AUTHENTICATION OF EDIBLE BIRD'S NEST USING ADVANCED ANALYTICAL TECHNIQUES AND MULTIVARIATE DATA ANALYSIS

CHUA YONG GUAN PETER

(B.Sc.(Hons.), NUS)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARMENT OF CHEMISTRY

NATIONAL UNIVERSITY OF SINGAPORE

2014

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety, under the supervision of Professor Li Fong Yau Sam, (in the laboratory S5-02-05), Chemistry Department, National University of Singapore between 03-08-2013 and 10-03-2014.

I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has not been submitted for any degree in any university previously.

The content of the thesis has been partly published in:

1) Metabolite profiling of edible bird nest using GCMS and LCMS Chua, Y.G., Bloodworth, B.C., Leong, L.P. and Li, F.Y.S, Journal of Mass Spectrometry, 2014, 28,1-14.

Chua Yong Guan Peter		10/03/2014
Name	Signature	Date

ACKNOWLEDGEMENT

First, I would like to express my grateful appreciation to my Ph.D supervisor Professor Li Fong Yau Sam and co-supervisor Dr Leong Lai Peng for their support in my Ph.D program in the last 5 years. Prof Li has taught me skills of designing and performing scientific research work that are of the highest quality. In addition, he has provided me with encouragement and sound advice whenever I needed them. I am also grateful to Dr Leong who often took time off from her busy schedule to discuss with me on my research work. Through these discussions, I was able to obtain useful scientific advice and knowledge which greatly aids in my understanding on the field of food authentication.

I would also like to express my gratitude to National University of Singapore (NUS) and Singapore-Peking-Oxford Research Enterprise for Water Eco-efficiency (SPORE) for providing me with the research scholarship for my PH.D studies and all the staff from the Chemistry Department in NUS for the administrative support. Next, I would like to thank Applied Sciences Group at the Health Science Authority (HSA) and Shimadzu (Asia Pacific) Pte Ltd for providing me with the chance of conducting my research work in their laboratories and Eu Yan Seng (Singapore) for agreeing to sponsor the edible bird's nest for the project.

Throughout the course of my PH.D , it has been a great pleasure to work with the post-graduates from Prof Li's laboratory - Dr Fang Guihua, Dr Ji kaili, Dr

TABLE OF CONTENTS

<u>SUMMARY</u>

The authenticity issue involving edible bird's nest (EBN) has affected the consumer's confidence. Instead of relying on current techniques, new analytical methods are developed and applied in combination with multivariate data analysis to tackle the problem of authenticity. Hotelling T2 range plot illustrated that it is feasible to identify EBN as well as, to differentiate matries similar to EBN, thereby resolving the issues of quality control and species of origin of EBN. Classification of EBN according to its coloration, country of origin and production site could be done with metabolite fingerprinting and supervised score plots. Moreover, score plots based on the data from gas chromatography mass spectrometer demonstrated better prediction abilities. In qualitative discrimination, principal component analysis is able to discriminate EBN from spiked samples at the level of 0.5 %. In quantitative analysis, accurate prediction of spiked sample was shown to detect as low as 1 % of adulterants.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AVA	Agri-Food and Veterinary Authority of Singapore
AAS	Atomic absorption spectrometer
ANOVA	Analysis of variance
CE	Collision energy
CS	Constant sum
CXP	Collision exit potential
DC	Direct current
DP	Declustering potential
DNA	Deoxyribonucleic acid
DModXPS	Distance to model X predicted score
EP	Entrance potential
EI	Electron impact
EU	European Union
EAA	Essential amino acid
EBN	Edible bird's nest
ESI	Electrospray ionisation
ELISA	Enzyme-linked immunosorbent assay
FP	Focusing potential
FAO	Food and Agricultural Organization
FDA	Food and Drug Administration
GC/MS	Gas chromatography mass spectrometer
GC-FID	Gas chromatography flame ionisation detector
HSA	Health Science Authority of Singapore

LC	Liquid chromatography
IT	Ion trap
LOD	Limits of detection
LOQ	Limits of quantification
LC/MS	Liquid chromatography mass spectrometer
ICP-MS	Inductively-coupled plasma mass spectrometry
LC/MSMS	Liquid chromatography tandem mass spectrometer
m/z	Mass to charge ratio
MS	Mass spectrometer
MRM	Multi reaction monitoring
N.A.	Not available
NMR	Nuclear magnetic resonance
OPLS-DA	Orthogonal projections to latent structures discriminant analysis
PC	Principal component
Par	Pareto scaling
PCA	Principal component analysis
PDO	Protected designated or origin
PGI	Protected geographical indication
PLS	Projections to latent structures
PMP	1-phenyl-3-methyl-5-pyrazolone
PLS-DA	Projections to latent structures discriminant analysis
Q^2	Goodness of prediction
R^2	Goodness of fit
Rf	Radio frequency
RSD	Relative standard deviation

RMSEE   Root mean square error of estimation

RMSEP   Root mean square error of prediction

SDS-PAGE   Sodium dodecyl sulphate polyacrylamide gel electrophoresis

S.D.   Standard deviation

TMS   Trimethylsilane

TOF   Time of flight

TSG   Traditional specialities guaranteed

Uv   Unit variance

VIP   Variable importance for the projection

WHO   World Health Organization

2-DE   Two-dimensional electrophoresis

LIST OF SYMBOLS

P	Loading matrix between X and T
P_o	Loading matrix between X and T_o
P_p^T	Loading matrix between X and T_p
C	Loading matrix between Y and T
C_p	Loading matrix between Y and T_p
m	m^{th} sample in prediction set
X	Matrix of N samples and K variables
Y	Matrix of N samples and L variables
A	Number of principal components
N	Number of samples in X or training set
M	Number of samples in prediction set
K	Number of variablesin X or training set
YPredPS	Predicted Y values from the prediction set
YPred	Predicted Y values from the training set
PRESS	Predictive residual sum of squares
T_p	Predictive score matrix of X
E	Residual variance matric of X
F	Residual variance matric of Y
T	Score matrix of X
t	Score on the score plot
SS	Sum of square
$\tilde{x}_{kn}$	Scaled X variable
T_o	Y orthogonal score matrix

YVar	Y values from the training set
YVarPS	Y values from the prediction set
x_{kn}	X variable
$\bar{x}_k$	X variable average
s_k	X variable standard deviatio

CHAPTER 1. INTRODUCTION AND LITERATURE REVIEW

The natural food item known as edible bird's nest (EBN) is defined as EBN (known as yàn wō in Chinese) is a food ingredient secreted from the two sublingual glands of an *Aerodramus* genus, or more commonly known as swiftlet. It is recognized as a delicacy, medicine and an important agricultural product. Despite its widespread consumption and economic importance, information on safeguarding quality and safety of EBN is far from being adequate. This opens up to an immense opportunity for fraud of the food item. To address this problem, authentication of EBN is required.

Food authentication is the act of verifying a food item to see that it complies with its labelled description to ensure the quality and safety of the food item.[1] To do so, an approach –identification, classificationand discrimination of the food item and its spiked form, was devised and adopted in this thesis. This would provide a systematic way of tackling the problem of fraud and also give way to logical explanations on the inherent difference in EBN. The approach would require the support from various analytical methods to ensure the success of authentication. For this reason, it is vital to develop accurate and precise analytical methods capable of providing reliable data prior to study of the food item. The use of analytical methods would generate a large data set, making reasonable deductions difficult. As such, multivariate data analysis would be employed to reduce the complexity and facilitate the interpretation of obtained results.[2] A graphical display of the general workflow of this thesis is illustrated in Figure 1.

Hopefully, the implementation of such an authentication approach would eventually be able to safeguard the quality and safety of EBN and provide greater insights into food items labelled as EBN.

```
┌─────────────────────────┐
│   Determine the type of │
│      authentication     │
└─────────────────────────┘
            ⇩
┌─────────────────────────┐
│ Devise the analytical method │
└─────────────────────────┘
            ⇩
┌─────────────────────────┐
│   Analysis of samples with │
│      analytical method    │
└─────────────────────────┘
            ⇩
┌─────────────────────────┐
│  Tabulation of data obtained │
│    from analytical method   │
└─────────────────────────┘
            ⇩
┌─────────────────────────┐
│     Process the data with │
│   multivariate data analysis │
└─────────────────────────┘
```
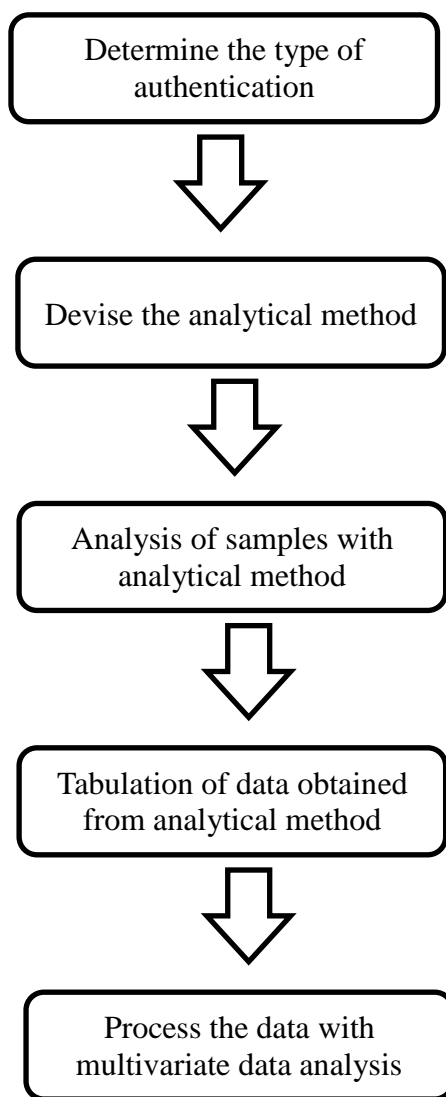
Figure 1 General workflow for the authentication of EBN in this thesis.

1.1.Overview on food authentication

Food authentication is the act of combating against an age-long problem - food fraud. One of the earliest records on food authentication was published in the 19th century by Frederick Carl Accum to expose the act of deliberate addition of water into wine, beer, brandy and custards.[3] Besides the need to protect the economic interest of food suppliers and consumers, food authentication is deemed to be an important issue because of the serious impact brought about by food fraud in today's increasingly globalised world. A recent example is the melamine-tainted Chinese infant milk powder incident in 2008, which clearly underlined the severity of the threat of food fraud on consumers' health. In this unfortunate incident, the number of affected infants is estimated to be 300,000, and six of them died due to kidney damage.[4] Besides the issue on public health, there is an increasing pressure coming from consumers on the food suppliers to reveal more information on the food items so as to gain control of their diet.

To address the concerns of consumers and food suppliers, international organizations like Food and Agricultural Organization (FAO) and World Health Organization (WHO) have been set up to ensure the authenticity of food items.[5, 6]These organizations focus on devising scientific methods and safety guidelines to serve as a reference for countries to combat against food fraud. In Singapore, two regulatory bodies– Agri-Food and Veterinary Authority of Singapore (AVA) and Health Science Authority of Singapore (HSA),have been tasked to be in charge of the authenticity of food. To

perform the task, legalisation like the Food Act has been established in Singapore with reference to guidelines stated in Codex Alimentarius.[7] This official statement makes it necessary that all food items adhere to the safety regulations and be labelled according to their contents. However, merely enforcing food labelling may not be sufficient as it is prone to manipulation. Moreover, tonnes of food items are imported into and exported out of Singapore yearly, adding to the difficulty of authentication. Besides Singapore, many countries are facing a similar problem. Therefore, apart from enforcing accurate food labelling, it is also vital to develop reliable analytical methodologies to act as an independent and objective tool for the authentication of food items.

Food authentication is a broad and complex topic, due to the numerous factors influencing the food items from farm to fork. Thus, in this thesis, authentication is divided into three categories, each representing a unique aspect on the authentication of the core food item– EBN. The categories are as follows:

1) Identification of food items labelled as EBN based on species origin.
2) Classification of EBN according to its natural coloration, country of origin, production site.
3) Discrimination between genuine EBN and EBN samples spiked with adulterants.

The categories are formulated as mentioned to account for factors which may potentially affect the chemical composition of EBN – beginning with the swiftlets that secrete the EBN, followed by the environmental factors during the formation of EBN, and lastly the processes which EBN undergoes from harvest till its consumption. In addition, findings in these categories would be able to answer consumers' questions on the authenticity of EBN.

It is important to have a clear understanding of the categories before attempting to authenticate EBN. Hence, a detailed discussion on the individual categories would be done in the following sections.

1.1.1. Identification of food items

Identification of a food item is the process of verifying that the species origin of a food item complies with its food label. This process relies on the fact that food items originating from different species have unique chemical characteristics and dissimilarities in their physical and organoleptic properties, which could be detected by human inspection generally. Despite the high risks of being exposed, many unscrupulous food suppliers continue to mislabel food items, blinded by lucrative incentives.

A common example of a mislabelled food item is beef. Beef is one of the main sources of protein for human. Thus, it is not surprising that the increase in global population would drive the rise in the demand of beef. Beef often commands a price premium over the other kinds of meat reared due to its

longer rear time and higher popularity among people. In addition, the execution of the authentication of beef relies heavily on paper traceability. All these factors contributed to one of the biggest mislabelling case of horse meat as beef in Europe in 2013.[8] The "horse meat" scandal has offended many English and Jewish as horse meat is regarded as a taboo food.[9] Moreover, this has led to the entry of the veterinary drug phenylbultazone into human food chain. At the moment, there have been no reports on the health problems with regard to the incident but concerns have been raised on the impacts on the consumption of this mislabelled food item. Besides beef, food items such as butter, seafood and wheat has been reported to be mislabelled with other species of food items.[10-12] Although no such fraud cases have been reported on EBN, it is necessary for the regulatory bodies to take the precautionary measures as the promise of economic gains from mislabelling EBN maybe too alluring for food suppliers in the future.

1.1.2. Classification of food items

Apart from species origin, food items could also be differentiated according to intrinsic properties, such as their geographical origin, and other man-induced properties. These differences in food items are often not observable via human inspection. As such there is a need to devise another approach, known as classification, to address this problem.

The motivation behind the classification of food items is mainly driven by the needs of consumers. With improvement in personal financial ability and being

better informed about quality, consumers are no longer satisfied with food items that are safe for consumption. They are also concerned about the quality of the food items such as the organoleptic properties, geographical origin and production methods. Some consumers are willing to pay a premium price for food items that possess the quality they desire. This creates a favourable financial incentive for unscrupulous food suppliers to exploit upon. In view of the high possibility of fraud cases on food, several countries have created policies to protect their citizens. One of the most well-known policies is found in the European Union (EU) and they are known as protected designation of origin (PDO), protected geographical indication (PGI) and traditional specialities guaranteed (TSG).[13] The implementation of these policies hopes to guarantee that the food items are produced from certain geographical origins and/or by certain processing methods. Classification of food items has also been utilized as a branding method for food items. For instance, India has established a scheme known as Grapenet to ensure the quality and authenticity of grapes produced by Indian farmers and hopefully to be recognized by consumers as a reliable source of the food item.[14]

Recognizing the importance of classification, scientists have also attempted to develop new analytical methods to classify food items. The food items that have been targeted for analysis generally command a premium price or are widely consumed by people. Through the use of analytical methods, scientists were able to prove that it is feasible to differentiate between organic and non-organic food specifically, determining the type of processing methods for tealeaves and the geographical location of honey and ginseng.[15-18]

In the context of EBN, classification of the food item has been used as a basis for the variation in the prices of EBN. Consumers generally accept that EBN originating from caves in the wild commands a premium price than those produced on farms as they believe that the cave EBN is more nutritious than the farm EBN. However, currently there are no scientific evidences to support this belief. Thus, it is vital to make use of analytical methods to classify EBN so as to prevent the possibility of food suppliers from exploiting the lack of measures on EBN classification, thereby gaining unethical economic benefits.

1.1.3. Discrimination of genuine food items from its spiked form

Adulteration is known as the act of adding a substance that is not present in the food naturally. This act is performed by food suppliers to improve the physical and/or organoleptic properties of the food items so that the adulterated food item could be sold as the genuine one. These adulterants come in two forms – chemical and organic.[19]

With advancement in technology, food suppliers can easily obtain information on food authenticity tests conducted by regulatory bodies from the internet. Thus, in order to evade detection by authorities, food suppliers would select adulterants that have not been included in routine tests done by regulatory bodies. An adulterant is chosen such that a minute addition of it would not change the quality of food item to that of a genuine one. As a result of evading

detection, this deceitful act is usually uncovered only after the food item has been in the market for a period of time.

An example of such fraud is the 2008 Chinese infant milk scandal mentioned earlier. Authenticity test of the infant milk powder in China had relied on kjeldahl method to determine the protein level of the food item. This prompted unscrupulous milk suppliers to adulterate milk with a nitrogen rich compound known as melamine. Such an adulterant enables the reduction of the amount of milk in the food item without changing the overall protein content, making it the ideal adulterant. This unexpected adulterant was revealed after many infants were observed to have fallen sick after the consumption of the adulterated milk powder for some time. Another incident which took placed recently in 2013 was the adulteration of olive oil using grape seed oil and copper chlorophyll in Taiwan. As this complex mixture of adulterated olive oil did not lead to any significant health problems, it was not uncovered until one of the employees from the manufacturing company blew the whistle. What really shocked the public was that this adulteration had been going on for several years.[20]

Among the different food fraud cases, adulteration is the most common one for EBN. At the moment, it has been reported that a minute amount of adulterants such as agar agar, white jelly fungus and isinglass have been spiked into EBN.[21] Chemicals such as bleach and nitrate compounds have also been identified to be spiked into EBN to enhance its appearance.[22, 23]There is also a high possibility that there are many other adulterants that have not been

detected in EBN. In view of the complexity of the problem of food adulteration, there is a need to develop new and more sensitive analytical methods to safeguard the health and interest of both the consumers and suppliers.

1.2. Background information on EBN

EBN is the focus of study in this thesis. Thus, a literature review on EBN is performed so as to gain a better understanding on the food item. In the following sections, the origin, production sites, cleaning methods, economic importance and medicinal effects of EBN will be discussed in detail due to their importance and relevance to the thesis.

1.2.1. Origin of EBN

Swiftlets are small sized birds, weighing between 6 to 40 g are found mainly in the South East Asian countries or regions within this geographical range–from Andaman and Nicobar Island in Indian Ocean to the coastal regions of Malaysia, Thailand, Vietnam, Palawan Islands in the Philippines and the South-Eastern part of China (Figure 2). [24-27]

Figure 2 Map of EBN locations. Shaded in green are the areas where EBN has been sighted.

Swiftlets are observed to have a grey brown appearance and they possess a pair of short legs, a short bill and a wide gape.[28] They are creatures that exhibit colonial behaviour, thus it is not surprising that they are observed to be in large numbers in their nesting or hunting sites. Swiftlets are known to be capable of producing echolocating calls which aid them in hunting insects and navigating in total darkness.[29] One of the most intriguing features of swiftlets is that they secret a viscous mucus and use its solidified form as a construction material for its nest to house their eggs and hatchlings.[30] Furthermore, swiftlets are observed to be strongly affiliated to their nesting site, meaning that they would build their nest in the same site even if their nests are removed, either naturally or by man. The *Aerodramus* genus is known to be made up of four different types of swiftlets - *A. fuciphagus*, *A. maximus, A. maximus* and *A. unicolor*.[31] However, only EBNs produced from *A. fuciphagus* are selected for human consumption as their nests are mainly made up from their salivary

secretions and have small amounts of impurities like the swiftlet's feathers and its droppings.[32]

AN EBN that has just been harvested is observed to be in a half bowl shape. Its curvature is made up of fine strands of solidified swiftlet's saliva and numerous feathers while the ends are formed from thick and compact solidified secretions (Figure 3.). Such a structure enables the nest to withstand the weight of the swiftlets' eggs and hatchlings while staying firmly attached to the walls throughout the course of breeding.[33] Upon measuring and confirming with the EBN suppliers regarding EBNs general dimensions, the unprocessed EBN is revealed to possess an approximate length of $5.0 - 10.0$ cm, breadth of $3.5 - 6.0$ cm. Thickness at the curvature of the EBN stands at $0.5 - 1.0$ cm, thickness at the ends is about $1.0 - 2.5$ cm and weighting about $4.0 - 8.0$ g.
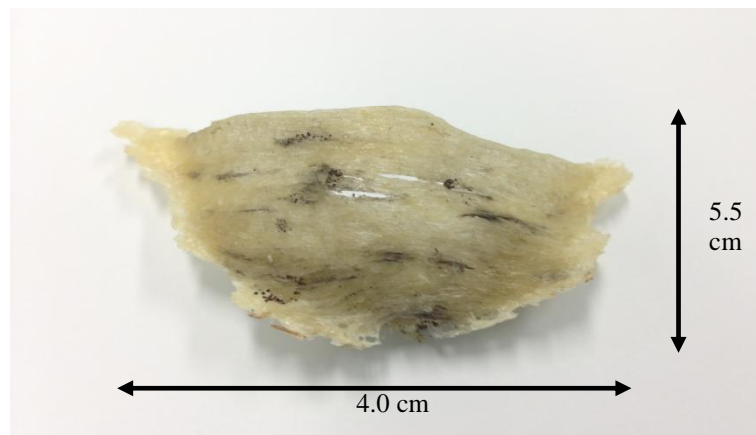
Figure 3 A picture of an unprocessed EBN. The dimensions are stated at the side of the EBN.

Although EBN is produced by the *Aerodramus* genus, this food item is observed to come in three kinds of coloration – white, orange and red (Figure 4). This difference in the coloration has been an age-long mystery. One of the most popular hypotheses is that the red coloration is induced by the blood of swiftlets which are exhausted from the construction of their nests. Other speculations include the diet of the swiftlets and the oxidation of iron from the drippings of the limestone caves in which the swiftlets residue. Recently, a scientific explanation was proposed for the differences in the coloration. According to But et al., a nitrate gas originating from the swiftlets' droppings induced the color to change from white to red in EBN.[34]
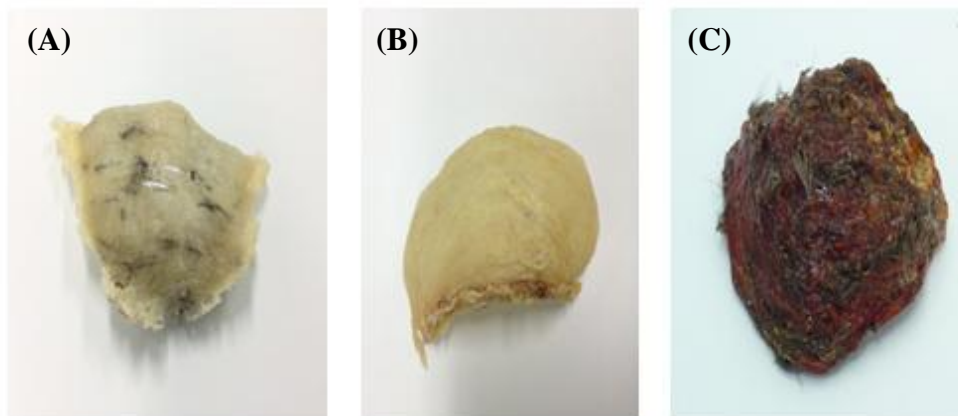
Figure 4 EBNs of different colorations – (A) white (B) orange and (C) red.

1.2.2. Production sites of EBN

The production site of EBN must a reliable shelter to protect the swiftlets from adverse weather conditions and their predators. It should also preferably be close to the swiftlet's food and water sources. Based on these criteria, swiftlets are often found to construct their nests in caves located at/near the coastal

regions or tropical rainforests. In addition, EBN is mostly found on smooth surfaces of inward inclining cave walls, located at least 2.5 m above the ground and a minimal distance of 1 m away from the cave entrance. This type of environment is often inaccessible to humans, thus bamboo poles, climbing ropes and harnesses are some of the common equipments used to harvest EBN. The most famous nesting sites of swiftlets are known to be the Niah cave and the Gomantong cave, located respectively in the states of Sarawak and Sabah in Malaysia (Figure 2).[35] These caves come in various shapes and sizes but they share a similarity of being dimly lit or in total darkness. This poor lighting condition provides an environment exclusive to a few animals such as swiftlets and bats acting as a natural form of protection against the swiftlet's predators. Caves of the swiftlets are also well known for a strong ammonia smell, generated from the massive amount of swiftlet's droppings accumulated on the ground. Although the droppings are a major deterrent to most creatures, especially humans, they are the main source of nutrient for insects.

Besides the natural cave environment, EBN has been sighted in areas populated with humans. One of the contributing factors is that rapid urbanisation has reduced the choice of available nesting sites for these swiftlets, as such these creatures had to source for new nesting sites within the urban environment. Besides that, the increasing availability of deserted buildings during the Asian economic financial crisis (1997 – 1998) has also attracted the swiftlets to construct their nests in empty buildings that have been void of human activities for a long period of time.[36] Furthermore, these buildings are similar to the environment of the swiftlets' natural habitat being

dark, damp and cooling. There are also numerous hard concrete surfaces that serve as a platform for the construction of the swiftlets' nests.

The discovery of EBN in buildings prompted the locals to build more vacant buildings to increase these nest construction activities. Over the years, this organized and systematic approach to harvesting EBN has grown into a sizeable industry known as swiftlet farming. However, this has also brought about several problems for residents staying near to the swiftlet farms. Besides the awful smell coming from the swiftlets' droppings, the loud chirping sounds produced by the swiftlets are deemed to be an intolerable disturbance to the residents.[37] Thus, the EBN farms had to be shifted to the suburbs and their hygiene conditions are closely monitored by the local authorities to ensure the safety and quality of the harvested food item, at the same time making sure that they do not become a hazard to the surrounding environment.[38] A picture of a typical EBN farm is displayed in Figure 5.

Figure 5 A man-made building used as EBN farm in Johor Bahru, Malaysia.

1.2.3. Methods utilized to process the EBN

Harvested EBN are not consumed directly as they contain alot of impurities such as swiftlets' droppings, twigs and/or dirt. As such, they would have to undergo a series of processes to separate the impurities from EBN. These processes generally include cleaning (removing the swiftlet's feathers and dirt from the EBN), moulding into a half bowl shape and drying prior to packaging as the sale food item.

In the cleaning stage, the EBN is first soaked in clean water until it softens and the tightly bound laminae partially loosens. Large impurities (such as dirt and feathers) floating on the water are then removed from the water. Once the EBN is softened – usually after 1 to 2 hours of soaking in water, the leftovers are scooped out and small feathers and dirt are removed manually using a pair of tweezers. This is the most tedious and time consuming step in the cleaning of EBN. A magnifying glass would be used at times to pick out the small impurities or to visually inspect the cleanliness of the food item. Often in the workstation, another bowl of clean water is placed at the side so as to wash the tweezer during the cleaning process.

Next, the EBN would be shaped into a half bowl structure with the aid of a plastic mould. To perform this task, the EBN is placed into the mould one strand at a time. After shaping, the food item is dried to remove excess moisture. Drying is generally performed with an oven and/or fan to ensure efficiency of the process. Finally, the cleaned EBN are packaged and sent to

the market for sale. A graphical display of the processing method used to clean the EBN is illustrated in Figure 6.

Figure 6 Graphical display of the processing method for EBN. (A) soaking of EBN in water (B) manual removal of the impurities from EBN (C) the moulds which shape the EBN into a half bowl shape. (D) the oven for the drying of EBN.

Currently, there are no harmonized cleaning methods for EBN, thus it is not surprising that there would be slight modifications to the methods used by different suppliers. The methods stated in this section are an outline of the cleaning methods based on personal visits to an EBN cleaning factory. A similar method would be applied to the cleaning of unprocessed EBN used in this research in the later chapters of the thesis.

1.2.4.Economic importance of EBN

EBN is an important export food item to many countries in South East Asia. It has been estimated that the trade of EBN earns up to several hundred million dollars (US dollars) in foreign exchange, making its production an important driving force to the economy of major suppliers like Indonesia, Malaysia and Thailand. The trade value of EBN has increased dramatically from approximately USD$170 million in 1989 to USD$380 million in 2004 and it is expected to grow further with the rising demands from East Asia countries.[39] Thus, in order to benefit from such trade growth, the Malaysian government has recently announced in the Malaysia Economic Transformation Programme that the EBN industry will receive prioritized public investment and policy support for its development.[40]

Based on a personal communication with the retailers in Singapore, it was found that the retail value of EBN can go up to as high as USD$15 per gram of EBN. Thus, it is not unexpected to get reports on the adulteration of EBN with cheaper food items as the economic incentive is attractive.

1.2.5.Health effects of the consumption of EBN

Consumption of the food item known as EBN could be traced back to as early as the Tang dynasty in AD 618.[41] This salivary secretion of the swiftlets has become popular because of the therapeutic effects claimed to be brought about by its consumption. One of the earliest records which describe the medicinal

properties of EBN was the Ben Cao Gang Mu written by Li Shizhen. To date, the consumption of EBN is believed by the Chinese community to be able to bring about effects: dissolve phlegm, relieve gastric troubles, aid renal functions, raise libido, enhance complexion, relieve asthma, suppress cough, cure tuberculosis, cure hematemesis, nourish the lungs, speed up recovery after surgery, strengthen the immune system, improve the growth of children, increase energy and metabolism, and improve concentration.[42]

In the last century, many scientists have devoted their effort to studying the health benefits brought about by the consumption of EBN. Kong et al. discovered epidermal growth factor (EGF) in EBN and demonstrated that EGF is capable of stimulating the growth of cells.[43] EGF is known to play an important role in the proliferation, differentiation and survival of the cells. With the discovery of EGF in EBN, it provided the scientific evidence on the beautifying quality of EBN. EBN is also found to be able to inhibit influenza virus due to the presence of N-acetylneuraminic acid residue in the glycans of EBN.[44] As such, consumption of EBN has been suggested to be useful in the prevention of influenza viral infection. Recently, it is shown that the EBN is also capable of reducing the production of tumour necrosis factor-alpha and nitric oxide, improving bone strength, decreasing the rate of osteoarthritis, lowering food-borne pathogens activities, preventing age related degenerative disease as well as inducing the proliferation of stem cells and corneal keratocytes.[45-51] Although the biological mechanisms have not been elucidated for many of these health effects, these research studies provided strong

evidences that the consumption of EBN would be able to provide a wide range of health benefits.

Despite the numerous health benefits, EBN was reported to be the most common cause for food induced anaphylaxis in the National University Hospital in Singapore. Studies indicated that a 66kDa protein in EBN, similar to egg white protein, was the main reason leading to the allergy.[52] Besides anaphylaxis, the consumption of EBN has also been reported to have led to eczema exacerbation in children suffering from atopic eczema.[53] Recently, a carcinogenic compound known as semicarbazide was detected in selected EBN products, which trigger the alarm on the safety on the consumption of EBN and its products in China.[22] Another recent health scare incident involved the detection of nitrite in the EBN sold in China. Potassium nitrile and sodium nitrile are a common preservative used in food but the excess consumption of these chemical may result in cancer.[54] This led to a ban in the import of EBNs from South East Asia countries from 2011 to 2013 to protect the consumers in China.[55] In view of the potential health benefits and hazards from the consumption of EBN, it is important to develop analytical methods for the assessment of this food item.

1.3.Analytical techniques applied for food authentication

Food is a complex matrix made up of compounds which exist in different forms and with chemical nature.[56] This requires sophisticated and high-end analytical instrument to detect and quantify compounds which have authentication properties in the food item. The international research group, TRACE has been established by EU to ensure the authenticity of food items from farm to fork. TRACE has demonstrated to be able to apply analytical techniques to authenticate food items such as olive oil, honey and lamb meat.[57-59] Besides TRACE, other international research programs such as Metabolomics for Plants, Health and OutReach(META-PHOR) have also been developing new analytical methods for food authentication.[60]This provides additional confidence in applying analytical techniques to authenticate EBN.

Analytical techniques that have been used in food authentication can be broadly divided into 3 categories – genetics-based technique, spectroscopic technique and chromatographic technique. A discussion on the advantages and disadvantages of their application on food authentication would be done in the later sections, with an emphasis on the chromatographic technique as it is the primary focus of this thesis

1.3.1. Deoxyribonucleic acid based techniques

Deoxyribonucleic acid (DNA) is a molecule containing the genetic codes of an organism. This genetic code is made up of a sequence of nucleotides which is

exclusive to the organism. This makes it a feasible target molecule to assess food authenticity.

A single DNA molecule codes for a large number of genes which makes it impractical to analyse the entire molecule. Nonetheless, it is feasible to utilize DNA by analyzing specific genes that are unique to the sample. This could be achieved with a technique known as polymerase chain reaction (PCR)which amplifies and replicates the genes.[61] Besides the high selectivity of PCR, this technique is chosen due to its high throughput and simplicity in analysis. As such, food forensic scientists have applied DNA based techniques to authenticate food items such as seafood, beef and peanut.[62-64]

However, this technique is restricted by temperature. DNA molecules degrade upon heating to a temperature of $100\,^{o}C$ or higher, resulting in an undetectable form.[65] Thus, this technique is limited to food items that have not undergone any form of processing. DNA based techniques are able to detect minute amounts of DNA in the sample due to their amplification capability, making it an excellent analytical tool in the identification of species origin. On the other hand, the amplification capability leads to the inability of the technique to resolve between genuine samples and spiked samples. This implies that the DNA based technique is applicable to a limited number of food authenticity cases.

1.3.2. Spectroscopic techniques

Spectroscopic techniques are based on the interaction between the molecules with electromagnetic radiation. This is a generic term for methods which employ electromagnetic radiation- ultraviolet spectroscopy and infrared spectroscopy. With data obtained from spectroscopic techniques, information such as the molecular bonds, type of compound and/or molecular structure could be deduced. Besides the wealth of information obtained, spectroscopic technique is known to offer an analytical process which requires minimal sample preparation, short analysis time and is non-destructive to the sample.[66] Thus, it is not surprising that this analytical technique has been commonly used in the authenticity of food items such as tea leaves, fruit juices and salmon.[67-69]

The main disadvantage of the spectroscopic technique lies in its selectivity of compounds as spectroscopic technique requires the targeted compound to be of a high purity percentage in order to identify it. This is usually not feasible as the matrix of food item is made up several types of chemical compounds. As such, identification of compounds by most spectroscopic techniques may be hindered, with the exception of NMR. This is because NMR is able to detect the small modifications to the external magnetic field caused by the surrounding molecules of a compound. However, the cost of utilizing such an analytical technique is high, thus limiting its popularity in food authentication.

1.3.3. Chromatographic technique

Chromatographic techniques involve the separation of compounds in a matrix. The purpose is to perform qualitative and/or quantitative analysis of individual compounds in the matrix. There are two different types of chromatographic techniques being dependent on the state of the compounds introduced into the instrument – gas chromatography (GC) or liquid chromatography (LC).

GC is a chromatographic technique used to analyze compounds which are volatile and thermally stable.[70] However, a sample does not need to be in a gaseous state. This is because the sample could be first vaporized under high temperature prior introduction into the GC capillary column by the carrier gas. The carrier gas is usually an inert gas such as helium or nitrogen. Compounds in the sample are separated in the column based on the difference in the partition equilibrium of individual compounds with the stationary phase and the carrier gas. Since the separation involves the compound, stationary phase and carrier gas, factors such as the compounds' chemical nature, nature of stationary phase and speed of carrier gas would directly influence the retention time of the compound.

On the other hand, LC is a technique which is suitable for compounds which are non-volatile and/or degrade when they are heated at high temperature.[70] A sample for LC analysis will be in its liquid state prior to analysis thus it could be mixed with the mobile phase directly prior entry into the column. Separation of the compounds takes place due to the difference in the affinity

of individual compounds with the stationary phase of the column and mobile phase. Although GC and LC function differently, both share a similar configuration starting with the chromatographic system, followed by the detector, and then the computer. The function of the detector is to translate the compounds into electrical signals that could be expressed as a graphical representation of time against signal. Furthermore, the detector could input another dimension into chromatogram, thus providing additional information on identifying the peak.

Mass spectrometer (MS) is a detector which could be connected to both GC and LC − the hyphenated technique is known as GC-MS and LC-MS respectively. Its frequent usage is due to its capability of ionizing the compounds and subsequently detecting the masses of the ionized compounds. Regardless of the chromatographic platform the MS is connected to, the configuration of the detector consists of three parts − ion source, mass analyzer and mass detector (Figure 7).

| GC or LC | ⇨ | Ion Source | ⇨ | Analyzer | ⇨ | Mass Dectector |
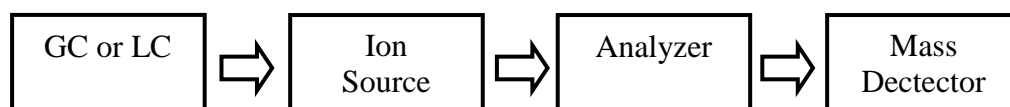
Figure 7 General configuration of the mass spectrometer for GC and LC.

Different types of ionization methods are employed in GC and LC. Electron impact (EI) is one of the most frequently used methods in GC to ionize gaseous compounds. During ionization, a voltage of 70 eV is supplied to the tungsten filament so that electrons could be produced. These high energy

electrons would collide with the gaseous compounds and cause the formation of fragment ions.[71] As such, EI is often termed as 'hard' ionization as no molecular ions are formed during the process. The fragment ion spectra generated could be compared with commercial libraries or standards to verify the identity of the compound. In contrast, electrospray ionization (ESI) is the more common ionization method for LC. The solution entering ESI is dispersed as an aerosol, then being stripped off its solvent with the use of nebulizing gases such as nitrogen to form a small droplet. The desolvation process would carry on until the droplet is reduced to a size such that the repelling coulombic force is larger than the surface tension holding the droplet together, which in turn triggers a the coulombic explosion. This would either yield smaller sized droplets or ions. In the former case, the droplet would have to undergo further coulombic explosions until the ions are formed.[72]

In order to identify the ions, a mass analyzer is utilized so that ions can be separated based on their mass to charge ratio (m/z).Although there are a few types of mass analyzers, all of them have to be operated in vacuum. In this thesis, the mass analyzers that are employed are single quadrupole, triple quadrupole, time of flight(TOF) and ion trap(IT).[70]

As its name suggests, the single quadrupole is made up of four metal rods, aligned parallel to each other. In this mass analyzer, a specific direct current (dc) voltage and a radio frequency (rf) voltage are applied to the metal rods so as to guide the selected ions to the detector while the rest of the ions would exit the single quadrupole. The triple quadrupole uses the same principle as

the single quadrupole, but the configuration of the mass analyzer starts with a single quadrupole followed by a mass collision cell and then another single quadrupole. This permits the selection of an ion, followed by further fragmentation of the selected ion, and finally selecting the daughter ions formed in the collision cell. Similarly, the IT applies a dc voltage and rf voltage to the ring and end electrode to create an oscillating electric field to separate the ions However, the IT would trap all the ions in the center of the mass analyzer and then forces the ions that are not of interest out of the ion trap. On the other hand, TOF accelerates all the ions with an electric field so that the ions would have the same kinetic energy. These ions would travel along a vacuum tube and they would be differentiated by the time in which they arrive at the detector where ions with lower m/z are considered faster and vice versa.

