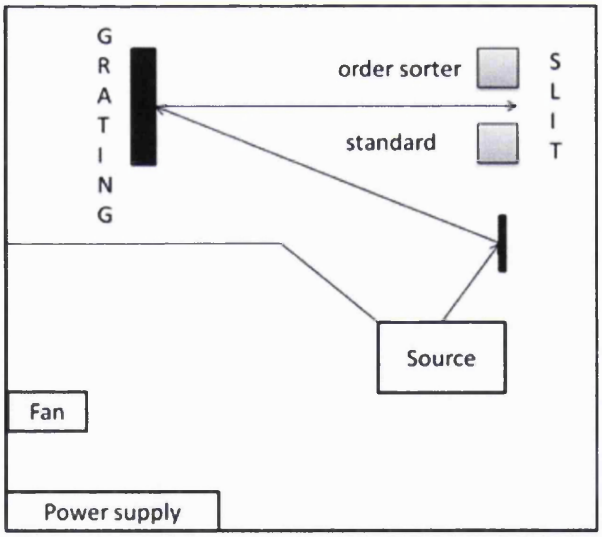


Figure 1.10 Schematic illustration of the FOSS NIRSystem 6500 monochromator [adapted from: FOSS, ref. 86].

1.5.2 Mass spectrometry (MS)

1.5.2.1 Principles

Mass spectrometry is an analytical tool used for measuring the molecular mass of a sample. The first step in mass spectrometric analysis is sample introduction

into the ionisation source, whereby the gas-phase ions of the sample molecules (M) are produced and ionised because ions are easier to manipulate than are neutral molecules:

$$M + e^- \rightarrow M^{\bullet+} + 2e^-$$

This molecular ion then undergoes fragmentations producing either a radical (R) and an ion with an even number of electrons (EE^+) or a molecule (N) and a new radical cation (OE^+):

$$M^{\bullet+} \rightarrow EE^+ + R^\bullet$$

$$M^{\bullet+} \rightarrow OE^+ + N$$

These two types of ions have different chemical properties and further fragmentations can take place on these primary product ions.

The ions are then extracted into the analyser region of the mass spectrometer where they are separated according to their mass-to-charge (m/z) ratios. The separated ions are detected, and this signal is sent to a data system where the m/z ratios are stored in the proportion of their relative abundance, producing the mass spectrum. Figure 1.11 shows the illustration of the fundamental principal of the instrument.

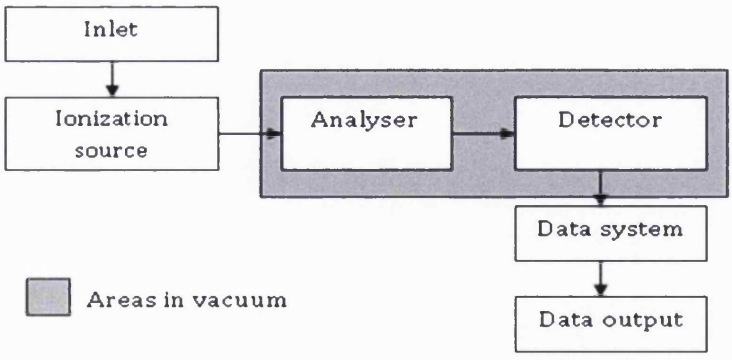


Figure 1.11 The key components of a mass spectrometry system [adapted from: Navigator training manual; ref. 88].

1.5.2.2 Sample introduction

Samples can be introduced in two ways: directly to the ionisation source (used in this research) or via chromatography en route, which allows the sample to be

separated into different components before further analysis. The latter methods usually involve the MS being coupled to another instrument: high pressure liquid chromatography (HPLC), gas chromatography (GC), or capillary electrophoresis (CE). Either way, two factors that determine which technique used are the choice of ionisation method and the type and complexity of the sample.

Ionisation source

Different types of ionisation methods have their own advantages and disadvantages [89]. Only two methods were used in this PhD work: electrospray ionisation (ESI)-MS with an ion-trap analyser and matrix-assisted laser desorption ionisation/time-of-flight (MALDI-TOF).

1.5.2.2(a) ESI

ESI is a soft ionisation technique that is useful when considering biological molecules that have a large molecular mass (polar molecules ranging from less than 100 Da to more than 1,000,000 Da), as this process does not fragment the macromolecules into smaller charged particles, but rather turns the macromolecule being ionised into small droplets.

The use of ESI analysis expanded after the success of works by Fenn *et al.* [90, 91]. In a standard ESI process [92], the sample is dissolved in a polar volatile solvent and pumped through a narrow, stainless steel capillary at a flow rate of between 1–10 μl min^{-1}. A high voltage of 3–5 kV is applied to the tip of the capillary, which is situated within the ionisation source of the mass spectrometer. This strong electric field causes the sample emerging from the tip to be dispersed into an aerosol of highly charged droplets. This process is aided by a coaxially introduced nebulising gas flowing around the outside of the capillary. This gas, usually nitrogen, helps direct the spray emerging from the capillary tip towards the mass spectrometer (Figure 1.12).

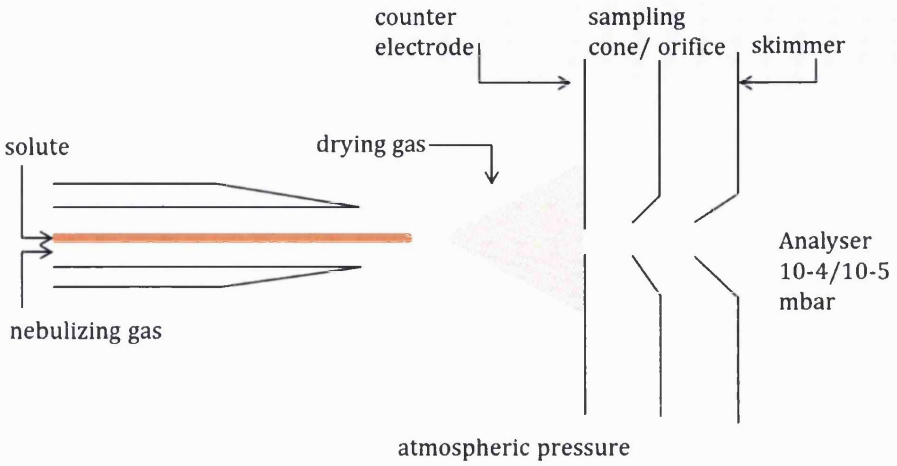


Figure 1.12 Standard ESI source [adapted from: Navigator training manual; ref. 88].

The charged droplets reduce in size by solvent evaporation (Figure 1.13), assisted by a warm flow of nitrogen known as the drying gas. As the droplet evaporates, the electrical field increases and ions move towards the surface until eventually the charged sample ions are free from solvent and are released from the droplets. The ions pass through a sampling cone or orifice into an intermediate vacuum region and from there through a small aperture into the analyser of the mass spectrometer, which is held under a high vacuum. The lens voltages are optimised individually for each sample.

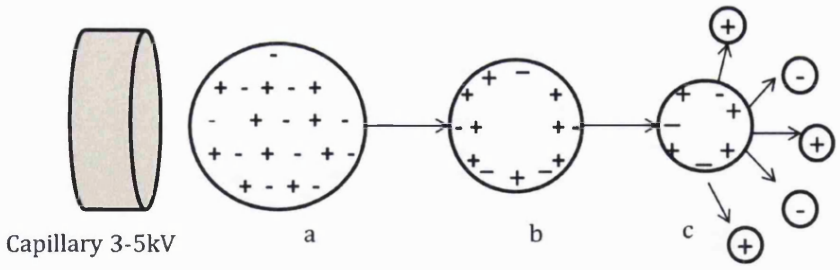


Figure 1.13 The ESI process. (a) Droplet containing ions released from the capillary, (b) as the droplet evaporates, the electrical field increases and ions move towards the surface, and (c) ions evaporated from the surface move into the analyser [adapted from: Navigator training manual; ref. 88].

Ion-trap analyser

The ion-trap analyser was invented by Paul and Steinwedel [93] and was modified to become a useful mass spectrometer by Stafford *et al.* [94]. It consists

of three hyperbolic electrodes: the ring electrode, the entrance end cap electrode, and the exit end cap electrode (Figure 1.14). These electrodes form a cavity in which it is possible to trap (store) and analyse ions. Both end cap electrodes have a small hole in their centres through which the ions can travel.

Ions produced from the source enter the trap through the quadrupole and the entrance end cap electrode. In the trapping cavity, a three-dimensional quadrupole potential field is produced from the overlap of an AC potential with the ring electrode RF potential. This traps ions of all masses in a stable oscillating trajectory within the cavity. Various voltages are applied to the electrodes to trap and eject ions according to their *m/z* ratios.

During detection, the electrode system potential is altered to produce instabilities in the ion trajectories and thus eject the ions in the axial direction. The ions are ejected in order of increasing *m/z* ratio, focused on the exit lens, and detected by the ion detector system.

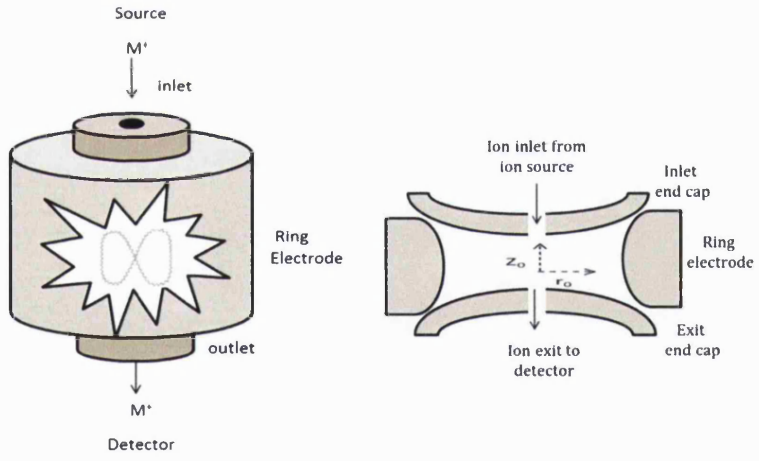


Figure 1.14 Schematic diagram of an ion-trap mass analyser. The trap is made up of the 2' end cap electrodes and the ring electrodes. Inside the trap, the ions rotate and oscillate in an eight-shape trajectory [adapted from: Hoffman; ref. 95].

In MS/MS analysis, the precursor ion is selected inside the trap where an inert gas is introduced for collision-induced dissociation (CID). After that, product ions are ejected for detection. Alternatively, product ions can be kept inside the trap and another CID reaction can be initiated; this repeat CID reaction can continue several times (denoted as MS^n in which n is the number of CID reactions). This can help differentiate molecules with similar structures.

1.5.2.2(b) MALDI

The principal of MALDI was introduced by Hillenkamp and Karas [96]. It is also a 'soft' ionisation method and deals well with thermolabile, non-volatile organic compounds, especially those of high molecular mass. This technique is characterised by easy sample preparation and large tolerance to contamination by salts, buffers, and detergents.

MALDI is achieved in two steps. In the first step, the sample is pre-mixed with a highly absorbing matrix compound (solution of small organic molecules). The matrix is used to transform the laser energy into excitation energy for the sample, which leads to the sputtering of analytes and matrix ions from the surface of the mixture. In this way, energy transfer is efficient and analyte molecules are spared excessive direct energy that may otherwise cause decomposition (Figure 1.15).

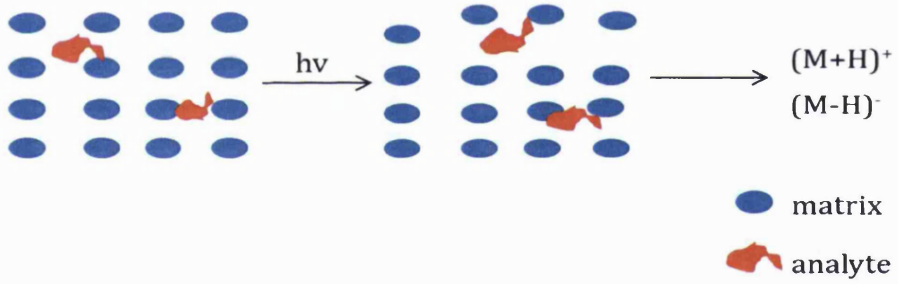


Figure 1.15 Principles of MALDI-TOF; first step. Sample is mixed with the highly absorbing matrix compound to increase the excitation energy of the sample [adapted from: Hoffman; ref. 95].

The second step occurs inside the source of mass spectrometry, whereby this solid solution is bombarded with intense laser pulses over a short duration to bring about sample ionisation. As the laser is fired, the energy arriving at the sample/matrix surface is optimised and the data are accumulated until (m/z) spectrum of reasonable intensity has been amassed. The TOF analyser separates ions according to theirm/z ratios by measuring the time it takes for ions to travel through a field-free region known as the flight or drift tube (Figure 1.16).

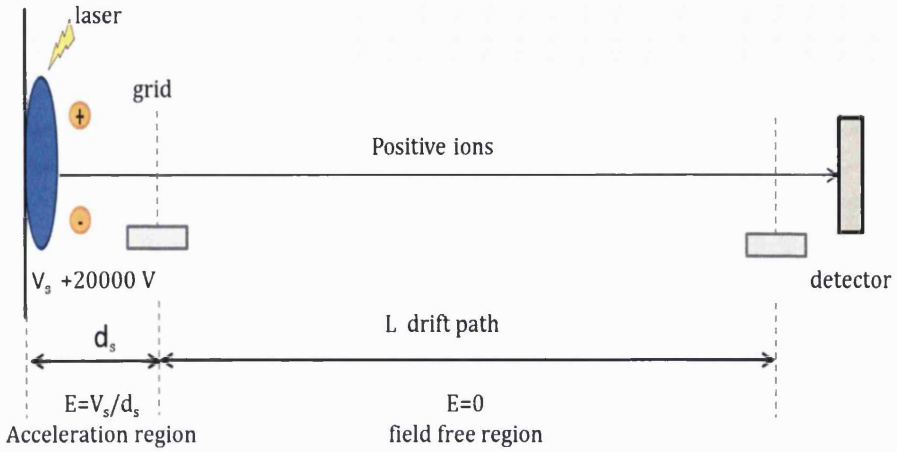


Figure 1.16 Principles of MALDI-TOF instrument; second step. After formation during a laser pulse, ions are subject to an applied electric field. Ions are continuously accelerated and drift in a field-free region where they travel in a velocity based on individual *m/z* ratios [adapted from: Hoffman; ref. 95].

TOF analyser

After ions are expelled from the source, they are accelerated towards the flight tube by the difference of potential applied between the electrode and the extraction grid. Before it leaves the source, an ion with mass m and total charge $q = ze$ is accelerated by a potential V_s. Its electric potential energy E_{el} is converted into kinetic energy E_k:

$$E_k = \frac{mv^2}{2} = qV_s = zeV_s = E_{el}$$

The velocity of the ions leaving the source is given by rearranging equation above:

$$v = (2zeV_s/m)^{1/2}$$

After the initial acceleration, the ion travels at a constant velocity to the detector. The time t needed to cover the distance L before reaching the detector is:

$$t = \frac{L}{v}$$

Replacing v the value in equation above:

$$t^2 = \frac{m}{z}\left(\frac{L^2}{2eV_s}\right)$$

This equation shows that m/z can be derived from t^2 (the terms in parenthesis being constant). It also shows that lighter ions move faster than do heavier ones.

1.5.3 Nuclear Magnetic Resonance (NMR)

1.5.3.1 Principles

NMR is a phenomenon that occurs when the nuclei of certain atoms are immersed in a static magnetic field and exposed to a second oscillating magnetic field. Some nuclei experience this phenomenon and others do not depending on whether they possess a property called spin.

Properties of Spin

Similar to electrical charge or mass, spin is a fundamental property of nature. Spin (I) comes in multiples of 1/2 and can be positive or negative. Protons, electrons, and neutrons possess spin. However, two or more particles with spins that have opposite signs can pair up to eliminate the observable manifestations of spin. Only nuclei with spin number ($I \neq 0$) can absorb/emit electromagnetic radiation.

Almost every element in the periodic table has an isotope with a non-zero nuclear spin. NMR, however, can only be performed on isotopes whose natural abundance is high enough to be detected. Table 1.5 lists some of the nuclei routinely used in NMR.

Another important parameter of nuclei with spin property is the magnetic moment (μ), which is a vector quantity that gives the direction and magnitude (or strength) of the 'nuclear magnet'. It can be expressed as:

$$\mu = \gamma I h/2\pi$$

where h is the Planck constant and γ is the gyromagnetic ratio, which depends on the nature of each nuclei. Different nuclei have different magnetic moments.

Table 1.5 Common nuclei for NMR analysis. Each element has an isotope with a non-zero nuclear spin and is present in high abundance.

Nuclei	Unpaired Protons	Unpaired Neutrons	Net Spin	γ (MHz/T)
^{1}H	1	0	1/2	42.58
^{2}H	1	1	1	6.54
^{31}P	1	0	1/2	17.25
^{23}Na	1	2	3/2	11.27
^{14}N	1	1	1	3.08
^{13}C	0	1	1/2	10.71
^{19}F	1	0	1/2	40.08

Magnetic energy and populations

The energy of a spin in a magnetic field will depend on the magnetic field (B_o) and μ:

$$E = -\mu - B_0$$

When the B_o field is applied, spins have two possible energy limits: in favour of the field (low energy state) and against it (high energy state). The energy is the dot product of the corresponding vectors and it can be represented by an energy level diagram (Figure 1.17):

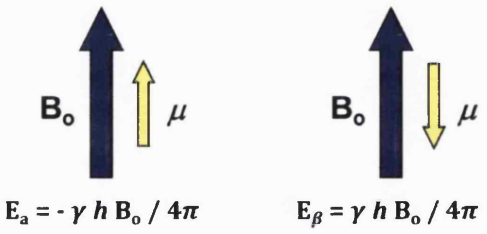


Figure 1.17 The energy level diagram showing two possible energy limits: low and high energy states.

The energy difference of the two levels, α and β and b, is:

$$\Delta E = \gamma h \beta_0 / 2\pi$$

From this equation, it is noted that ΔE corresponds to B_o. $\Delta E$is used to determine the population ratio, calculated using the Boltzmann distribution.

Boltzmann Statistics

At room temperature, the number of spins in the lower energy level (N^+) is slightly more than is the number in the higher energy (N^-). Boltzmann statistics are represented by:

$$N^-/N^+ = e^{-\Delta E/kT}$$

whereΔE is the energy difference between the spin states, k is Boltzmann's constant (1.3805×10^{-23} J/Kelvin), and T is the temperature (Kelvin). As the temperature decreases, the ratio (N^-/N^+) also decreases and as the temperature increases, the ratio approaches one.

The transition of the energy state (higher ↔ lower) based on the differences of the energy absorbed by the spins results in the NMR signal. This signal is thus proportional to the population difference between the states.

Precession

When a net magnetisation is placed in the XY plane, it rotates about the Z axis at a frequency equal to the frequency of the photon that would cause a transition between the two energy levels of the spin. This frequency is called the Larmor frequency, ω_o (in radian):

$$\omega_o = 2\pi v_0 = \gamma B_o$$

Precession is related to the *angular momentum, L,* associated with all nuclei whether magnetic or not (Figure 1.18).

Several magnetic fields act on the spins. One is B_o, which is constant in time and generates the precession at ω_o. The others fluctuate because of the molecular anisotropy and its environment, and this means that the spins sample all the possible orientations with respect to B_o during a certain period.

Orientations in favour of B_o have lower magnetic energy and these are slightly favoured. After a certain time (the *longitudinal relaxation*), a *net magnetisation*(M_o) pointing in the direction of B_o develops.

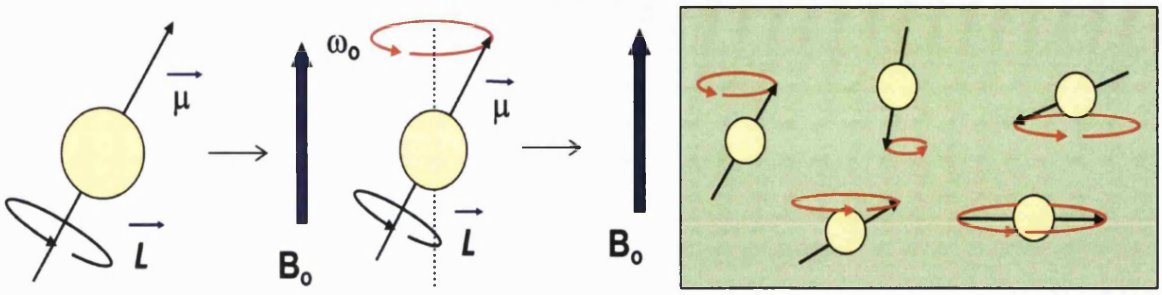


Figure 1.18 *Precession explained.* (a) Nuclei can be described as small magnetised tops that spin on their axes, (b) two forces act on the spins on the presence of magnetic field B_o: one tries to turn then towards B_o and the other wants to maintain its angular momentum, and (c) spins precess at the angle they were at when B_o started [adapted from: Moyna G.; ref 97].

Net magnetisation

Spins experiencing the same magnetic field are represented by a magnetisation vector, where each vector is proportional to $(N^+ - N^-)$. The vector sum of the magnetisation vectors from all the spins produces the *net magnetisation*.

T_1 Processes

At equilibrium and in the rotating frame of reference, the net magnetisation vector lies along the direction of the applied B_o and is called the *equilibrium magnetisation* M_o. In this conformation, the Z component of magnetisation $M_Z$equalsM_o and this is referred to as the *longitudinal magnetisation*. There is no *transverse magnetisation* (M_X or M_Y) at equilibrium.

When the nuclear spin system is exposed to the energy of a frequency equal to the energy difference between the spin states, the net magnetisation changes. It is possible to saturate the spin system with enough energy and make $M_Z=0$.

The time constant, which describes how M_Z returns to its equilibrium value, is called the *spin lattice relaxation time* (T_1). Thus, T_1 is required to change the Z component of magnetisation by a factor of e in a function of time after its displacement, t:

$$M_z = M_o(1 - e^{-t/T1})$$

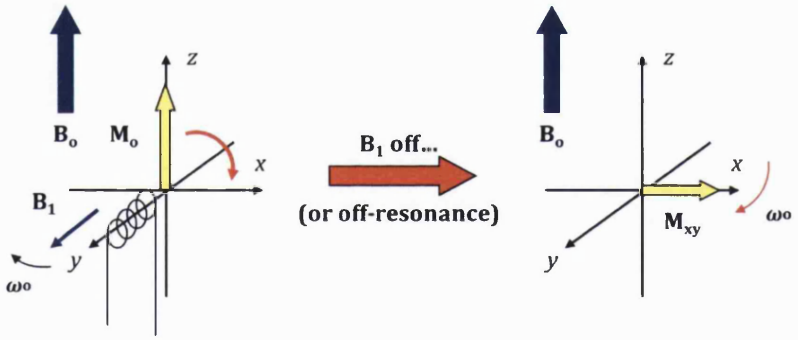
If the net magnetisation is placed along the –Z axis, it will gradually return to its equilibrium position along the +Z axis at a rate governed by T_1. The equation governing this behaviour as a function of time t after its displacement is:

$$M_z = M_o (1 - 2e^{-t/T1})$$

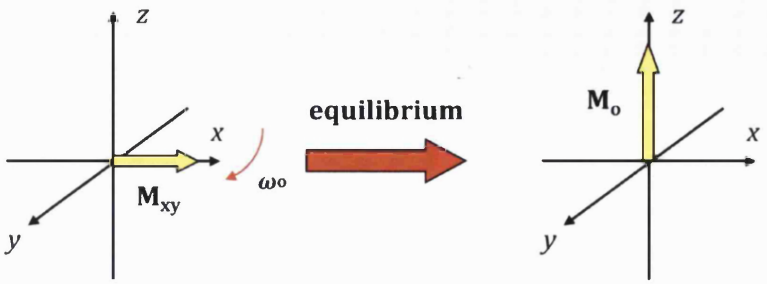
NMR excitation

The net magnetisation discussed so far is in the equilibrium system. To produce NMR signals, the system must move away from the equilibrium state. Its population (N_a/N_b) must be disturbed by absorbing some energy from the oscillating electromagnetic radiation from an alternating current.

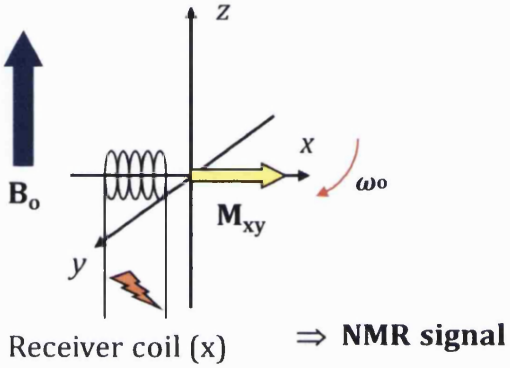
When the frequency of the alternating current is ω_o and the frequency of the right vector of B_1isω_o, a *resonant condition* is achieved. The alternating magnetic field and all the *M*s interact, torque is generated, and they rotate. Since they all rotate around the same amount, the macroscopic effect is that M_o rotates around the Y axis (in this case) and *transverse magnetisation* (M_{xy}) is generated.



Since we have altered the population ratio between energy levels (i.e., N_a / N_b), the system absorbs the energy, causing equilibrium changes. Since the individual spins keep precessing under the effect of B_o, the transverse magnetisation M_{xy} rotates around the **z** axis at the precession frequency, $\mathbf{w_o}$. In the absence of the external B_1, M_{xy} will try to go back to the **z** axis (M_o; equilibrium) by restoring the original N_a / N_b distribution. M_{xy} returns to the **z** axis precessing on the **xy** plane.



The oscillation of M_{xy} generates a fluctuating magnetic field that can be used to generate a current in a coil, which can be translated as the NMR signal:



1.5.3.2 NMR Scale

The NMR scale (δ, ppm) is a relative scale referring all signals in the spectrum to the signal of a particular compound:

$$\delta = \frac{\omega - \omega_{ref}}{\omega_{ref}}$$

The most common internal standard is tetramethyl silane (TMS) because it is inert, volatile, soluble in most organic solvents, and has 12 equivalent ^{1}Hs and four equivalent ^{13}Cs. Other types of references can also be used, such as the residual solvent peak, dioxane for ^{13}C, etc. Figure 1.19 (A) and (B) show the scales of different nuclei for proton and carbon, respectively.

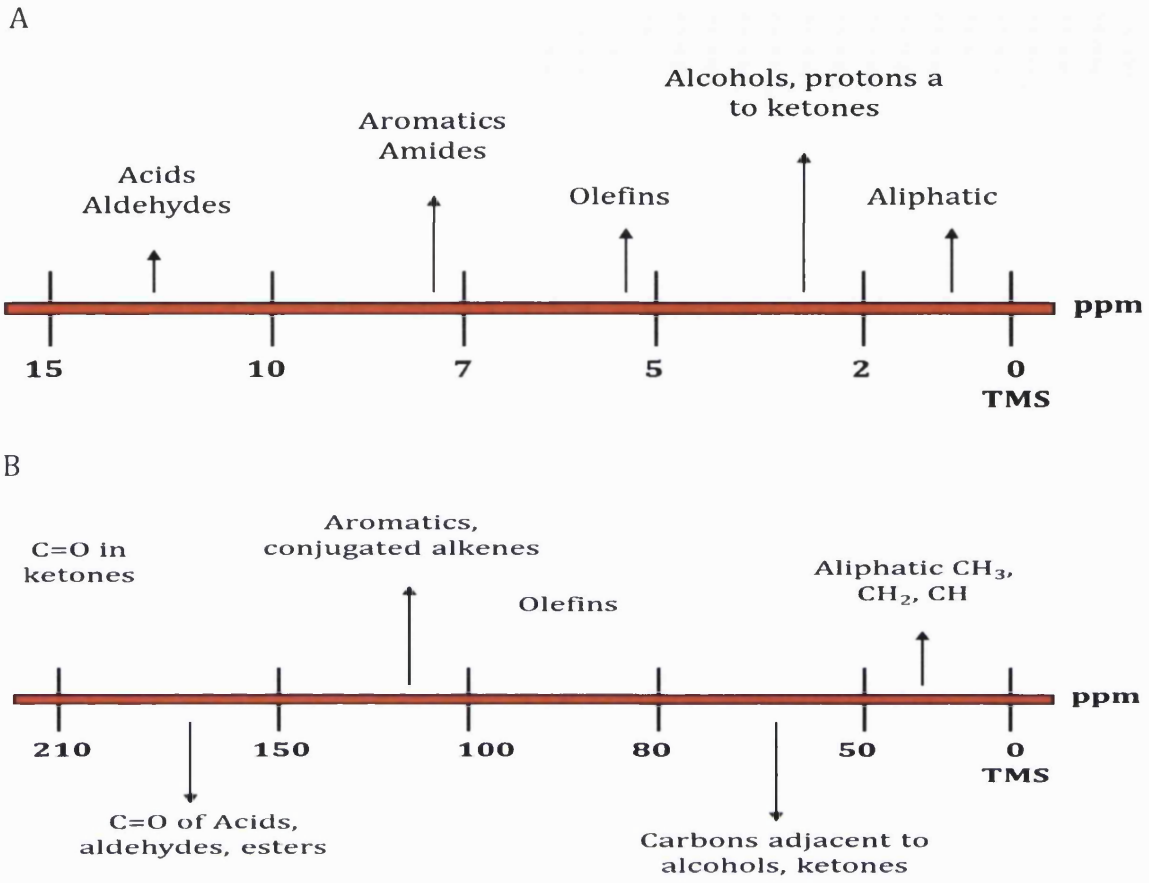


Figure 1.19 The signals in the spectrum corresponding to the signal of a particular compound on the NMR scale for proton (A) and carbon (B).

1.5.3.3 NMR instrumentation and experiments

The NMR instrument (Figure 1.20) in general consists of a magnet, a frequency generator, a detector, and a recorder. The magnet produces an electromagnetic environment for the samples, while the frequency generator creates the alternating current (at $\mathbf{w_o}$) that induces $\mathbf{B_1}$. The detector subtracts the base frequency (a constant frequency close to $\mathbf{w_o}$) from the output frequency, making it much easier to deal with this lower frequency, and then the information is recorded by the detector in a desirable format. Some of the experiments are listed in Table 1.6.

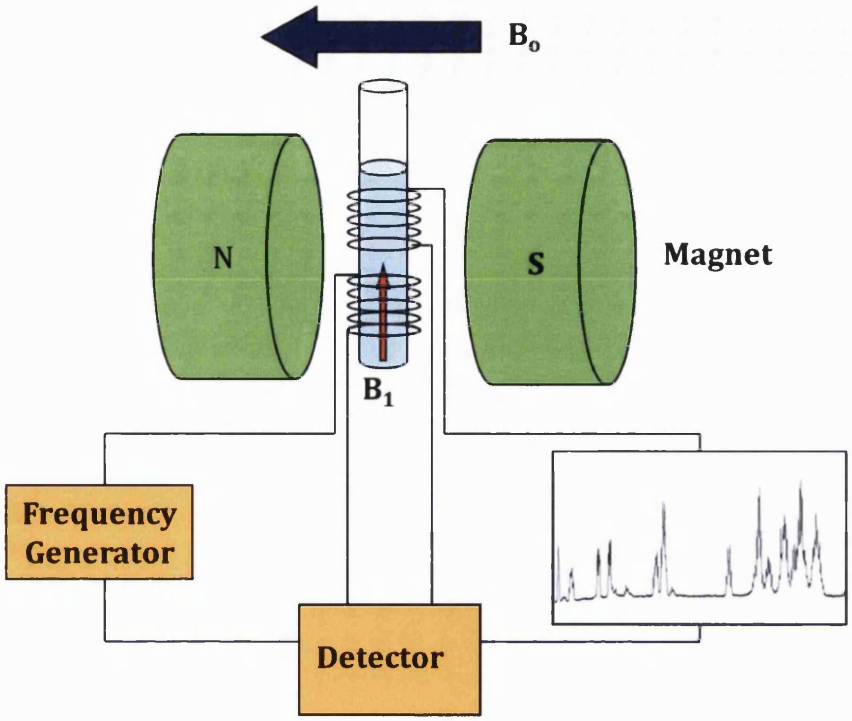


Figure 1.20 The main components of the NMR instrument: a magnet, a frequency generator, a detector, and a recorder [adapted from: Moyna G; ref.97].

Table 1.6 Other NMR experiments; its description and purposes.

Experiments	Description
COSY Correlation Spectroscopy	The original 2D experiment used to identify nuclei that share a scalar (*J*) coupling by correlating any high abundance homonuclear spins. The presence of off-diagonal peaks (cross-peaks) in the spectrum directly correlates the coupled partners
DEPT Distortionless Enhancement by Polarisation Transfer	A 1D experiment used to enhance the sensitivity and for editing of ^{13}C spectra. The experiment is typically run using different final proton pulse angles resulting in differing signs (+ve or -ve) for the various carbon resonances.
DOSY Diffusion Ordered Spectroscopy	A pseudo-2D NMR experiment that presents chemical shifts on one axis versus the self-diffusion coefficients of the solutes on the other. The diffusion coefficients are determined from the NMR signal intensity decays in a sequence of 1D spectra recorded with increasing amplitudes of pulsed field gradients, which are used to map the translational behaviour of the solutes.
HMBC Heteronuclear Multiple-Bond Correlation	A 2D experiment (closely related to HMQC) used to identify 'long-range' couplings (2- or 3- bonds) between protons and carbons.
HMQC Heteronuclear Multiple-Quantum Correlation	A 2D experiment used to correlate directly bonded carbon-proton nuclei. The correlations can be used to map known proton assignments onto their directly attached carbons.
TOCSY Total Correlation Spectroscopy	A 2D homonuclear correlation experiment used to analyse scalar (*J*) coupling networks between protons. It has a similar appearance to the 2D COSY spectrum but not limited to only directly coupled spins.

1.6 Chemometric analysis

The NIR instrument produces a vast amount of overlapping data in a single analysis. In order for this technique to become reliable, data should be able to be transformed into the desired information without compromising the time factor. Chemometrics is a discipline using mathematical and statistical methods for the selection of the optimal experimental procedure and data treatment of chemical analysis. It has become an important tool in NIR analysis where data can be processed in a fast and efficient manner to produce useful analytical information. Relevant information can be extracted using various multivariate analysis techniques, which normally allow samples with similar characteristics to be grouped together to establish classification methods for unknown samples, qualitative analysis, or quantitative analysis [98]. The use of NIR spectroscopy and chemometrics in pharmaceutical technologies was reviewed by Roggo *et al.* [99].

1.6.1 Spectra pre-processing

Spectra pre-processing is conducted to transform the spectra by removing unwanted variability that is not associated with the chemical information of the products. With these enhanced spectral features, it is easier to distinguish the similarity and differences between spectra. Many different types of pre-processing methods are available but only those used in this research are discussed here.

1.6.1.1 Standard normal variate (SNV)

The main objective of SNV pre-processing is to remove the multiplicative interference from light scatter, particle size effect, and path length variation [100, 101]. Figure 1.21 shows the representative spectra of paracetamol from three different manufacturers. The raw spectra (Figure 1.19A) show the effect of light scattering, which causes varying heights of the peaks along the three spectra despite having the same concentration. Similar sets of spectra after subjected to SNV pre-processing have undergone baseline shift and multiplicative scaling, resulting in the spectra coinciding with each other. Two regions (1400–1500 nm

and 1900–2500 nm), where the spectra differ greatly, correspond to the water peaks.

SNV transforms each spectrum by subtracting the mean and dividing by the standard deviation of the measured value. This is represented by:

$$SNV_i = \frac{y_i - \bar{y}}{\sqrt{\sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{n-1}}}$$

where $\bar{y}$ is the mean absorbance over the entire spectral range scanned, y_i is the absorbance at the i^{th} wavelength, and n is the number of wavelengths within the spectral range to which SNV is applied.

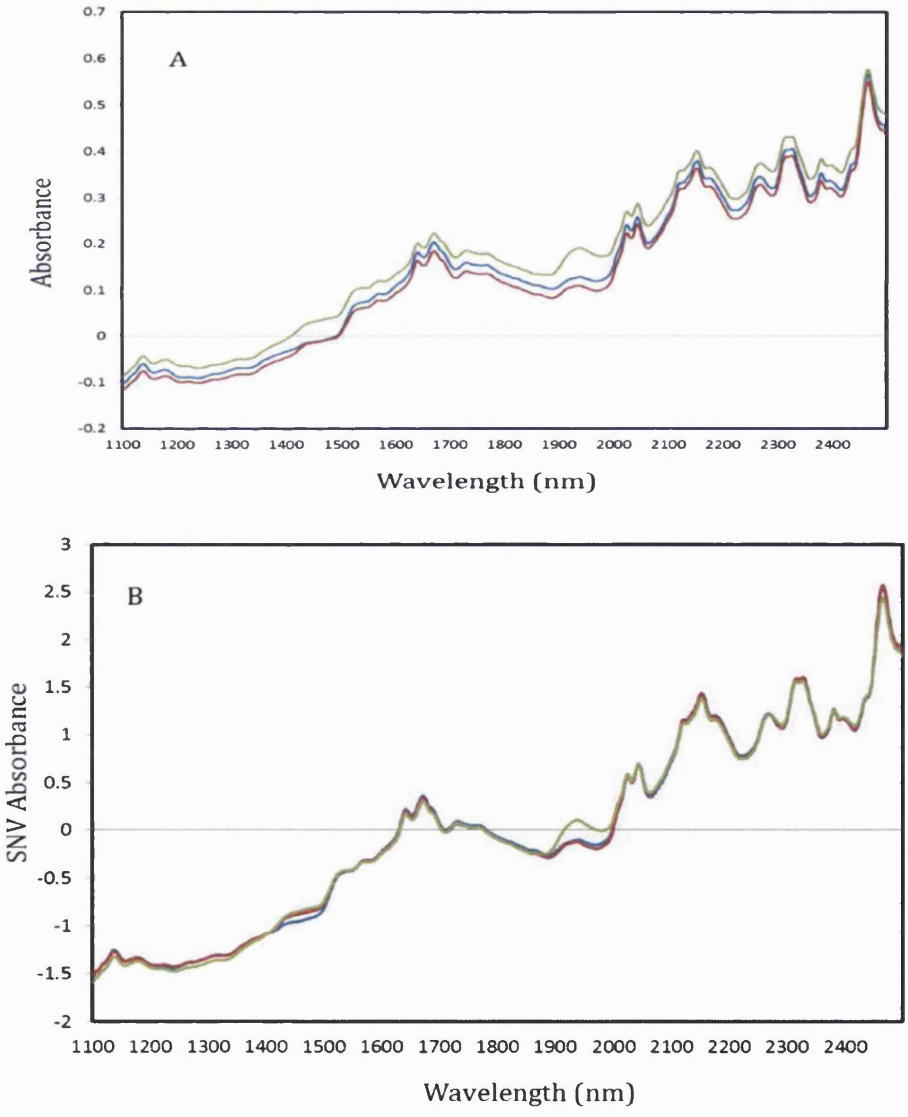


Figure 1.21 Representative spectra of the three brands of paracetamol in (A) raw spectra and (B) after SNV pre-processing.

1.6.1.2 Derivatives

The purpose of using derivative pre-processing is to increase spectra resolution and suppress the constant background effects. The first derivative (Figure 1.22A) is given by the slope at every point in the spectrum and this removes the constant offset along the spectrum. However, first derivatives produce peaks where the original spectrum had maximum slopes and cross zero where the original had a peak, and are thus rather difficult to interpret [102]. For this reason, second derivatives are often preferred.

First derivatives at wavelength λ_n may be calculated as:

$$\lambda_n{}^{1stDer.} = X_n - X_{n-1}$$

where X_n is the measured spectrum at λ_n.

Second derivatives are achieved by taking the derivative of the first derivatives. This is represented by:

$$\lambda_n{}^{2ndDer.} = X_{n-1} - 2X_n + X_{n+1}$$

Second derivatives are a measure of the change in the slope of the curve. As the first derivatives, they work by ignoring the baseline slope and offset, but are not affected by any linear 'tilt'. They have negative peaks where the original had a peak and are thus more readily comprehensible. Second derivatives can help resolve nearby peaks and sharpen spectral features (Figure 1.22B).

Taking derivatives decreases the scale and increases the noise [102]. This drawback can be overcome by averaging segments and leaving gaps between the points used for computation. Two common methods to calculate the derivative spectra are the gap-segment method and Savitzky-Golay [103].

The gap-segment method requires two parameters: gap size and segment size. An overly large gap or segment can result in the loss of important features, while significant noise will be introduced in a too small gap/segment.

The Savitzky-Golay algorithm is based on performing a least squares linear regression fit of a polynomial around each point in the spectrum to smooth the data. The derivative is then the derivative of the fitted polynomial at each point. The algorithm includes a smoothing factor that determines how many adjacent

variables will be used to estimate the polynomial approximation of the curve segment. This method is applied in this research.

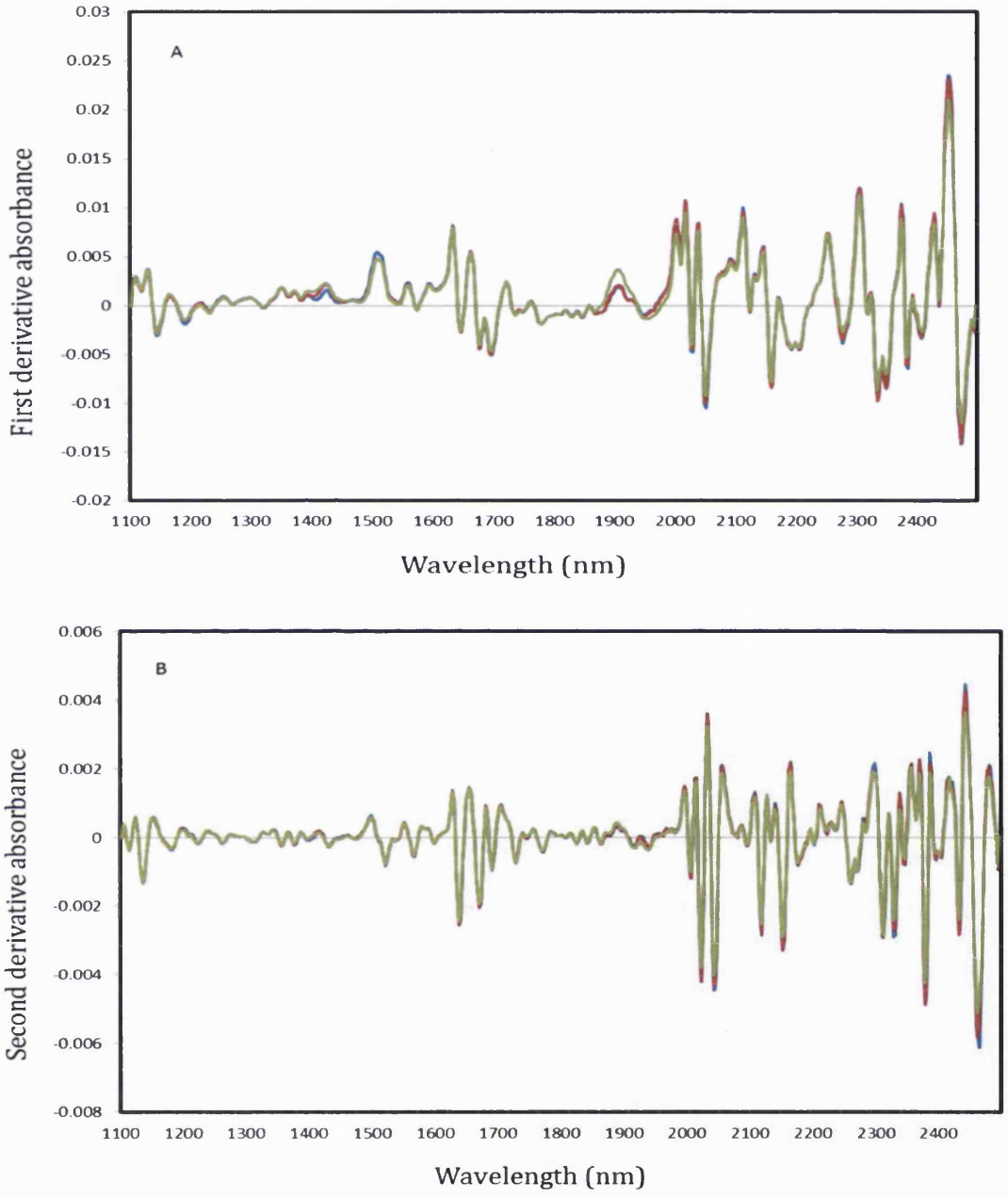


Figure 1.22 Representative spectra of the three brands of paracetamol after Savitzky-Golay pre-processing as (A) first derivative spectra and (B) second derivative spectra in three polynomial order using five smoothing points.

1.6.1.3 Baseline Correction, GIFTS (Auto-Levelling Method)

This method removes the baseline slope and offset using an iterative fitting process that discards points above a threshold and fits the remaining points to a straight line [104]. The algorithm used is commonly employed in the industry for pre-correcting data in IR spectral searching. A least squares line is fit through all the points in the trace. The number of points above and below the line is then counted. If there are fewer points above the line than those below, they are considered peaks and discarded. The process is then repeated until the number of points above the line is less than or equal to those below the line. The calculated line is then used as the baseline and subtracted from the trace.

1.6.2 Principal component analysis (PCA)

The NIR spectrum comprises intensity measurements at hundreds of wavelengths. Owing to the correlation among absorbance values within a spectrum, most information contained in spectra is redundant. Thus, a variable reduction method is needed to identify the unique spectral information from these many wavelengths.

PCA is the most widely used variable reduction method to search for a pattern in the data and express them in such a way as to highlight their similarities and differences [105]. These patterns mark the directions of maximum variability in the linear combinations of the original spectra that are orthogonal to each other and use them as new axes called 'principal components' (PCs). The new compressed data, having a reduced number of dimensions but without much loss of information, are then used as new variables to give an overview of the data. This overview assists in the visualisation of the data and it may reveal groups of observations, trends, and outliers.

In theory, PCA decomposes a data matrix $\boldsymbol{X}$ into a 'structure' part and a 'noise' part to reveal hidden phenomena. The data matrix $\boldsymbol{X}$ is made up of $\boldsymbol{N}$ objects (samples) and $\boldsymbol{K}$variables (wavelengths and wavenumbers). Within the matrix, the element $\boldsymbol{X_{ik}}$is the response of the $\boldsymbol{i^{th}}$ sample at the $\boldsymbol{k^{th}}$ wavelength. Using this notation, the rows of the matrix represent the spectra of individual samples and the columns represent the measurement at a single wavelength for all samples. By using PCA, a data table $\boldsymbol{X}$ can be modelled as (Figure 1.23):

$$X = TP' + E$$

where**T** is the score matrix, P' is the transpose of the loading matrix, and E is the matrix containing residuals (part of the data that cannot be described by the PCA model: noise). The PC scores of the first, second, third and so on components (t1, t2, t3...) are the columns of the score matrix **T**. These scores are the coordinates of the observations in the model (hyper-) plane, which are seen as new variables that summarise the old ones.

The loadings define the orientation of the PC plane with respect to the original X variable. The loadings inform how the variables are linearly combined to form the scores. This reveals the magnitude (large or small correlation) and the manner (positive or negative correlation) in which the measured variables contribute to the scores.

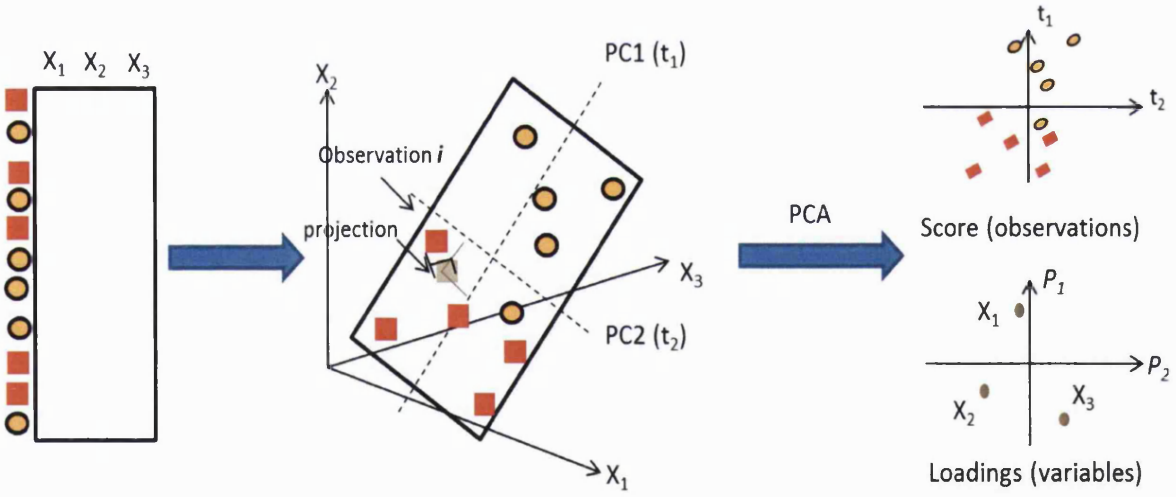


Figure 1.23 A PCA model is developed based on the approximation of the variation in a data table by a low dimensional model plane, which provides a score plot in which the relation among the samples in the model plane is visualised, and a loading plot, which describes the influence of the variables in the model plane and the relation between them [adapted from: Trygg*et al.*; ref. 106].

1.6.3 Soft independent modelling of class analogy (SIMCA)

SIMCA [107] is a supervised classification method based on PCA. The concept of SIMCA is based on constructing a separate PCA model for each known class of observations. These PCA models are then used to assign the class belonging to the

observations of unknown class origin by the prediction of these observations into each PCA class model where the boundaries have been defined by a 95% confidence interval. Observations that are poorly predicted by the PCA class model, hence, have large residuals, are classified being outside the PCA model, and do not belong to the class.

This technique is applied when there is a well-defined class of objects and all other objects are heterogeneous or two or more homogeneous classes of observations that are separately modelled by PCA. An unknown sample can be predicted as belonging to any or none of the class models by measuring the Euclidian distance of the sample to the model (e_i) and the Mahalanobis distance within the PC space (h_i). The calculation is shown diagrammatically, for two groups, in Figure 1.24.

Different plots can be used for assessing the SIMCA results. These plots differ in the ways the measurement of the sample distance is compared with the models. Two of the most common plots are:

a. Cooman's plot — Compares the orthogonal distances from the new objects to two different classes (models) at the same time in pair wise plots.

b. Membership plot — Compares both the sample distance to model (residual standard deviation) and the sample leverage (distance to model centre)

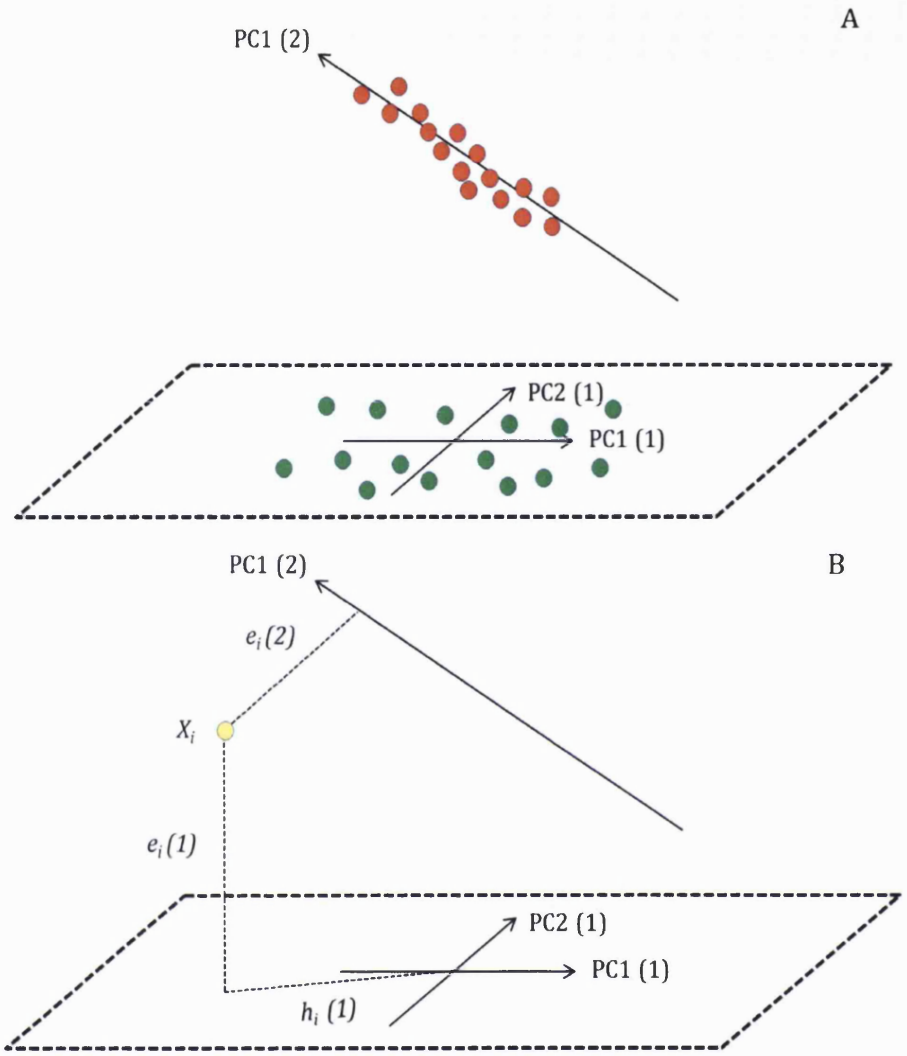


Figure 1.24 SIMCA for two groups. (A) Group 1 is modelled by two PCs, PC1(1) and PC2(1), while group 2 is modelled by one PC, PC1(2); (B) a new sample (X_i) is compared to both groups by projecting it onto the models – a plane for group 1 and a line for group 2. This gives the distancese_i (1) and h_i (1) for group 1 and e_i (2) and h_i (2) for group 2 [adapted from: Davies; ref. 108].

1.6.4 Hoteling's T2 ellipses

Hotelling's T2 ellipse defines the 95% confidence interval of the modelled variation [109]. It is a linear function of the leverage that can be compared with a critical limit using an F-test. This statistic can be included in the PCA score plot

and thus it is useful for the detection of outliers at the modelling or prediction stages.

The test was implemented for each dataset, according to the equation:

$$P[y = (x - \mu_x)^T S_{xx}^{-1}(x - \mu_x) < y_p] = 1 - \alpha$$

where$\mu_x$is the mean of the data and S_{xx}^{-1} is the inverse of the data covariance of matrix S_{xx}:

$$S_{xx} = (t - 1)^{-1} X_c {X_c}^T.$$

Where X_C is the centred data matrix T. The required Mahalanobis distance was calculated according to the equation:

$$q\,(X - \mu_x)^T S_{xx}^{-1}(x - \mu_x) < x^2_{0.99.t}$$

Where$q$is the number of dissolution curves in the reference and test lots and was compared with the corresponding chi square value at a 99% confidence level and t (number of data points per curve) degrees of freedom.

The 95% a confidence region (a= 0.05) was used to test the hypothesis of similarity between the sets of data based on the variability of the weighted scores of the reference set [200]. This confidence ellipse was obtained from Hotelling's test (equation above) and plotted according to:

$$P\left[\left(\frac{w_1^2}{d_{1X^2_{1-\alpha,r}}}\right) + \left(\frac{w_2^2}{d_{2X^2_{1-\alpha,r}}}\right) < 1\right] = 1 - \alpha$$

whered_1 and d_2 are eigenvalues of $\mathbf{S}_{XX}$, while w_1 and w_2 are elements of the matrix w= **B** (x − μ_x). Furthermore, the rows of **B** are eigenvectors of $\mathbf{S}_{XX}$. w_1 and w_2 provide information related to the orientation of the ellipse, which defines the axes' lengths as $(d_{1X^2_{1-\alpha,r}})^{0.5}$ and $(d_{2X^2_{1-\alpha,r}})^{0.5}$, respectively. The degrees of freedom (r) of the X^2 equal the number of selected PCs.

1.7 Hypothesis and Aims

Most of the current strategies for identification of the medicine by NIRS rely either on the direct comparison of spectra of the unknown sample to spectra of reference samples or on the prediction of the spectra based on the known composition of sample. However, lack of willingness of manufacturers to provide reference samples, samples from unknown origin, and medicine dispensed without proper packaging or labelling are some of the barrier in the development of reference spectra library. This prevents the analysis of tablets by established drug analysis and testing procedure by NIRS.

The present research attempts to overcome the limitations of previous methods by developing an effective screening strategy using a two tier approach, i.e. incremental NIR spectra database followed by PCA. It is a cost-effective strategy that can be used by casually trained personnel to produce a quick analysis on a random selection of samples based on their classification with similar brands, sources of products, and/or types of composition without relying on standard products from the manufacturer or reference compounds.

The applicability of the two-tier method was demonstrated for product identification, drug quality study, analysis of herbal preparation, and the detection of counterfeit and adulterated medicines.

Here, it is hypothesised that the use of an incremental spectral database of pharmaceutical and herbal formulations obtained on the Malaysian market will obviate the reliance on the repository of reference samples when analysing unknown samples. The concept of a universal model translated into a database system with built-in search algorithms should allow this two tier strategy for identification of a sample with minimum or no background information. The samples range from well-defined pharmaceutical formulations to highly variable herbal mixtures.

For this purpose, a series of steps were taken with different objectives:

1. Establish a sample preparation protocol and spectra acquisition conditions for NIR, NMR and MS methods to ensure the quality of spectra included in the database and reproducibility of the spectral search.

2. Assess the suitability of commercial databases for the purposes of storing and analysing of spectra acquired using different methods.
3. Evaluation of the search algorithms applied during full spectrum and peak searches of unknown spectra to show which provides most reliable results.
4. Validate the results of database searches using PCA and extensive analysis using NMR and MS.
5. Test the application of the two tier system on several unknown samples.
6. Examine the use of the developed database system for drug quality monitoring.

This work should provide an alternative post-marketing surveillance procedure to monitor products on the market by purchasing samples in pharmacies and/or using a handheld device to scan and immediately comparing the spectra with the database. Random analysis of these products will help establish the general quality of products and scan for the presence of adulterated and counterfeit drugs in the local market.

In this work, similar principals are applied in the database screening strategy for four different applications; product identification, drug quality study, herbal analysis, and the detection of counterfeit and adulterated medicines.

The same strategies will also be used to build up the spectral database of mass spectrometry and NMR data. These databases will be the basis for further analytical research and expansion at the National University of Malaysia.

CHAPTER 2

MATERIALS AND EXPERIMENTAL METHODS

2.1 Materials

2.1.1 Materials for drug database and analysis

The samples acquired for analysis in this thesis can be divided into three categories: conventional medicines, herbal medicines, and other external samples, which are given in Tables 2.1, 2.2, and 2.3, respectively. Details included the name of the active pharmaceutical ingredients (APIs), the brand name, manufacturer, and batch number if appropriate.

Table 2.1 Active ingredients, brand names, dosage, supplier and batch number (where available) for the common drugs used for spectra acquisition, chemometrics analysis and creation the spectral database. (MY= Malaysia, SG= Singapore, UK= United Kingdom).

APIs	Brand name	Dose (mg)	Type of preparation	Batch no.	Manufacturer (Country)
Amlodipine Besilate	Vamlo	5	tablet	2008638	Ranbaxy (MY)
Amoxicillin	Beamoxy	500	capsule	809092	CCM Pharma. (MY)
Amoxicillin	Amoxicillin Capsule	500	capsule	09C0272	Dynapharm (MY)
Amoxicillin	Beamoxy	500	capsule	*	CCM Pharma. (MY)
Amoxicillin	Amoxicap	500	capsule	AK04089	Hovid (MY)
Amoxicillin	Ospamox	500	tablet	157113	Sandoz (MY)

APIs	Brand name	Dose (mg)	Type of preparation	Batch no.	Manufacturer (Country)
Amoxicillin	Beamoxy	500	capsule	*	CCM Pharma(MY)
Amoxicillin	Amoxicap	500	capsule	AK04089	Hovid (MY)
Amoxicillin	Amoxicap	500	capsule	A01052	Hovid (MY)
Amoxicillin	Amoxicillin	500	capsule	*	Pharmaniaga Man. Bhd. (MY)
Amoxicillin	Amoxicap	500	capsule	AK04085	Hovid (MY)
Amoxicillin	Ospamox	500	tablet	157823	Sandoz (MY)
Amoxicillin	Amoxicap	500	capsule	AK 04085	Hovid (MY)
Amoxicillin	Amoxy	500	capsule	*	*
Amoxicillin	Amoxicillin	500	capsule	*	*
Amoxicillin	*	*	capsule	*	*
Amoxicillin	*	*	capsule	*	YSP Industries (MY)
Amoxicillin	*	*	capsule	*	*
Amoxicillin	Beamoxy	500	capsule	809091	CCM Pharma. (MY)
Amoxicillin	Beamoxy	250	capsule	*	CCM Pharma. (MY)
Amoxicillin	*	250	capsule	*	*
Amoxicillin	*	250	capsule	*	Pharmaniaga Man. Bhd. (MY)
Amoxicillin	*	250	capsule	*	Pharmaniaga Man. Bhd. (MY)
Amoxicillin	*	250	capsule	*	HD (MY)
Amoxicillin	*	250	capsule	*	Pharmaniaga Man. Bhd. (MY)
Amoxicillin	*	250	capsule	*	*
Amoxicillin	*	250	capsule	*	HD (MY)

APIs	Brand name	Dose (mg)	Type of preparation	Batch no.	Manufacturer (Country)
Becampicillin HCl	Becampi-cillin	400	film-coated tab	11001	Pharmaniaga Man. Bhd. (MY)
Carbimazole	Camazol	5	tablet	236811	*
Cefuroxime Axetil	Xylid	250	tablet	7P4186	Pharmaniaga Man. Bhd. (MY)
Cefuroxime Axetil	Xylid	250	tablet	7P4222	Pharmaniaga Man. Bhd. (MY)
Cefuroxime Axetil	Xylid	250	tablet	8A083	Pharmaniaga Man. Bhd. (MY)
Chloram-phenicol	Xepanicol	5ml	eye drop	244831	Xepa-Soul Pattinson (MY)
Chloram-phenicol	Xepanicol	5ml	eye drop	244831	Xepa-Soul Pattinson (MY)
Chloram-phenicol	Xepanicol	5ml	eye drop	244831	Xepa-Soul Pattinson (MY)
Dexa-methasone	Dexalone	0.5	tablet	809119	CCM Pharma. (MY)
Diclofenac sodium	Voren	50	enteric coated tablet	VRT5 TA005	YSP Industries (MY)
Diclofenac sodium	Voren	50	enteric coated tablet	VRT5 TA006	YSP Industries (MY)
Diclofenac sodium	Voren	50	enteric coated tablet	VRT5 TA007	YSP Industries (MY)
Enalapril maleate	Invoril	10	tablet	1179510 6	Ranbaxy (MY)
Escitalopram	Lexapro	10	tablet	2154266	H.Lundbeck (Denmark)
Ibuprofen	Ibuprofen	200	caplet	-	Galpharm (UK)
Hydroxy-propyl methyl cellulose	Eye Mo Moist	9ml	eye drop	8100133	GSK, KL (MY)

APIs	Brand name	Dose (mg)	Type of preparation	Batch no.	Manufacturer (Country)
HPMC	Eye Mo Moist	9ml	eye drop	8080058	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	710069	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	8100004	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	9010026	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	8080058	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	8080059	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	8100041	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	8050081	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	8030118	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	9020132	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	8030006	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	8100041	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	9020132	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	8100004	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	8110130	GSK, KL (MY)

APIs	Brand name	Dose (mg)	Type of preparation	Batch no.	Manufacturer (Country)
HPMC	Eye Mo Moist	9ml	eye drop	8110133	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	9020132	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	9010026	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	9010026	GSK, KL (MY)
HPMC	Eye Mo Moist	9ml	eye drop	6090174	GSK, KL (MY)
Mefenamic acid	Pontalon	500	film-coated tab	PLT TD014	YSP Industries (MY)
Mefenamic acid	Pontalon	500	film-coated tab	PLT TD011	YSP Industries (MY)
Mefenamic acid	Pontalon	500	film-coated tab	PLT TD012	YSP Industries (MY)
Meloxicam	Melonex	7.5	tablet	H8124	Intas Pharma. (India)
Paracetamol	Poro	500	tablets	*	YSP Industries (MY)
Paracetamol	Fepril	500	tablets	*	Idaman Pharma. (MY)
Paracetamol	Uphamol	500	tablets	309108	CCM Pharma.(MY)
Paracetamol	Uphamol	500	tablets	*	CCM Pharma. (MY)
Paracetamol	Uphamol	650	tablets	611210	CCM Pharma. (MY)
Paracetamol	Biogesic	500	tablets	26037314	Unam Corporation (MY)
Paracetamol	Biogesic	500	tablets	2106848	Unam

APIs	Brand name	Dose (mg)	Type of preparation	Batch no.	Manufacturer (Country)
				8	Corporation (MY)
Paracetamol	Biogesic	500	tablets	26062808	Unam Corporation (MY)
Paracetamol	Milidon	500	tablets	BN32790	Malayan Pharmacy (MY)
Paracetamol	Milidon	500	tablets	BN32177	Malayan Pharmacy (MY)
Paracetamol	Oralcet	500	tablets	61207	Scan Lab (MY)
Paracetamol	Paracil	500	tablets	PA07C640	SM (MY)
Paracetamol	Pamol	500	tablets	658-0758	Hovid (MY)
Paracetamol	Pritamol	500	tablets	*	Prime Pharm (MY)
Paracetamol	Ifimol	500	tablets	*	Unique Pharmacy (MY)
Paracetamol	Progesic	*	tablets	*	Xepa-Soul Pattinson (MY)
Paracetamol	Boots PCM	500	tablets	7S	Hamol Limited (UK)
Paracetamol	Boots Value	500	tablets	7S	Galpharm (UK)
Paracetamol	Lidl PCM	500	tablets	AU17037F	Bristol Lab (UK)
Paracetamol	Galpharm PCM	500	tablets	PU16G7G	Galpharm (UK)
Paracetamol	GSL PCM	500	tablets	8572	M&A(UK)
Paracetamol	Bristol Lab PCM	500	tablets	E67252F	Bristol Lab (UK)
Paracetamol	Uphamol A	500	tablets	309108	CCM Pharma. (MY)
Paracetamol	Pamol	500	tablets	658-0758	Hovid (MY)

APIs	Brand name	Dose (mg)	Type of preparation	Batch no.	Manufacturer (Country)
Paracetamol	Pritamol	500	tablets	*	Prime Pharm (MY)
Paracetamol	Ifimol	500	tablets	*	Unique Pharmacy (MY)
Paracetamol	Progesic	*	tablets	*	Xepa-Soul Pattinson (MY)
Paracetamol	Ifimol	500	tablets	*	Unique Pharmacy (MY)
Paracetamol	Pamol	500	tablets	*	Hovid (MY) (MY)
Paracetamol	Uphamol A	500	tablets	*	CCM Pharma. (MY)
Paracetamol	Fepril	500	tablets	*	Idaman Pharma (MY)
Paracetamol	Panamol	500	tablets	*	Beacons Pharmaceuticals Pte Ltd (SG)
Paracetamol	Medon	500	tablets	*	Winwa Medical Sdn Bhd (MY)
Polyvidone	Oculoect Fluid	10ml	eye drop	T090	Excelvision, (France)
Polyvidone	Oculoect Fluid	10ml	eye drop	T090	Excelvision, (France)
Polyvidone	Oculoect Fluid	10ml	eye drop	T090	Excelvision, (France)
Ranitidine	Histac	150	tablet	1922514	Ranbaxy (MY)
Risperidone	Respidon-2	2	film- coated tablet	B9029004	Torrent Pharm. (India)
Sibutramine hydrochloride monohydrate	Reductil	10	capsule	3000380	Abbott Lab

APIs	Brand name	Dose (mg)	Type of preparation	Batch no.	Manufacturer (Country)
Sildenafil Citrate	Viagra	50	film- coated tablet	7148300 26	Pfizer Inc. (Australia)
Simvastatin	Simvastatin	20	film-coated tab	8F236	Pharmaniaga Man. Bhd. (MY)
Simvastatin	Simvastatin	20	film-coated	8J089	Pharmaniaga
Simvastatin	Simvastatin	20	film-coated tab	8L064	Pharmaniaga Man. Bhd. (MY)
Sulpiride	Negatil	200	tablet	TG0820	MPI (MY)
Ticlopidine	Ticlopidine	250	film-coated tab	8P0321	Pharmaniaga Man. Bhd. (MY)
Ticlopidine	Ticlopidine	250	film-coated tab	8P031	Pharmaniaga Man. Bhd. (MY)
Ticlopidine	Ticlopidine	250	film-coated tab	8P034	Pharmaniaga Man. Bhd. (MY)
Antazoline HCl Tetryzoline HCl	Spersallerg	10ml	eye drop	416799	Excelvision (Swiss)
Antazoline HCl Tetryzoline HCl	Spersallerg	10ml	eye drop	416799	Excelvision (Swiss)
Antazoline HCl Tetryzoline HCl	Spersallerg	10ml	eye drop	416799	Excelvision (Swiss)
Loratadine/ pseudoepihedrine sulfate	Carinox	5/ 120	tablet	0701297	CCM Pharma. (MY)
$Mg(OH)_2$/ $Al(OH)_3$ gel/	Macgel	200/ 200/	chewable tablet	MAT SJ027	YSP Ind. (MY)

APIs	Brand name	Dose (mg)	Type of preparation	Batch no.	Manufacturer (Country)
dimethylpoly-siloxane		25			
$Mg(OH)_2$/ $Al(OH)_3$ gel/ dimethylpoly-siloxane	Macgel	200/ 200/ 25	chewable tablet	MAT SJ028	YSP Ind. (MY)
$Mg(OH)_2$/ $Al(OH)_3$ gel/ dimethylpoly-siloxane	Macgel	200/ 200/ 25	chewable tablet	MAT SJ029	YSP Ind. (MY)
Neomycin, Polymyxin B Sulfate	Maxitrol	5ml	eye drop	09H10A	Alcon Lab, PJ (MY)
Neomycin, Polymyxin B Sulfate	Maxitrol	5ml	eye drop	09H10A	Alcon Lab, PJ (MY)
Neomycin, Polymyxin B Sulfate	Maxitrol	5ml	eye drop	09H10A	Alcon Lab, PJ (MY)
Paracetamol, dihydroco-deine tartrate	Paramol	500/ 7.46	tablet	7H68231	SSL Int., Cheshire (UK)
Paracetamol/ Aspirin/ Caffeine	Extra Power Pain Reliever	200/ 300/ 45	tablet	51468A	Whafton Lab, North Devon (UK)

Table 2.2 The sample names, batch number, dosage form, manufacturer and types of formulation (where available) of the herbal medicines used to acquire spectra and enrich the spectral database.

Samples	Batch no.	Dosage form	Manufacturer (Country)	Type of formulation
Thomson Activated Ginkgo 40mg	*	tablet	Herbal Revival Sdn. Bhd. (MY)	Activated extract 40 mg
Thomson Activated Ginkgo 40mg	*	tablet	Herbal Revival Sdn. Bhd. (MY)	Activated extract 40 mg
Thomson Activated Ginkgo 40mg	*	tablet	Herbal Revival Sdn. Bhd. (MY)	Activated extract 40 mg
Thomson Activated Ginkgo 40mg	*	tablet	Herbal Revival Sdn. Bhd. (MY)	Activated extract 40 mg
Thomson Activated Ginkgo 40mg	*	tablet	Herbal Revival Sdn. Bhd. (MY)	Activated extract 40mg
Thomson Activated Ginkgo 40mg	GB77101	tablet	Herbal Revival Sdn. Bhd. (MY)	Activated extract 40 mg
Thomson Activated Ginkgo 40mg	*	tablet	Herbal Revival Sdn. Bhd. (MY)	Activated extract 40 mg
Thomson Activated Ginkgo 40mg	GB62702	tablet	Herbal Revival Sdn. Bhd. (MY)	Activated extract 40 mg
Thomson Activated Ginkgo 40mg	GB85301	tablet	Herbal Revival Sdn. Bhd. (MY)	Activated extract 40 mg

Samples	Batch no.	Dosage form	Manufacturer (Country)	Type of formulation
Thomson Activated Ginkgo 40mg	*	tablet	Herbal Revival Sdn. Bhd. (MY)	Activated extract 40 mg
Thomson Activated Ginkgo 40mg	GB77201	tablet	Herbal Revival Sdn. Bhd. (MY)	Activated extract 40 mg
Thomson Activated Ginkgo 40mg	GB84901	tablet	Herbal Revival Sdn. Bhd. (MY)	Activated extract 40 mg
Thomson Activated Ginkgo 40mg	GB85301	tablet	Herbal Revival Sdn. Bhd. (MY)	Activated extract 40 mg
Thomson Ginkgo extract 80mg	GE80101	capsule	Herbal Revival Sdn. Bhd. (MY)	Standardize extract 80 mg
Thomson Ginkgo extract 80mg	GE80101	capsule	Herbal Revival Sdn. Bhd. (MY)	Standardize extract 80 mg
Gincare (Shine)	TGAT R1021	tablet	YSP Industries (MY)	Standardized extract 40 mg
Gincare (Shine)	TGAT SB006	tablet	YSP Industries (MY)	Standardized extract 40 mg
Gincare (Shine)	TGAT SB004	tablet	YSP Industries (MY)	Standardized extract 40 mg
Hovid Ginkgo	AU06613	tablet	Hovid Pharm. (MY)	Standardized extract 40 mg
Hovid Ginkgo	A309563	tablet	Hovid Pharm. (MY)	Standardized extract 40 mg
Hovid Ginkgo	AH02515	tablet	Hovid Pharm. (MY)	Standardized extract 40 mg

Samples	Batch no.	Dosage form	Manufacturer (Country)	Type of formulation
Gold Life Ginkgo	H090401	tablet	Medipharm (MY)	Standardized extract 120 mg
Greenwave Ginkgo	GC80301	capsule	ProHealth (MY)	Activated extract 60 mg
Ginkolin	270195	tablet	Duopharma (MY)	GB Folium 2000 mg
Herbalton Ginkgo	*	tablet	Winwa Medical (MY)	Standardized extract 40 mg
Tongkat Ali Plus	C080904	capsule	Hurix (MY)	Mix preparations with 50 mg radix extract
Tonex Tongkat Ali	3109003	capsule	Polens, Selangor (MY)	Radix extract 300 mg
Gold Box Tongkat Ali	GBM03090 14	capsule	Orient Lab, Selangor (MY)	Radix extract 300 mg
Puteri Rembulan Tongkat Ali	TGA81001	capsule	Herbal Science (MY)	Radix extract 500 mg
Tongkat Ali Capsule	7064039	capsule	Orang Kampung (MY)	Mix preparations with 354mg radix extract
Tongkat Ali Capsule	7064039	capsule	Orang Kampung (MY)	Mix preparations with 354 mg radix extract
Tongkat Ali Capsule	8125039	capsule	Orang Kampung (MY)	Mix preparations with 354 mg radix extract
Tongkat Ali Capsule	8084882	capsule	Orang Kampung (MY)	Mix preparations with 354 mg radix extract
LKH Tongkat	G0207	capsule	De Choice, Sabah	Radix extract

Samples	Batch no.	Dosage form	Manufacturer (Country)	Type of formulation
Ali			(MY)	350 mg
Nu-Prep100 Tongkat Ali	NE090701	capsule	Phytes Biotek, Sel (MY)	Radix extract 100 mg
Chi Kit Teck Aun	CK060931	tiny spherical pellets	Eagle& Pagoda Brand Teck Aun Medical Factory (MY)	Mix preparations of chinese herbals
Chi Kit Teck Aun	CK081065	tiny spherical pellets	Eagle& Pagoda Brand Teck Aun Medical Factory (MY)	Mix preparations of chinese herbals
Chi Kit Teck Aun	CK081168	tiny spherical pellets	Eagle& Pagoda Brand Teck Aun Medical Factory (MY)	Mix preparations of chinese herbals
Chi Kit Teck Aun	CK081065	tiny spherical pellets	Eagle& Pagoda Brand Teck Aun Medical Factory (MY)	Mix preparations of chinese herbals
Chi Kit Teck Aun	CK060931	tiny spherical pellets	Eagle& Pagoda Brand Teck Aun Medical Factory (MY)	Mix preparations of chinese herbals
Chi Kit Teck Aun	CK090104	tiny spherical pellets	Eagle& Pagoda Brand Teck Aun Medical Factory (MY)	Mix preparations of chinese herbals
Chi Kit Teck Aun	CK060931	tiny spherical pellets	Eagle& Pagoda Brand Teck Aun Medical Factory (MY)	Mix preparations of chinese herbals

Samples	Batch no.	Dosage form	Manufacturer (Country)	Type of formulation
Chi Kit Teck Aun	CK090101	tiny spherical pellets	Eagle& Pagoda Brand Teck Aun Medical Factory (MY)	Mix preparations of chinese herbals
Po Chai	21112	tiny spherical pellets	Li Chung Shing Tong (Hong Kong)	Mix preparations of chinese herbals
Po Chai	21131	tiny spherical pellets	Li Chung Shing Tong (Hong Kong)	Mix preparations of chinese herbals
Po Chai	21135	tiny spherical pellets	Li Chung Shing Tong (Hong Kong)	Mix preparations of chinese herbals
Po Chai	21124	tiny spherical pellets	Li Chung Shing Tong (Hong Kong)	Mix preparations of chinese herbals
Po Chai	21129	tiny spherical pellets	Li Chung Shing Tong (Hong Kong)	Mix preparations of chinese herbals
Po Chai	21129	tiny spherical pellets	Li Chung Shing Tong (Hong Kong)	Mix preparations of chinese herbals

Table 2.3 List of samples used for method validations and detailed analysis with their names, type of preparation, source and the chapters where each sample will be discussed.