It is also critical to have a mass detector to convert the ions into electrical signals. A mass detector known as electron multiplier is used for this purpose. The dynodes in the mass detector would generate electrons upon ion impingement. These ions are then multiplied as they pass through the dynodes, thereby amplifying the electrical signal which is sent to the computer.The use of GC and LC coupled with mass spectrometers, GC/MS and LC/MS respectively, has been widely applied in all issues involving authenticity of food items as shown in Table 1.

Table 1 Examples of food authentication with the use of GC/MS and LC/MS.

Analytical Platform	Food	Application
GC/MS	Fish	Authenticate different Mediterranean fish species using fatty acid profiles.[73]
	Saffron	Determination of the geographical origin of saffron based on their volatile compounds.[74]
	Liquor	Distinguish brandy and whisky through organic acid profiling.[75]
LC/MS	Fruit juice	Differentiation of orange, apple and grapefruit juice based on their metabolite profile.[76]
	Beer	Differentiation of trappist beer and other brand of beers using metabolite fingerprinting.[77]
	Olive oil	Determination of metallo-chlorophylladulterant in olive oil.[78]

1.4. Multivariate data analysis

With the use of modern analytical instruments, it is not unusual to observe the analysis of one sample would generate thousands of signals within a short period of time. The advancements in analytical instruments, brought about by factors such as improvements in technical specifications of the chromatography instruments and the increase in sensitivity of the detectors, greatly improved the throughput of the analysis; at the same time this makes the interpretation of the obtained data more challenging. Besides that, detected variables are likely to be correlated to each other. Thus, it is no longer valid to rely solely on classical statistical methods such as Student's *t*-test and analysis of variance (ANOVA) to make comparisons between the samples because

these statistical methods assume that the variables are independent. In view of the complexity and size of the data, multivariate data analysis is employed as the main statistical tool to analyse and interpret the data in this thesis.[79]

In multivariate data analysis, samples' datasets are initially viewed as points in a multidimensional space - dimensions of the space are in accordance with the number of variables (n). In order to interpret the data with ease, the dimensionality is reduced. To perform the task, a hyper plane with a few axes, known as principal component (PC), is constructed. These PCs are new variables which are formed on the basis of the linear combinations of the variables in the data set and they determine the number of dimensions of the hyper plane. Subsequently, points in the multidimensional space are projected onto the established hyper plane.[80]

Prior to using the multivariate analysis, a workflow is established so that the analysis could be conducted in a systematic way. The general workflow for multivariate analysis is shown in Figure 8. Firstly, a group of samples known as the model set would be utilized to establish the multivariate model. This group of samples is selected to account for any possible analytical variations. Data from the model set is then subjected to data pre-processing such as normalization and scaling prior to multivariate analysis to ensure the accuracy of a multivariate score plot as a representation of the samples. Next, the type of multivariate model is selected. There are generally two types of multivariate model – unsupervised and supervised model, which would be discussed in detail in the later sections. Once the models are established, an

independent set of samples known as prediction set is utilized to validate the model. An excellent prediction result would prove that the model is accurate and precise.
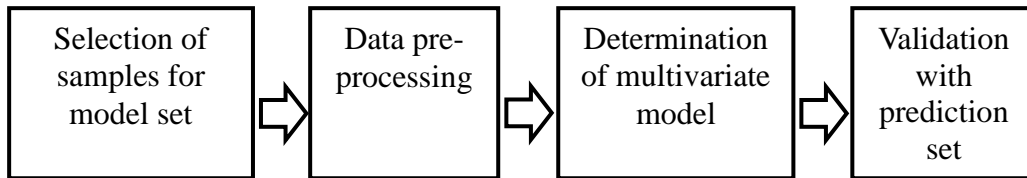
Figure 8 General workflow for a multivariate data analysis.

1.4.1. Scaling

Scaling is a data transformation process which involves changing the weights of the variables.[81] This is necessary as the data obtained may differ by orders of several magnitudes. In such cases, the variables with high signals would dominate the established model even thou they might not be responsible for the variation between the samples. This would in turn lead to problems such as misclassification of samples and/or misinterpretation of the data. However, with an appropriate scaling method, weights could be placed on more informative variables.[82]

Prior to the scaling of the variables, mean-centering would be applied to all the variables. In mean-centering, the variable (x) is subtracted from the average of the variable $(\bar{x})$. This ensures that the scaling is conducted only on the fluctuating portion of the data.

There are three main types of scaling methods – unit variance (uv) scaling, pareto (par) scaling and no scaling. Generally, scaling divides each variable by a scaling factor, and thereby changing the degree of fluctuations in different variables when they are expressed as a value relative to the scaling factor . Uv involves dividing the mean-centered variable by the variable's standard deviation(s). As a result, the influence of the variables with large fluctuations is reduced while those with small fluctuations will be amplified. Such approach is useful when variables with small fluctuations are more informative since the model is now no longer dominated by variables with large fluctuations. On the other hand, in no scaling, the variables only undergo mean centering. As such, the model would be greatly influenced by variables with large fluctuations. Par scaling is a compromise between the two aforementioned types of scaling. It involves dividing the mean-centered variable by the square root of the variable's standard variation. Upon scaling, the variables are adjusted to the same effect as uv scaling but with a lower impact due to the difference in the scaling factor. The formulae of the uv, par and no scaling are displayed below in equation (1), (2) and (3) respectively.

$$\tilde{x} = \frac{x - \bar{x}}{s}\,(1)$$

$$\tilde{x} = \frac{x - \bar{x}}{\sqrt{s}}\,(2)$$

$$\tilde{x} = x - \bar{x}\,(3)$$

where$\tilde{x}$ is the scaled variable, x is the original variable from the data set,$\bar{x}$ is the average of all the variables and s is the standard deviation of the variables.

1.4.2. Unsupervised model

Unsupervised models are multivariate models that do not have any prior information on the class membership of the samples. It is often constructed before the supervised models to provide an overview of the data trend and outliers.[83] There are several types of unsupervised models in multivariate analysis such as principal component analysis (PCA) and hierarchical clustering, however, PCA would be the focus in this thesis.

After scaling, the data set is expressed as a data matrix X which is made up of N rows and K columns, representing the samples and variables in the original data set respectively. Since PCA is a form of multivariate analysis, it is capable of reducing the dimensions of the data matrix X down to a very small number of PCs- the number of PCs is generally fewer than the number of variables in the data. PCs are orthogonal axes of the PCA score plot which determines the greatest variance in the multidimensional X space. This implies that each successive PC would explain as much as possible the variance unaccounted by the previous PCs. This PC formation process will carry on until all or most of the variances of the data matrix X have been accounted for. In addition, the PCs are deduced using linear least square method and they would have to pass through the center of the multidimensional X space.[84] The X matrix could be expressed as Equation (4):

$$X = TP + E \qquad (4)$$

where T is a score matrix for the variables in the data matrix X and the loading matrix P illustrates how the scores are related to the X variables. E is known as

the residual variance matric of the data matrix X and it explains the variance which is not represented by the PCs.

Graphically, the PCA score plot is illustrated as a plot with the PC axes labelled by the letter t followed by the PC number, eg. the first PC axis is t[1]. To evaluate whether an over fitting model has been constructed, model performance parameters such as goodness of fit (R^2) and goodness of prediction (Q^2) are utilized. R^2 accounts for the amount of variance in the data matrix X explained by the PC. Meanwhile, Q^2 illustrates the predictive capabilities of the model calculated based on Z fold cross-validation of the data matrix, where Z is taken as 7 which is the default value in the statistical software used in this thesis. In the cross validation, data matrix X is divided into 7 portions, with 6 portions being used to establish a score plot while the remaining portion serves as the prediction set.[85] This process is repeated until each portion has served as a prediction set for once. Both model performance parameters, R^2 and Q^2 vary between 0 and 1, where 1 indicates a perfect fit for the model. In addition, they are applicable to both the unsupervised and supervised models mentioned in this thesis.The equation to calculate R^2 and Q^2 is presented as Equation (5) and (6):

$$R^2X = 1 - \frac{SS(E)}{SS(X)} \qquad (5)$$

$$Q^2X = 1 - \frac{PRESS(E)}{SS(X)} \qquad (6)$$

where SS denotes the sum of square, and PRESS denotes the predictive residual sum of squares. E and X are the residual matrix of X and the data matrix X respectively.

Besides observing data trends, it is also interesting to understand how the scores (t) are related to the X variables. As such, the loading plot is used to deduce the variables responsible for the observations in the score plot. This is because the loading plot has the same shape as that of the score plot, meaning that the positioning of the scores and loading is correlated. The loadings for each variable are calculated by applying the mathematical function consine to the angle between the axes in multidimensional X space and the PC. Thus, the axis in the loading plot would have a range of 1 to -1, where 1 indicates variable with a strong positive impact while -1 denotes the variable with a strong negative impact.[86] Examples of the applications of PCA and its loading plots are summarized in Table2.

1.4.3. Supervised model

Supervised models are multivariate models which take into account the class membership during the construction of the model. As such the class of each sample has to be identified prior to the construction of the model.[87] In this thesis, the supervised models, which would be discussed in detail, are projection to latent structure (PLS), projection to latent structure discriminant analysis (PLS-DA) and orthogonal projection to latent structure discriminant analysis (OPLS-DA).

PLS is a multivariate technique which attempts to find a quantitative relationship between two data matrixes, X and Y.[88] The data matrix X is made

up of the signals from the variables while data matrix Y contains continuous variables of the sample. Similar to PCA, PLS determines a hyper plane of a lower dimension than the multidimensional X space. However, this hyper plane is constructed in a manner that the projected X in the hyper plane would be a good predictor of Y.[89] As such, the PCs of PLS score plot may not agree with the highest variance of the data matrix X. Since a PLS score plot is constructed based on the data matrix X which is correlated to data matrix Y, PLS can be represented by equation (7) and (8):

$$X = TP^T + E \qquad (7)$$

$$Y = TC^T + F \qquad (8)$$

where T is the score matrix of X, P is the matrix of weights describing the correlation between X variables and T, while C is the matrix of weights describing the correlation between Y variables and T. E and F are the residual variances for the data matrix X and Y respectively.

PLS is used to derive a calibration plot known as PLS regression. In this calibration plot, the y axis is represented by YVar and illustrates the values representing the continuous variables of the sample. On the other hand, the x axis is known as YPred and it shows the predicted values of the continuous variables based on the X variables. The application of PLS regression is often found in cases which require the prediction of continuous variables of the sample from a given set of X variables. A list of the applications of PLS regression is displayed in Table 2.

Besides the model performance parameters R^2 and Q^2, new calibration parameters are required for the assessment of the quality of PLS regression. These calibration parameters are known as root mean square error of estimation (RMSEE) and root mean square error of prediction (RMSEP).[86] RMSEE is the measure of precision of the PLS regression using the model set. On the contrary, RMSEP indicates the precision of the PLS regression against the external prediction set. In both RMSEE and RMSEP, parameters with values closer to zero indicate that the PLS regression has a higher precision. RMSEE and RMSEP are computed with Equation (9) and (10).

$$RMSEE = \sqrt{\frac{\sum_{i=1}^{m}(YVar - YPred)^2}{(N - A - 1)}} \quad (9)$$

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{m}(YVarPS - YPredPS)^2}{M}} \quad (10)$$

where YVar denotes the Y values from the training set, YPred denotes the predicted Y values from the training set, YVarPS denotes the Y values from the prediction set and YPredPS denotes the predicted Y values from the prediction set. N and M represents the number of samples in the training and prediction set respectively while A represents the number of PCs.

PLS could also be applied in discriminant analysis. In this case, it would be known as PLS-DA. PLS and PLS-DA are similar in terms of their calculation except for the data matrix Y. In PLS-DA, the data matrix Y is made up of dummy variables in binary codes – either zero or one. This is because Y matrix is generally made up of discrete variables which require numerical representations for differentiation. As such, during prediction, the unknown

sample will take on a value of one for a Y variable while the remaining Y variables will be given a value of zero in order to be assigned to a particular class. In comparison to PCA, PLS-DA involves the determination of the maximum covariance between the data matrix X and Y while PCA finds the maximum variance only in data matrix X. Thus, the PLS-DA score plot is able to maximise class separation as well as minimising the variation within the class which PCA might not be able to achieve.[90] As a result, PLS-DA is often favoured over PCA in terms of class separation and is frequently seen in the application of food classification (Table 2)

OPLS-DA is a multivariate method based on modifications of PLS. The development of the OPLS-DA involves a two steps process. Firstly, the OPLS-DA uses the information in Y to break down X into two parts, one describing the correlation of X and Y and the other describing the systematic variation in X that is orthogonal to Y. These two parts are commonly known as the predictive variation and orthogonal variation respectively. Next, the orthogonal variation is removed for the data matrix X and the remaining portion is utilized for the construction of the OPLS-DA score plot.[91] OPLS-DA could be expressed with equation (11) and (12).

$$X = T_P P_p^T + T_o P_o^T + E \qquad (11)$$

$$Y = T_P C_p^T + F \qquad (12)$$

where T_p describes the predictive score matrix for X, P_p^T denotes the matrix of weights describing the correlation between X variables and predictive score matrix T_p , T_o describes the Y orthogonal score matrix, P_o denotes the matrix of weights describing the correlation between X variables and Y orthogonal

score matrix T_o, C_p denotes the matrix of weights describing the correlation between Y variables and predictive score matrix T_p. E and F is the residual matrix of the data matrix X and Y respectively.

In comparison to PLS-DA, OPLS-DA requires lesser PCs to relate to the data matrix Y due to the removal of the orthogonal variation. This would lead to the establishment of a simplified model with a smaller prediction error in OPLS-DA.[92] Besides that, OPLS-DA would rotate the PCs so that the majority of the class discrimination would be observed in the first predictive PC. This reduces the complexity of interpreting the score plot and loading plot in OPLS-DA.[93] Adding to OPLS-DA advantages, this multivariate technique could also break down the orthogonal variation independently to obtain information within class variation, thereby permitting the establishment of more insightful information on the samples. Being a promising multivariate technique for sample discrimination, the application of OPLS-DA is noted in several scientific studies and they are listed in Table 2.

Table 2 Examples of food authentication with different types of multivariate data analysis.

Multivariate Data Analysis	Food	Application
PCA	Coffee	Differentiation of instant coffees produced by three different manufactories.[94]
	Hazelnut	Differentiation of raw from roasted hazelnuts.[95]
PLS	Wine	Predicting the age of Madeira wine.[96]
	Chocolate	Predicting the amount of lard in chocolate.[97]
PLS-DA	Spices	Determination of the type of Sudan dyes in spices.[98]
	Ginseng	Classification of ginseng based on the duration of steaming.[99]
OPLS-DA	Honey	Classification of honey according to botanical origins.[100]
	Tomato	Differentiation of Chinese and Italian tomato pastes.[101]

1.5. Scope of thesis

EBN is a food item that has been highly prized due to its perceived health benefits through consumption.[43, 44] Clinical trials on the benefits of EBN consumption have proven that it would promote the proliferation of human stem cells.[50] This resulted in numerous dubious cases of fraud cases of this food item in recent years. To address the complex authenticity issues on EBN, this thesis endeavours to explore feasible analytical methods and multivariate data analysis to authenticate EBN.

Identification of EBN is a crucial issue as it entails the determination of the species origin of a food item, thereby preventing the mislabelling of EBN. Amino acid and monosaccharide analysis were employed as the analytical methods to analyze EBN. With the aid of multivariate data analysis like PCA, the signature for each type of analysis would be established in hope of identifying food items that are labeled as EBN. The signature would also be compared against samples which are suspected to have a similar amino acid and monosaccharide content to EBN to prove that the signature is an exclusive one for EBN.

Classification of EBN is also another approach to authenticating EBN. In this part of the thesis, the possibility of classifying EBN in terms of coloration, country of origin and production site would be explored. Metabolite fingerprinting is utilized as the analytical method to analyze EBN. This analytical method is able to detect most of the compounds in EBN within a

single analysis, thus offering a high possibility of detecting compounds that are able to discriminate EBN from various classes. The data obtained from metabolite fingerprinting would be subjected to multivariate data analysis to establish the difference between the various classes of EBN and also to offer a logical explanation for their differences.

The adulteration of EBN, with a focus on organic adulterants, is the final food authenticity issue that would be looked into in this thesis. In this part of the thesis, discrimination of genuine EBN and EBN spiked with adulterants is sub-divided into qualitative and quantitative discrimination. In qualitative discrimination, the type of data pre-treatment, multivariate data analysis and analytical methods would be explored to determine the ideal combination for the detection of spiked EBN. Further, the correlation between the analytical methods and PLS regression would also be established to determine which analytical method is able to provide the lowest detectable spiking.

CHAPTER 2. IDENTIFICATION OF EDIBLE BIRD'S NEST WITH AMINO ACIDS AND MONOSACCHARIDES

2.1. Introduction

In spite of the long history of consuming EBN dating back to 618 AD, the first scientific study on EBN was conducted only in the early 20th century to determine the composition in the food item.[102] It was revealed in the study that the largest compositional fraction of EBN is protein, followed by carbohydrate. As such, the identification of EBN in the early days made use of classical methods like the Kjeldahl method to measure the protein content.[103] The use of classical methods for identification of food item is often deemed as laborious and time consuming. In addition, the Kjeldahl method suffers a major disadvantage of being easily manipulated by the addition of nitrogen-rich compound as reported in the 2008 Chinese melamine food scandal.[104] Hence, many new analytical methods have been developed, especially in the last decade, to identify EBN. These analytical methods can be divided into 3 categories – molecular biological methods, chemical analytical methods and elemental analytical methods.

Molecular biological based methods such as PCR, electrophoresis and enzyme-linked immunosorbent assay (ELISA) have been reported to be applied to identify EBN. An example is Dong et al. who made use of PCR to ascertain the presence of EBN's DNA so as to identify the food item.[105] Electrophoresis is one of the earliest method used to study the major biological

marcomolecule in EBN – protein. In the study made by Houdret et al., it was demonstrated that electrophoresis was capable of isolating the protein in EBN.[102] Although limited information was collected on the protein of EBN, this study laid the foundation for future work on EBN's protein. Recently, Yang et al. demonstrated the use of sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) to isolate two major protein molecules in EBN(106 KDa and 128 KDa).[106] After that, Yang et al. went on to perform a detailed analysis on the proteomic profile of EBN using two-dimensional electrophoresis (2-DE).[107]Yang et al. also utilized ELISA to detect two unique EBN's proteins so as to identify the food item.[106] Though the use of molecular biology based methods in identifying EBN has been successful to some extents, these methods have a restriction - the food item of interest must have intact biological molecules for detection, which is often not feasible in the case of a heavily processed food item like EBN.

The chemical analytical methods can be further divided into two sub-categories according to the type of analytical techniques spectroscopy and chromatography. In the case of spectroscopic techniques, an analytical instrument known as NMR is often utilized due to its capability of elucidating the structure of novel compounds. With the use of spectroscopic methods, novel sialic acid derived compound and different types of glycans, such as glucosaminoglycans N-linked glycans and O-linked glycans, have been identified in EBN.[108-112] Analytical techniques based on chromatography involve the separation of compounds with the use of a range of analytical platforms such as GC and LC coupled to different types of detectors. The

application of the GC and LC based analytical platforms have been reported in several studies to be able to determine a diverse class of compounds such as amino acids, monosaccharides, fatty acids and hormones in EBN.[113-115]

In elemental analytical methods, atomic absorption spectrometer (AAS) and inductively-coupled plasma mass spectrometry (ICP-MS) have been used to detect and quantify the elements in EBN. Matan et al. demonstrated the potential of using AAS and ICP-MS to profile the elements in EBN so as to identify the food item.[116] While the use of chemical analytical methods and elemental analytical methods is useful in identifying EBN, the major problem to the studies employing these methods lies in the inadequate number of samples to represent the entire spectrum of EBN. In view of the limitations and drawbacks of current studies, it is important to develop a new approach to identifying EBN.

Amino acids and monosaccharides are promising candidates for the identification of EBN as they are the monomeric constituents of proteins and carbohydrates in EBN.[117, 118] EBN is a food item that undergoes a series of man-induced processes prior to its sales in the market as mentioned in the earlier chapter.[119] These processes are likely to trigger the denaturation and break down of the polymers, protein and carbohydrate, which would cause inaccuracy in their quantitation of the polymers. The determination of both amino acids and monosaccharides is a relatively complex process, consisting of three main steps: hydrolysis of samples to liberate the monomers, derivatization and chromatographic analysis, however, they are preferred over

their polymers.[120] This is because amino acids and monosaccharides are relatively stable even when the samples are soaked in water and/or subjected to heat.[121, 122] Additional information on the type of protein and carbohydrate could also be provided by the analysis of amino acids and monosaccharides. Hence, amino acid and monosaccharide analysis are selected for this chapter, informed by the various above mentioned factors.

The analysis of amino acids and monosaccharides in EBN has been on-going for more than a decade, with most analyses focusing on the measurement of common amino acids and monosaccharides originating from protein and carbohydrate. As such, the amino acids and monosaccharides that are not generally detected in protein and carbohydrate will also be taken into account in this chapter.[123, 124] This would permit the understanding of different types of protein and carbohydrate molecules existing in the EBN, thus increasing the analytical significance of this chapter.

The aim of this chapter was to demonstrate the potential of amino acid and monosaccharide analysis, combined with multivariate data analysis, to establish a Hotelling T2 range plot. This plot was used to identify whether a sample is EBN or not based on species origin. The practical applicability of the Hotelling T2 range plot was evaluated with different types of EBN and samples suspected to possess a similar matrix to EBN(non-EBN). Furthermore, EBN was compared with other food items rich in protein and/carbohydrate to assess the nutritional value of the former. This comprehensive study of the amino acids and monosaccharides will hopefully not only be able to create a

new approach to identifying EBN, but also offer insightful information on the monomers (amino acids and monosaccharides) and polymers (protein and carbohydrate) of EBN.

2.2. Materials and methods

2.2.1. Information on the samples

In this chapter, a total of 138 analyses were conducted on the EBN. These EBNs were obtained from different suppliers in the local and overseas markets so as accurately represent the entire spectrum of the food item. All EBNs used in this chapter had been processed prior to the analysis. Three biological replicates of each EBN sample were performed in this chapter to ensure the precision of the results.

The non-EBN samples including milk, chicken egg yolk, chicken egg white, quail egg yolk and quail egg white were obtained from the local retailers. The rest of the non-EBN samples including fetuin, infant formula and human salvia were obtained from various sources. Fetuin from fetal bovine serum was purchased from Sigma-Aldrich (St Louis, MO) while the standard reference material infant formula, T2562, was purchased from FAPAS. The human saliva was obtained from a male human volunteer within a single day.

2.2.2. Chemicals and materials

Amino acid standards and the amino acid derivatization kit (EZ: faast^{TM} GC-FID amino acid analysis kit) were purchased from Phenomenex. Several chemicals were provided in the EZ: faast^{TM} kit. They were internal standard solution (norvaline), sodium carbonate, sodium hydroxide, N-propanol, derivatizing agent (chloroformate-type compound), isooctane and hydrochloric acid solution. The standard used consisted of the following 17 amino acids: alanine (1), glycine (2), valine (3), leucine (4), isoleucine (5), threonine (6), serine (7), proline (8), aspartic acid (9), methionine (10), 4-hydroxyproline (11), glutamic acid (12), phenylalanine (13), lysine (14), hydroxylysine (15), tyrosine (16) and cysteine (17). The bracket beside the name of individual amino acids contains the labels of amino acids in their chromatogram. The standards were stored in a freezer while chemicals in the kit were stored in a refrigerator at 4 °C prior to their usage.

The chemicals - 1-phenyl-3-methyl-5-pyrazolone (PMP) and trifluoroacetic acid were obtained from Aldrich (St. Louis, Missouri, America). Acetic acid was obtained from Merck, Darmstadi, Germany while hydrochloric acid was purchased from Normapur. Ammonia hydroxide was purchased from Panreac, Barcelona. Ammonia acetate (99.7 % purity) was obtained from Hayashi Pure Chemical Industrial Company, Osaka, Japan. The 7 monosaccharides – mannose (1), glucose (2), galactose (3), rhamnose (4), fucose (5), ribose (6) and xylose (7), used in the working standards were purchased from Sigma Aldrich (St. Louis, Missouri, America.).The bracket beside the name of

individual monosaccharides contains the labels of monosaccharides in their chromatogram. The internal standard for the monosaccharide analysis (D-glucose-$^{13}C_6$) was purchased from Cambridge Isotope Laboratories, America. All the other reagents utilized were of analytical grade.

2.2.3. Amino acid analysis

Five milliliters of concentrated hydrochloric acid (6 N) was used to hydrolyze 0.2 g of EBN for 24 hours. The hydrolyzed sample was filtered into a round bottom flask before evaporating it to dryness. The dried extract was reconstituted with 10 mL of 0.1 N hydrochloric acid and was stored in a refrigerator at 4 °C prior to derivatization.

Fifty microliters of the reconstituted sample were pipetted into a 1-mL vial and topped up to 1 mL with 0.1 N hydrochloric acid. After that, 100 μL of the diluted solution and 100 μL of internal standard were pipetted into a sample preparation vial and vortexed for 1 min. Then, a 1.5-mL syringe with a sorbent tip was immersed into the mixture solution. The mixture solution was drawn up the sorbent tip by pulling the syringe piston slowly to ensure that the amino acids in the mixture solution were attached to the sorbent. Next, 200 μL of deionsied water was added into the same sample preparation vial and was also drawn up the sorbent tip slowly. This step was to make sure that no amino acids remained in the sample preparation vial.

Subsequently, the 1.5-mL syringe was replaced with a 0.6-mL syringe. Close attention had to be paid to ensure the sorbent tip remained in the sample preparation vial at all times during the change of syringe. Next, a mixture solution of sodium hydroxide and N-propanol - consisting of 1 portion of N-propanol to 1.5 portion of sodium hydroxide, was prepared. Two hundred microliters of the mixture solution of sodium hydroxide and N-propanol were pipetted into the sample preparation vial. After that, the mixture solution was drawn up the 0.6-mL syringe and then all the liquid and sorbent particles were ejected into the sample preparation vial.

Next, 50 μL of the derivatizing agent was added to the mixture following by 100 μL of isoctane. The mixture was vortexed for 0.5 min after each reagent is added to into the preparation vial. Finally, 100 μL of hydrochloric acid solution was introduced into the preparation vial and vortexed for another 0.5 min. The emulsion was allowed to stand for 1 min so that two distinct layers would appear. The upper layer was then transferred into a 1-mL vial with a glass pipette for analysis. The amino acids standards used for the construction of the calibration curves were also derivatized in the same way as the sample prior to GC-FID analysis.

Analysis of the amino acids was performed on a Hewlett Packard 6890 Series GC system coupled to an Aligent Technologies 7683 series injector and flame ionisation detector (FID). The Zebron ZBAAA (10 m x 0.25 mm capillary GC column) was used for the GC separation. The carrier gas (helium) had a flow rate of 1.5 mL/min. The column oven temperature was programmed to start at

110 °C and to increase at 32°C per min to 320°C. The FID detector temperature was set at 320°C. Two microliters were injected by an auto sampler at an injection temperature of 250°C and a split level of 1:2. The software Chemstation was used to analyse the results obtained from the amino acid analysis. Result for each detected amino acid was normalized by dividing its concentration against the total concentration of all the 17 amino acids in that sample prior to exporting to multivariate analysis.

2.2.4. Monosaccharide analysis

Two milligrams of EBN were hydrolyzed in 2.0 mL of trifluoroacetic acid (2M) at 120 °C for 4 hours. After that, the hydrolyzed solution was cooled to room temperature prior to filtration into a round bottom flask with a 0.45 μm PTFE filter. Next, the hydrolyzed solution was evaporated to dryness and the dried extract was reconstituted with 3 mL of water.

Two solutions − 0.75 M of PMP dissolved in methanol and 50:50, v/v ammonium hydroxide to deionized water, were prepared for the derivatization reaction. Seven hundred and fifty microliters of PMP solution, 450 μL of ammonia hydroxide and 10 μL of the internal standard (100 mg/L) were added to 150 μL of the reconstituted monosaccharide solution. The mixture solution was then vortexed and incubated in a heating block at 70°C for 2 hours to permit derivatization to occur. Then the mixture solution was evaporated to dryness using a gentle stream of nitrogen gas. Finally, the mixture solution

was reconstituted with 150 μL of ACN and 850 μL of water prior to sending it to liquid chromatography tandem mass spectrometer (LC/MSMS) for analysis.

The calibration standards for the monosaccharide analysis were also derivatized in a same manner. A stock solution of mixture (100 mg/L) of 7 monosaccharides (mannose, glucose, galactose, rhamnose, fucose, ribose and xylose) and an internal standard solution (100 mg/L) of D-glucose-$^{13}C_6$ were prepared with deionized water prior to the analysis. All solutions were stored in the refrigerator at 4 °C until they were needed.

The analysis of monosaccharides were carried out on an Agilent 1100 LC system coupled to an API 3000 Q-Trap LC/MSMS with Turbolon Spray source (Applied Biosystems, Foster City, CA, USA). The analytical column Inertsil HPLC column (ODS-3, 3 μm particle size, 3 mm I.D. x 100 mm) produced by G.L. Sciences, Tokyo, Japan was used for the separation of the monosaccharides. The injection volume was 10 μL. Elution was carried out with a flow rate of 0.35 mL / min at 25 °C. The mobile phase consisted of 100 mM ammonia acetate (pH 6.9) with (A) 0 % of acetonitrile and (B) 60 % acetonitrile. The condition of the mobile phase was follows: 30 to 35 % B (0.0 -25.0 min), isocratic at 35 % B (25.0 -30.0 min), 35 to 55 % B (30.0 -50.0 min), 55 to 30 % B (50.0 -50.1 min), isocratic at 30 % B (50.1 -55.0 min). The mass spectrometer (MS) was equipped with an electrospray ionization interface operating in the positive mode. The MS parameters such as the multi reaction monitoring (MRM) transitions, declustering potential (DP), collision energy (CE), collision exit potential (CXP), entrance potential (EP) and

focusing potential (FP) are stated in Appendix 1. The electron spray ionization (ESI) temperature was 100 °C while the ion spray was maintained at 5500 V and the curtain gas pressure was maintained at 10 mbar. The software Analyst Software 1.4.2 was utilized to analyze the results obtained from the samples. The concentration of each monosaccharide was normalized against the total concentration of all 7 monosaccharides within the same sample before exporting for multivariate analysis.

2.2.5. Statistical analysis

The normalized data from the amino acids analysis and monosaccharides analysis were exported to SIMCAP+ software package version 13 (Umetrics, Umea, Sweden) for the construction of their respective principal component analysis (PCA) score plots. Significance calculation was based on a two-tailed Student's *t*-test performed with Microsoft Excel software (Microsoft Corporation, USA). Compounds with $p > 0.05$ are considered to have no significant difference.

2.2.5. Hotelling T2 range plot

A numerical representation of the PCA prediction model, Hotelling T2 range plot was utilized to visualize the T2 range values of the samples. This Hotelling T2range plot is a plot of T2 range value of the samples from the first principal component to the last principal component selected for the construction of the score plot.[125] The Hotelling T2 range plot contains a

statistically calculated value known as the T2 critical value to determine the presence of outliers at a particular level of confidence. The T2 critical value is a numerical representation of the confidence region in the PCA prediction model, referred to as the Hotelling T2ellipse. As such, any sample with a T2 range value larger than the T2 critical value implies that it is significantly different from the samples that were used to form the Hotelling T2 range plot and vice versa.[126]

The data for this chapter was divided into model set and prediction set. The model set was used to establish the Hotelling T2 range plot for EBN thus the selection of EBNs in this group played an extremely crucial role in the success of the study. EBNs for the model set were collected at random by the authors from an EBN farm located in Johor Bahru, Malaysia to ensure the integrity and quality of the food item from farm to fork. The model set also included biological replicates of the EBNs to factor in the variation of the analytical method. This was to ensure that samples which are dissimilar from the model set could only be attributed to differences in the samples rather than variation in the analytical method. The prediction set consisted of EBNs obtained from the same source as the model set. The prediction set was utilized to determine the optimal scaling method to be employed for the construction of the Hotelling T2 range plot. The established Hotelling T2 range plot was then evaluated with different types of EBN – EBN with different coloration, originating from different countries and/or cleaned with different processing methods. In addition, the evaluation was further extended to samples which were suspected to contain a similar matrix to EBN.

2.3. Results and discussion

2.3.1. Development and validation of an analytical method for the monosaccharide analysis of EBNs

In this chapter, two different types of analysis - amino acid analysis and monosaccharide analysis were utilized to examine EBN. The analytical method used to determine amino acids was a well-established one. It has been demonstrated by Ruiz et al. that the accurate quantitation of amino acids could be performed within a short analysis time of 10 mins with GC-FID.[127] Hence, this analytical method was employed in this chapter for the quantification of the amino acid content in EBN. On the other hand, the analysis of monosaccharides was not as straight forward as the amino acids. Studies have revealed that the monosaccharides originated mainly from the glycans of EBNs.[110, 112] As such, it is expected that the different monosaccharides may differ by a few order of magnitude due to the minute presence of several monosaccharides in the glycans.[109] Thus, the analytical method employed to analyze the monosaccharides should take into account this fact. However, at the moment, most analytical methods do not possess the required dynamic range to quantify the monosaccharides in EBN.[128] In addition, low resolution and long analysis time often hinder the accuracy and efficiency of the available analytical methods.[129] Hence, a new analytical method would have to be developed and validated prior to the analysis of the monosaccharides in EBN.

Acid hydrolysis was performed to release the monosaccharides in this chapter. This is an indiscriminating method which implies that monosaccharides would be released from all the oligosaccharides and/or polysaccharides in EBN.[130] Hence, besides the quantitation of common neutral monosaccharides in glycans - mannose, glucose, galactose, fucose, and xylose, it is also important to include monosaccharides like rhamnose and ribose as they exhibit a high possibility to be in oligosaccharide and/or polysaccharide.[131] After the liberation of monosaccharides, an analytical platform is required for the chromatographic analysis. GC has the advantage of resolving the derivatives of the monosaccharides but they are found to be unstable when exposed to moisture and would produce multiple peaks for a given sugar which complicates the analysis.[132] On the other hand, in LC the derivatives of monosaccharides are relatively stable in water but its major disadvantage lies in the low chromatographic resolution.[133] With the information gathered, LC/MSMS in MRM mode was selected to analyze the monosaccharides of EBN as this analytical platform was able to make use of the advantages provided by both the gradient flow of the mobile phase and the MRM mode to improve the resolution and sensitivity of the analysis.

However, it has been reported that MS detection would not be able to achieve the dynamic range for monosaccharide analysis as monosaccharides have poor MS/MS fragmentation.[134] Thus, a derivatization step was incorporated into the analytical method. The derivatizing agent chosen was PMP because the monosaccharide PMP-derivative possessed excellent fragmentation efficiency for MS detection.[135] In addition, the derivatization reaction would be

conducted under mild conditions and no isomerization of the derivatized monosaccharides would occur.[136] Referring to Figure 9, it is observed that chromatographic separation of mannose, glucose and galactose, rhamnose and fucose, ribose and xylose was achieved with the gradient flow of the mobile phase while co-eluting monosaccharides like rhamnose and ribose were differentiated with the MRM mode.
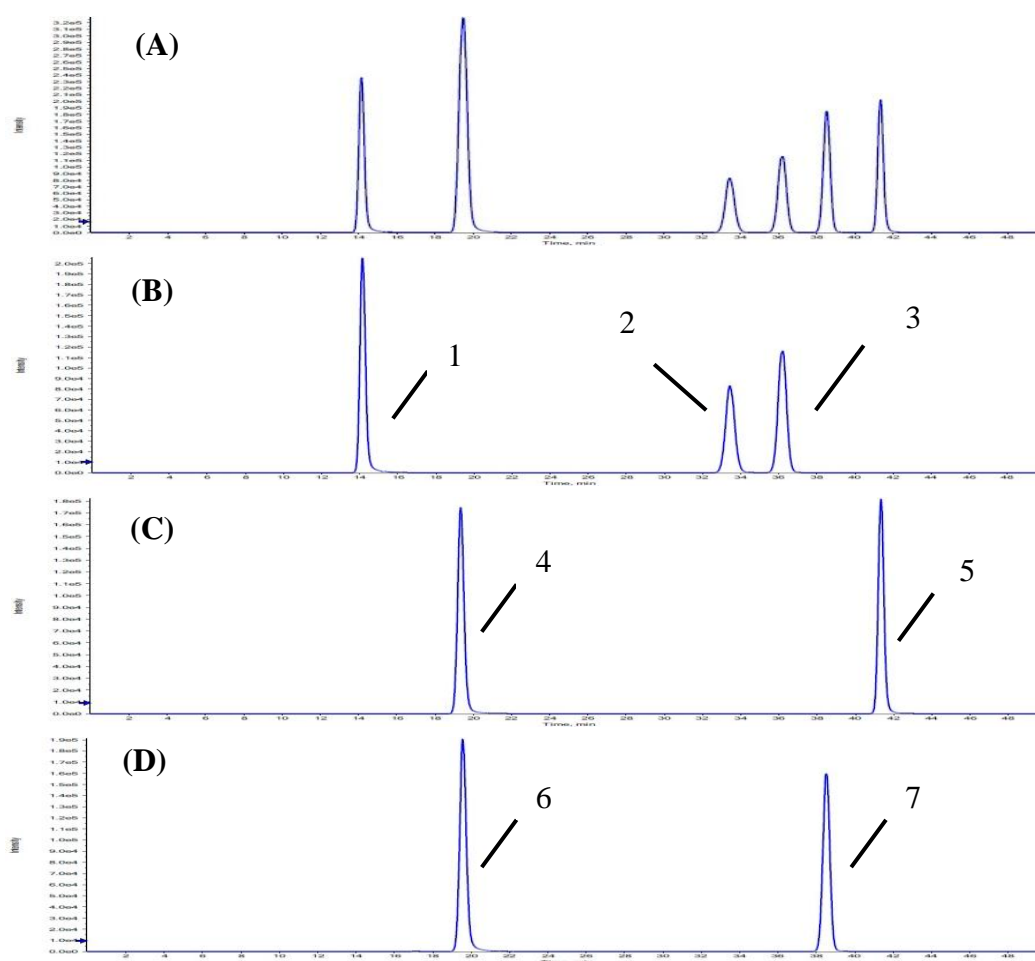

Figure 9 Chromatograms of monosaccharides in a 10000 ppm standard. (A) the TIC. (B) the XIC for mannose, glucose and galactose (m/z 511.4→175.1). (C) the XIC for rhamnose and fucose (m/z 495.3→175.1). (D) the XIC for ribose and xylose (m/z 481.2→175.0).The monosaccharides are labelled according to section 2.2.2.