Samples	Type of preparation	Source	Application
Kintop	Herbal slimming product	*NPCB* (2007)	Chapter 8
Jamu Ajaib	Herbal preparation for general health and strength	*NPCB* (2007)	Chapter 8
Tunglin	Anti-rheumatic	*NPCB* (2007)	Chapter 8
Pronoton	Herbal preparation for general health and well-being of men	*NPCB* (2010)	Chapter 8
Lami	Herbal slimming product	*NPCB* (2010)	Chapter 8
Eurycoma longifolia Complex	*Eurycoma longifolia*	Purchased via website (UK)	Chapter 7
Malaysian Ginseng Capsule	*Eurycoma longifolia*	Purchased via website (UK)	Chapter 7
Tongkat Ali	*Eurycoma longifolia*	Purchased via website (UK)	Chapter 7
Counterfeit Viagra®	Sildenafil Citrate (counterfeit)	NPCB (2007)	Chapter 8/ Appendix 7

2.1.2 Sampling Methods

Throughout the research, four sampling methods were used to obtain the samples. The selection of sampling method mainly depended on the objective of the research work in which the samples were used. This research was conducted without additional funding. Therefore, the cost of the samples was one of the considerations in the sampling procedure.

2.1.2.1 Convenience Sampling

Almost 80% the samples used to enrich the database were obtained through convenience sampling, where no specific sampling frame or strategy was required. This sampling technique was chosen because the researcher had limited time and resources for sample collection, because the procedure was inexpensive, and because it did not require a complete list of outlets in the selected area [80]. However, this method is not adequate to provide general ideas on the quality of medicines or the prevalence of counterfeit, substandard, or adulterated medicines in the Malaysian market.

Although sampling was not designed within a specific sampling frame, the sampling procedures followed the guidelines for drug sampling by the US Pharmacopeia Drug Quality Monitoring Program [111]. The selection of medicines included in this research was based on the list of counterfeit drugs detected in Asian countries [112] and limited information was available from Malaysian sources [15]. All samples were kept either in the manufacturer's original packaging or loose (as dispensed), and stored in the dry, dark environment and at an ambient temperature until tested.

Phase 1 of the sample collection was conducted in November 2007 prior to the beginning of the research work. Samples were purchased from pharmacies around Kuala Lumpur by the researcher, who posed as a customer. Different brands of products with the same APIs were selected from 10 different classes of medicines in 15 different pharmacies within this area. A minimum of 20 and five units of samples were purchased for each brand of tablet and liquid preparation, respectively.

Phase 2 of the sample collection was conducted in June 2009 and expanded the defined area of sampling to the three main cities of Malaysia (Figure 2.1): Kuala Lumpur (central), Penang (northwest), and Johor Bahru (southeast). The selection of these states was based on the higher prevalence of counterfeit medicines found in another survey [15].

Sample selection was divided into three categories: (i) products of many variations that had higher probability of being counterfeited antibiotics, (ii) two sets of commercial herbal products, namely *Ginkgo biloba* and *Eurycoma longifolia,* and (iii) products of different dosages, namely steroid creams and

inhalers. Altogether, 20 units of solid and 10 units for non-solid samples were purchased from 20 pharmacies in each defined area. There was a concern about purchasing antibiotics over the counter because, according to local law, a prescription is required for this class of medicines. However, out of 60 pharmacies sampled, only two outlets refused to prescribe without a prescription.

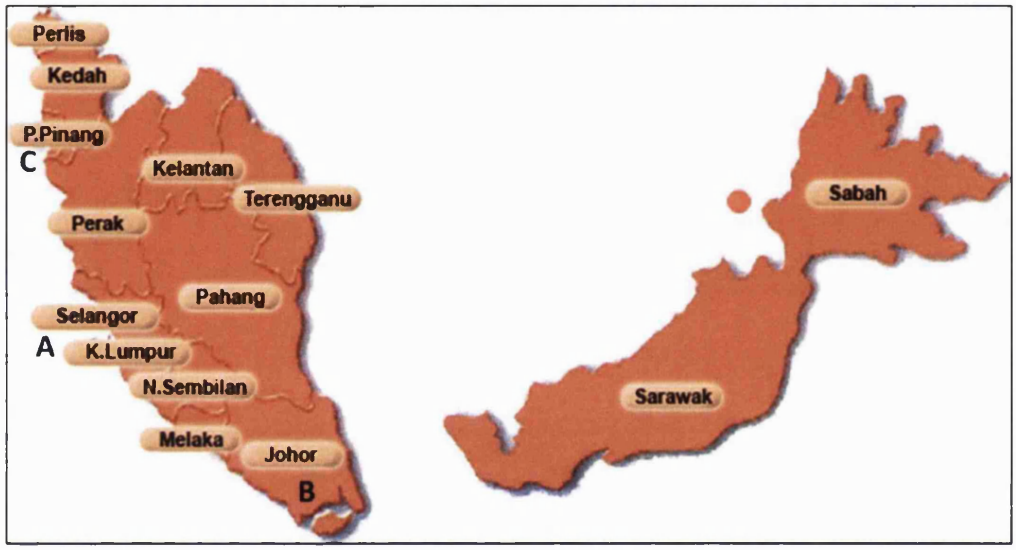


Figure 2.1 The sampling locations: (A) Kuala Lumpur (B) Johor Bahru, and (C) Penang.

2.1.2.2 Direct contact with local manufacturer

The first attempt to establish a relationship between the researcher and pharmaceutical companies in Malaysia was in February 2008, where a formal invitation to participate in the research was sent via email to 10 official contacts from the most reputable pharmaceutical companies in Malaysia. This email included a brief description about the on-going research and a request for participation by contributing samples to enrich the database. However, only three out of 10 responded and all declined to participate. A representative from one manufacturer said that it was regrettably unable to cooperate because of the UK immigration policy on drug importation.

The second request was made in June 2009 while the researcher was in Malaysia. A formal invitation was sent out to the 10 pharmaceutical companies via post followed by a telephone call. Four companies responded; two declined while the other two agreed to have a short discussion with the researcher and finally

supplied the minimum samples requested (Appendix 1). The other six companies did not respond.

2.1.2.3 Request to the National Pharmaceutical Control Bureau

Before commencing the research in 2007, a request was extended to the National Pharmaceutical Control Bureau (NPCB) for counterfeit and adulterated samples seized by the organisation, which will be included in this study. The agency responded to our request by supplying a small number of these samples (listed in Table 2.3). The first set of samples was handed over in November 2007 and the second set in July 2010 through informal channels and for research purposes (application in chapter 8).

2.1.2.4 Purchasing via the Internet

Three sets of herbal samples were purchased via the Internet from a reputable online health shop in the UK (sample listed in Table 2.3, application in chapter 7).

2.2 Sample Preparations

Throughout this PhD study, four types of instrumental analysis were used to analyse three different types of sample preparations: tablets, capsules, and liquids. For each technique and sample type, a standardised sample preparation method was used.

2.2.1 NIR Analysis

2.2.1.1 Intact tablet preparation

NIR analyses of intact tablets were conducted based on the proposed protocol for model construction aimed at the inspection of drugs and identification of counterfeit samples by Scafi and Pasquini [113]. All tablet and capsule samples were weighed prior to analysis.

For intact tablet preparations, samples were placed centrally on the sample stage over the light beam in the NIR instrument. Each side of the tablet was analysed four times while rotating them at 90° each time. Eight accumulations of 32 scans of reflectance spectra were recorded for each sample over the wavelength range 1100–2500 nm.

2.2.1.2 Capsule preparation

Capsules were emptied into 4 ml clear glass vials with snap caps 45 × 15mm (height × diameter) (WAT25051, Waters Ltd., Herts, UK). The glass vial was placed in the middle of the sample stage. Each sample was analysed four times, shaking the vials in between acquisition time to avoid the compaction of the powder materials. Eight accumulations of 32 scans of reflectance spectra were recorded for each sample over the wavelength range 1100–2500 nm.

2.2.1.3 Liquid preparation

For liquid samples, 0.5 ml of the sample was measured in a measuring cylinder and transferred directly into a quarts reflectance cell (Hellma, Essex, UK; outside diameter 30 × 50 (height × diameter), inside diameter 28 × 46 (height × diameter), volume 32 mL). A disc was immersed in the cell with care to ensure that no air bubbles were trapped between the cell wall and the disc. Three types of discs were used to provide reflectance in liquid samples: gold (0.5 cm thickness), silver (1.0 cm thickness), and plastic (1.0 cm thickness). The cells and discs were washed with methanol and HPLC-grade water (Fisher Scientific, UK), and dried using a hairdryer between samples. Sample acquisition was conducted similar to the powdered samples.

The experiment was performed under a monitored humidity and ambient temperature. The average spectra for each sample were taken for further chemometric treatments and for inclusion in the spectral database.

2.2.2 ESI Mass Spectrometry Analysis

All samples for electrospray analysis were crushed into a fine powder. Then, 1 mg and 2 mg of the sample were taken for conventional medicines and herbal preparations, respectively. The powdered samples were dissolved in 1 ml of HPLC-grade methanol (Fisher Scientific, Leicestershire, UK) in 1.5 ml eppendorf tubes. The solutions were sonicated for 15 minutes, followed by centrifugation at 13,000 rpm for 6 minutes.

A micro-pipette was used to transfer 1–5 µl of the sample solution into the tip of a metal-coated capillary PicoTip™ emitter with a 0.5 mm outer diameter and a 0.69 mm inner diameter (New Objective, Inc., Woburn, MA, USA). It was ensured that

the loaded sample was free from air bubbles. The tip of the capillary was cracked against a metal stopper on the stage of a light microscope to give a spraying orifice of about 5 μm. Samples were analysed using the static nano-spray procedure.

In the nano-electrospray source assembly, the tip tube was inserted through the centre of the nose cone tip mount assembly by hand initially and then gently inserted further into the assembly using tweezers. The X and Y knobs of the XYZ stage were adjusted to move the tip approximately to the centre of the heated capillary. The Z knob was used to move the tip towards the heated capillary surface until the end was approximately 5 mm away from the heated capillary.

The full spectrum was recorded in the 100–2000 scan range. For MS/MS analysis, the parent mass was fragmented using the normalised collision energy in the range of between 15% and 35%. The number of microscans was set at three. The maximum injection time was set at 200 ms.

2.2.3 MALDI Analysis

Sample analysis using MALDI was conducted only on the eye drop liquid preparation. The flowchart in Figure 2.2 shows the sample preparation steps involved in the MALDI analysis. Different types of matrix (tetrahydrofurane, dioxane, 2-4 dihydroxyphenylazo benzoic acid, and dihydroxy benzoic acid) and solvents (water, methanol, and ethanol) were tested to determine the suitability of the sample. A hydrophobic plate was used for analysis.

2.2.4 NMR Analysis

Altogether, 100 mg of the powdered samples were dissolved in 1 ml of deuterated methanol (GOSS Scientific Ins., Cheshire, UK) with 0.05% of tetramethylsilane (TMS) (GOSS Scientific Ins., Cheshire, UK) added as the internal reference. In some experiments where the sample was water-soluble, deuterium oxide (GOSS Scientific Ins., Cheshire, UK) was used as the solvent with added trimethylsilyl propionate(TSP) as the internal standard. Samples were sonicated for 15 minutes and ultra-centrifuged for another 5 min at 13,000 rpm. In total, 650 μl of the supernatant was transferred into the NMR tube. Thus, 256 scans

were acquired for each sample. TOPSPIN version 1.3 software was used for spectra processing. All spectra were phase-corrected, baseline-corrected, and calibrated using the internal reference TMS. NMR analyses included 1D ^{1}H-NMR, 2D ^{1}H-^{1}H COSY, 1D ^{13}CNMR, 2D ^{1}H-^{13}C HMQC, and HMBC experiments.

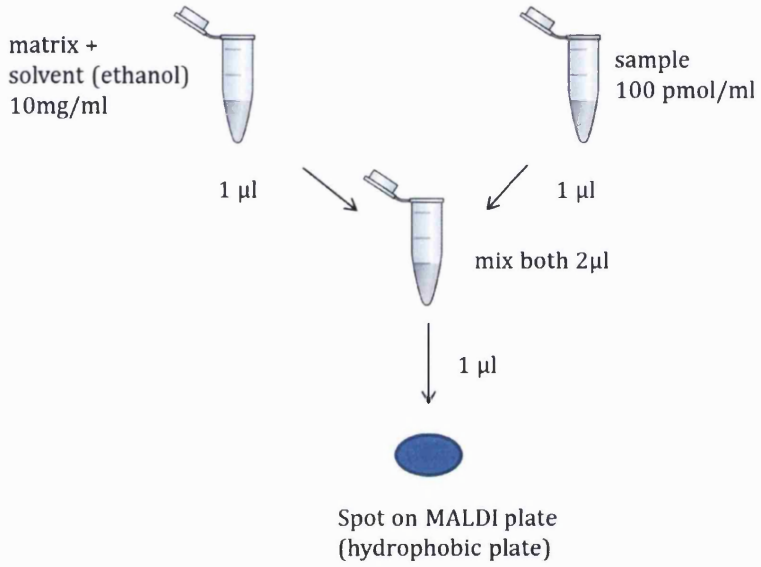


Figure 2.2 Step-by-step procedures in sample preparation for the MALDI analysis.

2.3 Instrumentation and System Check

2.3.1 FOSS 6500 Spectrophotometer

NIR analysis was conducted using a FOSS 6500 spectrophotometer (NIRSystems, Silver Springs, MD, USA) fitted with a Rapid Content Analyser. The Rapid Content Analyser detector unit (Figure 2.3) was horizontally aligned with the monochromator unit and then it had six lead sulphide detectors arranged in a circular orientation angled at 45° under the sample stage.

The instrument included a sample centring iris that was used to position the sample. Tablets were placed centrally on the sample stage and measured on both sides. The spectra were collected in reflectance mode, and all measurements were taken relative to the instrument's ceramic reference. A new reference spectrum was acquired hourly throughout the analysis. The experiment was conducted under a monitored humidity and ambient temperature.

2.3.1.1 NIR system checks

A performance test analysis was conducted every morning to ensure that the NIRSystems instrument was performing to the manufacturer's specifications. The performance test checked the instrument's parameters such as photometric noise and accuracy, bandwidth analysis, and wavelength accuracy. Wavelength linearization checked the internal reference peak position corresponding to the nominal positions. A set of linearization constants that minimised deviations from these nominals was derived. The instrument was controlled by Vision® Spectral Analysis Software for Windows (FOSS NIRSystems, version 2.11 or 3.1) for data acquisition.

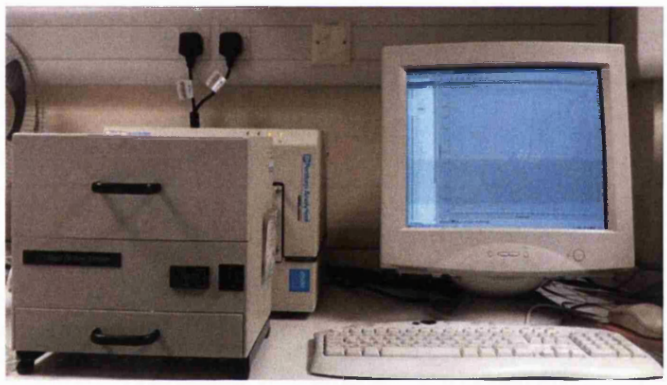


Figure 2.3 FOSS 6500 spectrophotometer with a Rapid Content Analyser detector in horizontal alignment with the monochromator unit.

2.3.2 Thermo LCQ Duo Mass Spectrometer

The Thermo LCQ-MS is a quadruple ion-trap liquid chromatography mass spectrometer equipped with ESI and atmospheric pressure chemical ionisation (APCI) ion sources (Figure 2.4). The basic operation of this instrument is through the generation and migration of ions from the ESI or APCI assembly into the ion-trap chamber.

In this research, samples were introduced into the ion source by direct infusion using a micro-capillary emitter without any chromatographic separation of components in the samples and then were passed into the MS detector. The ions

trapped in the ion-trap chamber were then sequentially ejected and detected by the MS detector. Positive and negative ion ESI mass spectra were acquired over the *m/z* range 50–2000 at a 200-unit resolution and with a peak width of 0.25 daltons/z with the detection sensitivity down to 10 femtomoles.

In a single MS analysis, the full-scan MS mode provides a full mass spectrum of analytes introduced into the ion source. Molecular weight information on pure compounds and compounds in the mixture were determined. This scan mode provided the maximum amount of information available from a single stage mass spectrometry experiment.

MS/MS Analysis

In the full-scan MS/MS mode, information was provided about the molecular structure of a particular ion. A specific ion can be selected, stored in the trap, and collided with helium gas in an attempt to cause fragmentation reactions. The spectrum of the fragment ions resulting from such collisions can then be obtained. In addition, one of these fragment ions can be selected, stored in the trap, and induced to undergo further collision-induced fragmentation [114]. This process can be repeated up to 10 times. This technique is important in structure determinations and confirmation because ESI and APCI mass spectra are characterised by the presence of molecular and/or adduct ions (e.g. the molecular ion coordinated with one or more ubiquitous Na ions).

2.3.2.1 Mass Spec system checks

The instrument was calibrated daily using Glu-Fib ((Glu^1) – Fibrinopeptide B human, Sigma) before any analysis was conducted. It was a 14 amino acid peptide with a mass of 1,570.57 Da. A full MS spectrum was acquired to assess the intensity of the peak for Glu-Fib (786.1 *m/z*, $[M+2H]^{2+}$). The minimum accepted intensity for this peak was around 400 cps. A MS/MS mass spectrum was obtained by fragmenting this peak with normalised collision energy of 35%. The fragmentation pattern was compared with the theoretical value. For both analyses, data were acquired for 3 minutes. This was important to ensure the mass accuracy of the instrument and the consistency of the data. LCQ tune, Xcalibur® software, version 1.2 (Thermo Scientific, Leicestershire, UK) permitted

the in-line acquisition and integration of HPLC dual-wavelength absorbance data along with mass spectral data.

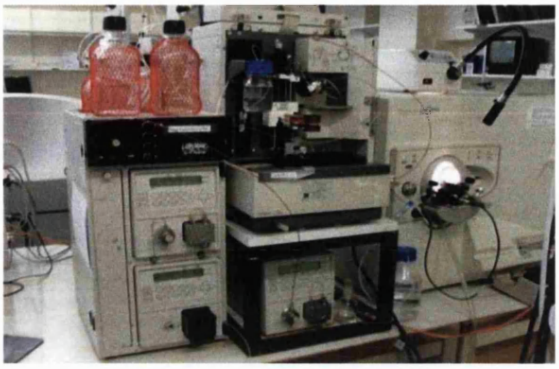


Figure 2.4 Thermo LCQ Mass Spectrometer.

2.3.3 Bruker Avance 500 MHz NMR spectrometer

The AV500 is equipped with broadband and triple resonance (^{1}H, ^{13}C, and ^{15}N) inverse probes (Figure 2.5). It is a two-channel NMR instrument, and TOPSPIN, version 1.3 (Bruker Biospin, Coventry, UK) software was used for the acquisition and processing of NMR spectra. ICON-NMR was used for the automation and control of the 'open access' setup. This incorporates an 11.74T ultrashield magnet and is equipped with both a triple axis-gradient and a 60-position BACS sample changer.

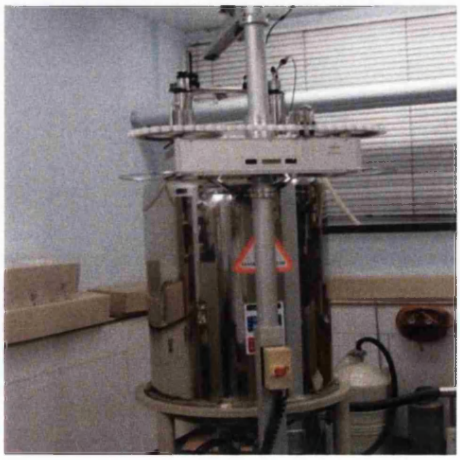


Figure 2.5 Bruker Avance 500 MHz NMR spectrometer.

2.3.4 Applied Biosystem Voyager-DE Pro Biospectrometry Workstation MALDI-TOF Mass Spectrometer

The MALDI-TOF system in Figure 2.6 (Applied Biosystem, Cheshire, UK) provides accurate mass measurement of peptides and polymer at a high sensitivity, up to less than 5 femtomoles. Positive and negative ion detection is possible over a *m/z* range in excess of 300 Da. The instrument is equipped with Data Explorer™ for flexible acquisition and data processing.

Figure 2.6 Applied Biosystem Voyager-DE MALDI- TOF mass spectrometer.

2.4 Software and Data Analysis

2.4.1 Commercial software

The Unscrambler® v9.8 (CAMO Software AS, Norway) was used to perform spectra pre-treatment and PCA and SIMCA analysis.

Spectra ID v9.0 (part of the GRAMS spectroscopy software suite, Thermo Fisher Scientific Inc., Leicestershire, UK) was used to build up the spectral database of medicines.

Amix-Viewer v3.5 (Bruker BioSpin, Coventry, UK) was used to generate a bucket table for NMR spectra prior to the application of the multivariate analysis. The creation of 'buckets' is a way to divide a 1D NMR spectrum to reduce the

dimensions of a dataset. Excel 2010 (Microsoft Corporation) was used for data representation.

2.4.2 In-house software

CUBIC is the format conversion program that was used to convert the spectra in .txt files to a .jdx format for storage or processing in other software. This program was written by Dr. R.D. Jee (The School of Pharmacy, London).

CHAPTER 3

SPECTROSCOPIC TECHNIQUES AND CHEMOMETRICS IN DRUG ANALYSIS

3.1 Introduction

This chapter highlights some experimental observations and considerations throughout the process of developing the database. The initial studies consist of recording the chemical profiles or fingerprints of various pharmaceutical products using three different types of analytical instruments: NIR, NMR, and MS. Focus were given to three different types of dosage forms: solids, powders, and liquids. The data were further used to observe the distribution and classification of samples and also to understand more about the effect of pre-treatment on chemometric analysis.

For each analytical technique, several variables were selected and monitored in order to determine the parameters, consistency, repeatability, and standardisation of any future analysis. Table 3.1 lists some examples of the variables that were important in the respective analytical methods. These are important to establish the validation procedures in future works.

Table 3.1 Some important variables in this analysis.

Technique	**Variables**
NIR	Sample preparations and methods
MS	Type of solvent Type of matrix (MALDI)
NMR	Sample solubility Sample concentration
Chemometrics	Types of pre-treatment analysis
Others	Samples storage and stability

The initial analysis on solids and powders (part I), liquids (part II), and special dosages (Part III) are described in the following sections.

3.2 Materials and Methods

3.2.1 Part I: Analysis of Paracetamol tablets

Eight sets of 500 mg paracetamol samples from five brand names (OR, Mil_A, Mil_B, Up_A, Up_B, Up_C, PC, and PR) were analysed by NIR and their spectra were acquired between the wavelength of 1100 and 2500 nm. First, the spectra of 10 intact tablets for each set were acquired. The tablets were then crushed individually using a pestle and mortar and kept in 4 ml glass vials (Waters Ltd.) for the spectra acquisition of the powdered samples. PCA was conducted on the raw, SNV, and second derivative spectra of these samples to compare product classifications and to observe the effects of crushing the samples (detailed descriptions of the procedure are given in section 2.2.1.1 for solids and 2.2.1.2 for powders (page 94 and 95).

3.2.2 Part II: Analysis of Eye drops

3.2.2.1 Drug samples

In total, 22 bottles of Eye Mo Moist 9 ml (GSK, Kuala Lumpur, Malaysia) were purchased by convenience from pharmacies in three states in Malaysia; Kuala Lumpur, Johor Bahru and Penang (Figure 3.1). This product is an ophthalmic lubricant that is used to moisturise and refresh dry, tired, and dull eyes [115]. Each 1 ml of the solution contains hydroxypropyl methylcellulose (HPMC; 3 mg), benzalkonium chloride (0.1 mg), and edetate disodium (1 mg). All samples were selected from different batch numbers to include production variability in analysis. Five additional brands of eye drops (Table 3.2) were included in the analysis for the comparison of spectra using the transflectance method.

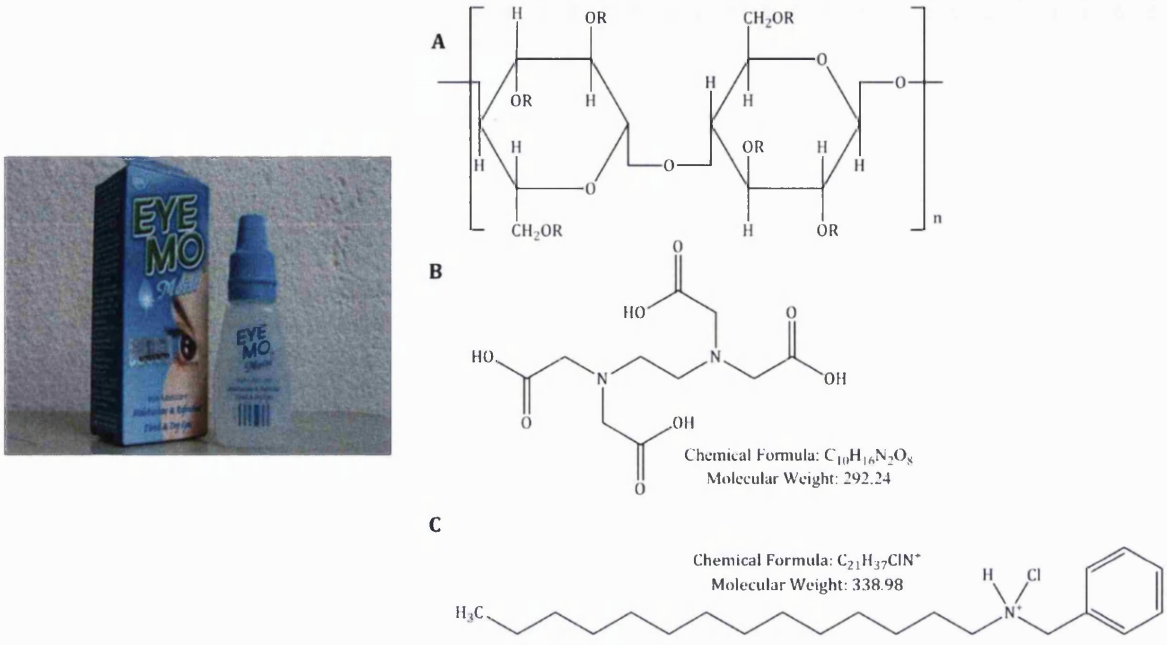


Figure 3.1 *Eye Mo Moist* 9 ml (GSK, Malaysia). Chemical structures of the main ingredients consisting of (A) HPMC (B) edetate disodium ($C_{10}H_{16}N_2O_8$, mwt. 292) and (C) benzalkonium chloride ($C_{21}H_{37}NCl$, mwt. 339).

Table 3.2 Sample list of other types of eye drops included for analysis.

Brand names	Main AI	Vol. (ml)	Manufacturer	Set
Spersallerg	Antazoline HCl, Tetryzoline HCl	10	Excelvision, AG (Swiss)	1
Oculoect Fluid	Polyvidone	10	Excelvision, (France)	2
Maxitrol	Neomycin, Polymyxin B Sulfate	5	Alcon Lab, PJ (M)	3
Xepanicol	Chloramphenicol	5	Xepa-Soul Pattinson (M)	4
Brolene	Propamidine isethionate	10	Sanofi Aventis, (UK)	5

3.2.2.2 MALDI Analysis

Mass spectra analysis was performed using a Voyager-DE Pro Biospectrometry Workstation MALDI-TOF MS. 2,5-dihydroxybenzoic acid was used as the matrix

with HPLC-grade water to dissolve the sample and matrix solution. Samples were analysed on hydrophobic MALDI plates in a positive linear mode for polymer detection and in a positive reflectance mode for the detection of substances with lower molecular weights (a detailed description of the procedure is given in section 2.2.3, page 96).

3.2.2.3 LC-MS Analysis

A representative sample was taken to be analysed by ESI-MS for the confirmation of the lower molecular weight molecules. The sample was diluted (10-fold) with methanol and analysed using direct injections in positive modes. Several peaks were selected for MS/MS analysis to see the fragmentation patterns (a detailed description of the procedure is given in section 2.2.2, page 95).

3.2.2.4 NMR Analysis

NMR measurements were taken using a Bruker Avance 500 MHz NMR spectrometer equipped with broadband and triple resonance (^{1}H, ^{13}C, and ^{15}N) inverse probes. 540 µl of the eye drop sample was pipetted into the NMR tube and 60 µl of D_2O with 0.1% of TSP was added. Samples were centrifuged. The sample was then analysed using the Bruker Avance 500 MHz NMR spectrometer using the water suppression by excitation sculpting with gradient method. The spectra were acquired using 256 scans, four dummy scans, a relaxation delay of 2, spectral width of 20 ppm, and referenced to the TSP signal. One sample was analysed on day 1, day 2, and on the following week (day 8) for stability assessment.

3.2.2.5 NIR Analysis

NIR reflectance analysis of the liquid sample was conducted by pouring about 3 ml of the clear liquid samples into glass vials for each sample. The vials were analysed in the standing position on the sample stage. Samples in original plastic bottle packaging and an empty bottle were analysed for comparison purposes.

The work was carried out by conducting transflectance analysis of five additional samples with different brand names and compositions. 2 ml of the sample was measured in a measuring cylinder and transferred directly into the quarts reflectance cell (Hellma, Essex, UK) with a reflector disc. Three different types of reflectors made of gold, silver, and plastic were used for comparison purposes

(Figure 3.2). The disc was immersed in the cell and placed opposite to the side of the light source on the sample stage. The cell and discs were washed with methanol and HPLC-grade water and then dried using a hairdryer between the samples. Spectra were collected over the wavelength range 1100–2500 nm. Data acquisitions were done in triplicate and averaged for each sample.

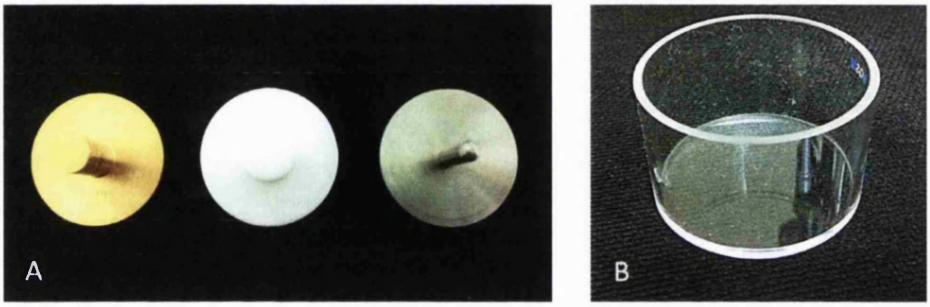


Figure 3.2 NIR transflectance analysis: (A) three types of the reflector discs; gold, plastic, and silver; (B) quarts reflectance cell.

3.2.3 Part III: Analysis of pellets: Chinese herbal preparations

Two brands of traditional Chinese herbal preparation, 'Po Chai Pills' (Li Chung Shing Tong, Hong Kong) and 'Chi Kit Teck Aun' pills (Eagle & Pagoda Brand Teck Aun Medical Factory, Malaysia) were analysed by NIR, MS, and NMR to compare within products. Both products were made from several herbs formed into tiny spherical pills about 4 mm in diameter, with one dose in each bottle or packet (Figure 3.3). They are traditionally used for the symptomatic relief of food stagnation, including epigastric or abdominal bloating, abdominal cramps, frontal headache, belching, hiccoughs, and nausea [116].

Although both products have similar physical properties, a closer look at the claimed ingredients on the packaging (listed in Table 3.3) showed that slightly different herbal components were used in these products. Although this type of product has a long history of usage, particularly in Asia, 'Po Chai Pills' have been associated with sibutramine [117] and diclofenac [118] adulteration in recent years.

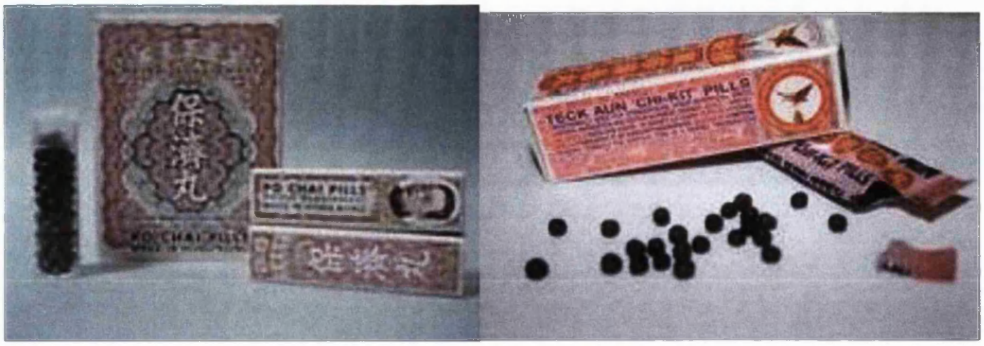


Figure 3.3 'Po Chai Pills' (A) and 'Chi Kit Teck Aun' pills (B). The herbal ingredients are prepared into tiny spherical pills and consumed as one dose per packet/ bottle.

Table 3.3 Listed ingredients for two types of Chinese herbal preparations; 'Po Chai Pills' and 'Chi Kit Teck Aun Pills'.

Samples	Manufacturer	Claimed ingredients
Po Chai Pills	Li Chung Shing Tong, Hong Kong	*Polyporus* (fu-ling) southern tsangshu (Cag-Zhu) rhizome, patchouly (Huo-Xiang) herb, fragrant *Angelica*, kudzu (Ga-Gen) root, *Magnolia* root bark, massa fermentata, job's tears seed, germinated rice, *Trichosanthes* root, field mint herb, halloysite, mum flower and red *Citrus* peel.
Chi Kit Teck Aun	Eagle& Pagoda Brand Teck Aun Medical Factory, Malaysia	*Pericarpium citri* R. *Radix angelicae* D, *Semen arecae*, *Rhizome altrastylodisM. fructus citri*, *Herba asari*, *Herba pogostemonis*, *Oleum menthae*, *Radix glycyrrhisae*, *Fructus chaenomies*, *Fructose amomi*, Herba mentha, Radix auklandiae, Poria

3.2.3.1 NIR Analysis

The challenge in analysing this set of samples by NIR was due to their unique sample dosage forms, namely tiny spherical pills. Analysing individual pills was impossible because of their size. Thus, samples were analysed per dose i.e. all the tiny spherical pills in each bottle or pack were transferred into a 4 ml clear glass

vial. Spectra were acquired using similar methods to powder analysis (section 2.2.1.2, page 95) in triplicate and the average was taken for further PCA.

3.2.3.2 NMR Analysis

The sample was then crushed finely using a pestle and mortar and dissolved in D_20 with TSP added as an internal standard. Samples were sonicated for 15 minutes and ultra-centrifuged for another 5 minutes at 13,000 rpm. 650 µl of the supernatant was transferred into the NMR tube. ^{1}H-NMR analysis was further conducted to compare products' compositions based on the spectra acquired.

3.3 Results and Discussion

3.3.1 Analysis of paracetamol tablet

3.3.1.1 NIR spectra analysis

The raw NIR spectra were mostly influenced by the physical properties of the samples, such as shape, surface structure, or marking, and the distribution of the APIs. This caused the absorbance recorded to be slightly shifted between spectra. The effect was more apparent on the samples analysed as intact tablets compared with powdered samples. Furthermore, the increased baseline of the spectra was also observed because of the scattering effects produced by the instrument (Figure 3.4A).

The SNV pre-treated spectra eliminated the effects of this absorbance shift but did not normalise the increased baseline. The spectral feature of paracetamol (PCM) in these spectra were in good agreement with previously published results [30] whereby the spectrum of paracetamol showed characteristic spectral features at about 1525 and 1625–1675 nm; this region seemed to be the least affected by the difference between intact and powdered samples (Figure 3.4B).

The second derivative correction improved the spectra by removing the increased baseline/absorbance shift and thus enhanced the spectral resolution [99], allowing easier comparison of the absorbance spectra of the two samples. Each spectrum was centred on zero with a segment size of 1 (Figure 3.4C).

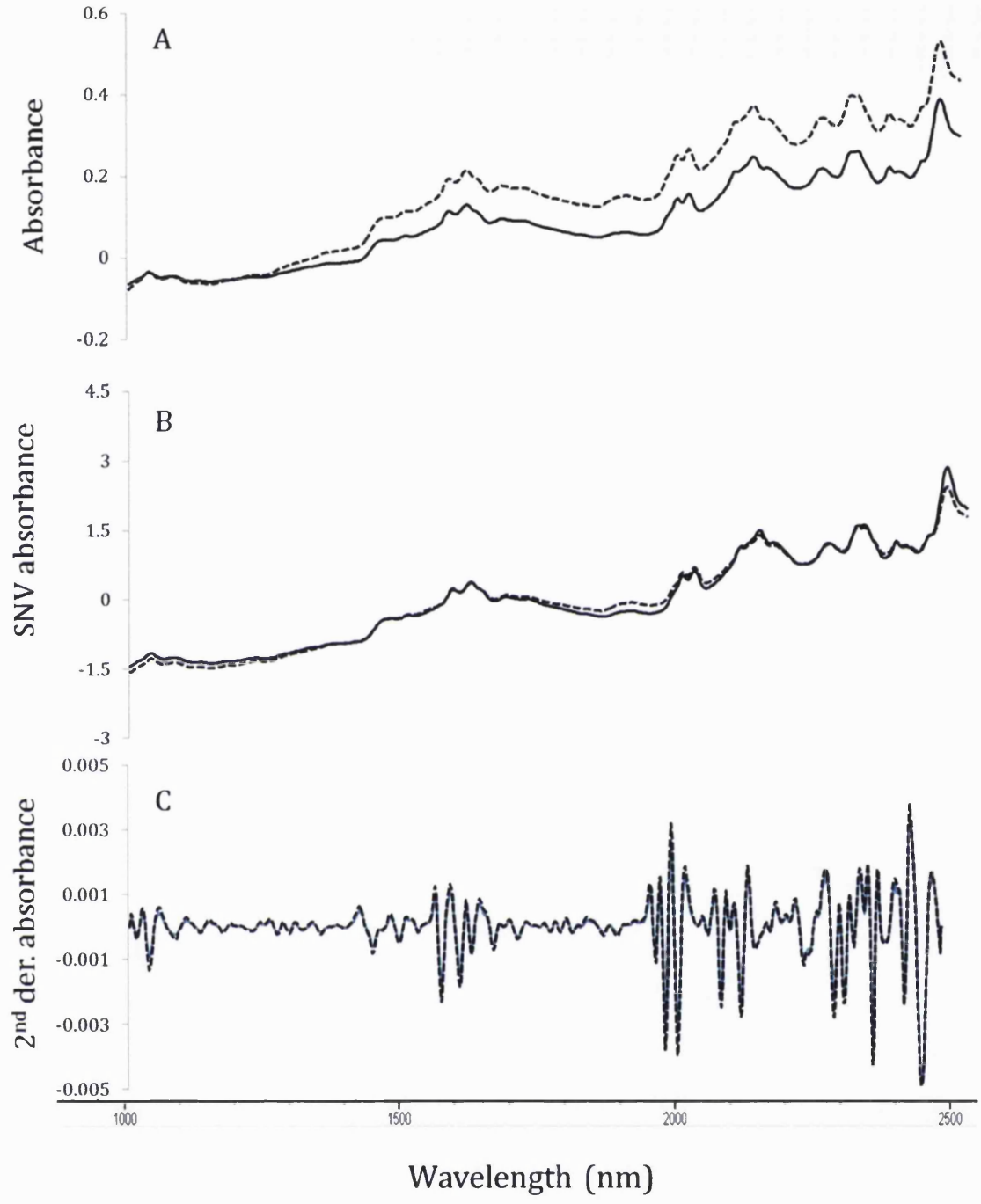


Figure 3.4 The NIR spectra of one paracetamol sample (MIL_B) as an intact tablet (solid line) and a powdered sample (dashed line) in three different formats: (A) raw, (B) SNV, and (C) second derivative.

3.3.1.2 PCA

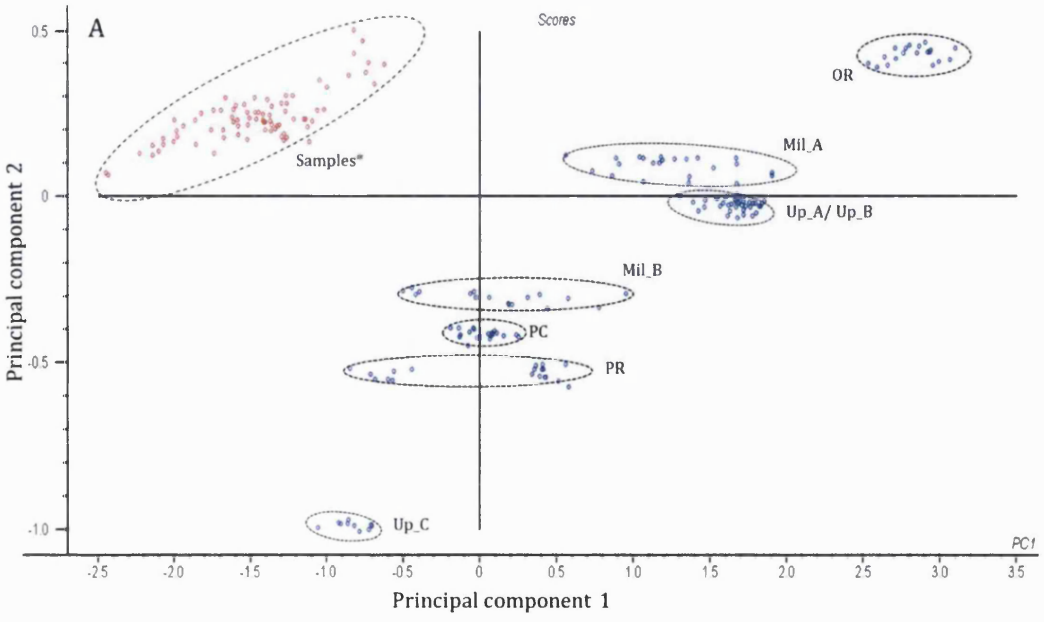
The PCA of the raw spectra showed that spectra for samples analysed as intact tablets form a distinctive classification unlike powdered samples (Figure 3.5A). This was because of the contribution of the physical characteristics of the intact samples in the spectra representation. Using the powdered samples standardised

the physical characteristics of each product to some degree and gave a better representation of the ingredients of the samples.

This observation was improved by using the SNV spectra; thus, the effects of the physical characteristics of the samples were ignored and classifications were made based on the differences between the chemical compositions in each sample set (Figure 3.5B).

By contrast, the second derivative spectra (Figure 3.5C) highlighted the major component of the sample, which in this analysis was the main API of the sample, paracetamol. As all the samples contained paracetamol, a general differentiation between intact and powdered samples was observed but without specific classifications according to source or different brand names.

Based on this observation, product classification for tablet samples can best be achieved when using raw spectra without pre-treatment. However, for the analysis of powder materials, using SNV spectra can classify the sample based on its chemical compositions. However, in this case, samples from the same source or using the same formulations could be classified together. This led to a poor classification between batches of samples. PCA using the second derivative spectra, however, was more suitable for classification between different types of medicines.



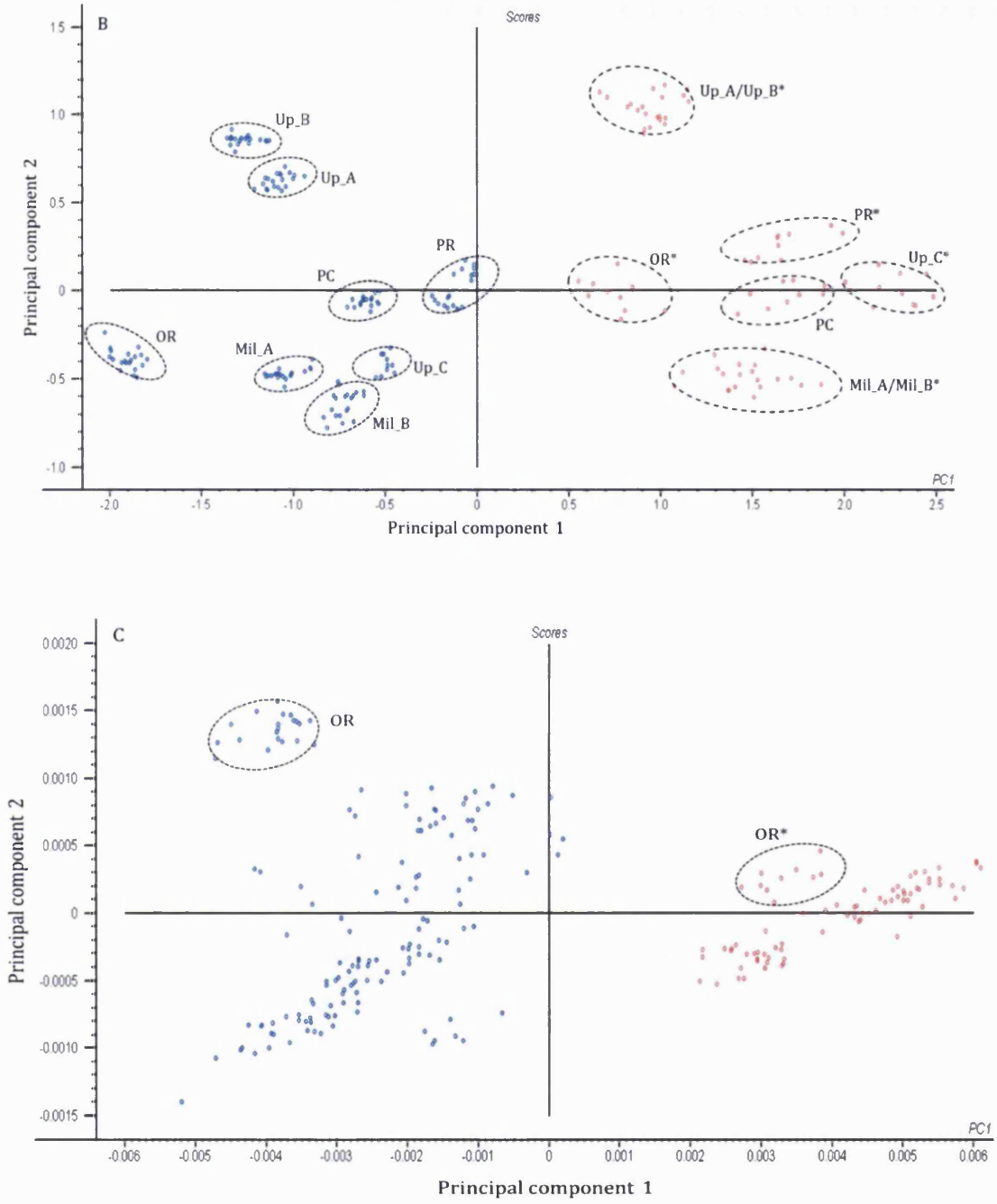


Figure 3.5 PCA analyses of paracetamol samples using (a) raw spectra showing all powdered* samples (red) clustered together without discrimination between different brand names unlike the tablet samples (blue), which show a distinct classification for each set (B) SNV pre-treated spectra showed good discrimination between intact and powdered samples and (C) second derivative spectra showed a good clustering between similar properties of drugs but were unable to differentiate between different sources.

3.3.2 Analysis of liquid: Eye Mo Moist (eye drop)

3.3.2.1 Mass spectrometry

The analytical parameters for the analysis of the liquid samples were observed in this work. The spectra fingerprints, stability profiles, and identification of the actives compounds of one type of eye drops, *Eye Mo Moist,* were also established in this work.

MALDI was used for the spectra profiling of this product because of its soft ionisation property. The analysis of large organic molecules such as polymers tends to be fragile and they can fragment when ionised by other ionisation methods.

Two methods were used: the reflectance method to identify compounds at the lower molecular weight level (Figure 3.6) and the linear method to identify the polymers (HPMC) at the higher molecular weight. HPMC is an ophthalmic lubricant and is the main high molecular weight composition of the product detected by mass spectrometry. Further optimization of the acquisition condition has to be carried out for the detection of HPMC in the linear method.

All the 22 samples analysed showed almost the same profiles as shown in Figure 3.6. Further analysis was conducted using ESI-MS/MS to identify the fragmentation patterns of the main peaks identified for the lower molecular weight compounds. Table 3.4 lists the peaks detected by ESI-MS and the fragmentation of the selected peaks by MS/MS. Most of these peaks corresponded to possible fragmentation ions derived from the chemical structures of the parent ions of edetate disodium and benzalkonium chloride as shown in Figures 3.7 and 3.8, respectively.

Table 3.4 Peaks detected by ESI-MS and the fragmentation of some peaks by MS/MS.

Analysis	Peaks detected	Fragmentation ions
ESI-MS	143, 165, 212, 263, 304, 332, 360, 381, 382, 408	
ESI-MS/MS	304	91, 136, 212 (@35)
	332	136, 331 (@ 25), 240 (@35)
	360	257, 287 (@25) 117, 268, 360 (@35)
	388	160, 328. 9 (@25) 119, 296.3 (@35)

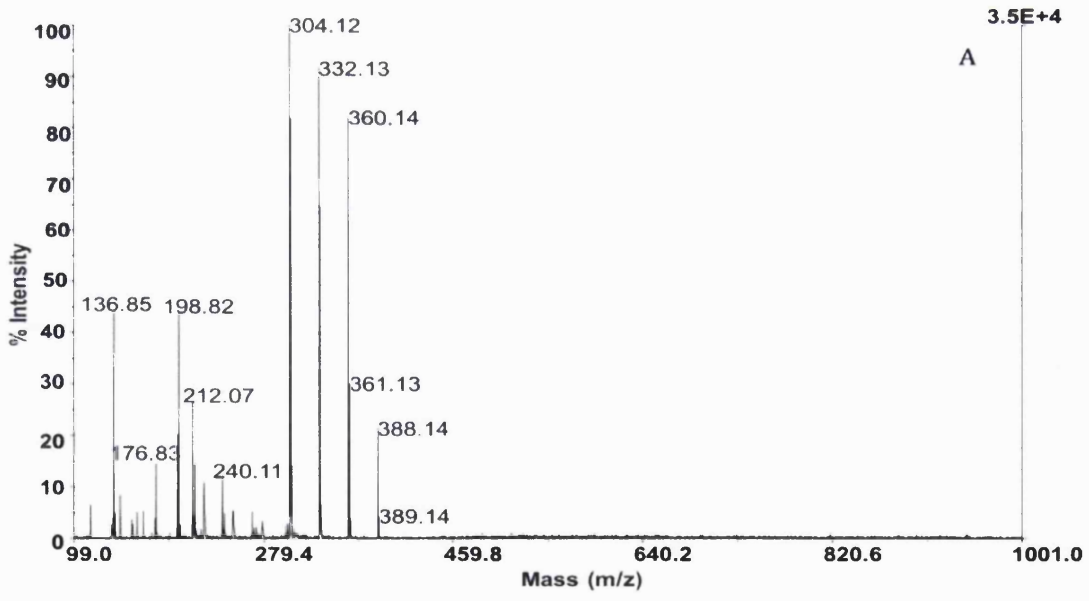


Figure 3.6 MALDI spectra for Eye Mo Moist analysed in reflectance mode, highlighting the lower molecular weight compound.

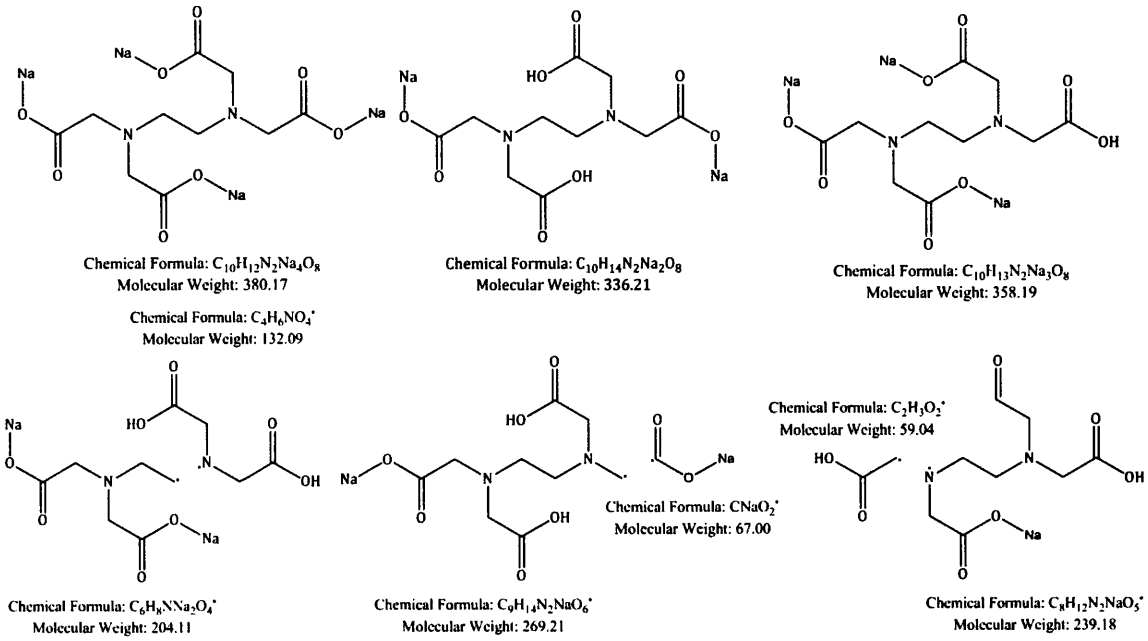


Figure 3.7 The chemical structures of the possible fragment ion patterns of the edetate disodium compound.

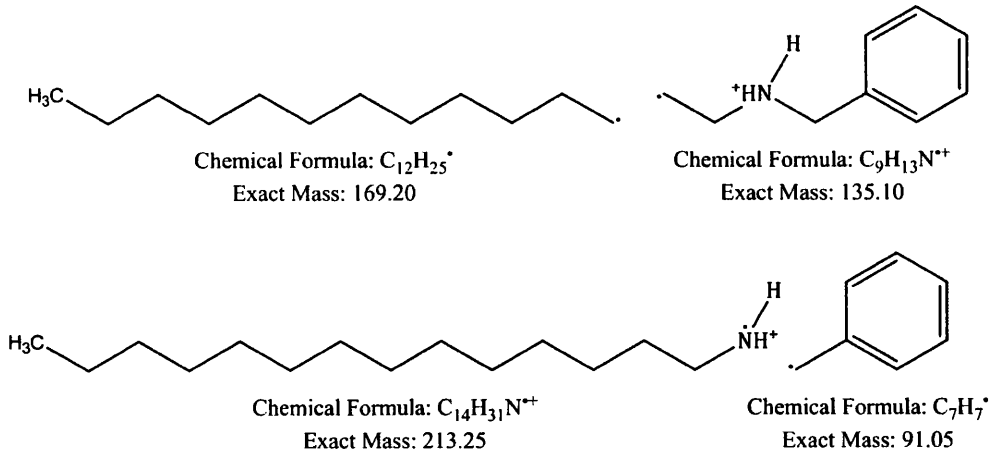


Figure 3.8 The chemical structures of the possible fragment ion patterns of the benzalkonium chloride compound.

3.3.2.2 NMR Analysis

The NMR spectrum (Figure 3.9) shows the peaks for HPMC present between 3.4 and 3.8 ppm and a few peaks of benzalkonium chloride at 1.2 ppm. Analysis of the sample was conducted on days 1, 2, and 8 after the first dilution on day 1 in order to observe the stability of the samples. The spectra were similar for days 1 and 2

but a degree of product aggregation was observed for the spectra acquired on the 8th day. This reflects the importance of the immediate analysis of samples. Using both techniques (NMR and MS) provided complementary results, where MS was used to identify small molecules and NMR confirmed the polymer that was difficult to identify by MS.

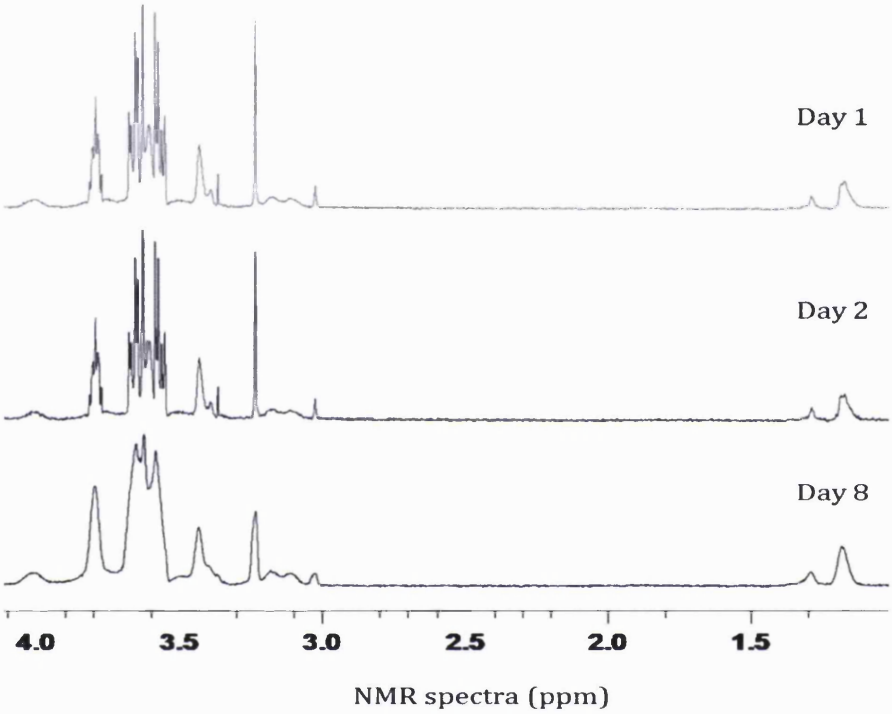


Figure 3.9 Stability study of samples. NMR spectra acquired at day 1 were similar to day 2 but showed a degree of aggregation after a week (day 8).

3.3.2.3 NIR Analysis

Early work on the NIR reflectance analysis of liquid preparations identified five possible constituents in ear drop preparations [119]. Subsequent works showed that this technique had been exploited for direct analysis through product packaging such as intact glass [120] and plastic [121] vials using the diffused reflectance and transmittance methods, respectively. It was expected that the container would contribute to the NIR absorbance spectrum. However, this effect could be normalised by incorporating the contribution in the calibration models.

While this research was used as an internal quality control of specific products, Chong *et al.* focused on a universal qualitative classification model that could identify products from different manufacturers [24]. This was possible using a multi-level classification model with more rigid parameters.

In the first analysis, diffused reflectance spectra were acquired for one type of eye drop, *Eye Mo Moist.* Figure 3.10 shows the spectra of three representative Eye Moist samples analysed in glass vials (red), original packaging (green), and an empty bottle with no liquid sample inside (blue). Spectra were recorded (A) in raw format (B) as SNV pre-treated spectra and (C) as second derivative spectra.

The spectra for the A and B samples did not differ much from one another. This was in agreement with PCA, whereby the samples seemed to cluster close to each other. Since spectra were acquired through the bottom of the bottles, every spectrum contains a contribution from the bottle. This may cause interference with future analysis that involves different brands of eye drops, which may or may not use similar types of plastic bottles. However, this factor can be normalised by using an additional application attached to the instrument (i.e. a fibre optic probe), which thus precludes the possibility of picking up the surface reflection from the glass, or by including the variation in the calibration model [24].

Another consideration in using this method for liquid analysis was that the first 1 mm contributes as much as 99% of the spectrum [122]. Thus, the separation of the water or oil layer on the glass window or on the inner surface of the plastic bottle could result in the reflectance spectra not representing the entire sample. Although this imposed a greater problem in terms of the identification of the product, the focus of this current work is to observe the distribution and classification of similar types of products.

Common methods for liquid analysis are transmission or transflectance [123]. The transflectance path of radiation was created by immersing a circular disc into the liquid sample to allow for a longer path length (1–2 mm) to be used to collect NIR spectra. Theoretically, this technique better represents the variation in components throughout the entire sample and not just on the surface [124]. Figure 3.11 shows the raw spectra of eight different eye drop samples analysed

using the transflectance method using gold (blue), silver (red), and plastic (green) discs.

The PCA of raw spectra shows the score plot distribution of samples according to disc type (Figure 3.12) in PC1 and PC2, indicating that the use of different discs affects the classifications. Higher numbers of PCs (PC5 and PC6) were required in order to reclassify the sample set according to specific brand names (Figure 3.13). Therefore, it is essential to use the same experimental setup in further work.

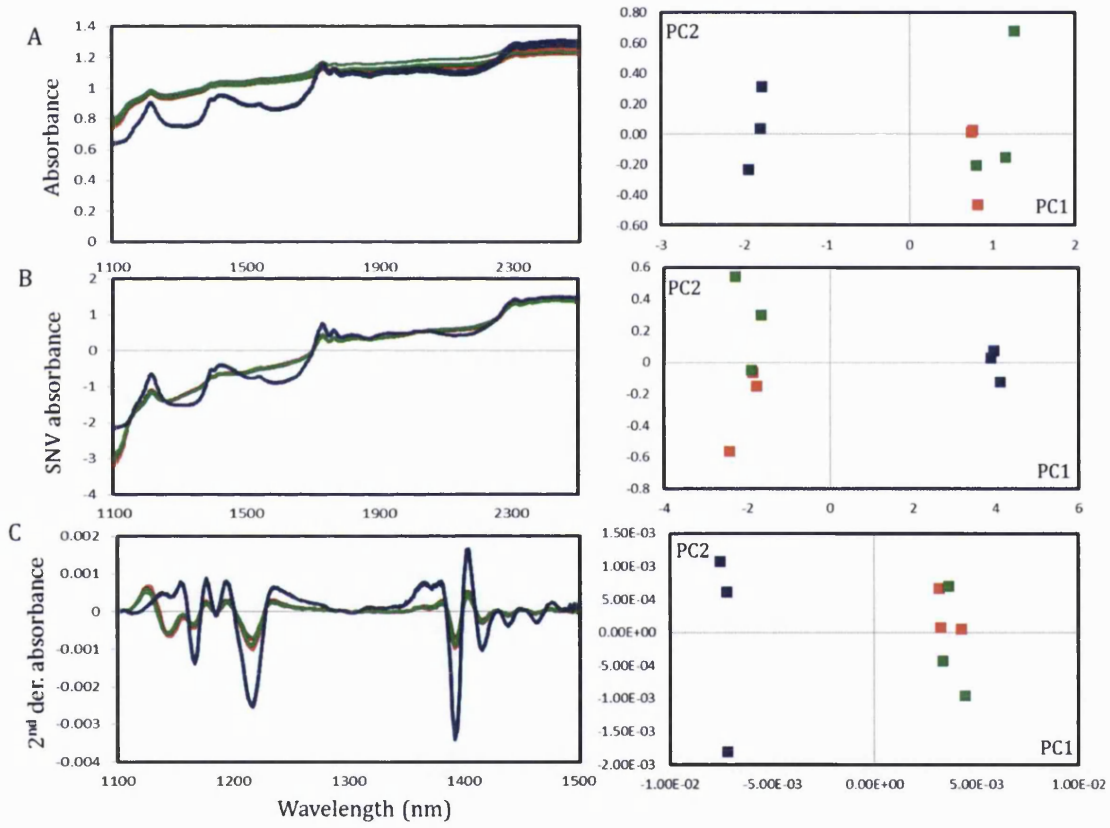


Figure 3.10 NIR spectra and PCA analysis of three *Eye Mo Moist* samples analysed in glass vials (red), in original packaging (green), and in an empty bottle with no sample (blue). Spectra were recorded in (A) raw, (B) SNV, and (C) second derivative formats. (PC= principal component).

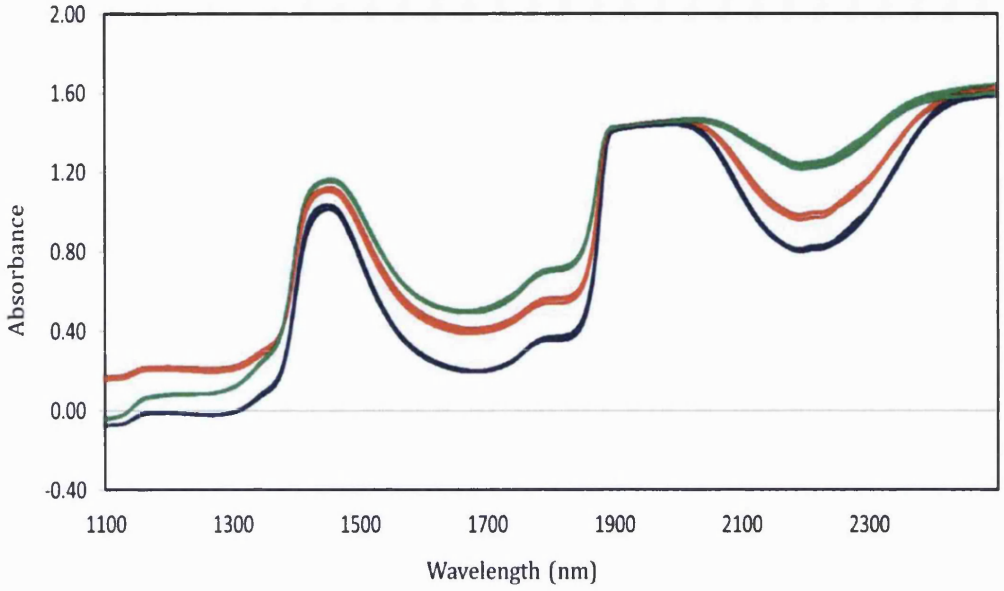


Figure 3.11 Line plot representing the NIR spectra of eight different types of eye drops analysed using the transflectance method using gold (blue), silver (red), and plastic (green) discs.

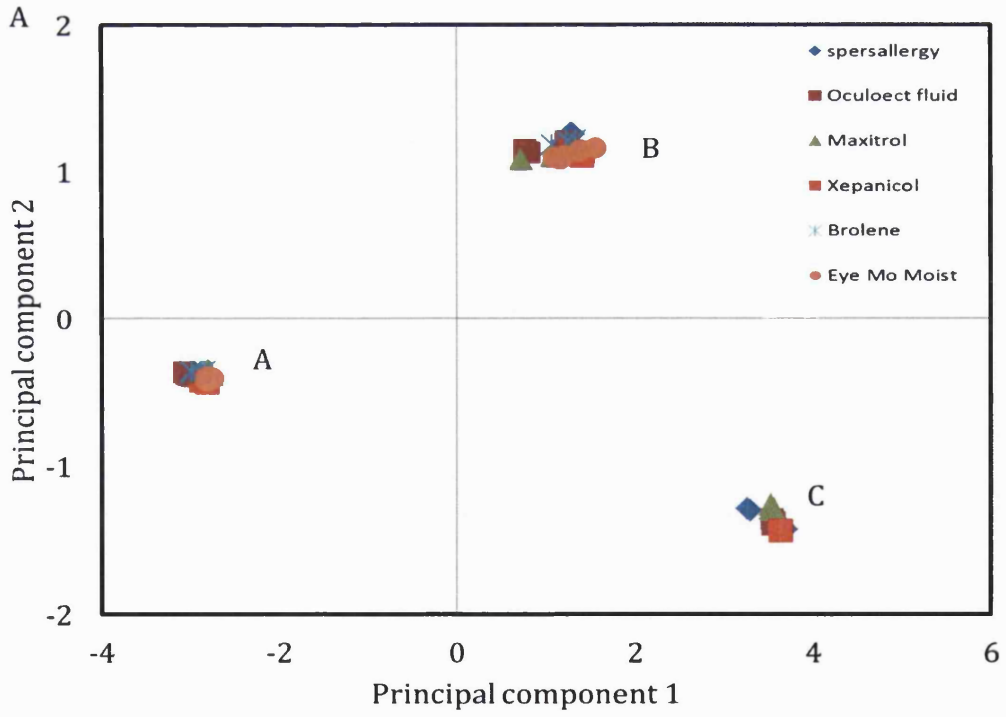


Figure 3.12 PCA score plot analysis shows the 6 different brands of sample clusters according to the types of discs; (A) gold (B) silver, and (C) plastic in PC1/PC2.

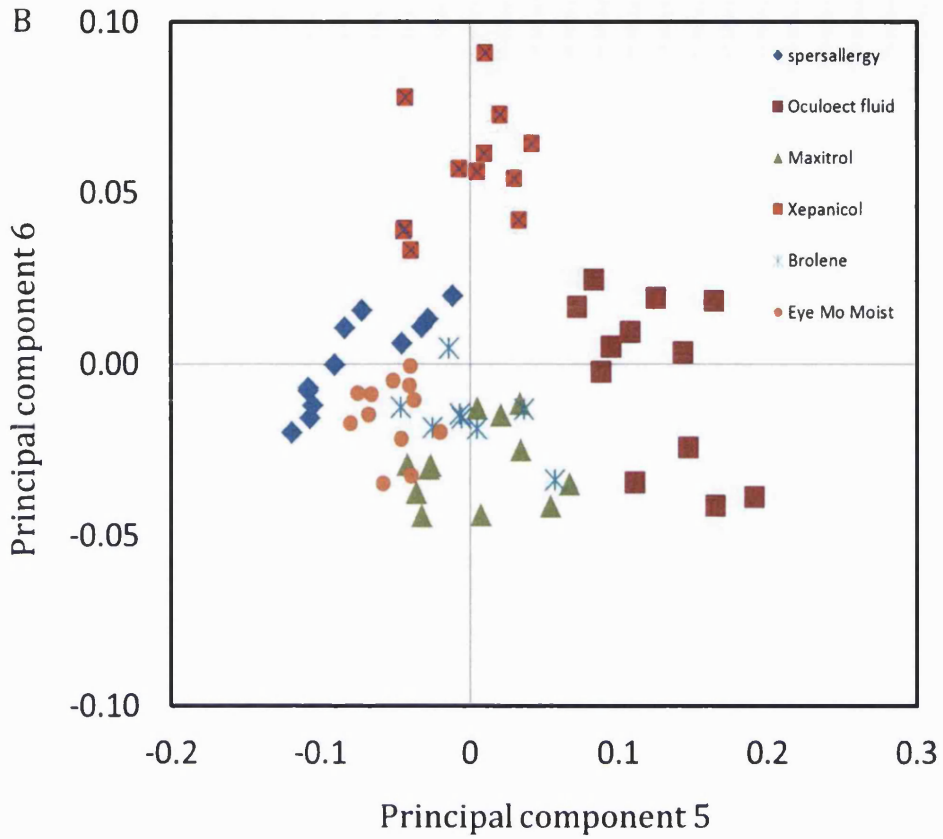


Figure 3.13 PCA score plot analysis shows better classifications of the samples according to the types of samples in PC5/PC6.

Transflectance analysis requires spectra to be pre-processed in order to exclude the effect of the discs on the spectra. In general, the spectra of the sample need to be extracted from the combination spectra of the sample and discs by deducting the spectra of the disc from the original spectra. This work, however, is not within the scope of this thesis. The higher PCs needed for sample classification could lead to more complicated procedures for spectra interpretation.

3.3.3 Analysis of pellets: Chinese herbal preparations

The two sets of products gave slightly different NIR spectra (Figure 3.14). This was expected based on the differences in the claimed ingredients on each packaging (Table 3.3).

Although there were few differences in spectra for the two types of herbal preparations, the PCA of the raw spectra shows a clear distinction between both sets of samples (Figure 3.15). In addition, further sample clustering was observed

within the 'Chi Kit Teck Aun' product (Figure 3.15 A and B) despite not seeing any visible discrepancy between its spectra.

The physical examination of the packaging and the samples themselves showed differences in packaging and the colour of the herbal pellets (Figure 3.16). NMR analysis was conducted to point out the chemical differences between the two sets of products. Additional peaks were observed (Figure 3.17) in two areas of the spectra: 2.0–2.2 nm and 6.9 nm. However, the exact identification of the chemical compositions of these formulations was beyond the scope of this work.

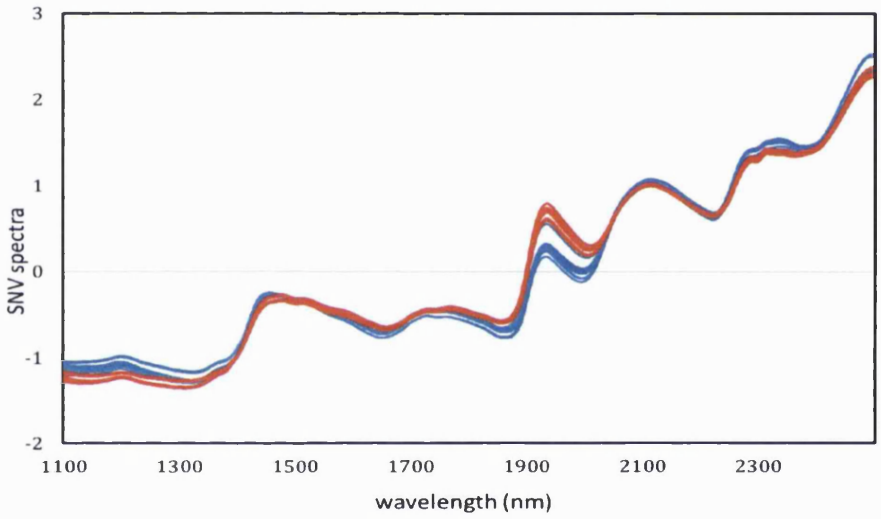


Figure 3.14 SNV spectra 'Po Chai' (red spectra) and 'Chi Kit Teck Aun' (blue spectra).

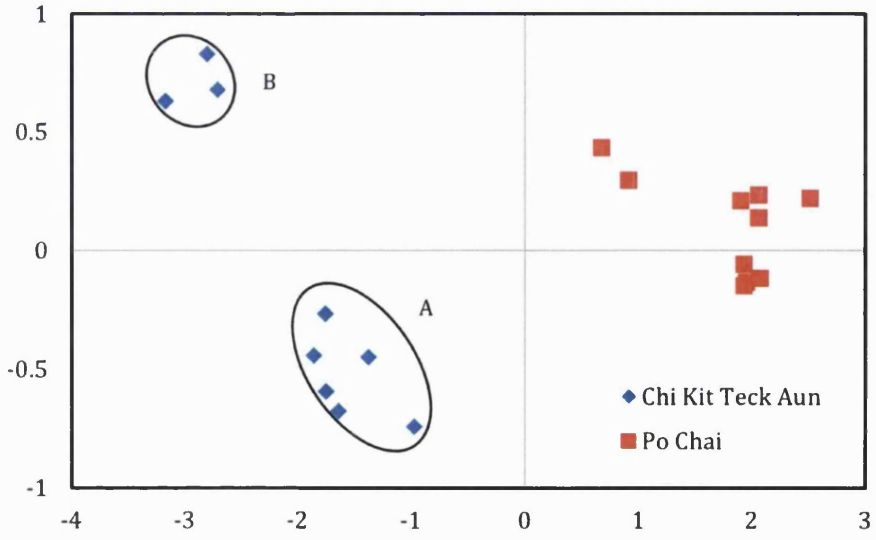


Figure 3.15 PCA distributions of samples 'Po Chai' and 'Chi Kit Teck Aun'. The spectra for samples 'Chi Kit Teck Aun' are further separated into two clusters.

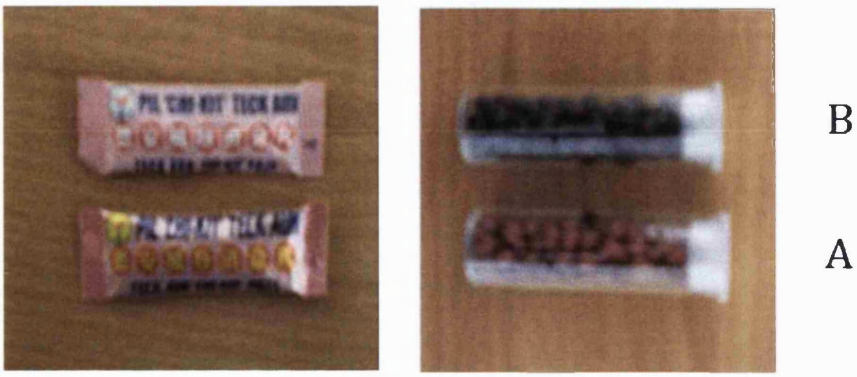


Figure 3.16 Different colours on the packaging and the 'Chi Kit Teck Aun' pellets that were distributed in two regions (A and B) in the PCA score plot.

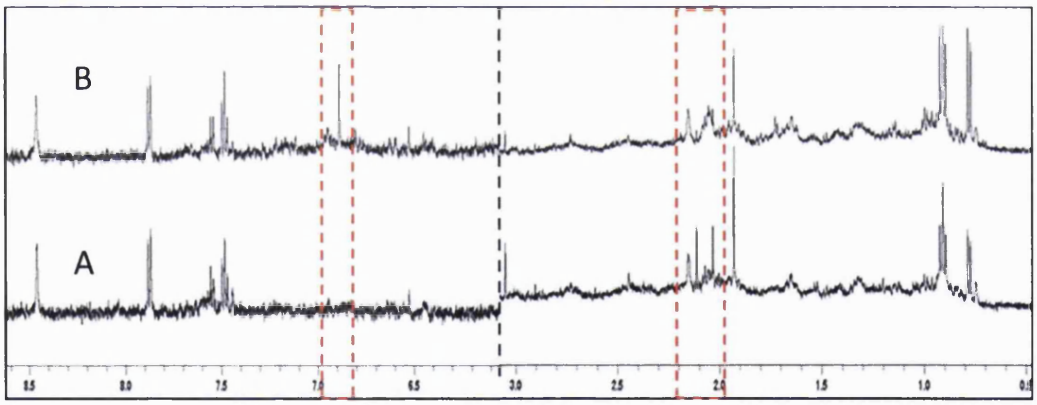


Figure 3.17 Comparison between the NMR spectra of 'Chi Kit Teck Aun' A and B shows the presence of additional peaks at two areas: at 2.0–2.2 ppm for sample A and at 6.9 ppm for sample B.

3.4 Conclusion

The compilation of the basic spectra fingerprints for products using different spectroscopic methods was conducted before transferring them into the database. One of the important parts of this research is to design a standard procedure for handling, storing, preparing, and finally analysing the samples followed by setting up the appropriate analytical procedures, parameters, and instruments.

NIR analysis allowed tablet samples to be analysed intact. Each set of tablets has its own physical characteristics. Therefore, the PCA of the original spectra showed distinct classifications between different brands of products. Crushing a solid preparation into powder form eliminated the physical characteristics of each product. This resulted in poor product clustering between different brand names of similar types of medicines. Powdered samples were analysed in glass vials.

The analysis of liquid materials used different methods (diffused reflectance, transmittance, and transflectance) and different ways (sample in the original plastic/glass bottle packaging versus glass vials). Different types of reflectors led to slightly different spectra produced. Data pre-treatment and PCA were observed to be more complicated in liquid transflectance analysis.

This chapter described only a few of the validation procedures for NIR analysis conducted prior to the real research work on the database. Other works such as the concentration study on NMR herbal analysis, effects of operator handling on analysis, the selection of bucket size in Amix analysis (for NMR), selection of different types of solvent and matrices (MALDI), instrument optimisation for mass spectrometric analysis (MS), selection and optimisation of the PCs, and the detection of outliers for paracetamol samples in PCA (chemometrics); some of these works are included in Appendix 2 and 4.

CHAPTER 4

DEVELOPMENT OF THE SPECTRAL DATABASE

4.1 Introduction

The identification of a molecule based on the comparison of the spectrum of an unknown sample with the reference spectra is a well-established technique [121]. In general, this approach involves an attempt to find spectral data in a reference library that is virtually identical to the spectral data of the unknown sample. If such a reference spectrum is found, the identity between the unknown and the molecule from the database is assumed. If no such reference is found in the library, the reference spectrum most similar to the spectrum of the unknown is retrieved [125]; however, further analysis would be needed to confirm or establish the identity of the unknown compound.

Earlier research describing methods for automatically matching unknown spectra with a library of reference spectra started by using punched cards configured that long rods could be used to extract the cards corresponding to similar compound, progressed to using automatic card sorters, and then using a binary representation of each spectrum after the computer-readable format established in the analytical community [62]. The search methods advanced with the development of various search strategies using peaks information [126] and further improved by using the chemical fingerprints of products [127].

An extensive overview of early work on spectroscopic databases was discussed by Wendy [53] followed by Davies and McIntyre [56]. In 1998, Lohninger provided a comprehensive collection of Internet resources based on his survey of existing databases and an overview on interpretation systems covering all major types of spectroscopy methods [57]. Many authors agree that these spectral database developments are essential because of the expansion of their use, from

sample identification to a fully integrated multi-spectra laboratory database system that includes structure elucidation, structure searching, and handling techniques [128]. Despite its potential, research on the advancement or improvement of spectroscopic database has been hampered due to a lack of funding and lack of interest [53].

In this work, databases were created using commercial software (GRAMS ID, Grams Suite Software, Thermofisher). Almost 5,000 spectra of different brands of 15 types of conventional medicines and three types of herbal preparations have now been recorded using three different instruments: NIR, NMR, and MS. However, because of the complexity of this research and time factor, only the applications on the NIR database for solid preparations are discussed extensively in this thesis.

The potential uses of the NIR spectral database in assessing pharmaceutical products are explored and demonstrated in this work. The exploitation of pattern recognition methods has led to the development of different types of search algorithms to utilise the information from broad and overlapping bands of NIR spectrum. Moreover, appropriate thresholds or cut-off values have been considered to ensure sufficient discrimination between different types of samples [129]. Blanco and Romero proposed using an NIR library as an alternative to qualitative analysis for identifying pharmaceutical raw materials [38]. The correlation coefficient algorithm has proven to be highly effective in identifying unknown materials by comparing their spectra against a library of 125 different raw materials.

However, a different approach was used in the development of the database of Malaysian medicine in which different brands of commercially available medicines are stored as the reference spectra instead of pure compounds. The rationale behind this strategy is based on the intended purpose of the database, which is to identify, classify, and monitor pharmaceutical products on a market based on their 'fingerprints', with a possibility of uncovering their sources or manufacturers.

The aim of the database design and search strategies was to compensate for the difficulty in obtaining authentic samples from manufacturers or other reliable sources, as was also experienced in this work. Since samples were sourced from

different locations, PCA and SIMCA were employed to ensure the authenticity of the samples before submitting them into the database. Furthermore, PCA and SIMCA were also used to validate this approach and ensure the correctness of the database search outcomes, since no similar work has previously been reported. The advantages of developing a database of commercially available products can benefit and enhance the work in the following two areas.

(1) Drug regulatory enforcement

The initial repository of spectral data could be collected during product registration by the DCA [130]. This could be added to currently established procedures. Full profiles of the products could then be added into the database and used for quick analysis during post-marketing surveillance and counterfeit medicine identification.

(2) Drug analytical studies in academic institutions

Prepared medicines could be obtained by liaising with manufacturers, community pharmacies, or hospitals to develop methods for monitoring the variability of products by time and location. Having a spectroscopic database of medicines is crucial because it can serve as a platform to develop and support other types of analytical research.

The objective of developing this database is to create a spectral data reservoir of the medicines available in the Malaysian market that can potentially be used as a screening technique for product identification (chapter 5), a drug quality study (chapter 6), herbal analysis and post-marketing surveillance (chapter 7), and the detection of counterfeit and adulterated products (chapter 8). These are, however, not the only possible uses of this type of database.

Although the ultimate objective of this database development is to create a customised search system based on structured query language programmes, because of the time constraints in the PhD programme, a commercial software (GRAMS Suite) was used as a starting point to investigate the feasibility of database use for the study of pharmaceutical products. A detailed procedure for creating a customised search system based on structured query language is discussed under the 'Future works' section.

In this chapter, the different types of search strategies and search algorithms available in the GRAMS software are evaluated based on their abilities to identify and classify unknown samples. The appropriate algorithms and cut-off values to differentiate between samples suitable for the intended use of this research are determined. A brief discussion on the quality control of the database is made at the end of this chapter followed by the proposed procedure for developing databases and performing searches of unknown products in the database.

4.2 Materials and methods

4.2.1 Materials

Two databases were specifically created to demonstrate and determine the best match algorithms and cut-off values for sample classification.

Database 1 consisted of spectra of 14 types of medicines (with the brand names in brackets) from different batches as the reference spectra; amoxicillin (14 different brand names), cefuroxime axetil (Xylid), diclofenac sodium (Voren), magnesium hydroxide/aluminium hydroxide (Macgel), mefenamic acid (Ponstan), paracetamol (14 different brand names), simvastatin (PMB Simvastatin), sibutramine HCl (Sibutramine), sildenafil citrate (Viagra), ticlopidine (PMB Ticlopidine), *Eurycoma longifolia* (Nuprep, herbal), *Ginkgo biloba* (g40, herbal), Chinese herbal mixture 1 (Po Chai), and Chinese herbal mixture 2 (PCTA).

Database 2 consisted of paracetamol (five brands; Bg_A, FP, Mil_A, OR, UP_A), amoxicillin (four brands; Amox_A, Amox_C, Amox_H, Amox_R), mefenamic acid (2 brands; MA_1, MA_16), *Ginkgo biloba* (three brands; Ginkgo_g40, Ginkgo_g120, Ginkgo_gd40), *Eurycoma longifolia* (three brands; TA_LKH, TA_hurix, TA_gb), and Chinese herbal mixtures (two brands; Pochai, PCTA).

4.2.2 Methods

4.2.2.1 NIR spectra acquisition and authentication

Figure 4.1 gives a summary of the step-by-step NIR spectral acquisition and authentication for a set of spectra that came from similar sources or manufacturers but comprised different batches. For the same brand of products,

variation between batches could be observed in the PCA score plots, but SIMCA should be able to classify them together. Samples that did not cluster with the rest of the group are not always outliers and should not be excluded. Such samples can provide information about inter-batch variations and, therefore, should be included in the database.

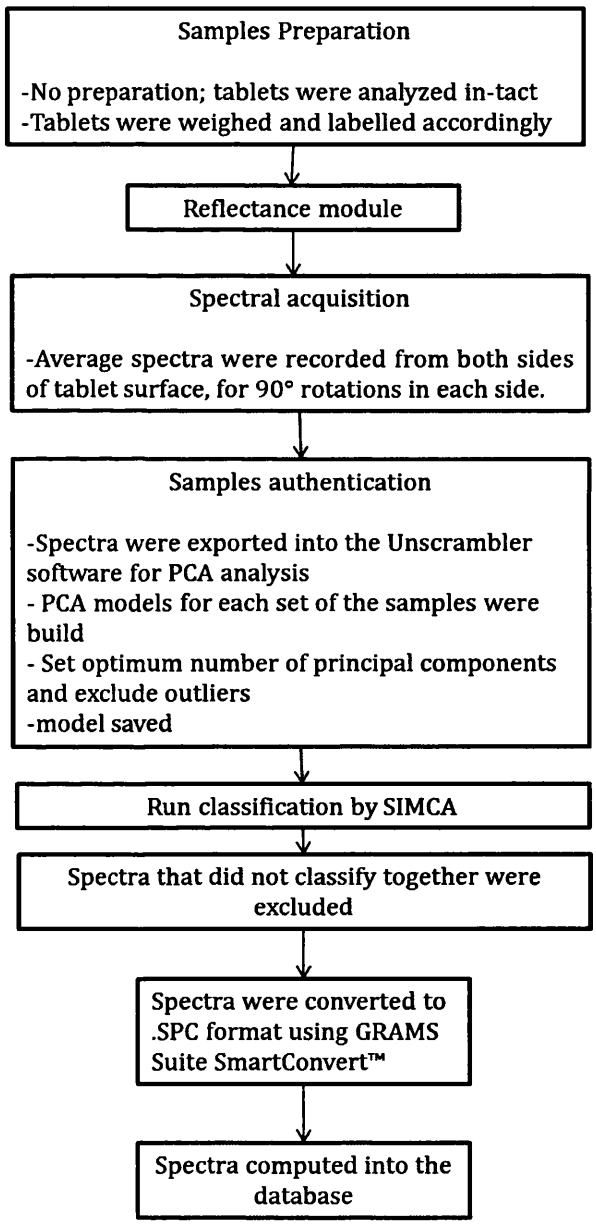


Figure 4.1 NIR spectra acquisition and authentication for a set of tablet samples from similar brands prior to inclusion in the database. Sample preparation ofother types of pharmaceutical dosage forms is described in section 2.2.1; page 94.

4.2.2.2 Database construction

The spectra were imported into the database together with the sample information, brand and proprietary name, batch number, expiry date, manufacturer name and address, origin, other excipients (where available), and description.

Before every search, the unknown spectra and other spectra in the database were set to be baseline-corrected to reduce the scattering effect that highly contaminates NIR spectra. This was performed using the auto-baseline correction algorithm that removes the linear baseline error of positive going peak data (GIFT auto levelling method, page 70) [103].

The database has a feature to mask or ignore some parts of the spectra that are insignificant or are constant variables (e.g. peaks for H_2O, CO_2, and CCl_4). However, this feature was not used in this particular work, since such variability can provide additional information about samples.

4.2.2.3 Database search methods

The spectral ID software allows database searches using three search methods: text, peak, and spectrum search. The text search was based on searching any text entry that is associated with a reference spectrum. This will not be discussed further in this work.

Figure 4.2 shows the outline of the search procedure involved in conducting a full spectrum and a peak search. For every search, the unknown spectra were always converted to the same format as that of the library entries. (i.e. SPC format).

4.2.2.3(a) Full spectrum search

The spectral ID software provides seven different algorithms for a full spectrum search: Euclidian distance, absolute value, first derivative absolute value, least squares, first derivative least squares, correlation, and first derivative correlation.

4.2.2.3(b) Peak search

The peak search system uses peak position and intensity level [69]. A forward search was conducted using the standard picking method with the level for minimum amplitude set at 0 and sensitivity set at 10.

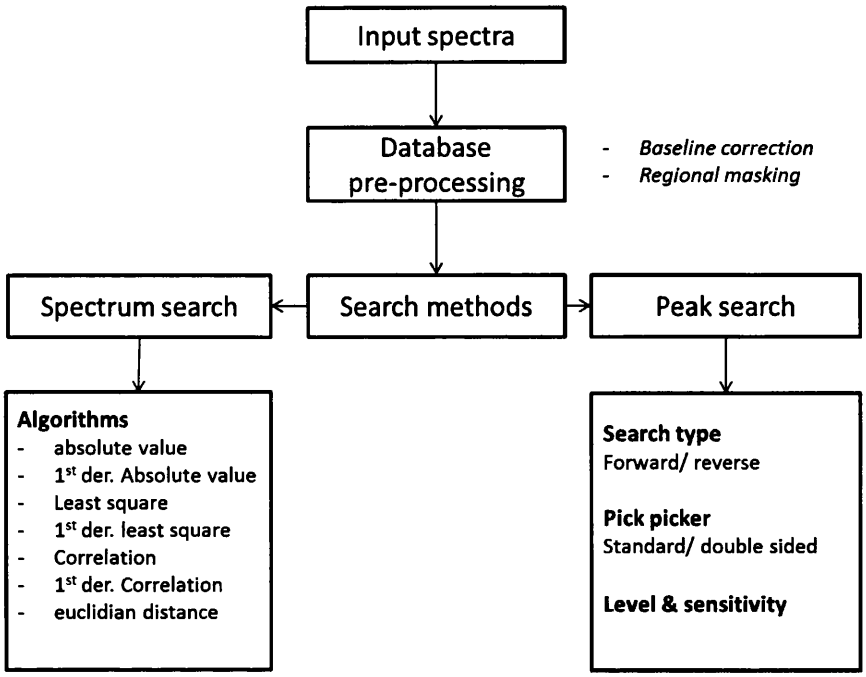


Figure 4.2 The search methods for optical spectra searches (NIR and NMR spectra) discussed in this work.

4.2.3 Cut-off point determination

Different techniques are required for identity searching and similarity searching [131]. Differentiation between the two classes of spectral searches was carried out by establishing some form of criteria based on a cut-off value or statistical threshold. In this work, these values were defined by performing successive spectral searches on similar types of medicines of known type and origin at the initial state and further expanding to a database of different types of medicines. If an unknown sample spectrum challenged against the library met these criteria, we had confidence that the sample had been correctly identified. The criteria were set to define four classes: best match, same brand/source/manufacturer, similar type, and different types of medicines. The same techniques for the identification and similarity assessment of unknown spectra were used to demonstrate the principles of the database.

4.3 Results and Discussion

The NIR database consists of a set of normalised spectra of selected medicines defining a set of reference spectra. Each spectrum comprised 700 data points and each was recorded in 32-bit resolution in a data range (1100–2500 nm) and with absorbance as the Y value. The peaks included in the peak table for each spectrum were selected using the standard single picker with default values of 0 and 10 for level and sensitivity, respectively. The same parameters were set for each new entry before adding it to the database, and each unknown spectrum had to have the parameters set to the reference spectra in the database for a search to be viable.

Throughout this PhD work, three databases were created for storing different types of the spectra (NIR, NMR, and MS). For each type of spectra, separate libraries were created within the database according to the class of drugs or herbal products. Spectral searches were conducted either by selecting all the libraries or by choosing selected libraries based on the objective of the search. However, in this chapter specific databases were created to test and demonstrate the capabilities of the database and the algorithms.

The reference spectra included in the database were obtained from different types of medicines purchased on the market and not from reference compounds or pure active pharmaceutical ingredients (APIs) as commonly used in other databases [38, 47, 31]. This is consistent with one of the objectives of this work. Spectra were acquired using samples 'as received' to reflect the real states of the medicines available on the market. Different types of medicines were included and, for every brand of a product, at least 20 averaged spectra were obtained and stored in the database to reduce variability among tablets. Only solid preparations were considered in this section.

4.3.1 Viewing the search outcomes

A database hit-list was generated for each search for the spectrum of an unknown sample (Figure 4.1). This list included the hit spectra with the best similarity to the unknown spectrum being searched. The similarity was indicated using the hit quality index (HQI), which was calculated using a specified method before the

search was carried out. The HQI represents the closeness of the match between the unknown spectra and a particular library entry.

For a full spectrum search, a perfect match gives an HQI of 0, while no match gives a worst case value of either 1.00 or 1.414 depending on the algorithm selected. For a peak search, this value ranges from 0% for no matching peaks to 100% for a perfect match. An example of the database hit-list is shown in Figure 4.3.

There are three main sections in the database hit-list. Section A is the spectrum display box. Here, the spectra for the best three hits are displayed in red (match spectrum), blue, and green spectra (next match spectra consecutively). In section B, the hit result box can display up to 20 spectra in the database that are closely matched to the unknown spectrum in order of their rank based on the HQI value. Other information regarding the selected reference spectrum from the library that was computed together with the spectra during the library building process are displayed in section C.

4.3.2 Database search algorithms

Previous research on the strengths and weaknesses of different types of search strategies has been reported [132, 133], which has mainly considered finding the best match to the unknown. For example, a confusion statistic has been used to assess the relative performances of different search algorithms [59]. The 'confusion factor' [134] provides a quantitative indication of the number of incorrect library entries that match the target spectra.

This method is useful for identifying a search strategy that indicates the mismatches of the spectra of similar compounds and the nearest matches. Furthermore, extensive modifications of the basic matching algorithms were applied to find the best-matched spectrum to the unknown.

In this work, a different approach was considered to compare the strengths of a search algorithm. Assessments were made based on the ability of an algorithm to identify the unknown (best-matched algorithm) and to classify the rest of the spectra in the library according to their types of medicines (best classification algorithm).

The direct comparison of search algorithms was performed based on the chemical fingerprints of the products using a single reference library as the source of all search files. Search performance was evaluated by examining the results of the validation and test searches using spectra of similar target products for all algorithms to allow a consistent comparison of several search techniques [59].

4.3.2.1 Full spectrum search

4.3.2.1(a) Best-matched algorithms

Successive spectral searches of several types of medicines were conducted to evaluate the results of the seven different algorithms provided in the software; Euclidian distance, absolute value, first derivative absolute value, least squares, first derivative least squares, correlation, and first derivative correlation. The examples of the outcomes of the search for unknown amoxicillin samples using these algorithms are listed in Figure 4.4. For this sample, the first derivative algorithms gave a better identification compared with other algorithms.

Further, we compared the outcomes of searches for another 13 different types of medicines used as the 'unknown' spectra in Figure 4.5. The unknown spectra were compared to the spectra of similar type of products that had already been saved in the database, but from a different batch set or from different manufacturers. A match between a pair of spectra leading to the correct identification of the medicine was marked as blue cells, while the red cells indicate a pair of spectra with incorrect identification.

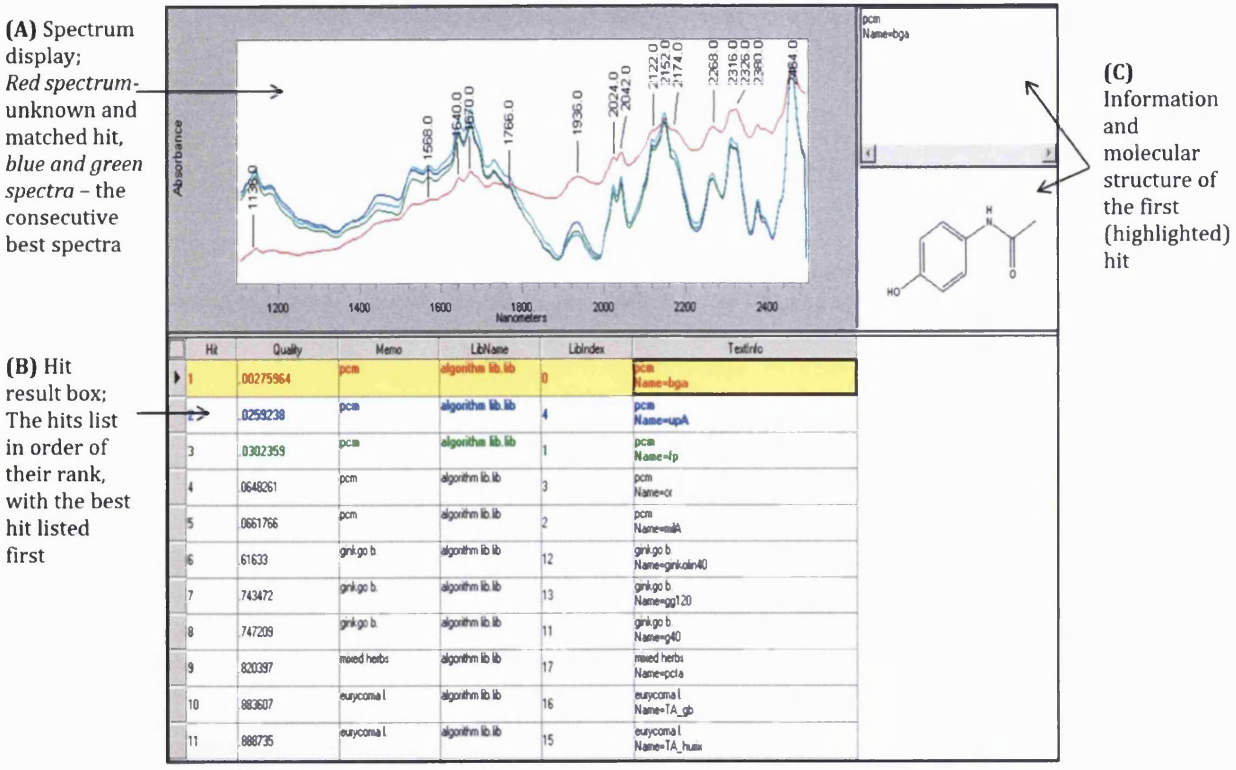
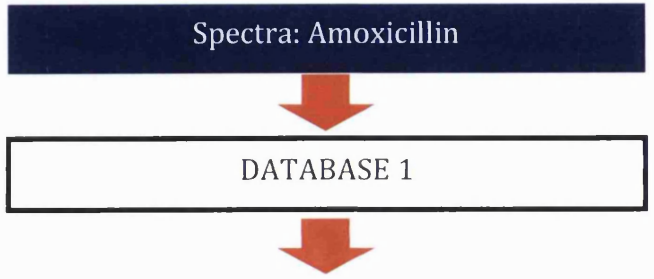


Figure 4.3 An example of the database hit-list. The spectrum display box shows the first three best-matched reference spectra to the unknown spectra with the peaks labelled (A). The hit result box can display up to 20 spectra in the database that are closely matched to the unknown spectrum based on HQI value (B). Information and the molecular structure of the highlighted hit spectra (C).



Search Outcomes

HIT#	[AV]	
	HQI	Sample ID
1	0.209705	Macgel
2	0.212474	Sibutramin
3	0.215512	TA_NUprep
4	0.235578	Viagra
5	0.248084	Pochai

HIT#	[1st AV]	
	HQI	Sample ID
1	0.004937	Amox_A
2	0.006836	NUprep
3	0.007067	Pochai
4	0.007202	Voren_05
5	0.007358	S.statin_39

HIT#	[LS]	
	HQI	Sample ID
1	0.070681	Macgel
2	0.071693	Sibutramin
3	0.076302	TA_NUprep
4	0.088897	Pcm_IF
5	0.099624	sildanafil

HIT#	[1st LS]	
	HQI	Sample ID
1	4.37E-05	Amox_A
2	0.000102	TA_NUprep
3	0.000105	Pochai
4	0.000109	Voren_05
5	0.000109	S.statin_39

HIT#	[1st C]	
	HQI	Sample ID
1	0.163155	Amox_A
2	0.788867	NUprep
3	0.792535	Pochai
4	0.838291	Voren_05
5	0.840106	S.statin_39

HIT#	[C]	
	HQI	Sample ID
1	0.931833	TA_NUprep
2	0.935747	Amox_A
3	0.955452	Pochai
4	0.97158	Ginkgo_g40
5	0.976671	Sibutramin

HIT#	[ED]	
	HQI	Sample ID
1	0.488872	PCTA
2	0.520935	Pochai
3	0.535116	Ginkgo_g40
4	0.543468	S.statin_39
5	0.54882	Voren_05

Figure 4.4 The top five search outcomes for the unknown amoxicillin sample using seven algorithms; absolute value [AV], first derivative absolute value [1st AV], least squares [LS], first derivative least squares [1st LS], correlation [C], first derivative correlation [1st C], and Euclidian distance [ED].

	Samples	AV	1st AV	LS	1st LS	C	1st C	ED
1	Amoxicillin							
2	Paracetamol							
3	Simvastatine							
4	Cefuroxime axetil							
5	Diclofenac sodium							
6	Mefenamic acid							
7	Sibutramine							
8	Ticlopidine							
9	Sildenafil citrate							
10	Macgel							
11	*Ginkgo biloba*							
12	*Eurycoma longifolia*							
13	Chinese herbal 1							
14	Chinese herbal 2							

Figure 4.5 Validation of the identification feature of the database for 14 different products using seven algorithms; absolute value [AV], first derivative absolute value [1st AV], least squares [LS], first derivative least squares [1st LS], correlation [C], first derivative correlation [1st C], and euclidian distance [ED]. Blue cells indicate correctly identified medicine based on the similar products in the database. Red cells indicate misclassification of medicine. Highlighted cells indicate that the spectra of the product from the same manufacturer were loaded into the database.

Figure 4.5 shows that the first derivative correlation algorithm [1st C] gave the best identification outcomes compared with the other algorithms. This is because one of the most common problems using NIR for analysing solid is the sloping baseline in the spectra, and this can be compensated for using such a derivative [135]. However, first derivative algorithms have one major drawback: if the compound is not in the library, the results will give a worse HQI match for spectra that have similar spectral features [62], which may prevent using the classification of products from different manufacturers of the same types of medicines.

4.3.2.1(b) Best classification algorithms

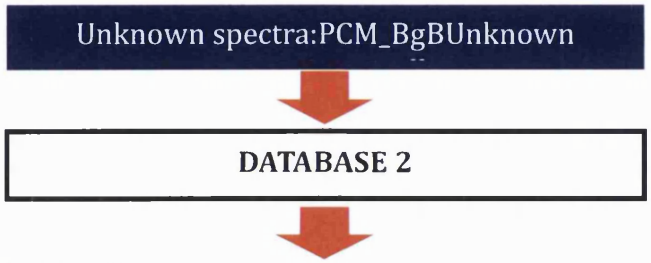
Subsequent searches were conducted to find algorithms that could best classify the rest of the reference spectra based on a search. The desirable outcome of a search algorithm application should indicate if the spectra of unknown products have a spectrum in the database that belongs to a product from (1) the same batch, (2) the same brand but a different batch, (3) similar types of medicines but from different brands, or (4) completely different products. The above outcomes would be achieved by comparing spectra with the reference spectra and predetermined cut-off values.

To find the best classification algorithm, the database was used to identify and then classify the paracetamol samples according to a specific manufacturer or source. Searches were carried out for one paracetamol product (Bg_B) against the reference spectra in database 2. Figure 4.6 shows the HQI of the first sample from each class of medicines when the spectral search was conducted for paracetamol Bg_B against the reference spectra in database 2 using different algorithms. A bar graph was developed using the HQI value to show the overall distribution of the reference spectra after each search (Figure 4.7). Lower HQI values indicated a higher similarity between two spectra.

The results showed that all seven algorithms correctly identified that spectra belonged to a paracetamol product; however, the correlation and Euclidian distance algorithms gave a higher degree of distinction between different types of medicines, especially compared with the first derivative correlation algorithm.

Although the euclidian distance algorithm gave a good result in sample classification, the overall performance of this algorithm in product identification was relatively poor (Figure 4.5). Thus, this algorithm was excluded at this stage. Further comparisons were made between the correlation and first derivative correlation algorithms to determine the most appropriate best-matched and best classification algorithm (Figure 4.8 A and B).

The results showed that the first derivative correlation algorithm misclassified three products in the search outcome (for all herbal samples (*G. biloba*, *E. longifolia*, and the Chinese herbal mixture) compared with two misclassifications by the correlation algorithm (*E. longifolia* and Chinese herbal mixture only)). Therefore, the correlation algorithm can be used in spectral searches when information on both the identification and classification are available. In cases where the HQI values for the correlation algorithm search provide inconclusive results, a combination of searches by using two algorithms can be used; first derivative correlation for sample identification and correlation for sample classification.



HIT#	Absolute value	
	HQI	Sample ID
1	0.004195	PCM_BgA
6	0.034899	TA_gb
7	0.035688	Ginkgo_g40
9	0.038999	MA_16
12	0.040864	Amox_A
17	0.06303	Pochai

HIT#	1st der. Absolute value	
	HQI	Sample ID
1	0.000103	PCM_BgA
6	0.002237	TA_gb
7	0.00226	Amox_H
10	0.002326	MA_1
14	0.002381	PCTA
15	0.002391	Ginkgo_g120

HIT#	Least square	
	HQI	Sample ID
1	2.02E-05	PCM_BgA
6	0.002075	TA_gb
7	0.00208	Ginkgo_g120
11	0.002983	Amox_A
12	0.003024	MA_16
17	0.005605	Po Chai

HIT#	1st der. Least square	
	HQI	Sample ID
1	0	PCM_BgA
6	1.27E-05	TA_gb
8	0.0000131	Amox_H
10	0.0000133	Ginkgo_g120
12	0.0000136	PCTA
15	0.000014	MA_1

HIT#	Correlation	
	HQI	Sample ID
1	0.00276	PCM_BgA
6	0.61633	Ginkgo_g40
9	0.820397	PCTA
10	0.883607	TA_gb
13	0.898465	MA_1
16	0.98185	Amox_H

HIT#	1st der. correlation	
	HQI	Sample ID
1	0.001518	PCM_BgA
6	0.89718	Ginkgo_g120
7	0.909758	TA_gb
8	0.909952	PCTA
9	0.911538	MA_1
15	0.932892	Amox_H

HIT#	Euclidian distance	
	HQI	Sample ID
1	0.02581	PCM_BgA
6	0.411143	Ginkgo_g40
9	0.479897	PCTA
11	0.522027	TA_hurix
14	0.600954	MA_1
15	0.606863	Amox_A

Figure 4.6 The top five search outcomes for the search of Bg_B spectra using different algorithms. The HQI value of the first sample identified in each class of medicines was recorded.

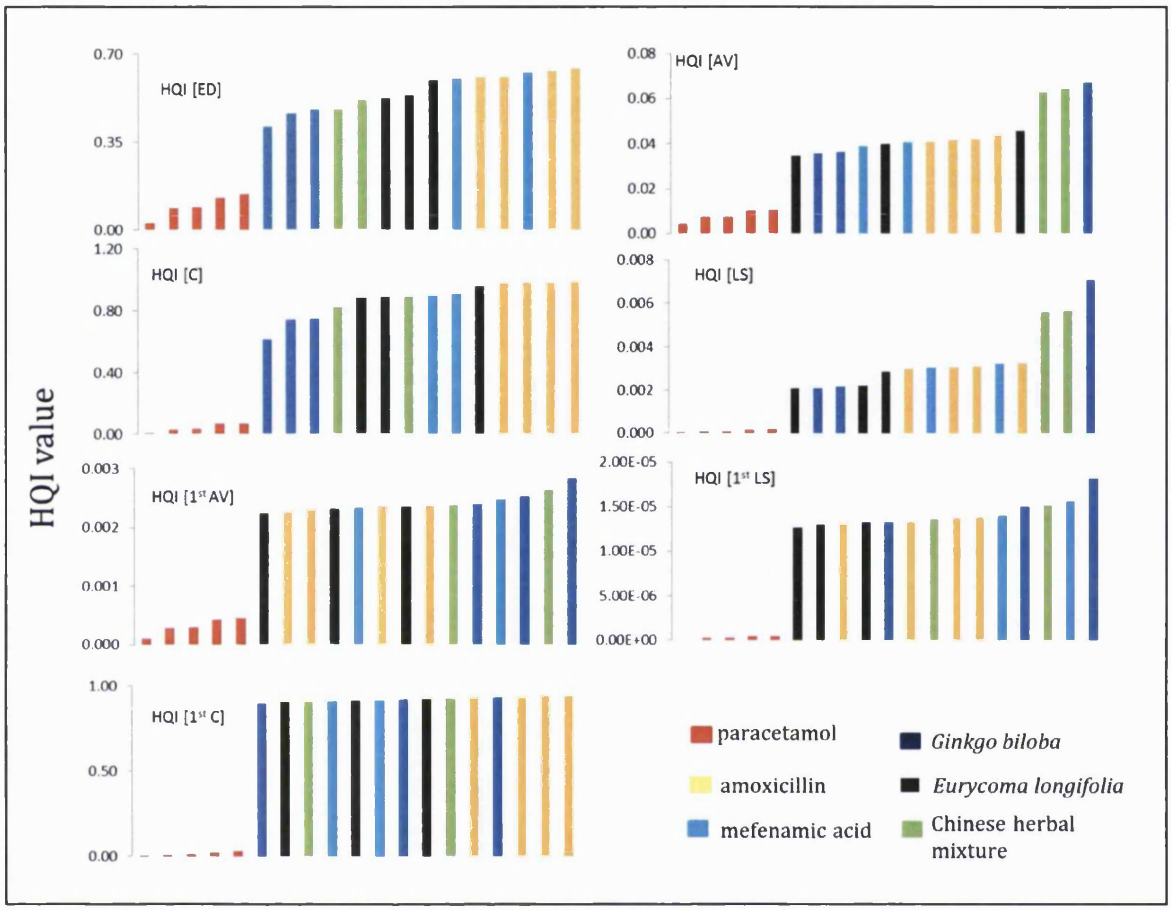


Figure 4.7 The search outcomes for the searches conducted using the spectrum of one brand of paracetamol (Bg_B) against a database consisting of different types of medicines. Seven algorithms were used; absolute value [AV], first derivative absolute value [1st AV], least squares [LS], first derivative least squares [1st LS], correlation [C], first derivative correlation [1st C], and euclidian distance [ED]. The scales are different for each algorithm based on the calculations of individual set.

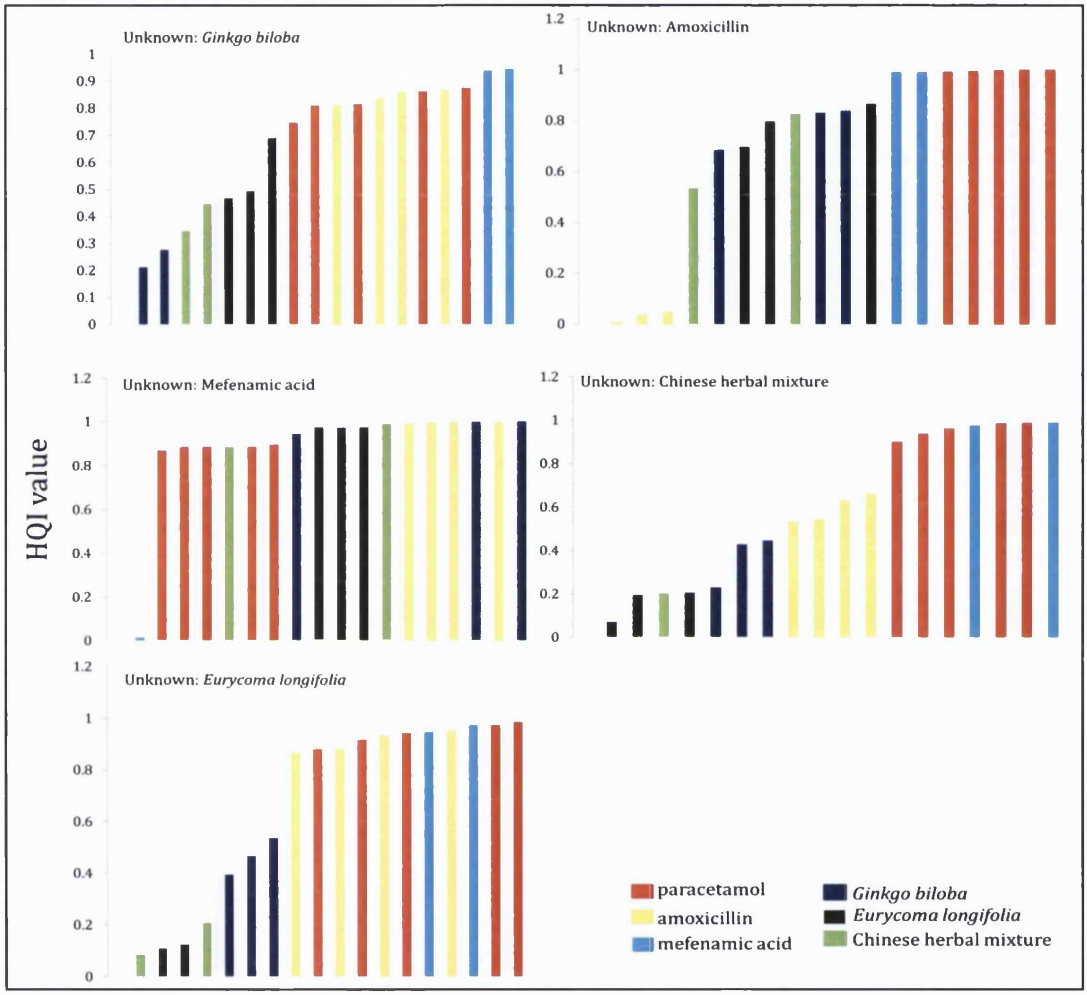


Figure 4.8A The search outcomes for each single type of medicines in a database consisted of the spectra of *Ginkgo biloba*, amoxicillin, mefenamic acid, Chinese herbal mixture, and *Eurycoma longifolia* using correlation algorithm. The scales are different for each algorithm based on the calculations of individual set.

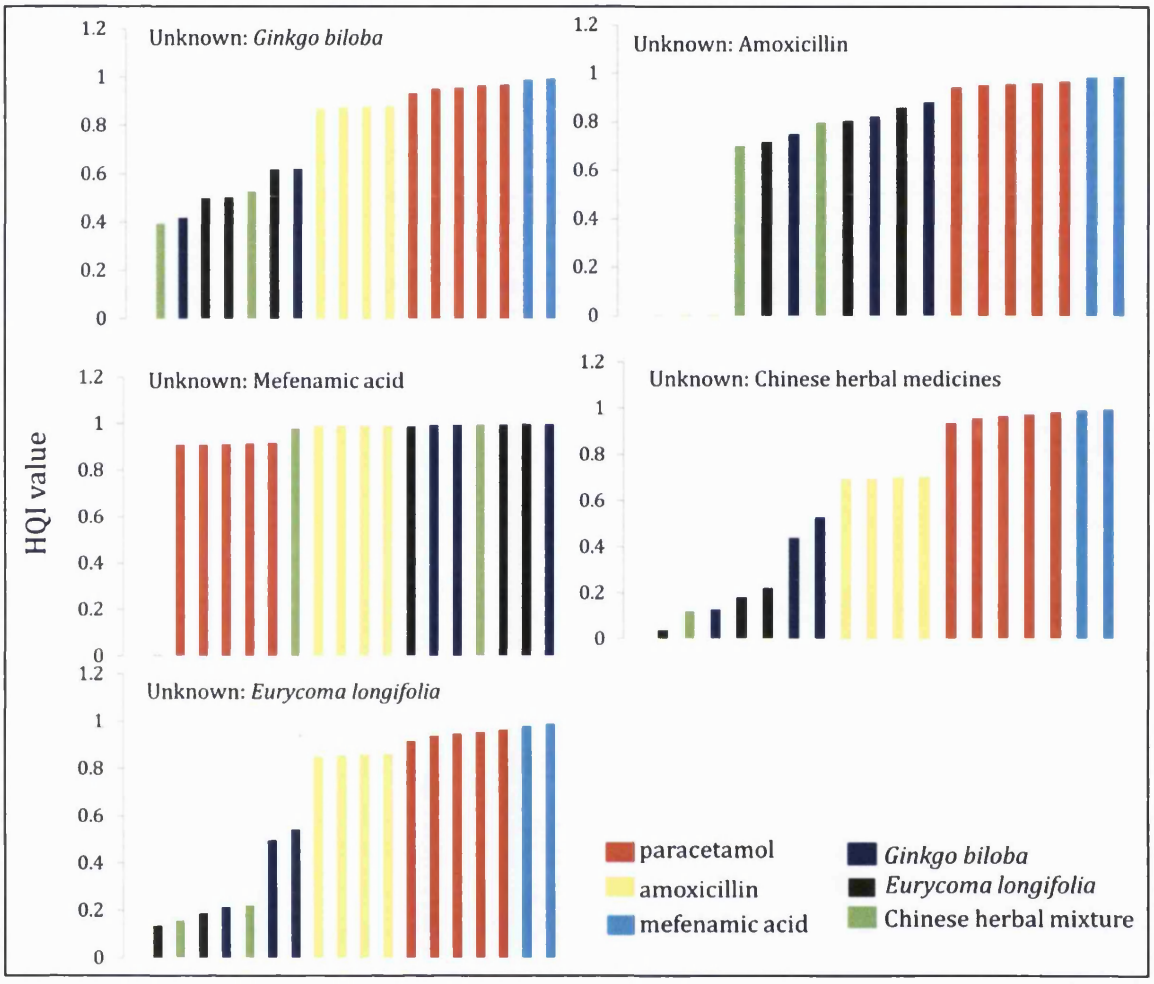


Figure 4.8B The search outcomes for each single type of medicines in a database consisted of the spectra of *Ginkgo biloba*, amoxicillin, mefenamic acid, Chinese herbal mixture, and *Eurycoma longifolia* using first derivative correlation algorithm. The scales are different for each algorithm based on the calculations of individual set.

4.3.2.1(c) Cut-off point determination

An extension to the study was to determine the cut-off values for the search results that classify a sample into one of four categories; matched sample, similar brand/source, similar class/type, or different products. The findings presented in the previous sections were used to propose cut-off values based on these classifications. Table 4.1 shows the outcomes of the paracetamol searches that were used to define the cut-off values. This indicates that it is possible to identify the correct types of medicines and, furthermore, to suggest the manufacturer of the examined samples if a similar product was present in the database.

The cut-off values for the HQI values shown in Table 4.2 are proposed to be suitable to be used for the classification of products with more than 75% of the tablet mass is the APIs. These values can be used for a spectrum search method on single or multi-file averaged spectra inputs.

For other types of products (e.g. potent drugs or complex mixtures such as herbal medicines), new cut-off values need to be determined. These values, however, should be used as a guide and only if there is an indication of what the product may be. These values may overlap, especially for atypical samples, and further work on the refinement of these values is needed as more products from classes are included into the database.

4.3.2.2 Peak Search

As a complementary approach for product identification, the peak search method was evaluated. The internal validation of the peak search methods was conducted using the same sets of unknown samples and the same databases as for the full spectrum search. Two sets of searches were conducted to observe the effect of applying the forward and reverse peak search techniques (Figure 4.9).

Table 4.1 The cut-off values (HQI) for each brand of paracetamol sample search based on the search conducted using the correlation algorithm. The database used was the same as that used in Figure 4.4. Samples marked with * are products of different batches from the reference spectra in the database.

Classification criteria	Unknown samples (paracetamol)								
	MIL_A	FP	OR	BG_A	UP_A	MIL_B*	BG_B*	UP_B*	FP_1*
Match	0	0	0	0	0	-	-	-	-
Similar brands/ sources	-	-	-	-	-	0.0028 (MILA)	0.0027 (BGA)	0.0014 (UPA)	0.004 (FP)
Similar types of medicines (U- upper range L- lower range)	0.011 (L) - 0.053 (U)	0.010 (L) - 0.023 (U)	0.012 (L) - 0.061 (U)	0.017 (L) - 0.061 (U)	0.011 (L) - 0.051 (U)	0.011 (L) - 0.052 (U)	0.025 (L) - 0.066 (U)	0.017 (L) - 0.061 (U)	0.0086 (L) - 0.032 (U)
Different types of medicines	0.77	0.71	0.77	0.63	0.70	0.78	0.62	0.69	0.73

Table 4.2 Proposed cut-off points for pharmaceutical formulation classification based on the degree of similarities and differences between the unknown spectrum and the spectra in the database.

Classification type	Search outcomes	Cut-off values (HQI)
I	Same batch (match)	<0.0001
II	Same brand/different batch	<0.0100
III	Same type of medicine/ different brand	<0.1000
IV	Different type of medicine	>0.2000

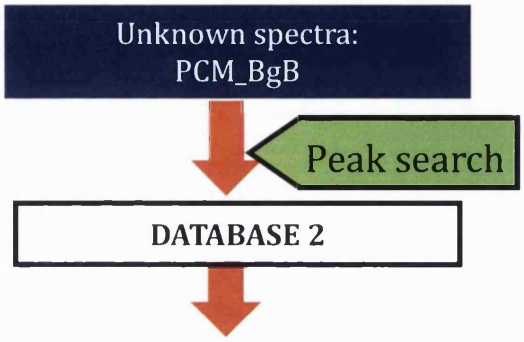
	Samples	A	B
1	Amoxicillin		
2	Paracetamol		
3	Simvastatin		
4	Cefuroxime axetil		
5	Diclofenac sodium		
6	Mefenamic acid		
7	Sibutramine		
8	Ticlopidine		
9	Sildenafil citrate		
10	Macgel		
11	*Ginkgo biloba*		
12	*Eurycoma longifolia*		
13	Chinese herbal 1		
14	Chinese herbal 2		

Figure 4.9 Internal validation of the database using the forward (A) and reverse (B) peak search techniques. Blue cells indicate a match between similar types of medicines to the similar types of spectra in the database. Red cells indicate a pair of spectra that did not meet the identification criteria. Highlighted products indicate the spectra of samples from different batch to the unknown were available in the database.

The table in Figure 4.9 shows that the reverse peak search strategy gives better spectra identification compared with the forward peak search. However, there are several well-known issues on using reverse peak search methods.

Reverse searching was designed to identify compounds in a mixture [50]. A good match can occur when all the peaks in a reference spectrum correspond to the peaks in the unknown spectrum, even though there are more peaks of the unknown that are not found in the reference. However, this approach could lead

to false positive results. Other limitations of this method are the difficulty in identifying the peaks from low concentration components, problems of identifying samples with sample spectra, and the presence of common side chains such as aliphatic hydrocarbon and certain chlorinated compounds [50].



HIT#	Forward peak search	
	HQI	Sample ID
1	31.59722	PCM_MilA
2	31.59722	PCM_OR
3	31.59722	PCM_UpA
4	31.25	PCM_FP
5	26.38889	PCM_BGA
6	18.57639	MA_1

HIT#	Reverse peak search	
	HQI	Sample ID
1	95	PCM_BgA
2	94.79167	PCM_MilA
3	94.79167	PCM_OR
4	94.79167	PCM_UpA
5	93.75	PCM_FP
6	47.32143	MA_1

Figure 4.10 The top six hit-lists for the searches conducted on paracetamol Bg_B in database 2 using forward and reverse peak searches. Both methods managed to produce positive identifications of the unknown but poor classifications between other brands of paracetamol in the database. Three different brand names of paracetamol (encircled) resulted in the same HQI using these search methods.

These problems were demonstrated using the peak search strategy to identify the spectra of paracetamol Bg_B in database 2. For the peak search, a perfect match between the unknown and reference spectra would produce a HQI of 100, while lower HQI values indicate higher dissimilarity between the unknown and the reference spectra. Figure 4.10 shows the top six hit-lists of the searches conducted using the forward and reverse peak search strategies.

Both peak search methods were able to identify the sample as paracetamol but were unable to differentiate between the different brands of paracetamol in the database as observed for the full spectrum search method (Figure 4.7). This is because the search strategy gave a strong match between the common peaks that are available in both the reference and the unknown spectra, while ignoring the rest of the peaks that could be used to differentiate between different brands of paracetamol. Figure 4.11 shows the complete distribution of paracetamol and other samples in the database after the searches.

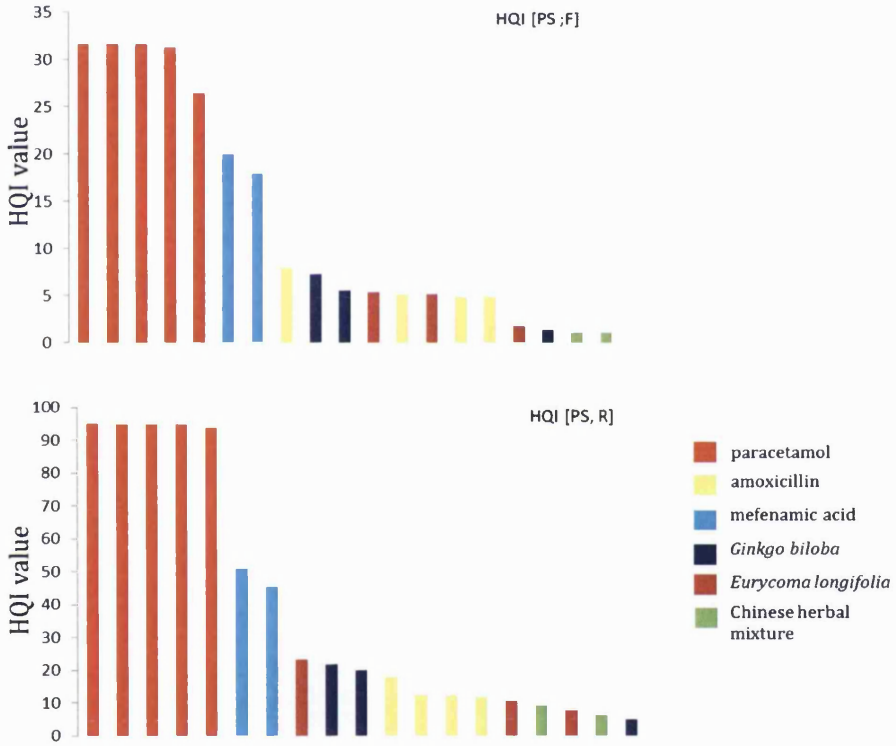


Figure 4.11 The outcomes of the searches conducted on paracetamol Bg_B in database 2 using forward [F] and reverse [R] peak searches. The reverse peak search gave better discrimination between similar types of samples in the database. The scales are different for each algorithm based on the calculations of individual set.

4.3.3 Quality control of the database

Although using the appropriate algorithms is important for ensuring a successful search, the basic quality of spectra also play a major role, even prior to the search being conducted. Lowry *et al.* [50] established a comprehensive quality control

protocol to ensure the quality of the library. This protocol comprised four areas: (i) establishing exact specifications, (ii) setting up a program for instrumentation certification, (iii) certifying sample purity and identification, and (iv) setting up a procedure for spectroscopic review.

Griffiths and Wilkins [136] proposed a three-digit quality index to evaluate the quality of digital IR reference spectra. The three aspects assessed were (i) sample authenticity and purity, (ii) sample preparation, and (iii) instrumental conditions. An algorithm for calculating this quality index was highly instrument-specific and was not discussed further.

Based on these references, a quality control procedure was designed according to the specifications of this research (Table 4.3). Adherence to this procedure is important in order to ensure the quality and consistency of spectra included in the database.

4.3.4 Rationale for the choice of software

Some of the advantages of using the GRAMS Suite Spectral ID Server include:

1. The database can be used to store spectra obtained by analytical techniques including IR, Raman, Mass Spec, UV-VIS, Fluorescence, and NIR. It supports most commercially available libraries including Aldrich, Bio-Rad Sadtler, Chemical Concepts, NIST, Thermo Nicolet, and Wiley.
2. The GRAMS Suite SmartConvert™ technology has the capability to automatically read, open, and search data files obtained by over 150 different analytical instruments.
3. All the databases were stored in a central location, which simplified storing, tracking, updating, and configuring processes.
4. Different search techniques and algorithms are available to choose from based on the suitability for the types of samples and intended purposes.

Table 4.3 The quality control procedure for spectra acquisition using NIR spectrometry.

Area	Specification	Description
(1) Analysis specification	Sample preparation	-Most of the solid samples were analysed intact without any sample preparations. Capsule and liquid samples were prepared in an identical way (section 2.2.1).
	Spectra acquisition	-Spectra acquisition was made in the exact way for each type of sample (sec. 2.2.1).
	Spectral range	-Spectral range was set between 1,100 to 2,500 nm (10,000–4,000 cm^{-1}).
(2) Instrument certification	Performance test	-Performance test of instrument was conducted every day to validate the instrument parameters such as photometric noise and accuracy, bandwidth analysis, and wavelength accuracy. Wavelength linearization checks that the internal reference peak position corresponds to the nominal positions.
	Reference spectra	-Reference using ceramic discs was run every hour during sample acquisition and retained for reference.
	Temperature and humidity	- Temperature and humidity was monitored and noted every morning and evening.
	Reproducibility	- Set of common spectra was run every time after any adjustment of the instrument was made (e.g. update processing software, change of lamp) to ensure spectra reproducibility
(3) Sample verification		-Sample identification and classification was validated by PCA and SIMCA.

A protocol was proposed to develop a customised spectral database of medicines (Figure 4.12) with the aim to directly compare and classify spectra of different pharmaceutical and herbal formulations without having to use reference standards. The database development steps (dotted line) included defining the scope and the intended use of the database, sample collection and preparation, spectra acquisition, and PCA and classification analysis for sample authentication prior to spectra storage in the database. The steps for the search using spectra obtained for the samples of unknown origin (full line) included defining the objective of the search, sample preparation, data acquisition, and data pre-processing.

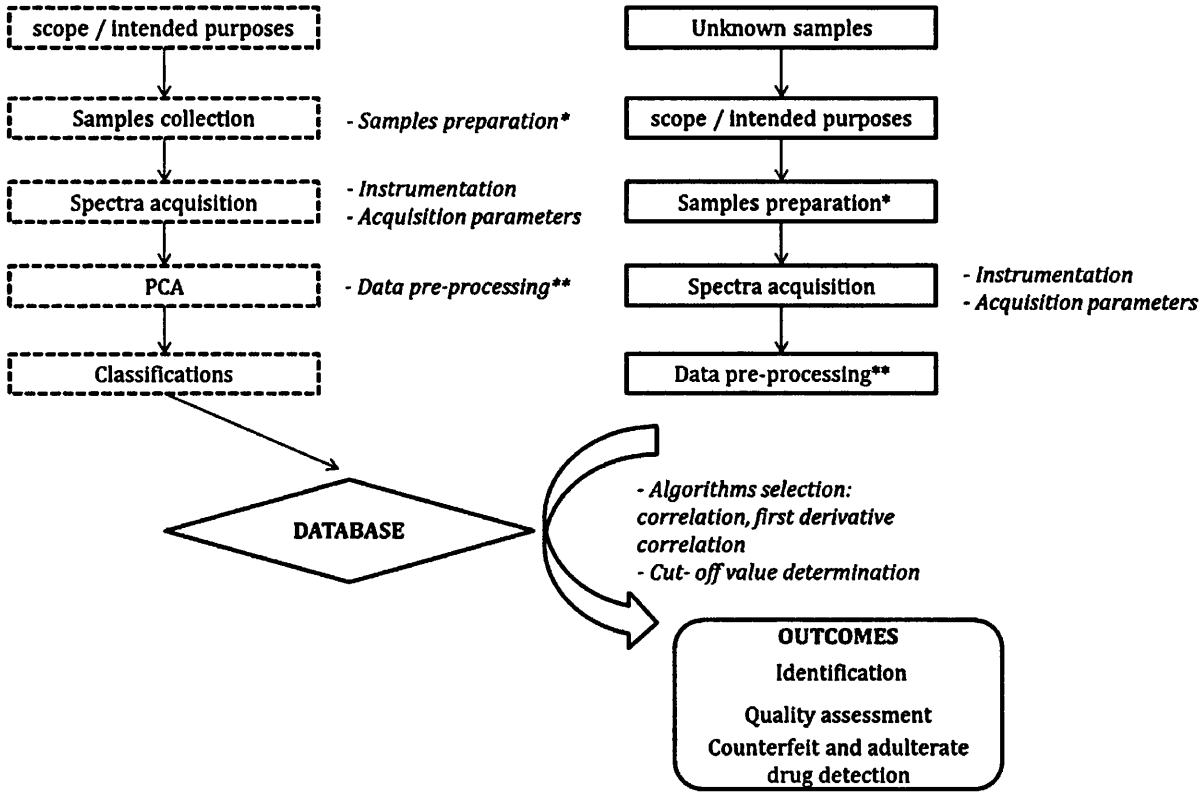


Figure 4.12 The flowchart showing a step-by-step procedure for developing databases and performing searches of unknown products against the database. The two main steps in the procedure were database development (dotted line) and identifying unknown spectra (full line).

4.4 Conclusion

The database analysis above has shown that using the full spectrum search method gave superior outcomes compared with the peak search method. Among the seven algorithms, correlation algorithms were the most suitable for both identification and classification purposes. In a real-life analysis, a combination of both search techniques is believed to be preferable because most information about the unknown samples would then be obtained.

A protocol was proposed for building a spectral database and its consequent use for sample identification and classification. It was established that all steps in the process carried out for the acquisition of the spectra of an unknown sample have to be similar to the steps for the samples already stored in the database (emphasised in italics). The determination of the right algorithm and cut-off value is important before the search begins. The spectra of unknown samples can be used to query the database and possibly classify the product if a similar formulation exists in the database. The new spectra and accompanied information are also stored in the database in order to enhance the results of future searches.

The quality of spectra in the database is another major factor in determining the success of spectral searches. For the proposed database technique, the quality control strategies are based on three important elements: sample, instrument, and analysis. These were implemented throughout the research in order to produce spectra of consistent quality and parameters for the database.

Paying enough attention to spectra quality during the initial development of the database is crucial to ensure the validity and reproducibility of the search outcomes. Although it was not discussed in this work, transferability between different instruments has to be taken into consideration when building a database to be used by agencies.

CHAPTER 5

IDENTIFICATION OF SPECTRA OF UNKNOWN SAMPLES

5.1 Introduction

It is crucial to ensure that the pharmaceutical products received by consumers are of good quality and safe and effective. The WHO recommends that each country should have a "central coordinating body with overall responsibility and accountability for all aspects of drug regulation for the entire country" [137]. In Malaysia, the Drug Control Authority (DCA) is responsible for this. Some of the DCA's duties include reviewing registration applications for drugs and cosmetics, licensing importers, manufacturers and wholesalers, post-marketing safety surveillance, and adverse drug reaction (ADR) monitoring [138].

Despite having strict regulations and the tough implementation of rules and law within the industry, the distribution of low quality pharmaceutical products on the market is still difficult to control. Poor quality products include substandard, adulterated and counterfeit medicines.

The illegal influx of medicines from different sources into the healthcare systems of developing countries presents the main challenge in maintaining the quality of medicines. This situation is further complicated by the dubious quality of many products sold over the internet, particularly herbal preparations. The surge of generic products also requires the continuous monitoring of product quality and bioavailability. For example, a recent search carried out on the Malaysian DCA website showed that there were more than 250 registered products containing paracetamol as the main or one of the main active pharmaceutical ingredients (APIs) available on the market [139].

The uses of the spectral database in combination with principal component analysis (PCA) for the quick identification of unknown sample and similarity assessment of products are shown in this work. The procedure is described in a flowchart in Figure 5.1.

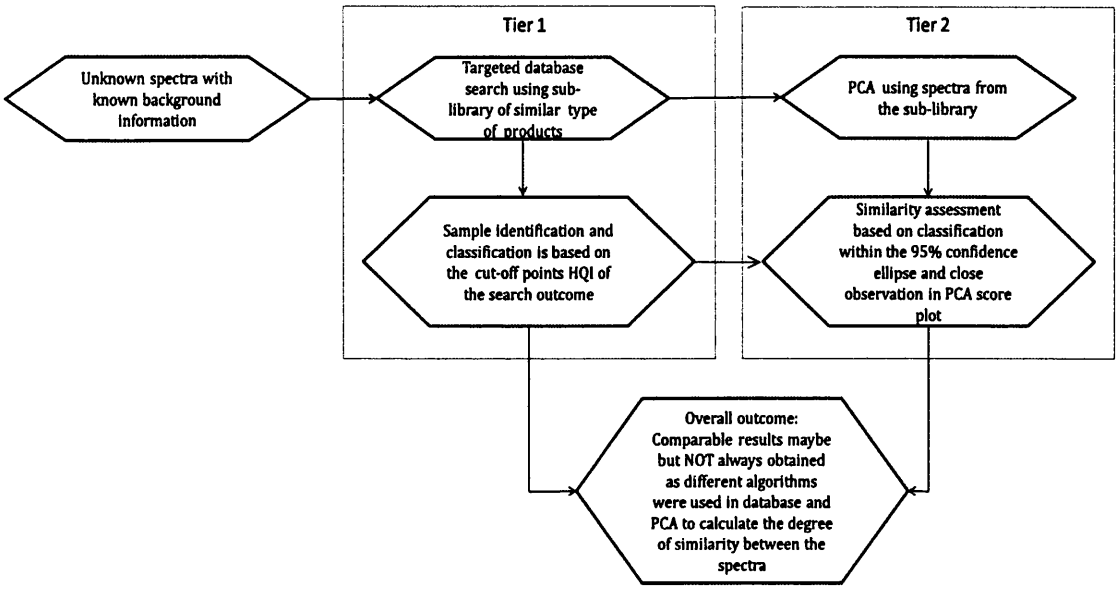


Figure 5.1 The flowchart showing the procedures involved in identification of an unknown sample using the two-tier screening method. Results from the database (Tier 1) and PCA (Tier 2) may or may not corresponding to each other since these applications use different algorithms to measure the degree of similarity between spectra. Similar findings in both methods validate the search outcome while two different findings will indicate the need for further analysis of the unknown sample.

In part one, qualitative analysis of different brands of paracetamol tablets purchased in pharmacies and supermarkets in Malaysia was used to demonstrate the proof of principle. Reproducibility study of the method was conducted by using two test sets: a set of different paracetamol samples acquired at the same time as the training set and an additional, new set of paracetamol samples acquired a year later. PCA was used to observe the distribution of samples in the score plot and its correlation with the database search outcomes.

In part two, the same procedure was used to observe the authenticity of amoxicillin drugs dispensed at local pharmacies in three main states within Malaysia; Kuala Lumpur, Johor Bahru and Penang. Both NIR and NMR spectra of this set of samples were acquired and compared in PCA.

5.2 Materials and methods

5.2.1 Part 1: NIRS and chemometric analysis of Malaysian and UK paracetamol tablets - a spectra database study

5.2.1.1 Drug samples

The sample set used in this study consisted of 16 of 500 mg paracetamol tablets (except for one batch with 650 mg tablets) purchased in Malaysia. Each batch contained 20 tablets and paracetamol was the only active ingredient. Three other products that either; contained other active ingredients in addition to paracetamol or did not contain paracetamol at all; were included in the test set as negative controls. After a year, a new set of tablets consisting of three new batches of the brands available in the database and one additional batch of paracetamol from a different manufacturer were purchased from the Malaysian market and their NIR spectra were recorded. These additional spectra were added to the database as test sets for identification purposes. Details of the samples are listed in Table 5.1.

All of the samples were purchased from pharmacies and supermarkets since attempts to obtain them directly from manufacturers were either ignored or rejected (for the Malaysian samples). The samples consisted of uncoated tablets of different shapes and colours and supplied either in the original blister packaging or as repackaged products

5.2.1.2 NIRS Analysis

The NIR diffuse reflectance analyses were carried out using a NIRSystems 6500 spectrophotometer, equipped with a Rapid Content Analyzer (FOSS NIRSystems, Silver Springs, USA). Sample preparation and analysis was described in section 2.2.1.1, page 94).

The original averaged NIR spectrawere converted into ASCII and then into JCAM-DX format using an in-house program. The spectral files were then imported into Unscrambler v9.7 software (CAMO, Oslo). The spectral database was created using GRAMS ID software (GRAMS Suite).

5.2.1.3 Principal Component Analysis (PCA)

The PCA treatment of the spectra decomposed them into scores and loadings for variables called principal components (PC). The main purpose of the PCA was to reduce the number of variables in order to represent a multivariate data table in a low dimensional space [140]. The PCA was conducted on the data with leverage correction as the validation method and the scaling factor was set as 1.

Table 5.1 Details of the samples used for the product assessment, database development and validation discussed in this work. A set of 20 tablets were analysed for each batch.

Purposes	Active pharmaceutical ingredients (APIs)	Amount (batch/es)	Source	Sample label
Calibration samples	Paracetamol 500 mg	15	Malaysia	BG_A, BG_B, BG_C, FP, IF, MIL_A, MIL_B, OR, PC, PG, PR, PM, PT, UP_A, UP_B
PCM 650	Paracetamol 650 mg	1	Malaysia	UP_C
Negative control	Paracetamol 500 mg, Dihydrocodeine tartrate 7.46 mg	1	UK	NCI (paramol)
	Paracetamol 200 mg, aspirin 300 mg, caffeine 45 mg	1	UK	NCII (epr)
	ibuprofen 200 mg	1	UK	NCIII (ibu)
Internal validation	Paracetamol 500 mg (internal)	3	Malaysia	IF, PM, UP_A
External validation	Paracetamol 650 mg (external 1)	3	Malaysia	IF, PM, UP_A
Unknown	Paracetamol unknown (external 2)	1	Malaysia	pcm_unknown (N)

5.2.1.4 Database Creation and Search

Raw NIR spectra were directly imported into Spectral ID v9.0 (part of the GRAMS spectroscopy software Suite, ThermoFisher Scientific Inc.) using the NSAS format. The database has smart-convert functions that automatically convert the NSAS

file to .SPC files. Other information computed together with the spectra included the brand and proprietary names, batch number, expiry date, manufacturer names and addresses, sample origin, other excipients (where available) and sample description.

Before each search, the unknown spectra and other spectra in the database were set for baseline correction to reduce the scattering effects that highly contaminate NIR spectra. This was done using the auto-baseline correction algorithm, which removes linear baseline errors of positive-going peak data (GRAMS).

A database search was conducted on the whole spectrum based on a correlation. A hit-list of the top matches was provided together with hit quality index (HQI) values. Low HQI values measured between the unknown spectra and the spectra in the library indicated a good match.

5.2.2 Part 2: Chemometric and spectra database analysis of amoxicillin tablets purchased randomly from the Malaysian market.

5.2.2.1 Samples and sampling

Samples of amoxicillin formulations were obtained from local pharmacies in three states of Malaysia; Kuala Lumpur, Johor Bahru and Penang (Figure 2.1). The researcher has purchased samples as a local buyer who complained of prolonged fever for more than a week. No ethical approval required as inform consent from the seller is not necessary in drug quality assessment study. Amoxicillin was requested from the pharmacist in 30 pharmacies through convenience sampling: 13 different sets of 500 mg and eight different sets of 250 mg and one set of 400 mg doses were dispensed and labelled accordingly (Table 5.2). Sample set E was in a tablet form while the rest of the samples were powders in capsules. Some of the samples were packaged in blisters while some were dispensed loose in plastic packaging labelled with only the generic name of the product.

Table 5.2 The amoxicillin samples purchased through convenience sampling in 30 pharmacies in Kuala Lumpur, Johor Bahru and Penang.

Sample label	Dosage	Manufacturer	Sample label	Dosage	Manufacturer
Amox A	500 mg	Hovid Pharm.	Amox Q	500 mg	Unlabelled (repackaged)
Amox B	500 mg	Dynapharm	Amox R	500 mg	CCM Pharm.
Amox C	500 mg	CCM Pharm.	Amox A2	250 mg	CCM Pharm.
Amox D	500 mg	Unlabelled (repackaged)	Amox B2	250 mg	Unlabelled (repackaged)
Amox E	500 mg	Sandoz	Amox C2	400 mg	Unlabelled (repackaged)
Amox G	500 mg	Hovid Pharm.	Amox D2	250 mg	Unlabelled (repackaged)
Amox H	500 mg	Hovid Pharm.	Amox E2	250 mg	Unlabelled (repackaged)
Amox I	500 mg	PMB	Amox F2	250 mg	Unlabelled (repackaged)
Amox M	500 mg	Unlabelled (repackaged)	Amox G2	250 mg	Unlabelled (repackaged)
Amox N	500 mg	Unlabelled (repackaged)	Amox H2	250 mg	Unlabelled (repackaged)
Amox P	500 mg	YSP Industries	Amox I2	250 mg	Unlabelled (repackaged)

5.2.2.2 NMR analysis

The tablet samples were crushed to fine powder and the capsules were opened up to obtain the drug powder. 10 mg of drug powder from each of the samples was dissolved in 1 ml of deuterated methanol with 0.5% tetramethylsilane (TMS) solution by agitation in a whirl mixer, followed by 15 minutes sonication and 5 minutes centrifugation at 13,000 rpm. Following this, 650 µl of clear solution was transferred into NMR tubes for ^{1}H-NMR analysis.

Each data set was scaled to the internal standard, phased and baseline corrected. After pre-treatment, the spectra were transferred to Amix-Viewer v3.5 (Bruker BioSpin). To reduce the number of data points and remove noise-related alternations, the "bucketing" function was used to generate a set number of integrated regions or "bins" of the data set [141]. The spectra were integrated based on the sum of intensities and three regions were excluded; -0.1 to 0.1, 3.2

to 3.4 and 4.5 to 5.0, which were regions that contained TMS, methanol and water peaks, respectively. The "binned" data generated from the original spectra were then exported as a spread-sheet suitable for importing into the statistical analysis software, The Unscrambler (CAMO).

5.2.2.3 NIRS analysis

Only the samples with 500 mg of the active ingredient and in a capsule dosage form were included for NIR analysis. Samples preparation and analysis was conducted as described in section 2.2.1.2, page 95).

5.2.2.4 Database construction and analysis

A library of the raw NIRS spectra of the amoxicillin samples was created. All samples were included in the database by using a similar procedure as discussed in section 5.2.1.4, page 155. A full search of the spectra was conducted on samples AmoxC2 and AmoxH2 by using the correlation algorithm.

5.3 Results and discussion – Part 1

The first part of this work focussed on obtaining the spectral fingerprints of each set of samples and observing the product variations in terms of raw NIR spectra and PCA. This information was correlated with the search outcomes in the database.