Two MRM transitions were utilized for the analysis of each monosaccharide. One MRM transition was selected to quantify the monosaccharide, while the other was used to validate its identity (Appendix 1). The precursor ion for the monosaccharides was revealed to be the bis-PMP derivative of each individual monosaccharide. This implies that during the derivatization reaction two PMP molecules would react with one monosaccharide molecule via a condensation reaction. Ions of m/z 175 and 187 were selected as the product ions for all the monosaccharides. The comparable m/z values of the product ions are attributed to the similar structure of the monosaccharide derivatives. Product ion with m/z 175 was deduced to be a PMP as it exhibited an identical molecular mass as a protonated PMP. On the other hand, the identity of product ion with m/z 187 was verified with the internal standard, D-glucose-$^{13}C_6$. It was observed that there was a difference of 1 m/z between the product ion of glucose and D-glucose-$^{13}C_6$. Based on the fact that D-glucose-$^{13}C_6$ is the isotopic form of glucose, the two molecules were expected to produce the same fragment ions but the m/z of the fragment ions would differ due to the variation in the mass of the carbon atom. As such, the product ion with m/z 187 was deduced to be made up of one methyl group from the monosaccharide moiety while the rest of the ion was from the PMP moiety (Figure 10).
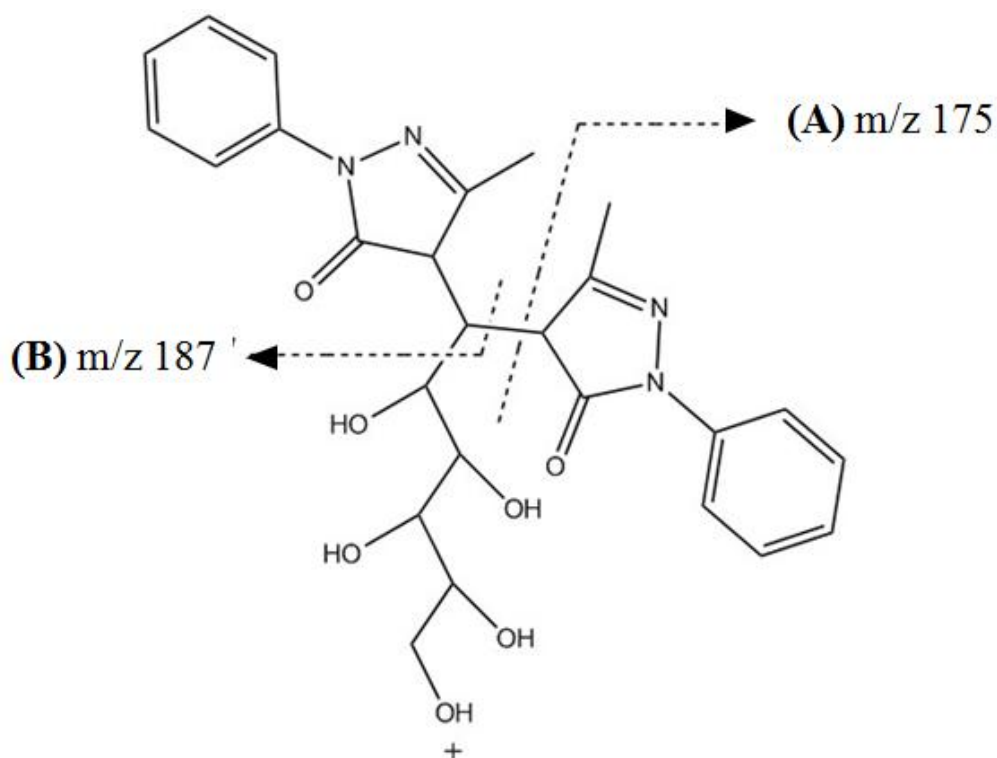
Figure 10 Proposed MS/MS fragmentation of the protonated molecule of glucose derivatized with PMP ([M+H]$^+$, m/z 511). (A) fragment of m/z 187 while (B) fragment of m/z 175.

The optimized analytical method was also applied monosaccharide standards with a concentration range of 50 - 10000 ppb. The linearity of the 7 monosaccharides was evaluated by inspecting the peak area as a function of compound concentration, with the aid of a regression line by the method of least square. Excellent correlation coefficients (> 0.99) were obtained from the calibration curves of the monosaccharaides. The retention time, limits of detection and limits of quantitation were tabulated in Table 3. The LODs and LOQs are observed to be lower than the reported literature values.[135] Moreover, the small fluctuations in the retention time and high correlation coefficient provided confidence for the quantitation of the 7 monosaccharides using the developed analytical method. Although a derivatization step had

been incorporated into the analytical method, this new analytical method has less steps than the recent one suggested by Hammad et al..[137]Hence, the developed analytical method could minimize the loss of monosaccharides and the possibility of errors occurring during sample preparation would be reduced. In addition, with the use of LC/MSMS, the analysis time was shortened and the dynamic range had increased as compared to similar analytical methods.[138]

Table 3 Retention time, linearity, limits of quantitation (LOQ) and limits of detection (LOD) for 7 monosaccharides (n = 6).

Compound Name	Equation	RSD (%)[a]	Linearity	LOD (ppb)	LOQ (ppb)
Mannose	y = 2.30x - 0.0634	1.93	1.000	10	5
Glucose	y = 1.35x + 0.0033	1.58	0.990	20	1 0
Galactose	y = 0.25x + 0.0011	2.09	1.000	20	1 0
Rhamnose	y = 2.19x - 0.0696	2.52	0.990	10	5
Fucose	y = 2.07x + 0.0180	0.58	1.000	10	5
Ribose	y = 2.47x - 0.0474	1.16	1.000	10	5
Xylose	y = 2.20x + 0.015	0.55	1.000	10	5

[a] represents the standard at concentration 50 ppb.
Limit of detection (LOD) is the lowest concentration of a compound which can be measured and it is calculated from the level of the compounds equivalent to three times the standard deviation of noise on analysis.
Limit of quantitation is the lowest concentration of a compound which can be quantified and it is calculated from the level of the compounds equivalent to ten times the standard deviation of noise on analysis.

Validation is an important requirement in the development of a new analytical method. It is the process of confirming that the analytical method under consideration has performance characteristics and capabilities consistent with

what the application requires. In this chapter, the analytical method developed was to quantitate the amount of 7 neutral monosaccharides from the carbohydrate component of EBN. Hence, a reference material, rather than spiking with standards, was employed in the validation.

The reference material, fetuin, was used to validate the monosaccharide analysis. Fetuin was utilized because of its similarity to EBN, it is a matrix rich in glycoproteins.[139] Besides, fetuin is a well-established glycoprotein that has been used in the validation of analytical methods involved in the quantitation of monosaccharides from carbohydrate.[129] At the moment, there is no matrix which permits the study for all 7 neutral monosaccharides in the carbohydrate. Thus, a partial method validation is often adopted to verify the suitability of a analytical method.[140] This is widely accepted due to the highly similar chemical structure and chemical nature of monosaccharides.

An intra- and inter-day study was conducted on the reference material for method validation. Results of the validation were in good agreement with the stated literature range of the reference material (Table 4).[122, 129, 137, 140] Excellent RSDs for the monosaccharides were achieved utilizing the analytical method (RSD < 5 %). Interestingly, several monosaccharides like glucose, rhamnose and xylose were noted to be present in the fetuin even though they are not part of the glycoprotein in fetuin. This is possibly due to monosaccharides attaching to the glycoprotein through glycation.[137]

Table 4 Validation result on the amino acid and monosaccharide analysis of fetuin (n =6).

Compound Name	Assigned Range (mg/g sample)	Intra-day			Inter-day		
		Mean (mg/g sample)	S.D.	RSD (%)	Mean (mg/g sample)	S.D.	RSD (%)
Mannose	25.15 $\pm$ 2.15	23.41	0.70	2.99	24.09	0.69	2.87
Galactose	39.45 $\pm$ 6.45	43.11	1.39	3.22	43.17	0.36	0.84
Fucose	1.00 $\pm$ 1.00	0.83	0.03	4.15	0.87	0.03	3.99
Glucose	N.A	10.30	0.34	3.27	10.33	0.03	0.26
Rhamnose	N.A	0.13	0.004	3.08	0.13	0.00	0.77
Ribose	N.A	0.00	0.00	0.00	0.00	0.00	0.00
Xylose	N.A	2.11	0.10	4.94	2.22	0.09	4.24

S.D. represents standard deviation.
N.A. represents not available.

2.3.2. Establishing the Hotelling T2 range plot to identify EBN

The amino acid and monosaccharide chromatograms representing the EBN are illustrated in Figure 11. Visual inspection of the amino acid chromatograms reveals that 17 amino acids were detected in the EBN. It was noted that several common proteinogenic amino acids such as tryptophan, asparagine and glutamine are not detected in EBN. This is due to the utilization of acid hydrolysis to liberate the amino acids before the chromatographic analysis. This approach would result in the complete hydrolysis of tryptophan and the deamidation of asparagine and glutamine into aspartic acid and glutamic acid respectively.[141] It was also observed from the amino acid chromatogram that there are no obvious differences between the EBNs from the model and prediction set. Meanwhile, the monosaccharide chromatogram shows that 7 monosaccharides were detected in EBN and that there was no clear distinction between the EBNs from the model and prediction set. This provides the confidence that it is viable to establish a T2 critical value with both types of analysis to identify EBN.
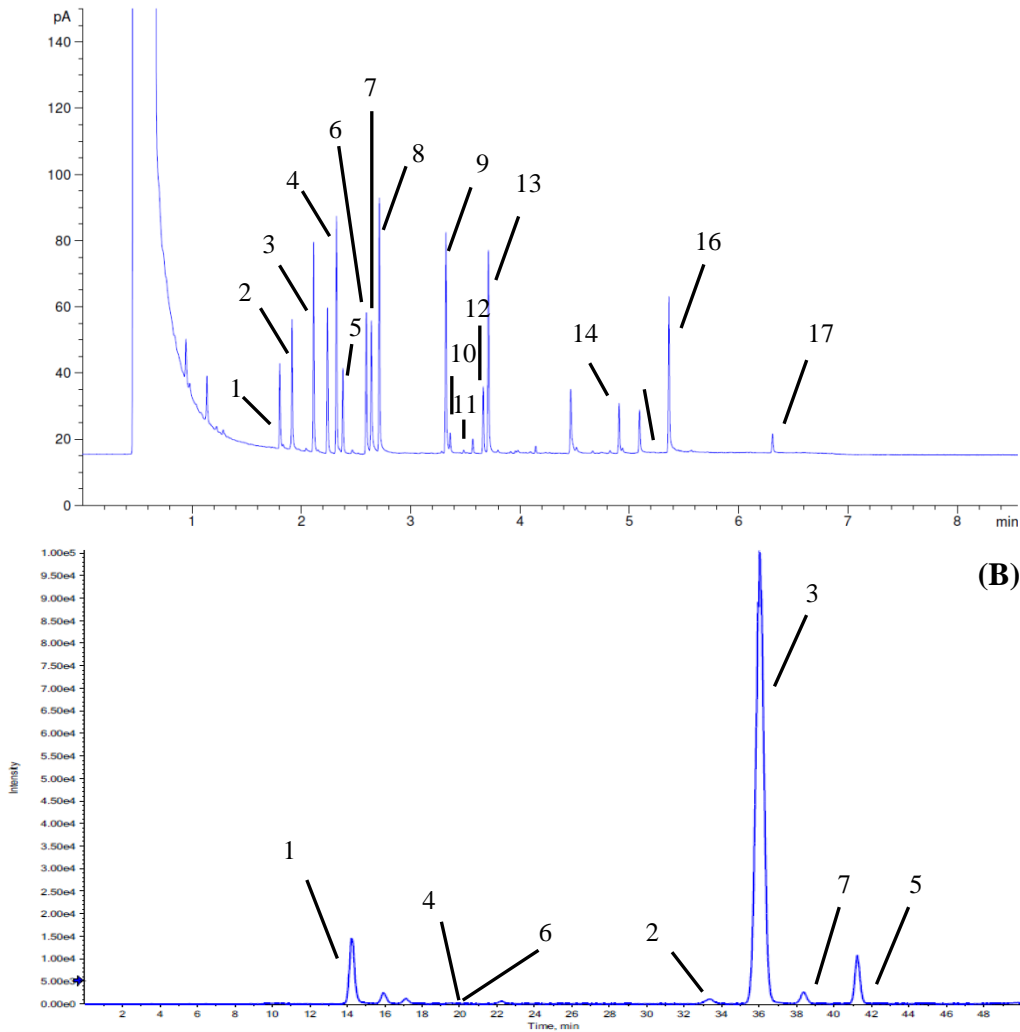
Figure 11 Chromatograms of EBN. (A) the amino acid data - and (B) monosaccharide data. The amino acids and monosaccharides are labelled according to section 2.2.2.

Amino acids and monosaccharides have been used in the past to differentiate food items that are suspected to have a similar matrix.[142] Such studies rely on the fact that the polymers of amino acids and monosaccharides - protein and carbohydrate, would exhibit variations if they are produced from different species. However, in this chapter, a different approach was utilized - the homogeneity of amino acids and monosaccharides. EBN is a food item that is known to be produced by *Aerodramus* genus. Thus, EBNs are expected to

exhibit a homogeneous amino acid and monosaccharide signature as they are originated from the same species origin.

In this chapter, the data from amino acid and monosaccharide analysis was combined with PCA to identify the EBN. PCA is known to be capable of reducing the complex data from individual amino acid and monosaccharide chromatograms into a single score on the score plot. The scores are then utilized to deduce the confidence region known as Hotelling T2 ellipse. Studies have shown that the scaling method of the variables has a direct impact on the prediction ability of the score plot. As such, three different scaling methods − uv scaling, par scaling and no scaling were performed to determine the ideal scaling method to be employed to identify EBN. The PCA score plots for the different scaling methods are displayed in Appendix 2. It was observed that no score was found located beyond the Hotelling T2 ellipse, at 95 % confidence level. In addition, the PCA score plots exhibited satisfactory R^2 and Q^2 that were generally greater than the level of 0.5 (Table 5).

The Hotelling T2 range plots of the respective PCA score plots were plotted to visualize the graphical display in their numerical form. It was evident from the T2 range plots that none of samples from the model set has a T2 range value greater than the T2 critical value (Appendix 3). Thus, an independent group of EBNs known as prediction set would be used to evaluate the different scaled score plots. The result from the amino acid data shows that application of uv scaling would result in a Hotelling T2 range plot that could accurately predict

all the EBN while par and no scaling would create score plots with low prediction ability of EBNs (Table 5). Similarly, the score plots based on the monosaccharide data also demonstrated that uv scaling is the ideal scaling method (Figure 12). Uv scaling involves the division of each variable by its standard deviation. Upon scaling, the variables would have the same weight prior to being subjected to multivariate analysis, thus preventing the dominance of variables with high fluctuations.[143] This indicates that the Hotelling T2 range plots from both data sets have to be unbiased towards all the input variables to be able to predict EBN.
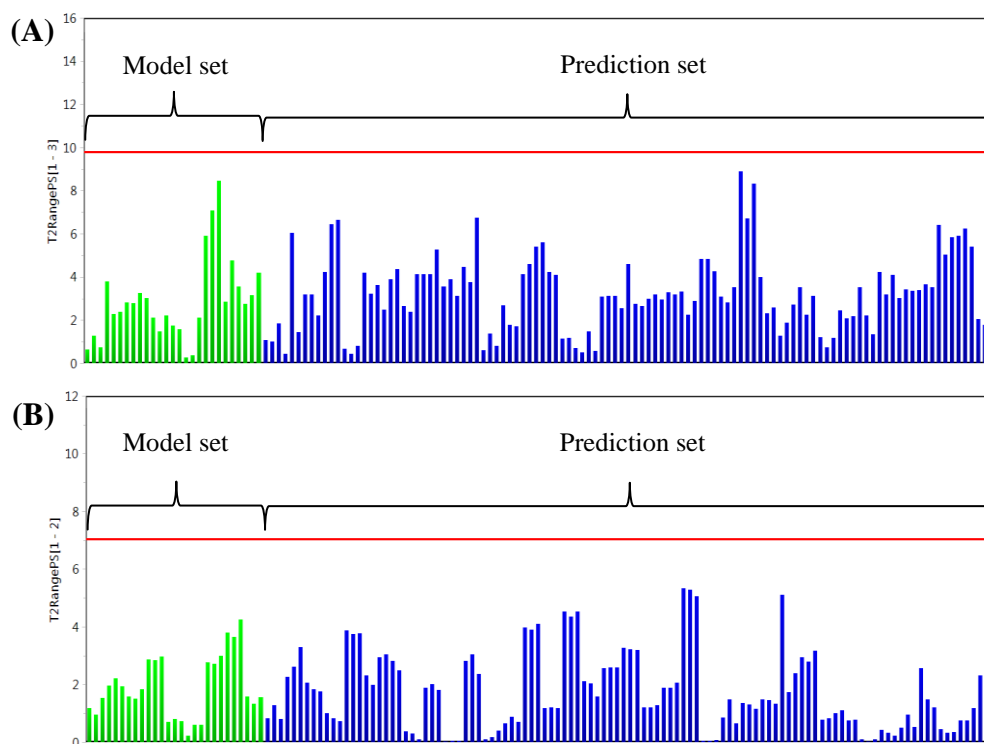
Figure 12 Hotelling T2 range plot of the model set(▇) and prediction set (▇). Plot (A) constructed based on amino acid data subjected to uv scaling. Plot (B) constructed based on monosaccharide data subjected to uv scaling. Both plot (A) and (B) contain a red line representing the critical value.

The amino acid and monosaccharide content of EBN could be significantly affected by the difference in EBN – coloration, country of origin and/or type of cleaning method applied to the EBN. Thus, the optimized Hotelling T2 range plots from both data sets were used to perform a comprehensive study on the different types of EBN. The prediction ability results in Table 5 show that different types of EBN could be accurately predicted by the plots from both data sets, thereby pointing out the feasibility of identifying EBN with amino acids and monosaccharides combined with Hotelling T2 range plot. The success of predicting the different types of EBN also demonstrated that the amino acid and monosaccharide content is not significantly affected by the variation in the coloration, country of origin and/or type of cleaning methods applied to the food item.

Table 5 Model performance parameters and prediction ability of Hotelling T2 range plots using different types of scaling methods on amino acid and monosaccharide data.

	Amino acid			Monosaccharide		
	Uv scaling	Par scaling	No scaling	Uv scaling	Par scaling	No scaling
Number of PCs	3	3	3	2	2	2
R2(X)	0.791	0.824	0.999	0.826	0.906	0.999
Q2(X)	0.524	0.456	0.995	0.560	0.711	0.248
T2 critical value	9.779	9.779	9.779	7.041	7.041	7.041
Prediction ability (%)	100	98.2	75.7	100	88.3	83.8

Recently, several research studies have demonstrated the existence of common glycoproteins among the EBN with different coloration or country of origin.[107, 114] Thus, since the amino acids and monosaccharides originated mainly from the protein and carbohydrate portion of EBN, the similarity in amino acids and monosaccharides could be attributed to the glycoproteins in EBN. In addition, it is suggested the quantity of these glycoproteins is not significantly affected by factors such as coloration, country of origin and/or type of cleaning methods applied to the EBN.

2.3.3. Evaluation of Hotelling T2 range plot with non-EBN

The Hotelling T2 range plots from the amino acid and monosaccharide data were established with 17 amino acids and 7 monosaccharides respectively. These amino acids and monosaccharides are known to be common building blocks for protein and carbohydrate. As such, it is possible that there are samples which contain similar amino acid and monosaccharide content as EBN. Thus, non-EBN samples - milk, chicken egg yolk, chicken egg white, quail egg yolk quail egg white, saliva, fetuin and infant fomula are utilized to determine the differentiating capabilities of the Hotelling T2 range plots. These samples were selected because they are known to contain a high quantity of amino acids and monosaccharides or they were produced in a similar manner to EBN.[142, 144, 145]

From the examination of the Hotelling T2 range plot in Figure13, it was observed that the T2 range values for all the non-EBN samples were larger

than the T2 critical value for both amino acid and monosaccharide data, thereby affirming that the Hotelling T2 range plot from both data sets are objective parameters to identify EBN based on species origin. Furthermore, the result suggests that the common glycoproteins in EBN do not exist in all the samples tested.
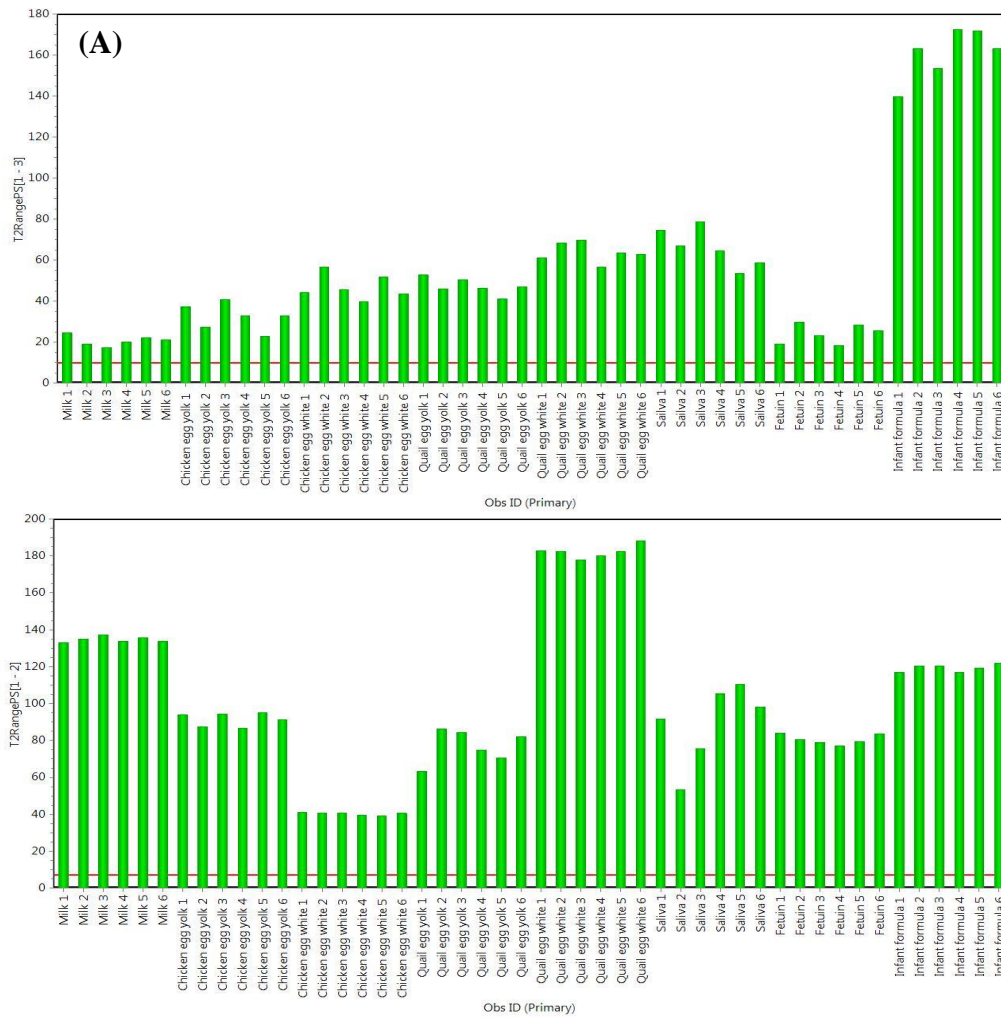

Figure 13 Hotelling T2 range plot of different types of samples. (A) Plot based on amino acid data. (B) Plot based on monosaccharide data.

Contribution plot is also applied to visualizing the variable contribution difference between the EBN and non-EBN.[146] Based on the contribution plots on both the amino acid and monosaccharide data, it was observed that the different types of samples would contain several variables which were clearly separated from the origin (Appendix 4 and 5). Another interesting observation was that samples such as egg yolk and egg white from chicken and quail exhibited a similar variable profile for their respective data sets. This is not unexpected because these samples are similar products, originating from avian. However, in the case of milk and infant formula – both originating from cows, the contribution plots for the amino acid data illustrated a dissimilar profile while the opposite is exhibited in the plots based on monosaccharides. This could be attributed to the fact that infant formula is often enriched with amino acids to meet the needs of infant development (Figure 14).[147] Hence, the contribution plot reaffirms that only samples with a similar amino acid or monosaccharide content to EBN could be identified with the Hotelling T2 range plots.
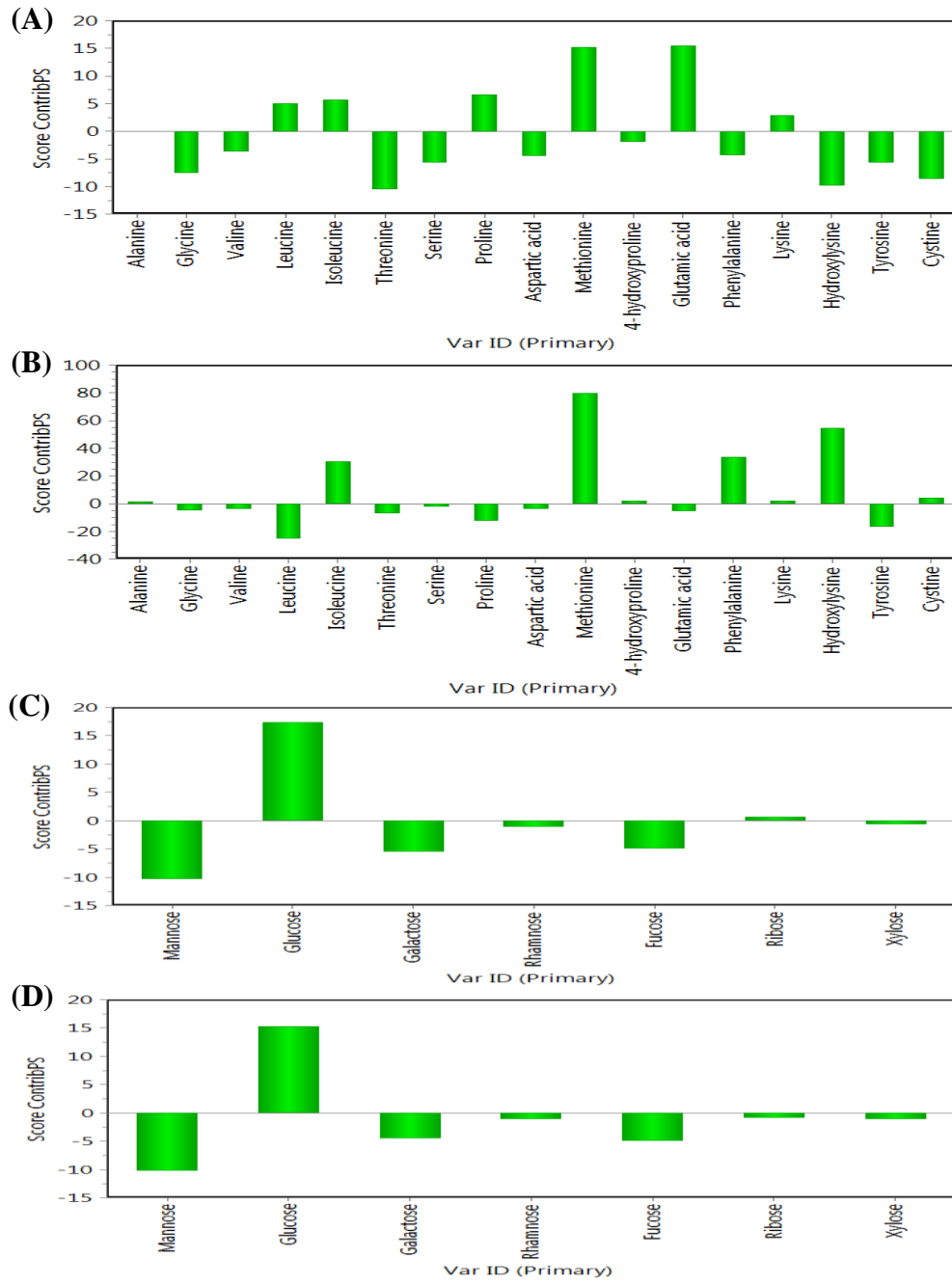
Figure 14 Contribution plots of milk and infant formula. (A) milk and (B) infant formula from amino acid data. (C) milk and (D) infant formula from monosaccharide data.

Comparing the two types of analysis employed in this chapter, the amino acid analysis uses GC-FID while the monosaccharide analysis makes use of the LC/MSMS. Despite a difference in the type and monetary value of the analytical instruments utilized, both types of analyses were able to produce reliable results. The sample preparation in the amino acid analysis involves an extraction method taking up to 24 hours and a number of steps to derivatize the amino acids. On the other hand, the monosaccharides are liberated using a hydrolysis method within 4 hours and then derivatized using a one-step procedure prior to analyzing with LC/MSMS. This greatly reduces the complexity of the sample preparation method of the monosaccharide analysis. However, the analysis time taken to achieve chromatographic separation in the monosaccharide analysis was much longer than the amino acid analysis, which in turn leads to a lower throughput. As there is no single analysis method which is deemed to be more favourable than the other, the type of analytical methods selected ultimately depends on the availability of the chemicals and instruments, as well as any deadline to be met by the users because both types of analysis are capable of producing Hotelling T2 range plots that are able to identify EBN and distinguish EBN from samples with similar matrix (non-EBN) .

2.3.4. Assessment of amino acid and monosaccharide contents of EBN

In terms of the amino acids, aspartic acid, proline and threonine were observed to have the highest concentration in EBN (Appendix 6). This result is in line with the findings published by Theng et al., and thus increasing the reliability of the results on amino acids.[148] Aspartic acid (representing aspartic acid and asparagine) and proline are classified to be non-essential amino acids, implying that they could be synthesized by the human body. On the other hand, threonine is noted to be an essential amino acid (EAA), thus the only way for human to obtain this amino acid is through the consumption of food containing the amino acid. EAAs are vital in the well-being of human's health as EAAs are the precursors for protein synthesis and they are required as intermediates in various pathways of metabolism.[149] In fact, due to the importance of EAA, the World Health Organization, Food and Agriculture Organization of the United Nations and the United Nations University have set a recommended daily amount for EAAs.[113] The tabulation of amino acids also indicates the presence of two non-proteinogenic amino acids namely 4-hydroxyproline and hydroxylsine in EBN. These two amino acids are often used as indicators for the existence of collagen. Thus, it is proposed that the EBN contains collagen.[124, 150]

The nutritional value of EBN was also assessed by comparing EBN with protein- and/or carbohydrate-rich food items such as milk, chicken egg yolk, chicken egg white, quail egg yolk, quail egg white and infant formula. Generally, all the amino acids have a higher concentration in EBN than the

rest of the food items with the exception of methionine in infant formula. Thus, EBN is a better provider of amino acids than the protein rich food items. The total amino acid content was found to be larger in EBN (Figure 15). This indicates that protein constitutes a larger compositional fraction in the EBN matrix relative to the protein rich food items. EBN is also more enriched in EAAs than the protein and/or carbohydrate rich food items. Studies have shown that a higher EAA content would provide more building blocks for protein synthesis as well as increase the rate of the protein synthetic pathway.[151] As such, apart from the epidermal growth factor in EBN, the high EAA content is suggested to support the growth effects caused by the consumption of EBN.
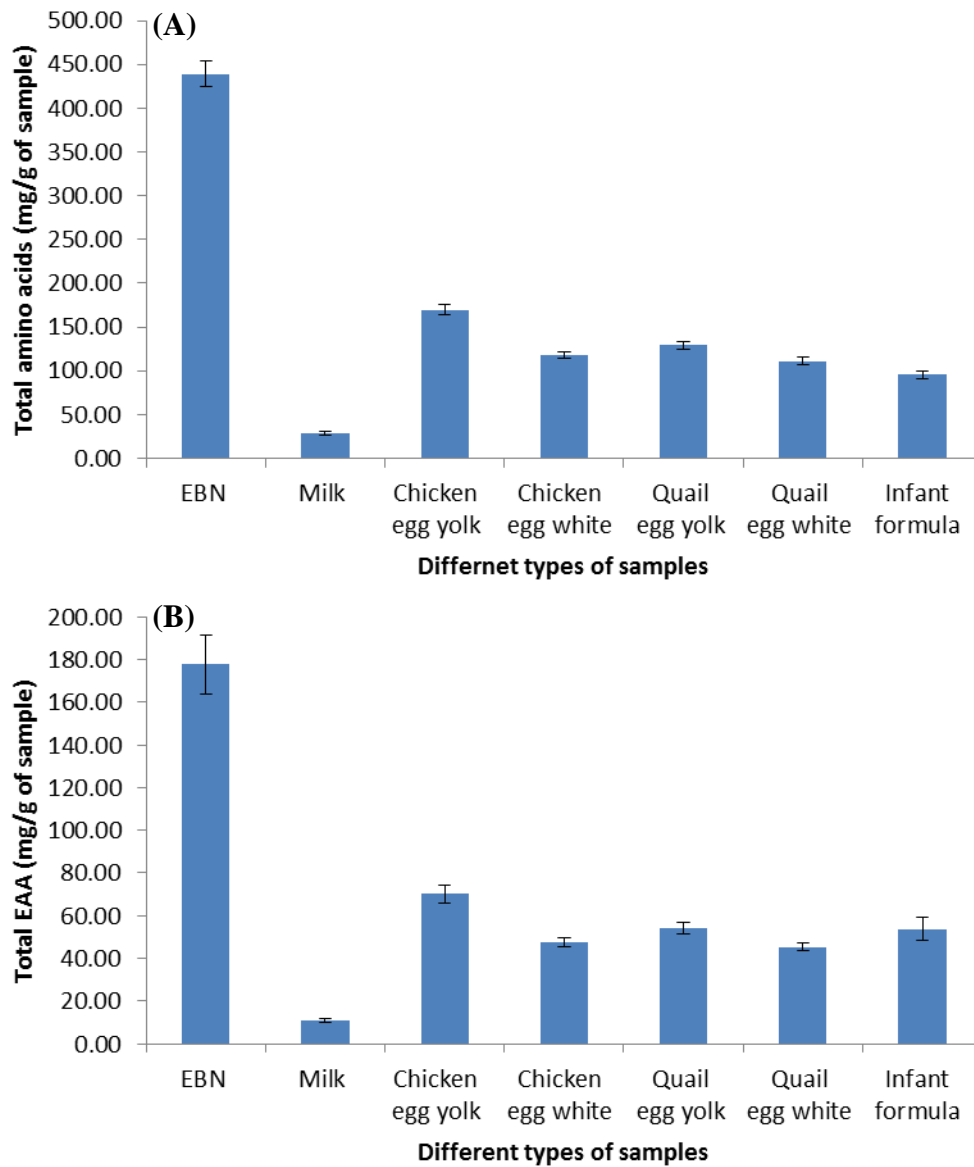
Figure 15 Column plots of (A) total EAA and (B) total amino acids of different types of samples.

In the case for the monosaccharide data, the most prominent monosaccharide in EBN is galactose (111.85 μg / g of sample), which is likely to be derived from the glycans in EBN. This deduction was made as several studies pointed out that galactose is the major neutral monosaccharide in the glycans of EBN.[110-112] Besides galactose, mannose, glucose, rhamnose fucose and xylose were suggested to be derived from glycans as well. Ribose is the only monosaccharide that has not been demonstrated to be present in the glycans of EBN. Its presence is most probably due to the DNA in EBN.[105] It was noted from Table 6 that the monosaccharides have a much larger variation as compared to the amino acids – RSD of monosaccharides range from about 50 to 90 % while it is less than 20 % in amino acids. As mentioned earlier, monosaccharides are derived generally from the glycans. Glycans have linear or branched monosaccharide structures protruding out from the protein molecule.[152] Thus, glycans might have been lost due to degradation during the cleaning process of EBN. In addition, Li et al. demonstrated that proteoglycans of EBN could be extracted with water.[109] This implies that the loss of proteoglycans during the cleaning process of EBN could also have contributed to the larger RSD values in the monosaccharides.

Comparing the monosaccharides of EBN and the different types of food items, all 7 monosaccharides were detected in EBN while only certain monosaccharides were present in the rest of the food items (Table 6). This diversified monosaccharide profile in EBN could be attributed to the glycans detected in the food item mentioned previously. EBN has a significantly lower concentration of glucose than milk and infant formula. This is because both

milk and infant formula contain a large amount of lactose.[144] This reasoning is further supported by the ratio of glucose and galactose in milk and infant formula – approximately 1:1 which coincides with lactose. Besides milk and infant formula, the glucose concentration in EBN is also found to be significantly lower than in chicken and quail egg white. Egg white of chicken and quail are known to be the source of energy for the avian embryonic growth.[153] As such, it was not surprising that the concentration of glucose is elevated in the chicken and quail egg white. Although the concentration of individual monosaccharides varies between the different types of samples, EBN generally has a higher overall monosaccharide content, with the exception of infant formula. This could be due to the presence of a rich lactose content in infant formula to satisfy the nutritional and energy requirements for infant growth.[154, 155]

Table 6 Concentration of monosaccharide of EBN and the different types of food items.

Compound	Concentration of monosaccharide for different food items (mg / g of sample)						
	EBN[a]	Milk[a]	Chicken Egg Yolk [a]	Chicken Egg White [a]	Quail Egg Yolk [a]	Quail Egg White [a]	Infant formula [a]
Mannose	10.53 ± 5.97	0.00 ± 0.00	1.59 ± 0.15	2.84 ± 0.18	2.01 ± 0.18	2.63 ± 0.26	0.53 ± 0.02
Glucose	4.51 ± 3.89	57.29 ± 1.49	2.19 ± 0.19	7.25 ± 0.54	2.82 ± 0.26	57.49 ± 4.91	532.28 ± 26.56
Galactose	111.85 ± 57.22	42.94 ± 0.80	0.82 ± 0.07	0.54 ± 0.04	1.07 ± 0.10	10.48 ± 1.13	520.51 ± 28.85
Rhamnose	0.19 ± 0.15	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Fucose	11.39 ± 9.91	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Ribose	0.18 ± 0.12	0.26 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Xylose	2.13 ± 1.91	0.66 ± 0.03	0.00 ± 0.00	0.52 ± 0.06	0.11 ± 0.01	0.73 ± 0.06	1.19 ± 0.05
Total monosaccharide	140.77 ± 75.32	101.15 ± 2.11	4.60 ± 0.40	11.15 ± 0.71	6.01 ± 0.50	71.33 ± 6.18	1054.52 ± 53.53

* represents no significance difference, $p > 0.05$

[a] represents n = 6.

2.3.5. Quality control of EBN with OPLS-DA

In section 2.3.4., it has been established that EBN is a better amino acid and monosaccharide provider than most food items, thereby implying a difference between the quality of EBN and other protein- and/or carbohydrate-rich food items. Thus, it is possible to utilize the existing amino acid and monosaccharide data for quality control of these food items to protect the interest of consumers and food suppliers.