5.3.1 NIR Analysis

All of the tablets contained a 75%-90% mass fraction of paracetamol relative to the tablet weight and hence it was expected that most of the peaks in the NIR spectra would be from paracetamol. The raw NIR spectral analyses indicated that all of the samples produced a similar spectroscopic fingerprint of paracetamol (Figure 5.2A), whereas they were different compared to the negative control samples (Figure 5.2B). The main difference between the spectra was the absorbance shift concerning the whole spectral range that corresponded to scattering effects and physical properties (e.g. particle size, shape or surface structure) of the samples. Moreover, a clear difference in absorbance over the whole range of wavelengths could be observed for one of the samples, PM. One of

the negative control samples, NC1, had some degree of similarity with the other paracetamol spectra due to a similar composition of the main active ingredient.

5.3.2 Principal Component Analysis (PCA)

The PCA was employed to validate the spectral interpretations and to provide more information about the chemical differences between the batches. These results indicated that the types of ingredients but not the amounts of active ingredients influenced the classification. One batch from the Malaysian samples, identified as sample PM, was not included in the 95% confidence interval set, indicating significantly different properties of the tablets in that batch, which may indicate different excipients or a different manufacturing process (Figure 5.3). The comparison of NMR spectra from this sample and the other paracetamol samples indicated that there was no significant difference in the chemical composition of the tablets that could lead to such difference in clustering of the NIR spectra (Appendix 3).

The PCA score plot in Figure 5.4 shows that all of the negative control samples were clearly excluded, whereas the 650 mg sample (UP_C) was well clustered with the other samples.

Inclusion of the new samples changed the PCA score plot, with a different clustering of the NIR spectra (PM and UP_C became well clustered with the rest of paracetamol tablets). The confidence ellipse expanded to accommodate new samples and two negative controls samples formed clusters away from the rest of the samples. The NC1 sample, which had 500 mg of paracetamol as the main active ingredient in addition to other active ingredients, was included in the 95% ellipse together with the paracetamol sample. However, its cluster was well separated from the rest of the group.

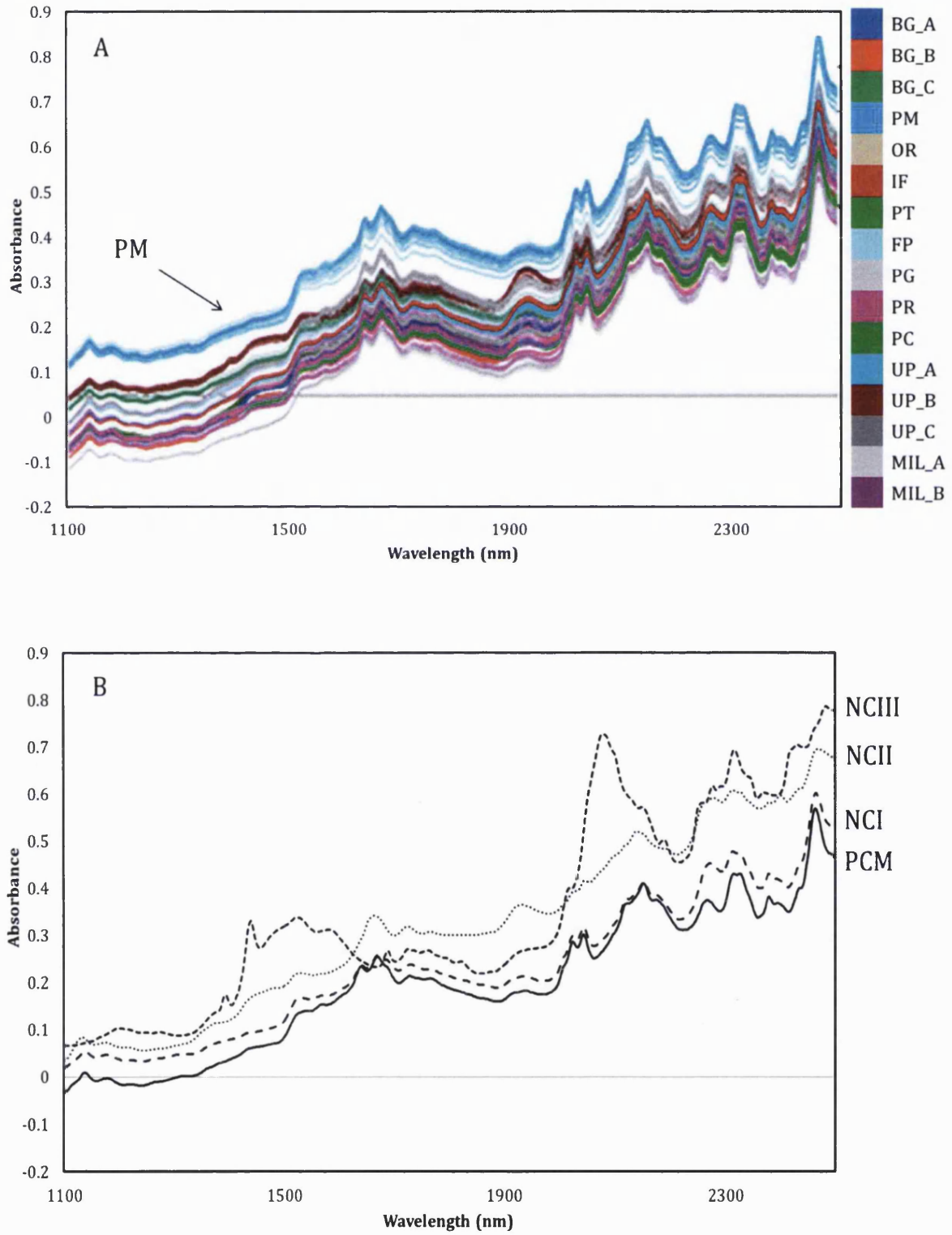


Figure 5.2 Mean NIR spectra in the reflectance mode of (A) 22 batches of paracetamol samples and (B) the representative paracetamol spectrum (PCM) compared to the negative control spectra (NCI, NCII and NCIII).

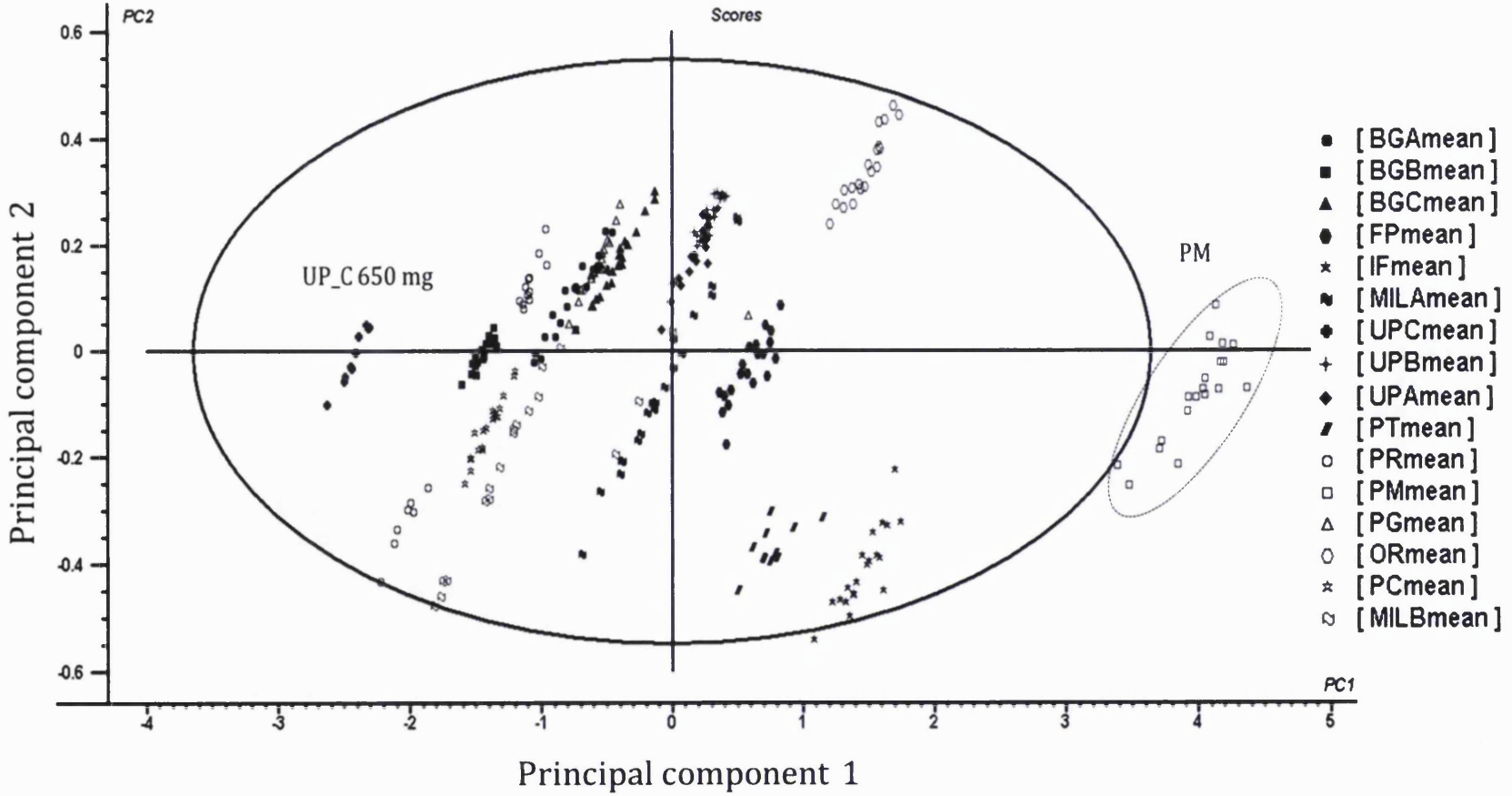


Figure 5.3 PC scores and 95% confidence interval ellipse for the raw reflectance spectra of 16 Malaysian paracetamol samples in 2PC (PC1=97% PC2=2%). PM is a set of samples that were significantly different from the rest of the samples. UP_C was the only sample with 650 mg paracetamol as the API.

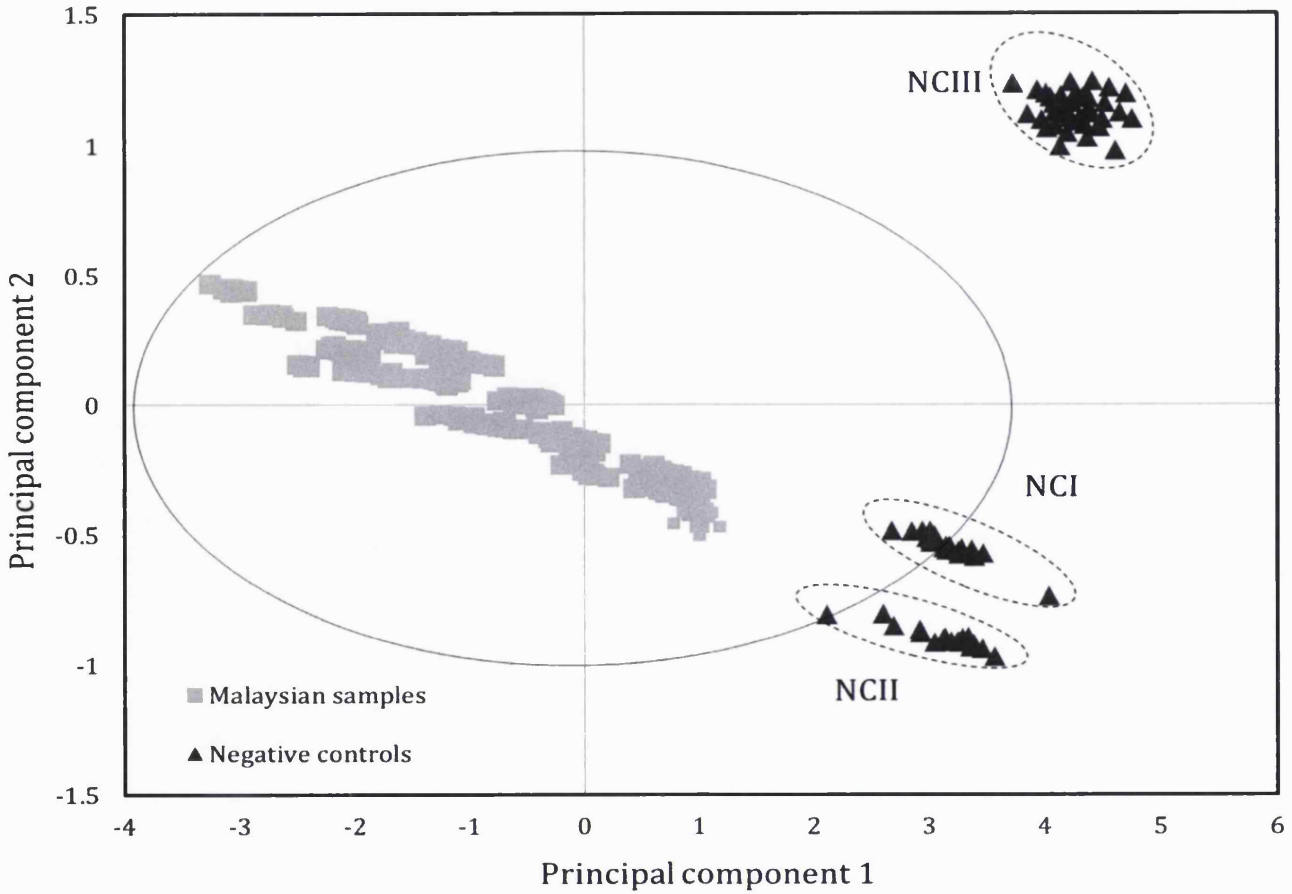


Figure 5.4 PCA scores plot with 95% confidence interval ellipse for the Malaysian paracetamol samples with the negative control samples (NCI=paracetamol +dihydrocodeine tartrate, NCII=paracetamol +aspirin +caffeine and NCIII=ibuprofen).

5.3.3 Database Construction

5.3.3.1 Database Library

All spectra were imported into a paracetamol database to consider the use for classification. In the database, all of the X-axes of the spectra represented the spectral wavelength in nanometres in the range of 1100 to 2500 and the Y-axes formed the spectral absorbance values. For each brand of paracetamol, 20 averaged spectral profiles with 700 individual points in each were combined as a multi-file in order to enhance the accuracy of the spectral search. The spectra were saved at a 32-bit resolution.

5.3.3.2 Database search

The spectra of selected samples were initially added to the paracetamol database. This set of spectra is known as the reference spectra. Two additional samples, namely BG_B and UP_B, were searched against this database using a correlation algorithm on the full spectra. These were used as the test or unknown spectra which were not initially included in the database. These samples were similar to reference samples, but from different batches. The results of these searches are presented in the hit-list form in Table 5.3.

Using the database, both of the unknown spectra were found to be similar to samples from corresponding manufacturers whose spectra were already added into the database. However, the HQI values were different due to product variations. The batch-to-batch variation was a more prominent source of variability compared to other factors such as sample positioning or time of analysis [142]. While this type of variability can be controlled by including spectra from different batches in the database to "normalize" sample variability, it also provides the advantage of identifying samples and their manufacturer.

The UP_B samples were classified as belonging to one of the similar products came from different batch in the database, UP_A (type II classification) and to another product, UP_C; hit#13 (type III classification). This was because this set of samples had a different dosage (650mg). This showed that the database was capable of distinguishing different doses of the same product.

The HQI values for the products of hit #14 in both hit-lists were high and indicated that these products may belong to different products or different

classes of medicines. This finding corresponded well with the earlier PCA observation (Figure 5.3) whereby sample PM in Table 5.3A was excluded from the 95% confidence interval ellipse, while the paramol sample in Table5.3B, was one of the negative control (NCI) which consisted of 500 mg paracetamol and 7.46 mg dihydrocodeine tartrate.

Table 5.3 The search outcomes for samples (A) BG_B was classified as similar brand/source to BG_C (HQI=0); BG_A (HQI=0.0038) and (B) UP_B was classified as similar to UP_A (HQI: 0.0009).

A

Hit #	HQI	Sample ID.
1	0	BG_C
2	0.0038	BG_A
3	0.0229	FP
4	0.0247	UP_A
5	0.0385	PT
6	0.0471	PC
7	0.0509	PR
8	0.0553	OR
9	0.0604	UP_C
10	0.0643	MIL_A
11	0.0667	PG
12	0.0669	MIL_B
13	0.0781	IF
14	0.1610	PM

B

Hit #	HQI	Sample ID.
1	0.0009	UP_A
2	0.0095	PT
3	0.0131	PC
4	0.0148	BG_A
5	0.0157	PR
6	0.0216	BG_C
7	0.0255	FP
8	0.027	PG
9	0.0287	UP_C
10	0.0313	MIL_B
11	0.0350	MIL_A
12	0.0392	IF
13	0.0460	OR
14	0.1783	Paramol (NC1)

5.3.4 Internal/External Validation

Internal and external validations were conducted to investigate the reproducibility and reliability of the database search. Three sets of samples from the initial sample set were re-analysed after one year and was considered as internal validation. Additionally, a set of three new samples of brands that were similar to the products with spectra in the database were obtained from the market (external validation 1). A set of new brand samples (pcm_unknown) was also subjected to a database search (external validation 2).

The spectra of all of the above samples were acquired and searched against the existing paracetamol database 12 months after the database was created and the initial study was conducted. During this period, the NIR instrument was serviced once and the processing software (VISION, Foss) was upgraded to a newer version.

The PCA distribution of the spectra of the validation sets with the other spectra available in the database can be seen in Figure 5.5. All of the new validation samples fell within the 95% confidence interval in the distribution range of the available spectra. However, the new sample (N) was excluded from the ellipse. This sample was actually paracetamol with a different dosage form; that was paracetamol granules in capsules.

Ideally, the internal validation should give a perfect match to the spectra in the database as similar samples were used to test the database. However, product degradation and instrument variations after servicing were two possible reasons for the variation in spectra between these samples. The external validation could have been affected by product variation, instrument variation, and the time of analysis.

Figure 5.6 A and B show the expansion of the PCA plot of the validation spectra compared to the original spectra. The validation spectra showed a shift to the left from the original spectra for all samples. This general shift in the spectra was probably due to the instrument variations before and after servicing. This raised the issue of spectra transferability, and it is possible to overcome this problem by fitting appropriate mathematical algorithms. Smith et al. [143, 144] successfully

demonstrated two methods for transferring the reflectance spectra of intact paracetamol between two instruments; one was by using a response surface calculated between the reflectance values of a set of six certified photometric standards measured on both instruments, with the full range of partial least squares (PLS) regression models subsequently being transferred. The second was correction of the spectra from the second instrument by utilizing the residual spectrum between the mean samples of the validation set measured on two instruments.

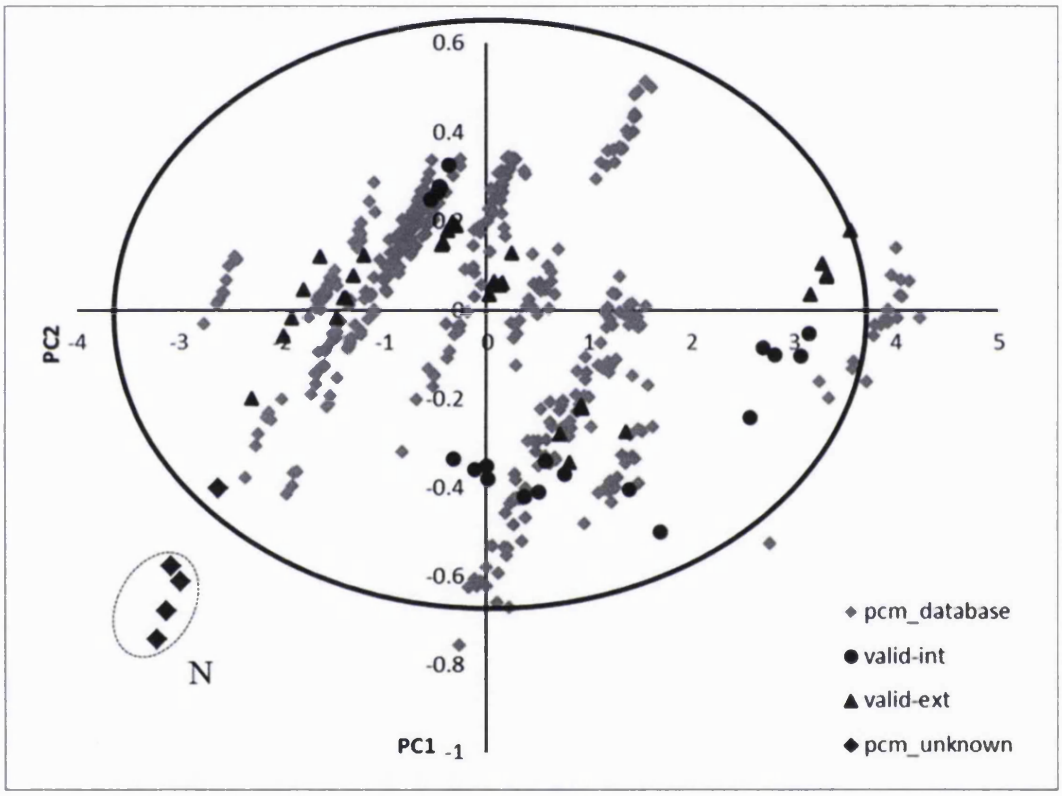


Figure 5.5 PC scores and 95% confidence interval ellipse to describe the distribution of NIR reflectance spectra of the samples used to create the library (pcm_database), the internal (valid-int) and external (valid-ext) validation samples, and the unknown sample (pcm_unknown). N is an unknown brand of paracetamol newly added into the database.

The representative spectra of IF, PM and UP_A that were used for the original and the internal and external validations are presented in Figure 5.7. Despite these drawbacks, the database searches identified all of the internal and external samples as paracetamol products and classified them as samples belonging to the

same source or similar brand names (type II). The HQI values are listed in Table 5.4.

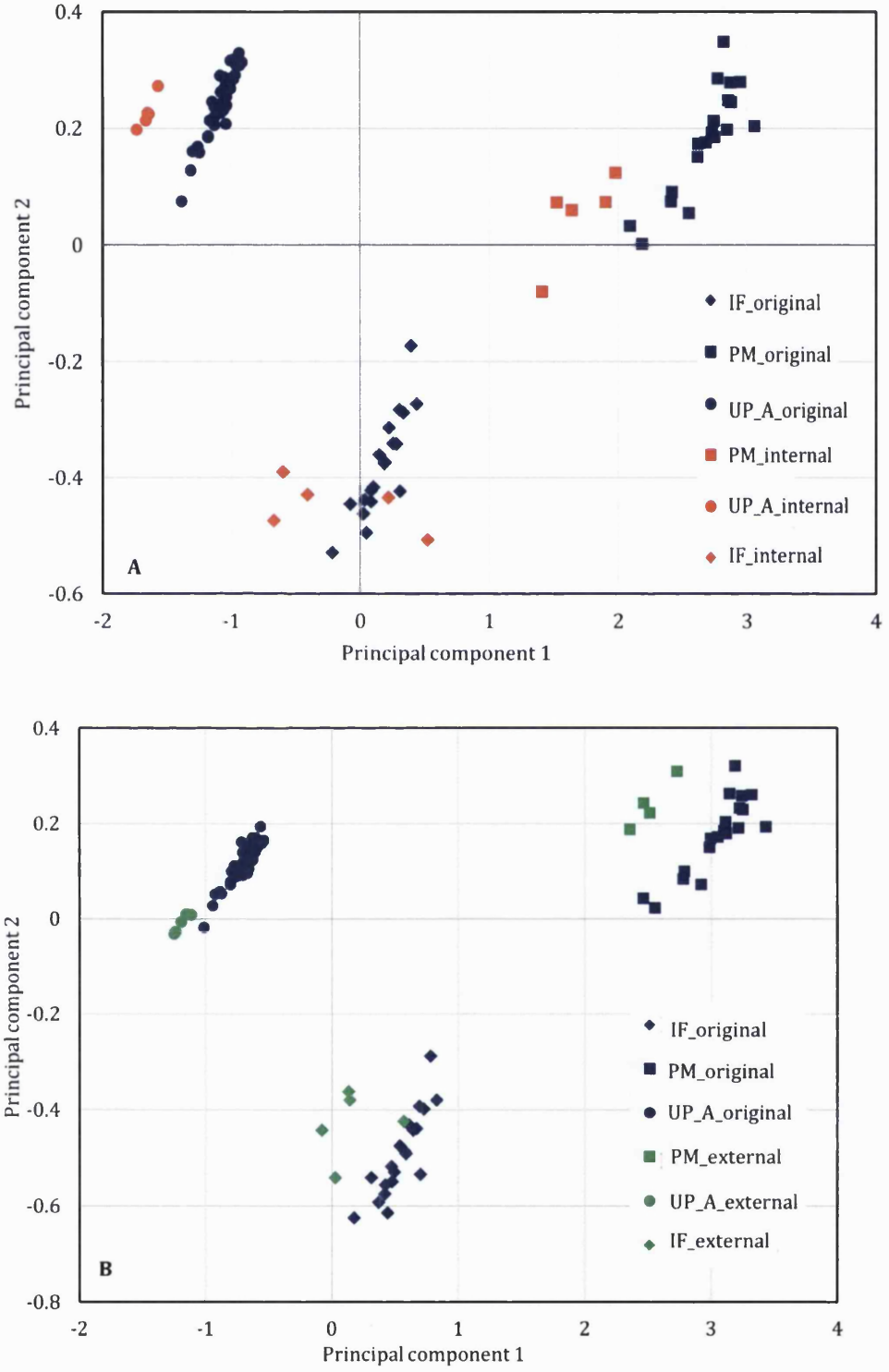
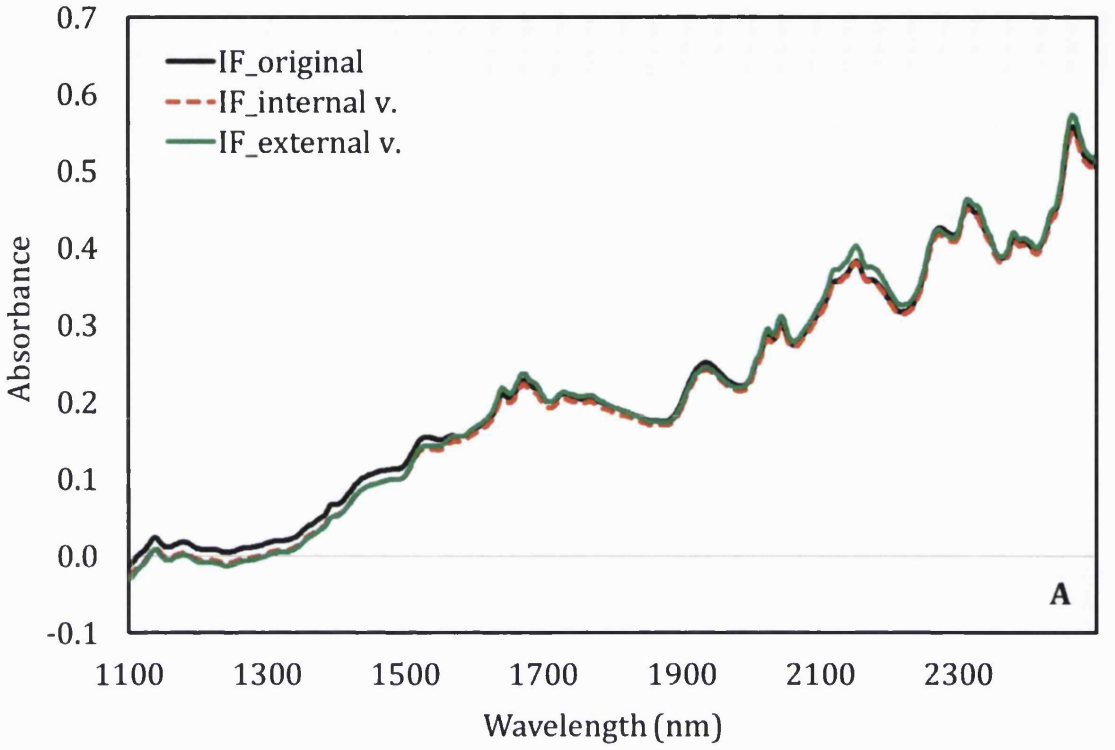


Figure 5.6 Expansion of the PCA plot for (A) internal validation spectra represented by points in red and (B) external validation spectra represented by points in green; compared to the original spectra.



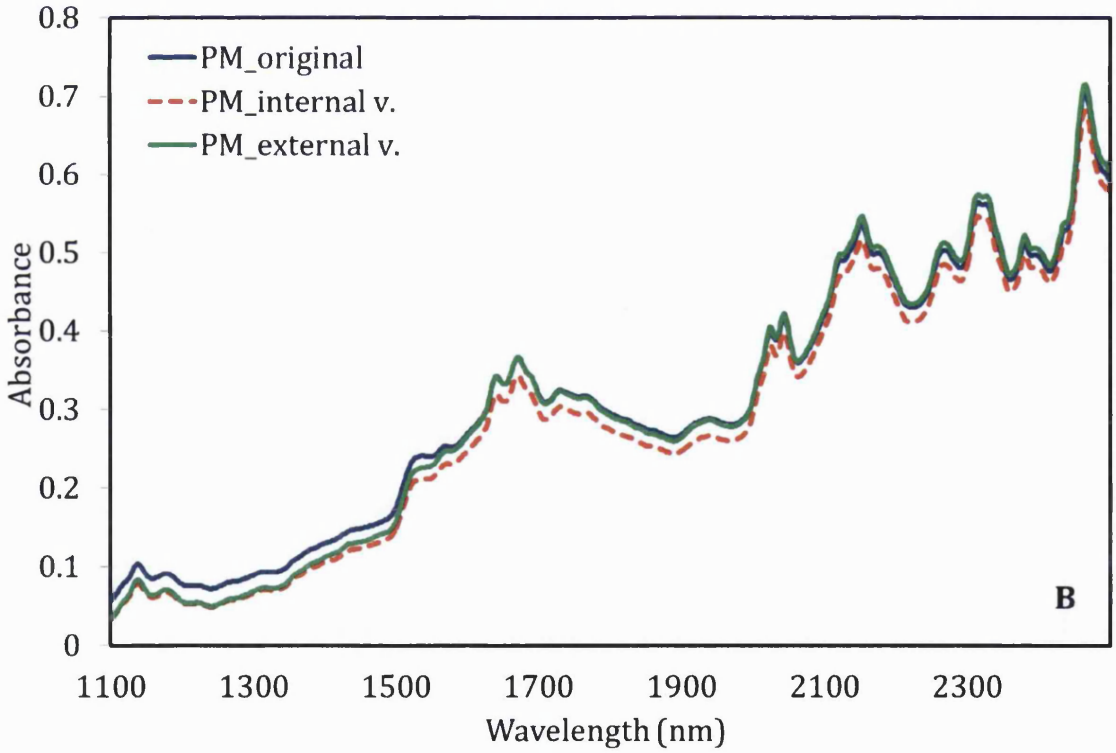


Figure 5.7 Comparison of the linear plots for the original, internal and external validation of the raw spectra of samples (A) IF and (B) PM.

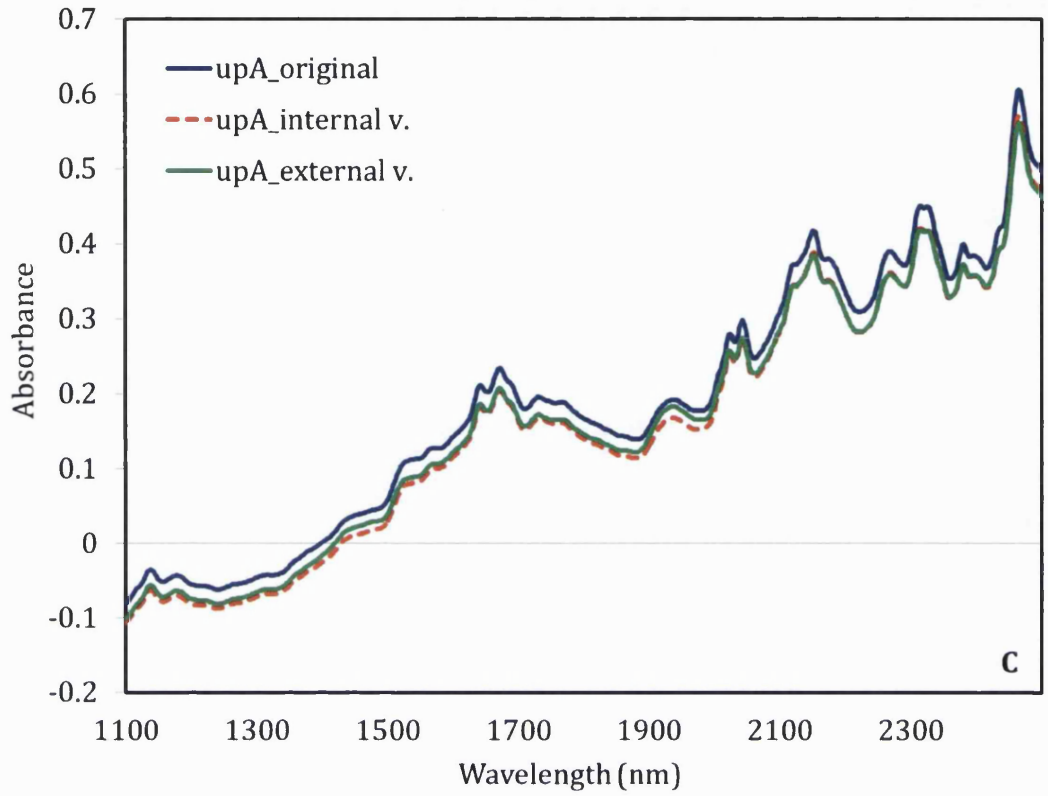


Figure 5.7C Comparison of the linear plots for the original, internal and external validation of the raw spectra of sample UP_A.

Table 5.4 The HQI for the internal and external validation sets of three brands of paracetamol available in the database. The analysis was conducted 12 months after the database was created.

Samples	HQI	
	Internal validation	External validation
UP_A	0.0023	0.0109
IF	0.0163	0.0336
PM	0.0008	0.0003

5.4 Results and discussion – Part 2

Although a physician's prescription is required by law to purchase any antibiotic, many outlets did not refuse to sell this medication upon its request without a prescription. Only two outlets declined to sell without prescription and one claimed not to have the medicine in stock.

The ill practice of dispensing antibiotics among doctors and pharmacies are quite common because of lax restrictions and enforcement. This has caused antibiotics to be sold over-the-counter to people who can afford them and use them as they find fit, which can lead to under- or overdosing themselves [145]. Overuse of this class of medicines will lead to drug resistance in disease-causing bacteria and thus undermine effectiveness of antibiotics.

5.4.1 The database search

The NIR spectra of all the samples were acquired and stored in the sub-database library of amoxicillin (reference spectra). A full spectrum search conducted on two samples, Amox C and Amox H, gave a match to similar reference spectra in the database followed by another match with the spectra of samples from the same brand name or manufacturer (Table 5.5 A and B). For both analyses, the spectra that belonged to Amox C2 were identified last, for which the HQI of this sample was very high (HQI= 0.9844). The cut-off value (HQI) to classify the sample as similar class or type of medicines was less than 0.1 as determined in earlier works (chapter 4, page 144).

This finding gave an indication that Amox C2 was not similar to the rest of the reference spectra in the database. Further multivariate analysis using NMR and NIR spectra were conducted to see the classification of this sample in PCA score plot.

Table 5.5 The search outcomes for samples (A) Amox C and (B) Amox H, showing a match to similar spectra in the database.

A

HQI	Sample ID	Manufacturer
0	Amox C	CCM Pharm.
0.0025153	Amox R	CCM Pharm.
0.0101471	Amox G	Hovid Pharm.
0.010824	Amox N	unknown
0.0114963	Amox Q	unknown
0.015749	Amox I	PMB
0.0168298	Amox D	unknown
0.0219945	Amox H	Hovid Pharm.
0.0316764	Amox M	unknown
0.0387016	Amox A	Hovid Pharm.
0.0444542	Amox E	Sandoz
0.0694963	Amox P	YSP ind.
0.0878048	Amox B	Dynapharm
0.9844023	Amox C2	PMB

B

HQI	Sample ID	Manufacturer
0	Amox H	Hovid Pharm.
0.005117	Amox D	unknown
0.0054111	Amox G	Hovid Pharm.
0.0061078	Amox I	PMB
0.0072581	Amox M	unknown
0.0095273	Amox A	Hovid Pharm.
0.0099787	Amox Q	unknown
0.0101312	Amox N	unknown
0.0193458	Amox P	YSP ind.
0.0219945	Amox C	CCM Pharm.
0.0284897	Amox R	CCM Pharm.
0.0304064	Amox E	Sandoz
0.0304679	Amox B	Dynapharm
0.9733783	Amox C2	PMB

5.4.2 Spectroscopic analysis

PC analysis of the original NIR spectra from the single database of amoxicillin shows a clear separation between samples Amox C2 from the rest of the spectra in the database (Figure 5.8). However, both were included in the 95% confidence interval ellipse.

Figure 5.9 shows the PCA analysis of the ^{1}H 1D NMR spectra of all of the samples collected after spectra pre-processing by Amix software. Three replicates of each sample were prepared for analysis. The PC analysis showed the exclusion of two sample groups from the 95% confidence ellipse: all three samples of C2 were excluded while only one of the three R5 samples did not cluster together with the rest of the samples. Thus, the misclassification of R5 was considered as a gross error and not as an outlier.

Both the PCA score plot for NIR and NMR spectra produced a similar pattern whereby the spectra of Amox C2 were excluded to the right of the rest of the samples (Figure 5.8 and Figure 5.9). PCA using NMR spectra was more specific; according to the peaks identified in each bucket; compared to PCA using NIR spectra which based on the principal component that represented the maximum variation in the set of spectra. Different specificity of the spectra used according to the spectroscopic techniques has led this sample to be included in the 95% confidence ellipse when analysed using NIR spectra and excluded when NMR spectra were used.

A closer look at the single spectrum of each sample (Figure 5.10) indicated that the NMR spectrum of C2 (red line) was different from the spectra of the rest of the samples. The C2 sample was identified as becampicillin based on a comparison with results in the literature [146]. Becampicillin is a prodrug of amoxicillin with improved bioavailability [147] and both drugs have high similarity of the main structure (Figure 5.11).

Due to the similarity of the mode of action and side effects of both drugs, the prescriber would have thought that there would be no harm in giving becampicillin instead of amoxicillin. However, the patient should have been informed regarding this replacement, but that was not done in this case.

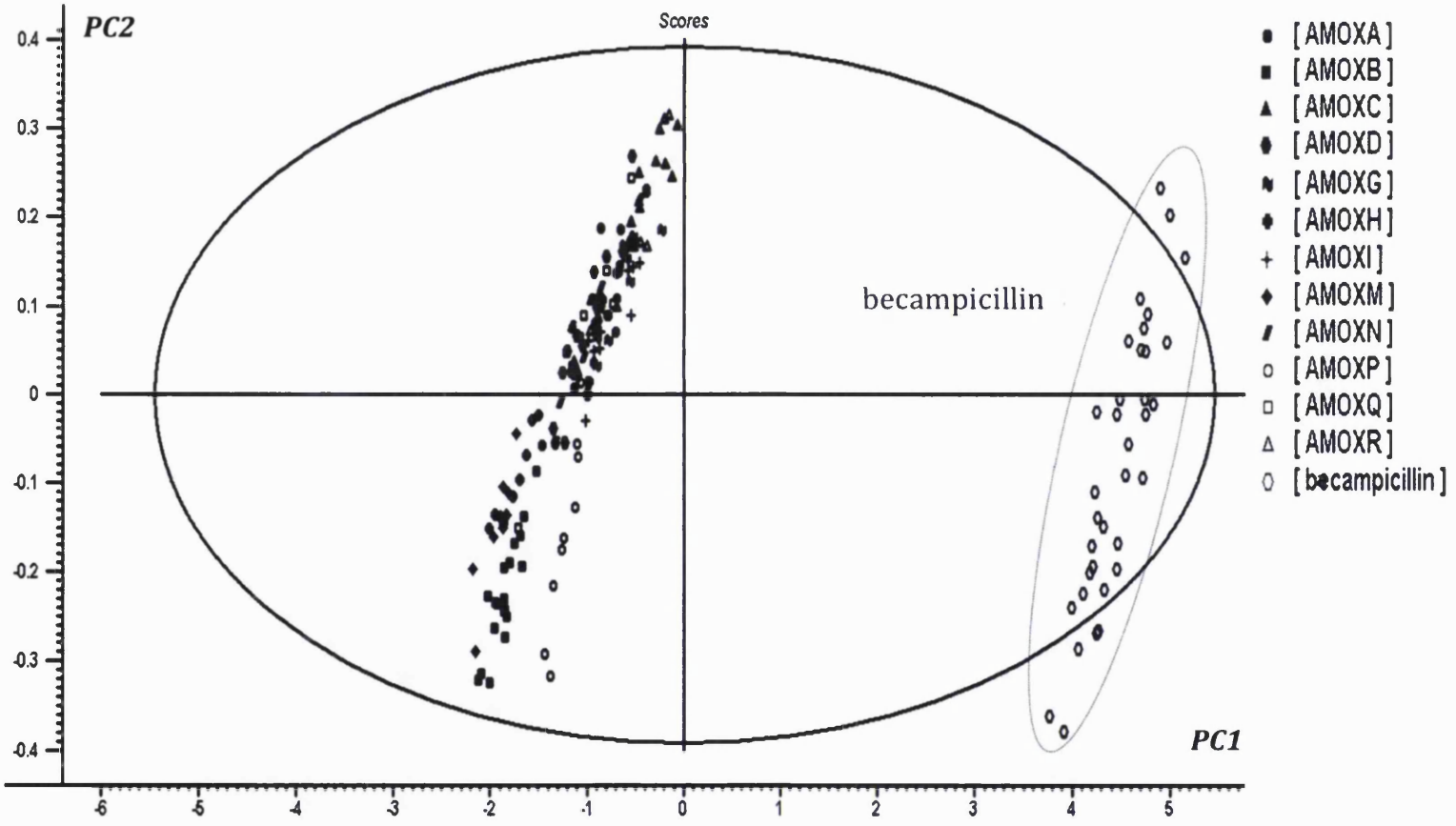


Figure 5.8 The PCA of the NIR spectra of amoxicillin and becampicillin clustered together in the 95% confidence ellipse

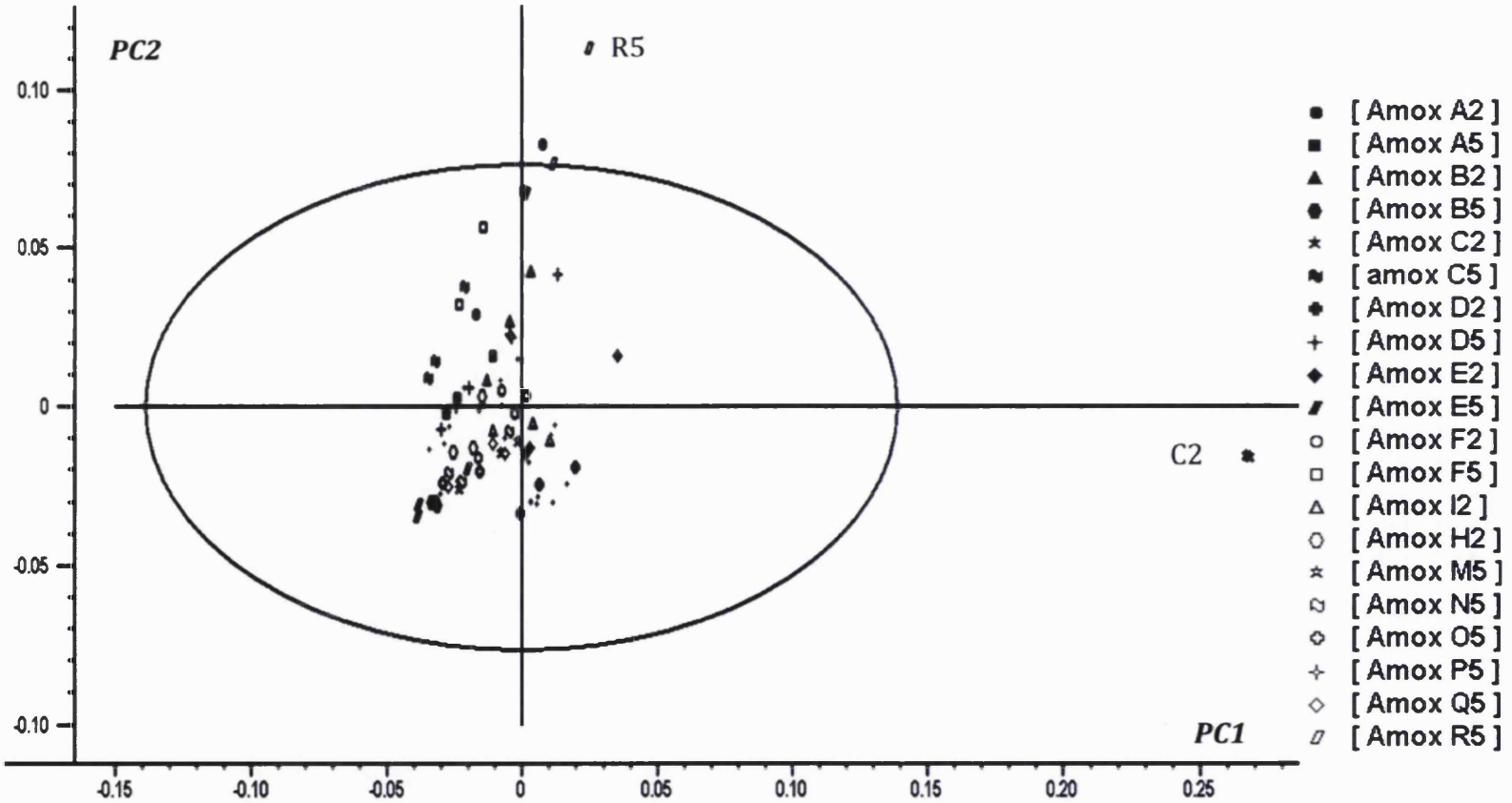


Figure 5.9 PC analysis of ^{1}H-NMR spectra of the amoxicillin samples after the bucketing process using Amix software. Sample C2 (becampicillin) was excluded from the 95% confidence ellipse.

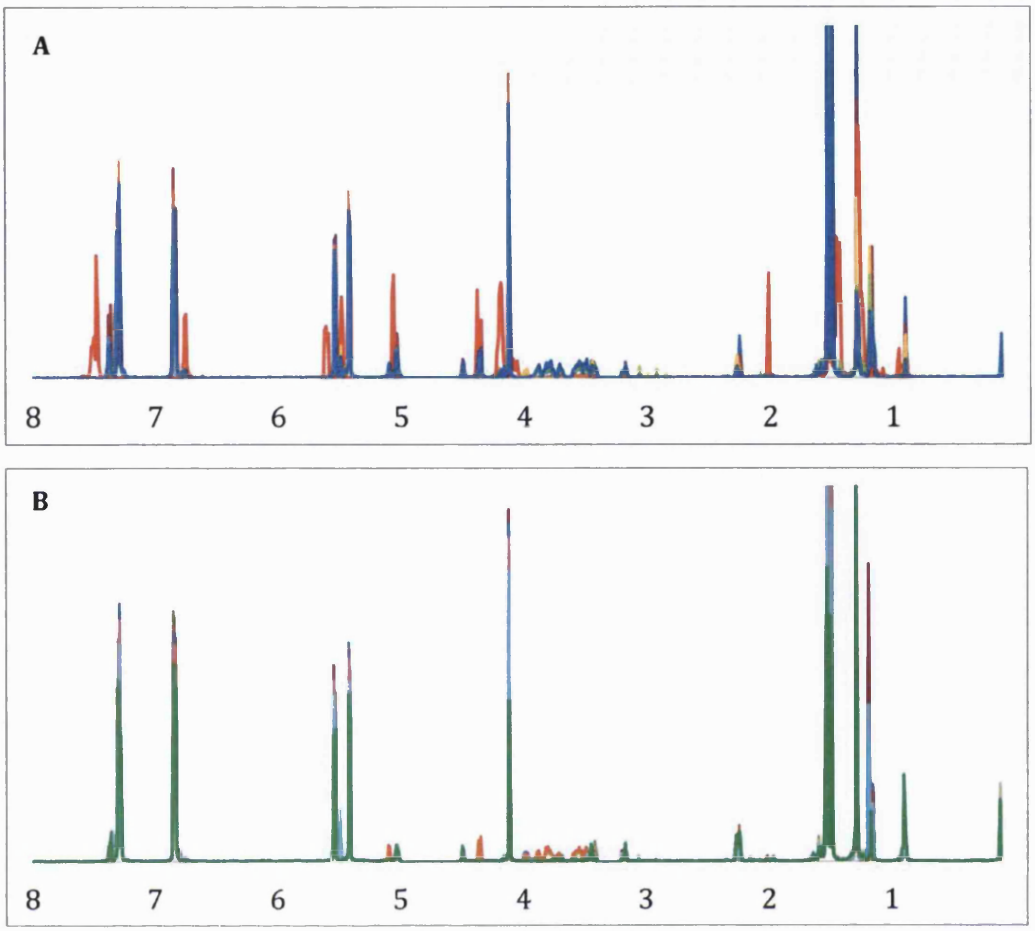


Figure 5.10 Compilation of bucketed ^{1}H-NMR spectra of the amoxicillin samples for (A) 250 mg and (B) 500 mg. The spectrum in red belongs to sample C2 (becampicillin).

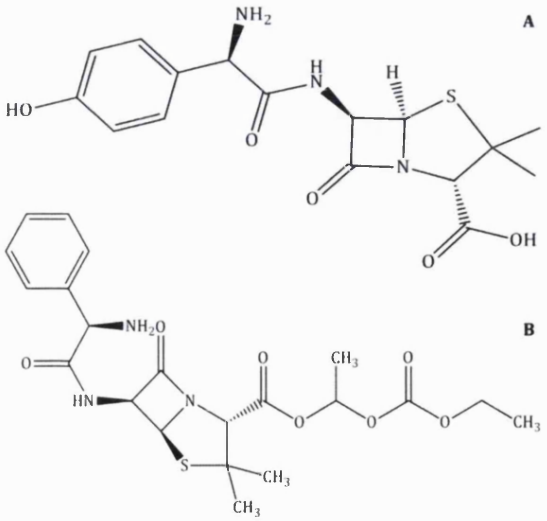


Figure 5.11 Chemical structures of (A) amoxicillin and (B) becampicillin.

5.5 Conclusion

The NIR spectral database of paracetamol was proven to be reliable for identification and quick product screening procedures. The database also managed to classify samples according to their type of dose, dosage form and product variability according to the cut-off point determined in chapter 4.

PCA analysis of same set of spectra has provided similar findings as the database search outcome. This process has not only validated the database search but also allowed the graphical observation of the findings. However, there will be cases where both the database and PCA give different outcomes. This is well expected as both analyses used different methods in calculating the degree of similarity between the spectra. This provides more information of the unknown sample and it is another advantage of the two tier screening procedure.

The internal and external validation conducted after 12 months proved that the database search gave reproducible results. The PCA allowed differences in intra-batch and inter-batch variability for different products to be observed. Although some degree of change to the spectra was observed after one year, the database successfully classified the tablets correctly using the pre-determined cut-off value. The issues of possible ambiguity could be resolved by using either chemometric analysis or other analytical chemistry techniques.

Analysis of amoxicillin samples purchased via a random sampling strategy from the Malaysian market demonstrated that patients do not always receive what they have asked for. A quick chemometric analysis of the pool of spectra could identify the different spectra without the need of a standard from the manufacturer.

Building a database of medicines on the market can be a tedious process; however, once it has been established, the spectral analysis of drugs will be a relatively simple, less time-consuming and cost-effective procedure. Furthermore, the existence of a comprehensive spectral database will be of great benefit as a repository of data for further chemometric analyses and also for drug identification, drug quality surveillance and as a potential method of counterfeit and adulterated drug screening, particularly in cases where the original samples are difficult to obtain from the manufacturers.

CHAPTER 6

DRUG QUALITY STUDY

6.1 Introduction

In order to design and implement an effective strategy for preventing poor quality or counterfeit medicines flooding the industry it is very important to know the prevalence and distribution of these products on the pharmaceutical market. One of the ways to do this is by doing continuous or periodical monitoring of the products on the market.

Theoretically, there are two main concepts of pharmaceutical testing using NIR. One method used is to identify the components of the products based on special algorithms (quantitative analysis) [148]. Another method is used to determine whether or not the product is within the specification limits by applying supervised pattern recognition techniques (qualification analysis) [149, 150, 151, 152]. Qualification involves comparing the spectrum of the unknown sample with the spectra of similar products contained in the library. Unknown spectra that are confined to the specification set are similar to the reference spectra in the library.

The spectra database employed both concepts of NIR analysis. Similarity between unknown and other spectra in the database was measured using a specific algorithm and classified samples based on pre-determined threshold values.

The presence of a spectra database will help in the quick screening of products, where only suspect items are selected for further, more expensive analyses. Besides this, the spectra bank has made it possible for general spectra comparisons to be made between products from similar batches, productions or even between two countries.

Using a limited set of samples, two pilot studies were conducted to test whether or not the database could be used to identify inter-batch variations between

spectra (part A) and how well the reservoir of spectra collected could provide a general indication of the quality of the products (part B).

6.2 Materials and Methods

6.2.1 Part A – Ginkgo biloba

6.2.1.1 Samples

Thirteen batches of 40 mg Thomson Ginkgo biloba *(G. biloba)* products (Herbal Revival Sdn. Bhd.) were used for the inter-batch comparison. Twenty capsules from each batch were emptied into glass vials and analysed by NIRS.

6.2.1.2 NIR and chemometric analyses

The NIR analysis was conducted based on the procedure outlined in sections *2.2.1.1* and *2.2.1.2* (pages 94 and 95) for the analysis of intact tablet and powder samples, respectively. PCA was conducted using the raw spectra. The data was centred and validated using the leverage correction. The 95% confidence ellipse (hotelling T2) was included in the score plots to reveal potential outliers, i.e. samples that were outside the ellipse.

6.2.1.3 Database development and search

Spectra for each batch of samples were individually computed into specific libraries according to their type. The absorbance spectra were recorded at a wavelength of 1100 to 2500 nm with 700 data points and at a 16-bit resolution. Searches were conducted using correlation algorithm for the full spectrum. Both sets of samples were included in two databases; Paracetamol_MY and Paracetamol_UK for the Malaysian and UK samples, respectively.

6.2.2 Part B – Paracetamol

The sample sets consisted of 16 batches of Malaysian (section 5.2.11, page 154).Six batches of UK 500 mg paracetamol tablets were added into the samples collection. These new samples were labelled as UK_bts, UK_lidl, UK_mrs, UK_vh, UK_wtr and UK_gsl. .

6.2.2.1 NIRS, SIMCA and Database analysis

The NIR analysis was conducted based on the procedure outlined in section *2.2.1.1*(page 94) for the analysis of intact tablets.

PCA was developed for each batch of samples. The PC models were individually constructed for each batch of the samples. After the first run, the presence of outliers, groups, clusters and trends was determined based on observations of the score plots. At this stage, the outliers detected belonged to the same population but they were badly described by the model. The optimum number of PCs was determined based on the total explained variance plot.

SIMCA was applied as in the original method [107]. Classification was performed by all of the PC models against one sample as a class model with the level of significance set at 95%. The distance versus leverage plot was evaluated.

6.3 Results and Discussion

6.3.1 Part A – Ginkgo biloba

The PC analysis of 13 batches of 40 mg Thomson *G. biloba* is displayed in Figure 6.1.One set of samples, batch 7, was located outside of the 95% hotelling T2 ellipse. This indicated that out of 13 batches, one set of samples was different to the rest. Figure 6.2 shows the SNV plot of the spectra, highlighting the representative sample from batch 7 in red. The difference between the spectra can be observed in the regions of 1400-1600 nm and 2050-2150 nm. These regions mainly correlated with the water content present in the samples, which can cause problem in the production of *G. biloba* products[153].

Further analyses of representative samples from batches 1, 7 and 9 were conducted using NMR to identify whether or not there were any differences in the composition of these samples.

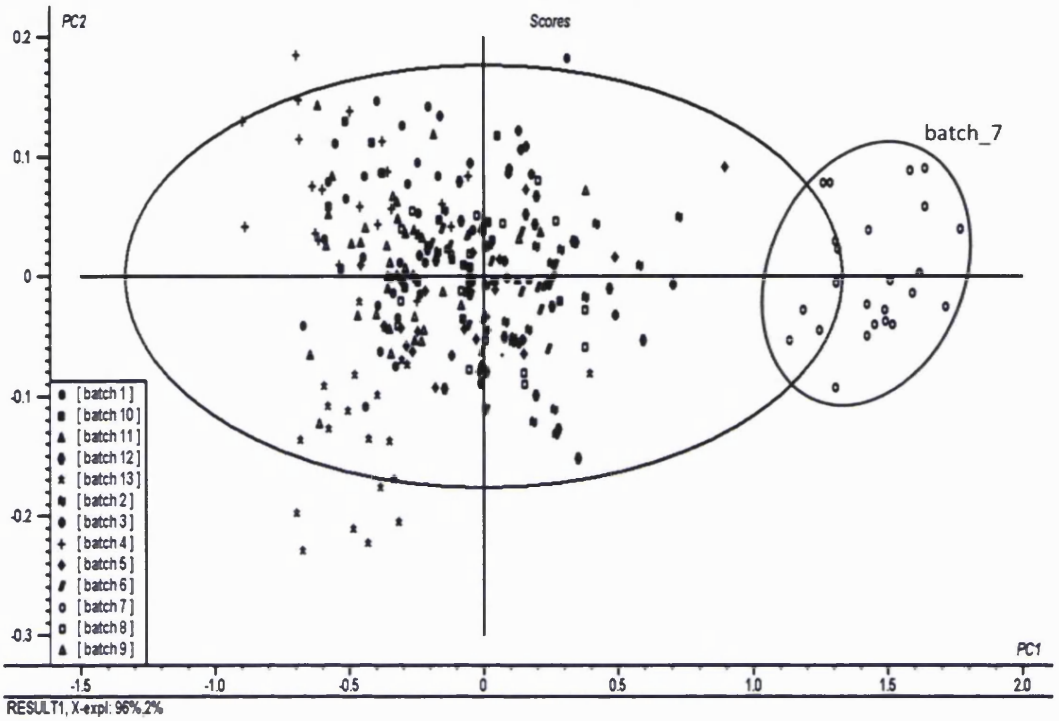


Figure 6.1 PCA score plot of 13 batches of 40 mg Thomson *G. biloba*. The spectra from samples in batch 7 were excluded from the confidence ellipse.

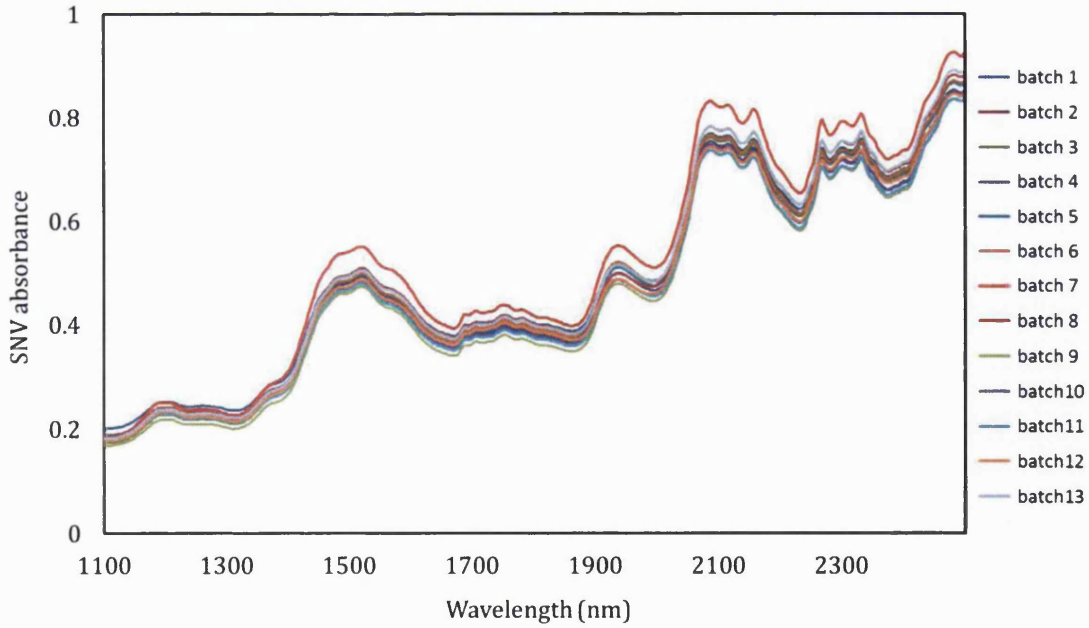


Figure 6.2 The SNV spectra of 13 batches of Thomson *G. biloba*, with the representative sample from batch 7 highlighted in red.

Figure 6.3 shows that the NMR spectra of a representative sample from batch 7 were different when compared to the spectra of samples from batches 1 and 9. Additional peaks in the spectra of batch 7 can be observed between the region 1.0-1.4 ppm and 6.0-8.0 ppm. Complete identification of these peaks was beyond the scope of this work. However, ^{1}H-NMR cannot be used to quantify the water content in samples, thus the above results cannot be directly correlated with the initial results obtained by NIR.

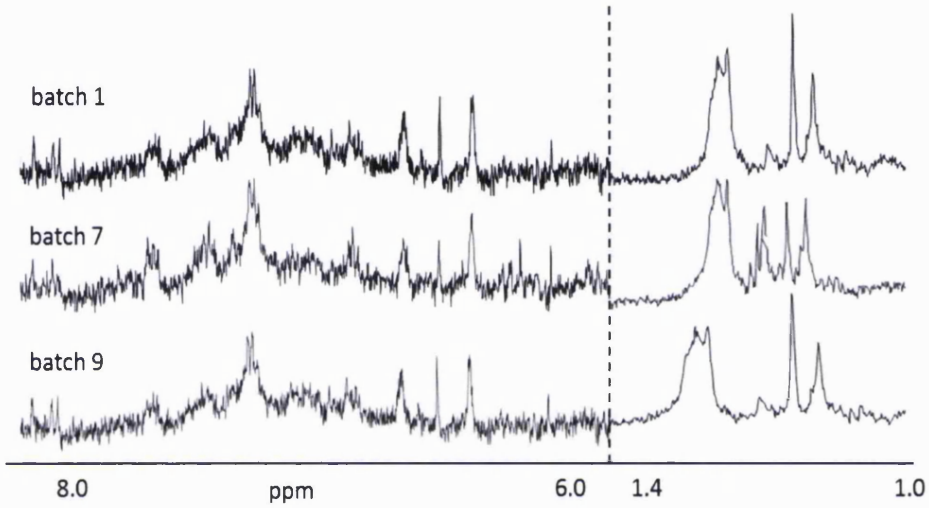


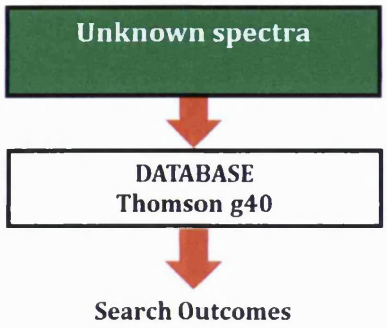
Figure 6.3 ^{1}H-NMR spectra of representative samples from batches 1, 7 and 9 in deuterated methanol at 298 K (expansions in the regions of 1.0 to 1.4 ppm and 6.0 to 8.0 ppm show where the differences were observed).

6.3.1.1 Database analysis

The NIR database consisting of 20 spectra from each set of the 13 batches of Thomson *G. biloba* was developed. Using this database, the ability of the search strategy to provide some indication that batch 7 was different from the other samples was evaluated.

Figure 6.4 shows examples of a full search of four representative sets of spectra from batches 1, 5, 8 and 13, using a correlation algorithm. All of the searches resulted in a perfect identification of matching spectra in the database. Focus was given to the spectra of batch 7, which were identified as being the least similar to the search spectra of the other examples. The HQIs measured for the spectra of batch 7 were quite high compared to those of the other 12 batches of spectra in

the database. This indicates that HQI values can be used as an indicator of the quality of the samples investigated, but further work is needed to develop relevant criteria for quality monitoring.



Spectra search: g40_1		
Hit #	HQI	Sample ID
1	.	g40_1
2	0.0001567	g40_13
3	0.0001714	g40_10
4	0.0002542	g40_6
5	0.000284	g40_9
6	0.0002909	g40_11
7	0.0006525	g40_2
8	0.000714	g40_5
9	0.0014321	g40_8
10	0.001605	g40_3
11	0.001826	g40_4
12	0.0022092	g40_12
13	0.0061028	g40_7

Spectra search: g40_5		
Hit #	HQI	Sample ID
1	.	g40_5
2	0.000279	g40_6
3	0.000525	g40_2
4	0.000616	g40_9
5	0.000631	g40_10
6	0.000645	g40_11
7	0.000714	g40_1
8	0.001073	g40_13
9	0.001124	g40_3
10	0.001183	g40_4
11	0.003724	g40_8
12	0.004899	g40_12
13	0.005528	g40_7

Spectra search: g40_8		
Hit #	HQI	Sample ID
1	.	g40_8
2	0.000309	g40_12
3	0.001003	g40_13
4	0.001432	g40_1
5	0.001927	g40_10
6	0.002134	g40_9
7	0.002138	g40_11
8	0.002353	g40_6
9	0.003057	g40_2
10	0.003724	g40_5
11	0.00462	g40_3
12	0.005048	g40_4
13	0.008015	g40_7

Sample search: g40_13		
Hit #	HQI	Sample ID
1	.	g40_13
2	0.000157	g40_1
3	0.000236	g40_10
4	0.000357	g40_6
5	0.000423	g40_9
6	0.000527	g40_11
7	0.001003	g40_8
8	0.001073	g40_5
9	0.001133	g40_2
10	0.001574	g40_12
11	0.002392	g40_3
12	0.002656	g40_4
13	0.006907	g40_7

Figure 6.4 An example of the database search outcomes for the identification of samples from four different batches of Thomson *G. biloba* product; g40_1 (batch 1), g40_5 (batch 5), g40_8 (batch 8) and g40_13 (batch 13). Sample searches were conducted using the correlation algorithm for the full spectrum.

6.3.2 Part B – Paracetamol

This example shows how the bank of spectra collected in the database could provide general observations on product quality by using chemometric analysis.

Prior to the SIMCA analysis, PCA models were developed for each set of samples, excluding the outliers. The procedures are illustrated in Figure 6.5, as proposed in earlier works (Appendix 4).

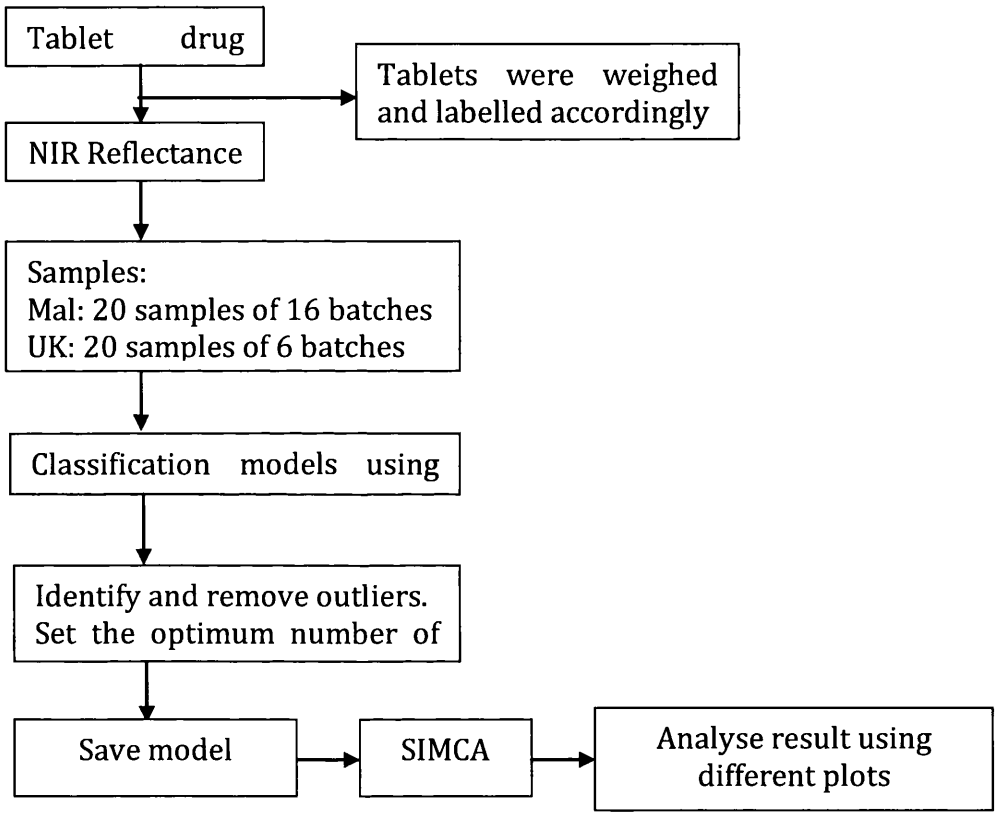


Figure 6.5 Flow chart showing the procedures in the analysis of PCM tablets by NIRS.

All of the PC models were projected into the PC space based on their distance from the sample PM as the class model. The PM sample was selected as the model class due its distinct characteristics in the NIR analysis and PCA (Chapter 5). A clear distinction of the variability between the Malaysian and UK samples can be observed in the distance versus leverage plot in Figure 5.6. All Malaysian samples clustered well within the batches but were scattered from the leverage range of 0 to 120 while the UK samples clustered together within only the leverage range of

15 to 35. Additionally, the 650 mg batch was well separated from the other batches.

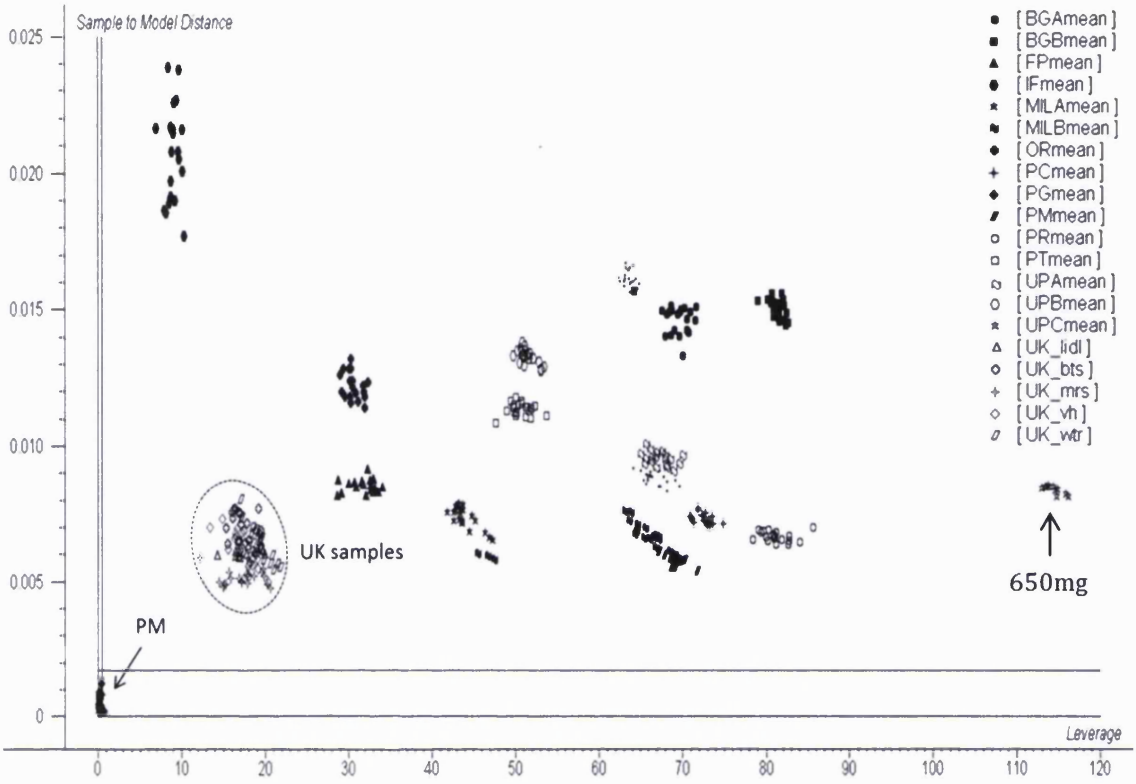


Figure 6.6 The distance versus leverage plot showing the relative distance of the samples to sample PM (class model). All of the UK samples were clustered together in the region between the leverage range of 15-35.

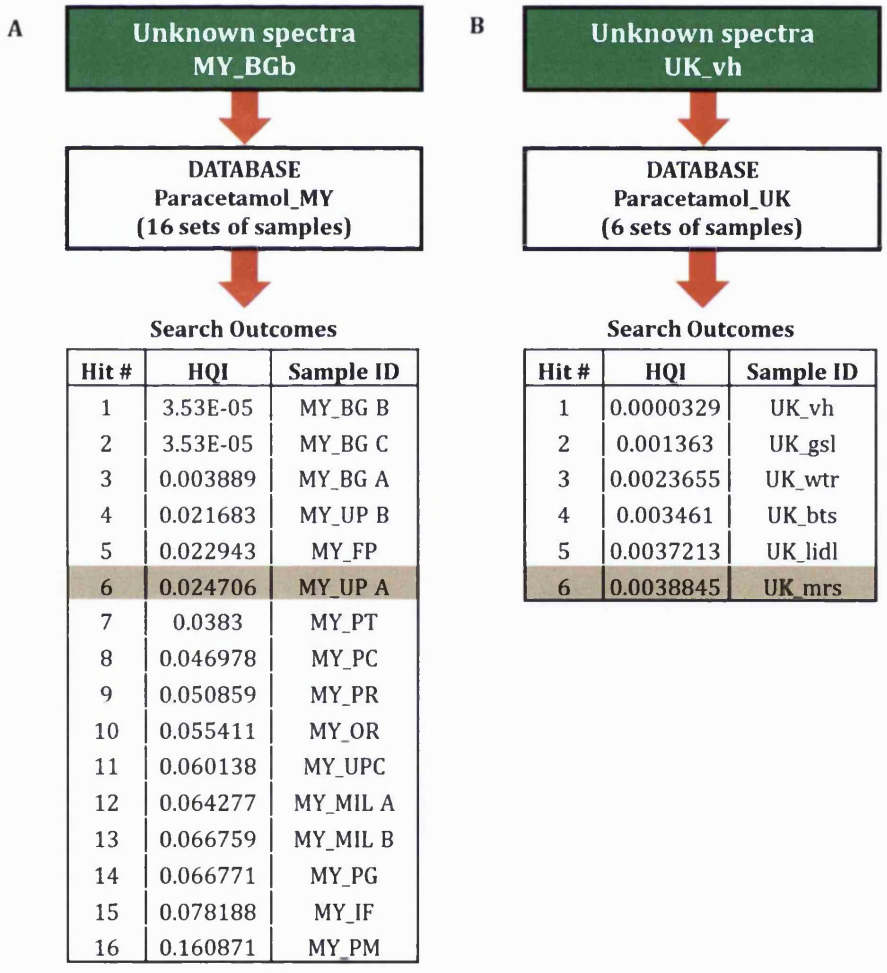
Candolfi et al. [107] have demonstrated that the number of PCs included in the classification models is an important factor for SIMCA and that using pre-processing data decreases the within-class variance and increases the between-class variance. Taking the first factor into consideration, we used the raw NIR spectra as all of the data were sufficiently described by only one and two PCs. Despite compromising the second factor, our results did not appear to be affected.