In this quality control, multiple variables are utilized as they are more objective descriptors of the food item, in comparison to a single variable. To make use of the multiple variables, OPLS-DA is applied to the amino acid and monosaccharide data to reduce the dimensions as well as to maximize the separation of food items. This would result in the ease of determining EBN and other protein- and/or carbohydrate-rich food items. OPLS-DA score plots constructed based on the amino acid and monosaccharide data (Figure 16A and B) exhibited satisfactory goodness of fit and prediction (R2X (cumulative)$_{amino\ acid}$ = 0.996, R^2Y (cumulative) $_{amino\ acid}$ = 0.932, Q2X (cumulative) $_{amino\ acid}$ = 0.861 ; R2X (cumulative)$_{monosaccharide}$ = 0.999, R^2Y (cumulative)$_{monosaccharide}$ = 0.531, Q2X (cumulative)$_{monosaccharide}$ = 0.544). Clear discrimination between EBN and the other protein- and/or carbohydrate-rich food items is observed, thus illustrating the feasibility of using amino acid and monosaccharide data combining with OPLS-DA to differentiate food items with dissimilar quality.
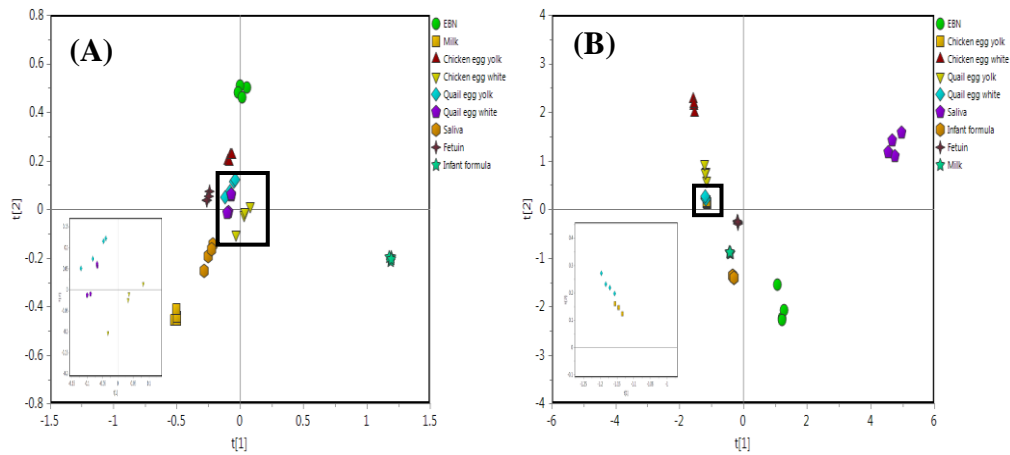
Figure 16 OPLS-DA score plot for the quality control of EBN. (A) based on the amino acid data and (B) based on the monosaccharide data. The highlighted portion for each score plot is display on the bottom left.

Validation of OPLS-DA plots constructed from both data sets is conducted with prediction sets made up of biological replicates from each type of food item (n = 18). An excellent validation result of 100 % was achieved with these prediction sets. This suggests that amino acid and monosaccharide analysis is capable of reproducing the EBN and protein- and/or carbohydrate-rich food items results for quality control purposes.

2.4. Conclusion

In this study, the approach of combining amino acids and monosaccharides with multivariate analysis to establish Hotelling T2 range plots for EBN was adopted to identify the food item based on species origin. However, it was found that there was no suitable analytical method to determine the monosaccharide content of EBN. As such, an analytical method was

developed and validated prior to establishing the Hotelling T2 range plot. The approach was observed to be successful in identifying and predicting different types of EBN. This outcome revealed that EBN has a similar amino acid and monosaccharide content, which is likely to be due to a common group of glycoproteins in the food item. The approach also illustrated a clear distinction between EBN and samples that contained a similar matrix (non-EBN)when compared. Thus, verifying the practical applicability of a Hotelling T2 range plot to identify EBN. The result also suggested that the common group of glycoproteins in EBN is unique to the food item.

Besides using amino acids and monosaccharides for the identification of EBN, the nutritional value of EBN could be determined through comparison with food items rich in protein and/ carbohydrate. The measurement of amino acids revealed that EBN contains a higher protein and EAA content. In addition, the EAA content was suggested to have contributed to the growth properties brought about by the consumption of EBN. In terms of the monosaccharides, EBN was found to contain a more diversified monosaccharide profile, thus suggesting that EBN is a more comprehensive provider of monosaccharides than the other food items. The amino acid and monosaccharide data could also be utilized in the quality control of EBN by combining the data sets with OPLS-DA.

CHAPTER 3. CLASSIFICATION OF EDIBLE BIRD'S NEST WITH METABOLITE FINGERPRINTING

3.1. Introduction

Protein and carbohydrate are known to be the major compositional fraction of EBN – taking up 60 % and 30 % of the mass of the food item respectively.[113, 116]Thus it is not surprising that protein and carbohydrate are common targets of interest for the quality assessment of EBN. An example is Cheung et al. has found that it is possible to classify the different grades of EBN through the quantitation of sialic acid.[156] Apart from protein and carbohydrate, EBN is also made up of lipids and other small molecules. So far information on these other compounds in EBN is limited, thus metabolite fingerprinting was employed in this chapter to determine all the possible chemical components in EBN.

Metabolite fingerprinting is a powerful tool for the comprehensive analysis of a collection of small molecular metabolites of the matrix, which could be influenced by genetic modification, pathological stimuli or environment.[157] Metabolite fingerprinting is frequently used in the analysis of biological matrixes to identify potential biomarkers for diseases.[158] In recent years, its application has also gained popularity in food classification.[159] This is because metabolite fingerprinting is an untargeted analytical approach which enables it to cover all or most of the metabolomic molecular information of the food item, and thereby increasing the analytical significance of the analysis. In fact,

metabolite fingerprinting is recognized by the WHO and United States Food and Drug Administration (FDA) as a strategy to assess the quality of food products and thus increasing our confidence in using this approach to analyse EBN.[160]

The success of metabolite fingerprinting relies heavily on the type of analytical platforms used. The most extensively utilized analytical platforms are NMR and mass spectrometry coupled with chromatographic methods such as GC/MS and LC/MS. NMR is commonly used as it is a non-destructive and unbiased analytical method, requiring minimal sample preparation and relatively short analysis time as compared to the mass spectrometry techniques.[161] However, in this chapter, mass spectrometry techniques coupled with LC and GC are chosen for the classification of EBNs. This is because mass spectrometry is able to provide better sensitivity and better resolved peaks in comparison to NMR.[162] Adding to the advantages of mass spectrometry, the identification of metabolites by GC/MS is facilitated by the use of mass spectra libraries while in LC/MS, the process is aided with the utilization of high mass accuracy and information on the MS/MS fragmentation.

It is very time consuming to interpret the large amount of chromatograms generated by metabolite fingerprinting with GC/MS or LC/MS.[163] Thus, in order to convert the graphical chromatograms into a comprehensible format, the data from metabolite profiling is usually combined with multivariate data analysis such as the unsupervised model, PCA, or a supervised model such as

PLS-DA and OPLS-DA. PCA is often the preferred multivariate approach as it does not assume the class of the given samples prior to the analysis. This approach, however, might lead to poor classification results because it is unable to resolve minute differences between the different classes of samples. In such cases, supervised models would then be utilized to achieve a better classification.[164]

The aim of this chapter was to demonstrate the use of metabolite fingerprinting to identify the chemical components in EBN. Although metabolite fingerprinting on food products is not novel as mentioned, it would be a first application to EBN so as obtain an alternative perspective on this food item. The chapter also aims to further extend the results obtained from metabolite fingerprinting, combined with multivariate data analysis, to classify EBNs according to their coloration, country of origin and production sites. This is especially crucial as consumers are unable to distinguish between EBNs that are from the same classification category due to their similar morphology. They have to instead rely on food labels on EBN products and their prices to determine the type of EBN they are purchasing. Score plots constructed were further validated with external prediction sets to explore the possibility of replicating the outcome and also the impact of the EBN processing method on the metabolite composition. The metabolite fingerprinting of EBN will, hopefully, not only be able to classify EBNs and also offer logical explanations on the differentiation of EBNs.

3.2. Materials and methods

3.2.1.Sample information

In this chapter, a total of 152 EBNs were analyzed. These EBNs came in different color and originated from various countries and production sites to represent the entire spectrum of EBN available in the market. The EBNs were blended into tiny pieces and stored at 4°C prior to analysis.

The GC/MS and LC/MS data were divided into model and prediction sets. The model set represented EBNs used to construct the score plots. It consisted of EBNs obtained directly from the suppliers. The model was also made up of biological replicates of the samples to account for the analytical variations. This permitted the authors to be certain about the information on the EBNs, thereby increasing the confidence level of the results. In this chapter, the prediction set was subdivided into two groups − prediction set A and prediction set B. Prediction set A comprised of EBNs provided from the suppliers whom the model set was obtained. The purpose of prediction set A was to evaluate the accuracy of the score plot. Prediction set B was made up of processed and non-processed EBNs purchased from the local market. The non-processed EBNs were cleaned with the procedures stated in Section 1.2.3. On the other hand, the processed EBNs were biological replicates of the model set. As such, the use of prediction set B enabled the assessment of the effects of cleaning procedures on the classification of EBNs and the precision of the score plot.

3.2.1. Chemicals and materials

Acetonitrile and formic acid of HPLC grade were obtained from Sigma-Aldrich (St.Louis, Missouri, America). Deionized water was purified by the MilliQ system (Millipore, Massachusetts, America). Standard compounds such as lauric acid, myristic acid palmitic acid, stearic acid, arachidic acid and cholesterol were purchased from Sigma-Aldrich Inc. (St. Louis, Missouri, America) while propanetriol was purchased from Merck (Germany). Solvents such as chloroform and methanol were obtained from Kanto Chemical Co., Inc. 2-8 Nihonbashi Honcho 3-Chrome Chuo-Ku, Tokyo, Japan and toluene was obtained from Merck (Germany). The derivatizing agent N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) with 1 % trimethylchlorosilane was purchased from Restek (Benner Circle, Pennslyvania, America).

3.2.2. Analysis with GC/MS

Five hundred milligrams of EBN were sonicated with 10 mL of chloroform:methanol (1:1) for 180 min to extract the metabolites. The solvent mixture was then filtered through a 0.45 μm PTFE membrane to remove the EBN and impurities. Subsequently, 1 mL of the solvent mixture was concentrated under a gentle stream of nitrogen. After that, 100 μL of toluene (dried over anhydrous sodium sulfate) was added to the dried mixture so that no moisture was retained. The mixture was then evaporated to dryness using nitrogen. Next, 100 μL of BSTFA and 100 μL of pyridine were pipetted into the dried mixture and heated in a 70°C water bath for 30 min to permit

derivatization to occur. Finally, 100 μL of the derivatized sample was sent to GC/MS for analysis.

Analysis of the metabolites was performed on a Shimadzu QP2010 Plus (Shimadzu, Kyoto, Japan). A DB-5MS fused capillary column (30 m x 0.25 mm i.d.; Agilent J&W Scientific, Folsom, California, America) chemically bonded with 5% phenyl-methylpolysiloxane and 95 % dimethylpolysiloxane cross-linked stationary phase (0.25 μm film thickness) was used. Helium was used as the carrier gas at 1.2 mL min^{-1} and the injector split ratio was set to 1:5. An injection volume of 1 μL was used and the solvent cut-off time was 5 min. The injector and source temperatures were at 250 and 200 °C, respectively. The column oven temperature was kept maintained at 60 °C for 3 min, increased at 7 °C min^{-1} to 140 °C where it was kept for 4 min and further increase at 5 °C min^{-1} to 310 °C where it remained for 6 min. The mass spectrometer was operated in electron impact (EI) ionization at 70eV. Data acquisition was performed in the full scan mode from mass to charge (m/z) 50 to 650 with a scan time of 0.5 s. Chromatogram acquisition and tentative compound identification was performed with Shimadzu GCMSsolution (version 2.71) software.

3.2.3. Analysis with LC/MS

The extraction procedure used for LC/MS analysis was identical to the one mentioned in section 3.2.2. However, in the LC/MS analysis, the concentrated

solvent mixture was reconstituted with 400 μL of the methanol prior to analysis.

To perform the metabolic profiling, the LC/MS system utilized was the Shimadzu LC/MS– IT-TOF which comprised of Prominence LC system (Shimadzu, Kyoto, Japan), electrospray ionization source (ESI) and a hybrid mass spectrometer combining a quadrupole ion trap (IT) and time of flight (TOF). The HPLC separation was performed on a Synergi Polar-RP column (50 mm x 2 mm IC, 2.5 μm particle size, Phenomenex).The mobile phase consisted of (A) 0.1 % formic acid in distilled water and (B) 0.1 % formic acid in acetonitrile. A gradient elution was utilized for the chromatographic separation of the metabolites. The conditions of the mobile phase elution were as follows: 5 to 100 % B (0.0-10.0 min), isocratic at 100 % B (10.0-14.0 min), 100 to 5 % B (14.0-14.1 min), isocratic at 5 % B (14.1-19.0 min). The total flow rate was set at 0.3 mL min^{-1} throughout the analysis. The injection volume was 2 μL. The column temperature was maintained at 40 °C. Mass acquisition of the metabolites was performed at positive (ES+) and negative ESI (ES-) modes to detect various metabolites with different ionization behaviour. The mass spectral was recorded over a m/z range from 150 – 1000 to ensure a wide coverage of the metabolites. The operating conditions of the mass spectrometer is as follows: heating block temperature 200 °C, curved desolvation line temperature 250 °C; nebulizing nitrogen gas flow 1.5 L/min, ion accumulation time for ES+ is 30 ms and 100 ms for ES-.

3.2.4. Pre-processing of GC/MS data

The data collected from GC/MS was processed with Shimadzu GCMSsolution (version 2.71) software. A reference compound table was constructed to tabulate the peak area of the detected peaks in the chromatograms. In this table, a single m/z (target ion) was utilized to quantify each metabolite. Additionally, 2 reference ions were selected to be used along with the target ion to validate the identity of the metabolite. The identification of the peaks was conducted with National Institute of Standards and Technology (NIST08 and 08s), Wiley EI mass spectra library and reference standards. Peaks with mass spectra having a similarity index (SI) higher than 70 % were assigned with compounds names while the rest of peaks were labelled as unknown. The peak area for each detected peak was normalized against the sum of the peak area within the same sample before exporting for multivariate analysis.

3.2.5. Pre-processing of LC/MS data

The LC/MS data collected was analysed with Shimadzu LCMS solution Version 3.60 software. Similarly to the GC/MS analysis, a reference compound table was constructed by registering the peaks from the EBN chromatograms. Each peak was assigned with an arbitrary identity based on their retention time and m/z. The compound table was used to extract and tabulate the area of the detected peaks in the EBN chromatograms. During data tabulation, only ions which demonstrate the same RT (tolerance of 0.01 min) and m/z value as the ions (tolerance of 5 ppm) in the compound table

were considered the same and their peak areas were calculated. The peak area for each detected peak was then normalized against the sum of the peak area in that sample.

In order to identify the metabolites, the Formula Predictor software version 4.3 (Shimadzu Corp, Kyoto, Japan) was used to deduce the empirical formulae. To perform the deduction, the software takes into account of the accurate ion mass (MS), experimental fragmentation pattern (MS/MS) and isotopic profile. In addition, the accurate mass of the metabolites was searched against several online databases: LIPIDS MAPS (http://www.lipidmaps.org/), Metlin database (http://metlin.scripps.edu/index.php), Massbank (http://www.massbank.jp), Human Metabolome Database, (http://www.hmdb.ca/), and Chemspider (http://www.chemspider.com/).Compounds with an error lesser than ±5 ppm and sharing the same empirical formulae deduced by Formula Predictor software were compiled into a list of probable metabolite in Microsoft Excel software (Microsoft Corporation, America). This list was then narrowed down to a single compound for individual metabolites by comparing the experimental MS/MS spectra with those provided either in online databases and/or scientific journals. It should be noted that in this chapter the indisputable confirmation of the metabolites with the use of chemical standards was not performed, as such, the identification performed was considered as tentative.

3.2.7. Statistical analysis

The normalized data from LC/MS and GC/MS was exported to SIMCAP+ software package version 13 (Umetrics, Umea, Sweden) for the construction of their respective multivariate analysis score plots. In this chapter, two different types of multivariate analysis: principal component analysis (PCA) and orthogonal projection to latent square discriminant analysis (OPLS-DA), were utilized for the classification of the EBNs. All data used was uv scaled prior to the construction of the score plots. A significant level of the normalized data was calculated with one-way Anova test in Microsoft Excel software (Microsoft Corporation, America). Compound with a p value lesser than 0.05 is considered to be statistically significant ($p < 0.05$).

3.3. Results and discussion

3.3.1. Profiling of metabolites using GC/MS

Prior to the analysis using GC/MS, derivatization of the metabolites was performed with BSTFA. This derivatization replaces the active hydrogen in −OH group or −NH group with a trimethylsilane (TMS) moiety in the metabolites to improve their volatility and thermal stability for GC/MS analysis.[165] The profiling of EBNs with GC/MS revealed that there are 43 metabolites (Figure17).
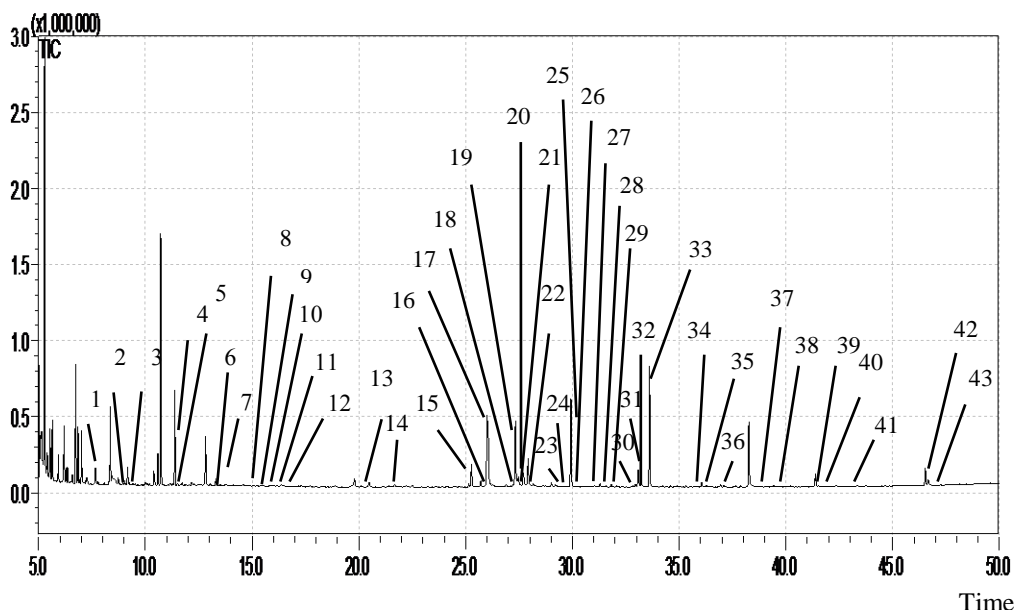
Figure 17 A Typical total ion chromatogram of EBN analysed with GC/MS.

Of the 43 metabolites, 42 compounds were assigned with a compound name while 1 compound was labelled as 'unknown' since the SI of this compound did not achieve 70 % when the mass spectra was compared with NIST and Wiley library. The identification of the metabolites is tentative until they are confirmed with standards. Hence, standard compounds were analysed and the retention time and mass spectra were compared to the detected metabolites to verify their identities (Table 7).

The 43 identified metabolites belonged to a diverse chemical class of organic acids, amino acids, sugars, sugar alcohols, fatty acids, fatty acid amides and steroids. This illustrates that the use of GC/MS for metabolite profiling is desirable as it is capable of detecting a wide range of chemical classes. Based on the elution order of the detected fatty acids, it was observed that fatty acids with longer alkyl chain and lesser number of double bonds eluted out of the column at a later retention time. This is because a relatively non-polar

capillary column was used for the chromatographic separation of the metabolites. It was also revealed that the hexadecanoic acid, E-9-octadecadienoic acid and the octadecanoic acid were the three most abundant metabolites in the GC/MS analysis. The detection of these metabolites in EBN was not unexpected in light of previous work conducted on EBN.[114] An interesting metabolite known as thymol-beta-D-glucopyranoside was been identified in the GC/MS analysis. This metabolite has been reported to be an effective agent against food borne bacteria like *Staphylococcus aureus* and *Escherichia coli*.[166] Hence, the identification of this metabolite supports the study on the antibacterial effects of EBN and also the Chinese belief that the consumption for EBN aids in the treatment of diarrhoea.[48]

Table 7 Information on the metabolites analyzed by GC/MS.

No	RT (Min)	Compound Name	SI	Normalized area (%)
1	6.852	Hexanoic acid	95	2.922 ± 4.921
2	8.723	3-Hydroxybutyric acid	97	0.214 ± 0.330
3	8.894	Heptanoic acid	89	0.851 ± 0.338
4	10.885	Octanoic acid	88	0.124 ± 0.161
5	11.118	Propanetriol[a]	96	7.657 ± 11.352
6	12.797	Nonanoic acid	95	0.196 ± 0.214
7	13.436	Isocitric lactone	79	0.006 ± 0.045
8	14.581	Decanoic acid	92	0.150 ± 0.154
9	14.893	3-Hydroxyoctanoic acid	84	0.045 ± 0.173

Table 3.1 continue

No	RT (Min)	Compound Name	SI	Normalized area (%)
10	15.226	Hydroxybutanedioic acid	95	0.210 ± 0.537
11	15.807	5-Oxoproline	95	0.723 ± 1.832
12	16.035	4-Aminobutanoic acid	87	0.056 ± 0.199
13	19.948	Dodecanoic acid[a]	96	0.629 ± 0.371
14	22.008	Xylitol	95	2.104 ± 3.569
15	24.047	Undecanedioic acid	70	0.069 ± 0.146
16	25.266	Tetradecanoic acid[a]	93	1.898 ± 0.971
17	25.28	2-acetamido-2-deoxy-beta-D-glucopyranoside	78	4.477 ± 6.781
18	25.496	Unknown 1	64	0.277 ± 0.526
19	26.293	D-Glucoheptono-1,4-lactone	84	0.750 ± 1.890
20	26.889	2-acetamido-2-deoxy-alpha-D-glucopyranoside	74	1.837 ± 2.932
21	27.071	D-Glucitol	95	0.497 ± 0.809
22	27.474	Pentadecanoic acid	75	1.567 ± 1.411
23	28.919	(Z)-9-Hexadecenoic acid	84	0.570 ± 0.591
24	29.036	(E)-9-Hexadecenoic acid	84	0.683 ± 0.563
25	29.654	Hexadecanoic acid[a]	96	29.105 ± 8.316
26	30.104	Oleanitrile	93	0.215 ± 0.437
27	30.179	D-Galactitol	84	0.826 ± 2.897
28	30.186	Myo-Inositol	90	0.500 ± 0.877
29	31.407	Heptadecanoic acid	90	0.715 ± 0.295

Table 3.1 continue

No	RT (Min)	Compound Name	SI	Normalized area (%)
30	32.551	(Z)-9,12-Octadecadienoic acid	95	1.270 ± 1.714
31	32.772	E-9-Octadecadienoic acid	96	10.316 ± 8.295
32	32.84	E-11-Octadecadienoic acid	95	0.983 ± 0.865
33	33.242	Octadecanoic acid[a]	96	18.260 ± 6.743
34	35.058	9-Octadecenamide	93	0.087 ± 0.319
35	35.176	Arachidonic acid	73	0.061 ± 0.208
36	36.525	Eicosanoic acid[a]	81	0.686 ± 0.493
37	38.366	2-Monopalmitin	79	0.007 ± 0.036
38	38.867	1-Monopalmitin	95	0.109 ± 0.226
39	40.975	Thymol-beta-D-glucopyranoside	88	5.418 ± 7.185
40	41.315	1-Monooleoylglycerol	90	0.102 ± 0.297
41	42.442	Tetracosanoic acid	93	0.259 ± 0.226
42	46.116	Cholesterol[a]	94	3.125 ± 1.709
43	47.169	Ergosterol	88	0.110 ± 0.193

[a] represents metabolites identified using standards.

3.3.2. Profiling of metabolites using LC/MS

A total of 35 metabolites were detected when EBNs were profiled with LC/MS (Figure 18). Out of the 35 metabolites, 15 metabolites were tentatively identified (Table 8).
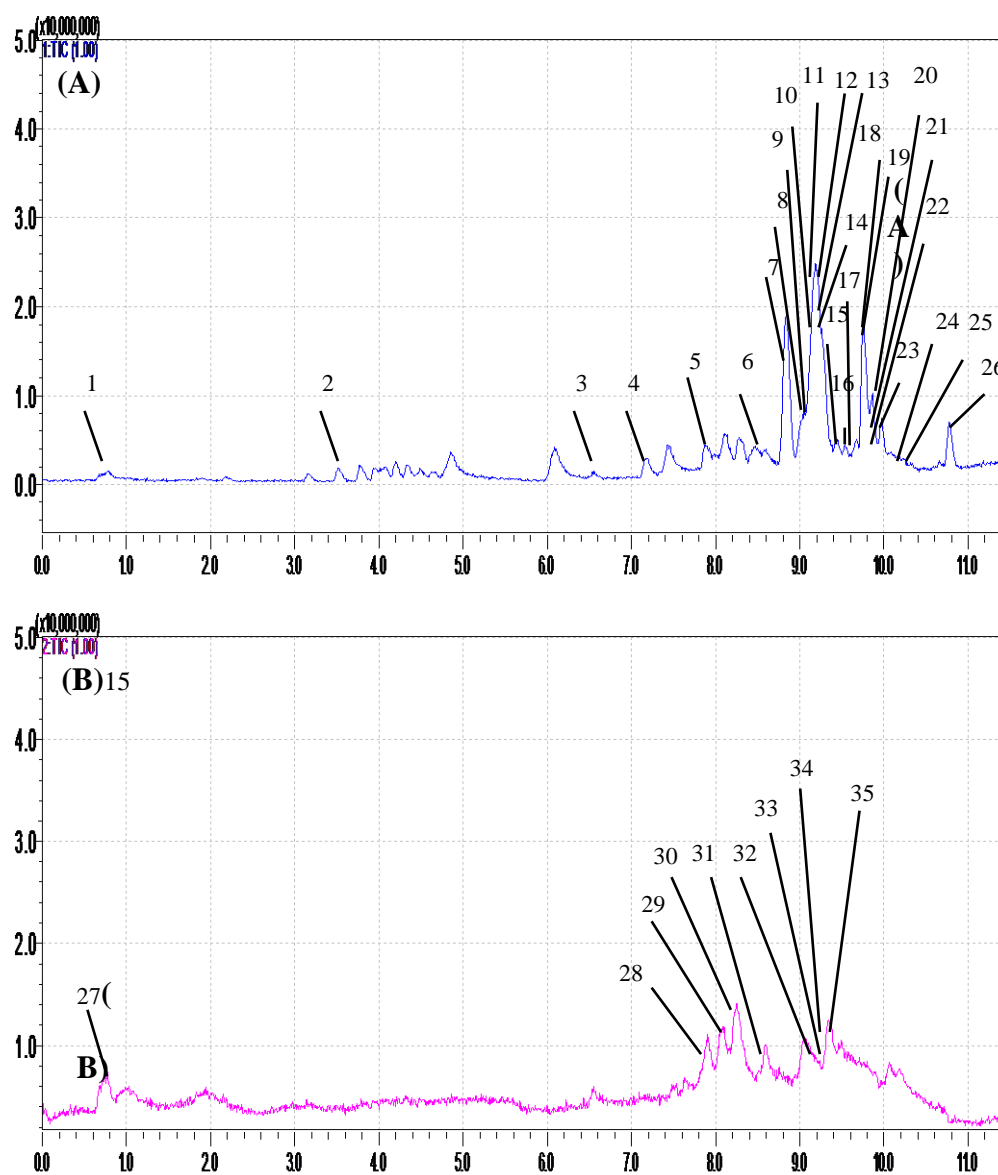
Figure 18 Typical total ion chromatogram of EBN analysed with LC/MS. (A) is in ES+ mode and (B) is in ES- mode.

It was found that a class of compound known as fatty acid amide, which contains a carboxamide head group and an acyl tail chain, possesses characteristic fragmentation pattern coinciding with the MS/MS mass spectra of some of the metabolites. It is known that the saturated fatty acid amide would exhibit a characteristic fragment ion of m/z 102 while the monounsaturated fatty acid amide would produce fragment ions of $[M+H-NH_3]^+$, $[M+H-NH_3-H_2O]^+$ and also alkyl chain fragments of m/z 135 and 121.[167] With this information, metabolites such as myristamide, palmitic amide, oleamide, docosenamide and docosanamide were identified. The search for the metabolite with m/z 401.3394 in Formula Predictor revealed that the empirical formula corresponds to $C_{27}H_{44}O_2$. However, this empirical formula was shared among several compounds. As such, the experimental fragmentation pattern was utilized and it was found that it coincided with the reference MS/MS spectra of a compound in the literature.[168] The metabolite was finally identified as 25-hydroxyvitamin D3. A similar workflow was also adopted for the identification of the metabolite with m/z 385.3452. Using the high mass accuracy, empirical formula and experimental fragmentation pattern, the metabolite was identified to be 7-dehydrocholesterol.[169]

Table 8 Information on the metabolites analyzed by LC/MS.

No.	RT (min)	Ion	Measured Mass (m/z)	Predicted Mass (m/z)	Mass Accuracy (ppm)	Molecular Formula	MS/MS (m/z)	Tentative Identification
Positive mode								
1	0.71	$[M + Na]^+$	226.0688	226.0699	4.87	$C_9H_9N_5O$	167.0255, 148.0384	3-Phenyl-5-ureido-1,2,4-triazole
2	3.67	$[M + H]^+$	432.2780	432.2776	0.93	$C_{14}H_{37}N_7O_8$	221.1394, 177.1128	Unknown 1
3	6.88	$[M + H]^+$	288.2883	288.2897	4.86	$C_{17}H_{37}NO_2$	270.2784, 106.0859	Heptadecasphinganine
4	7.13	$[M + H]^+$	335.2742	335.2739	0.89	$C_{17}H_{40}N_2P_2$	229.2121, 195.1209, 133.0863	Unknown 2
5	7.43	$[M + H]^+$	349.2904	349.2896	2.29	$C_{18}H_{42}N_2P_3$	195.1223, 177.1087, 133.0872	Unknown 3
6	7.92	$[M + H]^+$	228.2319	228.2322	1.31	$C_{14}H_{29}NO$	102.093	Myristamide
7	8.03	$[M + H]^+$	341.2679	341.2686	2.05	$C_{20}H_{36}O_4$	296.2942	Unknown 4
8	8.18	$[M + H]^+$	254.2466	254.2478	4.72	$C_{16}H_{31}NO$	237.2222, 219.2094, 149.1328, 135.1189, 121.1022, 109.1044	Palmitoleamide
9	8.19	$[M + H]^+$	479.2528	479.2540	2.50	$C_{28}H_{34}N_2O_5$	351.2073	Unknown 5
10	8.37	$[M + H]^+$	280.2630	280.2635	1.78	$C_{18}H_{33}NO$	263.2391, 245.2286, 205.0890, 149.1372, 133.1069, 107.8709	Linoleamide

Table 3.2 continue

No.	RT (min)	Ion	Measured Mass (m/z)	Predicted Mass (m/z)	Mass Accuracy (ppm)	Molecular Formula	MS/MS (m/z)	Tentative Identification
11	8.42	$[M + H]^+$	256.2631	256.2635	1.56	$C_{16}H_{33}NO$	200.1912, 130.1108, 116.0957, 102.0818	Palmitic amide
12	8.47	$[M + H]^+$	496.4115	496.4122	1.41	$C_{29}H_{49}N_7$	479.3886, 283.1709	Unknown 6
13	8.49	$[M + H]^+$	452.3881	452.3887	1.33	$C_{25}H_{50}N_5P$	435.3634, 239.1488	Unknown 7
14	8.63	$[M + H]^+$	282.2777	282.2791	4.96	$C_{18}H_{35}NO$	265.2528, 247.2429, 117.1641, 163.1489, 149.1366, 135.1178, 121.1026, 109.1050	Oleamide
15	8.85	$[M + H]^+$	284.2941	284.2948	2.46	$C_{18}H_{37}NO$	248.2492, 164.1548 150.1359, 134.0948, 135.1240, 102.0882	Stearamide
16	9.00	$[M + H]^+$	310.3101	310.3104	0.97	$C_{20}H_{39}NO$	293.2875, 275.2733, 228.2323, 191.1780, 177.1691, 163.1447, 149.1344, 135.1194	Oleoyl ethyl amide
17	9.14	$[M + H]^+$	401.3394	401.3414	4.98	$C_{27}H_{44}O_2$	383.3313, 365.3221, 253.1949, 239.1782, 197.1318, 175.1141, 159.1167, 147.1163, 131.0898	25-Hydroxyvitamin D3

Table 3.2 continue

No.	RT (min)	Ion	Measured Mass (m/z)	Predicted Mass (m/z)	Mass Accuracy (ppm)	Molecular Formula	MS/MS (m/z)	Tentative Identification
18	9.20	$[M + H]^+$	312.3255	312.3261	1.92	$C_{20}H_{41}NO$	102.0929	Icosanamide
19	9.34	$[M + H]^+$	338.3401	338.3417	4.73	$C_{22}H_{43}NO$	321.3151, 303.3047, 256.2526, 191.1790, 177.1651, 163.1488, 149.1323, 135.1162, 121.1026	Docosenamide
20	9.43	$[M + H]^+$	364.3571	364.3574	0.82	$C_{24}H_{45}NO$	282.2738, 247.2470, 196.1713, 135.1169	Unknown 8
21	9.55	$[M + H]^+$	340.3578	340.3574	1.18	$C_{22}H_{45}NO$	322.3267, 304.3187, 164.1541, 102.0782	Docosanamide
22	9.64	$[M + H]^+$	366.3721	366.3730	2.46	$C_{24}H_{47}NO$	349.3453, 331.3392, 284.2955, 163.1587, 149.1313, 135.1178, 121.1013	Tetracosenamide
23	9.69	$[M + H]^+$	391.2827	391.2843	4.09	$C_{24}H_{38}O_4$	279.1580, 149.0245, 167.0339	Unknown 9
24	10.08	$[M + H]^+$	396.4189	396.4200	2.77	$C_{26}H_{53}NO$	256.2624, 186.6763	Unknown 10

Table 3.2 continue

No.	RT (min)	Ion	Measured Mass (m/z)	Predicted Mass (m/z)	Mass Accuracy (ppm)	Molecular Formula	MS/MS (m/z)	Tentative Identification
25	10.15	$[M + H]^+$	385.3452	385.3465	3.37	$C_{27}H_{44}O$	367.3341, 255.2066, 241.1975, 227.1807, 213.1659, 201.1624, 171.1172,145.1041	7-Dehydrocholesterol
26	10.78	$[M + H]^+$	663.4498	663.4480	2.71	$C_{35}H_{61}N_5O_5P$	607.3903, 551.3273, 495.2654, 439.2034	Unknown 11

Negative mode

No.	RT (min)	Ion	Measured Mass (m/z)	Predicted Mass (m/z)	Mass Accuracy (ppm)	Molecular Formula	MS/MS (m/z)	Tentative Identification
27	0.73	$[M - H]^-$	520.9037	520.9042	0.96	$C_{14}H_{11}N_2O_{10}P_5$	384.9305, 361.8709, 248.9557, 204.9667	Unknown 12
28	6.80	$[M - H]^-$	311.1652	311.1653	0.32	$C_{20}H_{24}O_3$	183.0094, 119.0491	Unknown 13
29	7.04	$[M - H]^-$	325.1819	325.1809	3.08	$C_{21}H_{26}O_3$	183.0092, 119.0502	Unknown 14
30	7.28	$[M - H]^-$	339.1977	339.1966	3.24	$C_{22}H_{28}O_3$	183.0091, 119.0496	Unknown 15
31	8.11	$[M - H]^-$	465.2965	465.2974	1.93	$C_{27}H_{49}P_3$	327.9113	Unknown 16
32	8.60	$[M - H]^-$	248.9590	248.9592	0.80	$C_4H_{14}O_2P_4S$	168.8926	Unknown 17
33	8.75	$[M - H]^-$	384.9333	384.9331	0.52	$C_6H_{15}O_9P_5$	248.9592, 154.9748	Unknown 18
34	8.75	$[M - H]^-$	378.9149	378.9148	0.26	$C_7H_5N_6O_5PS_3$	296.6843, 190.9479, 174.9589	Unknown 19
35	8.81	$[M - H]^-$	666.0552	666.0551	0.15	$C_{21}H_{34}N_7O_2P_5S_3$	645.1210, 205.1309	Unknown 20

To date, there are only a limited number of metabolites with known MS/MS fragmentation. Thus, in order to improve the identification rate of the metabolites, identification of the metabolites was also performed by matching the molecular structure with the experimental MS/MS fragmentation. An example is the metabolite with a precursor ion at m/z 226.0688. As expected, the search for this empirical formula of m/z 226.0688 produced several potential hits. The experimental fragmentation pattern was then utilized to narrow the search field and it affirmed that the identity of the metabolite is 3-phenyl-5-ureido-1,2,4-triazole. This is because the fragment ion with m/z 167.0255 corresponded with the loss of urea moiety from the molecular ion while the fragment ion with m/z 148.0384 was as a result of the loss of phenyl moiety (Figure 19). Using this prediction approach, metabolites such as heptadecasphinganine, palmitoleamide, linoleamide, stearamide, oleoyl ethyl amide, icosanamide and tetracosenamide were identified (Table 8).

The use of LC/MS for metabolite profiling identified a number of metabolites belonging to chemical classes like the fatty acid amide, vitamin D metabolite and lipid. These metabolites were reported to be found in EBN for the first time. The identification of these metabolites opens up the possibility of new explanations to the medicinal properties of EBN. Oleamide, a metabolite, was found to be able to induce sleep, increase food intake and prevent inflammation of microglial cells [170, 171] Thus, oleamide is likely to be responsible for several medicinal properties of EBN, such as improvements in concentration, body energy and the immune system. The metabolite, doceosenamide, was also detected in EBN. Studies have shown that minute

amounts of doceosenamide promotes angiogenesis in skeletal muscle and mouse dorsal air-sac.[172, 173] Thus, the presence of doceosenamide is suggested to be associated with the health benefits of EBNs like an increase in the rate of post-surgery recovery and the nourishment of lungs. Research on EBN has proven that its extract leads to an increase in bone strength and the calcium concentration in bones.[46] This observation could be attributed to the presence of 25-hydroxyvitamin D3 in EBNs as it is essential in calcium intake, leading to healthier bones.[174]
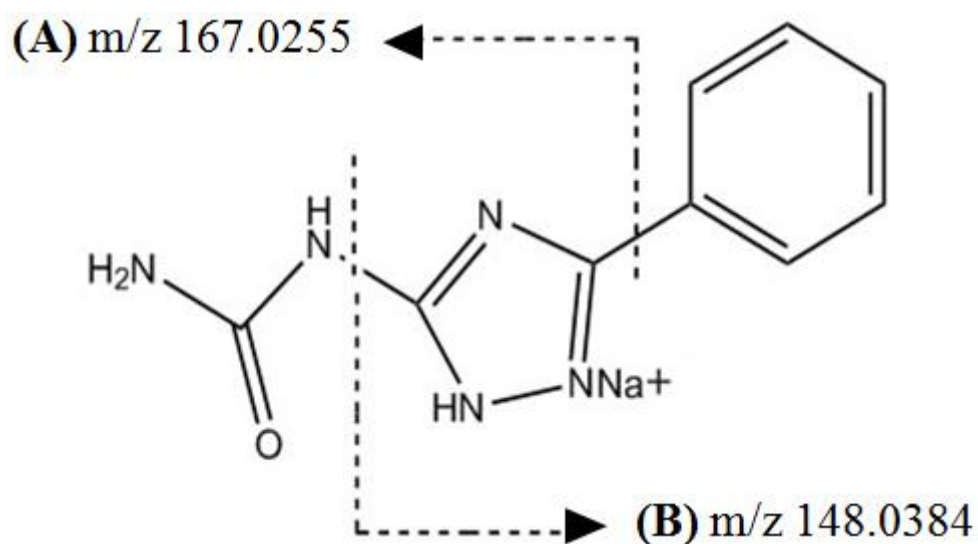
Figure 19 Proposed MS/MS fragmentation of the sodium adduct molecule of 3-Phenyl-5-ureido-1,2,4-triazole ([M+H]$^+$, m/z 226.0688). (A) is the fragment of m/z 167.0255 while (B) is the fragment of m/z 148.0384.

3.3.3. Classification of EBNs with PCA

The GC/MS and LC/MS chromatograms representing the EBNs of different colors, countries and production sites are illustrated in Appendix 7 and 8. Visual inspection of the GC/MS chromatograms reveals the presence of minor differences when different groups of EBNs are compared. These differences were observed mainly between the RT 25.00 – 35.00 min in the GC/MS chromatogram (Appendix 9). Similarly, variations were also observed in the LC/MS chromatograms, within the region of RT 7.00 – 9.00 min (Appendix 10). This provides the confidence that it is viable to classify EBNs.

PCA was performed to reduce the detected metabolites in individual GC/MS or LC/MS chromatogram into a single point known as score on a score plot. The PCA score plots for the different types of classification based on the GC/MS and LC/MS data are displayed in Figure 20. All constructed PCA score plots depend on the optimization of the number of PCs to better represent the score plot to strike an optimal balance between the values of R^2 and Q^2. This resulted in the number of PCs to classify EBN based on color, country and production site being determined to be 12, 6 and 6 respectively for the GC/MS score plots while the number of PCs deduced for the LC/MS data was 6, 8 and 8 respectively. The PCA score plots representing the GC/MS and LC/MS analysis illustrate a poor classification pattern for all the different types of classification (Figure 20). In addition, the goodness of prediction for the respective score plots (Q2X (cumulative)$_{GC/MS, color}$ = 0.263, Q2X (cumulative)$_{GC/MS, country}$ = 0.305, Q2X (cumulative)$_{GC/MS, production site}$ = 0.305,

Q2X (cumulative)$_{\text{LC/MS, color}}$ = 0.377, Q2X (cumulative)$_{\text{LC/MS, country}}$ = 0.461 and Q2X (cumulative)$_{\text{LC/MS, production site}}$ = 0.461) is much lesser than the minimum acceptable value of 0.5, thereby indicating a poor prediction of the X data through the 7-fold cross validation procedure. This poor display of classification is due to the minute differences between the different types of EBNs. In order to classify EBNs, the supervised multivariate method known as OPLS-DA was utilized as it is known to maximize the separation between the different types of EBN. Moreover, OPLS-DA was selected due to its advantage of being able to predict the identity of unknown samples.
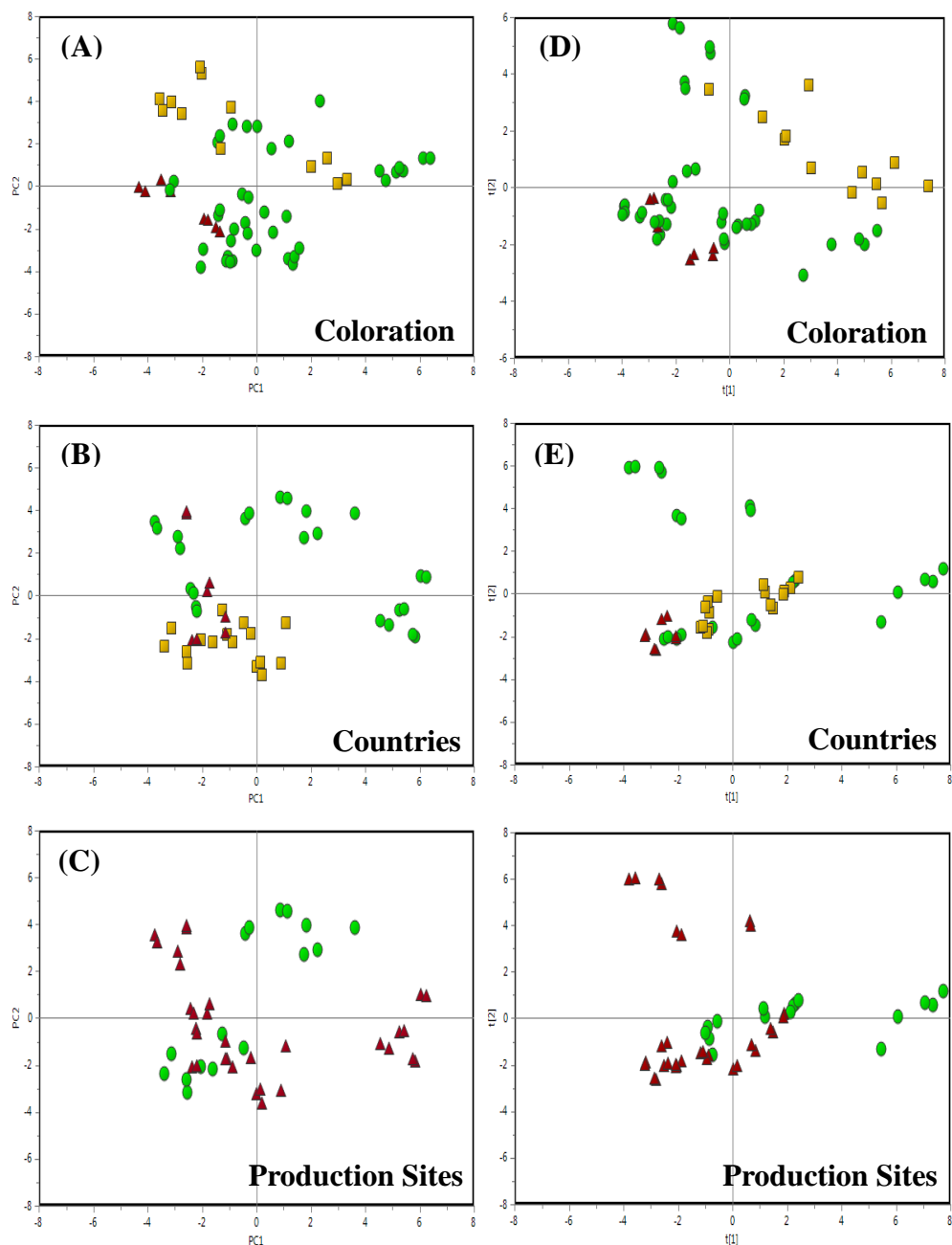
Figure 20 PCA score plots to classify EBNs. Score plots (A) to (C) are based on GC/MS data while score plots (D) to (F) are based on LC/MS data. (A) is the classification according to coloration. White EBN (⬤); Orange EBN (□); Red EBN (▲). (B) is the classification according to countries. Malaysian EBN (⬤); Indonesian EBN (□); Thai EBN (▲). (C) is the classification according to production sites. Farm EBN (⬤); Cave EBN (▲). (D), (E), (F) has the type of classification and representation symbols as (A), (B) and (C) respectively.

3.3.4. Classification of EBN based on color

EBN is a food item which comes in three different colors namely white, orange and red.[31] It has been noted that red EBN often commands a higher price compared to white EBNs as the Chinese claims that the red EBNs possess higher nutritional value.[34] However, these claims have not been supported by scientific evidences. Thus, in this chapter, classification of EBNs was conducted according to their coloration to determine whether significant difference can be found in their metabolite composition.

OPLS-DA score plots for both the GC/MS and LC/MS data (Figure 21A and B) obtained a relatively high value for their good goodness of fit and prediction (R2X (cumulative)$_{GC/MS, color}$ = 0.740, R^2Y (cumulative)$_{GC/MS, color}$ = 0.967, Q2X (cumulative)$_{GC/MS, color}$ = 0.868 ; R2X (cumulative)$_{LC/MS, color}$ = 0.714, R^2Y (cumulative)$_{LC/MS, color}$ = 0.894, Q2X (cumulative)$_{LC/MS, color}$ = 0.819). Examination of the OPLS-DA score plots indicated three distinctive clusters, representing the white, orange and red EBNs, thus demonstrating the feasibility of using OPLS-DA for the classification of EBNs according to their natural coloration. The scores representing each coloration group were observed to be in a tight cluster, indicating the high degree of similarity within the EBNs for each coloration group.

In this study, a loading plot was utilized to determine the metabolites responsible for the observations in the OPLS-DA score plots by superimposing the metabolites onto the score plot to relate the metabolites to

the scores. Based on the loading plot for GC/MS data (Figure 21C), the metabolites responsible for the differentiation were identified to be 2-acetamido-2-deoxy-β-D-glucopyranose and 2-acetamido-2-deoxy-α-D-glucopyranose (compound 17 and 20). The detection of 2-acetamido-2-deoxy-β-D-glucopyranose and 2-acetamido-2-deoxy-α-D-glucopyranose in the metabolite composition was a surprise as it was thought only to be present in the glycans of EBN.[109-112] Research studies have shown that the coloration change from white to red is due to a nitrile like vapor originating from the soil at the bird nest production site.[34] As such, it is proposed that the nitrile like vapor had dissolved in water, forming an acid which triggered the hydrolysis of glycans to release the metabolites. The presence of glucopyranoses may support the popular claim of EBNs beautifying properties as a research study has found that oral consumption of the glucopyranoses reduced the appearance of facial hyperpigmentation.[175]

On the other hand, examination of the loading plot based on LC/MS data (Figure 21D) reveals that 3-phenyl-5-ureido-1,2,4-triazole (compound 1) was responsible for the classification pattern. At present, the biological function of this metabolite has not been determined. Nonetheless, it is proposed that the metabolite may possess fungicidal ability similarity to its analogs.[176] However, further studies are required for verification.
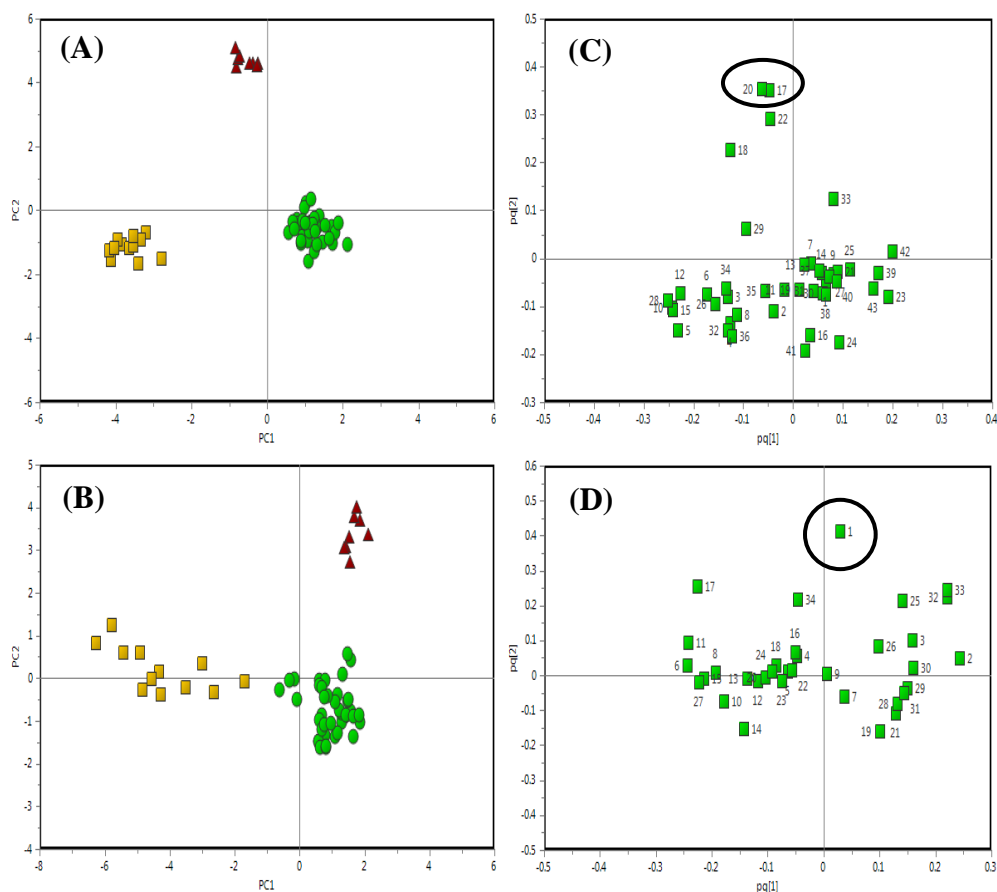
Figure 21 OPLS-DA score plot constructed for the classification of EBNs according to coloration. (A) and (B) represents the OPLS-DA score plots for the GC/MS data and LC/MS respectively. White EBN (●); Orange EBN (■); Red EBN (▲).The loading plots for the GC/MS data and LC/MS data to classify the EBNs according to their coloration are (C) and (D) respectively.

Referring to column plots (Figure 22), all three metabolites - 2-acetamido-2-deoxy-β-D-glucopyranose and 2-acetamido-2-deoxy-α-D-glucopyranose and 3-phenyl-5-ureido-1,2,4-triazole, share a common increasing trend starting from white to orange and to red, thereby implying that the percentage of the metabolite increases as the coloration of the EBN darkens. Although there was a significant difference in the percentage of the metabolites ($p < 0.05$),

overlapping was observed in the white EBN and orange EBN. This is because the coloration of the white and orange EBN was not uniform, meaning that the white EBN has orange spots on it or the orange EBN is of a lighter shade.
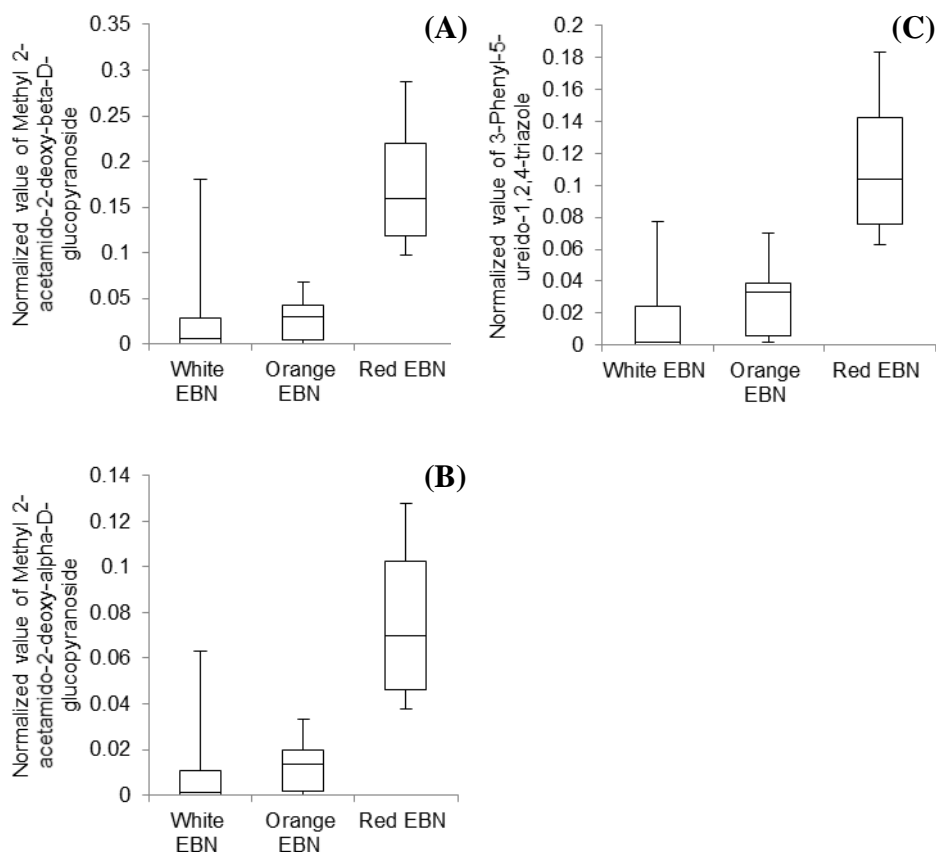
Figure 22 Box and whisker plot of (A) 2-acetamido-2-deoxy-β-D-glucopyranose, (B) 2-acetamido-2-deoxy-α-D-glucopyranose (C) 3-phenyl-5-ureido-1,2,4-triazole normalized area in EBN.

Validation of the OPLS-DA score plots constructed from both data sets was conducted with prediction set A and B. An excellent validation result of 100 % was achieved with prediction set A and B for both OPLS-DA score plots (Table 9). This underlines an excellent experimental reproducibility could be achieved with GC/MS and LC/MS data as well as the different EBN

processing methods do not have a significant impact on the metabolite profile of the EBNs with different colorations.

3.3.5. Classification of EBN based on country

EBNs are produced by the swiftlets in Southeast Asia countries like Malaysia, Indonesia and Thailand. Although these countries are subjected to similar weather conditions, the preservation of flora and fauna of these countries varies from one another due to the difference of the economic development in each country. Studies have shown that variations in environments would make an impact on metabolite composition of food products such as grape berry and ginseng.[177, 178] Thus, it is viable to make use of metabolite profiling to study the impact of the environment on EBN.

In this chapter, 48 EBNs were used to construct the OPLS-DA score plots (Figure 23A and B) for classification according to countries of origin, which are fewer as compared to the 60 EBNs used to construct the OPLS-DA score plot for the differentiation of different colored EBNs in Figure 22. This is because some of the EBNs whose country of origin was unknown were excluded from the score plot. It was evident from the OPLS-DA score plot based on GC/MS data (Figure 23A) that a clear discrimination of the EBNs was observed when they were classified according to their countries of origin ($R2X$ (cumulative)$_{GC/MS, country}$ = 0.718, R^2Y (cumulative)$_{GC/MS, country}$ = 0.923, $Q2X$ (cumulative)$_{GC/MS, country}$ = 0.721). Three distinct groups were also observed when an OPLS-DA score plot based on LC/MS data was constructed

(R2X (cumulative)$_{LC/MS, country}$ = 0.878, R^2Y (cumulative)$_{LC/MS, country}$ = 0.907, Q2X (cumulative)$_{LC/MS, country}$ = 0.762) (Figure 23B). The strong classification pattern in both OPLS-DA score plots suggests that significant differences existed in the metabolic profile of EBNs with regard to the country from where they were harvested.

Loading plots of the OPLS-DA score plots in Figure23A and B were displayed in Figure 23C and D respectively. In Figure 23C, the level of the majority of the fatty acids (17 out of 19 fatty acids) was shown to be elevated in Malaysian and Indonesian EBNs, which resulted in the differentiation of the Malaysian and Indonesian EBNs from the Thai EBNs along the PC2 axis in the OPLS-DA score plot in Figure 23A. This significant variation in the fatty acid content ($p < 0.05$) is most likely due to the environmental conditions. Malaysia, Indonesia and Thailand are agricultural countries, with Malaysia and Indonesia focusing on palm oil cultivation while Thailand concentrates on rice cultivation.[179, 180] Swiftlets are aerial insectivores that would usually hunt in close proximity to their habitat.[181] Thus, Malaysian and Indonesian swiftlets would feed on insects with a diet rich in palm oil whereas Thai swiftlets would prey on insects that have a rice based diet. It is known that palm oil has a higher fatty acid content than rice.[182, 183] Hence, Malaysian and Indonesian swiftlets would produce EBN with enhanced fatty acid contents compared to the Thai swiftlets whose diet consists of insects of lower fatty acid contents.

On the other hand, Malaysian EBNs and Indonesian EBNs were separated along the PC1 axis. This is because of the distribution of fatty acids in the

loading plot – 10 fatty acids in the region of Malaysian EBNs and 7 fatty acids in the region of Indonesian EBNs (Figure 23D). This elevated level of fatty acids in Malaysian EBN was expected because palm oil plantations take up a larger percentage of the land area in Malaysia relatively to Indonesia.[184] As such, the Malaysian swiftlets, which consume insects with higher fatty acid contents, are expected to produce EBNs with elevated fatty acid content.

Unlike the OPLS-DA score plot for GC/MS data, the OPLS-DA score plot for LC/MS shows that the Thai EBNs were separated from Malaysian and Indonesian EBNs along the PC1 axis (Figure 23B). The reason for this separation is because most of the detected fatty acid amides (9 out of 10 fatty acid amides) were distributed in Malaysian and Indonesian EBNs (Figure 23D). Malaysian EBNs and Indonesian EBNs were also separated along the PC1 axis but the distance between these two groups was relatively closer as compared to the Thai EBNs. The separation of Malaysian EBNs and Indonesian EBNs is due to 8 fatty acid amides being located in the region of the Malaysian EBNs while 2 fatty acid amides are situated in the region of the Indonesian EBNs. The result of the LC/MS data is in line with the GC/MS, where the number of discriminating variables - fatty acids and fatty acid amides, was largest in Malaysian EBNs followed by Indonesian EBNs and then Thai EBNs. This is because the precursor of a fatty acid amide is its corresponding fatty acid with the same number of carbons.[185] Therefore, based on the two sets of data, it is proposed that the swiftlet's diet would have a significant impact on the discriminating variables in the GC/MS and LC/MS data, for EBNs from different countries.
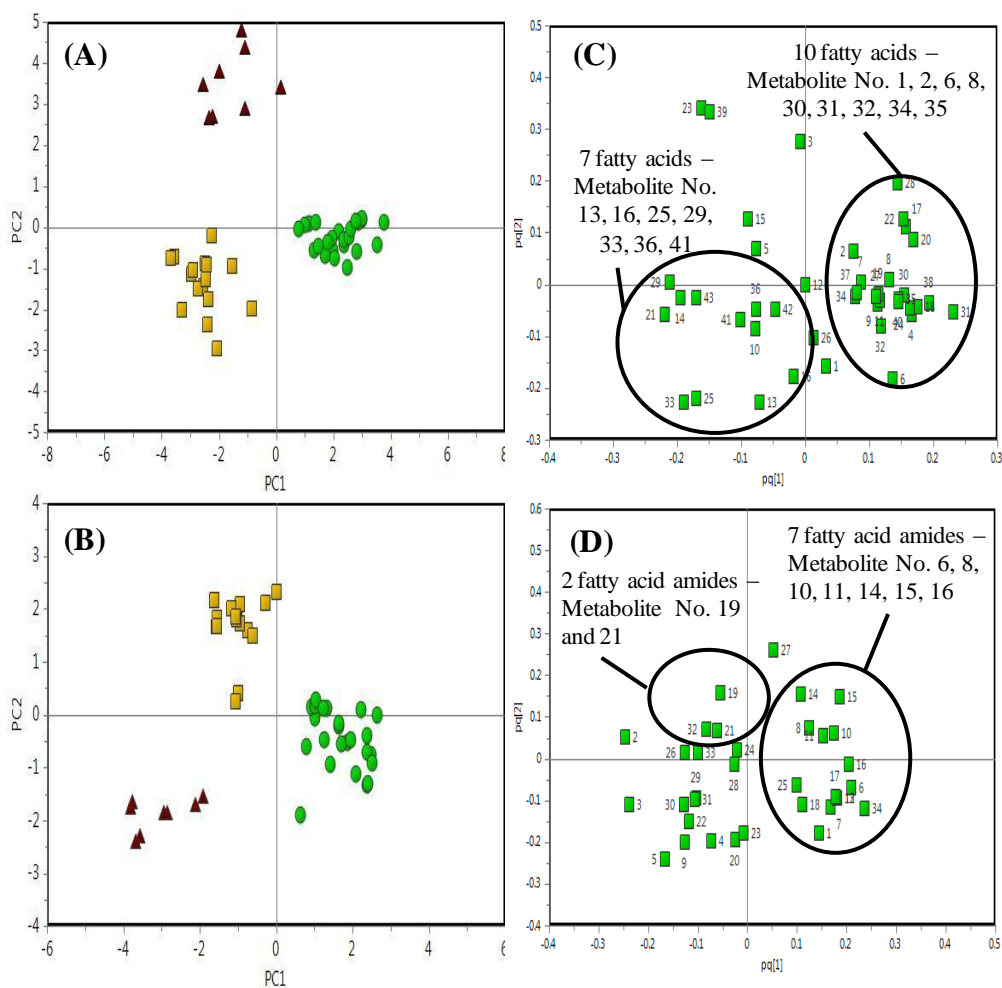
Figure 23 OPLS-DA score plot constructed for the classification of EBNs according to countries. (A) and (B) represents the OPLS-DA score plots for the GC/MS data and LC/MS respectively. Malaysian EBN (⬤); Indonesian EBN (⬛); Thai EBN (▲). The loading plots for the GC/MS data and LC/MS data to classify the EBNs according to their countries are (C) and (D) respectively.

Validation of the OPLS-DA score plots from the GC/MS and LC/MS data was performed to verify the classification results (Table 9). The validation result for the OPLS-DA score plot from GC/MS shows that all the EBNs from prediction set A and B were accurately classified according to their respective

country. In contrast, 50.0 % of prediction set A and 11.1 % of prediction set B were correctly predicted by the OPLS-DA score plot from LC/MS data. The poor classification rate of prediction set A could be attributed to the fatty acid amides as they were responsible for the classification pattern of EBNs. In the previous paragraph, it was demonstrated that swiftlets having a similar diet would produce a similar level of fatty acid amides but the validation result contradicts this. A possible attribution factor could be the variation in the biosynthetic pathway of the fatty acid amides in the swiftlets. Currently, there is no information regarding the synthesis of fatty acid amides in the swiftlets, thus future experiments could be conducted for verification. The validation results from prediction set B also indicated that the EBN processing method would significantly affect the LC/MS data.

While it is observed in this chapter that fatty acids are responsible for the differentiation of EBNs from different countries, Huda et al. reported no difference in their fatty acid content.[103] The discrepancy explained maybe by the use to of pre-processed EBNs in Huda et al. study, such EBNs could have contained the remains of insects or the swiftlets' droppings and thus affecting the fatty acid analysis. Besides that, a relatively long analysis time was utilized in this chapter for the detection fatty acids. This permitted the detection of more discriminating variables, thereby improving the rate of classification as compared to Huda et al. study.

3.3.6. Classification of EBN based on production site

Traditionally, EBNs have been harvested from natural lime stone caves located near rainforests or along the coastlines.[186] However, with the rapid increase in the demand of EBN, the traditional source of EBN is unable to meet the needs of the market. Thus, in the recent years, an alternative source for EBN known as EBN farming has been developed. EBN farming is a farming technique which involves the use of man-made buildings to attract swiftlets to inhabit and build their nests. Today, EBNs purchased in the market are either harvested from their natural habitat or a man-made environment and they are known as cave nests and farm nests respectively. The cave nest is known to command a higher price compared to the farm nest due to the risk involved during the collection of the cave nest.[31] However, the EBNs from caves or farms do not have any visible or physical differences, thus making them indistinguishable. Hence, in this chapter, classification of the farm nest and cave nest would be performed with metabolite profiling, combining with multivariate analysis to determine whether the difference in production sites would result in the variation of the metabolite composition.

In the classification of EBNs based on production sites, two OPLS-DA score plots were constructed. The model performance parameters for the GC/MS data were R2X (cumulative)$_{\text{GC/MS, production site}}$ = 0.561, R^2Y (cumulative)$_{\text{GC/MS, production site}}$ = 0.948, Q2X (cumulative)$_{\text{GC/MS, production site}}$ = 0.873 while the LC/MS data illustrates that R2X (cumulative)$_{\text{LC/MS, production site}}$ = 0.541, R^2Y (cumulative)$_{\text{LC/MS, production site}}$ = 0.968, Q2X (cumulative)$_{\text{LC/MS, production site}}$ =

0.927. From the examination of the OPLS-DA score plot from the GC/MS data, the farm and cave nests were clearly separated by PC1 axis, with the farm nest located in the left and the cave nest distributed at the right of the score plot (Figure 24A). Similarly, farm nest and cave nest exhibited a similar positioning in the OPLS-DA score plots constructed from the LC/MS data (Figure 24B).
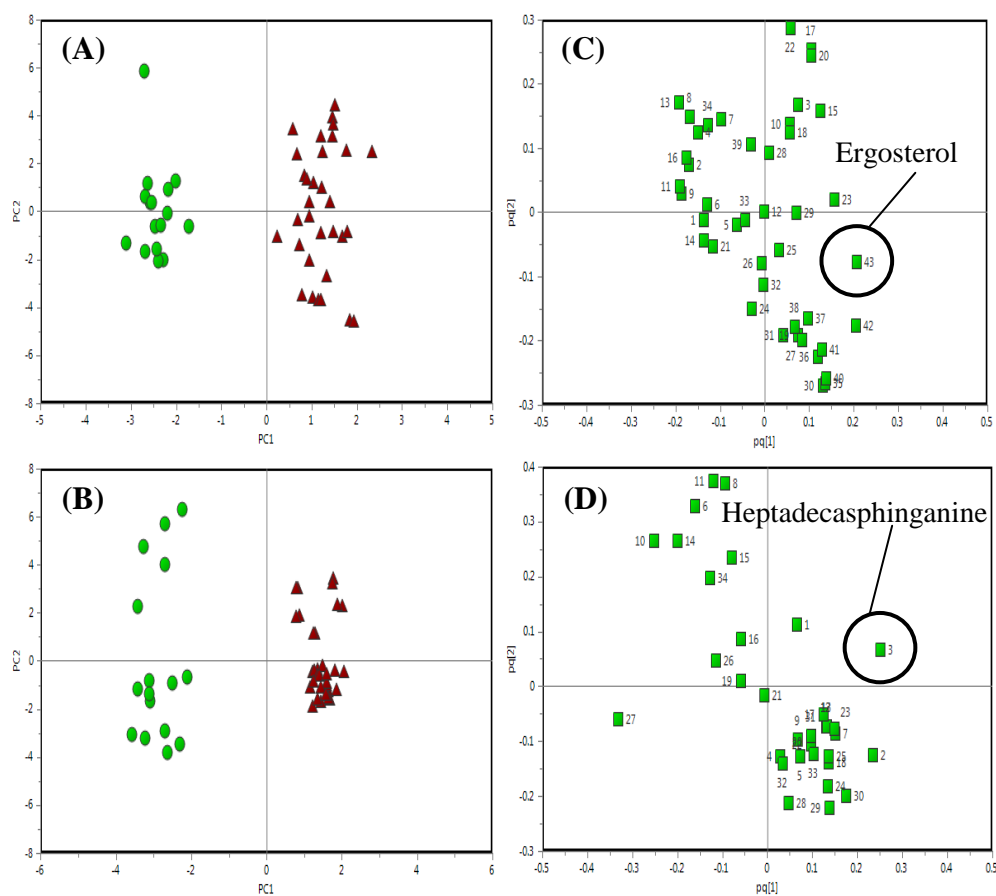
Figure 24 OPLS-DA score plot constructed for the classification of EBNs according to production sites. (A) and (B) represents the OPLS-DA score plots for the GC/MS data and LC/MS respectively. Farm EBN (○); Cave EBN (▲). The loading plots for the GC/MS data and LC/MS data to classify the EBNs according to their production sites are (C) and (D) respectively.

Analysis of the loading plot for the GC/MS data (Figure 24C) indicates the difference in cave and farm nests is due to ergosterol (compound 43). With the use of a box and whisker plot, it was further deduced that ergosterol is significantly higher in the cave nest (Figure 25A) Ergosterol is a major sterol component found in fungi and it is either absent or exists in a minute quantity in higher plants, therefore its detection in EBN can only be attributed exclusively to the presence of fungi.[187] The presence of fungi in the cave nest is proposed to be due to the soil in the cave being fertilized by the swiftlet's droppings, thereby creating a nutrition rich environment for the growth of the fungi. On the other hand, information obtained through personal observation of the EBN farms in Malaysia confirms that farm nests are harvested from an environment made up of only concrete and wood. Furthermore, for hygiene purposes, the swiftlets' droppings are cleared away on a regular basis. Hence, the environment for the farm nest is not ideal for the growth of the fungi and thus accounting for the absence of ergosterol.

With reference to the loading plot for the LC/MS data (Figure 24D), it was observed that heptadecasphinganine (compound 3) was responsible for the difference in farm and cave nests. Heptadecasphinganine has demonstrated the ability of inhibiting the growth of fungi.[188] Thus, the production of heptadecasphinganine in EBN is suggested to prevent fungi from affecting the health of the swiftlets and their hatchings. The significant higher percentage of heptadecasphinganine in the cave nest points out that the swiftlets would increase the production of heptadecasphinganine in EBNs in order to counter the thriving fungal population in the cave (Figure 25B).
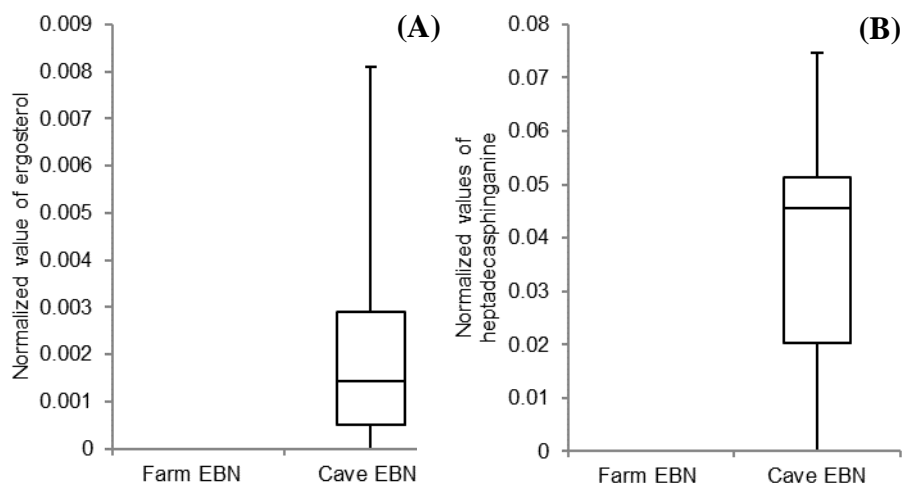
Figure 25 Box and whisker plot of (A) ergosterol and (B) heptadecasphinganine normalized area in EBN.

The validation result for the OPLS-DA score plot from GC/MS data illustrates that all the EBN samples in both prediction set A and B were correctly predicted. In contrast, the OPLS-DA score plot from LC/MS data shows that a perfect validation result of 100 % was achieved with prediction set A while only 76 % of the samples from prediction set B were classified correctly (Table 9). This suggests that processing methods had a significant impact on the LC/MS data but not on the GC/MS data.

Table 9 Validation results of the OPLS-DA score plots based on GC/MS data and LC/MS data.

Type of classification	GC/MS				LC/MS			
	Prediction set A		Prediction set B		Prediction set A		Prediction set B	
	Number of analysis	Prediction ability	Number of analysis	Prediction ability (%)	Number of analysis	Prediction ability (%)	Number of analysis	Prediction ability (%)
Color	40	100	50	100	40	100	50	100
Country	28	100	36	100	28	50	36	11.1
Production Site	28	100	50	100	28	100	50	76

Comparing the two analytical approaches for the classification of EBNs, GC/MS analysis was able to detect a wider range of metabolites such as organic acids, fatty acids, sugar and sterols compared to LC/MS analysis, which could only detect fatty acid amides, lipids and vitamin D metabolites. In addition, GC/MS based score plots possess a better prediction ability than the LC/MS ones. However, GC/MS requires additional steps in the sample preparation like the derivatization of the metabolites. Moreover, a longer analysis time was utilized in GC/MS to resolve the peaks.

3.4. Conclusion

In this chapter, a multi analytical platform, incorporating GC/MS and LC/MS was shown to be able to provide a comprehensive coverage of the metabolites in EBN and also identify novel components in the food item. The identification of metabolites provided logical explanations to some of the claimed medicinal effects of EBN. The metabolite profiles from GC/MS and LC/MS in combination with multivariate analysis, in particular OPLS-DA models, were demonstrated to be successful in classifying EBNs according to their coloration, countries and production sites. Multivariate also aided in the identification of the metabolites that were responsible for the differences in the EBNs. These metabolites pointed out that the environment in which the swiftlets inhabit played a vital role in all three types of classification of EBNs. Although both GC/MS and LC/MS are insightful approaches to understanding and classifying EBNs, validation results suggest that GC/MS produces better OPLD-DA models than LC/MS as the variation of the EBN processing

methods would not cause a significant variation in the metabolite composition analysed by GC/MS. The approach of metabolite profiling had shedded some light on the components of EBNs and its success in classifying EBNs also represents a paradigm shift in the quality control of this food item.

CHAPTER 4. DISCRIMINATION OF EDIBLE BIRD'S NEST WITH DIFFERENT ANALTYICAL METHODS AND MUTIVARIATE ANALYSIS

4.1. Introduction

According to the World Customs Institute, the food fraud industry has risen to about US$49 billion a year and it involves everything from fine food to boxed fruit juice.[189]. EBN is a food item which commands a premium price –price of per gram of EBN could be as high as USD$15 per gram (as of February 2014), thus it is not surprising that EBN has become the prime target for adulteration by food suppliers. These unscrupulous suppliers would often substitute genuine EBN with cheaper food items and/or introduce in chemicals to make the final food item more appealing to customers. To avoid falling prey to possible food fraud, consumers would often use their touch, sight and/or smell to assess the food item. However, these methods are subjective and prone to human errors, thus would not be able to serve their purpose. It is vital that the scope of food discrimination has to go beyond physical inspection and to take into account evidences from independent analytical methods to ensure the integrity of the food item.

Authenticity of EBN has been one of the main focuses on EBN in the past few years. In these studies, various analytical methods have been employed and they have exhibited different degrees of success in the discrimination of the food item. These analytical methods can be broadly divided into three main

groups – morphological method, molecular biological method and chemical analytical method.

Morphological method involves the use of a microscope to magnify and study the surface features of the sample in detail. It has been frequently used in the authentication of food items as it possesses advantages such as minimal sample preparation, short analysis time and non-destructive analysis. Marcone and Cheung's research group found that the distinct morphological differences could be observed between EBN and the common adulterants of the food item.[114, 156] However, this method could not distinguish genuine EBN from those which are adulterated, thus rendering it to be an ineffective tool to tackle fraud cases involving EBN.