6.3.2.1 Database Analysis

Figure 6.7 A and B shows a comparison of spectra searches between Malaysian and UK paracetamol samples and databases, respectively. The HQI representations of the Malaysian samples have a wider range, 3.53×10^{-5} to 0.16, compared to 3.29×10^{-5} to 0.0039 for the UK sets. This observation corresponded

well with the wider distribution of the spectral range for Malaysian samples in the SIMCA analysis.

One limitation in this example is the imbalance in the number of samples in the two databases, which may have influenced the HQI values. However, taking only the first six samples in both databases also showed a higher range in the Malaysian samples (HQI=0.024706) compared to the UK samples (HQI= 0.0038845). This finding shows that the observation of the HQI was able to provide some indication of the general quality of the similar classes of samples and highlights the use of the database for quality monitoring.



Hit #	HQI	Sample ID
1	3.53E-05	MY_BG B
2	3.53E-05	MY_BG C
3	0.003889	MY_BG A
4	0.021683	MY_UP B
5	0.022943	MY_FP
6	0.024706	MY_UP A
7	0.0383	MY_PT
8	0.046978	MY_PC
9	0.050859	MY_PR
10	0.055411	MY_OR
11	0.060138	MY_UPC
12	0.064277	MY_MIL A
13	0.066759	MY_MIL B
14	0.066771	MY_PG
15	0.078188	MY_IF
16	0.160871	MY_PM

Hit #	HQI	Sample ID
1	0.0000329	UK_vh
2	0.001363	UK_gsl
3	0.0023655	UK_wtr
4	0.003461	UK_bts
5	0.0037213	UK_lidl
6	0.0038845	UK_mrs

Figure 6.7 Database search outcomes for the identification of samples from (A) the Malaysian (MY) paracetamol database and (B) the UK paracetamol database. Comparison between six set of samples (in grey) indicated higher HQI value for Malaysian samples.

6.4 Conclusion

The results of this preliminary study on limited sets confirmed the general view that the quality of the products can be described in two ways. One is by using the spectral database itself. Similar types of samples will normally show HQI values that are very close to each other. If one of the samples has a higher HQI compared to the other similar types of samples, there is a higher possibility that this sample has different properties to the rest of samples. Further analysis is required to identify the reason for this observation.

Monitoring the distribution of samples in any of the SIMCA plots could also provide some graphical indications about the quality of the samples. Although there are many variables that affect spectra classification, measuring the distance of these spectra relative to a specific class model (in this example, PM) can be used as a quality indicator.

With a sufficient collection of samples from different batches enriching the incremental database, it is possible to determine the cut-off threshold that could classify a sample between being of acceptable/non-acceptable quality. This could be further developed as a method for monitoring pharmaceutical products.

Chapter 7

Construction of spectral database and analysis of herbal products

7.1 Introduction

The use of herbal medicines has a long therapeutic history and continues to expand rapidly across the world. According to the WHO, herbal treatments are the most popular among other alternative medicines and are highly profitable with annual revenues of US$ 5 billion (Western Europe, 2003-2004) and US$ 160 million (Brazil, 2007) and total sales of US$ 14 billion (China, 2005) [154].

The herbal medicines industry in Malaysia was expected to be worth US$ 2.5 billion by 2010 and the industry is growing faster than the general economy, at more than 15-20% per year [155]. From 2000 to 2005, annual sales of traditional medicines increased from US$ 385 million (RM 1 billion) to US$ 1.29 billion (RM 4.5 billion) [156]. The high market demands for herbal medicines and their over-the-counter availability require strict quality monitoring to ensure their safety and efficacy. Evaluation of the quality and safety of herbal medicines by the DCA of Malaysia emphasizes limits of contaminants such as heavy metals and micro-organisms and an absence of steroids and other adulterants [157].

Despite a stringent registration procedure, the analysis of 100 products that contain *Eurycoma longifolia* Jack (commonly known as Tongkat ali) purchased by random sampling on the Malaysian market showed that 36% and 8% of these products were contaminated by higher than permissible values of mercury [158] and lead [159], respectively. This has raised the call for stricter regulations on this type of medicine to ensure their safety and efficacy [160].

Simple, quick and cost-efficient methods are required to enable the continuous monitoring of commercial herbal products. However, analyses of herbal products

are generally more challenging than the analysis of their synthetic counterparts due to their nature and complex compositions, and quality control standards for raw materials and the standardization of finished herbal products can be difficult to establish [161]. Furthermore, there are various production factors, such as variations in the composition of raw materials, water dispersion, water contents and production processes that affect the quality parameters of these products [162].

This problem of a product's quality was amplified by the increased interest in purchasing pharmaceuticals via the internet, where many counterfeit, adulterated or substandard products have been distributed [163, 164]. The complexity of herbal formulations presents challenges for devising a reliable and rapid screening procedure for determining the quality of herbal products and their continuous monitoring.

Eurycoma longifolia (E. longifolia) is a herbal plant native to many south-eastern Asian countries. It is locally known as "tongkat ali" in Malaysia, "pasak bumi" in Indonesia and "cay ba binh" in Vietnam and it has traditionally been used as a general tonic for treating high blood pressure, tuberculosis, fever, diarrhoea, jaundice and dysentery [165].

Pharmacological evaluations of various compounds obtained from this plant showed that it exhibited antimalarial, anticancer, antiulcer, antimicrobial and aphrodisiac activities that were attributed to various quassinoids, squalene derivatives, biphenylneolignans, tirucallane-type triterpenes, canthine-6-one and carboline alkaloids isolated and characterized from different parts of the plant [166]. Although some of the pharmacological properties of this herb have been scientifically proven, the quality of its commercial products may not always meet standard criteria for the concentrations of active constituents due to variations in plant age, growth conditions and manufacturing processes [167]. Three of the main compounds that have been used as reference markers for standardization and quality control purposes are eurycomanone, 9-methooxycanthin-6-one, and 14-15-β-dihydroxyklaineanone (Figure 7.1) [165].

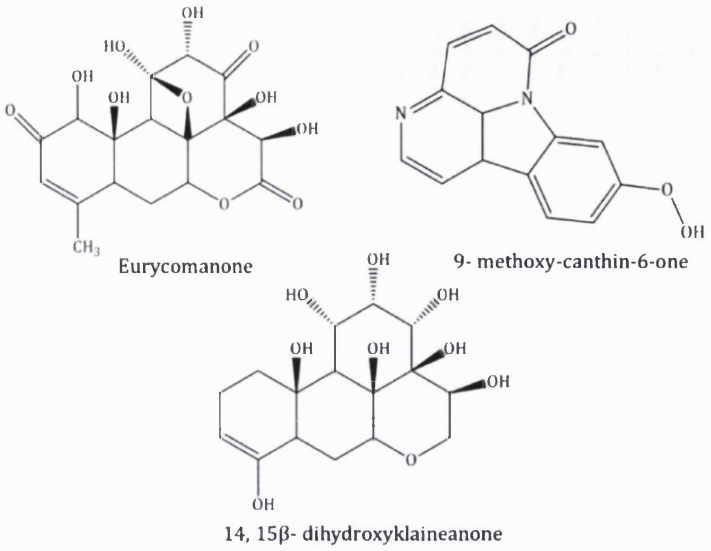


Figure 7.1 The chemical structures of eurycomanone, 9-methoxy-canthin-6-one and 14,15β-dihydroxyklaineanone, the compounds that are used as markers or as references for the standardization of *E. longifolia* products (adapted from: Zhari I. et al.; ref:165).

In this chapter, an experimental strategy was designed in order to demonstrate the use of the two-tier methods as a potential rapid screening tool for identification and classification of herbal products. This work highlights the possibility of different (contradicting) outcomes from these double analyses and indicates the advantages of use of 'single database' and 'universal database'.

A customized spectral database of different brand names of *E. longifolia* products sold in Malaysian market was created and validation of the proposed strategy was carried out using the three sets of test sample; *E. longifolia* products purchased via the internet from a UK-based vendor (Figure 7.2).

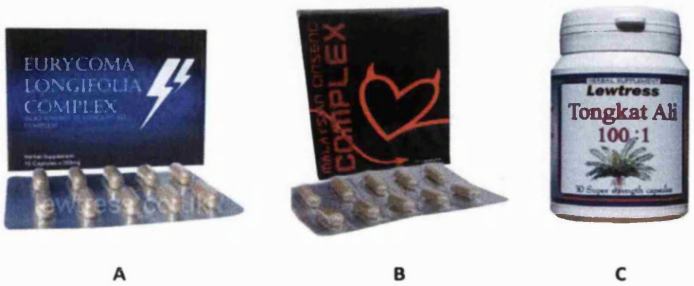


Figure 7.2 Three *E. longifolia* samples purchased via the internet; (A) *Eurycoma longifolia* Complex; ELC (B) Malaysian Ginseng Complex; MGC and (C) Lewtress Tongkat Ali; LTA.

7.2 Materials and Methods

The flowchart in Figure 7.3 shows the overview of the procedure used in this work with the direction for the complete description of method in each stage.

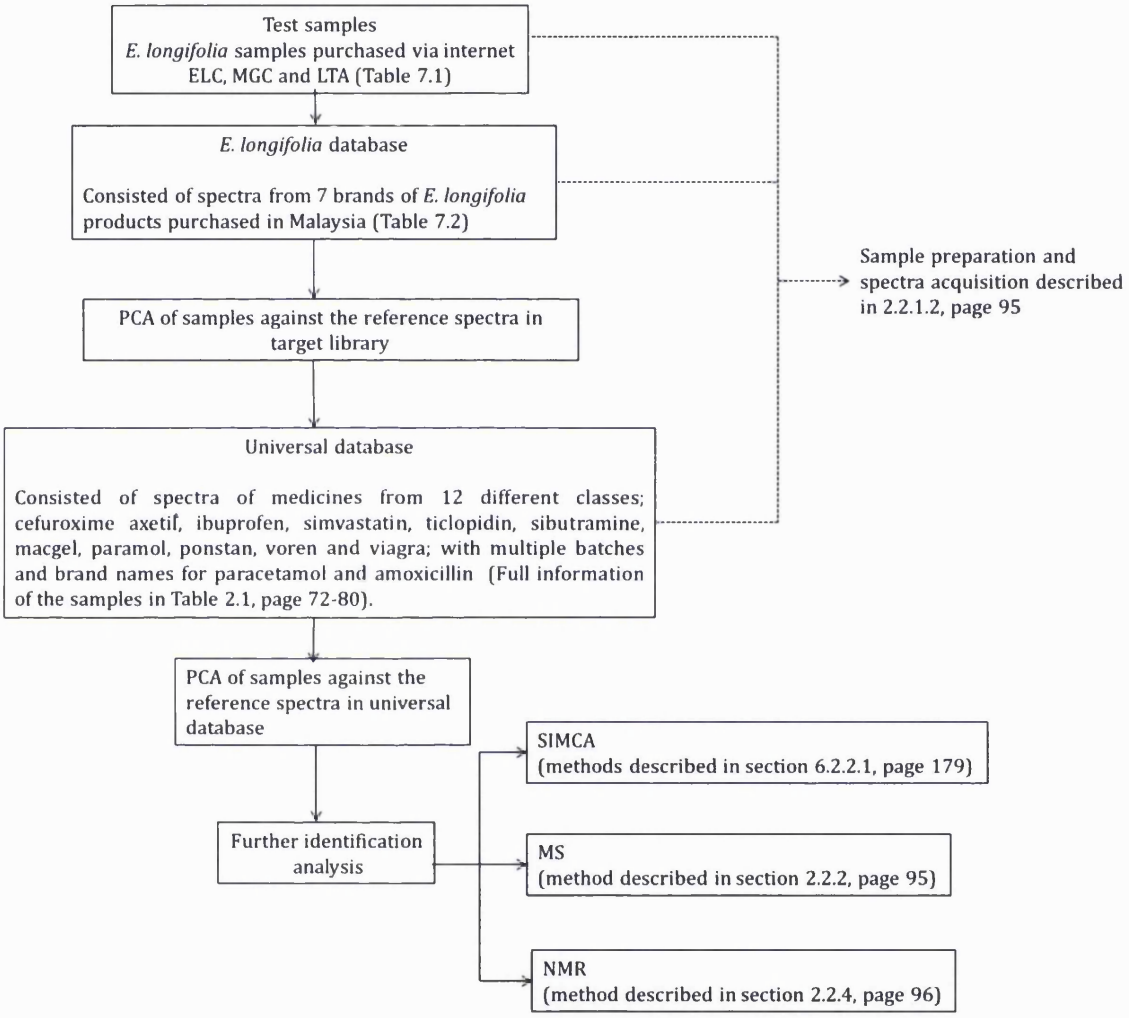


Figure 7.3 Flowchart showing the steps in screening three types of test samples (ELC, MGC and LTA) using single database of *E. longifolia* products, universal database and PCA. Further identifications were made using SIMCA (soft independent modelling of class analogy), MS (mass spectrometry) and NMR (nuclear magnetic resonance). The complete procedure can be referred to the following sections noted.

7.2.1 Samples

Commercial herbal products stated as containing herbal extracts of *E. longifolia* were purchased in Malaysia using the convenient sampling method [68]. Three

sets of *E. longifolia* samples were purchased via the internet from UK vendors [168]. Table 7.1 lists the ingredients of these products as stated on the products' packaging. Apart from these samples, 20 capsules from each of the 7 brands of *E. longifolia* were included in a customized spectral database of *E. longifolia*products (Table 7.2).

Table 7.1 List of products with their names, dosage forms, places of manufacture and ingredients as stated on the labels of samples purchased via the internet from the UK.

Sample	Label	Dosage form	Place of manufacture	Ingredient as stated on product label
Eurycoma longifolia Complex	ELC	capsule	Malaysia	*Eurycoma longifolia* 154 mg *Flos carthami* 24.5 mg *Rhizoma Cucurmae longae* 49 mg *Gingko biloba* extract 49 mg Epimedii 24.5 mg Cistanches 24.5 mg *Astragalus membranaceus* 24.5 mg
Malaysian Ginseng Complex	MGC	Capsule	Malaysia	Malaysian ginseng complex 157.5 mg *Fructus Tribulus terestris* 52.5 mg Herba Epimedium 52.5 mg *Radix Smilax myosotiflora* 28 mg *Radix Panax ginseng* 24.5 mg *Fructus Coriandrum sativum* 17.5 mg *Trigonella foenum graecum* 17.5 mg
Lewtress Tongkat Ali 100:1	LTA	Capsule	Indonesia	500 mg pure 100:1 Tongkat Ali root extract

Table 7.2 List of seven brands of *E. longifolia* products with their names, labels and the main active ingredients as stated on the labels and used as references spectra in the *E.longifolia*database.

No.	Samples	Label	Type of formulation
9	*Tongkat Ali Plus*	TA<hrx>	Mixed preparation with 50 mg Radix extract
10	*Tonex Tongkat Ali*	TA<tnx>	Radix extract 300 mg
11	*Gold Box Tongkat Ali*	TA<gb>	Radix extract 300 mg
12	*Puteri Rembulan Tongkat Ali*	TA<pr>	Radix extract 500 mg
13	*Tongkat Ali Capsule*	TA<ok>	Mix preparations with 354 mg radix extract
14	*LKH Tongkat Ali*	TA<lkh>	Radix extract 350 mg
15	*Nu-Prep100 Tongkat Ali*	TA<np>	Radix extract 100 mg

7.2.2 Database Construction

A database of NIR spectra was created using commercial software, GRAMS ID (Grams Suite Software, v 9.0, Thermofisher). Raw NIR spectra in the NSAS format were directly imported using the smart-convert function and saved in an SPC format. Other information recorded together with the spectra included the brand and proprietary names, batch numbers, expiry dates, manufacturer names and addresses, sample origins, other excipients (where available) and a description of the samples. The unknown spectra and all spectra in the database were baseline corrected to reduce the scattering effects that can highly contaminate NIR spectra. This was done using the auto-baseline correction algorithm, which removes linear baseline errors of positive peak data.

7.2.3 Chemometrics

Principal Component Analysis (PCA), an unsupervised classification method, was performed on the NIR spectra using Unscrambler 9.7 software (CAMO). The raw spectra in the JCAM-DX format were imported and the models were validated using leverage correction.

The Soft Independent Modelling of Class Analogy (SIMCA) method was used for a supervised classification of unknown samples against the pre-determined PCA models. The SIMCA analysis was also carried out using the Unscrambler 9.7 software (CAMO). At the same time, the orthogonal distances from the new objects to two different classes (models) were presented in Cooman's plots. The significance level was set at 95%.

7.3 Results and Discussion

Spectroscopic fingerprints obtained by NIR provided a distinct representation of individual herbal formulations despite their similar chemical entities since it reflected not only the active constituents but also all of the other chemical compositions/excipients and it also indicated the variability in the production processes used. The average NIR spectra of the *E. longifolia* samples after SNV transformation are shown in Figure 7.4. The spectra for three types of the test set samples (ELC, MGC, and LTA) were shown in red. The spectra for ELC and LTA can be viewed clearly however, the spectrum for sample MGC was not visible, most probably due to its overlap with other spectra.

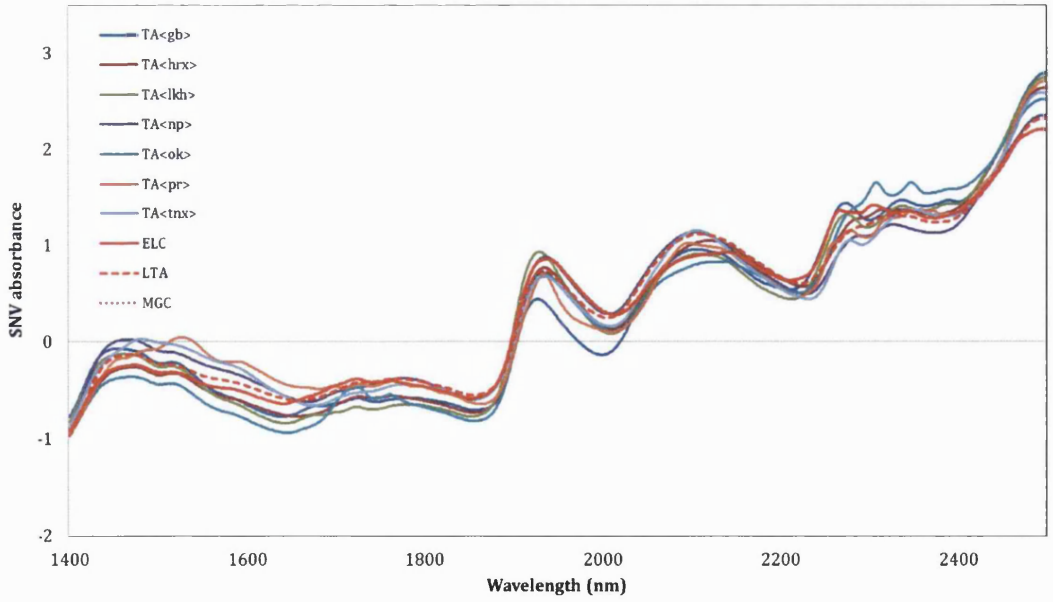


Figure 7.4 Average SNV spectra from the different commercial products of *E. longifolia* in the wavelength range of 1400-2400 nm.

7.3.1 Spectra search using single product database (E. longifolia database)

A single database contains spectra of a specific type of products or class of medicines. In this study, the single product database consisted of herbal preparations with *E. longifolia* as the main (if not only) active component. The adequate use of this type of database was only when the main components of the sample are known. This procedure was used to verify and to confirm if the sample belongs to a particular class of products and can also be used to further identify the sample according to the manufacturers or batches available in the database. The use of cut-off points were required to perform classification [162].

The average NIR spectra of the three test samples (samples *ELC,LTA*, and *MGC*) bought over the internet were imported into the single product database; consisting of different brands of *E. longifolia* products. Full spectrum search was conducted on these samples against the database to confirm their similarity using correlation algorithm for the whole spectrum. The least squares dot products of the unknown spectra were compared with the spectra in the database after being centralized to their respective means. A hit-list of the top matches is provided together with hit quality index (HQI) values. A low HQI value indicated a good match between the unknown spectra and their matching spectra in the database. The cut-off values used for classification were determined in section 4.3.2.1 (C), page 144.

The database hit-list (Table 7.3) showed that all three samples were classified as being most similar to the *E. longifolia* products. However, the HQI values observed for two of the samples; ELC (0.1133419) and MGC (0.1297498) were higher compared to LTA (0.0464). According to the cut-off values determined previously; similar class or type of sample should have HQI less than 0.1 (classification type 3). The finding indicated that there was possibility of these samples were different.

Further analysis was conducted by performing PCA of all the samples (reference spectra and test samples). The PCA score plot in Figure 7.5 showed that each brand of the herbal products clustered according to the manufacturers. Large variations of herbal samples from different productions were mainly due to the different of extraction methods and use of excipients in some of the samples.

However, all spectra including the test sample (labelled) were grouped in the 95% confidence limit with other *E. longifolia* samples in the database.

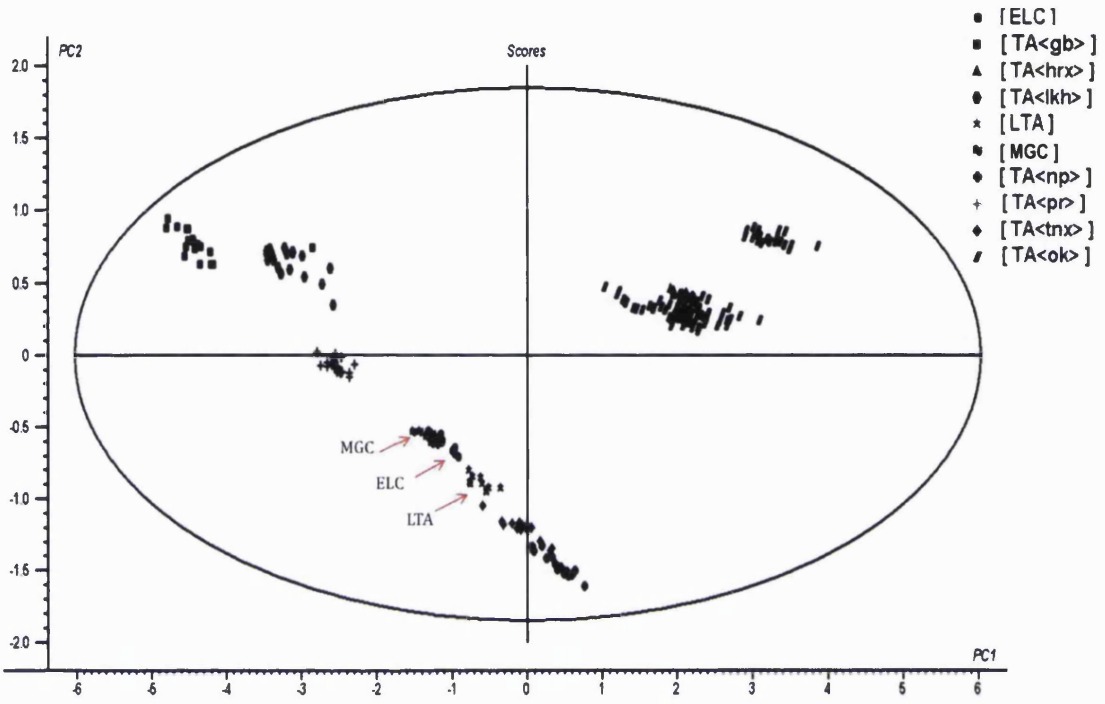


Figure 7.5 PCA score plots of the NIR spectra of commercial *E.longifolia*herbal formulations purchased randomly on the Malaysian market. The test samples; ELC, MGC and LTA (pointed) were clustered within the 95% confidence limit of other samples.

The database search outcome and the observation of the PCA score plot did not give agreeable results in this analysis. Database indicated that there was some dissimilarity of the spectra based on the higher HQI while PCA indicated that the test samples are similar to reference spectra as both sets clustered within the 95% confidence ellipse. The contradicting outcome was possible as the database and PCA used different methods in assessing the similarity between the spectra.

The database performed direct (physical) comparison of the test spectra with the reference spectra from the sub-library using specific algorithms which was correlation coefficient in this work. On the other hand, PCA was conducted using latent variables which were the principal components (PC), derived from the maximum variability of the spectra set. The PCA score plot was observed based on the variability presented by PC1 and PC2 only and not the whole spectrum. In

this case, PC1 and PC2 would probably represent the active components of *E. longifolia*.

One advantage of the two tier method is that although both methods did not give similar indication in some cases, information obtained from both analyses could be useful in identification the unknown spectra. Also, the lack of graphical representation in database analysis was compensated with PCA which would aid the understanding of the outcomes.

From these findings, it could be deduced that all the test samples contained *E.longifolia*extract as one of the main components. Two of the test samples with HQI above the threshold, ELC and MGC may also contained other compounds that are significantly different from the rest of the samples. To confirm this, another search was conducted against the universal database that consisted of over 3000 NIR spectra of common medicines acquired in our laboratory for fifteen types of Malaysian pharmaceutical products.

7.3.2 Spectra search using universal database.

The universal database is the combination of all the single product databases and it contained all the spectra acquired throughout the research. In a larger database that includes different types of medicines, it might be possible to detect the presence of other compounds or adulterants that may be present in sample ELC and MGC. For this purpose, the use of cut-off points was no longer relevant but similarity was indicated based on the highest rank on the database hit-list.

Spectra search of the test samples against the universal library indicated some degree of similarity of ELC and MGC with sildenafil citrate (Table 7.4). PCA analysis on test spectra and reference spectra from the library showed a close cluster of sample ELC and MGC with sildenafil citrate (Figure 7.6). With these findings, it can be hypothesised that both samples ELC and MGC were contaminated with an active substance, sildenafil citrate. Further analysis was conducted using MS, NMR and SIMCA.

Table 7.3 The hit-list showing the database search outcomes for samples (A) *ELC*(B)*LTA* and (C) *MGC* against the *E. longifolia* database. The HQI and sample in bold showed slightly higher values that the cut-off threshold determined in chapter 4 (page 144).

A) ELC

Hit #	HQI	Sample ID
1	**0.1133**	**TA<hrx>**
2	0.1707	TA<np>
3	0.1881	TA<tnx>
4	0.1995	TA<gb>
5	0.2049	TA<pr>
6	0.2247	TA<lkh>
7	0.2355	TA<ok3>
8	0.25	TA<ok4>
9	0.2504	TA<ok2>
10	0.376	TA<ok1>

B) LTA

Hit#	HQI	Sample ID
1	0.0464	TA<np>
2	0.0571	TA<tnx>
3	0.117	TA<pr>
4	0.1564	TA<hrx>
5	0.2434	TA<gb>
6	0.3421	TA<lkh>
7	0.3725	TA<ok3>
8	0.3853	TA<ok2>
9	0.3878	TA<ok4>
10	0.5127	TA<ok1>

C) MGC

Hit #	HQI	Sample ID
1	**0.1297**	**TA<hrx>**
2	0.1564	TA<np>
3	0.181	TA<tnx>
4	0.2036	TA<pr>
5	0.2212	TA<gb>
6	0.2498	TA<lkh>
7	0.2595	TA<ok3>
8	0.2745	TA<ok4>
9	0.2754	TA<ok2>
10	0.4046	TA<ok1>

Table 7.4 The hit-list showing the database search outcomes for samples (A) *ELC*and (B) *MGC* against the general customized database with the top-ten closest match. Both samples *ELC* and *MGC* showed highest similarity with sildenafil citrate sample (Viagra).

A) ELC

Hit #	HQI	Sample ID
1	**0.217919**	**viagra01**
2	0.247666	vor05
3	0.330835	vor07
4	0.333971	vor06
5	0.362105	mac28
6	0.37943	mac29
7	0.381483	mac27
8	0.441572	sim39
9	0.462623	AmoxP
10	0.468526	Amox E

B) MGC

Hit #	HQI	Sample ID
1	**0.233734**	**Viagra01**
2	0.250584	vor05
3	0.341273	vor07
4	0.343788	vor06
5	0.374505	mac28
6	0.391807	mac29
7	0.393594	mac27
8	0.430849	sim39
9	0.444474	Amox E
10	0.444864	AmoxP

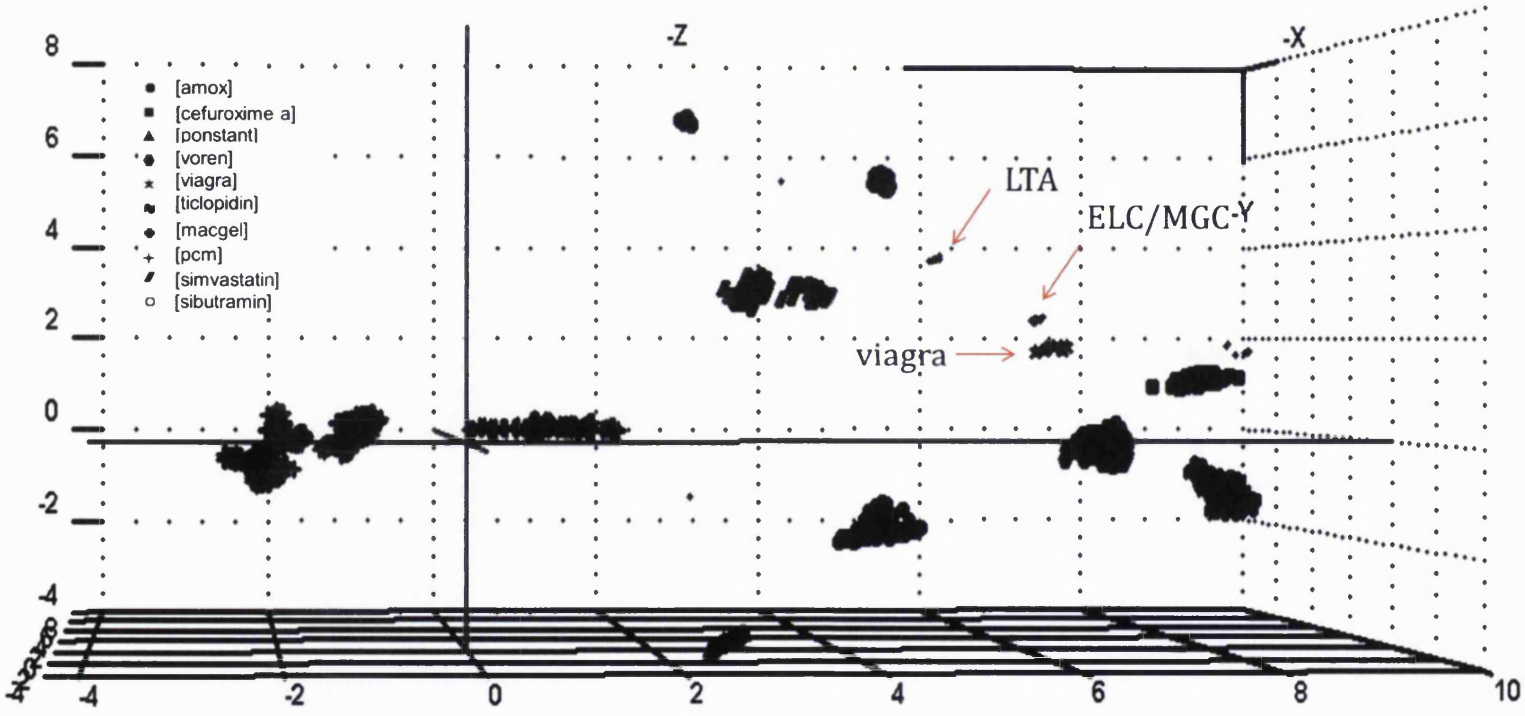


Figure 7.6 PCA score plot showing the distribution of the test samples purchased as *E. Longifolia*compared with the spectra of conventional medicine samples. Samples ELC and MGC were clustered close to the cluster of sildenafil samples (Viagra).

7.3.3 SIMCA

Classifications by SIMCA (Figure 7.7) showed that two of the internet samples, *MGC* and *ELC*, were actually the same product despite bearing different brand names and different ingredients on the packaging of both products (refer to Table 7.2). The primary packaging (blister packs) of both products and the distributor shown on the secondary packaging looked the same and both products were purchased from the same vendor. Based on the product declarations, the *MGC*product was 20% more potent than the *ELC* product and, consequently, the *MGC* product was sold at a higher price. However, our findings indicate that these two products actually had a very similar composition.

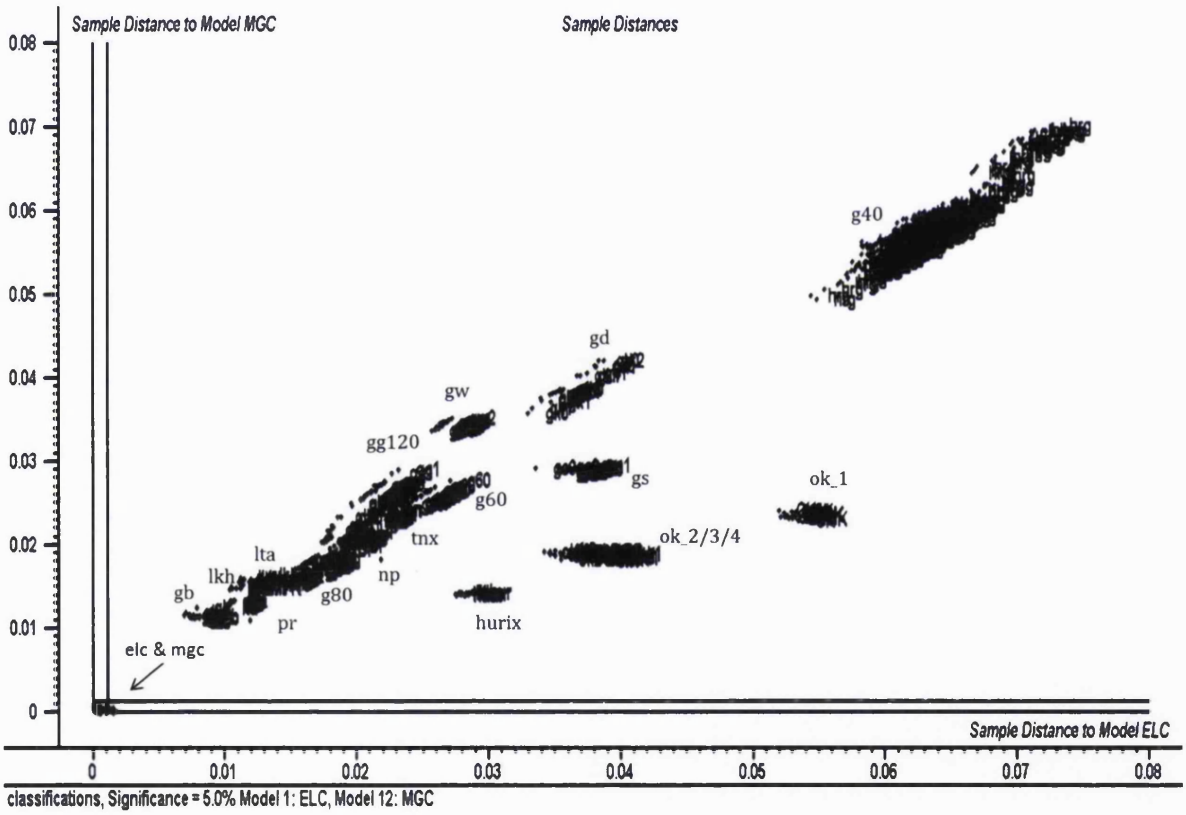


Figure 7.7 The Cooman's plot derived from the SIMCA analysis showing that two of the internet samples, *ELC* and *MGC*, were from products with a similar classification. The significance level was set at 95%.

7.3.4 Mass Spectral Analysis

The ESI- MS analysis of *LTA* showed mass spectra that were of a similar pattern to the rest of the *E. longifolia* samples in the database. However, the mass spectra of samples *ELC* and *MGC*were clearly different from spectra of the rest of the *E.*

*Longifolia*sample set. . A presence of a major peak of $[M+H]^+$ion peak at *m/z* 522.1 ESI-MS spectrum indicated that the herbal product was spiked with a synthetic compound (Figure 7.8a). This peak could possibly represent the adulterant in the sample and denoted as compound 1 onwards. The other peaks present at higher *m/z* value were the aggregates of this compound and its fragments.

The mass fragmentation pattern (Figure 7.8b) was similar to that of thiohomosildenafil [169], which indicates that compound 1 might be structurally relevant to thio-sildenafil analogues [170, 171]. An NMR analysis was conducted in order to confirm this hypothesis.

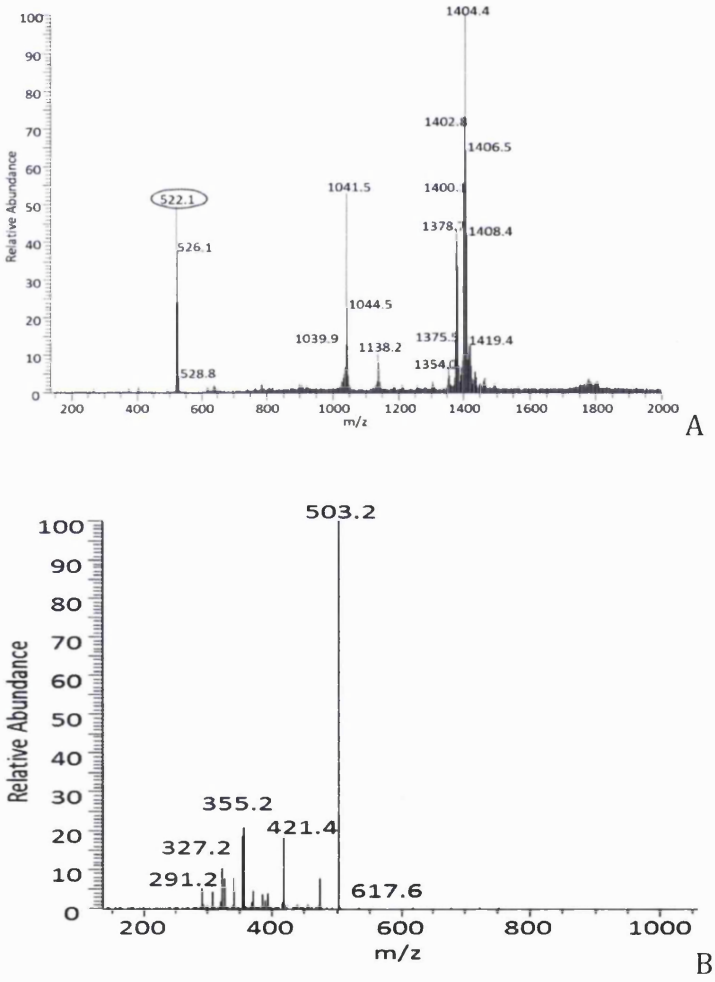


Figure 7.8 ESI-MS spectra of sample *ELC* in (A) full mass spectrum showing the $[M+H]^+$ion peak of compound 1 at *m/z* 522.1 (circled) (B) MS/MS fragmentation of peak 522.1 with normalized collision energy of 35%.

7.3.5 NMR Analysis

Phosphodiesterase-5 inhibitors (PDE-5i) are widely used for the treatment of erectile dysfunction (ED). Sildenafil, tadalafil and vardenafil are three types of PDE-5i that have been approved by the FDA for this indication [172]. However, health and herbal supplements have been found on the market contaminated by not only these three substances but also by their analogues [169, 173, 174].

NMR spectroscopic analysis of commercial herbal and dietary formulations revealed that 8 out of 17 samples contained approved PDE-5i or related compounds, including sildenafil, tadalafil, vardenafil, hydroxyhomosildenafil, thiosildenafil, and thiomethisosildenafil [173]. Another analysis of a supposedly 100% natural product for ED that was purchased over the counter and directly from the manufacturer via the internet revealed that the samples analysed contained an average of 55 mg sildenafil per capsule [174]. Figure 7.9 shows the chemical structure of sildenafil citrate.

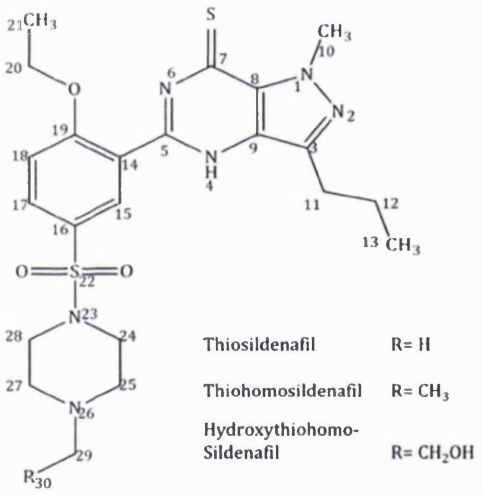


Figure 7.9 Chemical structure of sildenafil and its two analogues; thiohomosildenafil and hydroxythiohomosildenafil.

The ^{1}H and ^{13}C signals of the samples (after simple methanol extraction) were given in Figure 7.10 and 7.11A, respectively. In these spectra, the assignments of the NMR signals for compound 1 (in Table 7.5) were made based on the reference to previous literature of related work [169] which showed that compound 1 were very similar to thiohomosildenafil [169], except for the presence of H-29, H-30, C-29 and C-30; H-29 and H-30 were observed as a triplet at 2.95 ppm and as

amultiplet at 3.75 ppm in ^{1}H-NMR (Figure 7.10). Two different carbon signals at 60.00 ppm and 57.52 ppm were also observed in ^{13}C and DEPT spectra (Figure 7.11). Based on this information, compound 1 was classified as hydroxythiohomosildenafil ($C_{23}H_{32}N_6O_4S_2$, MW 521.0).

Table 7.5 Assignments of the NMR spectra of compound 1.

Carbon no.	^{1}H (ppm)*	^{13}C (ppm)*	DEPT **	COSY	HMQC	HMBC
3	-	147.12	0	-	-	H11,H12
5	-	148.83	0	-	-	H15,H18
6	-	-	-	-	-	-
7	-	174.05	0	-	-	H10
8	-	133.56	0	-	-	H10
9	-	133.80	0	-	-	H11
10	4.46 (3H, s)	39.51	3	-	H10	-
11	2.89 (2H, t)	28.36	2	H11/H12	H11	H12,H13
12	1.82 (2H, m)	23.29	2	H12/H11, H13	H12	H11,H13
13	1.00 (3H, t)	14.29	3	H13/ H12	H13	H11,H12
14	-	122.84	0	-	-	H18
15	8.36 (1H, s)	130.92	1	H15/ H17	H15	H17
16	-	128.98	0	-	-	H18
17	7.92 (1H, d)	131.57	1	H17/H15. H18	H17	H15
18	7.39 (1H, d)	114.86	1	H18/H17	H18	-
19	-	161.96	0	-	-	H15,H18,H20
20	4.34 (2H, q)	67.22	2	H20/H21	H20	H21
21	1.53 (3H, t)	14.96	3	H21/H20	H21	H20
24, 28	3.11 (4H, m)	45.34	2	H24-H28/ H25-H27	H24,28	-
25, 27	3.26 (4H, m)	53.05	2	H27-H27/ H24-H28	H25.27	H29
29	2.95 (2H, t)	60.00	2	H29/H30	H29	H30
30	3.75 (2H, m)	57.52	2	H30/H29	H30	H29
OH	1.27 (1H, s)	-	-	-	-	-

*ppm in Methanol d-$_4$**number of attached protons

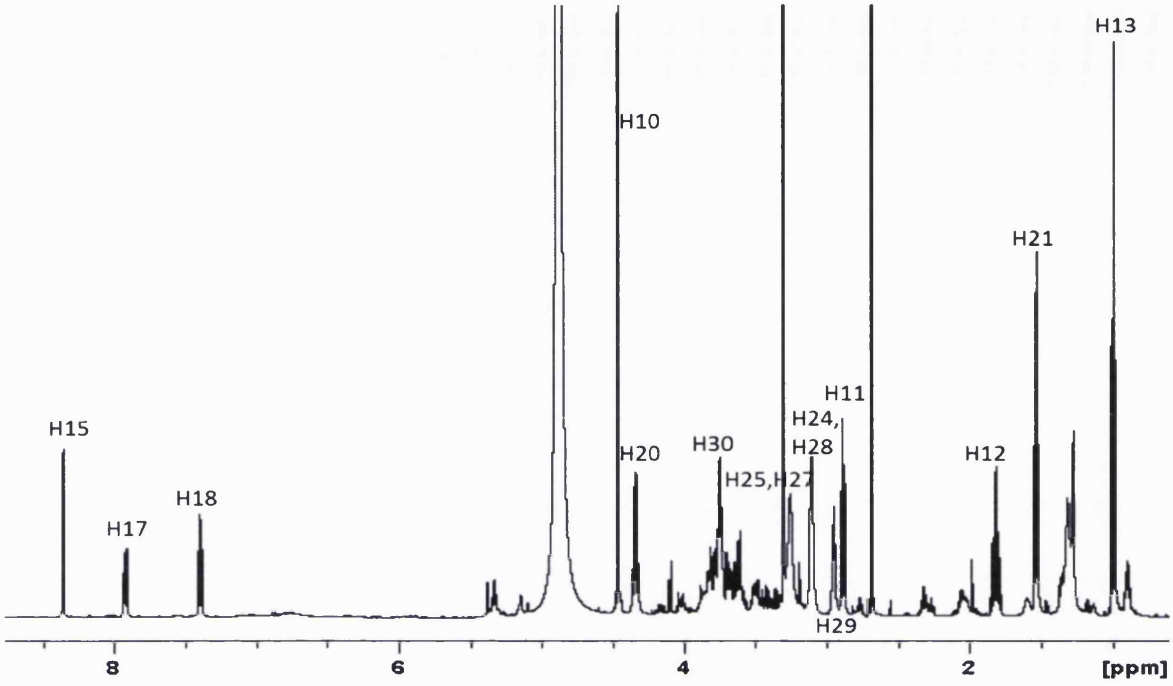


Figure 7.10 ^{1}H-NMR spectrum of compound 1.

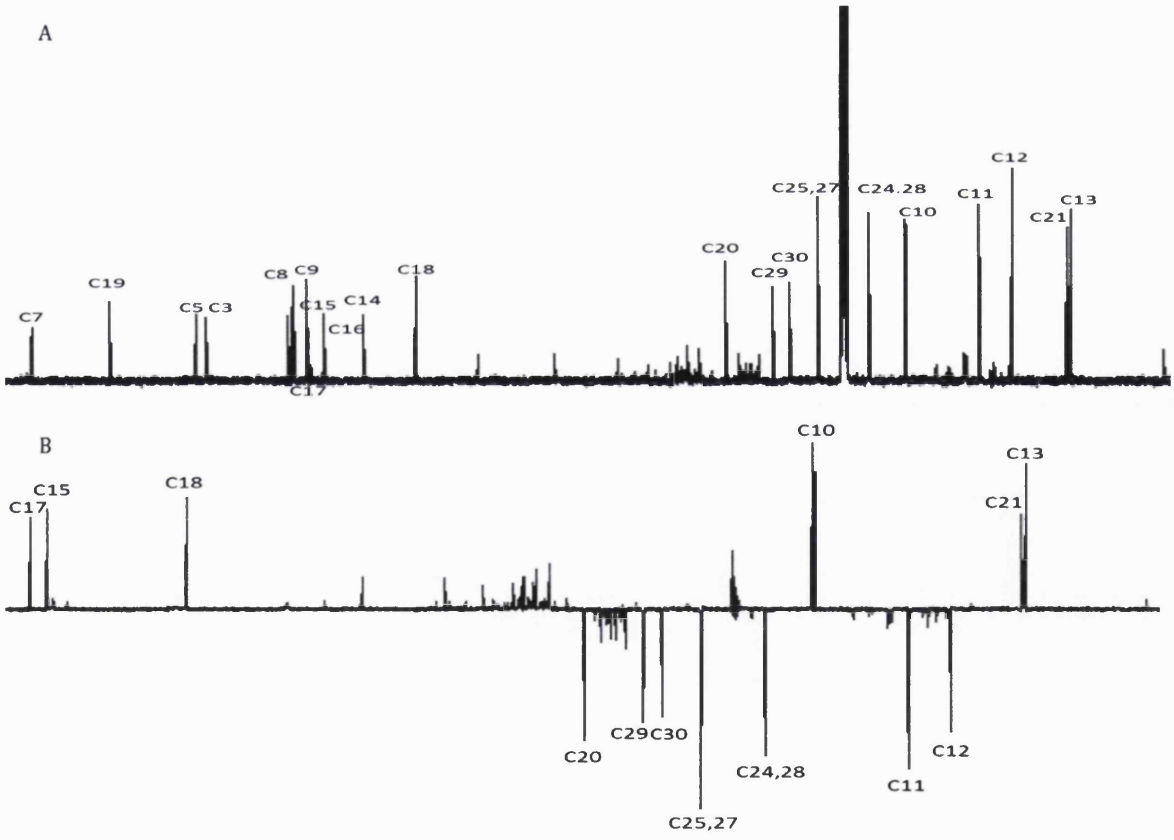


Figure 7.11 ^{13}C (a) and DEPT (b) spectra of compound 1.

7.4 Conclusion

The application of two-tier screening method using NIR spectral database and PCA in analysing herbal preparation purchased via internet was presented in this work. All steps in the procedure and their outcomes are illustrated in Figure 7.12. Double screening analysis were conducted using firstly, the single product database (A) followed by the universal database (B) has indicated the presence of sildenafil-like compound in two of the test samples ELC and MGC in addition to the *E. longifolia* compound as purported. The use of single product database was justified because it was assumed that the composition of the test samples was known.

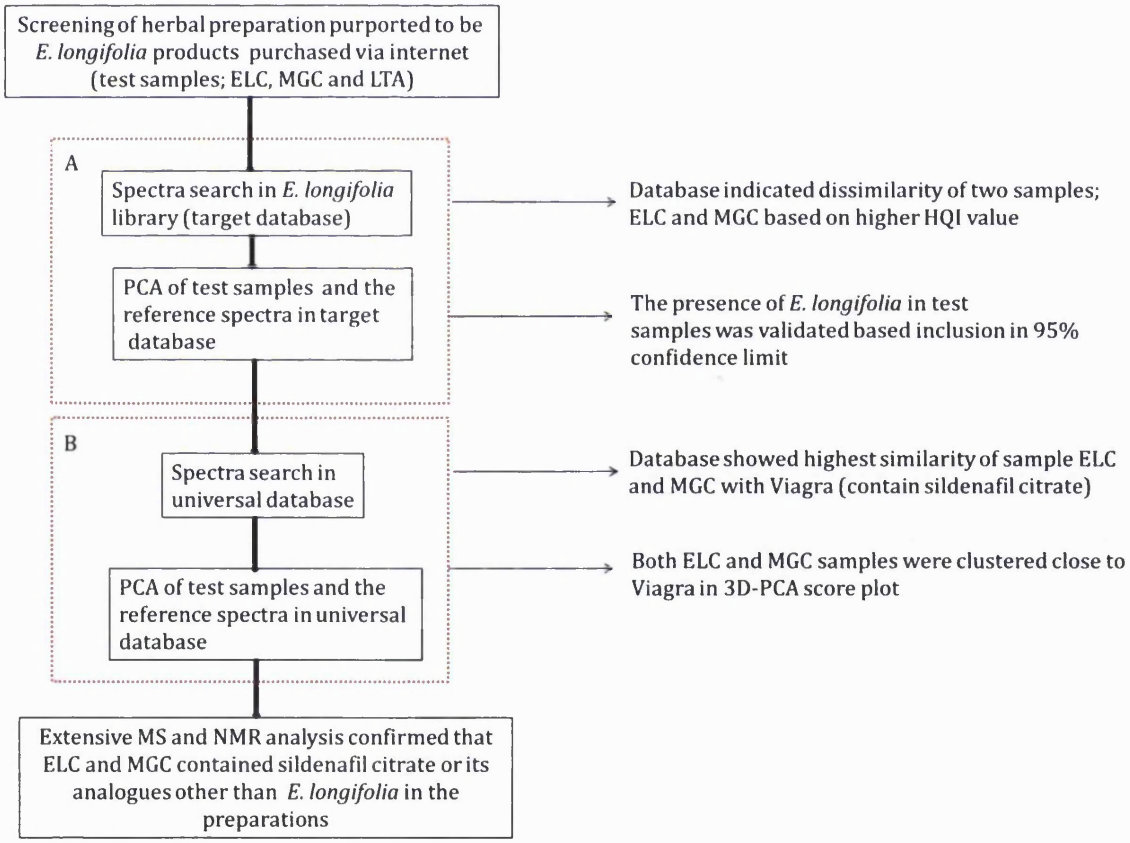


Figure 7.12 The flowchart of the two-tier screening protocol and outcomes of the analysis of *E. longifolia* samples using the (A) single product database; and (B) universal database. The combined use of two databases indicated the presence of sildenafil-like compound in ELC and MGC samples.

This method has allowed quick screening on the test samples to verify their content as labelled despite not having the spectra of those products in the database. This is one of the advantages of this procedure whereby analysis can be conducted without depending on the reference product or standard active compounds in the database. Other than that, the method was able to screen the presence of undeclared compound(s) in the products. One of the limitations of the database was the lack of graphical representations which was well compensated by PCA.

This strategy can be used to build incremental spectral databases of products on different markets as repositories of data for further monitoring and chemometric analysis. These could be used for product identification, drug quality surveillance and as potential methods for counterfeit and adulterated drug screening, particularly in cases where reference samples are difficult to obtain.

CHAPTER 8

SPECTROSCOPY ANALYSES OF ADULTERATED AND COUNTERFEIT SAMPLES

8.1 Introduction

Counterfeit drugs were first addressed at an international health meeting 20 years ago and methods for the prevention of counterfeit and substandard pharmaceuticals began to be discussed in 1988 [175]. Although the definition of counterfeit medicines by the WHO [7] is globally accepted, many countries still have their own interpretation and classification of these types of products [176]. For example, in addition to the common definition, products unregistered with the National Pharmaceutical Control Bureau (NPCB) are also classified as counterfeit in Malaysia but not in some countries [177]. These variations can cause difficulties in assessing the extent of the problem and thus become a barrier against combating counterfeit medicines at the international level [178].

An update of the prevalence of counterfeit medicines in Malaysia for the years 2010 and 2011 was not reported in any scientific journals. However, the first half of 2011 showed no reduction in this problem with the exposure of several cases [179, 180]. In one raid close to the main city of Kuala Lumpur, about 400 products consisting of common medicines, traditional herbal products and cosmetics, with a cost of up to half a million ringgit, were seized from 17 outlets. Among these products were counterfeit paracetamol, eye drops, and liniment medications disguised as being from reputable brand names [181, 182].

In addition to this problem, the adulteration of herbal medicines appears to be on the increase [183]. Adulterated medicines can be defined as products that have been mixed with some other substance(s) so as to reduce or increase their quality or strength. From 2006 to 2008, 17 products indicated for men's health, 6 products for weight loss, 5 cough medicines and 4 products for joint pain were

found to be adulterated with synthetic drugs in order to increase their pharmacological effects [16]. This has the potential to be an exceptionally dangerous problem as herbal products are highly complex and the pharmacology of conventional single chemical-herbal drug combination is poorly understood.

The fight against this health threat is considered seriously by the authorities in each country. Different policies and analytical strategies have been designed to help with this issue [184,185]. Table 8.1 shows some of the different methods that have been described for counterfeit drug detection, together with their advantages and disadvantages [80]. In general, these techniques can determine the contents of products, although they require ample time for analysis and expertise for their use, and tablets are destroyed, preventing their use as evidence if further action is desired.

In recent years, NIR spectroscopy used with Raman spectroscopy as a screening procedure for counterfeit drug detection has been further developed [113, 186, 187]. This technique is favourable due to its simple procedure and rapid analysis, as sample preparation is not necessary. The non-destructive nature of the analysis allows samples to be retained for subsequent analysis. One disadvantage of this technique is the overlapping of NIR bands, which requires spectral interpretation aided by appropriate chemometric analysis.

In this chapter, the application of the two tier screening method in identifying samples that were counterfeited, adulterated or with unknown active ingredients was demonstrated. The spectra of six adulterated and counterfeit products seized by the NPCB were searched against the universal database to find the possible undeclared ingredient in each sample. PCA score plot was observed to point out the potential conventional medicines used as undeclared ingredient.

Previous studies demonstrated the detection of counterfeit drugs by NIR using a large database consisting of wide variety of proprietary medicinal products (PMPs) with a large number of samples of each type [188] or a limited number of samples in each class [113, 186, 189], and also a large database that focused on certain types of medicines only [187, 190, 191]. In most of these works, spectral comparison was made by developing PCA models of similar type of samples and using classification method like factorial discriminant analysis or partial least square analysis to compare the unknown spectra with all the reference spectra in

the model. One main disadvantage of this technique was that new models need to be developed every time new reference spectrum is included in the library.

For the two-tier screening method proposed, no models development was required for spectra comparison. The unknown spectra acquired can be directly compared to selected or all spectra in the database. Once the unknown is identified, the new spectra can be directly included in the database. This method is a simple, fast and high throughput screening procedure which can be used to identify potentially 'bad' samples for further extensive, time consuming and expensive analysis on not widely accessible instruments like MS and NMR.

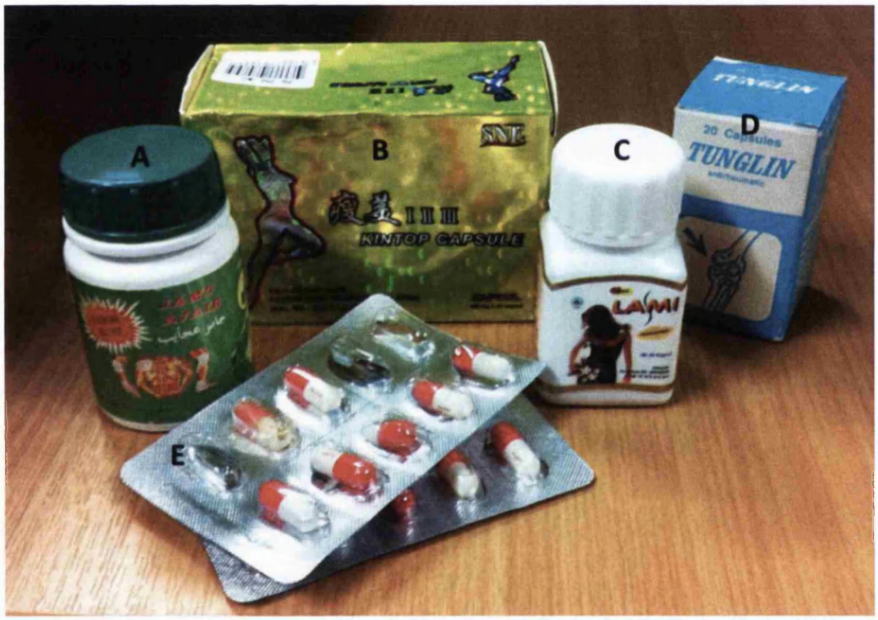


Figure 8.1 The adulterated samples included in analysis (A) Jamu Ajaib (B) Kintop (C) Lami (D) Tunglin and (E) Pronoton.

Table 8.1 Different techniques used for counterfeit drug detection with their advantages and disadvantages [68].

Technique	Type	Advantages	Disadvantages
Direct ionization mass spectrometry	Desorption electrospray ionization (DESI)	Non-destructive. Rapid. Identification and quantitation of AIs in large mass range.	Medium complexity instrumentation. Requires MS instrument.
	Direct analysis in real time (DART)	As above. Simple spectra. Suited for small molecules.	As above.
MALDI mass spectrometry	Single stage TOF	Allows fast determination of molecular weight and purity of biological pharmaceuticals.	Not widespread. Sophisticated and costly instrumentation.
	Tandem TOF	As above, plus possibility of quantification and determination of protein sequence.	Not widespread. Sophisticated and costly instrumentation.
Inductively coupled plasma spectrometry	Optical emission	Provides elemental composition. Large dynamic range.	Expensive. High maintenance costs. Requires constant supply of high purity gases.
	Mass spectrometry	Provides elemental composition, very sensitive for detection of trace impurities. More selective than optical.	As above.
Ion mobility spectrometry	----	No vacuum needed, low maintenance, compact instrument.	Moderately expensive, medium resolution, identity is verified by comparison with standards.
Nuclear magnetic resonance spectroscopy	Various	Many experiments possible with one instrument.	Quantitation difficult Moderate sensitivity, requires a lot of the sample. High cost.
X-ray diffraction	-------	Enables identification of inorganic and organic excipients, and crystallinity.	Extremely high cost, skilled operator required.

8.2 Materials and Methods

Figure 8.2 shows the steps involved in two-tier screening analysis. Complete sample preparation can be referred to related sections as given.

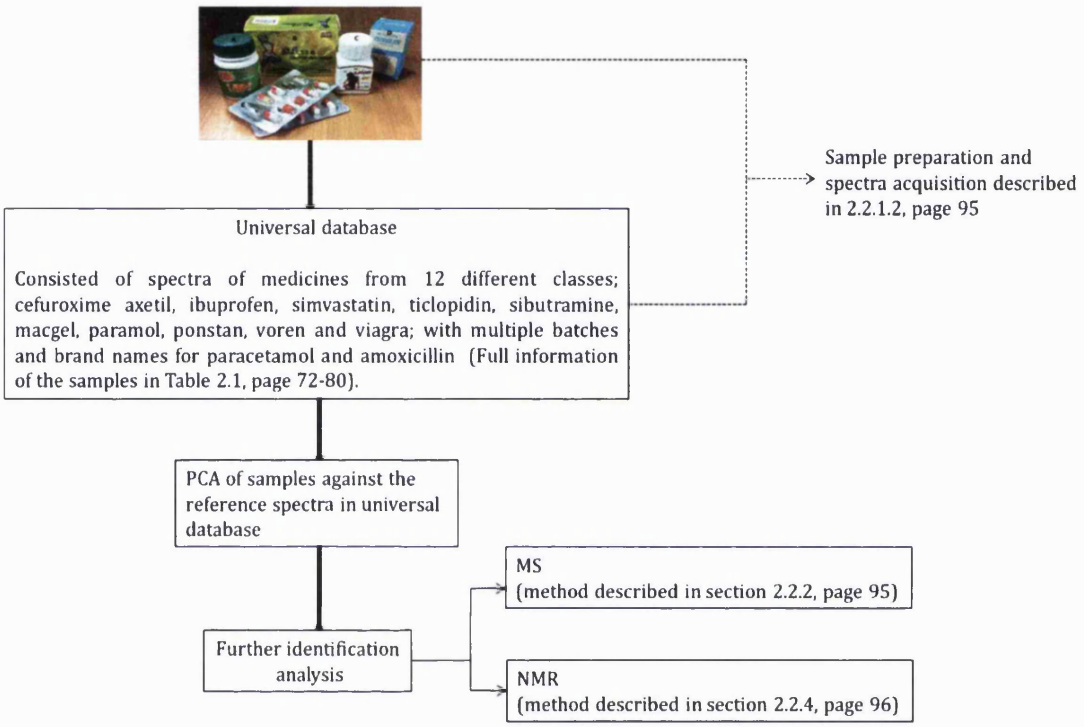


Figure 8.2 The flowchart showing two-tier screening procedure used in this work and the references to sections describing full method of sample preparation.

8.2.1 Materials

Five sets of adulterated samples (Figure 8.1) and one set of counterfeit Viagra were received from representatives of the NPCB, Malaysia, on two separate occasions in 2008 and 2010. These samples were seized in different raids and their details are listed in Table 8.2.

8.2.1.1 Database Reference Spectra

Medicines from 12 different classes; cefuroxime axetil, ibuprofen, simvastatin, ticlopidin, sibutramine, macgel, paramol, ponstan, voren and viagra; with multiple batches and brand names for paracetamol and amoxicillin, were included as the reference spectra in the library. All samples were either analysed

as intact tablets or in the powdered form (refer to the sample list in Chapter 2 for full chemical names of the samples).

8.2.2 Database Search

Raw NIR spectra were directly imported into Spectral ID v9.0 using the smart-convert program available in the database. All spectra were baseline correctedandentered into the database with other information including brand and proprietary names, batch numbers, expiry dates, manufacturer names and addresses, sample origins, other excipients (where available) and sample descriptions.

The database search was conducted using the correlation algorithm for full spectrum search. The hit-list of the top matches was provided together with hit quality index (HQI) values. Low HQI values between the unknown spectra and spectra in the library indicated a good match.

Name	Sample type	Listed ingredients	Indications	Dosage forms	Distributor	Details
Kintop Capsule	Herbal prep.	Ho Shou Wu/ Polygonum multiflorum 150 mg Fructus Crataegi/Crataegus pinnatifida 80 mg Polygonatum Sibiricum Rhizoma 150 mg Rhizoma Rheum palmatum 20 mg	Traditionally used to lose weight and for incontinence	Capsule	SNE Marketing Sdn. Bhd.	Samples were registered with NPCB and display original hologram on box
Jamu Ajaib	Herbal prep.	Schizonepetae Herbs 100 mg Radiz Gyathulae 70 mg Asari Cum Recens Herbs 70 mg Radix Clematidis 100 mg Radix Sileris 65 mg Radiz Dispari 65 mg	1. To reduce back pain, knee pain and shoulder pain 2. To reduce bloating 3. To increase general health and strength 4. To treat leg and hand Numbness	Capsule	Perniagaan Abd. B. Ismail, Kedah Malaysia	Samples had a fake registration number on the bottle and did not have the official hologram
Tunglin		N-Acetyl-aminophenol 500 mg Salicylamide 125 mg Vitamin B1 10 mg Vitamin B6 10 mg	Management of rheumatic disease; rheumatoid arthritis, osteoarthritis, ankylosing spondylitis, gout and painful musculo-skeletal conditions.	Capsule	Rugra Medex Chem, West Germany	Sample was not registered and had no official hologram

Lami	Herbal prep.	Guazumae ulmifolia folium 150 mg Murryyae paniculata folium 100 mg Morinda citrifolia fructus 50 mg Punicae granati cortex 50 mg Curcumae heynianae rhizome 50 mg Parameriae barbata cortex 50 mg Zingiber purpurei rhizome 50 mg	To lose weight and fat	Capsule	Herbalindo SM, Indonesia	Sample was not registered and had no official hologram
Pronoton (May First)	Herbal prep.	Not available	Traditionally used for general health and well-being of men	Capsule	Not available	Samples were received in blister packs. No ingredients or manufacturer information were available.

Table 8.2 List of the suspect counterfeit/adulterated samples confiscated by the NPCB between 2007 to 2010,used in this research.

8.3 Results and Discussion

8.3.1 Identification of adulterated samples

8.3.1.1 Pronoton

Database search

NIR spectra of *Pronoton* samples were acquired and compared against the reference spectra in the database. The first five hits of the search are shown in Table 8.3. The results indicated that the spectra were most similar to a conventional medication containing sildenafil citrate. Sildenafil citrate is a synthetic phospodiesterase-5 (PDE-5) inhibitor that is indicated for the treatment of erectile dysfunction. However, the HQI between both spectra was quite high (0.2119), which made identification by a direct comparison impossible.

This was due to the presence of components in the herbal mixture leading to differences in the NIR spectra. Further observations were made on the PC analysis of the second-derivative spectra, where both spectra, for *Pronoton* and Viagra, clustered together well in the three-dimensional PC space (Figure 8.3). This information was sufficient to deduce that *Pronoton* may have been adulterated with sildenafil or one of its analogues. Further analyses by MS and NMR were conducted in order to confirm the type of adulterant in this product.

Table 8.3 The database hit-list. The database search outcome indicated that the spectra of the *Pronoton* samples were most similar to sildenafil (Viagra).

Hit #	HQI	Sample ID	A.P.I
1	*0.211943*	*viagra01*	*Sildenafil citrate*
2	*0.218108*	*viagra02*	*Sildenafil citrate*
3	0.2635871	voren05	Diclofenac sodium
4	0.3491647	voren07	Diclofenac sodium
5	0.3519343	macgel28	$Al(OH)_3$/ $Mg(OH)_2$

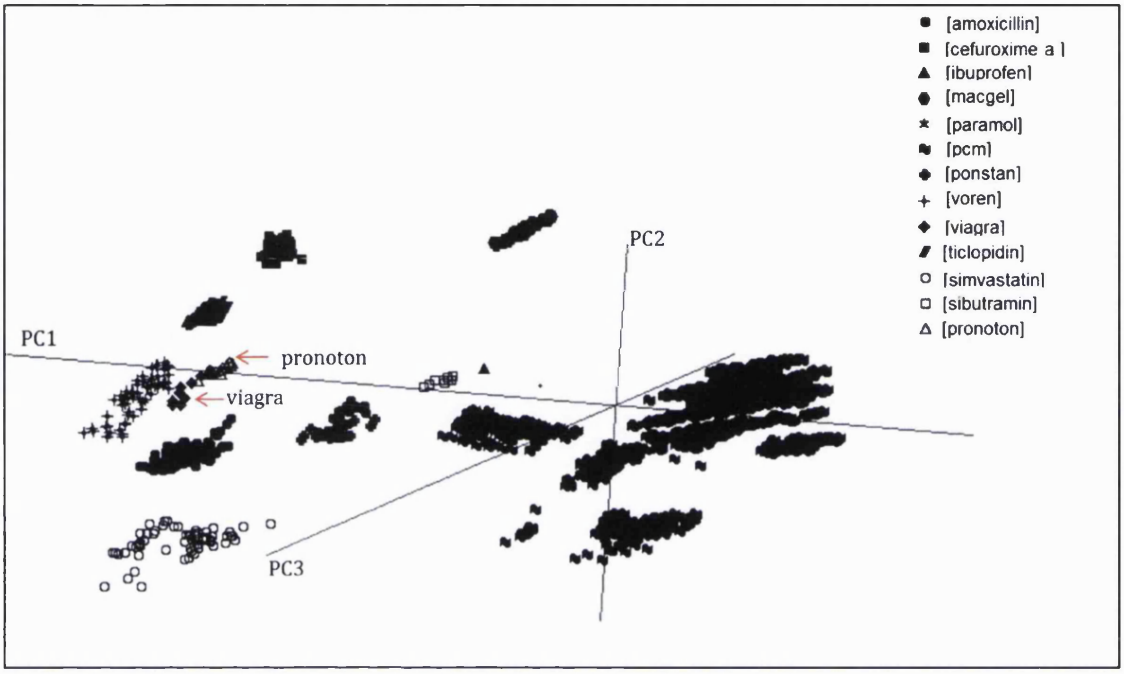


Figure 8.3 PCA score plot showing the spectra of sample *Pronoton* clustered closely to sildenafil citrate.

Mass spectrometry analysis

The MS analysis of *Pronoton (*Figure 8.4A) identified the presence of the main peak of a known sildenafil analogue, hydroxythiohomosildenafil ($C_{23}H_{32}N_6O_4S_2$, MW521.0), with an $[M+H]^+$ion peak at *m/z* 522.07 [171]. This peak was confirmed as the peak for hydroxythiohomosildenafil based on its similarity to the MS/MS fragmentation pattern previously reported [170](Figure 8.4B).

NMR analysis

The assignments of the NMR spectra for the hydroxythiohomosildenafil compounds in *Pronoton* are shown in Table 8.4. The presence of H-29, H-30 at 2.83 and 3.72, respectively, in the ^{1}H-NMR spectra (Figure 8.5) and the two carbon signals at 60.20 ppm (C29) and 58.00 ppm (C30) observed in ^{13}C and DEPT spectra (Figure 8.6) were the markers of this compound and were used to differentiate it from sildenafil and its other derivatives [171]. Based on these analyses, it was further confirmed that the product *Pronoton* was adulterated

with hydroxythiohomosildenafil. The fact that the peaks of hydroxythiohomosildenafil were prominent in the NMR spectrum indicates their high relative concentration in the soluble methanol components, corresponding with the results of the database search and the PCA analysis, providing a good guide for further analysis using other analytical techniques.

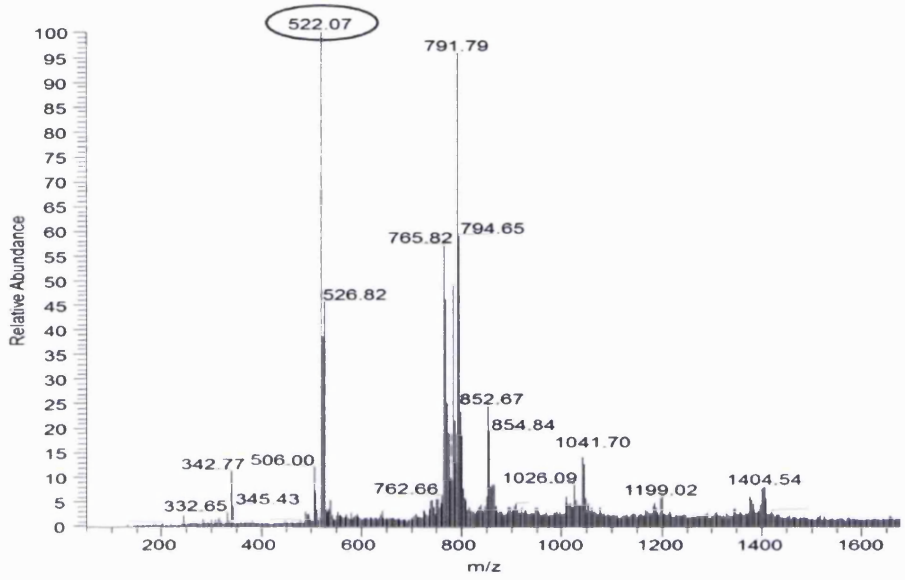


Figure 8.4A ESI-MS analysis of *Pronoton*: full spectrum showing the main peaks at *m/z* 522.07.

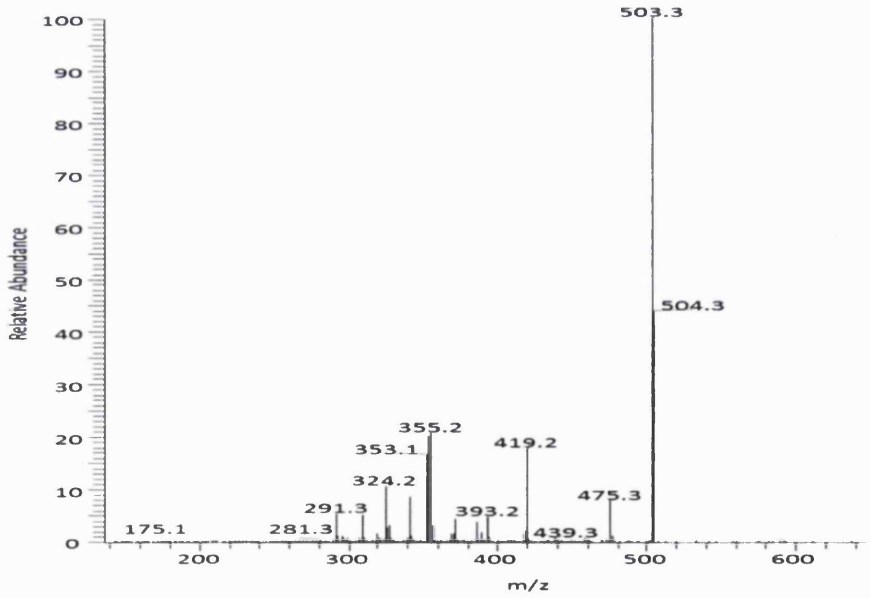


Figure 8.4B MS/MS fragmentation of peak 522.07 with normalized collision energy of 35.