Molecular biological methods such as polymerase chain reaction (PCR), electrophoresis and enzyme-linked immunosorbent assay (ELISA) have been reported to be utilized in the authentication of EBN. An example of the application of PCR was noted in Yuan et al. where EBN's deoxyribonucleic acid (DNA) was amplified by PCR so as to distinguish the food item from its adulterated form.[190] ELISA is also another popular method due to its high selectivity of proteins in complex matrices.[191] Recently, Yang et al. demonstrated the potential of ELISA in detecting low levels of EBN in food and cosmetics.[192] However, the success of the molecular biological methods is often restricted by their targeted analyte. In the case of PCR, DNA is known to degrade into fragments with lesser than 200 base pairs when subjected to temperature above 100 °C.[193] Meanwhile, in ELISA, proteins may be

denatured due to changes in the temperature and/or pH induced during food processing.[194]

In terms of chemical analytical methods, it can be further divided into two sub-categories – spectroscopy and chromatography. Hussin et al. made use of this fact to discriminate EBN.[195] Meanwhile, chromatography possesses the capabilities of separating, detecting and quantifying the compounds of interest, which are then expressed in terms of signals as a function of time in a graph. As such, chromatography has been utilized to create fingerprints of EBN and its adulterated form in several reports.[195-198] However, the use of chemical analytical methods carries the problem of insensitivity to low levels of adulterants and the requirement for a large number of genuine samples to establish the chemical fingerprints. In view of this, there is a need for the development of independent analytical techniques to ensure the integrity of EBN from farm to fork.

The aim of this chapter was to demonstrate the potential of different analytical methods - amino acids, monosaccharides and metabolites analysed by GC/MS and LC/MS, combined with multivariate analysis to discriminate EBN. Reports have illustrated that the common adulterants of EBN are agar-agar, white jelly fungus and isinglass.[114] Thus, apart from EBN and its common adulterants, binary mixtures of EBN and adulterant (known as spiked sample) would be included in the construction of multivariate models.[199] In this chapter, the discrimination of EBN and spiked samples was conducted in two parts - qualitative and quantitative discriminations. In qualitative discrimination, the

ideal data pre-processing and multivariate approach would be determined so that the best detectable spiking range could be attained. In addition, the prediction capability of the established multivariate model offers the possibility of discriminating EBNs from samples that are spiked with unknown adulterants. Meanwhile, quantitative discrimination serves as a secondary level of discrimination. After determining on the type of adulterants from qualitative discrimination, accurate measurement of the adulterant level could be performed in quantitative discrimination using projection to latent structure (PLS) regression. In this case, PLS regression for each type of adulterants would have to be established and validated prior to its discrimination of EBN. The established approach would then be utilized for the prediction of EBN mixed with more than one adulterant to take into account all possible authentication scenarios. This proposed approach for EBN discrimination will hopefully, not only be able to discriminate EBNs from the stated adulterants, but also function as a database for the discrimination of the food item.

4.2. Materials and methods

4.2.1. Information on the samples

EBNs used in this chapter were kindly sponsored by established EBN retailers in Singapore and Malaysia so as to ensure the authenticity of the EBN. The EBN was processed to remove the dirt and feathers prior to their analysis. EBN was homogenized through blending and stored under 4°C until they were needed for analysis. Three different types of adulterants – agar agar, white

jelly fungus (fungus), isinglass, were utilized in this chapter. These adulterants were mixed with EBN in varying proportions to create the binary mixtures for analysis. These binary mixtures were known as spiked samples and they were labelled with the adulterant name followed by the percentage of adulterant in the overall sample (eg agar agar 5 %). Samples which are mixtures of EBN and more than one adulterant were known as multiple spiked samples. The name of the adulterant was stated first followed by the overall percentage of adulterants in the sample (eg agar agar and fungus 10 %). The percentage of adulterants spiked into the sample was different from the binary mixture. In the case for samples containing two adulterants like agar agar and fungus 10 %, the percentage of agar agar, fungus and EBN in the mixture is 5 %, 5 % and 90 % respectively. However, if there were three adulterants in the sample, the proportion of the mixture changed. An example was agar agar, fungus and isinglass 10 % where the percentage of agar agar, fungus, isinglass and EBN was 3 %, 3 %, 4 % and 90 % respectively.

4.2.2. Chemicals and materials

Amino acid standards and the amino acid derivatization kit (EZ: faast^{TM} GC-FID amino acid analysis kit) were purchased from Phenomenex. Several chemicals were provided in the EZ: faast^{TM} kit. They were internal standard solution (norvaline), sodium carbonate, sodium hydroxide, N-propanol, derivatizing agent (chloroformate-type compound), isooctane and hydrochloric acid solution. The standard used consisted of the following 17 amino acids: alanine, glycine, valine, leucine, isoleucine, threonine, serine,

proline, aspartic acid, methionine, 4-hydroxyproline, glutamic acid, phenylalanine, lysine, hydroxylysine, tyrosine and cysteine. The amino acid standards were stored in a freezer while chemicals in the kit were stored in a refrigerator at 4 °C prior to their usage.

The chemicals - PMP, trifluoroacetic acid, acetonitrile and formic acid were obtained from Aldrich (St.Louis, Missouri, America). Acetic acid and toluene were obtained from Merck, Darmstadi, Germany while hydrochloric acid was purchased from Normapur. Ammonia hydroxide was purchased from Panreac, Barcelona. Ammonia acetate (99.7 % purity) was obtained from Hayashi Pure Chemical Industrial Company, Osaka, Japan. Solvents such as chloroform and methanol were obtained from Kanto Chemical Co., Inc. 2-8 Nihonbashi Honcho 3-Chrome Chuo-Ku, Tokyo, Japan. The derivatizing agent BSTFA with 1 % trimethylchlorosilane was purchased from Restek (Benner Circle, Pennsylvania, America).   Standard compounds such mannose, glucose, galactose, rhamnose, fucose, ribose and xylose were purchased from Sigma-Aldrich Inc. (St.Louis, Missouri, America) while propanetriol was purchased from Merck (Germany). The 7 monosaccharides (mannose, glucose, galactose, rhamnose, fucose, ribose and xylose) used in the working standards were purchased from Sigma Aldrich (St. Louis, MO, U.S.A.). The internal standard for the monosaccharide analysis (D-glucose-$^{13}C_6$) was purchased from Cambridge Isotope Laboratories, America.

4.2.3. Analytical methods

Four analytical methods were utilized to analyse all the samples mentioned earlier. These analytical methods were known as amino acid analysis, monosaccharide analysis, metabolite fingerprinting with GC/MS and metabolite fingerprinting with LC/MS. The procedures for amino acid analysis and monosaccharide analysis have been mentioned in detail in Chapter 2, sections 2.2.3 and 2.2.4. Meanwhile, the procedures for metabolite fingerprinting with GC/MS and metabolite fingerprinting with LC/MS has been stated in Chapter 3, sections 3.2.2 and 3.2.3. These analytical methods were further applied in this chapter to differentiate the genuine EBN from the spiked samples.

4.2.4. Statistical analysis

Data obtained from the different analytical methods were be exported to SIMCAP+ software package version 13 (Umetrics, Umea, Sweden) for the construction of their respective multivariate models. The confidence interval used for the models are 95 %.

4.3. Results and discussion

4.3.1. Determination of normalization approach for the different analye

Three different types of food items, namely agar agar, fungus and isinglass were utilized to create the spiked samples (Figure 26). These food items were chosen as adulterants because they had been reported to have impersonated EBN in the market, and more importantly they possess several common physical properties like EBN. Agar agar and fungus are food items from *Gelidium amansii* (algae) and *Tremella fuciformis* (snow fungus) respectively. They possess an exterior with a uniform white coloration and a gelatinous texture after being boiled.[200, 201] On the other hand, isinglass is a food item that is generally derived from the swim bladders of *Muraenesox cinereus* (eel).[202] It is tasteless when cooked and has a yellowish surface. Based on their physical characteristics, these adulterants bear a striking similarity to the authentic EBN in terms of their appearance and taste, which in turn makes it challenging to distinguish the authentic EBNs from the adulterated ones via human inspections
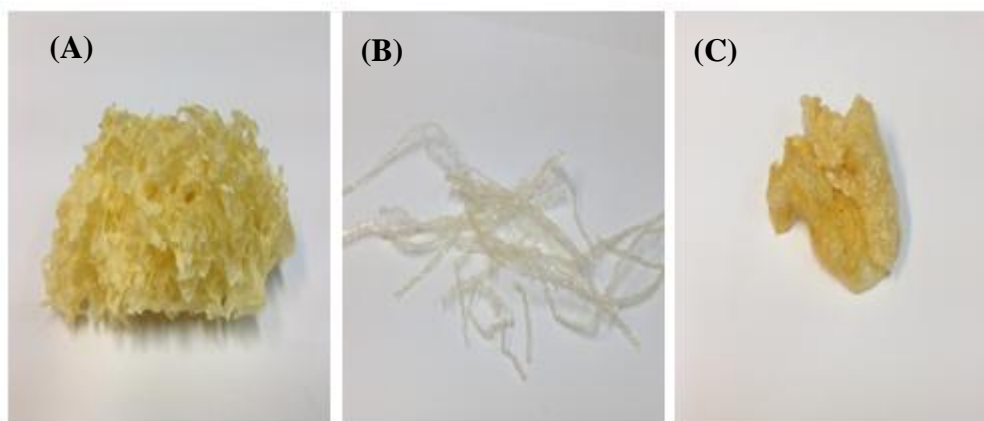
Figure 26 Display of (A) agar agar, (B) fungus and (C) isinglass.

Despite the comparable physical characteristic between the EBN and the adulterants, the adulterants are likely to exhibit a different chemical profile due to the fact that they originated from different species. Representative chromatograms obtained from EBN and the spiked samples are illustrated in Figure 27 to 31. Looking at the chromatograms, a considerable variation in the amount of amino acids and monosaccharides in the contents of EBN and its spiked samples could be seen (Figure 27 to 28). In the case of the metabolite chromatograms from GC/MS and LC/MS, additional differences were observed. Besides changes in the content of the EBN's metabolites, the spiked samples were observed to contain more compounds as the adulteration level increased (Figure 29 to 31). The distinctive chromatograms generated by the EBN and the spiked samples show that it is promising to utilize their chemical profile for the authentication of EBN (Appendix 11 - 14).
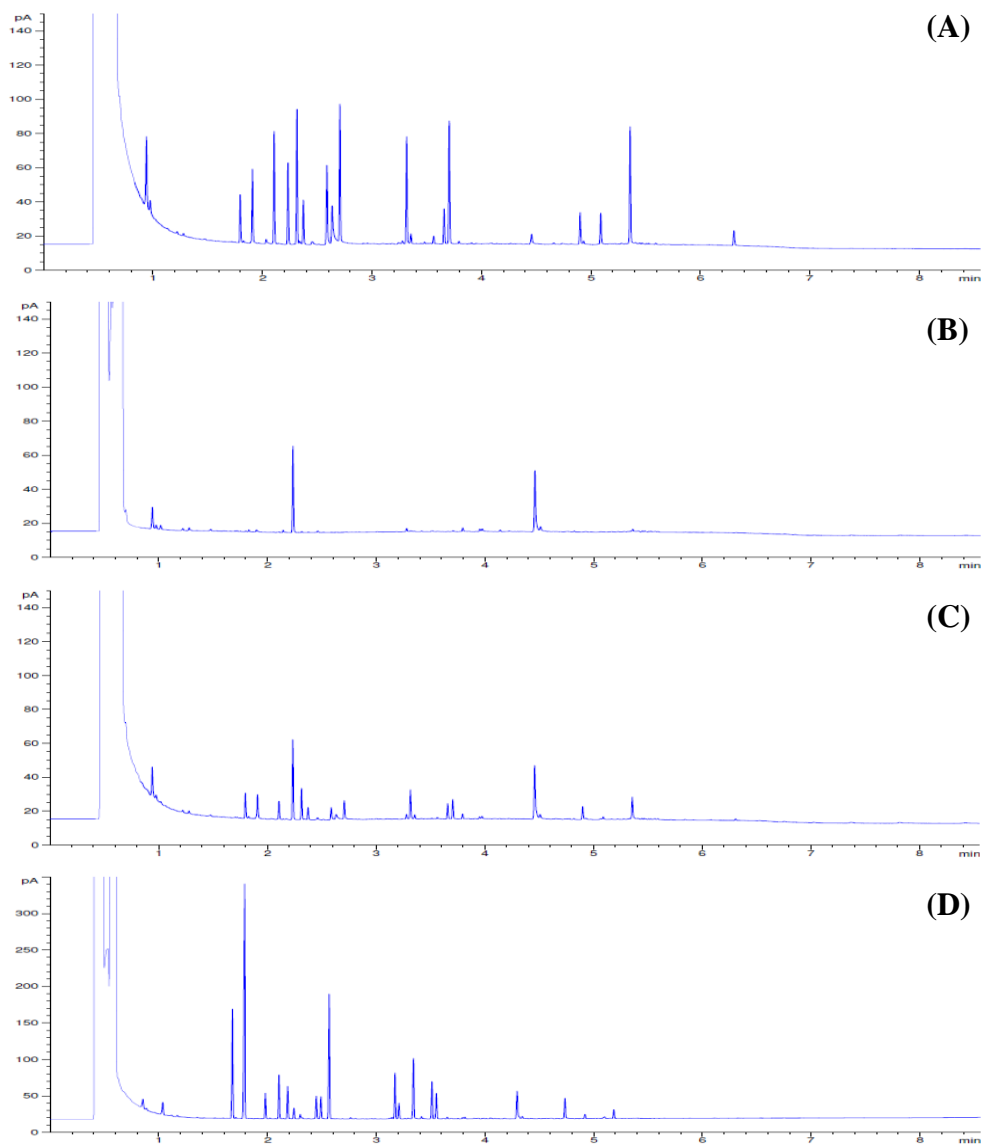
Figure 27 Chromatograms of EBN and its spiked samples based on amino acid analysis. Chromatogram (A) EBN while chromatogram (B), (C) and (D) samples spiked with agar agar, fungus and isinglass respectively.
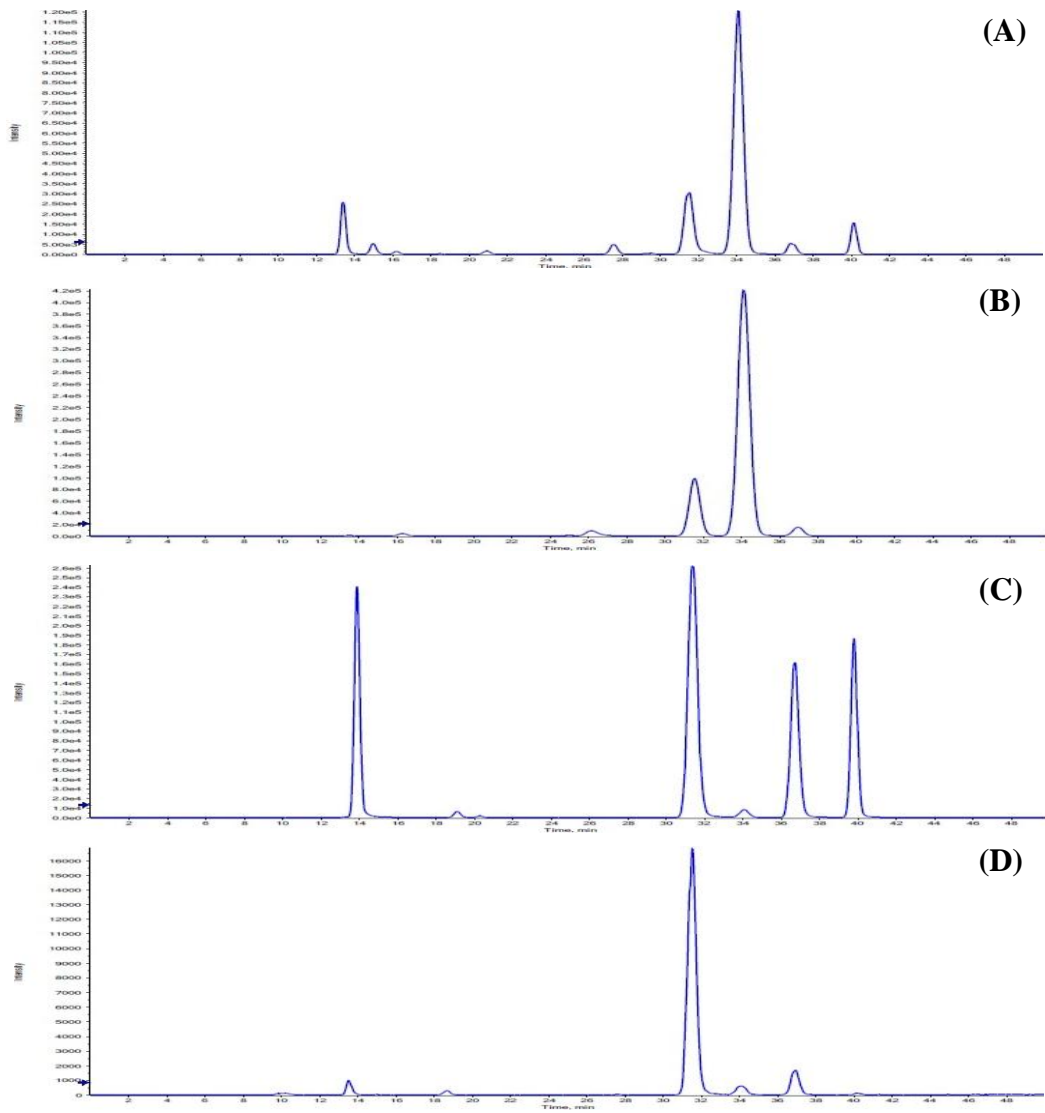
Figure 28 Total ion chromatograms of EBN and its spiked samples based on monosaccharide analysis. Chromatogram (A) EBN while chromatogram (B), (C) and (D) samples spiked with agar agar, fungus and isinglass respectively.
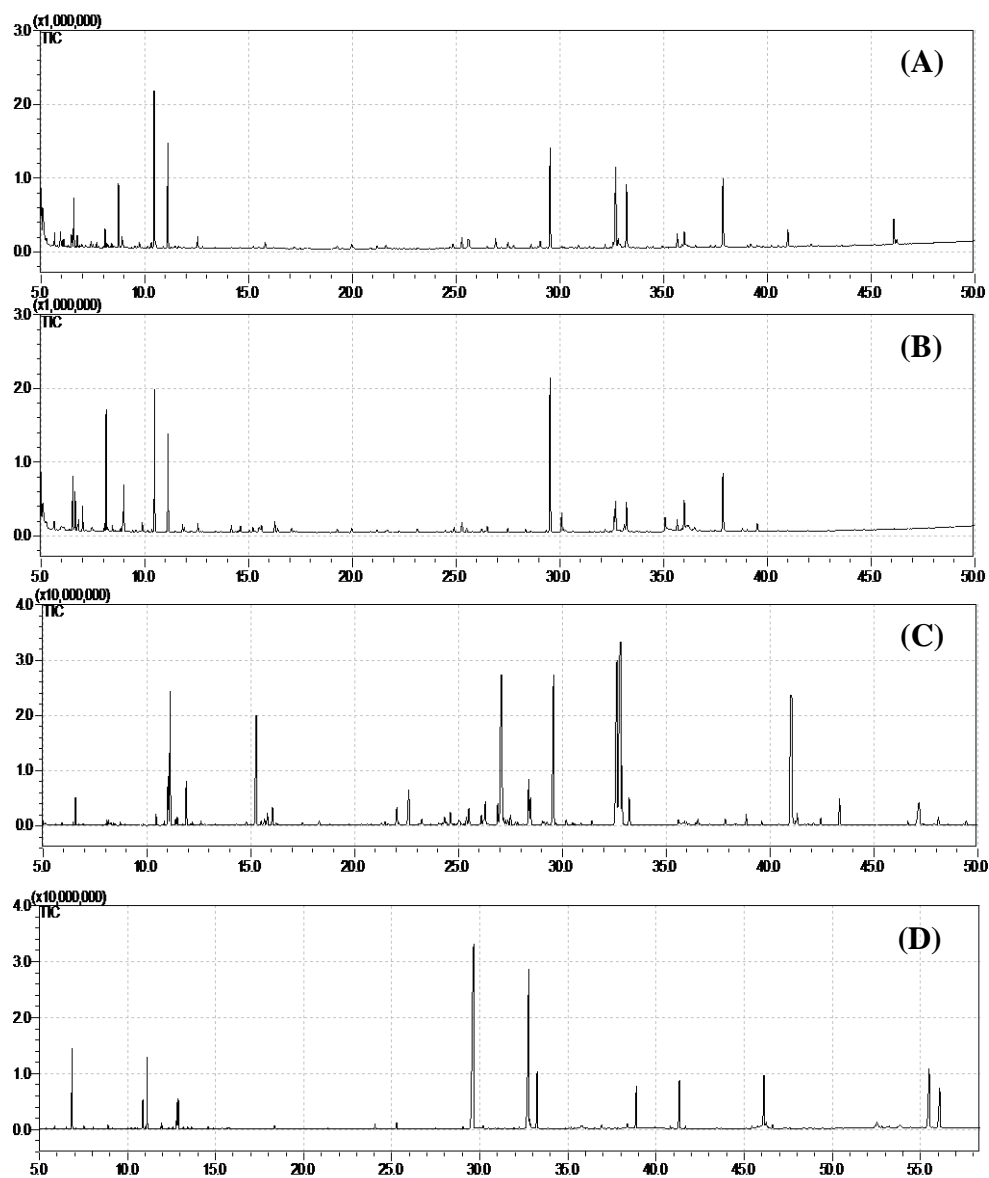
Figure 29 Total ion chromatograms of EBN and its spiked samples based on metabolite fingerprinting with GC/MS. Chromatogram (A) EBN while chromatogram (B), (C) and (D) samples spiked with agar agar, fungus and isinglass respectively.
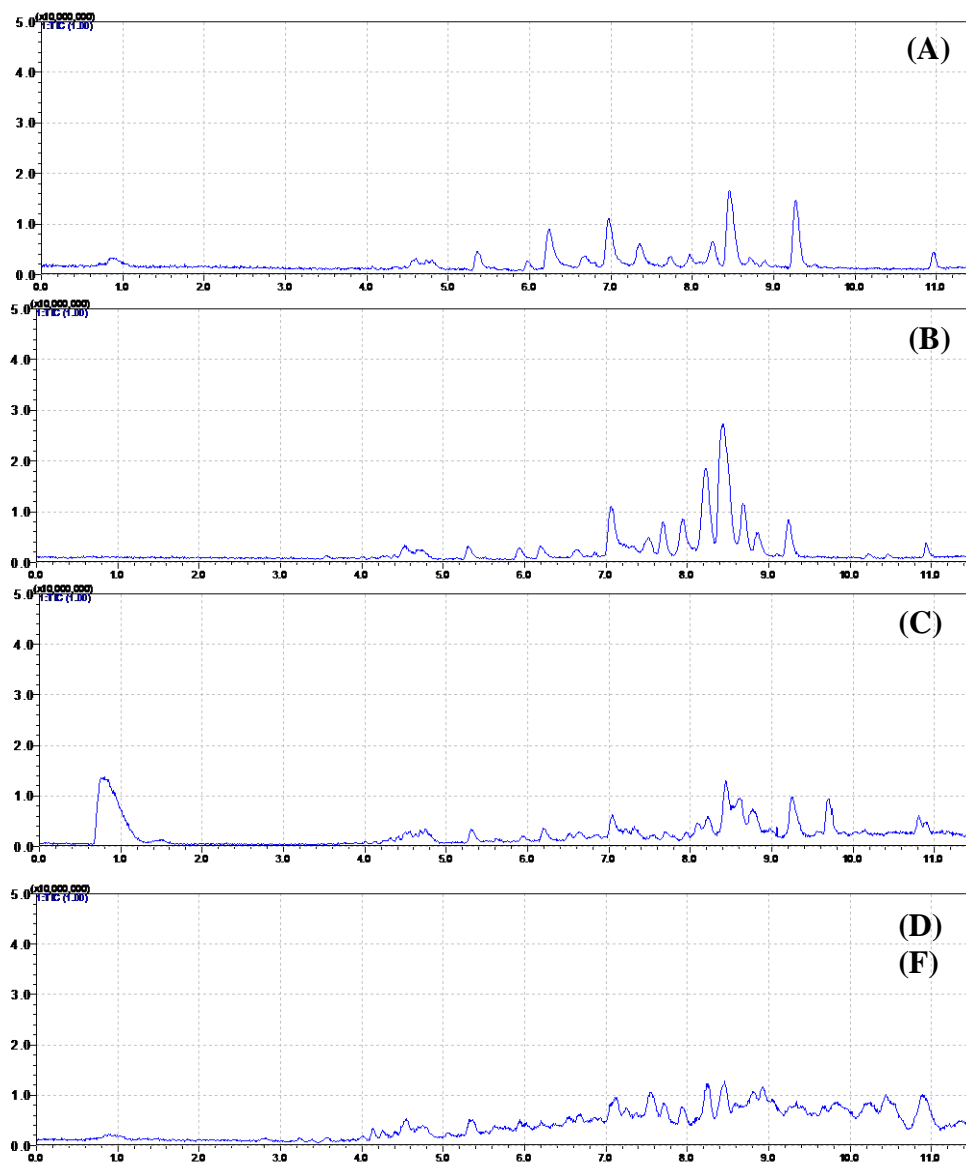
Figure 30 Total ion chromatograms of EBN and its spiked samples based on metabolite fingerprinting with LC/MS in ES+ mode. Chromatogram (A) EBN while chromatogram (B), (C) and (D) samples spiked with agar agar, fungus and isinglass respectively.
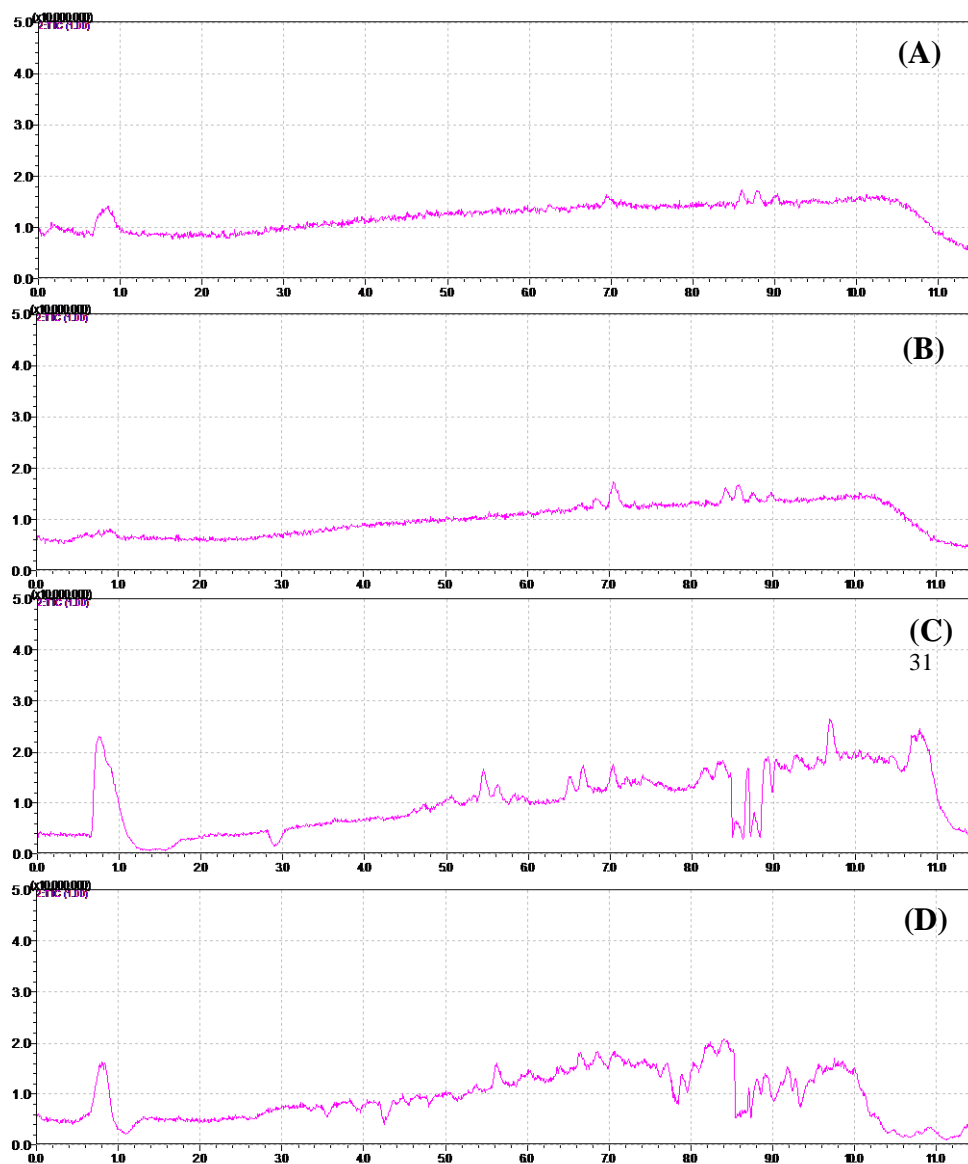
Figure 31 Total ion chromatograms of EBN and its spiked samples based on metabolite fingerprinting with LC/MS in ES- mode. Chromatogram (A) EBN while chromatogram (B), (C) and (D) samples spiked with agar agar, fungus and isinglass respectively.

With the vast amount of data acquired from the analysis, it was difficult to differentiate EBN from the spiked samples merely by inspecting the peak area or intensity of the analyzed samples. Hence, the data sets were subjected to multivariate analysis to observe the similarities or dissimilarities in the chemical profiles easily.[203] Prior to the construction of the score plots, the data sets underwent data pre-processing, which generally consisted of two parts – normalization and scaling.

Normalization is a common data pre-processing method prior to multivariate analysis.[204] Its application is deemed desirable it enables the direct comparison of data sets among the samples. There are two generally accepted and widely used approaches in normalization – normalizing to an internal standard and normalizing to a constant sum (CS). The internal standard approach uses a chemical compound other than the analyte(s) of interest to serve as a reference value, whereby the signal of analyte is expressed as a fraction of the reference signal of the internal standard.[205] This chemical compound generally possesses similar physical or chemical properties to the analyte(s) in the analysis. Thus, the addition of an internal standard serves to minimize errors due to sample preparation or the matrix effect. Although this approach has been proven to be successful in achieving a more accurate and precise analysis result, it incurs extra cost and time and is restricted to targeted analyte analysis.[206] As such, the internal standard approach is applied only to the amino acid and monosaccharide analysis rather than the metabolite analyse with GC/MS and LC/MS. This is because the amino acids and monosaccharides had been pre-determined while the metabolites were identified after the analysis.

Unlike the internal standard approach, the CS approach does not require any additional chemicals or steps in sample preparation; the entire approach involves only the mathematical manipulation of the data set. In such cases, each signal value for the sample is divided by the sum of all the signal values for that sample.[207] This would then eliminate the influence of the sample size over the measured analytes. The approach has been seen to be utilized in the data analysis of urine due to the variable nature of the volume of the sample produced.[208] However, CS is also known to affect the correlation structure of a data set, especially when the number of compounds varies for different samples.[158] This might lead to a different interpretation of the results. Considering both the advantage and disadvantage of CS approach, it would be applied to all four different data sets in this chapter to determine the most ideal data pre-processing method for the differentiation of the EBN and the different types of spiked samples.

Principal component analysis (PCA) score plots based on CS normalized data and without CS normalized data were constructed to determine the interpretable portion of the data sets. Preliminary PCA score plots illustrated that samples with a low percentage of adulterants were noted to be indistinguishable from authentic EBNs. In addition, some of the spiked samples were observed to overlap with each other in the score plots. As such, these samples were excluded from the subsequent multivariate analysis. The final PCA score plots obtained from the remaining samples demonstrated a clear discrimination between EBNs and the spiked samples (Appendix 15). Model performance parameters such as the R^2X(cumulative)and Q^2X

(cumulative) were of values greater than 0.8, thereby indicating that the established score plots were sound and reliable (Table 10).

Referring to the results derived from the amino acid and monosaccharide data, a distinct difference in the detectable spiking range was observed when CS normalization was applied (Table 10). The amino acid and monosaccharide data contained a fixed number of variables for all the samples, thus the adulterants would have to possess variables with a considerable concentration to induce a significant variation between the CS normalized data of EBNs and the spiked samples. Agar agar and fungus have a low amino acid content of 2 and 60 mg /g respectively, while isinglass contains a low monosaccharide concentration of 15 mg / g in comparison with EBN (300 mg / g of amino acids and 500 mg / g of monosaccharides). Thus, it was not surprising that an adulterant was detected to be as high as 80 % of the sample's weight when the adulterant only has a low variable concentration. In comparison, amino acid and monosaccharide data, which did not undergo CS normalization, illustrated a marked improvement in the detection of adulterants – the detectable spiking range is generally 10 – 100 %. Hence, it is proposed that these data are more sensitive to minor changes in the variable content, which result in the discrimination of samples spiked with a minute quantity of adulterants.

On the contrary, the metabolite data sets based on GC/MS and LC/MS exhibited a better detectable spiking range when processed with CS normalization (Table 10). Peaks representing adulterants were clearly observed in the metabolite chromatograms when each sample was spiked with

approximately 5 % of adulterant. However, this observation was only illustrated in the PCA score plot constructed with the metabolite data normalized to CS. As such, the result suggests that the application of CS normalization to the signal of the metabolites would be able to improve the clarity of discrimination power between EBNs and binary mixtures with a low percentage of adulterants in the PCA score plot.

Table 10 Model parameters and detection range results for different normalizing methods.

	Amino acid		Monosaccharide		Metabolite - GC/MS		Metabolite - LC/MS	
	CS normalization	Without CS normalization	CS normalization	Without CS normalization	CS normalization	Without CS normalization	CS normalization	Without CS normalization
Model parameters								
PC	11	17	6	6	5	12	23	20
R^2X (cumulative)	0.999	1	1	1	0.868	0.992	0.985	0.986
Q^2X (cumulative)	0.983	0.999	0.997	0.993	0.814	0.904	0.882	0.876
Detection spiking range								
Agar agar (%)	80 - 100	10 - 100	10 - 100	10 - 100	5 - 100	10 - 100	5 - 100	50 - 100
Fungus (%)	80 - 100	10 - 100	10 - 100	10 - 100	1 - 100	5 - 100	0.5 - 100	0.5 - 100
Isinglass (%)	10 - 100	10 - 100	80 - 100	10 - 100	0.5 - 100	0.5 - 100	0.5 - 100	0.5 - 100

4.3.2. Determination of the scaling method for the qualitative discrimination of EBNs and spiked samples

A multivariate model is known to be established with the variance of the data set. As such, it is vital to assess the variance of the raw data input as it plays a major role in the output of the multivariate analysis.[209] Variations in the raw data are broadly divided into two categories: technical variation and biological variation. Technical variation is contributed by the experimental procedures such as sample preparation or instrumental analysis, while biological variation is the inherent variation of the samples which could be induced by genetic composition, environment or human activities. In data analysis, it would be ideal that technical variations are completely removed from the raw data so that samples are discriminated according to their biological variations, only, so that meaningful interpretations could be made on the samples.[210]

Scaling is a data transformation process prior to the construction of the multivariate model.[211] It serves to modify the weights of the variables in the output multivariate model, thereby being capable of eliminating existing technical variations, as well as highlighting the biological variations. There are three scaling methods: unit variance (uv), pareto (par) and no scaling. These methods are noted to be frequently used to process the raw input data. These three different types of scaling would be applied to the data sets in this chapter (Appendix 16).

The initial PCA score plot established with EBNs and a wide range of spiked samples was observed to be unable to discriminate samples with a low adulterant level. Thus, these samples were removed from the respective data set prior to the construction of the final score plot. The final score plot results are summarized in Table 11. Good quality score plots were established with the different types of scaling in the various analytical analyses (R^2X (cumulative)and Q^2X (cumulative) were above 0.7 respectively). It was also revealed that uv was the most appropriate scaling method for all the different types of analytical methods. In addition, the detectable spiking range generally improves as the scaling method changes from no scaling to par scaling and then to uv scaling. Without scaling, the raw data set is exported for multivariate analysis after it is mean-centered. As such, the output model would be dominated by variables with high variance. On the other hand, par and uv scaling are applied in order to reduce the influence of the variables with a high variance; the latter has a relatively larger reduction. This suggests that the variables with large variances are not as informative as those containing small variances in the qualitative discrimination of EBNs and samples with a low level of adulterants.

Table 11 Model parameters and detection range results for different scaling methods.

	Amino acids			Monosaccharides			Metabolites - GC/MS			Metabolites - LC/MS		
	Uv scaling	Par scaling	No scaling	Uv scaling	Par scaling	No scaling	Uv scaling	Par scaling	No scaling	Uv scaling	Par scaling	No scaling
Model parameters												
PC	17	13	8	6	4	3	5	7	10	23	22	24
R^2X (cumulative)	1	1	1	1	1.000	1.000	0.868	0.950	0.998	0.985	0.994	0.999
Q^2X (cumulative)	0.999	0.974	0.990	0.993	0.960	0.984	0.814	0.786	0.961	0.882	0.924	0.967
Detectable spiking range												
Agar agar (%)	10 - 100	10 - 100	10 - 100	10 - 100	10 - 100	10 - 100	5 - 100	10 - 100	10 - 100	5 - 100	5 - 100	5 - 100
Fungus (%)	10 - 100	30 - 100	30 - 100	10 - 100	10 - 100	10 - 100	1 - 100	1 - 100	1 - 100	0.5 - 100	0.5 - 100	0.5 - 100
Isinglass (%)	10 - 100	10 - 100	10 - 100	10 - 100	50 - 100	50 - 100	0.5 - 100	0.5 - 100	5 - 100	0.5 - 100	1 - 100	1 - 100

4.3.3. Determination of the multivariate analytical method for the qualitative discrimination of EBNs and spiked samples

Besides selecting the data pre-processing approach, the choice of multivariate analysis method is also critical in the success of the authentication of EBNs. Multivariate analysis has been frequently utilized in the authentication of food items such as ginseng and infant milk powder as this multivariate approach is capable of determining the relationship between samples through reducing the data into a plot represented by a few principal components.[212, 213] To date, the most commonly used multivariate analysis methods are PCA, orthogonal projections to latent structures discriminant analysis (OPLS-DA) and projection to latent structure discriminant analysis (PLS-DA). These three different types of multivariate analysis methods would be employed to determine ideal score plots for the authentication of EBN (Appendix 17).