Table 8.4 NMR correlation of hydroxythiohomosildenafil in *Pronoton*.

Carbon no.	^{1}H (ppm)*	^{13}C (ppm)*	DEPT**	COSY	HMQC	HMBC
3	-	146.32	0	-	-	H11,H12
5	-	147.10	0	-	-	H15,H18
6	-	-	-	-	-	-
7	-	174.06	0	-	-	H10
8	-	133.56	0	-	-	H10
9	-	133.90	0	-	-	H11
10	4.46 (3H, s)	39.47	3	-	H10	-
11	2.88 (2H, t)	28.33	2	H11/H12	H11	H12,H13
12	1.82 (2H, m)	23.33	2	H12/H11, H13	H12	H11,H13
13	1.00 (3H, m)	14.49	3	H13/ H12	H13	H11,H12
14	-	122.44	0	-	-	H18
15	8.37 (1H, s)	130.92	1	H15/ H17	H15	H17
16	-	128.15	0	-	-	H18
17	7.91 (1H, dd)	131.60	1	H17/H15. H18	H17	H15
18	7.39 (1H, d)	114.71	1	H18/H17	H18	-
19	-	161.91	0	-	-	H15,H18,H20
20	4.33 (2H, q)	65.18	2	H20/H21	H20	H21
21	1.53 (3H, t)	14.28	3	H21/H20	H21	H20
24, 28	2.98 (4H, br. s)	45.62	2	H24-H28/ H25-H27	H24,28	-
25, 27	3.21 (4H, br. s)	53.18	2	H27-H27/ H24-H28	H25.27	H29
29	2.83 (2H, t)	60.20	2	H29/H30	H29	H30
30	3.72 (2H, m)	58.00	2	H30/H29	H30	H29
OH	1.27 (1H, s)	-	-	-	-	-

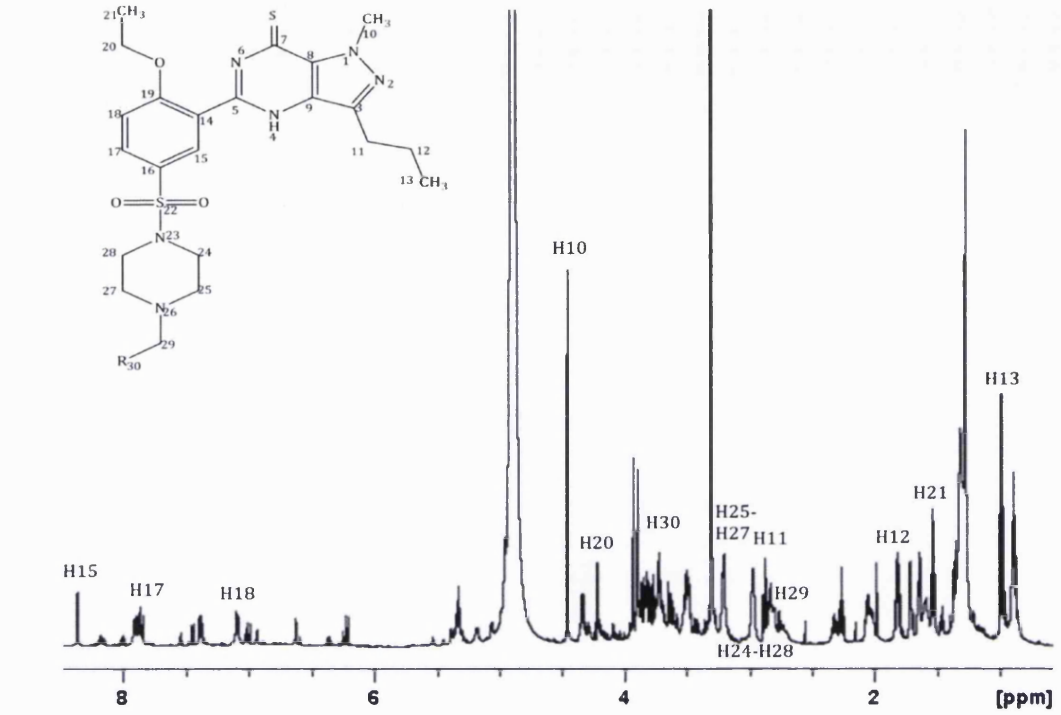


Figure 8.5 ^{1}H-NMR spectrum of hydroxythiohomosildenafil in *Pronoton.* (Insert: chemical structure of sildenafil. R= CH_2OH hydroxythiohomosildenafil).

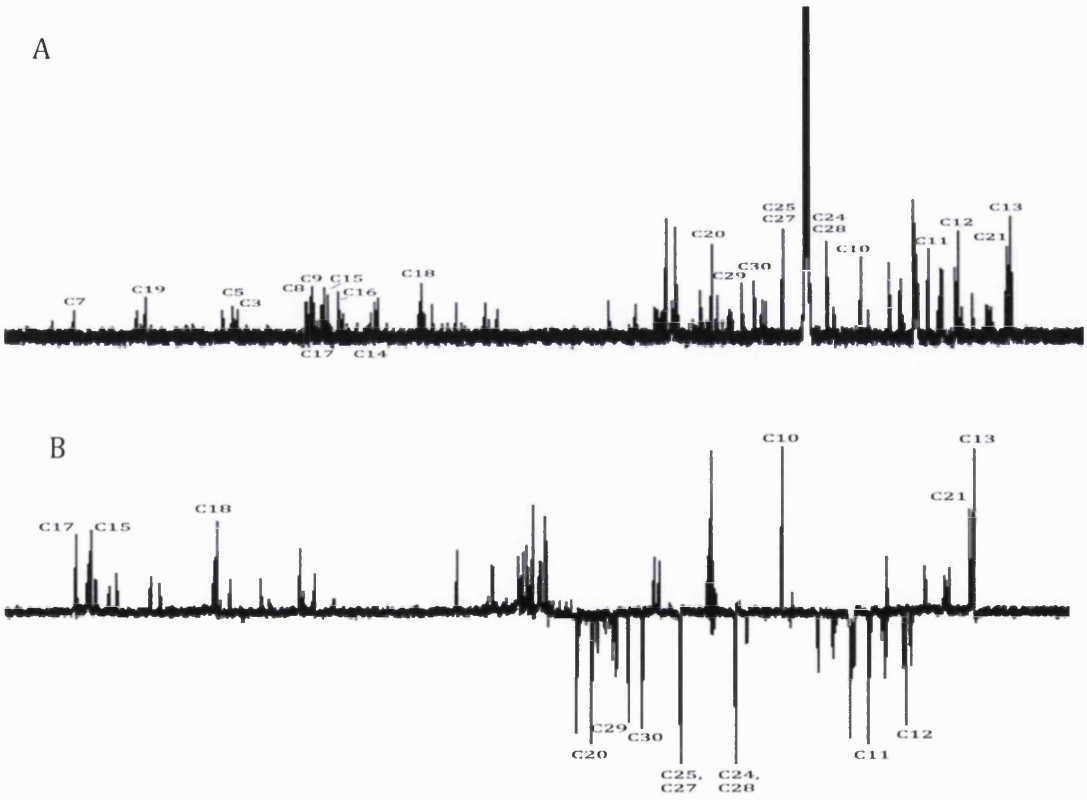


Figure 8.6 ^{13}C (A) and DEPT (B) spectra of hydroxythiosildenafil in Pronoton.

8.3.1.2 Jamu Ajaib

Database analysis

The database search outcome for *Jamu Ajaib* is listed in Table 8.5. The result indicated that the fingerprint spectra of *Jamu Ajaib* have some similarity with the spectra of Macgel and simvastatin, which both contain completely different compounds. Thus, the HQI could not be used to classify this product.

In PCA, the spectra of *Jamu Ajaib* were closely related to reference spectra of *Pontalon* (Figure 8.7). The main active ingredient of *Pontalon* is mefenamic acid and lactose is the main excipients of the medicines [192]. MS analysis was conducted to confirm the actual component of the herbal mixture.

Mass spectrometry analysis

The mass spectrometry analysis in Figure 8.8 shows the presence of peak at *m/z* 707.4; when further fragmented with the 25% normalized collision energy it gave another peak at *m/z*365.2. These peaks indicated the presence of lactose [193]. This finding could be related to the PCA observation, whereby the PC classifications were made based on lactose instead of the possible active components of the herbal formulation. Therefore, it could be deduced that *Jamu Ajaib* contained high amount of lactose with other herbal compounds or other ingredients at very low concentrations or not present at all. The NMR spectra of *Jamu Ajaib* (Appendix 5) confirmed the presence of peaks for lactose within the region 3.2- 4.0 ppm.

Table 8.5 The database hit-list. The database search outcome indicated that the spectra of the *Jamu Ajaib* samples were most similar to Macgel, a preparation containing aluminium hydroxide and magnesium hydroxide.

Hit #	HQI	Sample ID	A.P.I
1	0.3214652	macgel28	$Al(OH)_3$/ $Mg(OH)_2$
2	0.3269735	simvastatin39	Simvastatin
3	0.3321219	macgel27	$Al(OH)_3$/ $Mg(OH)_2$
4	0.3325946	macgel29	$Al(OH)_3$/ $Mg(OH)_2$
5	0.3835777	Sibu01 (sibutramine)	Sibutramin

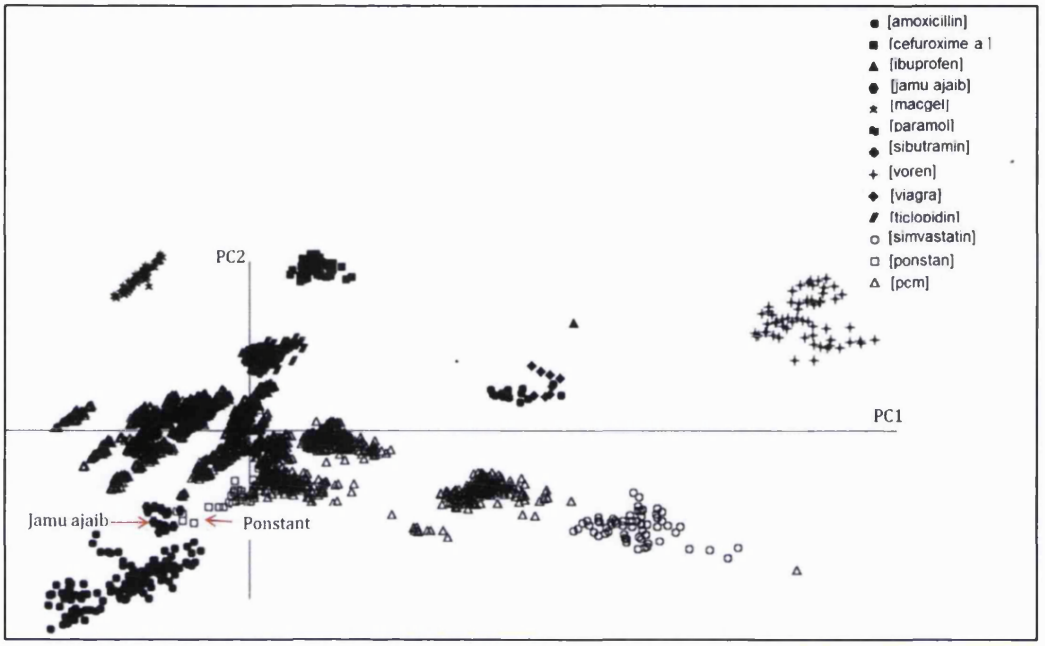


Figure 8.7 PCA score plot showing that the spectra of *Jamu Ajaib* clustered very closely to the spectra of *Pontalon*, which contained mefenamic acid as the main ingredient and lactose as the main excipient.

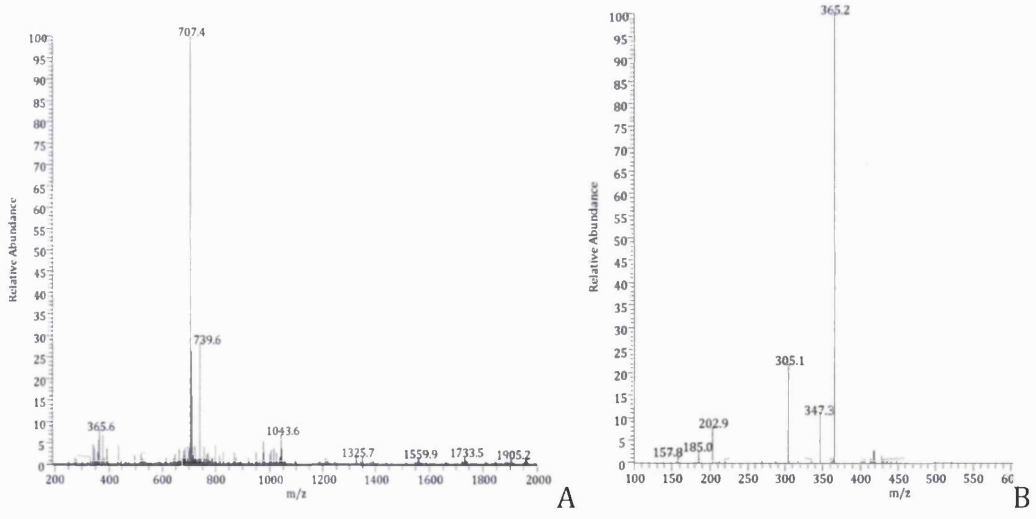


Figure 8.8 ESI-MS analysis of the *Jamu Ajaib* sample in (A) the full spectrum and (B) the MS/MS fragmentation of peak 707.60 @ 25% and the MS/MS/MS fragmentation of peak 365.20@ 22%.

8.3.1.3 Kintop

Database search

The database search for *Kintop* showed the highest similarity to the spectra of diclofenac sodium (Table 8.6). However, the HQI for the comparison of both spectra was higher than the average observed in the other analysis (the average value observed for the other analysis was below 0.4 - Appendix 6). Thus, this observation was invalid.

Figure 8.9 shows the distribution of second-derivative spectra of *Kintop* and the other reference spectra in the three-dimensional plot. One of the closest sets of spectra to *Kintop* was the reference spectra of a set of drugs that contained sibutramine. This provided an early indication that sibutramine may have been added to this product.

Sibutramine is one of the most common synthetic adulterants for herbal products used in weight loss [194]. Due to its opposing effects on peripheral and central sympathetic activity, this compound has been claimed to cause an increase in blood pressure [195] and arrhythmias [196]. It has been banned in most countries since January 2010 [197]. Figure 8.10 shows the chemical structure of sibutramine and its four analogues.

Table 8.6 The database hit-list. The database search outcome indicated that the spectra of *Kintop* were most similar to Voren, a preparation containing diclofenac sodium as the main active ingredient.

Hit #	HQI	Sample ID	A.P.I
1	0.4401639	voren07	diclofenac sodium
2	0.4447224	voren06	diclofenac sodium
3	0.4852104	voren05	diclofenac sodium

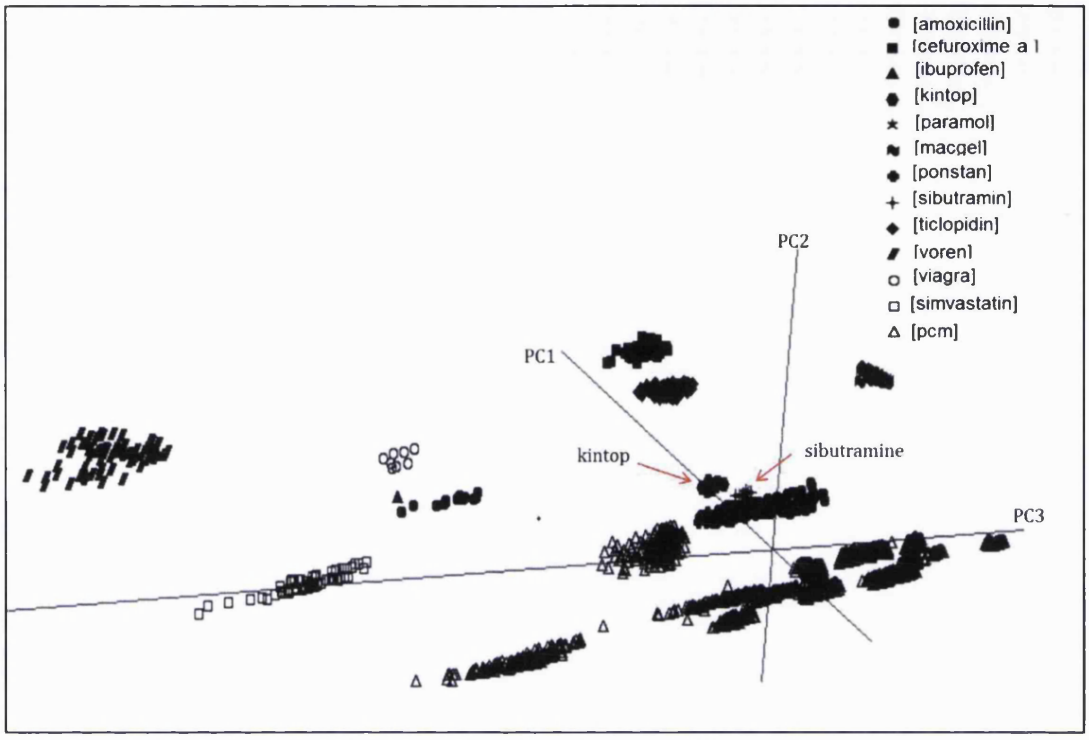
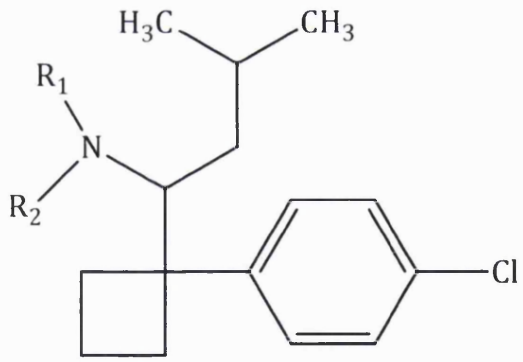


Figure 8.9 PCA score plot showing that the spectra of *Kintop* clustered very closely to the spectra of sibutramine, a conventional slimming product which was banned from the market.



$R_1 = CH_3$	$R_2 = CH_3$	Sibutramine
$R_1 = H$	$R_2 = CH_3$	N- desmethylsibutramine
$R_1 = H$	$R_2 = H$	N- didesmetylsibutramine
$R_1 = C_2H_5$	$R_2 = CH_3$	Homosibutramine

Figure 8.10 Chemical structure of sibutramine and its analogues.

Mass spectrometry analysis

The mass spectrometry analysis of the *Kintop* sample indicated the presence of one of the sibutramine analogues in the herbal mixture (Figure 8.11A). The compound N- desmethylsibutramine has one less CH_3 group than sibutramine (MW 280.18). This compound has a theoretical mass of 265, as observed in the spectra (M^+H^+ = 266) [198]. The fragmentation pattern of the peak at 266.4 (Figure 8.11B) was similar to that observed in a previous work by Zou et al. [199].

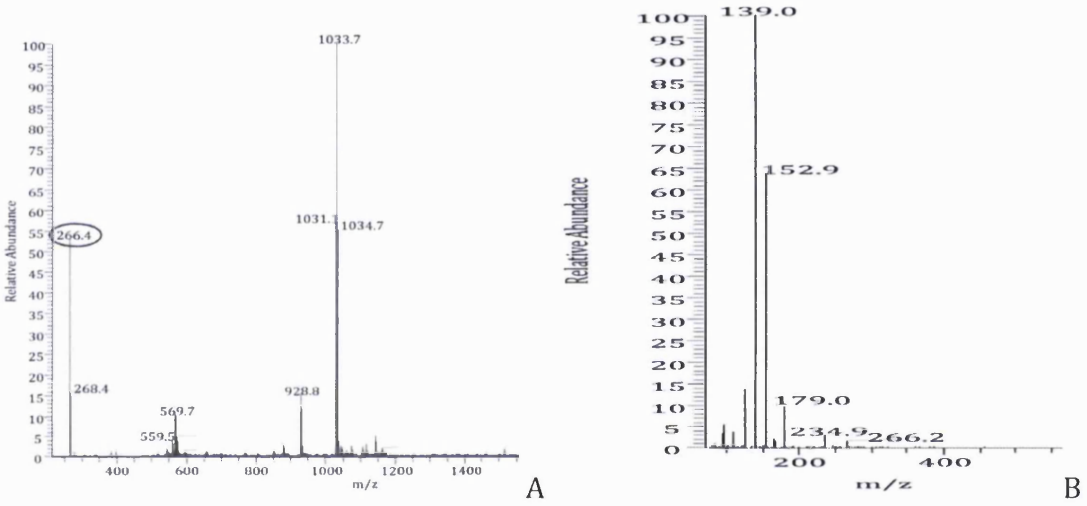


Figure 8.11 ESI-MS analysis of the *Kintop* sample in (A) the full spectrum and (B) the MS/MS fragmentation of peak 266.4 with normalised collision energy of 25%.

NMR analysis

Table 8.7 shows the NMR correlation of the N-desmethylsibutramine compound in *Kintop*. Two of the main differences found between the two compounds were, in the ^{1}H-NMR analysis, the presence of one CH_3 attached to an amine (1.38 ppm) and, in ^{13}C NMR, only three primary carbons were observed in the spectra instead of the four that are supposed to exist in sibutramine compounds (Figure 8.12 and 8.13).

These findings confirmed the presence of a sibutramine analogue in the database. Although the reference spectra of the drugs containing sibutramine were included in the database, the spectral search was unable to point out the similarity between these two products. One of the reasons for this was probably

due to the small amount of active ingredient, where only 10 mg of sibutramine was present in each capsule of the pharmaceutical product. Thus, a more detailed pre-treatment is suggested in order to determine the hidden similarity between the two sets of spectra.

Table 8.7 NMR correlation of sibutramine analogues in *Kintop.*

Carbon no.	^{1}H (ppm)*	^{13}C (ppm)*	DEPT**	COSY
1,2	0.98 (6H,d)	20.0,20.2	3,3	H-3
3	1.95 (1H,s)	27.0	1	H1,2 / H4
4	2.01 (2H,m)	38.6	2	H-5
5	1.28 (1H,br.s)	66.7	1	H-4
6	-	-	-	-
7	2.70 (3H,m)	34.1	3	-
8	-	50.5	0	-
9	2.46 (2H,m)	33.7	2	H-11
10	2.31 (2H,m)	33.4	2	H-11
11	2.54 (2H,m)	16.4	2	H-9/H-10
12	-	141.9	0	-
13,14	7.35 (4H,d)	130.6	1,1	H-15,H-16
15,16	7.43 (4H,d)	129.6	1,1	H-13,H-14
17	-	134.4	0	-

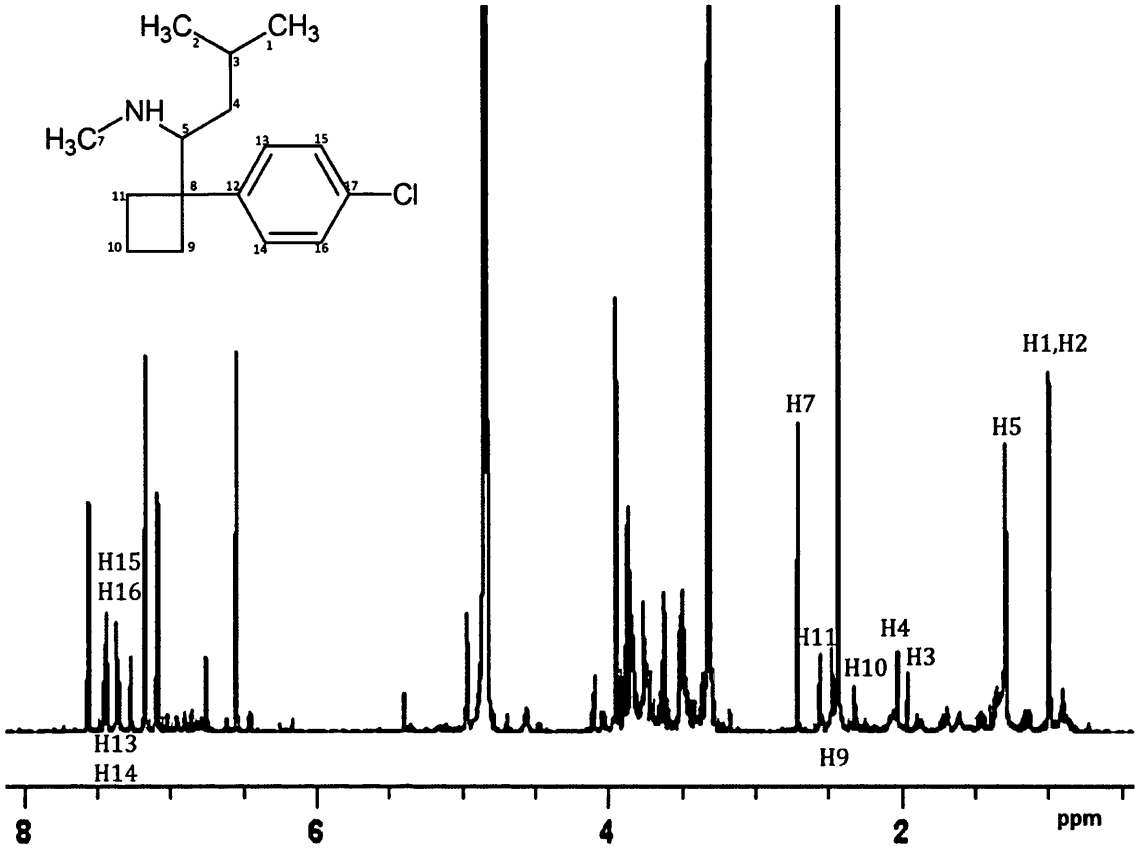


Figure 8.12 1H-NMR spectra for N-desmethylsibutramine in *Kintop* (insert: chemical structure of N- desmethylsibutramine).

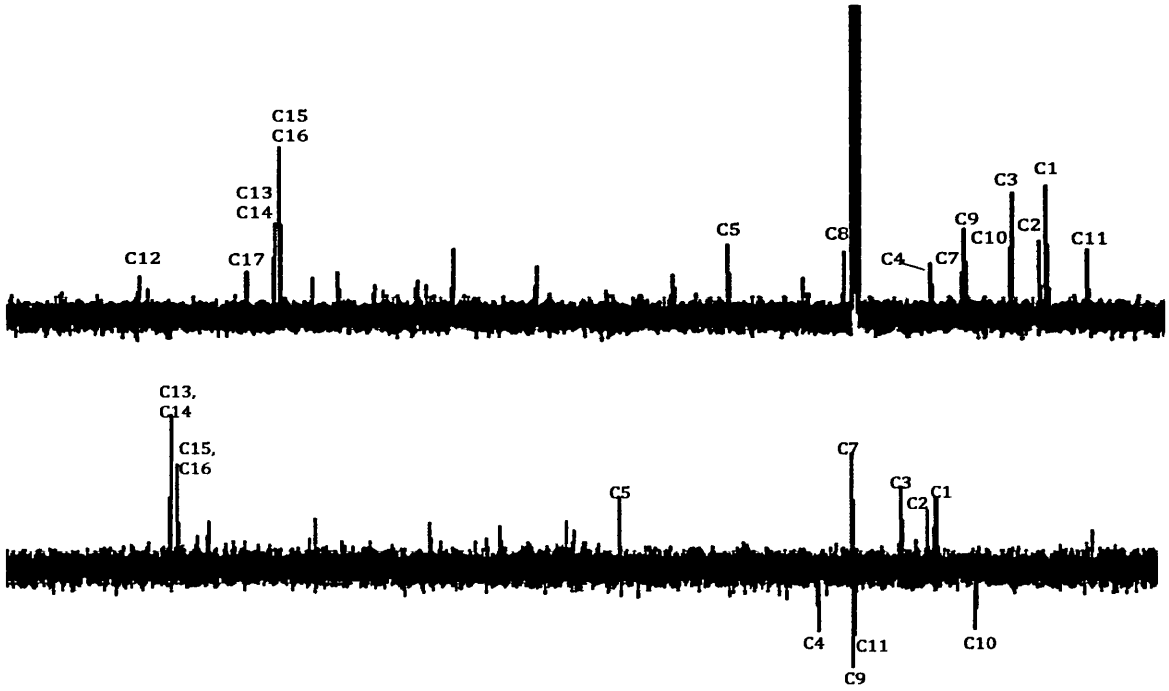


Figure 8.13 ^{13}C NMR (A) and DEPT (B) spectra for the analysis of *Kintop.*

8.3.1.4 Lami

Database Analysis

The database search showed that *Lami* shared some degree of similarity with Macgel (aluminium hydroxide and magnesium hydroxide) and amoxicillin (Table 8.8). A further observation was made on the distribution of the sample in the PCA score plot. The PC analysis of the second-derivative spectra showed that the *Lami* samples clustered well with a set of reference samples with sildenafil citrate as the main ingredient (Figure 8.14).

Table 8.8 The database hit-list. The database search outcome indicated that the spectra of *Lami* were similar to Macgel, a preparation containing aluminium hydroxide and magnesium hydroxideas the main active ingredient.

Hit #	HQI	Sample ID	A.P.I
1	0.3877507	Macgel28	$Al(OH)_3$/ $Mg(OH)_2$
2	0.3954883	Amoxicillin P	Amoxicillin
3	0.3991583	Amoxicillin E	Amoxicillin
4	0.4035932	Macgel 29	$Al(OH)_3$/ $Mg(OH)_2$

Mass spectrometry analysis

The mass spectrometry analysis revealed that the sample was a multi-component system. The two main peaks identified from the mass spectrum were *m/z* 536.3 and *m/z* 462.3 (Figure 8.16 A). Two compounds were suggested based on the masses observed and the results derived from the database and PC analyses (Figure 8.15).

The first peak of *m/z*536.3 may have belonged to an analogue of hydroxythiohomosildenafil ($C_{23}H_{32}N_6O_4S_2$, MW521.0) [171] with an extra CH_2. This compound had a molecular formula of $C_{24}H_{34}N_6O_4S_2$ and a molecular weight of 535.0. The ESI/MS/MS analysis of this peak (Figure 8.16 B) showed similarities with the fragmentation pattern of hydroxythiohomosildenafil in the literature [170].

The second peak detected was suggested to be oxytetracycline ($C_{22}H_{24}N_2O_9$ MW461) [200]. The MS fragmentation of this compound corresponded to the structure (Figure 8.16C). The peaks identified corresponded to the possible chemical fragmentation patterns of these compounds shown in Figure 8.17 A and B. This finding could be further confirmed by NMR but this was outside of the scope of the present study.

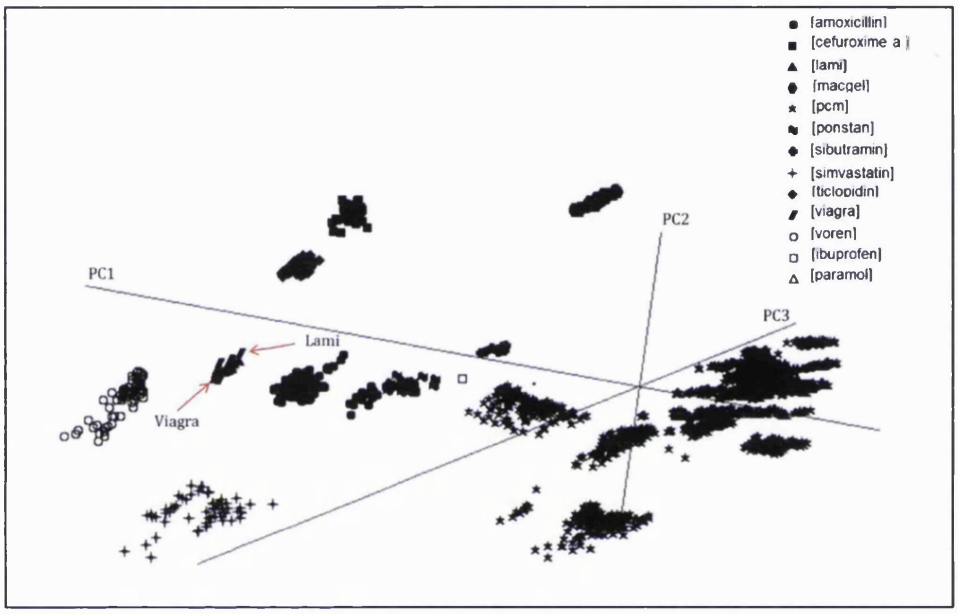


Figure 8.14 PCA score plot. Three-dimensional PC analysis of some of the spectra in the database showing that the *Lami* spectra clustered very closely to the Viagra cluster, which contains sildenafil citrate as the main ingredient.

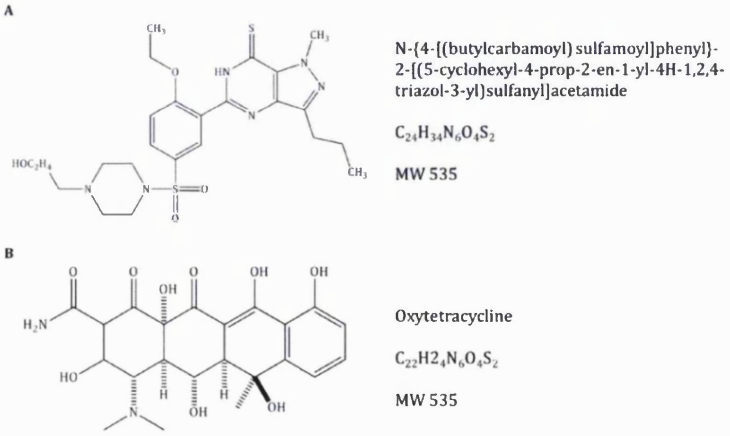


Figure 8.15 The chemical structures of (A) an analogue of hydroxythiohomo-sildenafil and (B) oxytetracycline.

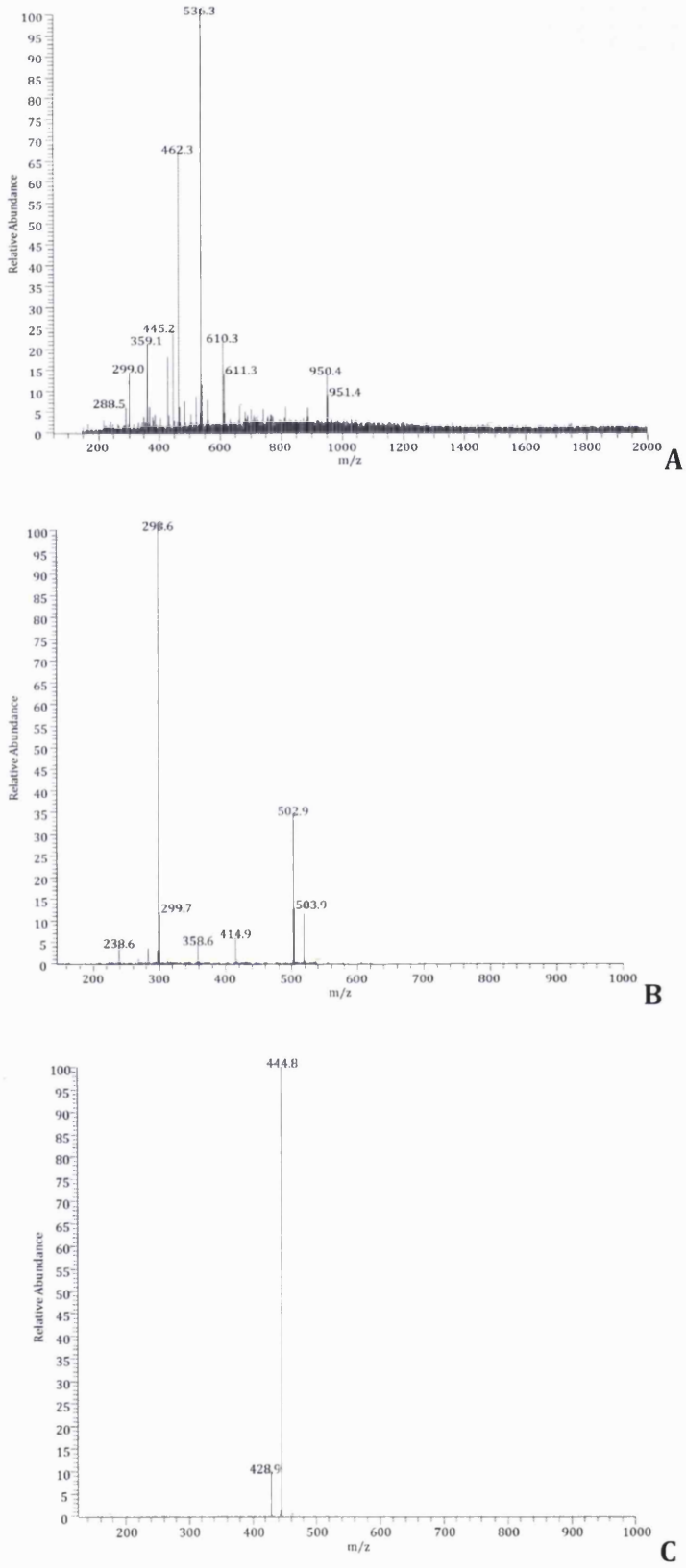


Figure 8.16 MS analysis of *Lami* showing (A) the full spectrum with the main peaks of 536.3 and 462.3 and MS/MS fragmentation of the peaks (B) 536.3 and (C) 462.3.

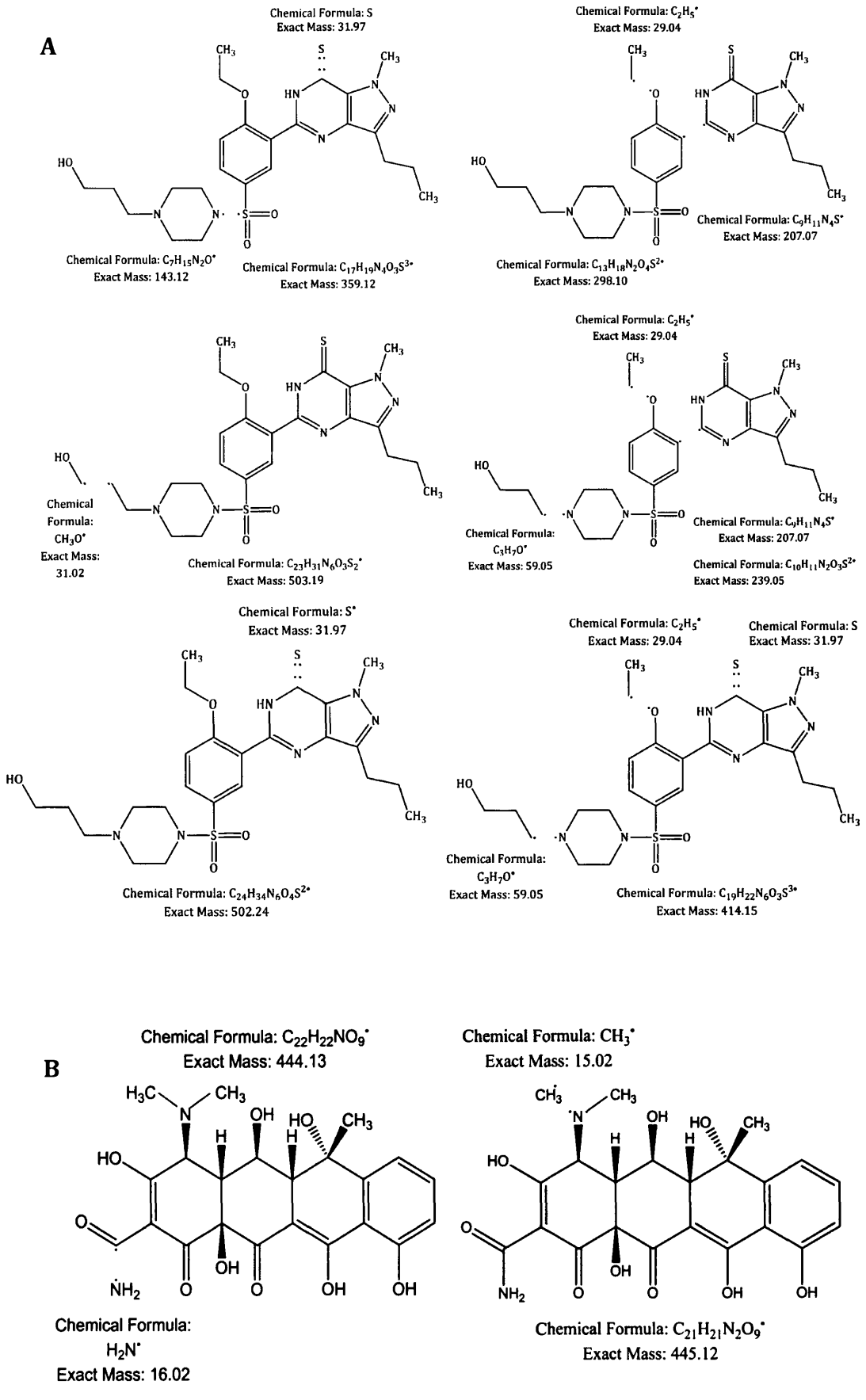


Figure 8.17 Fragmentation patterns of (A) an analogue of hydroxythiohomosildenafil and (B) oxytetracycline.

8.3.1.5 Tunglin

Database Analysis

Tunglin is an anti-rheumatic medication that is claimed to have N-acetyl-aminophenol and salicylic as the main ingredients as stated on the label. Database search indicated similarity of the sample with simvastatin (Table 8.9) while PCA showing a close cluster of Tunglin spectra to paracetamol (Figure 8.18). Although the results from PCA supported this claim, the database search did not share the same finding (Table 8.9).

Further analysis by NMR was conducted to determine the authenticity of this product. Figure 8.19 shows the arrows pointing out the possible peaks representing paracetamol in the 1H-NMR spectra of *Tunglin.* The peaks for the paracetamol compound were not clearly observed as the signals were poor, especially for the aromatic peaks. This indicated some degree of product degradation and this is probably the reason why the database search did not match the spectra of paracetamol and *Tunglin.* Further analysis using COSY experiment may be useful to confirm the connectivity between the aromatic protons. However, this work is not presented here.

Table 8.9 The database hit-list. The database search outcome indicated that the spectra of *Tunglin* were similar to simvastatin.

Hit #	HQI	Sample ID	A.P.I
1	0.2076202	sim39	simvastatin
2	0.2189064	Sibu01	sibutramine
3	0.2713994	sim36	simvastatin
4	0.2771198	mac28	$Al(OH)_3$/ $Mg(OH)_2$
5	0.283187	mac29	$Al(OH)_3$/ $Mg(OH)_2$

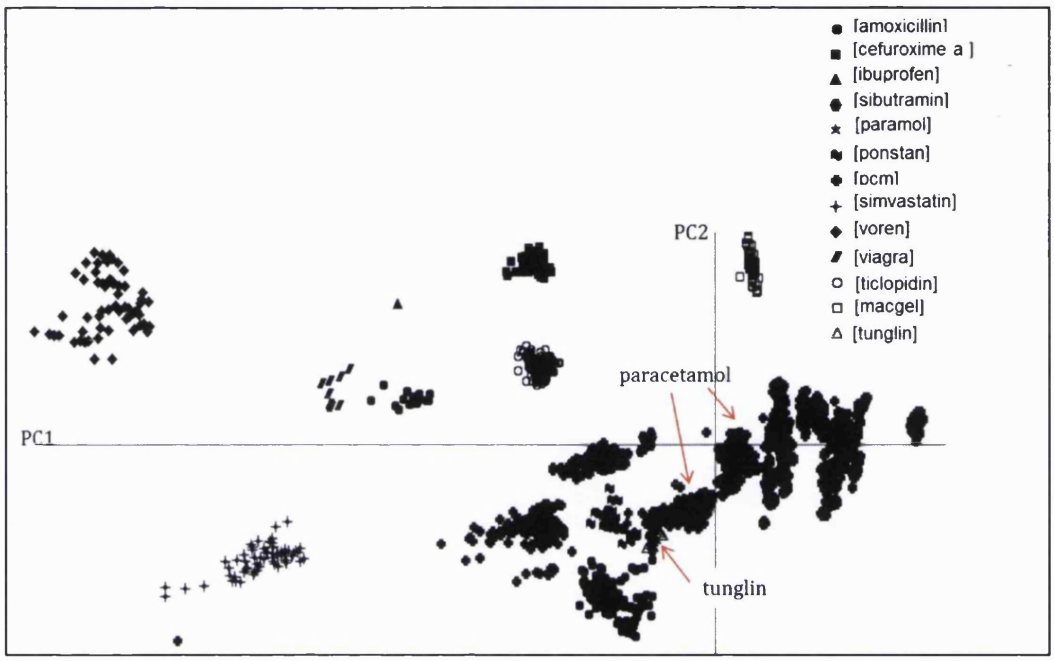


Figure 8.18 PCA score plot showing that the spectra of *Tunglin* sample clustered together with the spectra of paracetamol samples.

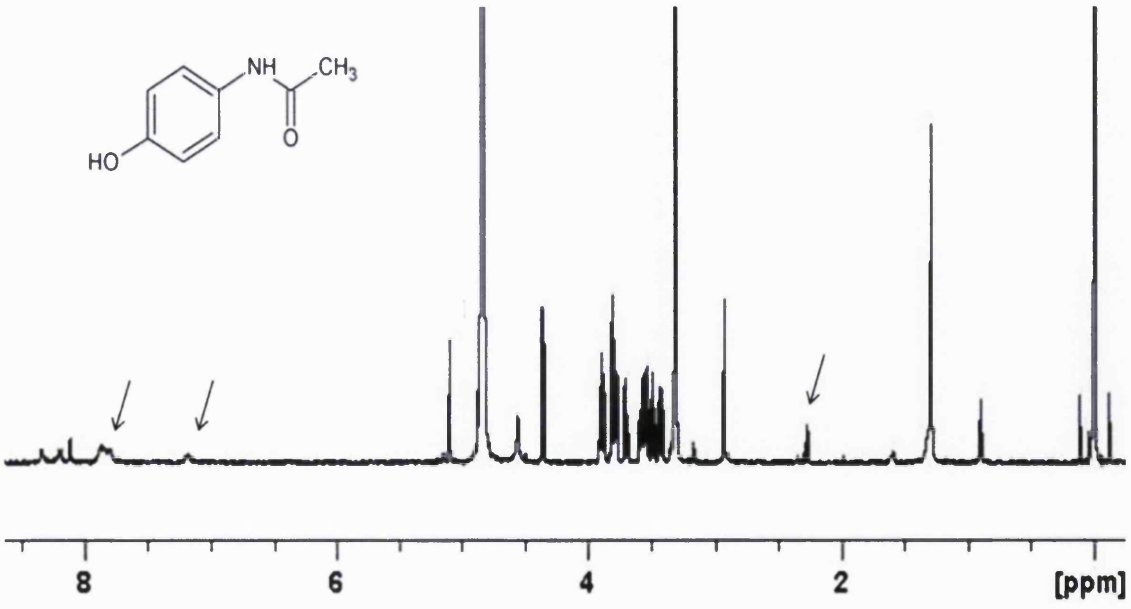


Figure 8.19 ^{1}H-NMR spectra for the sample *Tunglin*. Arrows indicate possible paracetamol peaks, although peak around 7.75 ppm may overlap with other compound.

8.3.2 Identification of counterfeit samples

The identification of counterfeit samples is a slightly simpler analysis than trying to figure out the adulterant in an adulterated sample. In general, samples of the same type and from the same source should have approximately the same fingerprint spectra in order to identify similar types of products in the reference spectra of the database. A cut-off value is an important indicator for differentiating between similar samples from different sources.

Other than this, similar products should cluster closely together in the PCA score plot. It is very unlikely that samples from a similar source would be distributed over a large range in the PC space, unless they were produced from a very poor manufacturing process or the samples and references were from different years and the products since degraded or their formulation was changed.

8.3.2.1 Viagra (Sildenafil citrate)

Database Analysis

Viagra is the most frequently counterfeited product in the whole world [201]. Different techniques have been developed to identify counterfeit Viagra [191]. Table 8.10 shows the search outcome for the spectra of suspected Viagra matched to the spectra of authentic Viagra in the database. A comparison was made based on the full spectrum range (Figure 8.20).

Table 8.10 The database hit-list. The database search outcome indicated that the spectra of the suspect Viagra were similar to the spectra of the original Viagra.

Hit #	HQI	Sample ID	A.P.I
1	0.0908532	Viagra	sildenafil citrate
2	0.1521653	Voren07	diclofenac sodium
3	0.1564789	Voren05	diclofenac sodium
4	0.1800504	Voren06	diclofenac sodium
5	0.3151498	Cef22	cefuroxime axetil

Although both spectra were matched, the hit quality value was high (0.09), indicating that the samples did not come from a similar source. This provided an early indication to show that while both samples had similar contents, they might

not have come from the same origin. The PC analysis (Figure 8.21) showed that the two sets of spectra did not cluster together.

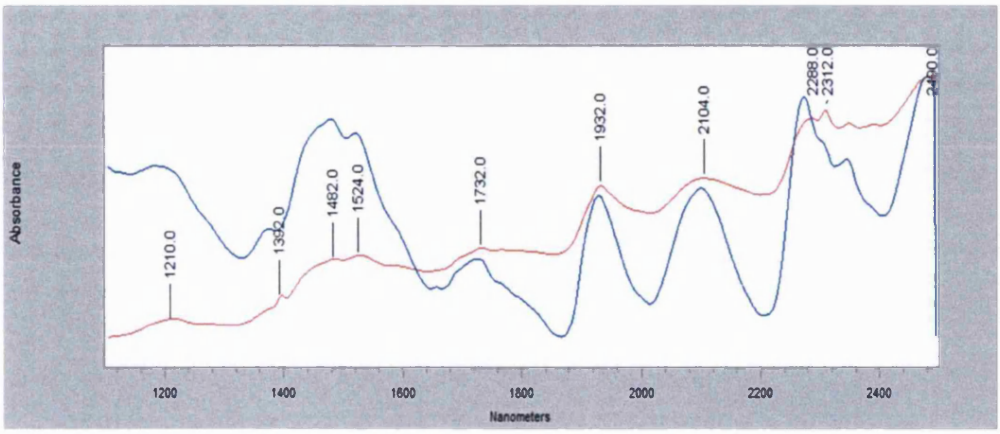


Figure 8.20 Comparison between the spectra of suspect (red) and original (blue) Viagra after baseline pre-processing. Ten major peaks can be seen in the spectral range.

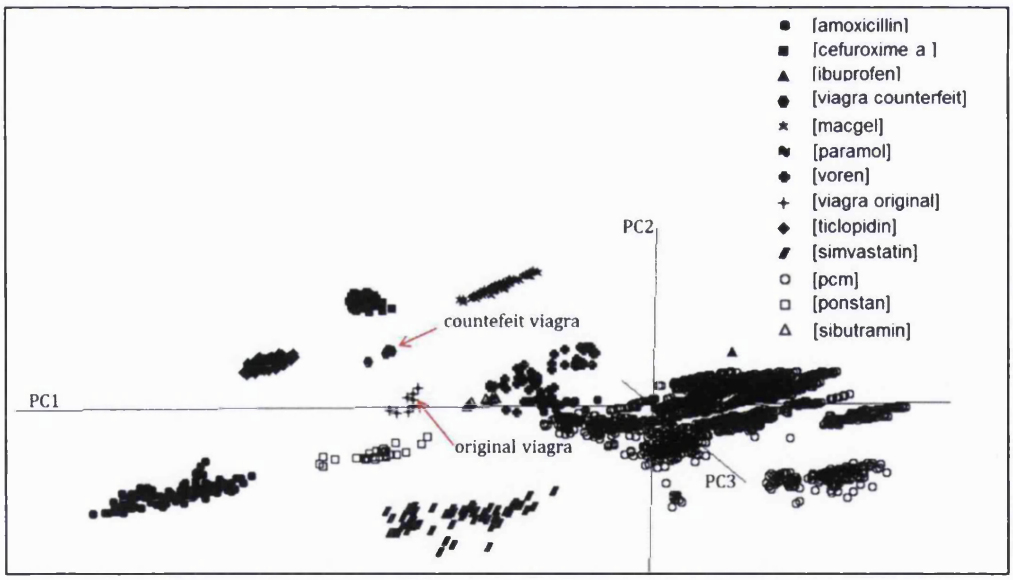


Figure 8.21 PCA score plot showing the spectra of the counterfeit Viagra did not cluster together with the spectra of the original Viagra.

Spectroscopic analysis

The mass spectroscopic analysis showed identical mass spectra for both the authentic and counterfeit samples. The presence of a sildenafil citrate peak was observed at *m/z*475.4 (data not shown). Further observations of the NMR spectra

highlighted the difference between both samples: the intensity of the peaks between region 3 to 4.5 ppm, which belong to the excipients in the formulations, particularly lactose and cellulose. The higher peaks in this region in the adulterated formulation (Figure 8.22) showed that more excipients were used in this formulation compared to the active ingredient. This led to the formulation of substandard medicines.

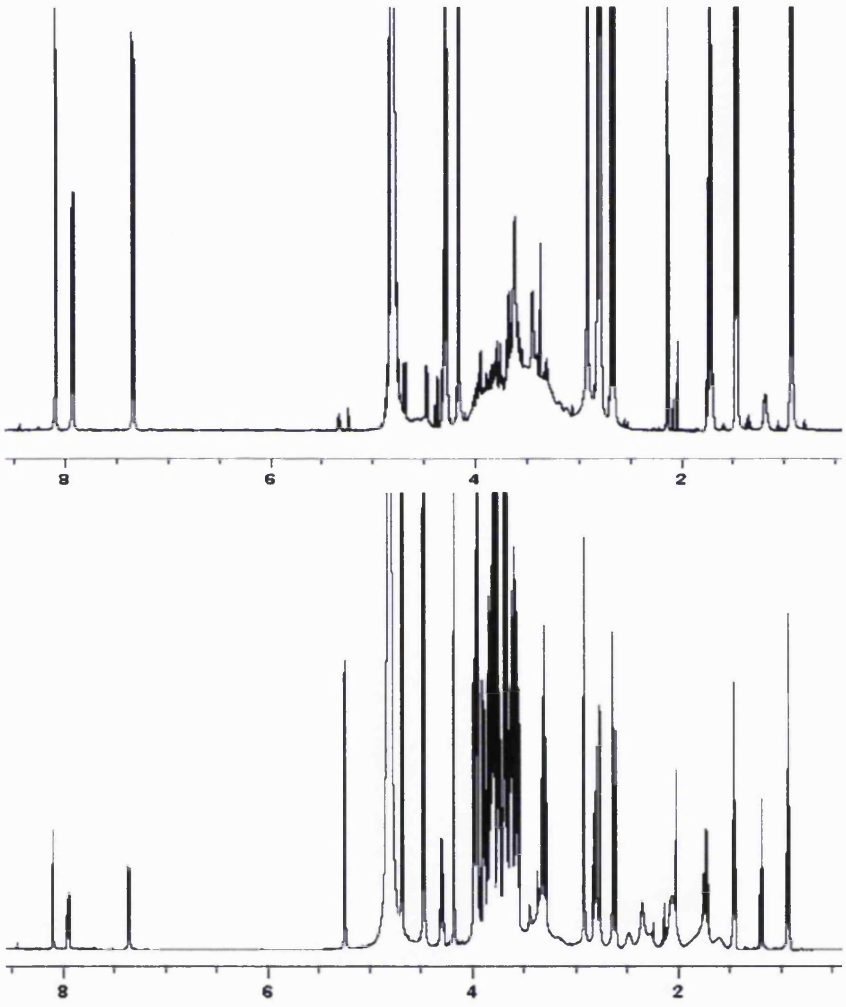


Figure 8.22 1H-NMR spectra for (A) the authentic Viagra (B) the suspected counterfeit Viagra. The major difference can be observed in the spectral range of 3.00 to 4.50 ppm.

8.4 Conclusion

The two-tier screening methods proved to be a suitable technique for quick, simple and cost effective screening procedure for the identification, validation and classification of pharmaceutical and herbal formulations.

Universal database consisted of spectra of conventional medicines was used in these analyses to find similarities with the spectra of seized products. In these cases, the use of cut-off value determined in earlier chapter was not relevant but spectra on the highest rank in the database search outcome was taken into consideration as guide for further analyses.

PCA was used to validate the database search outcome. As both the database and PCA used different ways in calculating the degree of similarity between spectra, the findings may not always be in agreement especially for complex samples like herbal preparation. However, some insights into possible sample compositions were obtained and this is the other advantage of the proposed method.

This was exemplified by the identification of sildenafil citrate in the *Protonon* and *Lami* samples. Identification was not limited to active ingredients only, it also included similarities in excipient contents, as mefenamic acid (*Pontalon)* was a search hit for *Jamu Ajaib* since both products mainly contained lactose. However, for other products, the database searches and HQIs were poor and results were over interpreted.

As expected, as this is only a screening procedure, further analysis by other analytical methods is required in order to identify the sample components or to find the actual chemical differences between samples. However, not all of the samples purchased or confiscated required tedious, extensive analyses for their identification. Therefore, this approach was proven to be a useful option for general screening purposes.

A current limitation of this database is the limited number of samples and their spectra. Expansion of the database will allow a better classification of suspect samples.

CHAPTER 9

CONCLUSION AND FUTURE WORK

9.1 Conclusion

The general objective of this research is to develop a simple, quick and cost effective drug screening procedure that can be used to identify and classify random samples based on composition, brand names or sources without relying on standard active compounds or products from the manufacturer.

A two-tier screening method using NIR spectral database and principal component analysis (PCA) has been developed and its applications for product identification (chapter 5), drug quality study (chapter 6), herbal analysis (chapter 7) and the detection of counterfeit and adulterated medicines (chapter 8) were demonstrated.

The advantages of using the two-tier method was that PCA can provide the graphical presentation of the analysis which was lacking from database searches and it also can be used to validate the database search outcome. However, database search outcome and the observation of the PCA score plot did not always give consistent results as the database and PCA used different methods in assessing the similarity between the spectra. In this case, although the validity of the search outcome cannot be ascertained, additional information of the unknown spectra was gained.

Identification and classification of unknown products are based on the using the spectral database and the hit quality index (HQI) calculated using correlation coefficient algorithm. The lower the HQI indicate higher similarity of the unknown with the reference spectra. Further classifications were made based on cut-off points determined.

Spectra similarity in the PCA score plot was observed based on the 95% confidence interval ellipse and distance of the spectra in the PCA score plot. PCA

was conducted using latent variables which were the principal components (PC), derived from the maximum variability of the spectra set. In this research, the PCA score plot was observed based on the variability presented by PC1 and PC2 only and not the whole spectra.

Some of the advantages of this technique compared to other NIR methods are that it has allowed samples to be identified without known background information which is difficult to do using other qualitative NIR techniques that require reference products to be included in the reference library for comparison. This is particularly useful in analysing drugs that were dispensed without proper labelling (dispensed in paper or plastic packaging instead in the original packaging) and more importantly in identifying counterfeit preparations which have fraudulent labelling.

Also, most of the other methods required the development of PCA models for each type of medicines in the library and these models need to be updated each time new spectra is included in the library. This approach was not used in the two tier screening method whereby a good spectra can be directly included in the incremental spectra database and spectral search were conducted on the spectra itself without the need of developing models.

The method is also useful for 'down-sizing' the number of suspected samples, selecting only the appropriate sample for further extensive analysis by other analytical methods and able to give some indication of the undisclosed components adulterating herbal preparations using the universal database.

Initial investigation on NIRS spectra acquisition proposed standard procedures and criteria for quality of spectral fingerprints of products from different dosage forms before incorporation in the spectral database. This has led to the development of validation procedures for solid, powder and liquid samples. The main observations from the preliminary studies were used to guide the database development:

(I) NIRS has well-known advantages of analysing solid samples with minimal or none sample preparation. Each set of tablets have its own physical characteristics which this information is contributed into the spectra acquired. Thus, the score plot distribution generated from the principal component

analysis of the original spectra showed distinct classifications between different brands of products. Furthermore, crushing a solid preparation into powder form would eliminate the physical characteristics of each product. This has resulted in poor products clustering between different brand names of similar type of medicines.

(II) Powdered samples were best analysed in a glass vials. Although the spectra produced might have the representation of the glass vial, this effects will be ignored when all the related samples were prepared and analysed using the same methods.

(III) Analysis of liquid materials could be done using few different methods; diffuse reflectance, transmittance and transflectance; and in few different ways; samples in the original plastic/glass bottle packaging or by using a standard glass vials. Different types of reflectors would lead to slightly different spectra produced. Data pre-treatment and PC analysis was observed to be more complicated in liquid transflectance analysis.

The next phase of the research focused on quality of information entered into the database. It was crucial that only spectra of acceptable quality and within the same specifications were included in the database to ensure the validity and reproducibility of the search outcomes. Therefore, consistency of the quality of each spectrum was maintained by adhering to the standard procedures developed for every step in the analysis. These were including:

(I) Standard procedures in handling, storing, preparing and analysis of the samples

(II) Setting up the appropriate analytical procedure, parameters and instruments validation

Throughout the work, quality assurance of the spectral database was built by incorporating the procedures before, during and after spectral acquisition (Table 4.3, page 149).

Table 4.3 The quality control procedures for spectra acquisition by near infra-red spectrometer.

Specifications	Descriptions
Sample preparation	- Solid samples were analysed intact without any sample preparations. -Capsule and liquid samples were prepared in a standard type of glass vials.
Sample acquisition	-Spectra were acquired the exact way for each type of samples
Spectral range	-Spectral range was set between 1100 to 2500 nm (10,000-4,000 cm^{-1})
Performance test	-Performance test of instrument was conducted every day- to validate the instrument parameters like photometric noise and accuracy, bandwidth analysis, and wavelength accuracy. Wavelength linearization checks the internal reference peak position correspond to the nominal positions.
Reference spectra	-Reference using ceramic disc was run every hour during sample acquisition and retained for reference
Temperature & humidity	- Temperature and humidity was monitored and noted every morning and evening
Reproducibility	- Set of common spectra was acquired every time after any adjustment of the instrument being made (e.g. update processing software, change of lamp) to ensure spectra reproducibility.
Sample verification	-Samples identification and classification was validated by PCA and SIMCA analysis.

The outcome of a search conducted depends on several factors such as spectra pre-processing, type of search strategy and most importantly the choice of algorithms used. Seven algorithms available in the database software were evaluated by conducting searches using different types of conventional and herbal medicines.

Analysis of different types of search strategy has shown that searches conducted using full fingerprint of products gave superior outcomes compared to the peak search method. Using the peak search strategy was prone to false positive outcomes as a good match can occur when all the peaks in reference spectrum are found in the unknown spectrum despite that the unknown is actually having more peaks that are not found in the reference.

One of the most common problems observed in the NIR spectra of solid products was the increased sloping baseline which affects proper identification of matching spectra. The first derivative types of algorithms were superior in this case compared to other algorithms as they normalized this baseline effect. However, despite of having this advantage, they normally gave poor classification of other compounds which are not present in the database but have similar structural and spectral features.

Searches conducted using different types of samples showed that among all seven algorithms, the correlation algorithm was the most suitable for both identification and classification purposes. It can be suggested that combination of two search algorithms, correlation and first derivative correlation, can be used as most information about the unknown samples would be obtained.

Further work on the database continued with setting up the cut-off limits that classified the other spectra in the database according to of four categories; matched sample, similar brand/source, similar class/type and different products (Figure 4.2, page 144).

The overall work on the development of the spectral database has led to the proposal of a simple protocol to build a spectral database and its consequent use for sample identification and classification. It was established that all steps in the process carried out for the acquisition of the spectra of an unknown samples have

to be similar to the steps for the samples already stored in the database (Figure 4.12, page 150).

Table 4.2 Proposed cut-off points for pharmaceutical formulation classification based on the degree of similarities and differences between unknown spectrum and spectra in the database.

Classification type	Search outcomes	Cut-off values (HQI)
I	Same batch (match)	<0.0001
II	Same brand/different batch	<0.0100
III	Same type of medicine/ different brand	<0.1000
IV	Different type of medicine	>0.2000

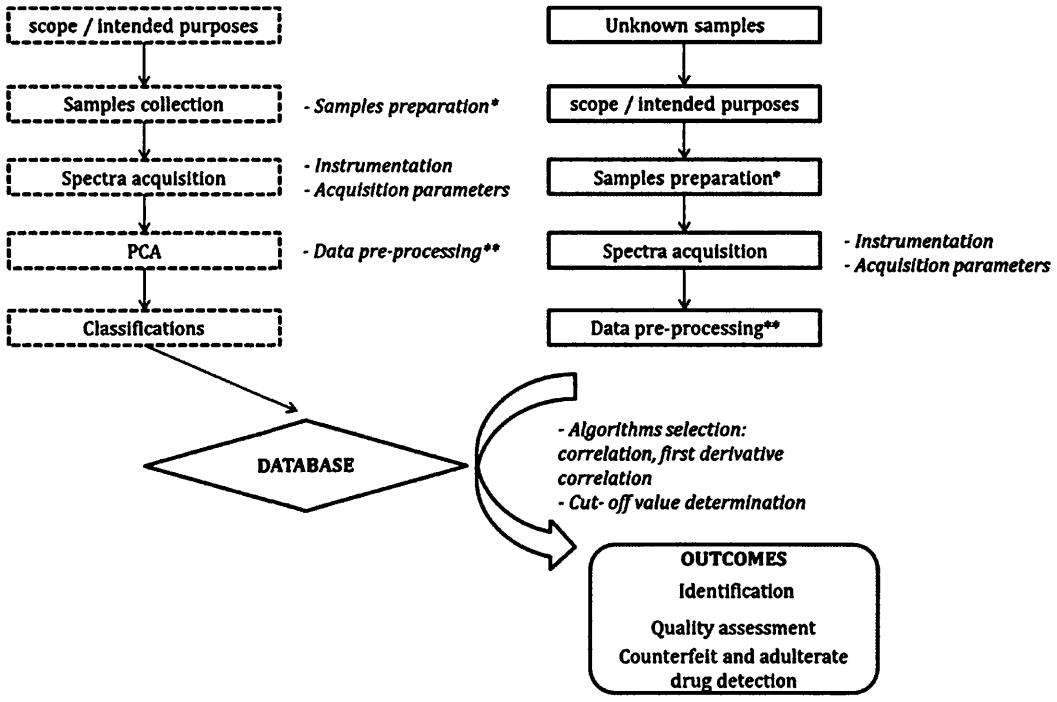


Figure 4.12 The flowchart showing a step-by-step procedure in developing databases and performing searches of unknown products against the database. The 2 main steps in the procedure was including database development (dotted line) and identifying unknown spectra (full line).

The work demonstrated on the NIR spectral database of paracetamol and amoxicillin (chapter 5) has proven that this technique was a reliable, quick screening procedure of products that can be used for product identification. The database also managed to classify samples according to their dose, dosage forms and product variability. This technique has eliminated the need to have standard samples from manufacturer for spectral comparison. This is particularly important for samples that were coming from unknown source or was not labelled appropriately, whereby the active ingredient was unidentified.

The internal and external validation conducted after 12 months proved that the database searches were giving reproducible results. PCA allowed observation of the differences in intra-batch and inter-batch variability for different products. Although some degree of changes to the spectra was observed after a year, the database successfully classified the tablets accordingly using the pre-determined cut-off value. The possible ambiguity issues could be resolved by using either chemometric analysis or other analytical chemistry techniques.

Analysis of amoxicillin samples purchased by random sampling from the Malaysian market was one of the examples to demonstrate that patients do not always receive what they wanted or what they think they should have. A quick analysis using chemometric on the pool of spectra has identified the different set of spectra without relying on reference compounds. This example has also highlighted the potential used of this technique as screening procedure for monitoring the products on the market and to select suspected samples in the market for other more time consuming and expensive analysis.

Another potential application of the spectral database demonstrated in this work was as a method for drug quality study. This could be done in two ways;

(I) Using the spectral database indicated that samples of similar type and composition will normally be represented by the HQI that are very close to each other. If one of the claimed samples has a higher HQI compared to the other samples, there was a higher possibility that this sample can have different properties than the rest of samples. However, further analysis was required to identify the cause of this observation.

(II) Monitoring the distribution of samples in SIMCA analysis could also give some indication about the quality of the samples. Although there are many variables that are affecting spectra classification, measuring the distance of these spectra relative to a specific product can be used as the quality indicator.

A collection of samples from different batches and enriching the incremental database with new samples would allow determination of the cut-off threshold that could classify the sample between the acceptable/non-acceptable qualities. This could be further developed as one of the methods for monitoring pharmaceutical products.

The spectral database also has proven to be useful in analysing more complicated samples like the herbal preparations. Analysis of herbal products were generally more challenging due to their nature and complex compositions. Multilayer analysis using the target and universal database has provided indication not only on the active components but also the hidden compound in the samples.

The spectra classifications of herbal preparations in PCA were less specific compared to other types of conventional medicines due to common chemical compositions shared by different types of herbals and more variation factors in herbal preparation. Despite observed weaknesses in the PCA analyses, the database searches correctly identified samples which contained adulterants based on higher HQI values. One of the ways to improve this weakness would be to select specific spectral regions that are unique for each type of herbal formulation and conducting the search using this fragment of the spectrum rather than full spectrum. The feasibility of this improvement will be tested in future works.

The application of the universal database and PCA also proved to be a promising technique for quick initial screening methods for counterfeit and adulterated drugs detection. Identification was not limited to active ingredients only but also including the composition of the excipients.

The results of these challenging searches carried out on products that were not present in the database had sometimes led to ambiguous answers, especially when complex herbal mixtures were tested. However, some insights about the

possible sample composition were obtained which informed the further analytical work.

However, for other products the database searches, the HQIs were poor and incorrect if results were over interpreted. As expected, the poor HQI required further analysis by other analytical methods to identify components of samples.

Almost 80% of the samples analysed in this work were obtained through convenience sampling; a simple, inexpensive and less structured sampling technique. Although data collected using this technique could not be generalized to describe the quality of medicines or to predict the prevalence of counterfeit, substandard or adulterated medicines in the Malaysian market, the data were sufficient to provide evidence of suitability of this approach in managing this issues.

Some of the main problems observed throughout this work were the quality of the generic products, inappropriate dispensing and labelling of medicines, quality of products sold via the internet, and adulteration of herbal products.

Although there were no counterfeit medicines detected throughout the process, it was not safe to say that the pharmaceutical market were free from these types of products as the numbers of medicines sampled were small and the sampling method was not extended to the whole country.

It has always been the role of the DCA and NPCB to regulate and monitor the safety of medicines and herbal preparations in the pharmaceutical market, this difficult task should be shared especially with the increasing number of pharmaceutical products on the market.

The spectral database developed in this work will provide a platform for developing a centre for pharmaceutical analysis at the National University of Malaysia (UKM). This centre will liaise with other organizations such as the Malaysian DCA, WHO and major pharmaceutical manufacturers in providing support by continuous monitoring of the quality of the pharmaceutical products. This project could provide open source solutions for the third world countries to fight against counterfeit, substandard and poor quality medicines.

An alternative scheduled surveillance procedure was suggested to monitor products on the current market by purchasing samples in pharmacies and or possibly use a handheld device to scan and immediately compare the spectra to the database, followed by the update of the database with a new spectrum. The incremental spectral database can be used as a repository of data for further chemometric analysis. Such repository information has to be secured to avoid potential misuse by counterfeit manufacturer.

9.2 Limitation of study

Some of the limitations of the current work:

(I) The main limitation of the current spectral database was having relatively small number of spectra of both conventional and herbal formulations. Further expansion of the database by adding spectra of different classes of products and increased number of samples for each set of products. This will give a better representation of the spectra and better classification of suspected samples in the database.

(II) All the spectra in this work were acquired using a single NIRS instrument as no access to other instrument was obtained for this work. Thus, the issues of transferability of instrument were not addressed here. However, in order for the database to be widely accepted, it is important to ensure that spectra from other type of NIRS could be matched to the reference spectra in this database.

(III) Classifications of the NIR spectra are made according to the chemical components that are in abundance in the products. Therefore, the unknown spectra search may give ambiguous outcomes for samples that have very similar chemical components with each other. Other than that, the presence of ingredients in small quantities may be hidden. In these cases, the NMR and MS analyses are required to clarify the ambiguity or to provide supplementary information.

9.3 Future work

9.3.1 Optimization of the NIR database

Further study will focus on different spectra pre- processing methods and the feasibility of using selected parts of the spectra for conducting the search.

(I) Pre-processing

Each spectrum in the current database was baseline corrected using the GIFT algorithm. This method was sufficient to removes the baseline slope and offset of the spectra. The effect of using other types of pre-processing techniques like second derivative and SNV on the spectra will lead to different classification of the spectra and thus leading to different database search.

(II) Fragments spectra

Spectra search using selected parts of the spectrum that is unique for each product instead of full spectrum. This is particularly useful for the analysis of herbal preparations as many of them shared similar type of common chemical compositions.

9.3.2 Development of MS and NMR database

As observed in this work, some database and chemometric analysis using the NIR spectra has led to ambiguous results due to overlapping features of chemicals in the NIR. These problems have to be further clarified using other analytical methods.

An alternative to this is a development of database consisting spectra of MS, LC-MS and NMR which can be used to compensate this limitation. These techniques produce more specific spectra/fingerprints for each product. However, both techniques are more costly, time consuming and require tedious sample preparations.

One major concern in the development of a LC-MS database lies in the spectra pre-processing. For every LC-MS chromatogram, retention time shifts occur between chromatographic fingerprints due to column aging, small variations in the mobile phase composition, the flow rate or the temperature. Consequently, the information of the corresponding peaks of two similar samples is not found in the same column. Thus, peak aligning or warping techniques are a necessary data pre-treatment for chromatographic fingerprints prior any other spectra manipulations (Figure 9.1) [202]. In figure 9.2, the flowchart showing the steps taken to process a chromatogram before it can be included in the database [203].

The post-run analysis will allow for the spectral characteristics of each peak for identification.

The collection of NMR spectra was more comparable with the use of internal standard, e.g. tetramethylsilane (TMS). The processed spectra can be analysed without further post-treatment or using AMIX software to exclude the unwanted part of the spectra. Solubility of samples and the choice of solvents are major consideration for NMR analysis.

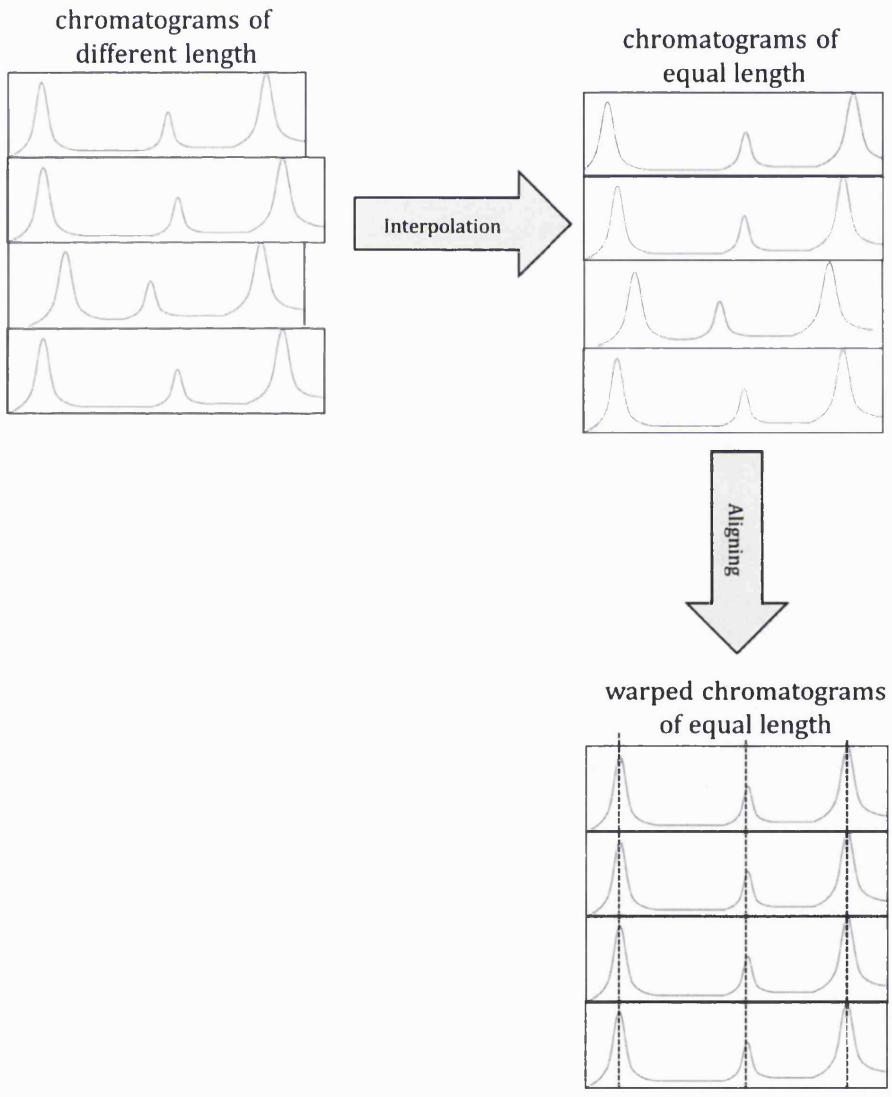


Figure 9.1 Interpolation of MS spectra to obtain fingerprints of equal length and alignment of the corresponding peaks [adapted from: Gong et. al., ref. 203].

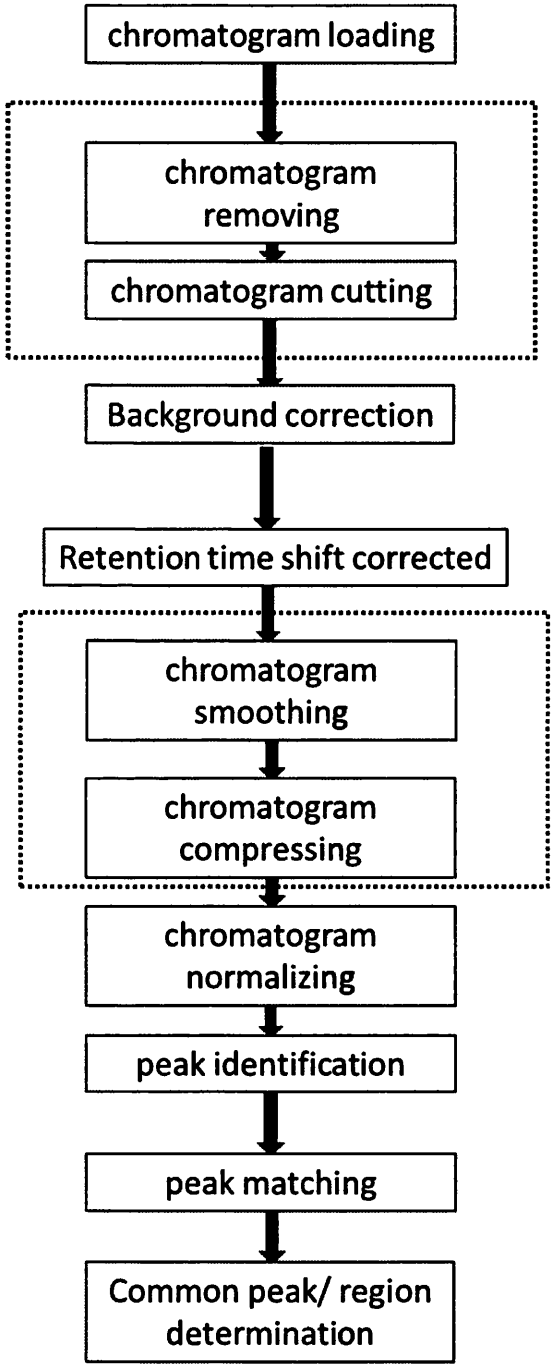


Figure 9.2 Flowchart showing the steps taken to process a chromatogram before it can be included in the database [adapted from: Alaerts et. al., ref. 203].

9.3.3 Development of spectral database using SQL program

The commercial spectral database has proven to be a very good tool in this research. However, the observed potential improvements can be carried out only on a flexible platform that is open to modifications. The next phase of this research is to create customized spectral databases using structure query language (SQL) program using a specific relational database management system (RDBMS) like Oracle, MySQL, and postgres SQL. SQL is a computer program that acts as the language or commands that provide ways to create tables, insert data, select data, delete data, update data, join table, create table schemas, defined functions etc. Other representations like SMILES and SMARTS can be used to record chemical structure information. The information entered into database can include images of the packaging as well as the formulations as well.

One of the main reasons that customized spectral databases are more preferable compared to the current database software used is to increase specificity of the database. This means that we would be able to exclude any functions in the current software and include or add only the desirable functions of the database. For example, some of the algorithms that have proven not to be useful in the database search could be excluded from the system. It is expected that with more specific functions, the database operation will be simpler.

Another reason for this customization is because of the interest to incorporate some chemometric functions like PCA and SIMCA in the database. This is to complement the database and can be used for spectra verification as demonstrated in the earlier chapters. The credibility of the database will be enhanced with the combination of the two functions; database search and chemometric analysis in a single program.

The first step in building a database using SQL is to create the schema that will represent the entity relationship diagram (ERD) of the database. A schema is a collection of tables and collection of schemas that will form a database. Based on this ERD, specific commands will be created for each relationship in the database schema using SQL. The SQL commands are instructions used to communicate with the database to perform specific task that work with data. These commands function in both areas; database creation and database search.

SQLcommandsaregrouped into four major categories depending on their functionality (Table 9.1) [204].

Table 9.1 The examples of SQL commands in four categories according to their specific functions.

Categories	**Functions**	**Example of commands**
Data Definition Language (DDL)	To create, modify, and dropping the structure of database objects.	CREATE, ALTER, DROP, RENAME, TRUNCATE
Data Manipulation Language (DML)	To store, retrieve, modify and delete data.	SELECT, INSERT, UPDATE, DELETE
Transaction Control Language (TCL)	To manage changes affecting the data.	COMMIT, ROLLBACK, SAVEPOINT
Data Control Language (DCL)	To provide security to database object.	GRANT, REVOKE

The ERD representing the main overview of the proposed spectral database system followed by the ERD for the proposed tacking schema and ERD for the database of a single product are illustrated in figure 9.3, 9.4 and 9.5, respectively. These ERDs represent a basis for developing a fully customizable online SQL spectral database with a fully customizable user graphical interface that will be used as a major repository of information of pharmaceutical and herbal products on the Malaysian market and potentially open to submission of information from other regions.

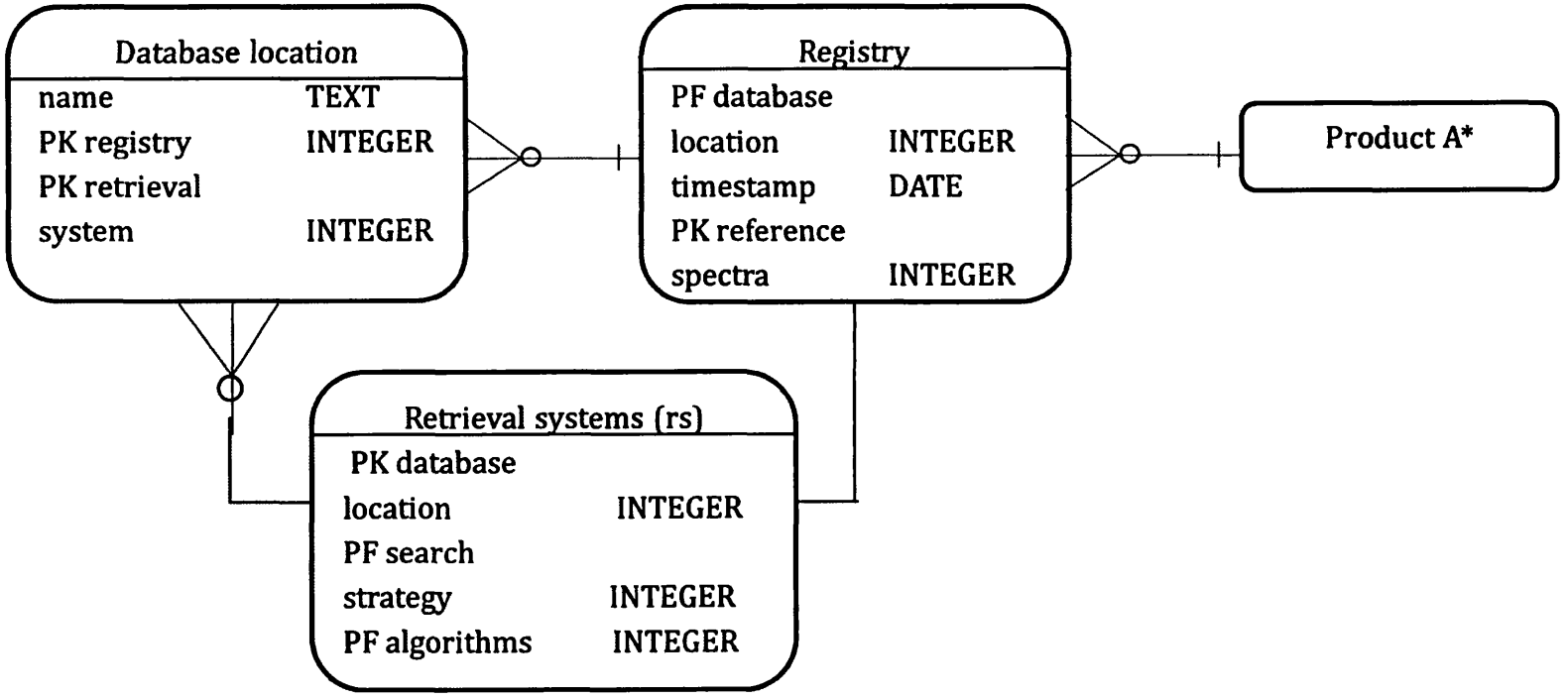


Figure 9.3 Entity relationship diagram (ERD) representing the main overview of the proposed spectral database system.

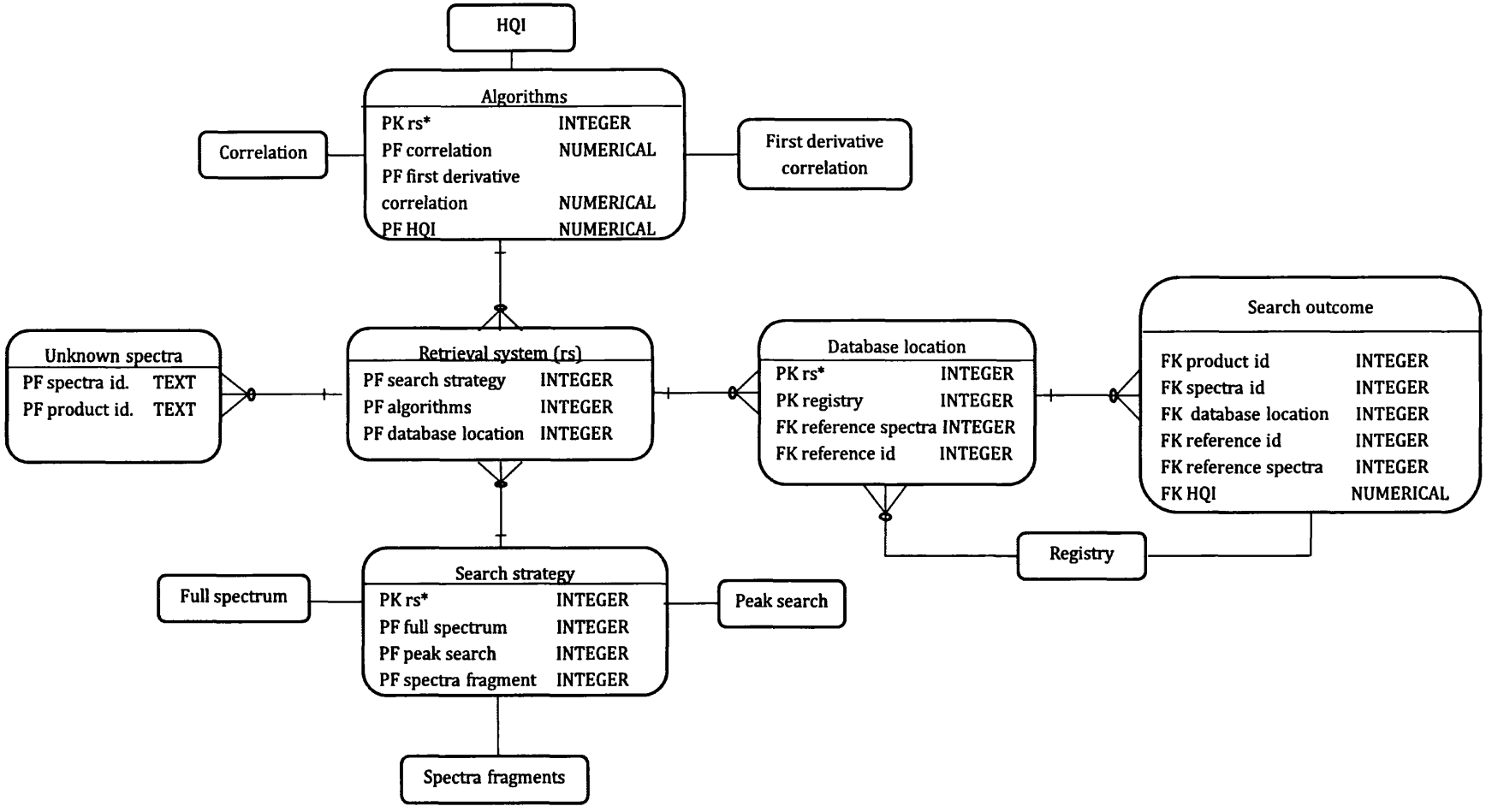


Figure 9.4 ERD for the proposed spectra tracking schema when unknown spectrum is presented to the database.

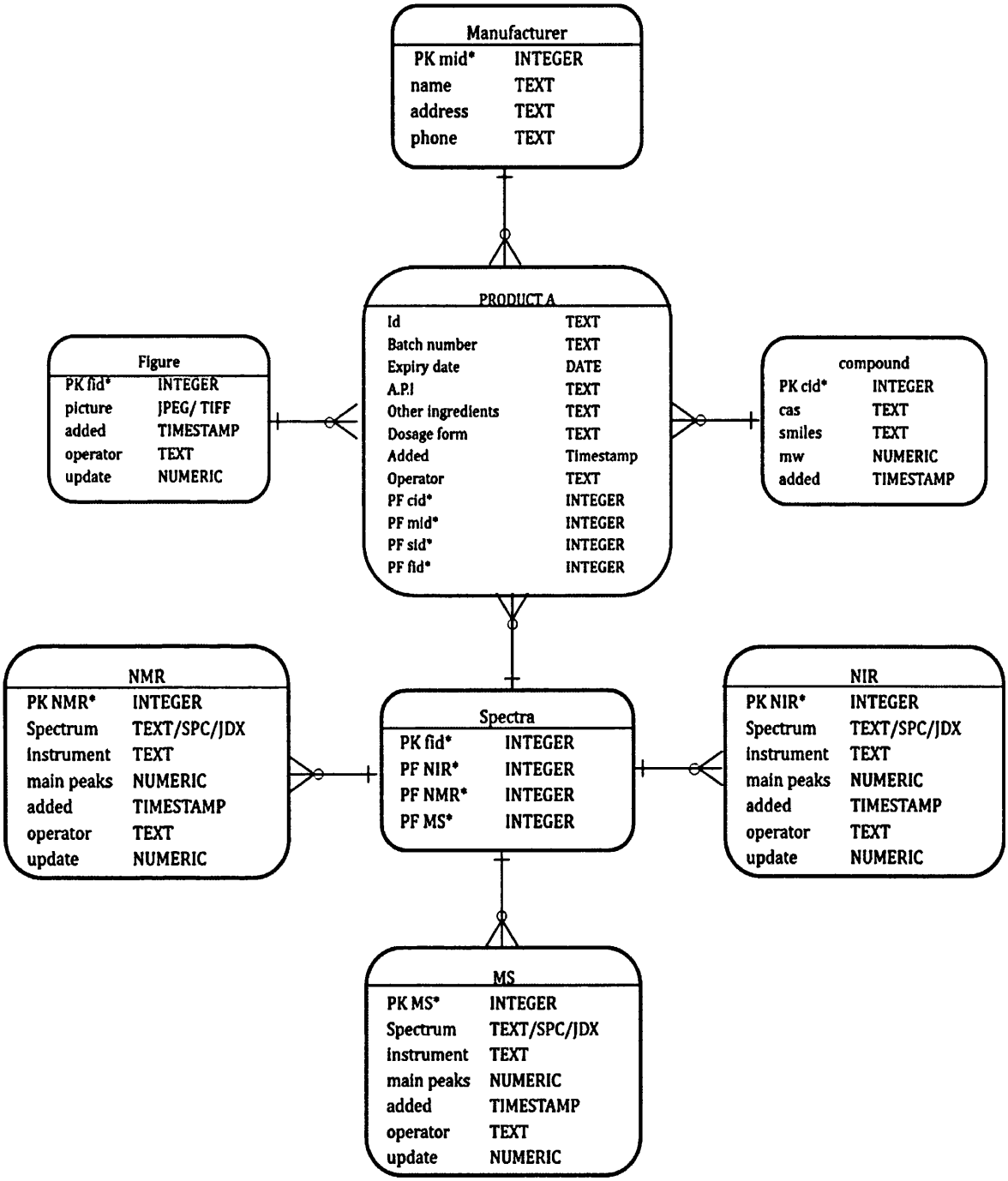


Figure 9.5 Entity relationship diagram for details information about of the product A included in the spectral library.

9.4 List of publications and presentations

Publications

Said MM, Gibbons S, Moffat AC, Zloh M. Near-infrared spectroscopy (NIRS) and chemometric analysis of Malaysian and UK paracetamol tablets: A spectral database study. International Journal of Pharmaceutics. 2011;415(1-2):102-9

Booker AJ, Zloh M, Said M, Suter A, Heinrich M. Metabolomic profiling of saw palmetto products using proton-NMR spectroscopy and multi-variate analysis. Planta Medica. 2011;77(12):1249. [

Said MM, Gibbons S, Zloh M. Chemometrics analysis of solid herbal products and development of spectral databases.Journal of Pharmacy and Pharmacology. 2010;62(10):1203-4. [Poster presentation, British Pharmaceutical Conference, 2010].

Said MM, Gibbons S, Zloh M. Qualitative comparison of Malaysian and UK paracetamol tablets by Near-Infra Red. Journal of Pharmacy and Pharmacology. 2009;61:A112-A3. [Poster presentation, British Pharmaceutical Conference, 2009].

Oral Presentations

Spectra database development of pharmaceuticals & herbal preparations; Malaysian perspective. PhD Research Day 2011. 15 Apr 2011. The School of Pharmacy, London UK.

Chemometric assessment of solid herbal products and development of spectral database. UKPharm Sci 2010. 1-3 Sept 2010, East Midlands Conference Centre, Nottingham UK.

REFERENCES

1. Bate R. Making a killing. Washington DC: The AEI Press; 2008.

2. Ismail M. Prevalence of counterfeit medicines in Malaysia, Pharmacy enforcement perspective.National Pharmaceutical Control Bureau (NPCB); 2009. [accessed 25 Oct 2010]. Available from: http://portal.bpfk.gov.my/aeimages/File/Seminar.../Chapter 12 Counterfeit.pdf

3. Wertheimer AI. Identifying and combating counterfeit drugs. Expert Rev. Clin.Pharmacol. 2008; 1 (3): 333-336.

4. Cockburn R, Newton PN, Agyarko EK, Akunyili D and White NJ. The global threat of counterfeit drugs: why industry and government must communicate the dangers. PLoS Med. 2005; 2(4): e100.

5. Combating counterfeit drugs: a report of the food and drug administration. US Food and Drugs Administration; 2004 [accessed 13 June 2007]. Available from: http://www.fda.gov/doc/initiatives/counterfeit/report02_04.html#scope

6. Forzley M. Combating counterfeit drugs: A concept paper for effective international co-operation; 2006 [accessed 27 Nov 2007]. Available from: http://www.who.int/medicines/counterfeit_conference/en/index.htm

7. WHO. Counterfeit medicines-factsheet; 2010 [accessed 27 May 2010]. Available from: http://www.who.int/mediacentre/factsheets/fs275/en/

8. US Food and Drugs Administration .SEC. 501. [21 USC §351] Adulterated drugs and devices; 2009 [accessed on 25 Apr 2011]. Available from: http://www.fda.gov/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCAct/FDCActChapterVDrugsandDevices/ucm108055.htm

9. Huang Y. Malaysia aims to boost economy with stepped up production of generic drugs. Pharm. Asia News. July 26, 2010.

10. Official website of Pharmaceutical Service Division, Ministry of Health Malaysia; 2011[accessed on 14 Dec 2010]. Available from: http://www.pharmacy.gov.my/index.cfm?&menuid=90&parentid=14

11. Malaysian National News Agency, New bill to help curb counterfeit medicines. Kuala Lumpur: New Straits Times; Sept 22, 2008.

12. Fight fake drugs. Kuala Lumpur: The Star; Sept 28, 2008.

13. Kaur M. Beware of fake medicines. Kuala Lumpur : New Straits Times; Aug 27, 2007.

14. Rahman EA. Malaysia's experience with Meditag.Asia Regulatory Conference on Asia's Role in Global Drug Development, Seoul, Korea; 28-

29 April 2011 [accessed 29 Jun 2011]. Available from: http://www.apec-ahc.org/files/tp201105/Dato_Eishah_A_Rahman_26April_Parallel_Track1_4pm.pdf

15. Counterfeit Medicines: Recommendation on measures to combat counterfeit medicines in Malaysia. Pharmaceutical Association Malaysia (PhaMa); 2006 [accessed 25 Oct 2007]. Available from: http://www.pharma.org.my/pdf document/CounterfeitMed PositionPaper vvfinal March15 pdf.pdf

16. Ministry of Health Malaysia. Traditional medicine manufacturers: Do not adulterate products (press release) [accessed 25 Oct 2010]. Available from: portal.bpfk.gov.my/view_file.cfm?fileid=395

17. Rodionova OY, Pomerantsev AL. NIR-based approach to counterfeit-drug detection. Trac-Trends in Analytical Chemistry. 2010;29(8):795-803.

18. Assi S, Watt RA, Moffat AC. Assay of ciprofloxacin in intact and powdered tablets by near-infrared spectroscopy. Journal of Pharmacy and Pharmacology. 2008;60:A7-A.

19. Assi S, Watt RA, Moffat AC. Identification of tablets through their blister packaging by near-infrared spectroscopy. Journal of Pharmacy and Pharmacology. 2009;61:A31-A.

20. Moffat AC, Assi S, Watt RA. Identifying counterfeit medicines using near infrared spectroscopy.Journal of near Infrared Spectroscopy. 2010;18(1):1-15.

21. O'Neil AJ, Jee RD, Lee G, Charvill A, Moffat AC. Use of a portable near infrared spectrometer for the authentication of tablets and the detection of counterfeit versions. Journal of near Infrared Spectroscopy. 2008;16(3):327-33.

22. Khan PR, Jee RD, Watt RA, Moffat AC. The identification of active drugs in tablets using near infrared spectroscopy. Journal of Pharmaceutical Science.1997; 3: 447-453.

23. Yoon WL, Jee RD, Charvill A, Lee G, Moffat AC. Application of near-infrared spectroscopy to the determination of the sites of manufacture of proprietary products. Journal of Pharmaceutical and Biomedical Analysis. 2004;34(5):933-44.

24. Chong X-M, Hu C-Q, Feng Y-C, Pang H-H.Construction of a universal model for non-invasive identification of cephalosporins for injection using near-infrared diffuse reflectance spectroscopy.Vibrational Spectroscopy. 2009;49(2):196-203.

25. Chong X-M, Hu C-Q, Feng Y-c, Pang H-H.Construction of a universal model for non-invasive identification of penicillins for injection using near-infrared diffuse reflectance spectroscopy.Vibrational Spectroscopy. 2009;51(2):313-7.

26. Hu CQ, Zou WB, Hu WS, Ma XK, Yang MZ, Zhou SL, et al. Establishment of a Fast Chemical Identification System for screening of counterfeit drugs of macrolide antibiotics. Journal of Pharmaceutical and Biomedical Analysis. 2006;40(1):68-74.

27. Yao J, Shi Y-Q, Li Z-R, Jin S-H. Development of a RP-HPLC method for screening potentially counterfeit anti-diabetic drugs.Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences. 2007;853(1-2):254-9.

28. Pang H-h, Feng Y-c, Hu C-Q, Xiang B-R. Construction of universal quantitative models for determination of cefoperazone sodium for injection from different manufacturers using near infrared reflectance spectroscopy.Spectroscopy and Spectral Analysis. 2006;26(12):2214-8.

29. Feng YC, Hu CQ. Construction of universal quantitative models for determination of roxithromycin and erythromycin ethylsuccinate in tablets from different manufacturers using near infrared reflectance spectroscopy.Journal of Pharmaceutical and Biomedical Analysis. 2006;41(2):373-84.

30. Eustaquio A, Blanco M, Jee RD, Moffat AC. Determination of paracetamol in intact tablets by use of near infrared transmittance spectroscopy. Analytica Chimica Acta. 1999;383(3):283-90.

31. Trafford AD, Jee RD, Moffat AC, Graham P. A rapid quantitative assay of intact paracetamol tablets by reflectance near-infrared spectroscopy. Analyst. 1999;124(2):163-7.

32. Blanco M, Peguero A. Analysis of pharmaceuticals by NIR spectroscopy without a reference method. Trac-Trends in Analytical Chemistry. 2010;29(10):1127-36.

33. Goodman C, Kachur SP, Abdulla S, Bloland P, Mills A. Drug shop regulation and malaria treatment in Tanzaniawhy do shops break the rules, and does it matter. Health Policy and Planning. 2007;22(6):393-403.

34. Blanco M, Coello J, Iturriaga H, Maspoch S, de la Pezuela C. Near-infrared spectroscopy in the pharmaceutical industry. Analyst. 1998;123(8):135R-50R.

35. Luypaert J, Massart DL, Heyden YV. Near-infrared spectroscopy applications in pharmaceutical analysis.Talanta. 2007;72(3):865-83.

36. Cogdill RP, Anderson CA and Drennan JK.Using NIR spectroscopy as an integrated PAT tool.Spectroscopy. Dec 2004; 19(12): 104-109.

37. Li H. PAT- Practical approaches to implement NIR in pharmaceutical processes; 2007.[accessed on 27 September 2011]. Available from: http://www.cvg.ca/Presentations/2007/PAT%20Solution.pdf

38. Blanco M, Romero MA. Near-infrared libraries in the pharmaceutical industry: a solution for identity confirmation. Analyst. 2001;126(12):2212-7.

39. European Pharmacopoeia, Council of Europe. Strasbourg 4rd edn, 2001.

40. Broad N, Graham P, Hailey P, Hardy A, Holland S,Hughes S, Lee D, Prebble K, Salton N and Warren P. Guidelines for the development and validation of near- infrared spectroscopic methods in the pharmaceutical industry. In: Chalmers JM and Griffiths PR, editors. Handbook of Vibrational Spectroscopy. Chichester, UK: John Wiley & Sons; 2002.

41. Plugge W, der Vlies V. The use of near-infrared spectroscopy in the quality control laboratory of the pharmaceutical industry.Journal of Pharmaceutical and Biomedical Analysis. 1992;10(10-12):797-803.

42. Plugge W, der Vlies V. Near-infrared spectroscopy as an alternative to assess compliance of ampicillin trihydrate with compendia specifications. Journal of Pharmaceutical and Biomedical Analysis. 1993;11(6):435-42.

43. Blanco M, Coello J, Iturriaga H, Maspoch S, delaPezuela C. Quantitation of the active compound and major excipients in a pharmaceutical formulation by near infrared diffuse reflectance spectroscopy with fibre optical probe. Analytica Chimica Acta. 1996;333(1-2):147-56.

44. John CT and Pixley NC.Methodology for NIR identification of pharmaceutical finished products with emphasis on negative controls and data driven threshold value selection.American Pharmaceutical Review. 2007;10(2): 120–124.

45. Dreassi E, Ceramelli G, Corti P, Lonardi S, Perruccio PL. Near-infrared reflectance spectrometry in the determination of the physical state of primary materials in pharmaceutical production.Analyst. 1995;120(4):1005-8.

46. Dreassi E, Ceramelli G, Savini L, Corti P, Perruccio PL, Lonardi S. Application of near-infrared reflectance analysis to the intergrated control of antibiotic tablet production.Analyst. 1995;120(2):319-23.

47. Gerhausser CI, Kovar KA. Strategies for constructing near-infrared spectral libraries for the identification of drug substances.Applied Spectroscopy. 1997;51(10):1504-10.

48. Reeves JB, Zapf CM. Spectral library searching: Mid-infrared versus near-infrared spectra for classification of powdered food ingredient. Applied Spectroscopy. 1999;53(7):836-44.

49. Bradfield KB and Forbes RA. Development and validation of an analytical method for identification of granulated nicarbazin by near infrared reflectance spectroscopy.Journal of Near Infrared Spectroscopy.1997; 1: 41-65.

50. van der Vlies C, Kaffka JK and Plugge W. Qualifying pharmaceutical substances by fingerprinting with NIR spectroscopy and PQS. Pharmaceutical Technology Europe.1996; 5: 43-48.

51. Dreassi E, Ceramelli G, Corti P, Perruccio PL, Lonardi S. Application of near-infrared reflectance spectrometry to the analytical control of pharmaceuticals: Ranitidine hydrochloride tablet production. Analyst. 1996;121(2):219-22.

52. Shah NK, Gemperline PJ. A program for calculating mahalanobis distance using principal component analysis. Trac-Trends in Analytical Chemistry.1989;8(10):357-61.

53. Warr WA. Spectral databases.Chemometrics and Intelligent Laboratory Systems. 1991;10(3):279-92.

54. ACD/NMR Database. Advance Chemistry Development; 2011 [accessed 3 Feb 2011]. Available from: http://www.acdlabs.com/products/dbs/nmr_db/

55. Hueni A, Nieke J, Schopfer J, Kneubuehler M, Itten KI. The spectral database SPECCHIO for improved long-term usability and data sharing.Computers & Geosciences. 2009;35(3):557-65.

56. Davies AN and McIntyre PS. Spectroscopic Databases. In: George WO, Steele D, editors. Computing Application in Molecular Spectroscopy. Cambridge: The Royal Society of Chemistry: 1995.p. 41-59.

57. Lohninger H. Spectral data bases and interpretation systems; 2010 [accessed on 5 Jan 2011]. Available at: http://www.lohningar.com/spectroscopy/dball.html

58. Antibase 2011. Wiley-VCH [accessed 3 Feb 2011]. Available from: http://www.wiley-vch.de/stmdata/pdf/flyer/Flyer_Antibase.pdf

59. Rasmussen GT, Isenhour TL. Library retrieval of infrared spectra based on detailed intensity information. Applied Spectroscopy. 1979;33(4):371-6.

60. Ruprecht M, Clerc JT. Performance analysis of a simple infrared library search system. Journal of Chemical Information and Computer Sciences. 1985;25(3):241-4.

61. Hanna A, Marshall JC, Isenhour TL. A GC/FT-IR compound identification system. Journal of Chromatographic Science. 1979;17(8):434-40.

62. Lowry SR, Huppler DA, Anderson CR. Database development and search algorithms for automated infrared spectral identification. Journal of Chemical Information and Computer Sciences. 1985;25(3):235-41.

63. Lowry SR. Automated spectral searching in infrared, raman and near-infrared spectroscopy. In: Chalmers JM and Griffiths PR, editors. Handbook

of Vibrational Spectroscopy. Chichester, UK: John Wiley & Sons; 2002.p. 1948- 1961.

64. Euclidean Distance algorithm. GRAMS Suite Help; 2009.

65. Absolute Value algorithm. GRAMS Suite Help; 2009.

66. Lowry SR, Huppler DA. Boolean logic system for infrared spectral retrieval.Analytical Chemistry. 1983;55(8):1288-91.

67. Woodruff HB, Lowry SR, Isenhour TL. A comparison of two discriminant functions for classifying binary infrared data.Applied Spectroscopy. 1975;29(3):226-30.

68. Woodruff HB, Ritter GL, Lowry SR, Isenhour TL. Pattern recognition methods for the classification of binary infrared spectral data.Applied Spectroscopy. 1976;30(2): 213-6.

69. Peak search. GRAMS Suite Help.2009

70. Heller SR. Conversational mass-spectral retrieval system and its use as an aid in structure determination. Analytical Chemistry. 1972;44(12): 1951.

71. WHO. WHO guidelines for sampling of pharmaceutical products and related materials.WHO technical report series No. 929; 2005.

72. Newton PN, Dondorp A, Green M, Mayxay M, White NJ. Counterfeit artesunate antimalarials in southeast Asia. Lancet. 2003;362(9378):169.

73. Bate R, Coticelli P, Tren R, Attaran A. Antimalarial Drug Quality in the Most Severely Malarious Parts of Africa - A Six Country Study. Plos One. 2008;3(5).

74. Syhakhang L, Lundborg CS, Lindgren B, Tomson G. The quality of drugs in private pharmacies in Lao PDR: a repeat study in 1997 and 1999. Pharmacy World & Science. 2004;26(6):333-8.

75. Kaur H, Goodman C, Thompson E, Thompson K-A, Masanja I, Kachur SP, et al. A Nationwide Survey of the Quality of Antimalarials in Retail Outlets in Tanzania.Plos One. 2008;3(10).

76. Amin AA, Snow RW, Kokwaro GO. The quality of sulphadoxine-pyrimethamine and amodiaquine products in the Kenyan retail sector.Journal of Clinical Pharmacy and Therapeutics. 2005;30(6):559-65.

77. Taylor RB, Shakoor O, Behrens RH, Everard M, Low AS, Wangboonskul J, et al. Pharmacopoeial quality of drugs supplied by Nigerian pharmacies. Lancet. 2001; 357(9272):1933-6.

78. Myatt M, Limburg H, Minassian D, Katyola D. Field trial of applicability of lot quality assurance sampling survey method for rapid assessment of

prevalence of active trachoma. Bulletin of the World Health Organization. 2003;81(12):877-85.

79. Robertson SE, Anker M, Roisin AJ, Macklai N, Engstrom K, Laforce FM. The lot quality technique: A global review of applications in the assessment of health services and diseases surveillance. World Health Statistics Quarterly. 1997; 50(3-4):199-209.

80. Newton PN, Lee SJ, Goodman C, Fernandez FM, Yeung S, Phanouvong S, et al. Guidelines for Field Surveys of the Quality of Medicines: A Proposal. Plos Medicine. 2009;6(3).

81. Ciurczak EW. Principles of near infrared spectroscopy. In: Burns DA and Ciurczak EW (Eds). Handbook of near infrared analysis. New York: Marcel Dekker; 1992.

82. Banwell CN and McCash EM.Fundamentals of molecular spectroscopy.4th ed. McGraw Hill; 1994.

83. Barton FE. Theory and principles of near infrared spectroscopy.Spectroscopy Europe. 2002; 14(1):12-18.

84. Burns DA. Historical development. In: Burns DA. Ciurczak EW (Eds.), Handbook of near-infrared analysis. New York: Marcel Dekker;1992:1-6.

85. Ellis JW. Molecular absorption spectra of liquids. Transactions of the Faraday Society.1929; 25:888

86. FOSS Training manual. FOSS Nirsystems;1998.

87. Yoon WL, Jee RD, Moffat AC, Blackler PD, Yeung K, Lee DC. Construction and transferability of a spectral library for the identification of common solvents by near-infrared transflectance spectroscopy.Analyst. 1999;124(8):1197-203.

88. Electrospray Ionization. Navigator connection information. Version 1.0

89. Chait EM. Ionization source in mass spectrometry.Analytical Chemistry. 1972;44(3):A77.

90. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. Science. 1989;246(4926):64-71.

91. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization-principles and practice. Mass Spectrometry Reviews. 1990;9(1):37-70.

92. Cody RB. Electrospray ionization mass spectrometry: history, theory and instrumentation. In: Pramanik BN, Ganguly AK and Gross ML (Eds). Applied electrospray mass spectrometry. New York: Marcel Dekker; 2002.

93. Todd JFJ. Ion trap mass spectrometer - Past, present, and future.Mass Spectrometry Reviews. 1991;10(1):3-52

94. Stafford GC, Kelley PE, Syka JEP, Reynolds WE, Todd JFJ. Recent improvement in ion trap technology.International Journal of Mass Spectrometry and Ion Processes. 1984;60(SEP):85-98.

95. de Hoffmann E and Stroobant V. Mass Spectrometry: principles and applications. Chichester: John Wiley & Sons Ltd; 2007.

96. Hillenkamp F, Karas M. Matrix-assisted laser desorption/ionisation, an experience. International Journal of Mass Spectrometry. 2000;200(1-3):71-7.

97. Guillermo Moyna. NMR lecture notes. University of Sciences (Philadelphia); 2008.

98. Blanco M, Villarroya I. NIR spectroscopy: a rapid-response analytical tool. Trac-Trends in Analytical Chemistry. 2002;21(4):240-50

99. Roggo Y, Roeseler C, Ulmschneider A. Near infrared spectroscopy for qualitative comparison of pharmaceutical batches. Journal of Pharmaceutical and Biomedical Analysis. 2004;36(4):777-86.

100. Blanco M, Coello J, Iturriaga H, Maspoch S, delaPezuela C. Effect of data preprocessing methods in near-infrared diffuse reflectance spectroscopy for the determination of the active compound in a pharmaceutical preparation. Applied Spectroscopy. 1997;51(2):240-6.

101. Davies AMC and Fearn T. Back to basics: removing multiplicative effects. Spectroscopy Europe. 2007; 19(4): 24-28.

102. Davies AMC. Back to basics: spectral pre-treatments-derivatives. Spectroscopy Europe. 2007;19(2):32-33.

103. Jee RD. Near infrared spectroscopy. In: Moffat AC, Osseiten MD, Widdop B, editors. Clarke's analysis of drugs and poisons.Pharmaceutical Press. 2004; v1: 358-367.

104. L.V. Azarraga and D. A. Hanna, GIFTS, Athens ERL GC/FT-IR Software and User's Guide; 1979.

105. Davies AMC. Back to basics: the principles of principal component analysis. Spectroscopy Europe. 2005; 17(1):20-23.

106. Trygg J, Gullberg J, Johansson AI, Jonsson P, Moritz T. Chemometrics in Metabolomics - An Introduction. Plant Metabolomics. 2006;57:117-28.

107. De Maesschalck R, Candolfi A, Massart DL, Heuerding S. Decision criteria for soft independent modelling of class analogy applied to near infrared data. Chemometrics and Intelligent Laboratory Systems. 1999; 47(1):65-77.

108. Davies AMC. Back to basics: multivariate qualitative analysis, "SIMCA". Spectroscopy Europe. 2008; 20(6):16.

109. Jackson JE. A Users Guide to Principal Components. New York: Wiley & Sons. 1991:45–47.

110. Maggio RM, Castellano PM, Kaufman TS. A new principal component analysis-based approach for testing "similarity" of drug dissolution profiles. European Journal of Pharmaceutical Sciences. 2008; 34(1):66-77.

111. Guidelines for Drug Sampling, USP DQI; 2006 [accessed 2 Feb 2008]. Available from: http://www.usp.org/pdf/EN/dqi/samplingGuidelines.pdf

112. M. McGinnis. Matrix of drug quality reports affecting USAID- assisted countries. US Pharmacopeia; 2009 [accessed 14 Jul 2009] . Available from: http://www.usp.org/pdf/EN/dqi/ghcDrugQualityMatrix.pdf

113. Scafi SHF, Pasquini C. Identification of counterfeit drugs using near-infrared spectroscopy. Analyst. 2001;126(12):2218-24.

114. Jonscher KR, Yates JR. Mixture analysis using a quadrupole mass filter quadrupole ion trap mass spectrometer. Analytical Chemistry. 1996;68(4):659-67.

115. Eye- Mo; 2007 [accessed 3 Mar 2011]. Available from: http://gsk.com.my/healthcare-eyemo.html

116. Po Chai Pills. Chinese Natural Herbs; 2010 [accessed 25 oct 2011]. Available from: http://www.chinesenaturalherbs.com/herb_pages/pochaipills.asp

117. Hong Kong Department of Health.Recall of Po Chai Pills (press release); March 2010 [accessed on 3 Mar 2011]. Available from: http://www.info.gov.hk/gia/general/201003/24/P201003240255_print.htm

118. Press release. Suspected counterfeit traditional product "Po Chai Pills" adulterated with diclofenac; 2010 [accessed 3 Mar 2011]. Available at: http://portal.bpfk.gov.my/aeimages//File/BUU/BUU_June_2010.pdf

119. Dubois P, Martinez JR, Levillain P. Determination of five components in a pharmaceutical formulation using near infrared reflectance spectrophotometry. Analyst. 1987;112(12):1675-9.

120. Kamat MS, Lodder RA, Deluca PP. Near-infra-red spectroscopic determination of residual moisture in lyophilized sucrose through intact glass phials. Pharmaceutical Research. 1989;6(11):961-5.

121. Broad NW, Jee RD, Moffat AC, Eaves MJ, Mann WC, Dziki W. Non-invasive determination of ethanol, propylene glycol and water in a multi-component pharmaceutical oral liquid by direct measurement through

amber plastic bottles using Fourier transform near-infrared spectroscopy. Analyst. 2000;125(11):2054-8.

122. Bergles E. The NIR First Overtone and Raman's Spectral Fingerprints: Telecom's Resurgence [accessed 3 Mar 2011]. Available from: www.bayspec.com

123. Williams P. Sampling, sample preparation, and sample selection. In: Burns DA and Ciurczak EW, editors. Handbook of near infrared analysis. New York: Marcel Dekker; 1992: 325.

124. Blanco M, Romero MA. Near infrared transflectance spectroscopy Determination of dexketoprofen in a hydrogel.Journal of Pharmaceutical and Biomedical Analysis. 2002;30(3):467-72.

125. Clerc JT, Zupan J. Computer-based systems for the retrieval of infrared spectral data. Pure and Applied Chemistry. 1977;49(12):1827-31.

126. Fox RC. Computer searching of infrared spectra using peak location and intensity data.Analytical Chemistry. 1976;48(4):717-21.

127. Tanabe K, Saeki S. Computer retrieval of infrared spectra by a correlation coefficient method. Analytical Chemistry. 1975;47(1):118-22.

128. Heller SR and Lowry SR. Library storage and retrieval methods in infrared spectroscopy. In: Meuzelaar HLC, Isengour T, Jurs PC (Eds). Computer-enhanced Analytical Spectroscopy; 1987.

129. Kramer K, Ebel S. Application of NIR reflectance spectroscopy for the identification of pharmaceutical excipients. Analytica Chimica Acta. 2000; 420(2):155-61.

130. Malaysian Drug Code; 2011 [accessed 5 Jun 2011]. Available from: http://www.pharmacy.gov.my/index.cfm?menuid=44

131. Affolter C, Clerc JT. Estimation of the performance of spectroscopic library search systems. Fresenius Journal of Analytical Chemistry. 1992;344(4-5):136-9.

132. Clerc JT, Pretsch E, Zurcher M. Performance analysis of infrared library search system. Mikrochimica Acta. 1986;2(1-6):217-42

133. Harrington PB, Isenhour TL. A quantitative measure of the reliability of searches of spectral libraries.Analytica Chimica Acta. 1987;197:105-19.

134. Grotch SL. Computer Identification of Mass Spectra using highly compressed spectra code. Analytical Chemistry. 1973;45(1):2-6.

135. Candolfi A, De Maesschalck R, Jouan-Rimbaud D, Hailey PA, Massart DL. The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra.Journal of Pharmaceutical and Biomedical Analysis. 1999;21(1):115-32.

136. Griffiths PR, Wilkins CL. Quality criteria for digital infrared reference spectra. Applied Spectroscopy. 1988;42(4):538-45.

137. Ratanawijitrasin S and Wondemagegnehu E. Effective drug regulation: a multicountrystudy. World Health Organization; 2002 [accessed 18 Apr 2011]. Available from: http://www.who.int/medicinedocs/pdf/s2300e/s2300e.pdf

138. Gross A and Weintraub R.Drug, device and cosmetic regulations in Malaysia. Pacific Bridge Medical; 2005 [accessed 3 Feb 2011]. Available from: http://www.pacificbridgemedical.com/publications/malaysia/2005_drug_device_and_cosmetic

139. Drug Control Authority. Ministry Of Health Malaysia; 2010 [accesses 9 Jan 2010].Available from: http://www.bpfk.gov.my/Search/Search_product.asp

140. Esbensen, K. Multivariate Data Analysis- in Practice (5thed.). Norway: CAMO Software AS; 2005.

141. Amix viewer and Amix software manual, version 3.8. Bruker BioSpin GmbH; 2008.

142. Candolfi A, Massart DL, Heuerding S. Investigation of sources of variance which contribute to NIR-spectroscopic measurement of pharmaceutical formulations. Analytica Chimica Acta. 1997;345(1-3):185-96.

143. Smith MR, Jee RD, Moffat AC. The transfer between instruments of a reflectance near-infrared assay for paracetamol in intact tablets.Analyst. 2002;127(12):1682-92.

144. Smith MR, Jee RD, Moffat AC, Rees DR, Broad NW. A procedure for calibration transfer between near-infrared instruments - a worked example using a transmittance single tablet assay for piroxicam in intact tablets.Analyst. 2004;129(9):806-16.

145. Gaik OB. Misuse and overuse of antibiotics: Who is to blame? [accessed on 5 Jun 2011]. Available from: http://www.twnside.org.sg/title/ong-cn.htm

146. Bahring-Kuhlmey SR. Bacampicillin. Drugs of Today. 1977;13(10):413-7.

147. Bodin NO, Ekstrom B, Forsgren U, Jalar LP, Magni L, Ramsay CH, et al. Becampicillin: a new orally well absorbed derivative of ampicillin. Antimicrobial Agents and Chemotherapy. 1975;8(5):518-25.

148. Blanco M, Coello J, Iturriaga H, Maspoch S, Delapezuela C. Determination of ascorbic acid in pharmaceutical preparations by near infrared reflectance spectroscopy. Talanta. 1993;40(11):1671-6.

149. Blanco M, Cueva-Mestanza R, Peguero A. Controlling individual steps in the production process of paracetamol tablets by use of NIR spectroscopy. Journal of Pharmaceutical and Biomedical Analysis. 2010;51(4):797-804

150. Mark H. Qualitative discriminant analysis. In: Burns DA and Ciurczak EW, editors. Handbook of near infrared analysis. New York: Marcel Dekker; 1992:363-400

151. Lonardi S, Viviani R, Mosconi L, Bernuzzi M, Corti P, Dreassi E, et al. Drug analysis by near-infra-red reflectance spectroscopy. Determination of the active ingredient and water content in antibiotic powders.Journal of Pharmaceutical and Biomedical Analysis. 1989;7(3):303-8.

152. Gonzalez F, Pous R. Quality control in manufacturing process by near-infrared spectroscopy. Journal of Pharmaceutical and Biomedical Analysis. 1995;13(4-5):419-23.

153. Rosa SS, Barata PA, Martins JM, Menezes JC. Near-infrared reflectance spectroscopy as a process analytical technology tool in Ginkgo biloba extract qualification. Journal of Pharmaceutical and Biomedical Analysis. 2008;47(2):320-7.

154. WHO. Traditional Medicines factsheet, 2008 [accessed 3 Mar 2009]. Available from: http://www.who.int/mediacentre/factsheets/fs134/en/

155. Malaysian National News Agency. Local herbal industry registers annual sales Of RM4.5billio; 2005 [accessed 18 Dec 2007]. Available from: http://www.bernama.com.my/bernama/v3/news business.php?id=150892

156. Malaysian National News Agency. Local herbal market to hit RM 8 Billion by 2010; 2006 [accessed 18 Dec 2007]. Available from:http://bernama.com.my/bernama/v3/news.php?id=219357

157. Aziz Z. Herbal medicines: predictors of recommendation by physicians. Journal of Clinical Pharmacy and Therapeutics. 2004;29(3):241-6.

158. Ang HH, Lee EL, Cheang HS. Determination of mercury by cold vapor atomic absorption spectrophotometer in Tongkat Ali preparations obtained in Malaysia. International Journal of Toxicology. 2004;23(1):65-71.

159. Ang HH, Lee EL, Matsumoto K. Analysis of lead content in herbal preparations in Malaysia. Human & Experimental Toxicology. 2003;22(8):445-51.

160. Giri L, Andola HC, Purohit VK, Rawat MSM, Rawal RS, Bhatt ID. Chromatographic and Spectral Fingerprinting Standardization of Traditional Medicines: An Overview as Modern Tools. Research Journal of Phytochemistry.2010; 4: 234-241.

161. Liang YZ, Xie PS, Chan K. Quality control of herbal medicines.Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences. 2004;812(1-2):53-70

162. WHO. Traditional medicines (Regional Office for the Western Pacific); 2001 [accessed 3 Feb 2011]. Available from: http://www.wpro.who.int/internet/resources.ashx/RCM/RC5207.pdf

163. Baert B, De Spiegeleer B. Quality analytics of internet pharmaceuticals. Analytical and Bioanalytical Chemistry. 2010;398(1):125-36.

164. Jackson G, Arver S, Banks I, Stecher VJ. Counterfeit phosphodiesterase type 5 inhibitors pose significant safety risk. International Journal of Clinical Practice. 2010; 64(4): 497-504.

165. Zhari I, Norhayati I, Jaafar L. Malaysian herbal monograph (Vol. 1). Kuala Lumpur: Malaysian Monograph Committee;1999.

166. Bhat R, Karim AA. Tongkat Ali (Eurycoma longifolia Jack): A review on its ethnobotany and pharmacological importance. Fitoterapia. 2010;81(7):669-79.

167. Bandaranayake WM. Quality control, screening, toxicity and regulation of herbal drugs. In: Ahamad I, Aqil F and Owais M. Modern Phytomedicine: turning medicinal plants into drugs. Weinheim: Wiley- VCH Verlag GmbH; 2006.

168. Lewtress Herbal Health Shop; 2010 [accessed July 2010] Available at: http://www.lewtress.co.uk/

169. Zou P, Hou P, Oh SS-Y, Chong YM, Bloodworth BC, Low M-Y, et al. Isolation and identification of thiohomosildenafil and thiosildenafil in health supplements. Journal of Pharmaceutical and Biomedical Analysis. 2008;47(2):279-84.

170. Lai K-C, Liu Y-C, Liao Y-C, Lin Y-L, Tsai L-Y, Lin J-H, et al. Isolation and Identification of Three Thio-sildenafil Analogues in Dietary Supplements. Journal of Food and Drug Analysis. 2010;18(4):269-78.

171. Li L, Low MY, Aliwarga F, Teo J, Ge XW, Zeng Y, et al. Isolation and identification of hydroxythiohomosildenafil in herbal dietary supplements sold as sexual performance enhancement products. Food Additives and Contaminants Part a-Chemistry Analysis Control Exposure & Risk Assessment. 2009;26(2):145-51.

172. US Food and Drug Administration. Information for Healthcare Professionals Sildenafil (marketed as Viagra and Revatio), Vardenafil (marketed as Levitra), Tadalafil (marketed as Cialis); 2011. [accessed 5 Jun 2011] Available from: http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm124841.htm

173. Balayssac S, Gilard V, Malet-Martino M, Martino R. 2-D DOSY and 3-D DOSY-COSY 1H NMR as efficient tools for analysis of herbal drugs and identification of possible adulterants. Abstracts of Papers of the American Chemical Society. 2009;237:208-.

174. Sabucedo AJ, Gutierrez MA, Mueller KC, Bellissima BL, Hsu YL, Rose S, et al. Sex, lies, and Niagra. Jama-Journal of the American Medical Association. 2004;291(5):560-2.

175. Counterfeit drugs -Guidelines for the development of measure to combat counterfeit drugs. World Health Organization; 1999 [accessed 27 Nov 2007]. Available from: http://www.who.int/medicines/publications/counterfeitguidelines/en/

176. General information on counterfeit medicines. World Health Organization; 2011 [accessed 10 Apr 2011]. Available from: http://www.who.int/medicines/services/counterfeit/overview/en/index.html

177. Say no to fake drugs; 2011. [accessed 30 Jul 2011]. Available from: http://firsttelepharma.com.my/node/42.

178. Report of the situation of counterfeit medicines based on data collection tool; WHO regions for Africa and Eastern Mediterranean. World Health Organization; 2011 [accessed 10 Apr 2011]. Available from: http://apps.who.int/medicinedocs/documents/s18385en/s18385en.pdf

179. Bernama. Fake Chinese medicine in the market.The Star; May 2011 [accessed on 27 May 2011]. Available from:http://thestar.com.my/news/story.asp?file=/2011/5/23/nation/20110523184728&sec=nation

180. Alias, WNH. 94 jenis ubat tiruan dirampas.Utusan Malaysia; 26 May 2011 [accessed on 27 may 2011]. Available from:http://www.utusan.com.my/utusan/info.asp?y=2011&dt=0527&pub=Utusan_Malaysia&sec=Dalam_Negeri&pg=dn_13.htm

181. Semua ini palsu atau tiruan. Kuala Lumpur; 19 April 2011 [accessed on 27 may 2011]. Available from: http://ondscene.my/pengguna/9699

182. Ubat Palsu. Kuala Lumpur; 19 April 2011[accessed on 27 may 2011]. Available from: www.tv3.com.my/beritatv3/tajuk_-utama/Ubat-Palsu-Pula.html

183. Moideen SVK. Malaysian experience in herbal medicines testing.National Pharmaceutical Control Bureau (NPCB); 2008. [accessed 25 Oct 2010]. Available from: portal.bpfk.gov.my/view_file.cfm?fileid=338

184. WHO. Measures to combat counterfeit medicines; 2011 [accessed 27 May 2010].Available from: http://www.who.int/medicines/services/counterfeit/overview/en/index2.html

185. Fernandez FM, Green MD, Newton PN. Prevalence and detection of counterfeit pharmaceuticals: A mini review. Industrial & Engineering Chemistry Research. 2008;47(3):585-90.

186. Rodionova OY, Houmoller LP, Pomerantsev AL, Geladi P, Burger J, Dorofeyev VL, Arzamastsev AP. NIR spectrometry for counterfeit drug detection: A feasibility study. Analytica Chimica Acta.2005; 549:151–158.

187. Olsen BA, Borer MW, Perry FM, Forbes RA. Screening for counterfeit drugs using near-infrared spectroscopy.Pharmaceutical Technology. 2002; 6(6): 62-71.

188. Storme-Paris I, Rebiere H, Matoga M, Civade C, Bonnet PA, Tissier MH, et al. Challenging Near InfraRed Spectroscopy discriminating ability for counterfeit pharmaceuticals detection. Analytica Chimica Acta. 2010;658(2):163-74.

189. Westenberger BJ, Ellison CD, Fussner AS, Jenney S, Kolinski RE, Lipe TG, et al. Quality assessment of internet pharmaceutical products using traditional and non-traditional analytical techniques. International Journal of Pharmaceutics. 2005;306(1-2):56-70.

190. de Peinder P, Vredenbregt MJ, Visser T, de Kaste D. Detection of Lipitor (R) counterfeits: A comparison of NIR and Raman spectroscopy in combination with chemometrics. Journal of Pharmaceutical and Biomedical Analysis. 2008;47(4-5):688-94.

191. Vredenbregt MJ, Blok-Tip L, Hoogerbrugge R, Barends DM, de Kaste D. Screening suspected counterfeit Viagra (R) and imitations of Viagra (R) with near-infrared spectroscopy. Journal of Pharmaceutical and Biomedical Analysis. 2006;40(4):840-9.

192. Ponstan® Capsules (mefenamic acid) - consumer medicine information. Pfizer Australia Pty Ltd. Mar 2011 [accessed 27 May 2011]. Available from: http://www.betterhealth.vic.gov.au/bhcv2/bhcmed.nsf/pages/pfcponsc/$File/pfcponsc.pdf

193. Berger S and Sicker D. Classics in spectroscopy: isolation and structure elucidation of natural products. Weinheim: Wiley-VCH;2009.

194. Abbott .Advisory Committee Briefing Document for Meridia®(sibutramine hydrochloride monohydrate) 15 September 2010. Available at: http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/EndocrinologicandMetabolicDrugsAdvisoryCommittee/UCM225573.pdf(Accessed on 7 January 2011)

195. Heusser K, Tank J, Diedrich A, Engeli S, Matta S, Kruger N, et al. Influence of sibutramine treatment on sympathetic vasomotor tone in obese subjects. Clinical Pharmacology & Therapeutics. 2006;79(5):500-8.

196. Harrison-Woolrych M, Clark DWJ, Hill GR, Rees MI, Skinner JR. QT interval prolongation associated with sibutramine treatment. British Journal of Clinical Pharmacology. 2006;61(4):464-9.

197. Cheung BMY. Drug Treatment for Obesity in the Post-Sibutramine Era.Drug Safety. 2011;34(8):641-50.

198. Lai K-C, Liu Y-C, Tseng M-C, Lin Y-L, Lin J-H. Isolation and identification of a sibutramine analogue in a healthy food for weight loss.Journal of Food and Drug Analysis. 2007;15(1):20-4.

199. Zou P, Oh SS-Y, Kiang K-H, Low M-Y, Bloodworth BC. Detection of sibutramine, its two metabxolites and one analogue in a herbal product for weight loss by liquid chromatography triple quadrupole mass spectrometry and time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry. 2007;21(4):614-8.

200. Bogusz MJ, Hassan H, Ai-Enazi E, Ibrahim Z, Al-Tufail M. Application of LC-ESI-MS-MS for detection of synthetic adulterants in herbal remedies. Journal of Pharmaceutical and Biomedical Analysis. 2006;41(2):554-64.

201. Simeon Bennett. Pfizer: Civil suits for drug counterfeiters. Bloomberg Businessweek; 8 Jul 2010 [accessed 27 May 2011]. Available from: http://www.businessweek.com/magazine/content/10_29/b4187021988297.htm

202. Gong F, Wang BT, Chau FT, Liang YZ. Data preprocessing for chromatographic fingerprint of herbal medicine with chemometric approaches. Analytical Letters. 2005;38(14):2475-92.

203. Alaerts G, Dejaegher B, Smeyers-Verbeke J, Vander Heyden Y. Recent Developments in Chromatographic Fingerprints from Herbal Products: Set-Up and Data Analysis. Combinatorial Chemistry & High Throughput Screening. 2010;13(10):900-22.

204. Zupan J. Computer supported data bases. New York: Howard/Wiley; 1986.

BIBLIOGRAPHY

1. Palmer KE.Transferability of near infrared spectral libraries for the identification of pharmaceutical actives and excipients [PhD thesis]. London: The School of Pharmacy, University of London; 2008.

2. Gunther H. NMR Spectroscopy: Basic principles, concepts and applications in chemistry. Chichester: John Wiley & Sons Ltd; 1998.

APPENDIX

Appendix 1

The letters of acceptance from two pharmaceutical companies,supported this research by supplying some samples to be included in the database.

enriching life together

pharmaniaga

MS ISO 9001 REG. NO. AR2175 MS ISO 14001 CERT. NO. FM54201121 OHSAS 18001 CERT. NO. RQ0880111121

Our ref : PMB/BD/0601
Date : 22-June-2009

School of Pharmacy,
University of London,
United Kingdom.

Attn : Ms Mazlina Said,

RE: INVITATION TO PARTICIPATE IN A RESEARCH PROJECT ENTITLED "DESIGN A SPECTRAL DATABASE OF MALAYSIAN MEDICINE"

We are interested for participate the above research by providing 3 sample of Pharmaniaga product. The detail as per below:-

1. Ticlopidine 250mg			
a) Batch number	:8P0321	8P031	8P034
b) Expiry	:Dec 2011	Dec 2011	Dec 2011
c) Qty	:20 tablet	20 tablet	20 tablet
2. Cefuroxime Axetil 250mg (Xylid)			
a) Batch number	:7P4186	7P4222	8A083
b) Expiry	:Dec 2010	Dec 2010	Jan 2011
c) Qty	:30 tablet	30 tablet	30 tablet
3. Simvastatin 20 mg			
a) Batch number	:8F236	8J089	8L064
b) Expiry	:Apr 2011	Jul 2011	Sept 2010
c) Qty	:30 tablet	30 tablet	30 tablet

We hope all data will be treated with confidentiality. Looking forv.ard to receive the research result from your PhD student.

Thank you.

Yours sincerely,
PHARMANIAGA MANUFACTURING BERHAD

Mohd Zul Efendi Bin Hassan
Senior Executive, Business Development.

Pharmaniaga Manufacturing Berhad (50016-D)
A subsidiary of Pharmaniaga Berhad

11A, Jalan P/1, Kawasan Perusahaan Bangi,
43650 Bandar Baru Bangi, Selangor Darul Ehsan, Malaysia.
Tel: 603-8925 7880 Fax: 603-8925 6177 Website: www.pharmaniaga.com
Portal: www.ehealth4all.com

UEM GROUP

Y.S.P. INDUSTRIES (M) SDN. BHD. (192593 U)
16th Floor, Plaza VADS,
No. 1, Jalan Tun Mohd. Fuad,
Taman Tun Dr. Ismail, 60000 Kuala Lumpur, Malaysia.

603-7727 6390 (12 lines)
603-7727 6701
info@yspsah.com

Y.S.P. Industries (M) Sdn. Bhd.
Lot 3, 5 & 7, Jalan P/7,
Section 13, Kawasan Perindustrian,
Bandar Baru Bangi,
43000 Kajang,
Selangor Darul Ehsan.

Ms. Mazlina Said,
The School of Pharmacy,
University of London,
29-39 Brunswick Square,
London WC1N 1AX

June 18, 2009

Dear Sir/Madam,

PARTICIPATE IN A RESEARCH PROJECT ENTITLED " DESIGN A SPECTRAL DATABASE OF MALAYSIAN MEDICINE"

This is the sample for a research project that is aiming to build a spectral database of Malaysian Pharmaceutical products with the product ID and batch no.:

1) MAT – SJ027, SJ028, SJ029
2) PLT – TD011, TD012, TD014
3) VRT5 – TD005, TD006, TD007.

Please sign and acknowledge receive sample on our copy of this letter. Do call us if any inquiries. Thank you.

Best regards,

.....
(Chong Mung Soon)
QC Lab Manager
Tel : 03-89251215 ext. 110

Acknowledge Receive by:

.....
(MAZLINA MOHD SAID

Customer Copy

Appendix 2a

The selection of solvent and the optimum sample concentration for the analysis of '*Kintop*' by NMR.

Objectives

1. To compare the efficiency of two types of solvents; deuterium oxide (D_2O) and deuterated methanol (methanol-d_4).
2. To determine the optimum concentration of the sample needed to produce a clear and stable NMR spectrum.

Materials and Methods

Drug Sample

A herbal preparation named '*Kintop*' was used in this analysis. Kintop was used traditionally as an anti-obsity and this supplement is sold as capsules formulation. According to the label on the packaging, this product contains herbal extract from different sources such as *Radix Polygonymultiflory*, *Radix astradyali*, *Radix noto* ginseng, *Rhizoma rhei* and *Spica purnellae*.

Sample preparation

20 capsules were weighed individually and then the powder samples from each capsule were emptied into a beaker. Twelve different concentrations; 1 mg, 5 mg, 10 mg, 20 mg, 30 mg, 40 mg, 50 mg, 60 mg, 70 mg, 80 mg, 90 mg and 100 mg were prepared by dissolving the powder sample in 1 ml of two different solvents; D_2O and methanol-d_4, with added internal standards TSP (tetramethylsilyl propionate) and TMS (tetramethylsilane), respectively. The solutions were mixed for 1 minute and then subjected to sonication for 15 minutes, followed by 5 minutes centrifugation at 13,000 rpm. 650 µL of the supernatant for each sample was transferred into a 5-mm NMR tube for analysis.

NMR analysis

NMR spectra were acquired using BRUKER AVANCE500 MHz spectrometers equipped with TOPSPIN (v 1.3) for data analysis. Spectra were recorded as the

following parameters acquisition time, 3.17 s; spectral width, 10330.57Hz; relaxation delay, 6 s; and 256 scans. The area of peaks in the spectra were calibrate against the area of the internal standard peak.

Results and discussion

The spectra of different concentrations dissolved in D_2O and methanol-d_4 are illustrated in Figure 1 and 2, respectively. More peaks were observed in the spectra dissolved in methanol-$d_4$indicating that the sample was more soluble in this solvent.

There are quite a different spectral patterns for the two sets of sample particularly in the following regions;

The region between 5.0ppm to 7.5ppm of D_2O set of spectra shows broad peaks with relatively lower intensities as compared to the spectrum of Methanol-d_4. The same region in the spectrum obtained by using methanol-d_4 solvent gives very clear and sharp peaks.

In the region between 0.5ppm to 3.0ppm, the set of spectra in methanol-$d_4$had a higher number of broader peaks, while the spectra in D_2O had fewer but sharper signals.

Following this work, seven peaks from the D_2O set of spectra and nine peaks for the methanol-d_4 set of spectra were selected and intergrated. Table 1 and 2 shows the position of the selected peaks (ppm) and their corresponding intensities for the D_2O and methanol-d_4 set of spectra, respectively. For each peak, a graph of the sample concentration against their corresponding intensities was plotted to observe the degree of solubility of the compound represented by each peak (Figure 3 and 4).

The set of D_2O spectra showed the increase of the intensities as the concentration increased. However, peaks in the Methanol-d_4 set of spectra exhibited unexpected behaviour. For example, the second peak shows only minor increase in its intensity corresponding to the increment of concentration. Going further, the intensity of peak 3 was found to be decreasing with increasing concentration. Moreover, almost all the peaks except for peak 1 showed only a minor increase over concentration.

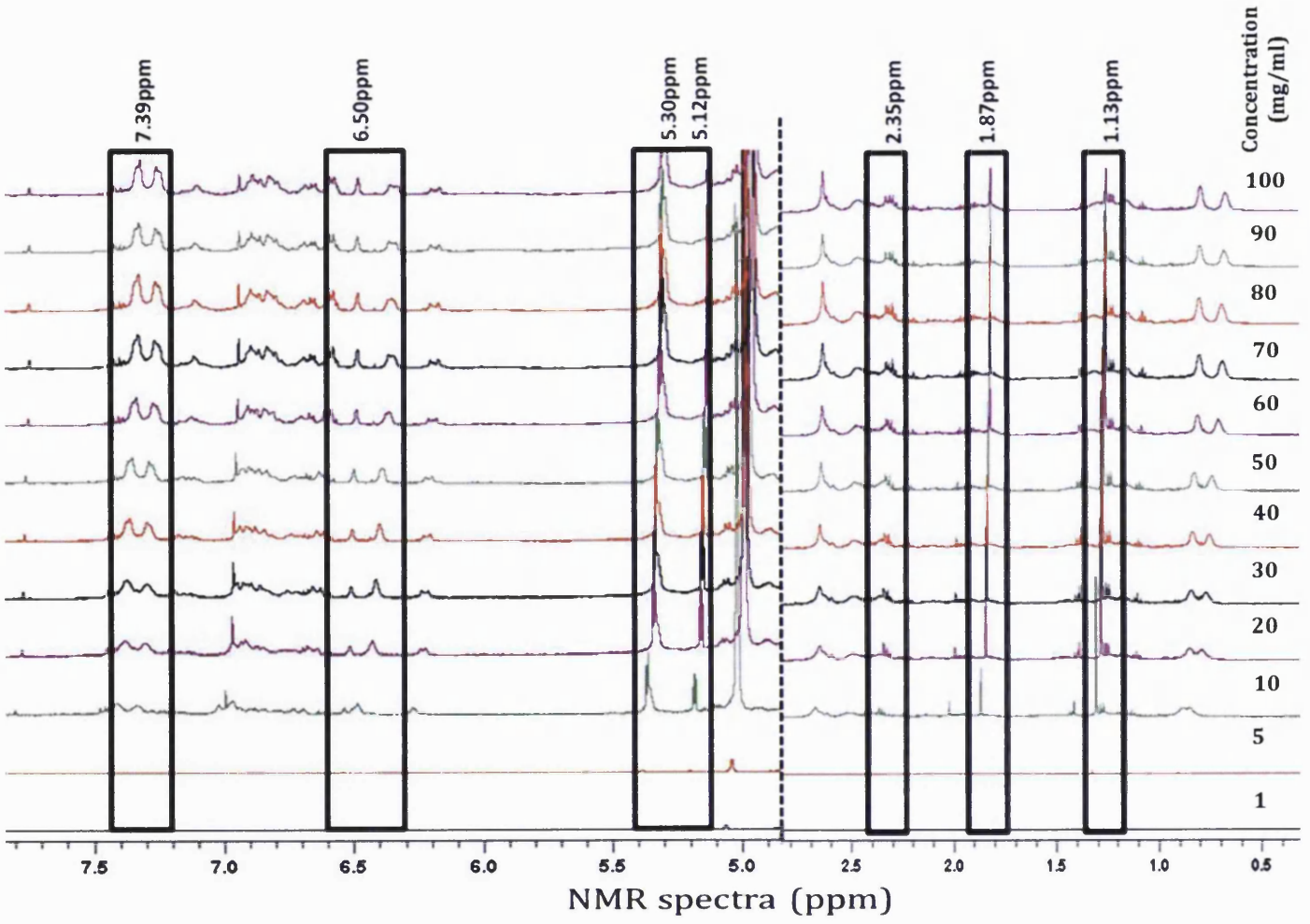


Figure A2-1 Overlap of spectra of different concentrations of Kintop in D_2O.

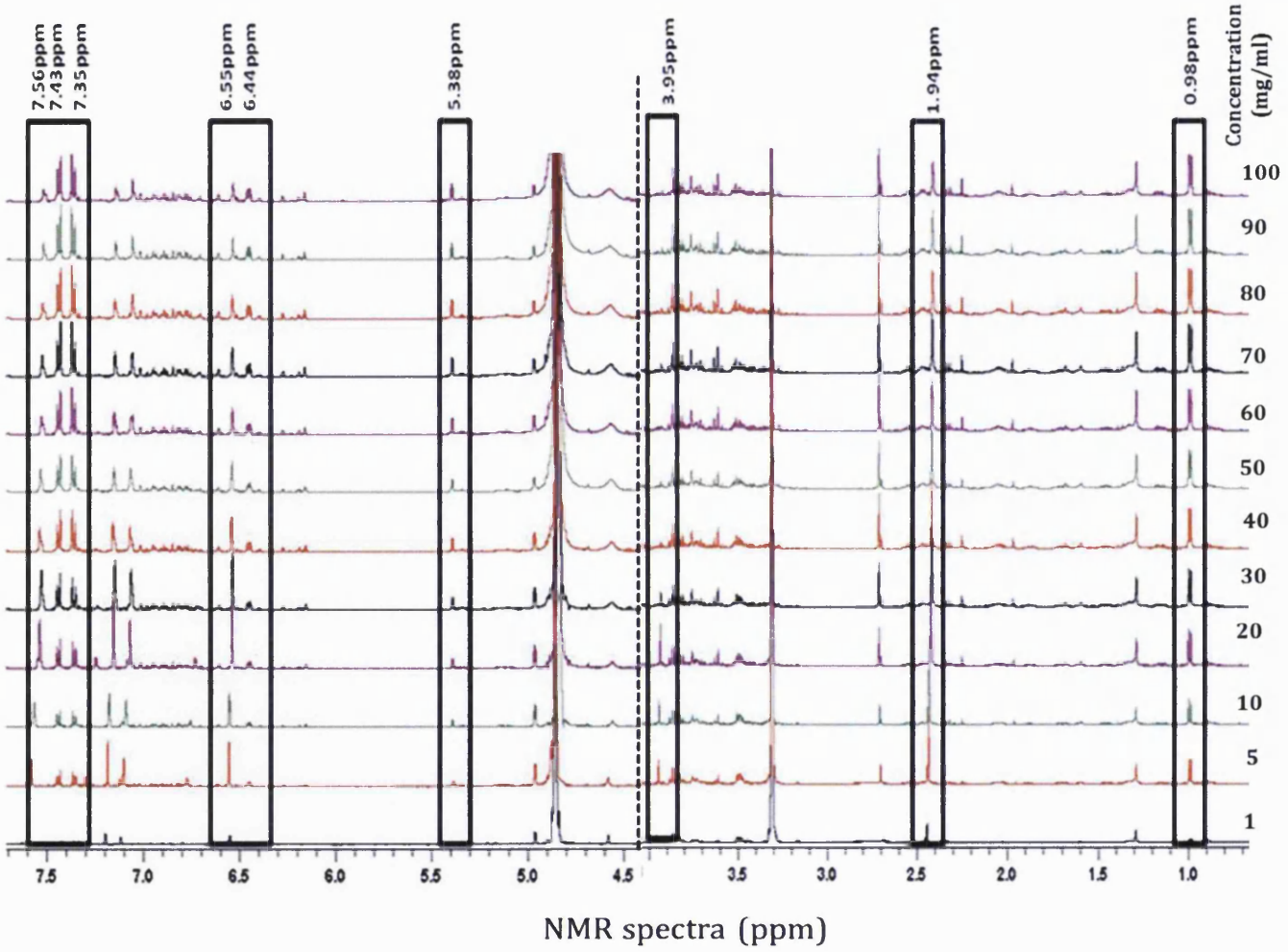


Figure A2-2 Overlap of spectra of different concentration of Kintop in methanol-d_4.

Table A2-1 The selected peaks from the NMR spectra of Kintop in D_2O with their corresponding intensities.