The preliminary score plots display tight clustering between EBN and samples with a low spiking level, thereby indicating that these low spiking levels were insufficient to induce a significant change in the respective variable response. Thus, the final score plots did not include these samples. Referring to the final score plots in Figure 32, it was observed that clear separation was achieved between EBN and the spiked samples. Moreover, samples spiked with the same adulterant were separated from EBN in a similar direction. Furthermore, the separation distance was noted to be dependent on the spiking level. The model performance parameters - R^2X (cumulative), Q^2X (cumulative) and R^2Y (cumulative) are above 0.4, thereby indicating that the quality of the final

score plots was satisfactory (Table 12). The assessment of the score plots was also conducted with external prediction sets. An excellent 100 % correct prediction rate was achieved by all the score plots, thus indicating good specificity.
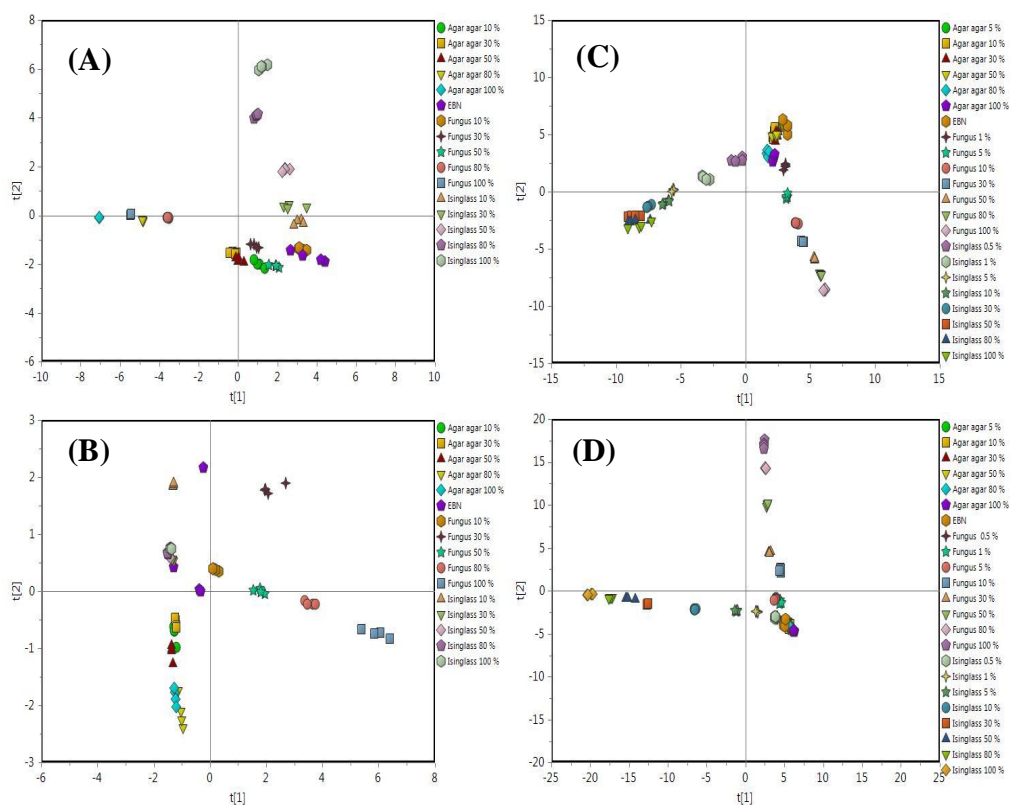
Figure 32 PCA score plots to distinguish between EBN and spiked samples. (A) based on amino acid data. (B) based on monosaccharide data. (C) based on metabolite fingerprinting with GC/MS. (D) based on metabolite fingerprinting with LC/MS

Coomans' plot was also employed to visualize the grouping between the EBN and spiked samples, In the Coomans' plot, two axes are utilized to represent the D-critical distance (P = 0.05) of the respective PCA model.[214] This D – critical distance serves as a boundary for EBN and spiked samples, with the boundary for EBN at the left side while the boundary for spiked samples at the bottom of the plot. Thus, if a predicted sample of a certain distance to model X predicted score (DModXPS) value enters the boundary of another type of sample, it would be deemed as a misclassification (graphically, the predicted sample would be spotted inside the rectangle defining the model of another type of sample). The plots for all the different analytical analysis exhibited a clear distinction between EBN and spiked samples, displayed in Figure33 to 36. As such, this result, together with the validation result on the external prediction set, provides reassurance on the quality of the outcome in Table 12.
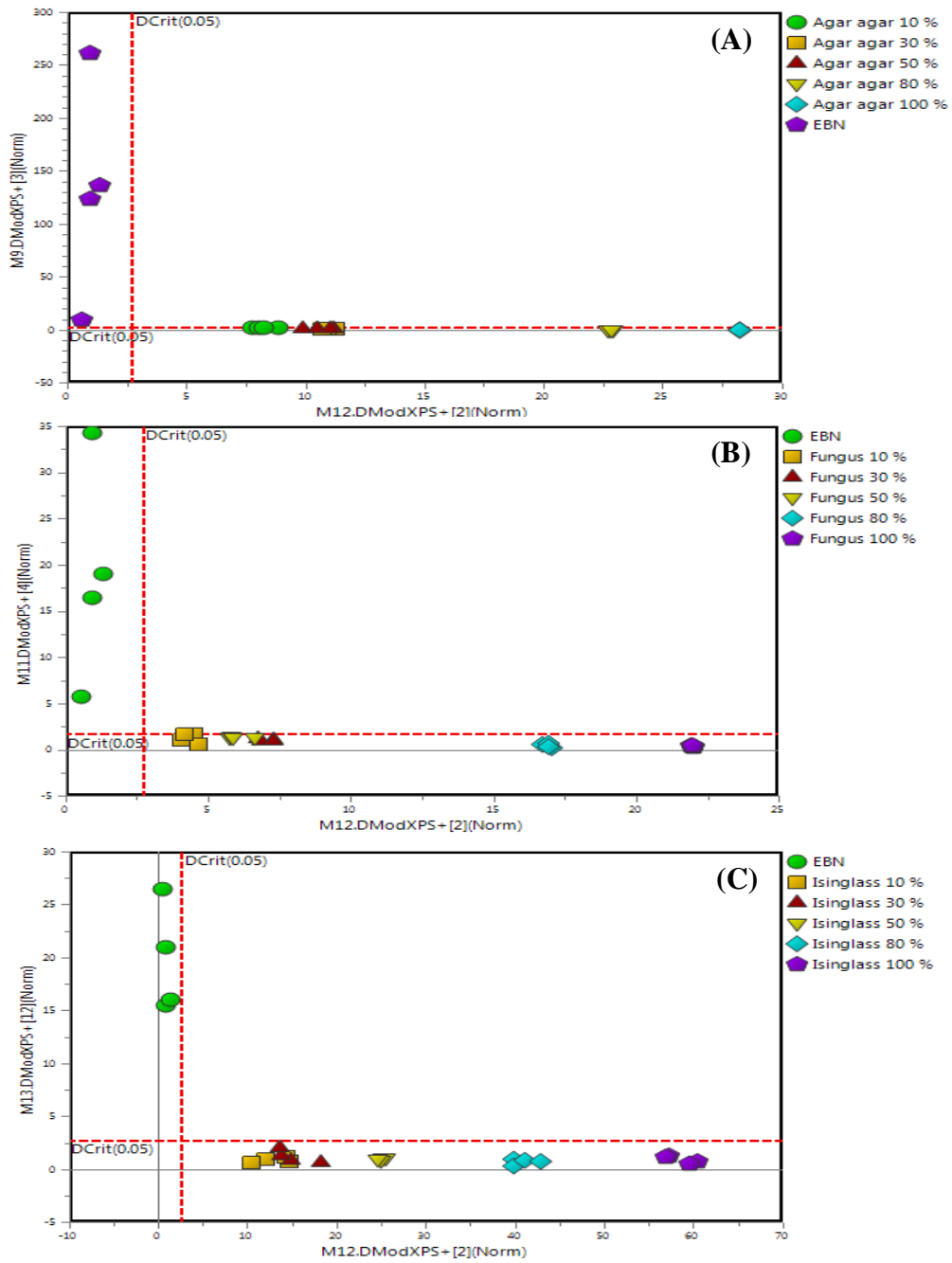
Figure 33 Coomans' plot of EBN and spiked samples based on amino acid analysis. The spiked samples in each plot containing the following adulterant: (A) agar agar, (B) fungus and (C) isinglass.
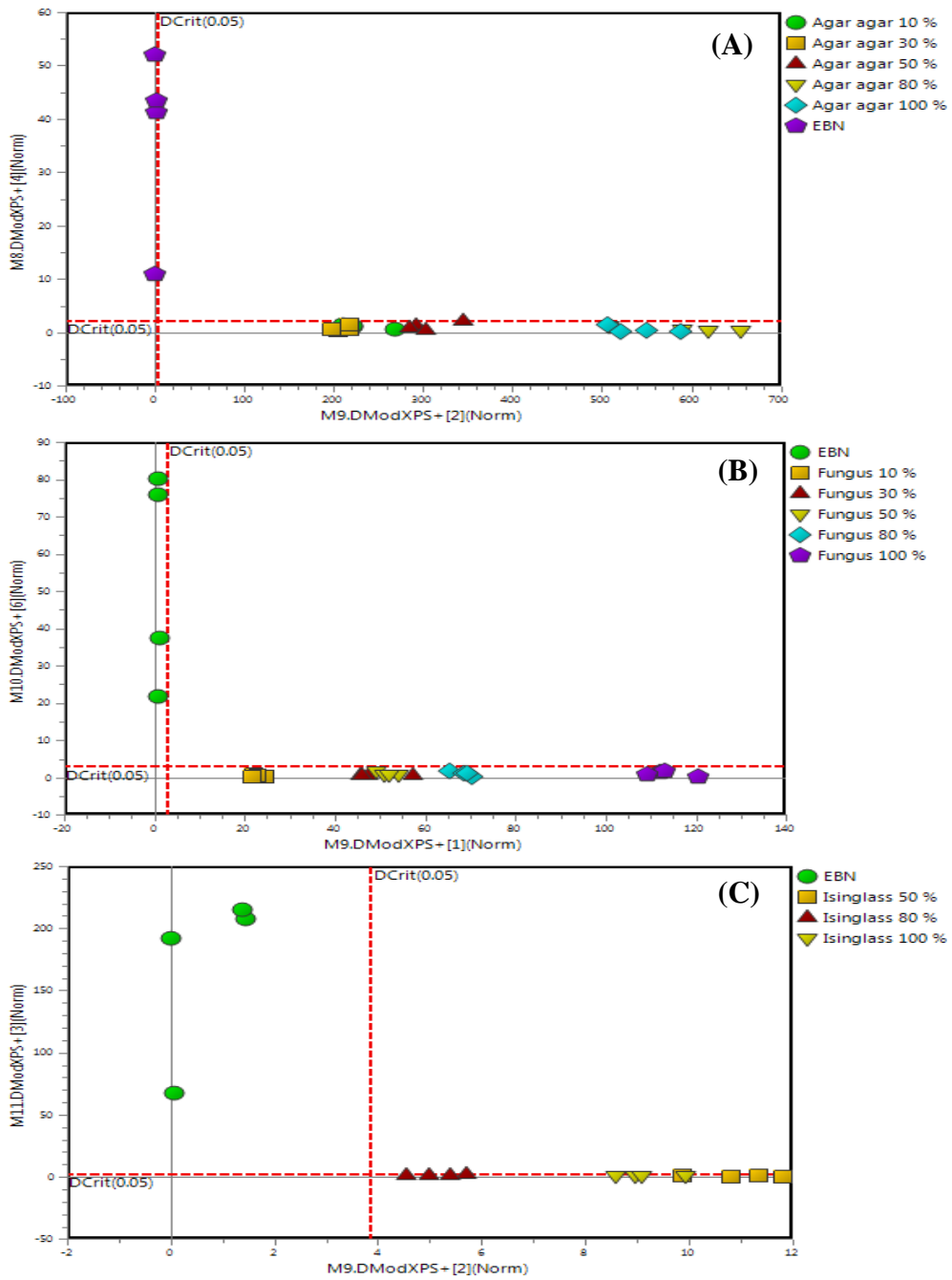
Figure 34 Coomans' plot of EBN and spiked samples based on monosaccharide analysis. The spiked samples in each plot containing the following adulterant: (A) agar agar, (B) fungus and (C) isinglass.
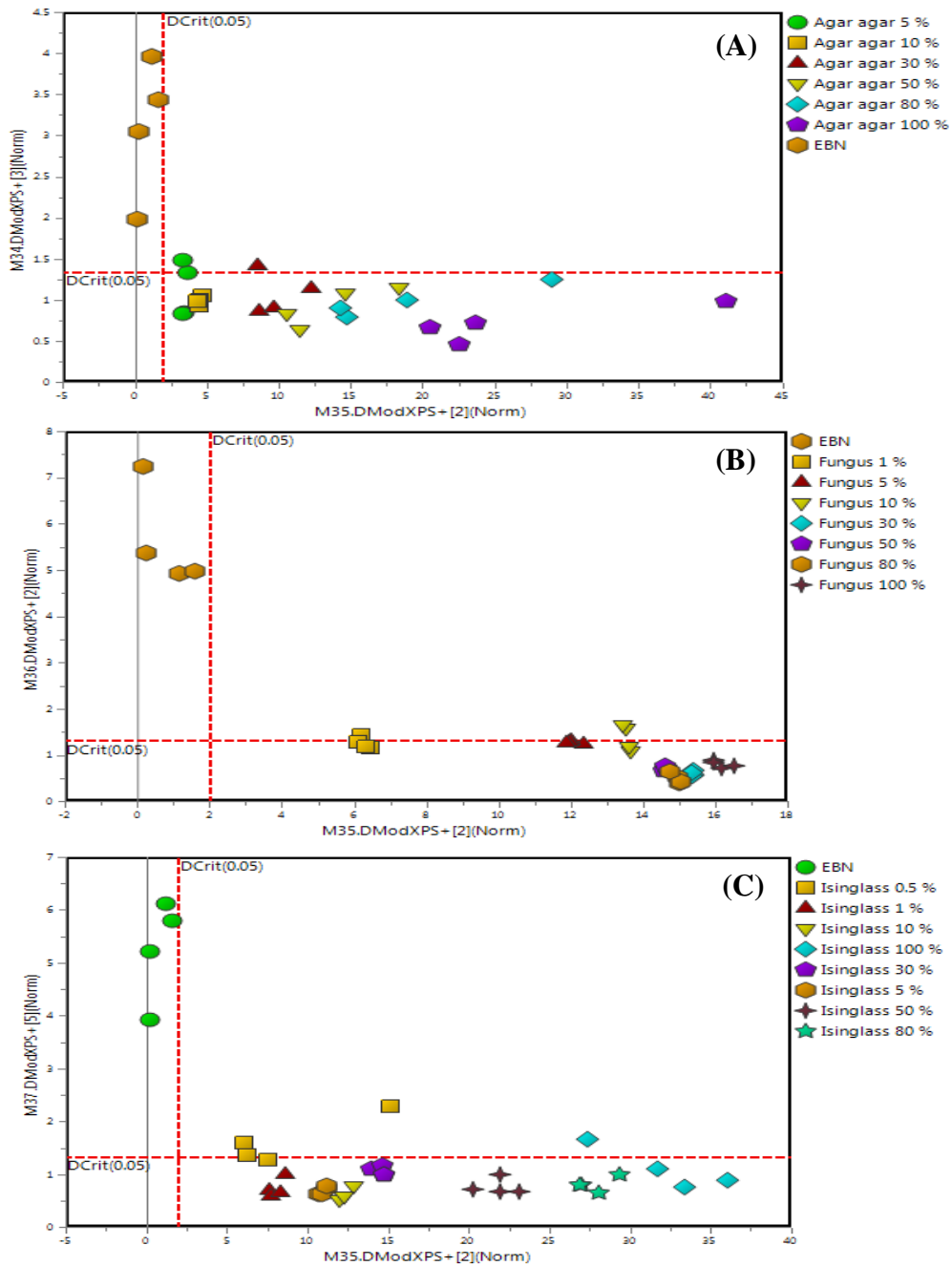
Figure 35 Coomans' plot of EBN and spiked samples based on metabolite fingerprinting with GC/MS. The spiked samples in each plot containing the following adulterant: (A) agar agar, (B) fungus and (C) isinglass.
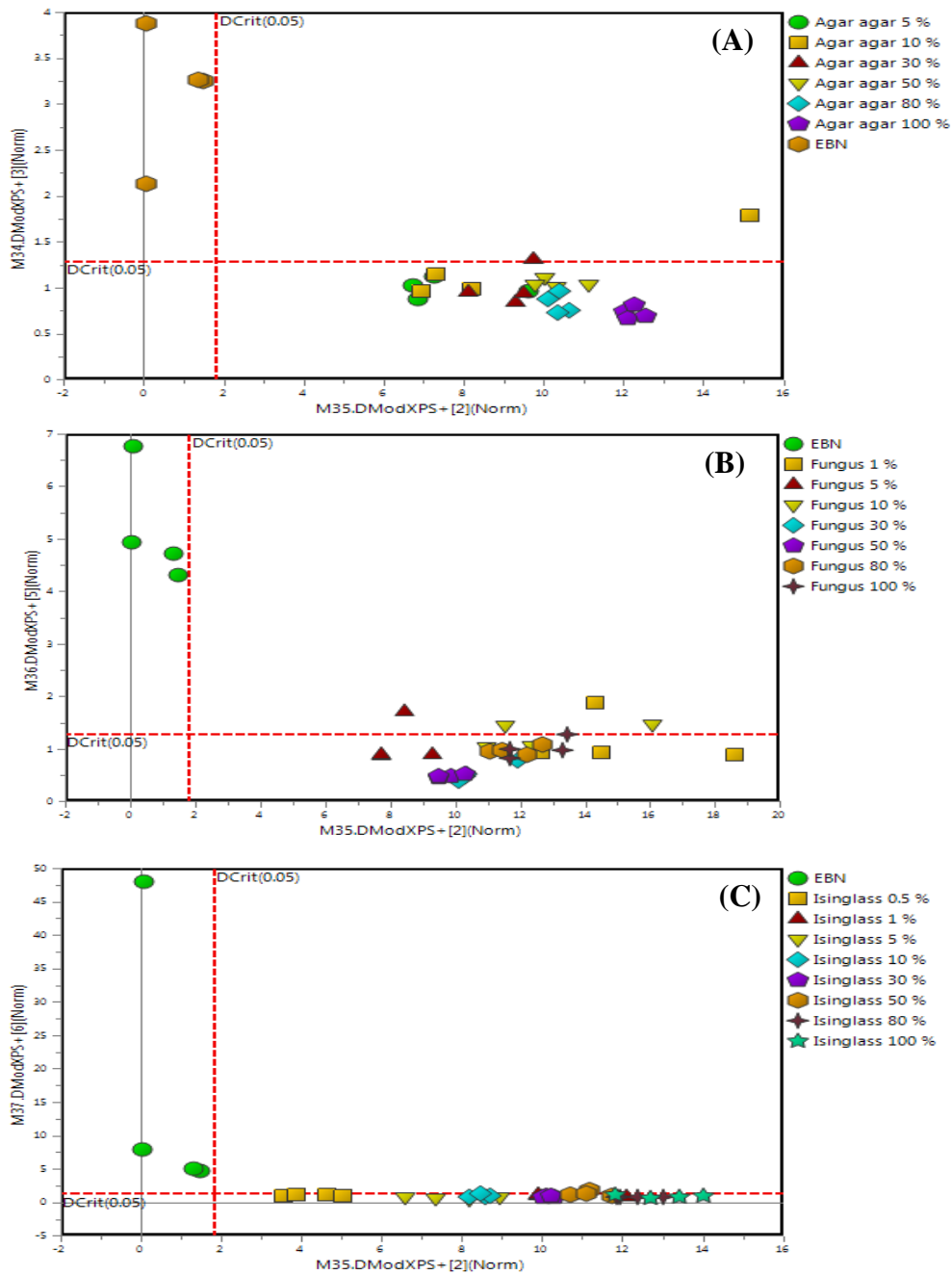
Figure 36 Coomans' plot of EBN and spiked samples based on metabolite fingerprinting with LC/MS. The spiked samples in each plot containing the following adulterant: (A) agar agar, (B) fungus and (C) isinglass.

Based on Table 12, it was observed that even though different multivariate approaches were applied to the same data, the detectable spiking range was revealed to be the same. In such cases, the simplicity of the execution of the multivariate approach would take priority. There are two main groups of multivariate approach – unsupervised and supervised methods. PCA is classified under the unsupervised method, thus information on the class of the samples is not required prior to the construction of this type of score plot.[215, 216] On the other hand, supervised methods like OPLS-DA and PLS-DA entails an additional step of indicating the classes of the samples because the distance between the various class memberships would be maximised during the establishment of these two types of score plot.[217] This was suggested to be able to provide better classification of the samples as the incorporation of orthogonal data may increase the noise in the score plot. Hence, in addition to the PCs representing the co-variation portion, the number of PCs for the orthogonal portion of the data would have to be optimized in the case of the OPLS-DA score plots. With this information, PCA was found to offer the most ideal multivariate solution for the data representing amino acid, monosaccharide and metabolite analysis with GC/MS and LC/MS.

Table 12 Model parameters and detection range results for different multivariate methods.

	Amino acids			Monosaccharides			Metabolites - GC/MS			Metabolites - LC/MS		
	PCA	OPLS-DA	PLS-DA	PCA	OPLS-DA	PLS-DA	PCA	OPLS-DA	PLS-DA	PCA	OPLS-DA	PLS-DA
Model parameters												
PC	17	6+6+2	13	6	3+4+3	7	5	21+0+0	32	23	10+12+5	36
R^2X (cumulative)	1	0.999	1	1.000	1.000	1.000	0.868	0.975	0.997	0.985	0.969	0.994
Q^2X (cumulative)	0.999	0.434	0.437	0.993	0.300	0.292	0.814	0.563	0.612	0.882	0.666	0.844
R^2Y (cumulative)		0.548	0.630		0.404	0.404		0.803	0.919		0.653	0.981
Prediction ability (%)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Detectable spiking range												
Agar agar (%)	10 - 100	10 - 100	10 - 100	10 - 100	10 - 100	10 - 100	5 - 100	5 - 100	5 - 100	5 - 100	5 - 100	5 - 100
Fungus (%)	10 - 100	10 - 100	10 - 100	10 - 100	10 - 100	10 - 100	1 - 100	1 - 100	1 - 100	0.5 - 100	0.5 - 100	0.5 - 100
Isinglass (%)	10 - 100	10 - 100	10 - 100	10 - 100	10 - 100	10 - 100	0.5 - 100	0.5 - 100	0.5 - 100	0.5 - 100	0.5 - 100	0.5 - 100

Comparing the different analytical methods, amino acid and monosaccharide analysis have the advantages of quantifying fewer variables compared to the metabolite analysis. The disadvantages of the amino acid and monosaccharide analysis lie in the long and tedious sample preparation, making it a less productive analytical method and increasing the chances of human errors. The metabolite analysis is an untargeted profiling study, thus no standards will be needed, thereby reducing the analysis cost, preparation time and overall run time. More importantly, metabolite analysis generates a large number of variables, which in turn leads to a better detectable spiking range relative to the amino acid and monosaccharide analyse. As such, the metabolite analysis provides a better platform to differentiate EBN and spiked samples than the amino acid and monosaccharide analysis. Between the two analytical platforms utilized in the metabolite analysis, LC/MS involves a more costly instrument than the GC/MS while GC/MS has a lower throughput. Since both analytical platforms do not exhibit a distinct advantage over the other and they produce comparable qualitative discrimination results, the ideal analytical platform to distinguish EBN from the spiked samples would be dependent on the amount of time available and funding of the intended user.

4.3.4. Quantitative discrimination of EBN and spiked samples with PLS regression

On top of qualitative discrimination, it is also vital to achieve the quantitative discrimination between EBN and spiked samples. This is because it would permit the understanding of the limits of the approaches and more importantly, it may provide the authorities a reliable approach to accurately quantitating the amount of adulterants spiked in the food item. PLS regression is a supervised multivariate calibration method which utilizes numerous variables to construct the regression line for prediction. This would result in the data analysis time to be considerably reduced as the building of multiple regression curves can be averted. The capabilities of PLS regression in evaluating the quality of food have been demonstrated in several literatures. [218, 219] Hence, in this chapter, PLS regression would be employed to determine the quantity of adulterants in the spiked sample (Appendix 18).

Model performance parameters such as R^2X (cumulative), Q^2X (cumulative), R^2Y (cumulative), of the PLS regression models have a value higher than 0.6, thus indicating that good quality models were established with the various analytical analyse (Table 13). The calibration parameters such as R^2, RMSEE and RMSEP are also tabulated in Table 13. In all the analytical analyse, PLS regressions were revealed to exhibit slopes close to 1, intercepts close to 0 and R^2 values larger than 0.98, thereby pointing out low biases and absence of systematic regression errors in the established curves. The predictive power of the PLS calibration curves was evaluated with RMSEC and RMSEP.[220]

Values for the predictive parameters were generally lower than 4 with the exception of the RMSEP values from amino acid and monosaccharide data. This inconsistency in the predicted samples indicates the presence of high variations in the amino acid and monosaccharide data. The analytical methods employed in this chapter to quantitate amino acids and monosaccharides have been developed for EBN. Thus, these analytical methods might not be able to accurately reflect the quantity of the amino acids and monosaccharides in the spiked samples. Besides that, the adulterants contained a minute quantity of amino acids and monosaccharides as mentioned earlier. This might have further aggravated the fluctuations of the results of the spiked samples.

With reference to Table 13, the ideal analytical method to quantitate the amount of adulterants in the spiked samples is metabolite analysis with GC/MS. In this analytical method, satisfactory model performance and calibration parameters were achieved and most importantly, it provided the best detectable spiked range (Figure 37). As expected, the metabolite analysis outperformed the amino acid and monosaccharide analyse due to a larger number of discriminating variables detected in the metabolite analysis. Although the number of variables detected through metabolite analysis with GC/MS is much lower than the LC/MS - GC/MS, containing 63 variables while LC/MS contains 157 variables, GC/MS was determined to have a better detectable spiking range than LC/MS. This could be attributed to the higher sensitivity of GC/MS towards the discriminating variables in the adulterants, which in turn leads to lower detected adulterant level in the PLS regression. Another observation was that the detectable spiked range for the metabolite

analysis was narrower in the quantitative discrimination compared to the qualitative discrimination. This is because spiked samples with a low percentage of adulterants (generally samples with lesser than 1 % of an adulterant) were found to overlap with each other in the PLS regressions. Thus, these samples were excluded from the final calibration curves as it was not possible to determine the actual value of adulterants in these samples. A possible attribution could be the generic extraction method employed. Utilization of such a method might not have been able to accurately extract all the metabolites in the spiked samples, especially when the adulterant was a minute amount.
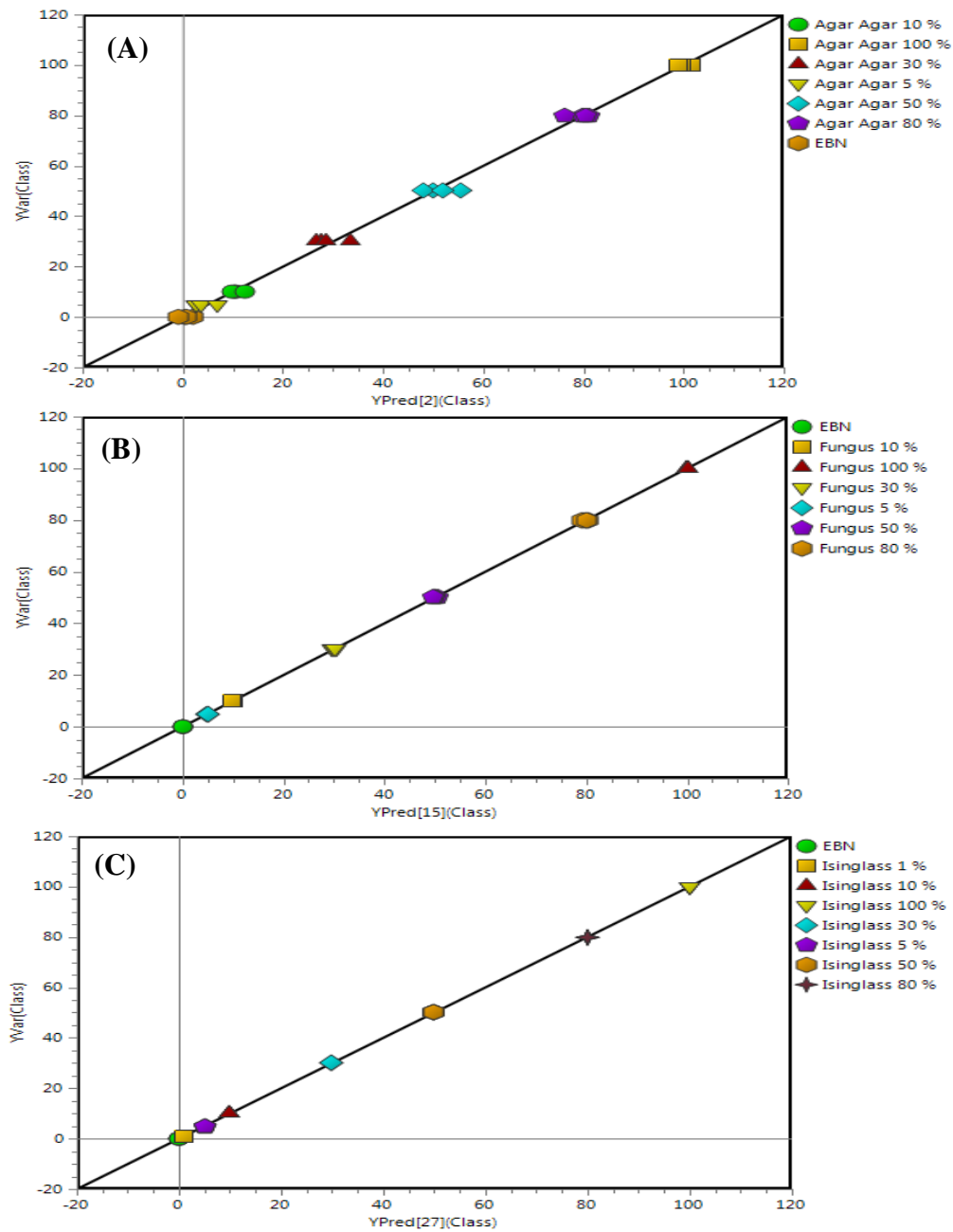
Figure 37 PLS regression plots of EBN and spiked samples based on metabolite fingerprinting with GC/MS. The spiked samples in each plot containing the following adulterant: (A) agar agar, (B) fungus and (C) isinglass.

Table 13 Statistical values of PLS regression for the quantitative discrimination between EBN and spiked samples.

	Amino acids			Monosaccharides			Metabolites - GC/MS			Metabolites - LC/MS		
	Agar Agar	Fungus	Isinglass	Agar Agar	Fungus	Isinglass	Agar Agar	Fungus	Isinglass	Agar Agar	Fungus	Isinglass
PC	7	7	2	6	7	7	2	15	27	15	6	15
R^2X (cumulative)	0.998	0.995	0.858	1.000	1.000	1.000	0.656	0.997	0.999	0.978	0.960	0.988
R^2Y (cumulative)	0.995	0.998	0.991	0.993	0.988	0.997	0.997	1.000	1.000	1.000	1.000	1.000
Q^2X (cumulative)	0.988	0.986	0.986	0.986	0.969	0.891	0.994	0.993	1.000	1.000	0.999	0.999
Slope	1	1	1	1	1	1	1	1	1	1	1	1
Intercept	1.66 $\exp.^{-6}$	-2.82 $\exp.^{-6}$	-1.85 $\exp.^{-6}$	1.89 $\exp.^{-6}$	-1.54 $\exp.^{-6}$	-1.47 $\exp.^{-6}$	2.98 $\exp.^{-6}$	1.75 $\exp.^{-6}$	-2.87 $\exp.^{-6}$	1.41 $\exp.^{-6}$	3.64 $\exp.^{-6}$	-1.07 $\exp.^{-6}$
R^2	0.995	0.998	0.991	0.993	0.988	0.997	0.997	1.000	1.000	1.000	1.000	1.000
RMSEE	2.962	1.872	3.573	3.573	4.765	2.832	2.199	0.393	0.003	0.020	0.785	0.060
RMSEP	3.068	3.754	4.067	11.667	8.997	8.766	1.886	1.753	3.759	3.288	1.263	1.210
Detectable spiking range (% w/w)	10 - 100	10 - 100	10 - 100	10 - 100	10 - 100	50 - 100	5 - 100	5 -100	1- 100	10 - 100	5 -100	1- 100

4.3.5. Variable importance for the projection (VIP) plot for the spiked samples

Besides discrimination of EBN and spiked samples, PLS regression is able to offer quantitative measurements of the discriminatory power of each variable with the use of VIP plots.[221] Mathematically, VIP is the weighted sum of squares of the PLS weights, which explains the amount of Y-variance (percentage of adulterant in the sample) in each latent variable.[222] In the VIP plot, variables with VIP values greater than 1 are considered as having meaningful contribution in the PLS model whereas variable assigned with a VIP value 0.5 or smaller would be deemed to have small or no influence on the established model.[223] The variables in the VIP plot for each PLS model would be arranged in ascending order for the ease of reference. It was revealed that the number of high influence variables (VIP greater than 1) in the PLS model based on amino acid, monosaccharide and metabolite analyse with GC/MS and LC/MS are 38, 14, 54 and 163 respectively.

Referring to the VIP plots on the amino acid analysis, samples spiked with agar agar had the highest number of amino acids which exhibited major discriminatory power between EBN and the spiked samples (Table 14). This was expected as agar agar had the lowest amino acid content (2 mg/g) among the adulterants. An interesting amino acid, 4-hydroproline, was identified only in the VIP plot for samples spiked with isinglass due to the presence of collagen – a 4-hydroproline rich protein, found exclusively in animals.[224]

In the case of the monosaccharide analysis, samples spiked with agar agar and fungus illustrated that the factor responsible for the difference between EBN and the spiked samples was mainly glucose. This could be attributed to the fact that these two adulterants originated from agar agar and fungus, thus they would contain a high content of cellulose - polysaccharide whose primary component is glucose.[225] On the other hand, the discriminating variable between isinglass and EBN was revealed to be galactose. This is likely due to the numerous glycoconjugates present in the food item (Table 15).[226]

In addition, the VIP plot could also be used to identify characteristic variables of the adulterant. For instance, common metabolites in agar agar – hexadecanoic acid and E-9-octadecenoic acid, were observed to be one of the main metabolites responsible for the discrimination of EBN and agar agar in the VIP plot based on metabolite analysis with GC/MS (Table 16).[227] This outcome was in line with the GC/MS chromatogram where the major discrimination appeared between 25 to 35 min. On the other hand, metabolites which are characteristic of fungus were generally found to have a VIP value greater than 1 (1,2-dipalmitin and 1,3-dipalmitin), with the exception of glucitol, 5-alpha-ergost-8 (14)-en-3-beta-ol and ergosterol.[228] It was a surprise that glucitol had a VIP value lesser than 1 as it is one of the most distinctive peak in the fungus's chromatogram. This could be due to the effects of normalization which changed the correlation structure of the data. 5-Alpha-ergost-8 (14)-en-3-beta-ol and ergosterol are fungus sterols which are used as indicators for the plant.[229] Their limited influence could be attributed to their small quantity size when there was a minute amount of fungus spiked into the

sample. In isinglass, metabolites with VIP values greater than 1 were mostly fatty acids - hexanoic acid, arachidonic acid, octanoic acid, tetracosanoic acid, eicosanoic acid and (Z)-9,12-octadecadienoic acid. These fatty acids are commonly found in meat and cooking oil, which fits the characteristic of isinglass.[230] Hexadecanoic acid and E-9-octadecenoic acid are major metabolites in isinglass, however, they were noted to have a low VIP values. This is attributed to the high abundance of these two metabolites in the composition of EBN.

Other than the metabolite analysis with GC/MS, the LC/MS platform also exhibited characteristic metabolites. Although the detected ions were not identified in the LC/MS analysis, the LC/MS platforms revealed that samples spiked with isinglass contained the largest number of variables with high discriminatory power (samples spiked with agar agar, fungus or isinglass contain 32, 84 and 163 variables with VIP values greater 1 respectively). This is probably due to the frying of the isinglass prior to its sales on the market. Cooking oil is known to be generally derived from sources like olives, peanuts and seeds.[231] As such, frying of isinglass might have added metabolites from a different source from isinglass to the final food item, thereby leading to a large number of discriminating variables in the VIP plot (Appendix19).

Table 14 Summary of amino acids with VIP values greater than 1.

No.	Variable	VIP	No.	Variable	VIP	No.	Variable	VIP
I) Agar agar			II) Fungus			III) Isinglass		
1	Methionine	1.08	15	Methionine	1.08	27	Tyrosine	1.10
2	Serine	1.07	16	Tyrosine	1.07	28	Leucine	1.10
3	Tyrosine	1.07	17	Valine	1.05	29	Alanine	1.10
4	Lysine	1.06	18	Proline	1.05	30	Phenylalanine	1.10
5	Glutamic acid	1.05	19	Phenylalanine	1.05	31	Glycine	1.10
6	Isoleucine	1.05	20	Aspartic acid	1.05	32	4-hydroxyproline	1.10
7	Valine	1.05	21	Leucine	1.04	33	Isoleucine	1.08
8	Leucine	1.04	22	Threonine	1.04	34	Valine	1.08
9	Phenylalanine	1.04	23	Isoleucine	1.03	35	Proline	1.08
10	Alanine	1.04	24	Lysine	1.03	36	Methionine	1.08
11	Glycine	1.04	25	Glutamic acid	1.03	37	Threonine	1.04
12	Proline	1.04	26	Glycine	1.01	38	Serine	1.02
13	Aspartic acid	1.03						
14	Threonine	1.03						

Table 15 Summary of monosaccharides with VIP values greater than 1.

No.	Variable	VIP	No.	Variable	VIP	No.	Variable	VIP
I) Agar agar			II) Fungus			III) Isinglass		
1	Glucose	1.14	6	Ribose	1.09	11	Galactose	1.33
2	Fucose	1.09	7	Glucose	1.09	12	Fucose	1.29
3	Galactose	1.07	8	Mannose	1.09	13	Glucose	1.18
4	Mannose	1.04	9	Fucose	1.07	14	Mannose	1.04
5	Xylose	1.04	10	Xylose	1.06			

Table 16 Summary of metabolites analysed with GC/MS with VIP values greater than 1.