Conc. (mg/ml)	Peaks (ppm)						
	1 (1.13)	2 (1.87)	3 (2.35)	4 (5.18)	5 (5.36)	6 (6.5)	7 (7.39)
1	-	-	-	-	-	-	-
5	-	0.22	-	0.47	0.96	0.41	0.51
10	0.35	0.50	0.86	0.78	2.18	1.03	2.71
20	0.65	1.11	1.98	1.93	4.42	2.58	5.69
30	0.88	1.70	2.82	2.39	5.63	3.01	7.33
40	0.90	1.97	3.25	2.46	6.68	3.02	9.30
50	0.93	2.41	4.21	3.03	7.12	2.91	10.82
60	1.37	3.55	6.76	5.29	12.26	5.46	16.77
70	1.70	4.85	8.70	5.87	14.66	5.95	19.92
80	1.76	5.18	10.34	6.86	15.90	6.06	22.82
90	2.01	6.09	10.63	8.63	18.61	7.46	24.26
100	2.09	6.74	13.26	8.89	20.02	6.89	28.30

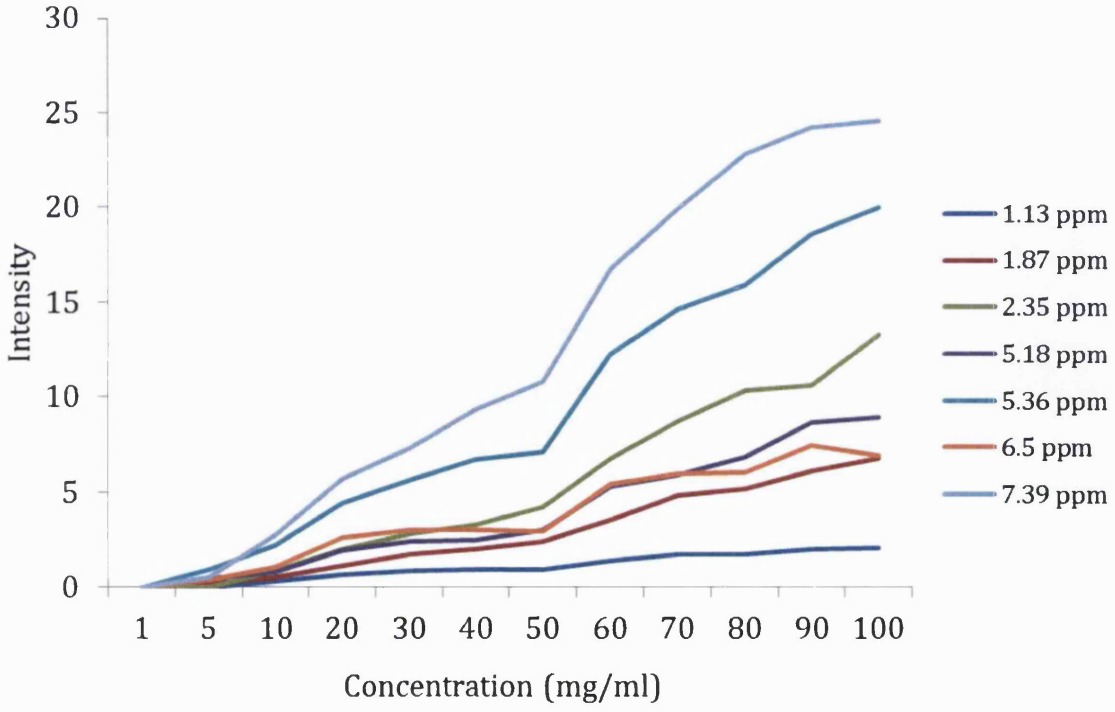


Figure A2-3 Effect of concentration on the intensity of peaks in NMR spectra of Kintop in D_2O.

Table A2-2 The selected peaks of the NMR spectra of Kintop in methanol-d_4 with their corresponding intensities

Conc. mg/ml	Peaks (ppm)								
	1 (0.98)	2 (1.95)	3 (3.95)	4 (5.38)	5 (6.44)	6 (6.55)	7 (7.35)	8 (7.43)	9 (7.56)
1	0.033	-	0.013	0.004	-	0.022	0.007	0.10	0.029
5	0.200	0.006	0.084	0.019	0.0243	0.105	0.066	0.069	0.132
10	0.209	0.010	0.104	0.024	0.033	0.140	0.068	0.073	0.172
20	0.362	0.018	0.144	0.038	0.057	0.237	0.125	0.138	0.283
30	0.606	0.028	0.091	0.065	0.091	0.313	0.206	0.237	0.340
40	0.931	0.042	0.064	0.088	0.132	0.272	0.319	0.353	0.287
50	1.441	0.080	0.069	0.127	0.195	0.312	0.472	0.517	0.338
60	1.453	0.067	0.043	0.143	0.202	0.296	0.486	0.541	0.322
70	1.592	0.083	0.045	0.159	0.228	0.300	0.553	0.623	0.321
80	1.876	0.088	0.045	0.195	0.276	0.305	0.636	0.737	0.331
90	2.439	0.118	0.086	0.223	0.340	0.338	0.816	0.926	0.348
100	2.691	0.138	0.077	0.263	0.383	0.383	0.932	1.046	0.421

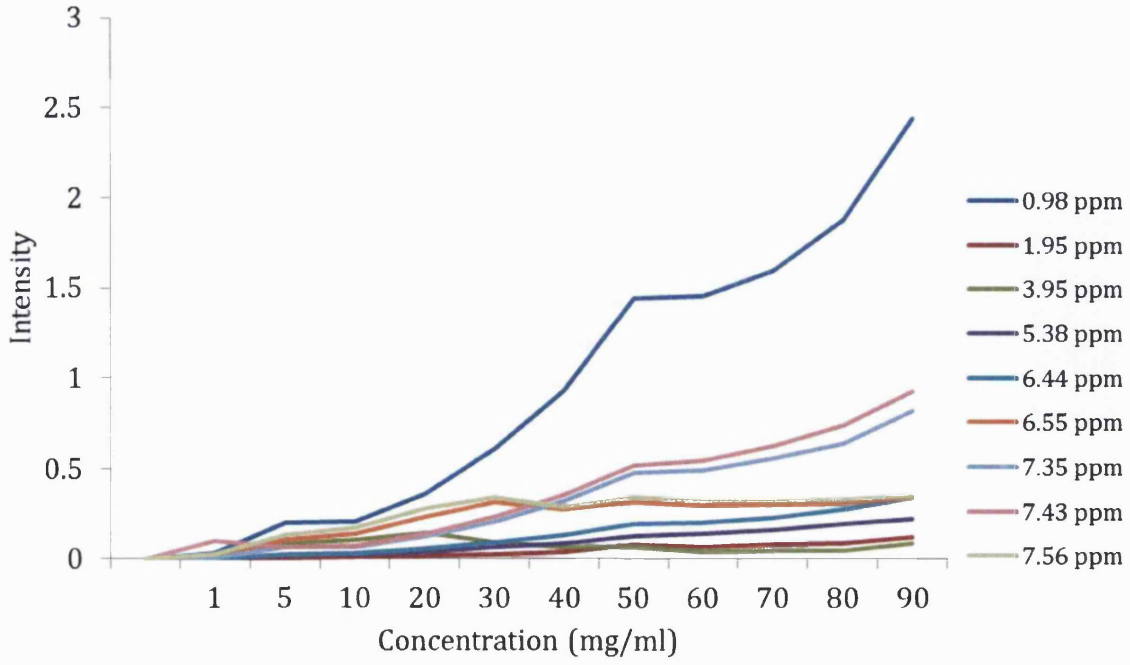


Figure A2-4 Effect of concentration on the intensity of peaks in NMR spectra of Kintop in methanol-d_4.

'*Kintop*' is a complex mixture of herbal extracts which most probably contained both the hydrophilic and hydrophobic compounds that leads to different NMR profiles in two different solvents. Thus, it would be difficult to observe all compounds using a single solvent. However, the use of mixture of solvents may complicate the analysis of spectra. Thus, it may be necessary to make compromises and select an adequate solvent for a herbal mixture that would enable analysis of bioactive compounds or compounds of interest from the mixture. From these observations, it can be said that as the concentration of a sample increases, the intensity of almost all peaks increase as well, but not at the same rate.

In the set of spectra of Kintop dissolved in D_2O, the solubility of polar components was found to be proportional to the concentration added. All the peaks (except for peak 1) showedan increase of peak area in samples at a concentration of 50mg/ml and above.

For the spectra of Kintop in Methanol-d_4, signal at peak 1 showed gradual increased in intensity with concentration. However, peak 2 and peak 3 were very much at the same level with increased concentration with a slight downfall in the slope beyond the sample concentration of 50 mg/ml. This may be because of the compounds associated with these peaks were not completely soluble in this solvent. The other peaks showed moderate increased in intensities with concentration.

The change of chemical shift for some peaks of NMR spectra was observed with increasing concentrations as the result of changes of pH or ionic strength of samples with higher amounts of sample in the solution. Thus it is essential to use similar concentrations of samples in order to be able to compare spectra of different samples.

Conclusion

The herbal sample 'Kintop' consisted of more lipophilic compounds based on the presence of more peaks observed in methanol-d_4 solvent compared with in D_2O. Thus, it may be sufficient to use this solvent in order to build up the chemical fingerprint of this sample. However, this information may not be enough for full identification of all components.

For future experiments, 50 mg/ml could be considered as the acceptable concentration for acquiring spectra of different commercial Kintop samples as acceptable intensities were observed in both experiments. In general, it is important to carry out concentration studies of herbal formulations to find the optimum conditions for the acquiring NMR spectra of such samples.

Appendix 2b

The effects of sample preparation by different operators on 1D ^{1}H NMR spectra

Objective

To observe the NMR spectra variations in herbal analysis conducted by three different operators.

Materials and Methods

Samples

Six types of herbal preparations were used for analysis. Details on the samples' name, main herbal components and claimed indications are listed in Table A2-3.

Table A2-3 Details on the herbal preparations used in this analysis and

Sample name	Sample label	Claimed indication/s
Kintop	*Kn*	Traditionally used to lose weight and for incontinence
Jamu Ajaib	*JA*	To reduce back pain, knee pain and shoulder pain, reduce bloating, to increase general health and strength, to treat leg and hand numbness
Tunglin	*Tg*	Management of rheumatic disease; rheumatoid arthritis, osteoarthritis, ankylosing spondylitis, gout and painful musculo- skeletal conditions.
Holland& Barret Saw Palmetto	*HB*	
Viridian Saw Palmetto	*Vir*	To support healthy prostate function
Phytopharmica Saw Palmetto	*Phy*	

Operators

The different operators known as: OP1, OP2, and OP3 took part in performing the analyses on the same day following the same procedure and utilizing the same lab equipment.

NMR Analysis

10 mg of the samples were dissolved in 1 ml of methanol+0.5% TMS solution, followed by 15 minutes sonication and 5 minutes centrifugation at 13,000 rpm. 650 µL of clear solution was pipette out from all the samples using the same pipette and 1D ^{1}H NMR spectrum were acquired for all of them. The spectra generated were subjected to phase correction and processing using TOPSPIN software.

Results and discussions

Figure 1A and B shows the NMR spectra of the six sets of herbal samples prepared by three different operators, labelled as A (OP1), B (OP2) and C (OP3) for each set. Comparison of the spectra was made after the spectra calibration using the internal standard, TMS (δ=0 ppm).

The different of spectral pattern among all samples can only be observed on two samples; JA and HB. In those two samples, the sample spectrum prepared by OP1 was different from the other two operators. Spectra inconsistency can arise from the difference of sample composition in each capsule, or degree of sample dissolution in the solvent used, or due to errors done by the operators during sample preparation. The spectra of other samples showed consistency between the three operators.

The three set of samples; HB, Vir, and Phy; that contained saw palmetto as the main ingredient (Figure 1B) showed the present of common peaks in each set of sample, but with different intensity. This reflects the consistency of the active contents that present in the herbal preparations. Additional peaks observed in sample HB may be due to the other ingredients of the sample or contamination of the samples prepared by OP1.

Conclusion

This work showed that sample handling by operators can be one type of variables that affecting the quality of the spectra. Consistency throughout the sample preparations and spectra analysis should be maintained to ensure reproducibility of the spectra pattern on similar type of samples.

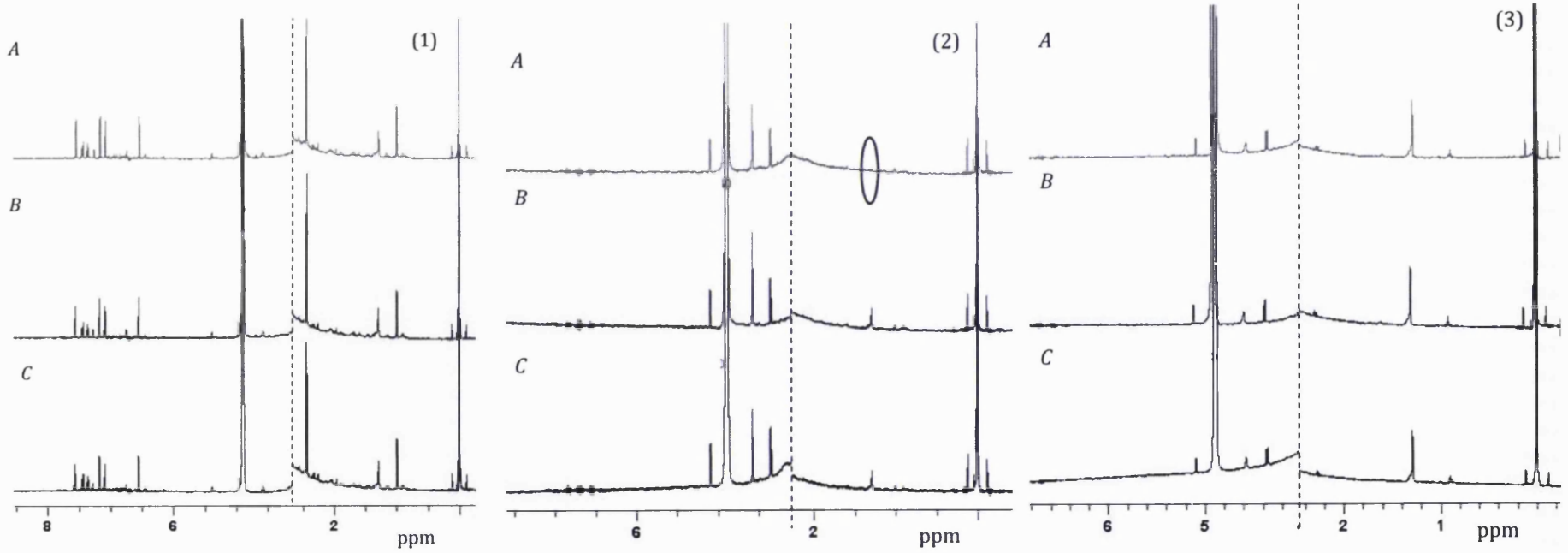


Figure A2-5 The comparison of 1D ^{1}H NMR spectra of samples (1) Kintop (2) Jamu Ajaib and (3) Tunglin, prepared by three different operators; A, B and C.

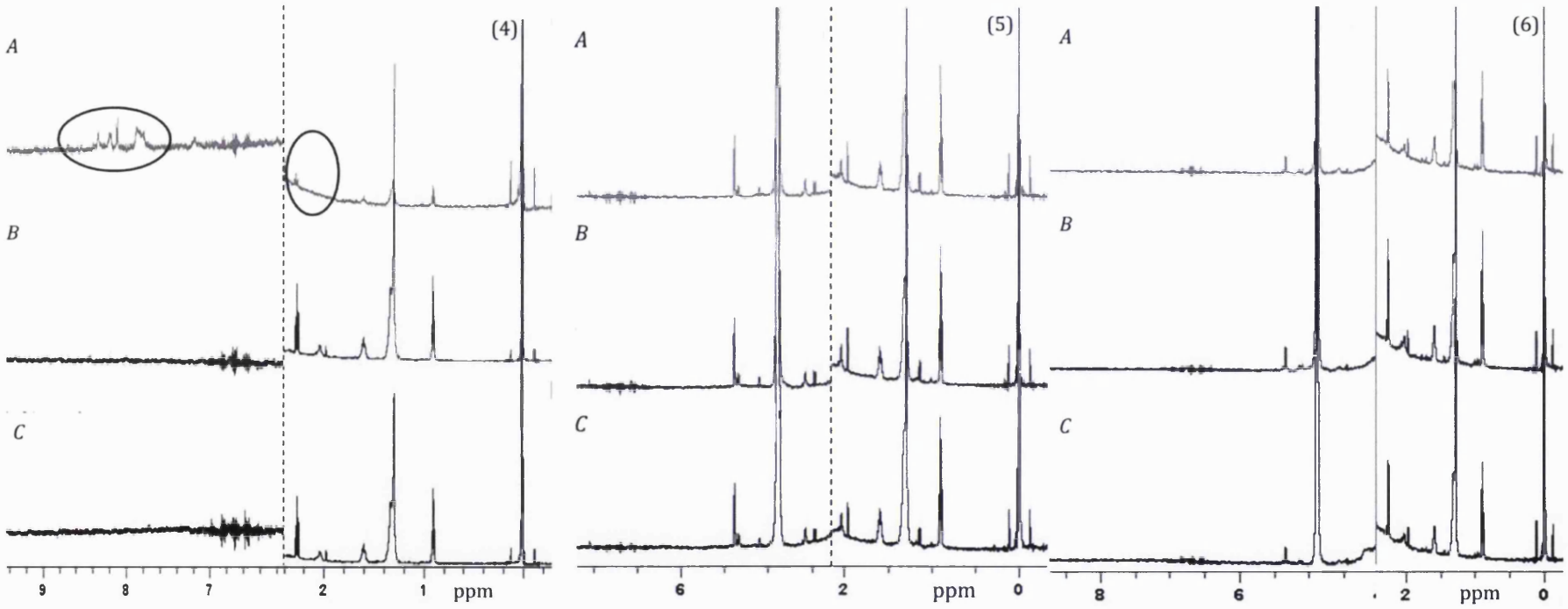


Figure A2-6 The NMR spectra comparison between samples (4) H&B (5) Psy and (6) Vir, prepared by three different operators; A, B and C.

Appendix 3

The NMR analysis conducted on a Bruker Avance 500 MHz spectrometer. Spectra acquisition and processing was done by TOPSPIN 1.3 software. Analysis was performed on 4 selected samples- representative samples from Malaysian and UK batches, a batch of sample with higher content of paracetamol (650mg) and sample PM - aimed to identify the chemical differences among these samples. Similar spectra with only minor differences in region I (figure 5A and B) were observed. This indicated that the contents of excipient soluble in methanol-d_4 were similar in all these products. We believe that slight differences spectra patterns in the region between 1.9ppm- 2.3 ppm is not a cause for sample PM to be classified outside of 95 % interval. The results of classification are probably due to manufacturing processes which lead to the differences in quality of tablets.

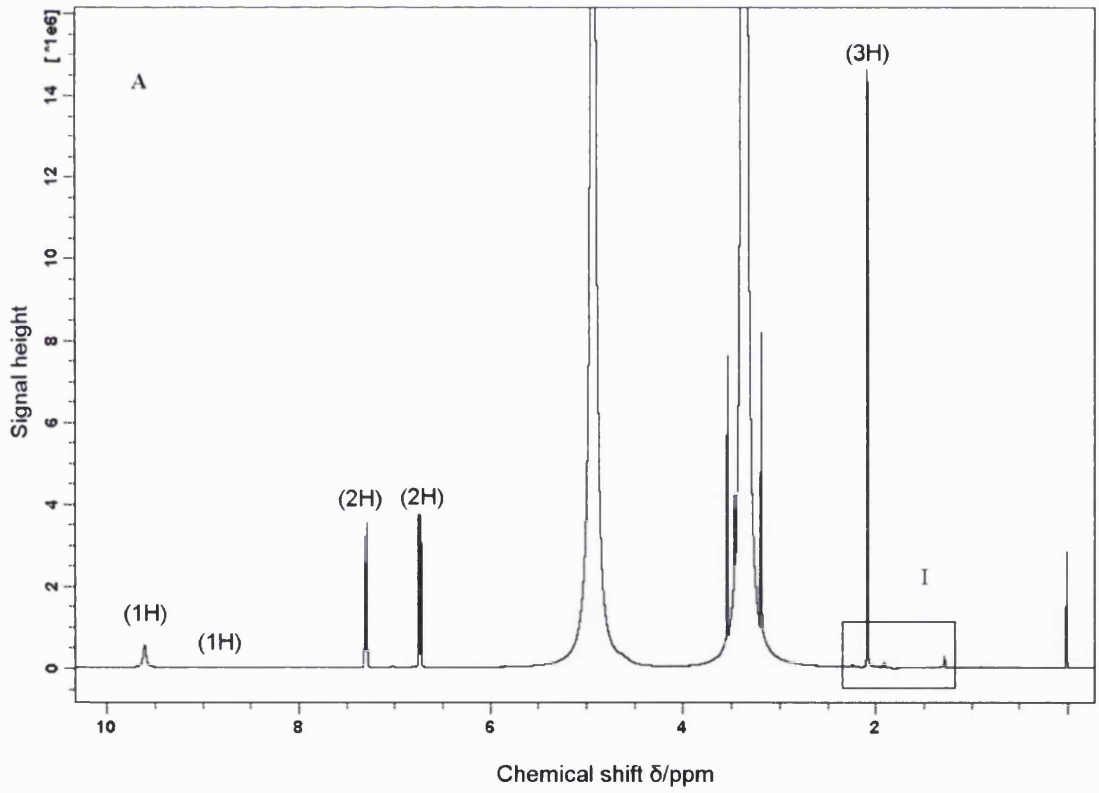


Figure A3-1 The 1D ^1H NMR spectrum of paracetamol tablet dissolved in methanol-d_4.

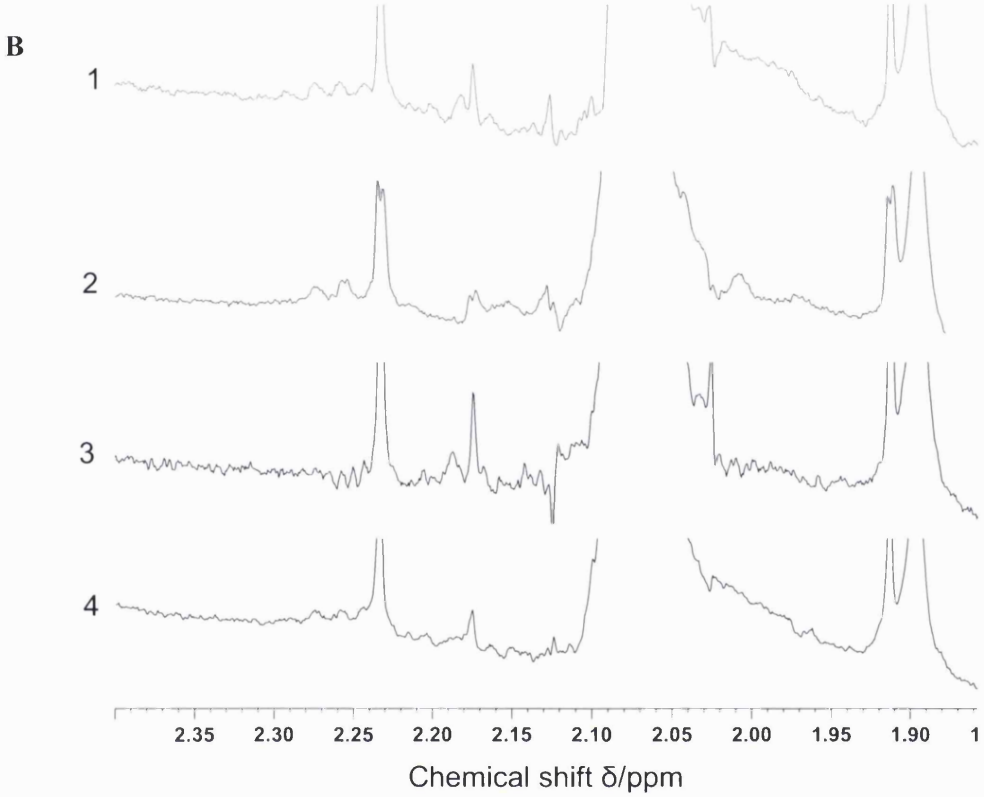


Figure A3-2 The expansion of the 1D ^{1}H NMR spectra samples from 4 categories; Mal 500mg (1), Mal 650mg (2), PM (3), UK 500mg (4) showed varying signals in region I (1.6-2.8 ppm).

Appendix 4

The establishment of validation methods for analyzing paracetamol tablets; development of PCA models for paracetamol samples.

Objective

To establish validation methods for analyzing paracetamol tablets.

To develop PCA model for each set of paracetamol sample.

Materials and methods

Drug samples

The paracetamol samples used in this experiment can be divided into two sets, based on the origin of production; in Malaysia and in the United Kingdom. Malaysian's samples set is consisting of 15 batches of 500mg tablets, 1 batch of 650mg tablets (marked *). It is believed that these tablets were produced by 11 different manufacturing companies. The UK samples consisting of 6 batches of 500mg PCM tablets from 3 different manufacturers. For each set, 20 tablets from each batch were used for analysis. More information on these samples is shown in Table 1 and Table 2.

Table A4-1 Details on PCM samples from UK manufacturers.

Brand names	Label	Manufacturer
Boots	BTS	Hamol Limited
Boots- value heath	VH	Galphram
Paracetamol Lidl	LIDL	Bristol Lab
Galpharm	MRS	Galpharm
GSL	GSL	M&A Pharm.
Bristol Lab	WTR	Bristol Lab

Table A4-2 Details on PCM samples from Malaysian's manufacturers

Brand names	Label	Manufacturer
Poro	PR	YSP industries
Fepril	FP	Idaman Pharma
Uphamol	UPA, UPB, UPC*	CCM Pharm.
Biogesic	BGA, BGB, BGC	UNAM Corp.
Milidon	MIL A, MIL B	Malayan Pharm.
Oralcet	OR	Scan Lab
Paracil	PC	SM
Pamol	PM	Hovid Pharma
Pritamol	PT	Prime Pharm.
Ifimol	IF	Unique Pharm.
Progesic	PG	Xepa Patisson

NIR analysis

The analyses were carried out using a FOSS NIRSystems 6500 spectrophotometer equipped with a Rapid Content Analyzer (Silver Springs, MD, USA). 8 accumulations of 32 scans on average (4 on each side, of different rotation), were recorded for each sample over the wavelength between 1100 – 2498 nm. The spectra were collected in reflectance mode and all the measurements were made relative to the instrument's ceramic reference. New reference was acquired hourly throughout the analysis. The experiment was conducted under a monitored humidity and ambient temperature. The instrument is controlled by Vision® Spectral Analysis Software for Windows (FOSS NIRSystems, version 2.11) for data acquisition. The average of each spectrum, taken in the original form, was saved in ASCII format and then converted to JDX files by using an in-house program. The spectra files were then imported into the Unscrambler 9.7 software (CAMO, Oslo) for chemometric manipulation.

Principal Component Analysis (PCA)

PCA was conducted on the data with leverage correction as the validation method and scaling factor was set as 1. PCA treatment of a set of spectra will decompose them into scores and loadings for variables called principal components (PC). After the first run, the presence of outliers, groups, clusters or trends were determined based on the observation of the score plots. The removal of outliers was made one by one for samples with more than one outlier. Several runs were made and successively exclude the outliers before the data set was satisfactorily characterized. The optimum number of PC was then determined based on the total explained variance plot. The PCA model was saved, and later used in SIMCA application

Results and discussion

PCA model was created for each of the sample sets using the raw spectra obtained. A model is consisting of 4 plots; score, loading, influence and residual variance. The score plots show the locations of the samples along each model component, and used to detect sample patterns, groupings, similarities or differences. In this plot, outliers have been identified by observing a sample point which is abnormal compared to the major part of the data. They were usually

points that are apparently does not belong to the same population, or that are badly described by a model. The software was equipped with outlier detection methods. However, most of the assigned outliers were due to error in the data. Thus, these outliers were further being investigated to determine if they were the real outliers or not before being removed.

Other than pattern of the score plot, residuals and leverage was also being observed as the indicators for outlier. Residuals measure how well samples or variables fit the model determined by the components. Samples with a high residual are poorly described by the model, which nevertheless fits the other samples quite well. Leverages, on the other hand, measure the distance from the projected sample (i.e. its model approximation) to the centre (mean point). Samples with high leverages have a stronger influence on the model than other samples. An influential outlier (high residual + high leverage) is the worst case; however it can easily be detected using an influence plot.

Once the outliers have been removed from the plot, the optimum number of principal component is decided. This is usually the number of PCs that gives the clearest break point in the total residual or explained variance plot (Esbensen, 2006). Figure 6A and B will demonstrate the development of PCA model for sample BGB, before and after the removal of the outlier, respectively.

There are 7 outliers identified by the software. However, it was clearly seen on the score plot that only sample 1 is located furthers from the rest of the samples. The influence plot showed that sample 1 has a very high residual value despite low in leverage. Removal of sample 1 also marked the increase of the percentage of explained variance. These were the indicators that pointed out sample 1 a real outlier. The other data points identified as outliers were removed successively after sample 1. However, the percentage of explained variance did not increase after the data point removal, which indicated that other samples were not the true outliers.

Therefore, a new model was created without sample 1 (figure 6B). The optimum principal component of this model was determined at PC3, where maximum variation is described by the model.

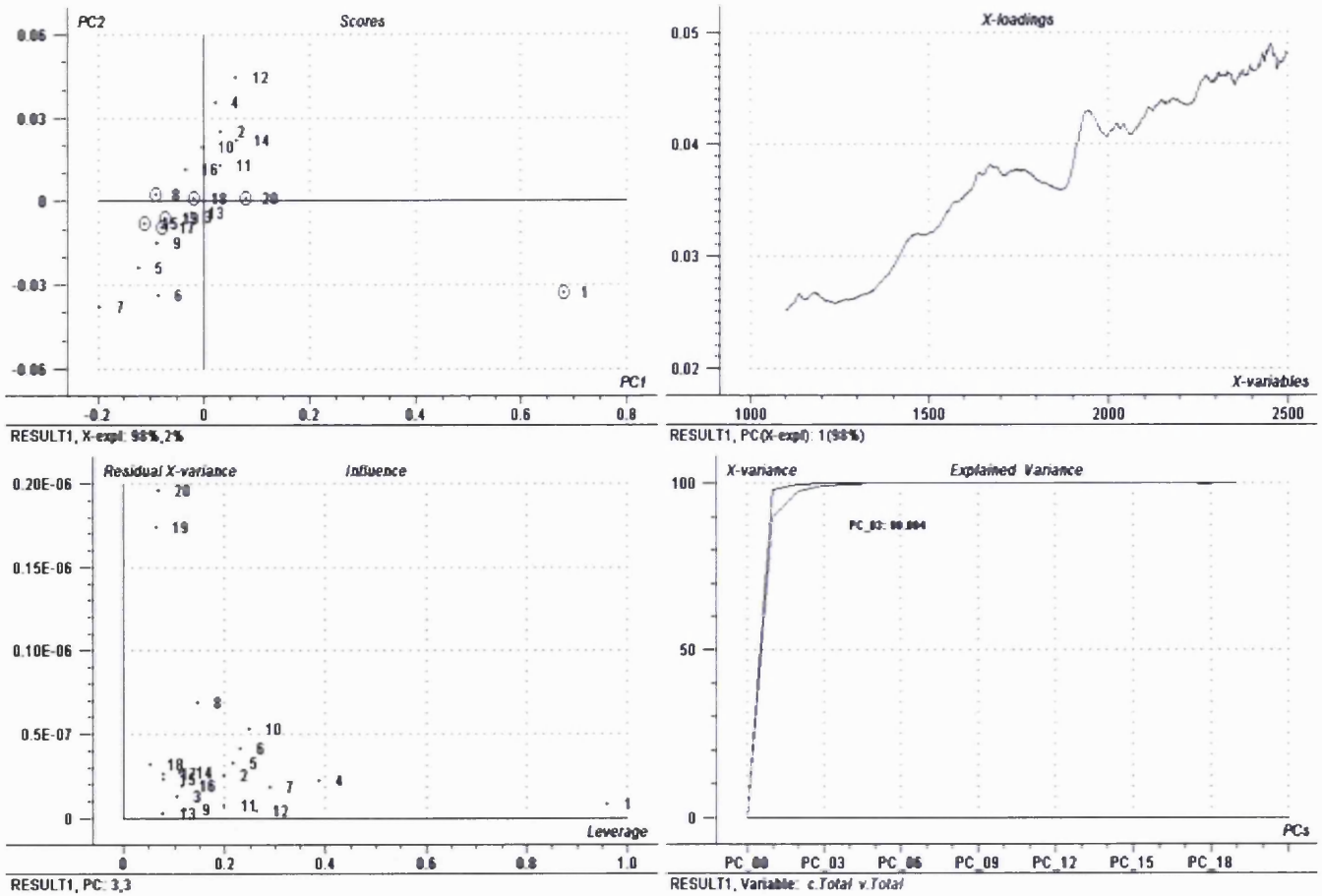


Figure A4-1 PCA model for 20 average spectra of BGB samples. From top left going clockwise- score, loading, influence and residual plots.

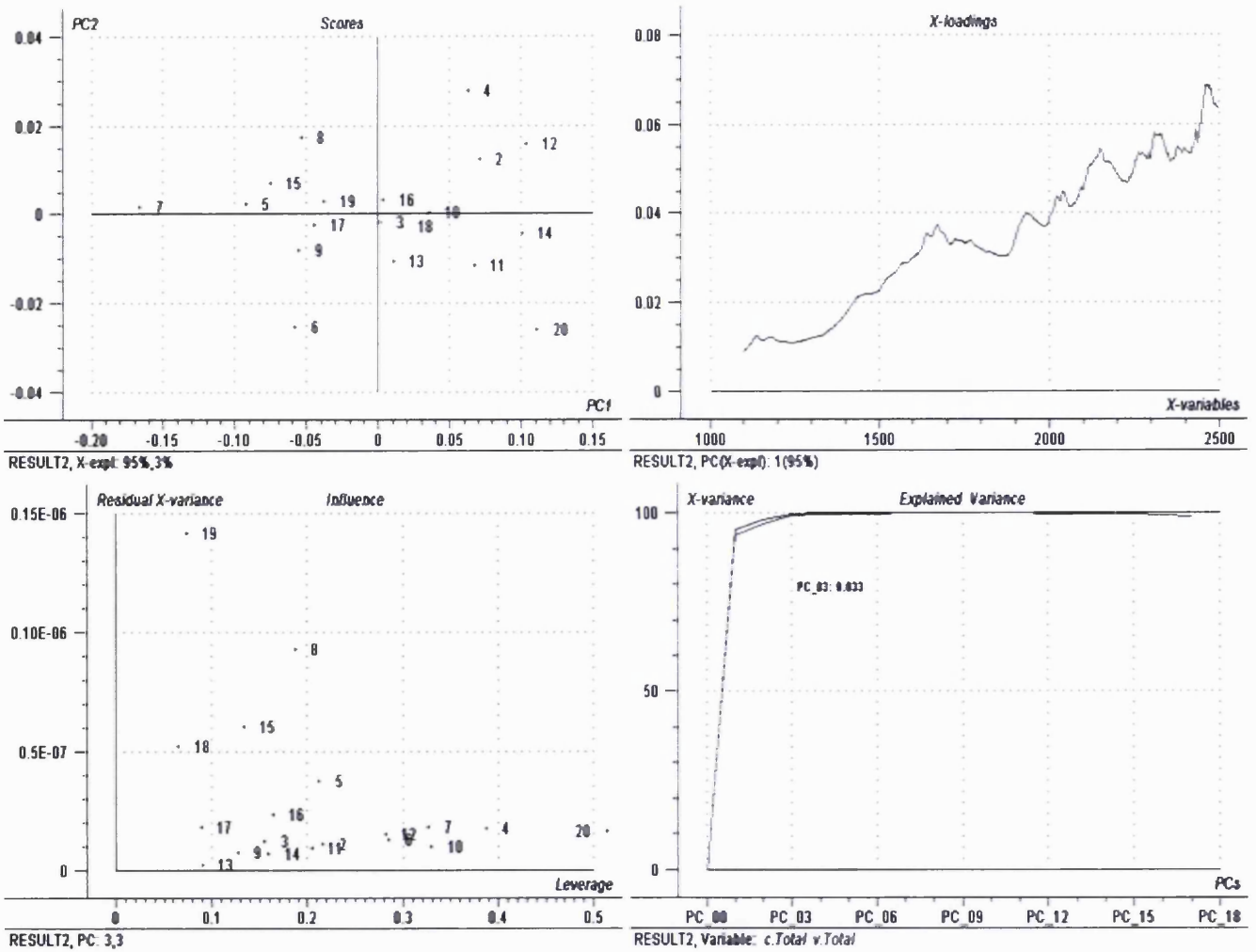


Figure A4-2 PCA model of 19 average spectra of BGB samples after the removal of one outlier (sample 1.)

The same process was repeated for all the samples. The outliers assigned by the software, true outliers selected, and the optimum principal component for PCA model is as stated in table 10A and B, for UK and Malaysian samples, respectively.

Table A4-3 The assigned/ selected outliers and PC for PCA model of the UK samples

Samples	Assigned outliers	Selected outliers	PC
VH	13,16,17,19	-	2
LIDL	1, 8	-	3
BTS	15	15	2
WTR	13	13,15	2
GSL	2,8,10,11,18,19	-	2
MRS	14	-	2

Table A4-4 The assigned/ selected outliers and PC for PCA model of Malaysian samples

Samples	Assigned outliers	Selected outliers	PC
BGA*	1, 17	-	1
BGB*	1	1	3
BGC*	7, 16	-	1
UPA#	2,3,10,12	3	2
UPB#	4, 7, 16	-	2
UPC# (10 samples only)	-	-	3
IF	2	20	3
OR	14	-	2
PC	6, 11,13,18	6	2
PG	4, 6, 13,17, 20	6	2
PM	10, 17	10	2
PR- distributed into 2 clusters (separate model)	-	-	2
Model A(run 3) Model B(run 4)	4	-	4
PT (11 samples only)	8	-	2
FP	6,11,14	13	2
MIL A^	5, 11	-	1
MIL B^	10,15	1,2,3	1

It can be seen from table 10A and B that some of the samples (e.g. VH, GSL, UPA, PC, and PG) have too many outliers detected by the software application. This is because

the automatic outlier detection limits is set at a very low level. This limit is not suitable for NIR spectra which are usually very precise and therefore causing many objects to be displayed as outlying. The distances of the samples from the limit are determined by studying the outlier warnings list. This information is taken into consideration in judging the real outliers.

Throughout the PCA analysis, principal component up to 4 is sufficient in representing all the variables of the samples. The main difficulty in developing a model is to justify the true outliers. While vast experience is needed before one is competence in determining the outliers, the method used in this analysis is purely based on their significant geometrical distance from the rest of the samples, samples with high leverage and high residual x-variance. The removal of the true outliers will cause the percentage of explained variance to increase (near to100%) or percentage of residual variance to drop (near to 0%). However, the number of principal components does not necessarily reduce after the outlier removal.

Conclusion

The validation method for analysing paracetamol tablets has been established in this work. In the analysis by NIR, few samples have been identified as the outliers by algorithms implemented in the Unscrambler statistical/chemometrics software. However, 'fingerprint' analysis needs to be conducted before a sample can be deduced as a true outlier. PCA models can be used for further classification analysis using SIMCA.

Appendix 5

NMR spectra: Jamu Ajaib

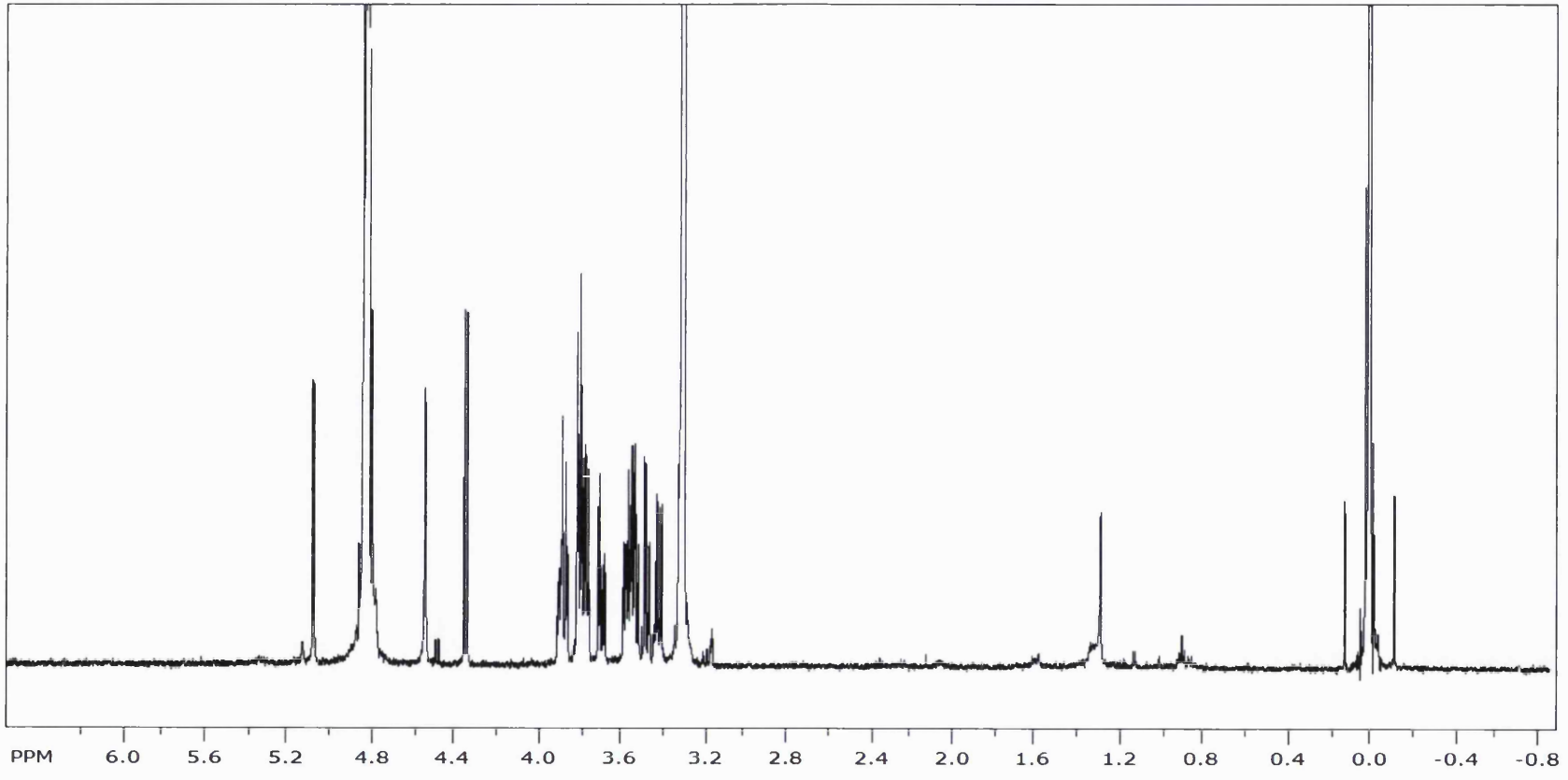


Figure A5-1 The NMR spectra of a herbal product, *Jamu Ajaib*.

Appendix 6

Determination of maximum acceptable HQI value for the identification of herbal mixture.

The NIR analysis of herbal products is more challenging compared to their synthetic counterparts as most of the actives shared the same chemical components which caused the spectra to be overlapped and thus, leading to ambiguous results. Therefore, using the cut-off values determined in Chapter 4 may be less suitable for these types of samples. This work represented the initial effort in establishing the cut-off point in the analysis of complex herbal mixture, which focused on determining the maximum acceptable HQI point that lead to acceptable database search outcome. Observation was made on the database search for the identification of five complex mixture samples as presented in this thesis.

Table A6-1 Database search outcome of five complex herbal mixtures showing the first three hit samples with their HQI value.

Unknown Samples	*HQI*		*Sample ID*
Pronoton	**1**	**0.2119430**	**viagra01**
	2	0.2181080	viagra02
	3	0.2635871	voren05
Jamu Ajaib	**1**	**0.3214652**	**macgel28**
	2	0.3269735	simvastatin39
	3	0.3321219	macgel27
Lami	**1**	**0.3877507**	**Macgel28**
	2	0.3954883	Amoxicillin P
	3	0.3991583	Amoxicillin E
Tunglin	**1**	**0.2076202**	**sim39**
	2	0.2189064	Sibu01
	3	0.2713994	sim36
Counterfeit Viagra	**1**	**0.0908532**	**Viagra**
	2	0.1521653	Voren07
	3	0.1564789	Voren05

Based on this observation, the HQI for the first hit for the unknown sample in each example was noted to be within the range of 0.09 to 0.388. Thus, based on these and few other examples conducted in this research, the acceptable HQI for identification of complex herbal mixture was set to be 0.4 or below. Spectral identification with HQI more than 0.4 is considered to be unacceptable and will be ignored to avoid over-interpretation of the search outcome.

Appendix 7

Quantification of sildenafil citrate by NMR

Objective

To quantify the content of sildenafil citrate in counterfeit and original Viagra sample.

Materials and methods

Sample

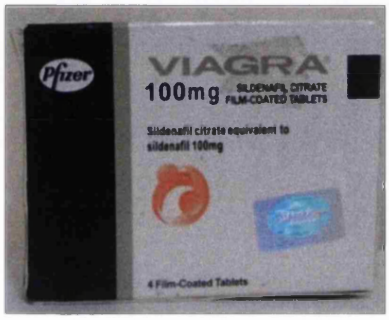


Figure A7-1 Counterfeit Viagra sample used in analysis

A set of four tablets of counterfeit Viagra sample obtained from the NPCB were used for analysis. A genuine sample purchased in Malaysian pharmacy was used for comparison.

NMR analysis

Quantification of the sample was made based on two reference compound; paracetamol and TSP. NMR analysis was conducted by dissolving 20 mg of the powdered Viagra counterfeit and original sample in 1ml of D2O+0.03% TSP solution. The sample was prepared following general procedure of NMR sample preparation as discussed earlier.

Quantification of sildenafil citrate in counterfeit Viagra sample

Quantity of sildenafil $m_A$was calculated by using formula:

$$m_A = m_B \times \frac{N_B}{N_A} \times \frac{\alpha_A}{\alpha_B} \times \frac{M_A}{M_B} \qquad [1]$$

Where, A represent sildenafil citrate, B represent TSP, α is the area under selected peaks, N is the numbers of hydrogen responsible for selected peaks, $m_A/m_B$and MA/MB are the weight and molecular masses of sildenafil and paracetamol respectively.

Results and discussion

Quantification using paracetamol as reference

Four different peaks associated with –CH3, and –CH2 (groups at C-13, C-21, C-20 and C-15 positions of sildenafil citrate) were used for calculation based on equation 1. Figure 2 shows the integrated signals of these peaks in genuine (A) and counterfeit (B) Viagra sample after standardization with the peak of the reference compound, TSP.

The concentration of TSP in D2O was 0.039% indicated the amount of TSP in 1 ml of solution is 0.39 mg. The value of TSP was substituted in equation (1) and the amount of sildenafil in both sample sets were calculated as described in Table 1.

Table A7-1 Quantification of sildenafil using four different signals with reference to TSP.

Selected signal of sildenafil	Number of proton associated with signal	Value of integral	Quantity of sildenafil in Genuine Viagra	Quantity of sildenafil in Counterfeit Viagra
0.92ppm (C-13)	3	7.64	42.6 mg	26.14 mg
1.45ppm (C-21)	3	7.65	42.6 mg	27.39 mg
4.28ppm (C-20)	2	5.51	46.0 mg	26.58 mg
8.08ppm (C-15)	2	2.43	40.65 mg	25.6 mg

Based on the finding in Table 1, the average weight of sildenafil citrate in the genuine sample was calculated to be 43 mg. This was less than the actual weight of the sample (50 mg). However, this value is acceptable based on the consideration that there may be some un-dissolved sample or the sample remained in the remaining solvent.

However, the average amount of sildenafil in counterfeit Viagra was found to be about 26 mg, which is only about 25% of label claimed, suggesting that the counterfeit formulation contains lower amount of active ingredient.

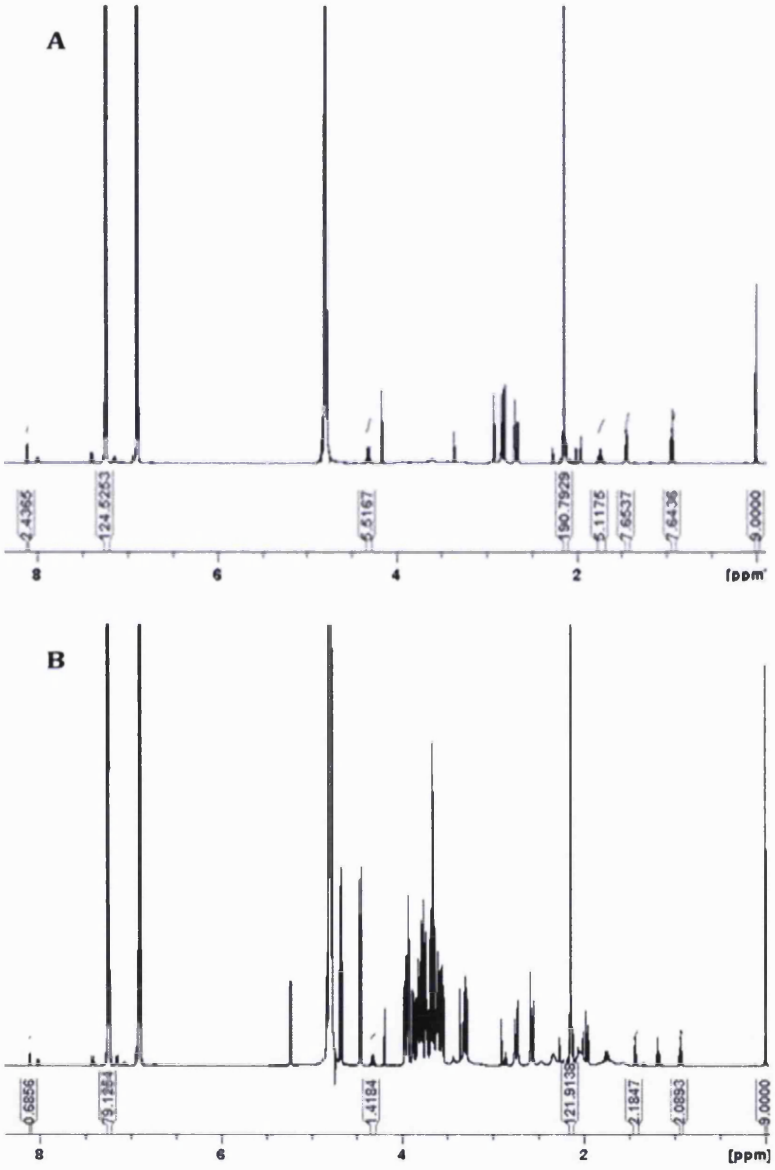


Figure A7-2 Integrated signals of sildenafil in the spectrum of genuine (A) and counterfeit (B) Viagra with reference to TSP.

Conclusion

Quantification of the sildenafil citrate shows that the counterfeit sample contained less active ingredient that purported.

APPENDIX 8

H [ED]		H[AV]		H[1D.AV]		H[LS]	
Sample ID	**HQI**	**Sample ID**	**HQI**	**Sample**	**HQI**	**Sample ID**	**HQI**
pcm	0.0258	pcm	0.0041	pcm	0.0001	pcm	2.02E-05
pcm	0.0850	pcm	0.0072	pcm	0.00028	pcm	7.98E-05
pcm	0.0873	pcm	0.0076	pcm	0.00029	pcm	8.64E-05
pcm	0.1263	pcm	0.0102	pcm	0.00042	pcm	0.000149
pcm	0.1401	pcm	0.0105	pcm	0.00045	pcm	0.000171
ginkgo b.	0.4111	eurycoma l.	0.0348	eurycoma l.	0.00223	eurycoma l.	0.002075
ginkgo b.	0.4619	ginkgo b.	0.0356	amoxicillin	0.00226	ginkgo b.	0.002075
ginkgo b.	0.4783	ginkgo b.	0.0366	amoxicillin	0.0023	ginkgo b.	0.00216
mixed herbs	0.4798	mefenamic a.	0.0389	eurycoma l.	0.00231	eurycoma l.	0.002183
mixed herbs	0.5166	eurycoma l.	0.0399	mefenamic a.	0.00232	eurycoma l.	0.002839
eurycoma l.	0.5220	mefenamic a.	0.0408	amoxicillin	0.00235	amoxicillin	0.002983
eurycoma l.	0.5344	amoxicillin	0.0408	eurycoma l.	0.00236	mefenamic a.	0.003024
eurycoma l.	0.5951	amoxicillin	0.0418	amoxicillin	0.00236	amoxicillin	0.003068
mefenamic a.	0.6009	amoxicillin	0.0419	mixed herbs	0.00238	amoxicillin	0.003086
amoxicillin	0.6068	amoxicillin	0.0440	ginkgo b.	0.00239	mefenamic a.	0.003211
amoxicillin	0.6082	eurycoma l.	0.045984	mefenamic a.	0.00247	amoxicillin	0.003259
mefenamic a.	0.6251	mixed herbs	0.06303	ginkgo b.	0.00252	mixed herbs	0.005605
amoxicillin	0.6332	mixed herbs	0.064315	mixed herbs	0.00264	mixed herbs	0.005671
amoxicillin	0.6416	ginkgo b.	0.067266	ginkgo b.	0.00283	ginkgo b.	0.007079

Table A8-1 The complete hit-list supplementing the result in Figure 4.7

H[1D.LS]		H[C]		H[1D.C]	
Sample ID	**HQI**	**Sample ID**	**HQI**	**Sample ID**	**HQI**
pcm	.	pcm	0.00276	pcm	0.001518
pcm	2E-07	pcm	0.025924	pcm	0.005866
pcm	2E-07	pcm	0.030236	pcm	0.011604
pcm	4E-07	pcm	0.064826	pcm	0.022627
pcm	4E-07	pcm	0.066177	pcm	0.029595
eurycoma l.	1.27E-05	ginkgo b.	0.61633	ginkgo b.	0.89718
eurycoma l.	0.000013	ginkgo b.	0.743472	eurycoma l.	0.909758
amoxicillin	1.31E-05	ginkgo b.	0.747209	mixed herbs	0.909952
eurycoma l.	1.33E-05	mixed herbs	0.820397	mefenamic a.	0.911538
ginkgo b.	1.33E-05	eurycoma l.	0.883607	eurycoma l.	0.915846
amoxicillin	1.34E-05	eurycoma l.	0.888735	mefenamic a.	0.917531
mixed herbs	1.36E-05	mixed herbs	0.891663	ginkgo b.	0.920671
amoxicillin	1.38E-05	mefenamic a.	0.898465	eurycoma l.	0.928256
amoxicillin	1.39E-05	mefenamic a.	0.906837	mixed herbs	0.928485
mefenamic a.	0.000014	eurycoma l.	0.958541	amoxicillin	0.932892
ginkgo b.	0.000015	amoxicillin	0.98185	ginkgo b.	0.933641
mixed herbs	1.52E-05	amoxicillin	0.986505	amoxicillin	0.937961
mefenamic a.	1.56E-05	amoxicillin	0.987111	amoxicillin	0.943137
ginkgo b.	1.82E-05	amoxicillin	0.988343	amoxicillin	0.943313

Table A8-1(cont.) The complete hit-list supplementing the result in Figure 4.7

Table A8-2 The complete hit-list supplementing the result in Figure 4.8A (correlation algorithm)

Search: *Ginkgo biloba*		**Search: Amoxicillin**		**Search: Mefenamic acid**		**Search: Chinese herbal mix.**		**Search: *Eurycoma longifolia***	
Sample ID	**HQI**	**Sample ID**	**HQI**	**Sample ID**	**HQI**	**Sample ID**	**HQI**	**Sample ID**	**HQI**
ginkgo b.	0	amoxicillinA	0	mefenamic a.	0	mixed herbs	0	eurycoma l.GB	0
ginkgo b.	0.212675	amoxicillinH	0.009527	mefenamic a.	0.013464	eurycoma l.	0.070151	mixed herbs	0.084951
ginkgo b.	0.277018	amoxicillinC	0.038702	pcm	0.867262	eurycoma l.	0.196011	eurycoma l.HUR	0.108079
mixed herbs	0.346116	amoxicillinR	0.049509	pcm	0.883726	mixed herbs	0.203838	eurycoma l.LKH	0.124555
mixed herbs	0.446046	mixed herbs	0.535063	pcm	0.885025	eurycoma l.	0.206921	mixed herbs	0.206921
eurycoma l.	0.467018	ginkgo b.	0.684595	mixed herbs	0.885351	ginkgo b.	0.231436	ginkgo b.	0.393283
eurycoma l.	0.493287	eurycoma l.	0.695737	pcm	0.885515	ginkgo b.	0.429346	ginkgo b.	0.467018
eurycoma l.	0.68995	eurycoma l.	0.797263	pcm	0.894864	ginkgo b.	0.446046	ginkgo b.	0.536188
pcm	0.746405	mixed herbs	0.826867	ginkgo b.	0.944875	amoxicillin	0.535063	amoxicillin	0.865134
pcm	0.808912	ginkgo b.	0.830808	eurycoma l.	0.972063	amoxicillin	0.546846	Pcm	0.879007
amoxicillin	0.812658	ginkgo b.	0.838565	eurycoma l.	0.9721	amoxicillin	0.632963	amoxicillin	0.879051
pcm	0.814898	eurycoma l.	0.865134	eurycoma l.	0.972999	amoxicillin	0.664575	Pcm	0.914775
amoxicillin	0.838565	mefenamic a.	0.990752	mixed herbs	0.989532	pcm	0.899669	amoxicillin	0.933039
amoxicillin	0.86083	mefenamic a.	0.991203	amoxicillin	0.991203	pcm	0.938701	Pcm	0.943173
pcm	0.861489	pcm	0.992988	amoxicillin	0.996638	pcm	0.962789	mefenamic a.	0.945857
amoxicillin	0.867457	pcm	0.995655	amoxicillin	0.999578	mefenamic a.	0.976912	amoxicillin	0.949817
pcm	0.874009	pcm	0.999073	ginkgo b.	0.999694	pcm	0.986818	mefenamic a.	0.972063
mefenamic a.	0.938166	pcm	0.999687	amoxicillin	0.999791	pcm	0.988402	Pcm	0.972201
mefenamic a.	0.944875	pcm	0.999986	ginkgo b.	0.999964	mefenamic a.	0.989532	Pcm	0.984808

Table A8-3 The complete hit-list supplementing the result in Figure 4.8B (first derivative correlation algorithm)

Search: *Ginkgo biloba*		**Search: Amoxicillin**		**Search: Mefenamic acid**		**Search: Chinese herbal mix.**		**Search: *Eurycoma longifolia***	
Sample ID	HQI	Sample ID	HQI	Sample ID	HQI	Sample ID	HQI	Sample ID	HQI
ginkgo b.	0	amoxicillinA	0	mefenamic a.	0	mixed herbs	0	eurycoma l.GB	0
mixed herbs	0.393869	amoxicillinC	0.004133	mefenamic a.	0.004232	eurycoma l.	0.037493	eurycoma l.LKH	0.135291
ginkgo b.	0.417504	amoxicillinR	0.005374	pcm	0.90709	mixed herbs	0.121089	mixed herbs	0.158753
eurycoma l.	0.499096	amoxicillinH	0.005563	pcm	0.907326	ginkgo b.	0.128873	eurycoma l.HUR	0.187557
eurycoma l.	0.503034	mixed herbs	0.70321	pcm	0.909774	eurycoma l.	0.181361	ginkgo b.	0.214134
mixed herbs	0.528895	eurycoma l.	0.720016	pcm	0.914559	eurycoma l.	0.222266	mixed herbs	0.222266
eurycoma l.	0.618809	ginkgo b.	0.750073	pcm	0.916263	ginkgo b.	0.43893	ginkgo b.	0.499096
ginkgo b.	0.620284	mixed herbs	0.798672	mixed herbs	0.977169	ginkgo b.	0.528895	ginkgo b.	0.54285
amoxicillin	0.869231	eurycoma l.	0.80607	amoxicillin	0.987788	amoxicillin	0.693908	amoxicillin	0.84658
amoxicillin	0.877407	ginkgo b.	0.8246	amoxicillin	0.987935	amoxicillin	0.693982	amoxicillin	0.853383
amoxicillin	0.878824	eurycoma l.	0.858545	amoxicillin	0.988316	amoxicillin	0.702838	amoxicillin	0.858545
amoxicillin	0.881372	ginkgo b.	0.881372	amoxicillin	0.988338	amoxicillin	0.70321	amoxicillin	0.859744
pcm	0.934397	pcm	0.942682	eurycoma l.	0.989549	pcm	0.934941	pcm	0.915004
pcm	0.950645	pcm	0.951646	ginkgo b.	0.992126	pcm	0.956177	pcm	0.937746
pcm	0.956151	pcm	0.957228	ginkgo b.	0.993892	pcm	0.966308	pcm	0.947476
pcm	0.965258	pcm	0.961643	mixed herbs	0.994751	pcm	0.973201	pcm	0.954433
pcm	0.969587	pcm	0.969144	eurycoma l.	0.995423	pcm	0.981725	pcm	0.963883
mefenamic a.	0.988624	mefenamic a.	0.983938	eurycoma l.	0.997315	mefenamic a.	0.98836	mefenamic a.	0.979856
mefenamic a.	0.993892	mefenamic a.	0.987788	ginkgo b.	0.997813	mefenamic a.	0.994751	mefenamic a.	0.989549

Table A8-4 The complete hit-list supplementing the result in Figure 4.9A (peak search-forward)

Paracetamol		Amoxicillin		Simvastatin		Cefuroxime axetil		Diclofenac sodium		Mefenamic acid		Sibutramin	
HQI	Sam.ID	HQI	Sam.ID	HQI	Sam.ID	HQI	Sam.ID	HQI	Sam.ID	HQI	Sam.ID	HQI	Sam.ID
41.66667	PM	29.04412	AmoxD	27.84091	sim36	27.08333	cef22	30.59896	gs	36.45834	P57	15.07353	Sibu01
31.59722	MIL A	24.63235	IF	23.29546	sim39	21.35417	cef83	27.08333	cef 86	36.11111	P7	12.5	gk
31.59722	OR	22.79412	AmoxH	22.15909	sim64	20.83333	cef 86	27.08333	cef83	36.11111	P10	12.5	gh
31.59722	UPA	21.32353	AmoxG	20.45455	gk	18.22917	gbta	26.38889	cef22	36.11111	P25	12.13235	sim64
31.25	FP	20.58824	Amox E	20.45455	IF	18.22917	LKH	25	mac27	36.11111	P49	11.76471	sim36
31.25	PC	20.22059	AmoxR	19.88636	Sibu01	18.22917	gh	25	mac28	36.11111	P51	11.02941	sim39
31.25	PG	19.48529	AmoxC	17.61364	P3	17.1875	g80	24.30556	viagra ori	36.11111	P52	10.29412	nuprep
31.25	PT	19.11765	AmoxB	17.61364	P7	16.66667	Amox C2	23.61111	gh	36.11111	P53	8.823529	P49
31.25	UPC	18.38235	Amox A	17.61364	P10	15.10417	sim64	20.83333	mac29	36.11111	P54	8.823529	IF
31.25	PR	18.38235	P57	17.61364	P12	15.10417	viagra ori	20.22569	IF	36.11111	P54	8.455882	P7
30.90278	MIL B	18.01471	AmoxI	17.61364	P14	13.54167	g60	19.44444	Tonex	36.11111	P58	8.455882	P10
26.38889	BGA	18.01471	P7	17.61364	P19	13.02083	vor07	18.92361	tic31	31.59722	P12	8.455882	P12
26.38889	BGC	18.01471	P10	17.61364	P25	12.5	ok3	18.92361	tic32	31.59722	P14	8.455882	P14
24.04514	IF	18.01471	P25	17.61364	P37	12.5	mac27	18.7934	tic34	31.59722	P19	8.455882	P19
22.04861	P49	18.01471	P51	17.61364	P39	11.45833	mac29	18.75	putrem	31.59722	P39	8.455882	P25
21.35417	P7	18.01471	P52	17.61364	P43	10.41667	nuprep	18.05556	gg120	31.59722	P43	8.455882	P39
21.35417	P10	18.01471	P53	17.61364	P43	10.41667	OK1	15.97222	vor05	31.59722	P43	8.455882	P43
21.35417	P12	18.01471	P54	17.61364	P46	10.41667	gg120	15.97222	vor07	31.59722	P46	8.455882	P43
21.35417	P14	18.01471	P54	17.61364	P47	10.41667	mac28	14.58333	g80	31.59722	P47	8.455882	P46
21.35417	P19	18.01471	P58	17.61364	P48	10.41667	P57	14.58333	Amox E	31.59722	P48	8.455882	P47

Ticlopidin		Sildenafil citrate		Mcgel		*G.biloba*		*E.longifolia*		Chinese herb 1		Chinese herb 2	
HQI	**Sam.ID**	**HQI**	**Sam.ID**	**HQI**	**Sam.ID**	**HQI**	**Sam.ID**	**HQI**	**Sam.ID**	**HQI**	**Sam.ID**	**HQI**	**Sam.ID**
16.5625	IF	19.1666	ok2	28.125	mac2	23.66072	g40	41.0714	hurix	36.875	Sibu01	24.4791	LKH
15	tic34	19.1666	ok3	26.5625	Sibu1	18.75	mac27	37.5	Tonex	35	ok3	24.4791	ok3
13.875	AmoxD	19.1666	ok4	24.375	gh	18.75	sim64	35.7142	g80	30.625	gk	23.4375	hurix
13.875	AmoxG	18.75	hurix	24.375	sim64	16.07143	mac29	28.5714	OK1	29.375	hurix	22.3958	OK1
13.875	AmoxH	16.25	OK1	23.125	mac2	15.17857	mac28	28.5714	gk	27.53906	LKH	20.9635	tic31
13.5	tic31	15	viagra ori	23.125	mac2	14.73214	viagra	28.125	Sibu01	25.625	OK1	20.9635	tic32
13.5	tic32	14.1666	tic34	19.41406	BGA	13.83929	LKH	27.67857	ok2	23.75	vor07	20.8333	mac29
13.3125	cef 86	13.4375	Sibu01	19.41406	BGC	12.05357	gw	27.67857	ok3	22.5	Amox E	16.1458	sim36
13.3125	cef83	13.3854	gs	19.41406	FP	10.26786	gs	27.67857	ok4	22.5	AmoxR	16.1458	sim39
11.9843	P57	13.0208	gh	19.41406	MIL B	9.821429	gk	26.78572	gg120	22.03125	mac29	16.1458	sim64
11.875	Amox A	12.9166	gk	19.41406	PC	9.375	gg120	26.78572	mac27	21.875	AmoxC	15.6575	IF
11.6718	P7	12.9166	sim36	19.41406	PG	9.375	Sibu1	26.33929	gh	21.875	AmoxD	15.625	AmoE
11.6718	P10	12.9166	sim39	19.41406	PM	9.375	sim39	26.33929	mac28	21.875	AmoxG	15.625	vor07
11.6718	P25	12.9166	IF	19.41406	PT	8.92857	putrem	25.89286	nuprep	21.875	AmoxH	15.1041	gbta
11.6718	P49	12.5	g80	19.41406	UPA	8.48214	sim36	25.2232	gs	21.875	AmoxQ	13.6718	tic34
11.6718	P51	12.5	gg120	19.41406	UPC	8.03571	gh	24.1071	viagra	21.25	sim64	13.5416	vor05
11.6718	P52	12.2916	mac27	19.41406	PR	8.03571	Amox A	23.2142	g60	18.75	ok2	13.2812	Tonex
11.6718	P53	12.1875	mac28	19.375	viagra	8.03571	AmoxG	20.6473	LKH	18.75	ok4	10.9375	Amox A
11.6718	P54	11.6666	tic31	18.78906	MIL A	8.03571	AmoxH	19.6428	ckta	18.75	gg120	10.4166	Amox C
11.6718	P54	11.6666	tic32	18.78906	OR	8.03574	AmoxQ	18.3035	cef 86	18.125	Amox A	9.89583	putrem

Table A8-4(cont.) The complete hit-list supplementing the result in Figure 4.9A (peak search-forward)

Table A8-5 The complete hit-list supplementing the result in Figure 4.9B (peak search reverse)

Paracetamol		Amoxicillin		Simvastatin		Cefuroxime axetil		Diclofenac sodium		Mefenamic acid		Sibutramin	
HQI	Sam.ID	HQI	Sam.ID	HQI	Sam.ID	HQI	Sam.ID	HQI	Sam.ID	HQI	Sam.ID	HQI	Sam.ID
95	BGA	69.79167	IF	64.0625	sim39	65	cef22	68.84766	gs	77.67857	P3	46.875	sim39
95	BGC	54.86111	AmoxD	61.25	sim36	54.6875	gbta	56.25	mac27	72.91667	P57	43.75	nuprep
94.79167	MIL A	51.04167	AmoxI	46.875	nuprep	54.16667	g60	56.25	mac28	72.22222	P7	42.70834	Sibu01
94.79167	OR	50	Amox E	45	gk	51.5625	g80	48.75	cef 86	72.22222	P10	42.5	gk
94.79167	UPA	49.10714	AmoxR	40.625	sim64	51.25	cef83	48.75	cef83	72.22222	P25	40	sim36
93.75	FP	48.75	BGA	37.5	IF	50	cef 86	47.5	cef22	72.22222	P49	35.41667	gh
93.75	PC	48.75	BGC	36.45834	Sibu01	37.5	ok3	46.875	mac29	72.22222	P51	34.375	putrem
93.75	PG	48.4375	AmoxH	34.375	putrem	37.5	mac27	43.75	pc	72.22222	P52	34.375	sim64
93.75	PM	47.32143	AmoxC	27.67857	P3	36.45834	gh	43.75	Tonex	72.22222	P53	28.125	gs
93.75	PT	46.42857	AmoxB	27.67857	P37	36.25	viagra ori	43.75	viagra ori	72.22222	P54	27.5	BGA
93.75	UPC	45.3125	AmoxG	26.25	gg120	34.375	mac29	42.1875	putrem	72.22222	P54	27.5	BGC
93.75	PR	41.66667	AmoxM	24.21875	P12	31.25	nuprep	35.9375	vor05	72.22222	P58	25	cef22
92.70834	MIL B	41.66667	AmoxN	24.21875	P14	31.25	OK1	35.41667	gh	71.09375	P12	25	mac29
71.875	IF	41.66667	FP	24.21875	P19	31.25	vor07	32.8125	g80	71.09375	P14	25	IF
54.46429	P37	41.07143	AmoxQ	24.21875	P39	31.25	mac28	32.5	gg120	71.09375	P19	23.75	gg120
48.95834	P45	40.625	UPA	24.21875	P43	30.20833	sim64	30.33854	IF	71.09375	P39	23.4375	mac27
47.65625	P12	39.0625	Amox A	24.21875	P43	28.57143	Amox C2	29.16667	ckta	71.09375	P43	23.4375	mac28
47.65625	P14	39.0625	sim39	24.21875	P46	26.5625	hurix	28.90625	hurix	71.09375	P43	22.91667	FP
47.65625	P19	37.5	AmoxP	24.21875	P47	26.5625	gs	28.75	vor07	71.09375	P46	22.91667	UPA
47.65625	P39	35.9375	P2	24.21875	P48	25	gg120	28.125	sim39	71.09375	P47	21.875	MIL A

Table A8-5(cont.) The complete hit-list supplementing the result in Figure 4.9B (peak search reverse)

Ticlopidin		Sildenafil citrate		Mcgel		*G.biloba.*		*E. longifolia*		Chinese Herb 1		Chinese Herb 2	
HQI	**Sam.ID**	**HQI**	**Sam.ID**	**HQI**	**Sam.ID**	**HQI**	**Sam.ID**	**HQI**	**Sam.ID**	**HQI**	**Sam.ID**	**HQI**	**Sam.ID**
70.3125	mac28	71.875	ok2	70.3125	mac29	65.625	mac27	71.875	hurix	71.875	hurix	51.5625	OK1
69.01042	IF	71.875	ok3	57.8125	mac27	56.25	mac29	65.625	Tonex	67.1875	ok3	50	ok3
66.25	cef 86	71.875	ok4	57.8125	mac28	55.20834	g40	62.5	g80	61.45834	Sibu01	48.4375	hurix
66.25	cef83	70.3125	hurix	45.3125	sim39	53.125	mac28	54.16667	g60	61.25	gk	48.4375	sim39
57.8125	mac27	60.9375	OK1	44.27084	Sibu01	43.75	sim64	50	OK1	56.25	OK1	43.75	mac29
57.8125	sim39	50.19531	gs	40.625	putrem	42.1875	gw	48.4375	ok2	46.875	ok2	40.625	pc
56.25	gs	48.4375	sim39	40.625	gh	41.25	viagra	48.4375	ok3	46.875	ok4	39.84375	Tonex
53.57143	tic34	47.91667	ckta	40.625	sim64	35.9375	gs	48.4375	ok4	45.70313	mac29	38.75	sim36
48.4375	mac29	46.875	g80	39.0625	gs	32.8125	sim39	46.875	mac27	43.75	pc	32.29167	sim64
48.21429	tic31	46.09375	mac27	38.82813	BGA	31.25	putrem	46.09375	mac28	41.25	vor07	31.31511	IF
48.21429	tic32	45.70313	mac28	38.82813	BGC	27.5	gk	45.83334	ckta	37.5	gg120	31.25	tic31
47.91667	sim64	45	viagra ori	38.75	viagra ori	26.25	gg120	45.3125	nuprep	35.9375	vor05	31.25	tic32
46.25	sim36	40.625	putrem	36.25	gk	25	nuprep	44.14063	gs	35.41667	sim64	29.6875	putrem
45	cef22	40.625	Tonex	32.35677	FP	23.75	sim36	40.625	pc	35.15625	Tonex	29.6875	ok2
42.96875	AmoxG	38.75	gk	32.35677	MIL B	23.4375	OK1	40	gk	32.14286	AmoxR	29.6875	ok4
42.96875	AmoxH	38.75	sim36	32.35677	PC	21.875	Sibu01	37.5	gg120	31.25	nuprep	26.5625	gw
40.3125	BGA	37.5	pc	32.35677	PG	21.52778	LKH	33.75	viagra ori	31.25	AmoxC	25	vor07
40.3125	BGC	37.5	gg120	32.35677	PT	18.75	ok2	32.8125	Sibu01	31.25	AmoxQ	23.75	viagra ori
39.28572	AmoxQ	35.41667	g60	32.35677	UPA	18.75	ok3	30.72917	gh	31.25	vor06	23.4375	gs
39.0625	gk	33.59375	Sibu01	32.35677	UPC	18.75	ok4	25.625	cef 86	30.35714	Amox E	22.91667	LKH

Table A8-6 The complete hit-list supplementing the result in Figure 4.10 (forward and reverse peak search)

Peak search					
Forward			Reverse		
Hit #	HQI	Sam.ID	Hit #	HQI	Sam.ID
1	31.59722	pcm	1	95	pcm
2	31.59722	pcm	2	94.79167	pcm
3	31.59722	pcm	3	94.79167	pcm
4	31.25	pcm	4	94.79167	pcm
5	26.38889	pcm	5	93.75	pcm
6	18.57639	mefenamic a.	6	47.32143	mefenamic a.
7	16.49306	mefenamic a.	7	41.96429	mefenamic a.
8	7.291667	ginkgo b.	8	21.875	ginkgo b.
9	6.597222	amoxicillin	9	21.25	ginkgo b.
10	5.902778	ginkgo b.	10	14.84375	amoxicillin
11	3.819444	amoxicillin	11	12.5	eurycoma l.
12	3.472222	amoxicillin	12	12.5	mixed herbs
13	3.472222	amoxicillin	13	9.375	eurycoma l.
14	2.777778	eurycoma l.	14	8.928572	amoxicillin
15	2.604167	eurycoma l.	15	8.928572	amoxicillin
16	2.083333	eurycoma l.	16	8.59375	amoxicillin
17	1.736111	ginkgo b.	17	8.333334	mixed herbs
18	1.388889	mixed herbs	18	6.25	ginkgo b.
19	1.388889	mixed herbs	19	5.208333	eurycoma l.