No.	Variable	VIP	No.	Variable	VIP	No.	Variable	VIP
I) Agar agar			II) Fungus			III) Isinglass		
1	(E)-9-Hexadecenoic acid	1.31	22	Hydroxybutanedioic acid	1.33	41	Hexanoic acid	1.60
2	Thymol-beta-D-glucopyranoside	1.30	23	Unknown 2	1.33	42	9-Octadecenamide	1.49
3	D-Glucoheptono-1,4-lactone	1.30	24	Succinic acid	1.33	43	Hydroxybutanedioic acid	1.42
4	Cholesterol	1.29	25	1,3-Dipalmitin	1.32	44	Glycerophosphate	1.42
5	Hexadecanoic acid	1.28	26	E-9-Octadecadienoic acid	1.32	45	5-Oxoproline	1.34
6	E-9-Octadecadienoic acid	1.25	27	Unknown 4	1.27	46	Undecanedioic acid	1.30

Table 4.7 continue

No.	Variable	VIP	No.	Variable	VIP	No.	Variable	VIP
I) Agar agar			II) Fungus			III) Isinglass		
7	Decanoic acid	1.25	28	Glycerophosphate	1.26	47	Arachidonic acid	1.27
8	Hydroxybutyric acid	1.24	29	4-Aminobutanoic acid	1.25	48	Octanoic acid	1.25
9	Unknown 7	1.23	30	E-11-Octadecadienoic acid	1.23	49	1-Monooleoylglycerol	1.19
10	Octanoic acid	1.21	31	D-Glucoheptono-1,4-lactone	1.21	50	Unknown 7	1.13
11	Tetradecanoic acid	1.20	32	Hexadecanoic acid	1.18	51	Tetracosanoic acid	1.12
12	5-Oxoproline	1.19	33	1,2-Dipalmitin	1.18	52	(Z)-9-Octadecenenitrile	1.12

Table 4.7 continue

No.	Variable	VIP	No.	Variable	VIP	No.	Variable	VIP
I) Agar agar			II) Fungus			III) Isinglass		
13	Xylitol	1.19	34	Xylitol	1.16	53	Sebacic acid	1.12
14	Heptadecanoic acid	1.13	35	Unknown 3	1.13	54	Eicosanoic acid	1.11
15	(Z)-9-Hexadecenoic acid	1.13	36	2-Monopalmitin	1.11	55	(Z)-9,12-Octadecadienoic acid	1.00
16	Nonanoic acid	1.09	37	Myo-Inositol	1.09			
17	Succinic acid	1.08	38	D-Galactitol	1.03			
18	(Z)-9-Octadecenenitrile	1.08	39	D-Turanose	1.01			

Table 4.7 continue

No.	Variable	VIP	No.	Variable	VIP	No.	Variable	VIP
I) Agar agar			II) Fungus			III) Isinglass		
19	Octadecanoic acid	1.07	40	Pentadecanoic acid	1.01			
20	Methyl 2-acetamido-2-deoxy-alpha-D-glucopyranoside	1.04						
21	Methyl 2-acetamido-2-deoxy-beta-D-glucopyranoside	1.03						

The most suitable analytical method to authenticate EBN in this chapter is metabolite analysis with GC/MS as it demonstrated the best detectable spiking range. Compared to existing methods to authenticate EBN as mentioned earlier, the use of a single metabolite analysis is capable of performing both qualitative and quantitative discrimination of EBN and samples spiked with adulterants as low as 0.5 % and 1 % respectively. In addition, data pre-processing prior to the construction of the multivariate model serves to eliminate the variations which might be induced by the processing method of EBN. However, metabolites have to be extracted from EBN to undergo chromatographic analysis prior to obtaining the data for authentication. The time taken for sample preparation and analysis would result in a lower throughput method relative to most existing EBN authentication methods. Furthermore, the use of GC/MS might incur a higher experimental cost which may be undesirable for users with constraints in funding.

4.3.6. Qualitative and quantitative discrimination of EBN and multiple spiked samples

Most studies focus on the discrimination of one type of adulterants. In reality, however, more than one adulterant can be spiked into EBN. As such, it is desirable to discriminate samples spiked with multiple adulterants (known as multiple spiked samples) from EBN in a single measurement. It is evident from the section 4.3.3. and 4.3.4. that multivariate data models constructed based on metabolite analysis with GC/MS offer the best solution for discriminating EBN and spiked samples. Hence, these models are utilized to

determine the applicability of multivariate data models for qualitative and quantitative discrimination of EBN and multiple spiked samples.
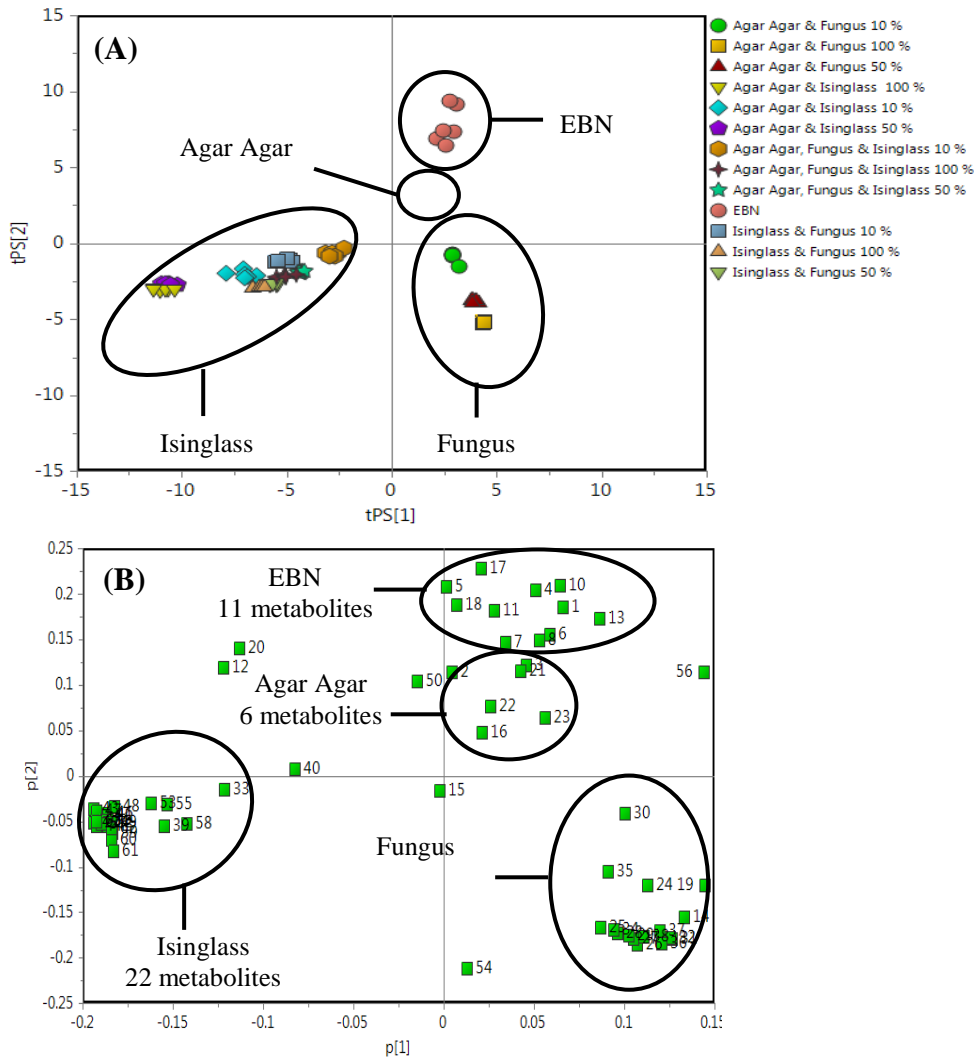
Figure 38 (A) Prediction results of multiple spiked samples using PCA score plot based on metabolite fingerprinting with GC/MS. (B) Loading plot of PCA score plot based on metabolite fingerprinting with GC/MS.

The prediction result for qualitative discrimination illustrates its feasibility to discriminate EBN from multiple spiked samples (Figure 38a). The result also shows that the multiple spiked samples are situated in the region accounting for a single type of adulterant, rather than in between regions of adulterant. This is due to the difference in the number of metabolites responsible for the sample's position in the PCA score plot – number of metabolites in the region for agar agar, fungus and isinglass is 6, 16 and 22 respectively (Figure 38b). Adulterants with a larger number of metabolites would have a greater influence in the metabolite composition of the multiple spiked samples. This in turn leads to the sample being situated in a region that accounts for the adulterant with a higher influence.

Referring to Table 17, RMSEP for the multiple spiked samples are higher in comparison to the spiked samples. This indicates that accurate prediction of the percentage of individual adulterants in multiple spiked samples could not be performed using PLS regression. The RMSEP for fungus prediction in samples spiked with agar agar and fungus is 19.88, much lower than agar agar prediction (167.74). RMSEP is affected by metabolites with VIP values greater than 1 and the common metabolites between the adulterants that are spiked into the sample. Based on these criteria, the number of metabolites affecting RMSEP for the fungus and agar agar prediction is determined to be 8 and 13 respectively (Appendix 20). The lower number of metabolites in fungus prediction implies that the extent of influence for agar agar is lower than fungus, thereby accounting for the large difference in RMSEP.

Table 17 RMSEP for multiple spiked samples.

Type of PLS Regression	RMSEP (%)			
	Agar agar and fungus[a]	Agar agar and isinglass[a]	Isinglass and fungus[a]	Agar agar, fungus & isinglass[a]
Agar Agar	167.74	147.42	N.A.	84.48
Fungus	19.88	N.A.	54.82	65.80
Isinglass	N.A.	279.71	198.00	84.02

[a] represents n = 6
N.A represents not applicable

4.4. Conclusion

The approach to discriminating EBN by combining different analytical analyse - amino acids, monosaccharides and metabolites analysed by GC/MS and LC/MS, with multivariate analysis has not only proven to be feasible but also accurate and reliable. Prior to the construction of multivariate models, different CS normalizations were applied to each data set to achieve the best detectable scaling range. This indicates that application of CS normalization was dependent on the type of analytes. PCA with uv scaling was identified and employed to all data sets for the qualitative discrimination of EBN and samples spiked with adulterant at levels as low as 0.5%.

With the use of PLS regression, it is possible to accurately quantitate the level of adulterants down to 1 % of the spiked sample. In addition, discriminating variables from the respective data sets were identified with VIP plots. An

interesting point to note was that the metabolite data sets exhibited variables which differed from the characteristic of the adulterant due to changes in the correlation structure.

Besides being applied to binary spiked samples, the developed approach was also utilized to discriminate multiple spiked samples. The results suggest that the complexity of the sample does not hamper the qualitative discrimination of EBN. Hence, the approach is feasible in the qualitative discrimination of EBN and multiple spiked samples. The success of qualitative and quantitative discrimination indicates that authentication of EBN can be conducted in a two step approach – qualitative discrimination with a PCA score plot, followed by accurate quantitation of the adulteration level with PLS regression. Given the accuracy and reproducibility of this approach, it could be a highly promising tool that a regulatory body can harness to combat against fraud cases involving EBN.

CHAPTER 5. CONCLUSION AND FUTURE WORK

5.1. Conclusion

The invaluable food item known as EBN has often been plagued by authenticity issues. In this thesis, different analytical methods and multivariate data analysis were utilized to solve the authenticity issues that have been afflicted upon the EBN industry for many years. Authentication is divided into three main parts – (a) identification of EBN, (b) classification of EBN and (c) discrimination between genuine EBN and EBN spiked with adulterants (spiked samples) to facilitate the formation of systematic approaches in resolving the quality control of EBN.

In the identification of EBN, an approach that combines amino acid and monosaccharide analyse with multivariate data analysis was developed and validated to identify the food item based on species origin. Prior to the approach, an analytical method was established and evaluated (RSD lesser than 10 %) in order to qualify monosaccharides in EBN. This provides confidence for the reproducibility of the data to be used in the multivariate data analysis. The multivariate data analysis known as Hotelling T2 range plot, in combination with amino acid and monosaccharide data was successful in predicting the different types of EBN and differentiating various types of samples. This outcome suggests EBN contains a group of glycoproteins which is not affected by the EBN's coloration, country of origin and/or the processing method of the food item. In addition, the glycoproteins were shown

to be unique to EBN. EBN was also revealed to be rich in protein and EAA as well as a variety of monosaccharides wider than most food items. Besides identification of EBN, amino acid and monosaccharide data, in combination with OPLS-DA is proven to be a feasible approach in the quality control of EBN. The overall findings suggest that the amino acid and monosaccharide data provides information not only on the detected compounds and also valuable insights into the glycoproteins of EBN.

Metabolite fingerprinting was performed to investigate the chemical components in EBN and to utilize the data for the classification of EBN according to their coloration, countries and production sites. Forty-three metabolites and thirty-five metabolites were detected in GC/MS and LC/MS respectively. Metabolites were identified and their relationships with the medicinal properties of EBN were deduced. Classification of EBN was performed with OPLS-DA as each score plot demonstrated a clear discrimination pattern when EBN was classified. Several metabolites were found to be responsible for the discrimination of EBN in each classification category and the metabolites' variation was revealed to be associated with the environment. Validation of the score plots indicates that GC/MS was more suitable than LC/MS for classification as the processing methods of the EBNs would not cause a significant variation in the metabolites detected by GC/MS.

For detection of adulterants, an approach was developed and validated for the discrimination between genuine EBN and spiked EBN with different analytical methods and multivariate data analysis. The approach comprises

qualitative and quantitative discrimination of EBN and spiked samples. In the qualitative discrimination, it was found that the ideal normalization method was content dependent while the PCA model which undergoes unit variance UV scaling was applicable to all the analytical methods. The detectable spiking amount for the qualitative discrimination was as low as 0.5 %. Meanwhile, application of PLS regression shows that metabolite analysis with GC/MS was the most suitable analytical analysis for quantitative discrimination as this analysis was able to achieve an accurate detectable adulterant level down to 1 %. This developed approach was also able to perform qualitative discrimination at low levels of multiple spiked samples. The success of qualitative and quantitative discrimination between EBN, and low levels of spiked samples and multiple spiked samples illustrates the feasibility of employing just one analytical method for the authentication of EBN.

In conclusion, it is feasible to make use of data obtained from advanced analytical techniques in combination with multivariate data analysis to authenticate EBN in terms of identifying EBN from samples with similar matrices, classifying the different types of EBN and discriminating the food item from samples which are spiked with a minute quantity of adulterants. This may serve as an approach for use by government regulatory bodies to safeguard the quality of EBN in the interest of consumers and food suppliers.

5.2. Future work

The proposed approach of combining analytical methods and multivariate data analysis has successfully resolved the authenticity issues involving EBN. Nevertheless, this approach is not a perfect one. The analytical methods utilized require a considerable amount of analysis time prior to obtaining the data for food authentication. This may hamper its usefulness as a routine authentication method to be employed by regulatory bodies as it may take days to complete the analysis. However, this limitation can be potentially resolved with the use of NMR. This is because it is able to detect the required compounds with a shorter analysis in approximately five minutes, thus significantly improving the throughput. As such, it is proposed that future analytical methods based on NMR could be developed for the authentication of EBN.

In the metabolite fingerprinting of EBN, the number of metabolites identified were fewer than the detected ones. This might be due to the limited amount of information that GC/MS and LC/MS could provide to deduce the identity of the metabolites. To overcome this limitation, NMR could be applied to the metabolite fingerprinting of EBN as this analytical instrument is known to be able to detect minor differences in the magnetic field to elucidate the structure of an unknown compound. With the utilization of NMR and in combination with the information obtained in this thesis, it is likely that more metabolites could be identified. This would enable a more comprehensive understanding on EBN and health benefits brought about by its consumption.

REFERENCES

1. Dennis, M. J., Recent developments in food authentication. *Analyst* **1998**, *123*, 151R-156R.

2. Chan, E. C. Y.; Koh, P. K.; Mal, M.; Cheah, P. Y.; Eu, K. W.; Backshall, A.; Cavill, R.; Nicholson, J. K.; Keun, H. C., Metabolic profiling of human colorectal cancer using high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy and gas chromatography mass spectrometry (GC/MS). *J. Proteome Res.* **2009**, *8*, 352-361.

3. Shears, P., Food fraud - A current issue but an old problem. *Br. Food J.* **2010**, *112*, 198-213.

4. Wikipedia 2008 Chinese milk scandal. http://en.wikipedia.org/wiki/2008_Chinese_milk_scandal#Source_of_contamination (accessed 3th March 2014).

5. Organization, W. H.O. http://www.who.int/en/ (accessed 4th March 2014).

6. Nations, FAO http://www.fao.org/home/en/ (accessed 4th March 2014).

7. Tan, A. S. L., An approach to building the case for nutrition policies to limit trans-fat intake - A Singapore case study. *Health Policy* **2011**, *100*, 264-272.

8. Wikipedia 2013 meat adulteration scandal. http://en.wikipedia.org/wiki/2013_meat_adulteration_scandal (accessed 4th March 2014).

9. Abbots, E. J.; Coles, B., Horsemeat-gate -The discursive production of a neoliberal food scandal. *Food, Culture and Society* **2013**, *16*, 535-550.

10.   De La Fuente, M. A.; Juárez, M., Authenticity assessment of dairy products. *Crit. Rev. Food Sci. Nutr.* **2005**, *45*, 563-585.

11.   Ortea, I.; Pascoal, A.; Cañas, B.; Gallardo, J. M.; Barros-Velázquez, J.; Calo-Mata, P., Food authentication of commercially-relevant shrimp and prawn species: From classical methods to Foodomics. *Electrophoresis* **2012**, *33*, 2201-2211.

12.   Armanino, C.; De Acutis, R.; Rosa Festa, M., Wheat lipids to discriminate species, varieties, geographical origins and crop years. *Anal. Chim. Acta* **2002**, *454*, 315-326.

13.   Reid, L. M.; O'Donnell, C. P.; Downey, G., Recent technological advances for the determination of food authenticity. *Trends in Food Sci. Tech.* **2006**, *17*, 344-353.

14.   Agricultural and Processed Food Products Export Development Authority,   I. http://www.apeda.gov.in/apedawebsite/Announcements/Traceability_System_HortiNet.htm (4th March 2014).

15.   Capuano, E.; Boerrigter-Eenling, R.; van der Veer, G.; van Ruth, S. M., Analytical authentication of organic products: An overview of markers. *J. Sci. Food Agric.* **2013**, *93*, 12-28.

16.   Fujiwara, M.; Ando, I.; Arifuku, K., Multivariate analysis for1H-NMR spectra of two hundred kinds of tea in the world. *Anal. Sci.* **2006**, *22*, 1307-1314.

17.   Consonni, R.; Cagliani, L. R., Geographical characterization of polyfloral and acacia honeys by nuclear magnetic resonance and chemometrics. *J. Agric. Food Chem.* **2008**, *56*, 6873-6880.

REFERENCES

18.   Mihalov, J. J.; Marderosian, A. D.; Pierce, J. C., DNA identification of commercial ginseng samples. *J. Agric. Food Chem.* **2000**, *48*, 3744-3752.

19.   Cordella, C.; Moussa, I.; Martel, A. C.; Sbirrazzouli, N.; Lizzani-Cuvelier, L., Recent developments in food characterization and adulteration detection: Technique-oriented perspectives. *J. Agric. Food Chem.* **2002**, *50*, 1751-1764.

20.   Wikipedia 2013 Taiwan food scandal. http://en.wikipedia.org/wiki/2013_Taiwan_food_scandal (accessed 4th March 2014).

21.   Marcone, M. F.; Yada, R. Y., Evidence for the phosphorylation and glycosylation of the amaranth 11S globulin (Amaranthin). *J. Food Biochem.* **1997**, *21*, 341-369.

22.   Xing, Y. N.; Ni, H. G.; Chen, Z. Y., Semicarbazide in selected bird's nest products. *J. Food Prot.* **2012**, *75*, 1654-1659.

23.   Paydar, M.; Wong, Y. L.; Wong, W. F.; Hamdi, O. A. A.; Kadir, N. A.; Looi, C. Y., Prevalence of nitrite and nitrate contents and its effect on edible bird nest's color. *J. Food Sci.* **2013**, *78*, T1940-T1947.

24.   Aowphol, A.; Voris, H. K.; Feldheim, K. A.; Harnyuttanakorn, P.; Thirakhupt, K., Genetic homogeneity among colonies of the white-nest swiftlet (Aerodramus fuciphagus) in Thailand. *Zoolog. Sci.* **2008**, *25*, 372-380.

25.   Sankaran, R., The status and conservation of the Edible-nest Swiftlet (Collocalia fuciphaga) in the Andaman and Nicobar Islands. *Biol Conserv* **2001**, *97*, 283-294.

26.   Phach, N. Q.; Voisin, J. F., On an ecological form of the white-nest swiftlet Aerodramus fuciphagus (Aves, Apodidae) breeding in houses in Vietnam. *Revue D Ecologie-La Terre Et La Vie* **2007**, *62*, 49-57.

27.   Tompkins, D. M., Impact of nest-harvesting on the reproductive success of black-nest swiftlets Aerodramus maximus. *Wildl. Biol.* **1999**, *5*, 33-36.

28.   Stimpson, C. M., A 48,000year record of swiftlets (Aves: Apodidae) in North-western Borneo: Morphometric identifications and palaeoenvironmental implications. *Palaeogeogr., Palaeoclimatol., Palaeoecol.* **2013**, *374*, 132-143.

29.   Medway, L., The function of echonavigation among swiftlets. *Anim Behav.* **1967**, *15*, 416-420.

30.   Lee, P. L. M.; Clayton, D. H.; Griffiths, R.; Page, R. D. M., Does behavior reflect phylogeny in swiftlets (Aves: Apodidae)? A test using cytochrome b mitochondrial DNA sequences. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 7091-7096.

31.   Koon, L. C.; Cranbrook, E. o., Swiftlets of Borneo − Builders of edible bird nest. *Sabah: Natural History Publication (Borneo) SDN., B.H.D.* **2002**, pp. 1-171.

32.   Langham, N., Breeding biology of the edible-nest swiftlet Aerodramus-Fuciphagus. *Ibis* **1980**, *122*, 447-461.

33.   Ramji, M. F. S.; Koon, L. C.; Rahman, M. A., Roosting and nest-building Behaviour of the white-nest Swiftlet Aerodramus Fuciphagus (Thunberg) (Aves: Apodidae) in farmed colonies. *Raffles Bull. Zool.* **2013**, 225-235.

34.   But, P. P. H.; Jiang, R. W.; Shaw, P. C., Edible bird's nests-How do the red ones get red? *J. Ethnopharmacol.* **2013**, *145*, 378-380.

35.   Hobbs, J. J., Problems in the harvest of edible birds' nests in Sarawak and Sabah, Malaysian Borneo. *Biodivers Conserv* **2004**, *13*, 2209-2226.

36.   Ibrahim, S. H.; Teo, W. C.; Baharun , A., A study of suitable habitat for swiftlet farming. *UNIMAS E-Journal of Civil Engineering* **2009**, *1*.

37.   Mamduh, S. M.; Shakaff, A. Y. M.; Saad, S. M.; Kamarudin, K.; Kamarudin, L. M.; Zakaria, A.; Kamarudin, H.; Ezanuddin, A. M. M.; Sa'Ad, F. S. A.; Nooriman, W. M.; Abdullah, A. H., Odour and hazardous gas monitoring system for swiftlet farming using wireless sensor network (WSN). *Chemical Engineering Transactions* **2012**, *30*, 331-336.

38.   Online, T. S. Swiftlet industry in Sarawak ruffles feathers http://www.thestar.com.my/story.aspx/?file=%2f2009%2f12%2f11%2fnation%2f20091211123920 (4th March 2014).

39.   Leung, C. Y., Three billions market competition for edible bird's nest shop. *Economic Digest* **2004**, *1197*, 68 -69.

40.   Economic Transformation Programme Handbook In Department, P. M., Ed. Performance Management and Delivery unit Malaysia 2010; Vol. 15, pp 513-550.

41.   Syahir, F. A. S.; Shakaff, A. Y. M.; Zakaria, A.; Abdullah, M. Z.; Adom, A. H.; Ezanuddin, A. A. M. Edible bird nest shape quality assessment using machine vision system.*Proceedings - 3rd International Conference on Intelligent Systems Modelling and Simulation* **2012,** 325-329.

42.   Wong, R. S. Y., Edible bird's nest: Food or medicine? *Chin. J. Integr. Med.* **2013**, *19*, 643-649.

43. Kong, Y. C.; Keung, W. M.; Yip, T. T.; Ko, K. M.; Tsao, S. W.; Ng, M. H., Evidence that epidermal growth factor is present in swiftlet's (Collocalia) nest. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **1987**, *87*, 221-226.

44. Guo, C. T.; Takahashi, T.; Bukawa, W.; Takahashi, N.; Yagi, H.; Kato, K.; Hidari, K.; Miyamoto, D.; Suzuki, T.; Suzuki, Y., Edible bird's nest extract inhibits influenza virus infection. *Antiviral Res.* **2006**, *70*, 140-146.

45. Vimala, B.; Hussain, H.; Nazaimoon, W. M. W., Effects of edible bird's nest on tumour necrosis factor-alpha secretion, nitric oxide production and cell viability of lipopolysaccharide-stimulated RAW 264.7 macrophages. *Food Agric. Immunol.* **2012**, *23*, 303-314.

46. Matsukwa, N.; Matsumoyo, M.; Bukawa, W.; Chihi, H.; Nakayama, K.; Hara, H.; Tsukahara, T., Improvement of bone strength and dermal thickness due to dietary edible bird's nest extract in ovariectomized rats. *Biosci., Biotechnol., Biochem.* **2011**, *75*, 590-592.

47. Chua, K. H.; Lee, T. H.; Nagandran, K.; Md Yahaya, N. H.; Lee, C. T.; Tjih, E. T. T.; Abdul Aziz, R., Edible bird's nest extract as a chondro-protective agent for human chondrocytes isolated from osteoarthritic knee: In vitro study. *BMC Complement. Altern. Med.* **2013**, *13*.

48. Saengkrajang, W.; Matan, N.; Matan, N., Antimicrobial activities of the edible bird's nest extracts against food-borne pathogens. *Thai J Agric Sci* **2011**, *44*, 326-330.

49. Kim, K. C.; Kang, K. A.; Lim, C. M.; Park, J. H.; Jung, K. S.; Hyun, J. W., Water extract of edible bird's nest attenuated the oxidative stress-induced matrix metalloproteinase-1 by regulating the mitogen-activated protein kinase

and activator protein-1 pathway in human keratinocytes. *Journal of the Korean Society for Applied Biological Chemistry* **2012**, *55*, 347-354.

50.   Roh, K. B.; Lee, J.; Kim, Y. S.; Park, J.; Kim, J. H.; Park, D., Mechanisms of edible bird's nest extract-induced proliferation of human adipose-derived stem cells. *Evid. Based Complement. Alternat. Med.* **2012**, 1-11.

51.   Zainal Abidin, F.; Hui, C. K.; Luan, N. S.; Mohd Ramli, E. S.; Hun, L. T.; Abd Ghafar, N., Effects of edible bird's nest (EBN) on cultured rabbit corneal keratocytes. *BMC Complement. Altern. Med.* **2011**, *11*.

52.   Goh, D. L. M.; Chua, K. Y.; Chew, F. T.; Seow, T. K.; Ou, K. L.; Yi, F. C.; Lee, B. W., Immunochemical characterization of edible bird's nest allergens. *J. Allergy Clin. Immunol.* **2001**, *107*, 1082-1088.

53.   Hon, K. L. E.; Leung, T. F.; Kam, W. Y. C.; Lam, M. C. A.; Fok, T. F.; Ng, P. C., Dietary restriction and supplementation in children with atopic eczema. *Clin. Exp. Dermatol.* **2006**, *31*, 187-191.

54.   Ward, M. H.; Pan, W. H.; Cheng, Y. J.; Li, F. H.; Brinton, L. A.; Chen, C. J.; Hsu, M. M.; Chen, I. H.; Levine, P. H.; Yang, C. S.; Hildesheim, A., Dietary exposure to nitrite and nitrosamines and risk of nasopharyngeal carcinoma in Taiwan. *Int. J. Cancer* **2000**, *86*, 603-609.

55.   Ramli, N.; Azmi, S. M. N., Food safety governance: Standard operating procedure on controlling of nitrite level, handling and processing of edible bird's nest. *Australian Journal of Basic and Applied Sciences* **2012**, *6*, 301-305.

56.   Wishart, D. S., Metabolomics: applications to food science and nutrition research. *Trends in Food Science and Technology* **2008**, *19*, 482-493.

57.   Camin, F.; Larcher, R.; Nicolini, G.; Bontempo, L.; Bertoldi, D.; Perini, M.; Schlicht, C.; Schellenberg, A.; Thomas, F.; Heinrich, K.; Voerkelius, S.; Horacek, M.; Ueckermann, H.; Froeschl, H.; Wimmer, B.; Heiss, G.; Baxter, M.; Rossmann, A.; Hoogewerff, J., Isotopic and elemental data for tracing the origin of European olive oils. *J. Agric. Food Chem.* **2010**, *58*, 570-577.

58.   Stanimirova, I.; Üstün, B.; Cajka, T.; Riddelova, K.; Hajslova, J.; Buydens, L. M. C.; Walczak, B., Tracing the geographical origin of honeys based on volatile compounds profiles assessment using pattern recognition techniques. *Food Chem.* **2010**, *118*, 171-176.

59.   Camin, F.; Bontempo, L.; Heinrich, K.; Horacek, M.; Kelly, S. D.; Schlicht, C.; Thomas, F.; Monahan, F. J.; Hoogewerff, J.; Rossmann, A., Multi-element (H,C,N,S) stable isotope characteristics of lamb meat from different European regions. *Anal. Bioanal. Chem.* **2007**, *389*, 309-320.

60.   Hall, R. D., Food metabolomics: META-PHOR- A new European research initiative. *Agro Food Ind Hi-Tech* **2007**, *18*, 14-16.

61.   Mafra, I.; Ferreira, I. M. P. L. V. O.; Oliveira, M. B. P. P., Food authentication by PCR-based methods. *Eur. Food Res. Technol.* **2008**, *227*, 649-665.

62.   Bossier, P., Authentication of seafood products by DNA patterns. *J. Food Sci.* **1999**, *64*, 189-193.

63.   Laube, I.; Spiegelberg, A.; Butschke, A.; Zagon, J.; Schauzu, M.; Kroh, L.; Broll, H., Methods for the detection of beef and pork in foods using real-time polymerase chain reaction. *Int. J. Food Sci. Technol.* **2003**, *38*, 111-118.

64.   Hird, H.; Lloyd, J.; Goodier, R.; Brown, J.; Reece, P., Detection of peanut using real-time polymerase chain reaction. *Eur. Food Res. Technol.* **2003**, *217*, 265-268.

65.   Martinez, I.; Aursand, M.; Erikson, U.; Singstad, T. E.; Veliyulin, E.; Van Der Zwaag, C., Destructive and non-destructive analytical techniques for authentication and composition analyses of foodstuffs. *Trends in Food Science and Technology* **2003**, *14*, 489-498.

66.   Monakhova, Y. B.; Kuballa, T.; Lachenmeier, D. W., Chemometric methods in NMR spectroscopic analysis of food products. *J. Anal. Chem.* **2013**, *68*, 755-766.

67.   Le Gall, G.; Colquhoun, I. J.; Defernez, M., Metabolite profiling using H-1 NMR spectroscopy for quality assessment of green tea, Camellia sinensis (L.). *J. Agric. Food Chem.* **2004**, *52*, 692-700.

68.   Cuny, M.; Vigneau, E.; Le Gall, G.; Colquhoun, I.; Lees, M.; Rutledge, D. N., Fruit juice authentication by 1H NMR spectroscopy in combination with different chemometrics tools. *Anal. Bioanal. Chem.* **2008**, *390*, 419-427.

69.   Aursand, M.; Standal, I. B.; Praél, A.; McEvoy, L.; Irvine, J.; Axelson, D. E., 13C NMR pattern recognition techniques for the classification of atlantic salmon (Salmo salar L.) according to their wild, farmed, and geographical origin. *J. Agric. Food Chem.* **2009**, *57*, 3444-3451.

70.   Skoog, D. A.; Holler, F. J.; Crouch, S. R., *Principles of Instrumental Analysis*. Thomson Brooks: 2007; p 1-1039.

71.   Chan, E. C. Y.; Pasikanti, K. K.; Nicholson, J. K., Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry. *Nat. Protoc.* **2011**, *6*, 1483-1499.

72.   Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246*, 64-71.

73.   Passi, S.; Cataudella, S.; Di Marco, P.; De Simone, F.; Rastrelli, L., Fatty acid composition and antioxidant levels in muscle tissue of different Mediterranean marine species of fish and shellfish. *J. Agric. Food Chem.* **2002**, *50*, 7314-7322.

74.   Anastasaki, E.; Kanakis, C.; Pappas, C.; Maggi, L.; del Campo, C. P.; Carmona, M.; Alonso, G. L.; Polissiou, M. G., Geographical differentiation of saffron by GC-MS/FID and chemometrics. *Eur. Food Res. Technol.* **2009**, *229*, 899-905.

75.   Park, Y. J.; Kim, K. R.; Kim, J. H., Gas chromatographic organic acid profiling analysis of brandies and whiskeys for pattern recognition analysis. *J. Agric. Food Chem.* **1999**, *47*, 2322-2326.

76.   Vaclavik, L.; Schreiber, A.; Lacina, O.; Cajka, T.; Hajslova, J., Liquid chromatography-mass spectrometry-based metabolomics for authenticity assessment of fruit juices. *Metabolomics* **2012**, *8*, 793-803.

77.   Mattarucchi, E.; Stocchero, M.; Moreno-Rojas, J. M.; Giordano, G.; Reniero, F.; Guillou, C., Authentication of trappist beers by LC-MS fingerprints and multivariate data analysis. *J. Agric. Food Chem.* **2010**, *58*, 12089-12095.

78.   Aparicio-Ruiz, R.; Riedl, K. M.; Schwartz, S. J., Identification and quantification of metallo-chlorophyll complexes in bright green table olives by high-performance liquid chromatrography-mass spectrometry quadrupole/time-of-flight. *J. Agric. Food Chem.* **2011**, *59*, 11100-11108.

79.   Goodacre, R.; Broadhurst, D.; Smilde, A. K.; Kristal, B. S.; Baker, J. D.; Beger, R.; Bessant, C.; Connor, S.; Capuani, G.; Craig, A.; Ebbels, T.; Kell, D. B.; Manetti, C.; Newton, J.; Paternostro, G.; Somorjai, R.; Sjöström, M.; Trygg, J.; Wulfert, F., Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics* **2007**, *3*, 231-241.

80.   Trygg, J.; Holmes, E.; Lundstedt, T., Chemometrics in metabonomics. *J. Proteome Res.* **2007**, *6*, 469-479.

81.   Cloarec, O.; Dumas, M. E.; Trygg, J.; Craig, A.; Barton, R. H.; Lindon, J. C.; Nicholson, J. K.; Holmes, E., Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in 1H NMR spectroscopic metabonomic studies. *AnaCh* **2005**, *77*, 517-526.

82.   Van den Berg, R. A.; Hoefsloot, H. C. J.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J., Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **2006**, *7*.

83.   Ni, Y.; Mei, M.; Kokot, S., Resolution of high performance liquid chromatographic fingerprints of rhizoma curcumae by application of chemometrics. *J. Liq. Chromatogr. Related Technol.* **2011**, *34*, 1952-1964.

84.   Brereton, R. G., Chemometrics for Pattern Recognition. *John Wiley & Sons, Ltd, Chichester, UK.* 2009, pp 1-504.

85.   Eastment, H. T.; Krzaonowski, W. J., Cross-validatory choice of the number of components from a principal component analysis. *Technometrics* **1982**, *24*, 73-77.

86.   Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; J.Trygg; Wikström, C.; Wold, S., *Multi- and Megavariate Data Analysis Part I Basic Principles and Applications*. MKS Umetrics AB: 2006, pp.1-419.

87.   Gad, H. A.; El-Ahmady, S. H.; Abou-Shoer, M. I.; Al-Azizi, M. M., Application of chemometrics in authentication of herbal medicines: A review. *Phytochem. Anal.* **2013**, *24*, 1-24.

88.   Eide, I.; Zahlsen, K., A novel method for chemical fingerprinting of oil and petroleum products based on electrospray mass spectrometry and chemometrics. *Energy Fuels* **2005**, *19*, 964-967.

89.   Wold, S.; Sjostrom, M.; Eriksson, L., PLS-regression: a basic tool of chemometrics. *Chemometrics Intellig. Lab. Syst.* **2001**, *58*, 109-130.

90.   Berrueta, L. A.; Alonso-Salces, R. M.; Héberger, K., Supervised pattern recognition in food analysis. *J. Chromatogr.* **2007**, *1158*, 196-214.

91.   Bylesjo, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J. K.; Holmes, E.; Trygg, J., OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemometrics* **2006**, *20*, 341-351.

92.   Trygg, J.; Wold, S., Orthogonal projections to latent structures (O-PLS). *J. Chemometrics* **2002**, *16*, 119-128.

93.   Wiklund, S.; Johansson, E.; Sjöström, L.; Mellerowicz, E. J.; Edlund, U.; Shockcor, J. P.; Gottfries, J.; Moritz, T.; Trygg, J., Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *AnaCh* **2008**, *80*, 115-122.

94.   Charlton, A. J.; Farrington, W. H. H.; Brereton, P., Application of 1H NMR and multivariate statistics for screening complex mixtures: Quality

control and authenticity of instant coffee. *J. Agric. Food Chem.* **2002**, *50*, 3098-3103.

95.   Locatelli, M.; Coïsson, J. D.; Travaglia, F.; Cereti, E.; Garino, C.; D'Andrea, M.; Martelli, A.; Arlorio, M., Chemotype and genotype chemometrical evaluation applied to authentication and traceability of "tonda Gentile Trilobata" hazelnuts from Piedmont (Italy). *Food Chem.* **2011**, *129*, 1865-1873.

96.   Pereira, A. C.; Reis, M. S.; Saraiva, P. M.; Marques, J. C., Madeira wine ageing prediction based on different analytical techniques: UV-vis, GC-MS, HPLC-DAD. *Chemometrics Intellig. Lab. Syst.* **2011**, *105*, 43-55.

97.   Che Man, Y. B.; Syahariza, Z. A.; Mirghani, M. E. S.; Jinap, S.; Bakar, J., Analysis of potential lard adulteration in chocolate and chocolate products using Fourier transform infrared spectroscopy. *Food Chem.* **2005**, *90*, 815-819.

98.   Di Anibal, C. V.; Ruisánchez, I.; Callao, M. P., High-resolution 1H Nuclear Magnetic Resonance spectrometry combined with chemometric treatment to identify adulteration of culinary spices with Sudan dyes. *Food Chem.* **2011**, *124*, 1139-1145.

99.   Toh, D. F.; New, L. S.; Koh, H. L.; Chan, E. C. Y., Ultra-high performance liquid chromatography/time-of-flight mass spectrometry (UHPLC/TOFMS) for time-dependent profiling of raw and steamed Panax notoginseng. *J. Pharm. Biomed. Anal.* **2010**, *52*, 43-50.

100.   Aliferis, K. A.; Tarantilis, P. A.; Harizanis, P. C.; Alissandrakis, E., Botanical discrimination and classification of honey samples applying gas chromatography/mass spectrometry fingerprinting of headspace volatile compounds. *Food Chem.* **2010**, *121*, 856-862.

101.   Consonni, R.; Cagliani, L. R.; Stocchero, M.; Porretta, S., Triple concentrated tomato paste: Discrimination between Italian and Chinese products. *J. Agric. Food Chem.* **2009**, *57*, 4506-4513.

102.   Houdret, N.; Lhermitte, M.; Degand, P.; Roussel, P., Purification and chemical study of a glycoprotein from Collocalia. *Biochimie* **1975**, *57*, 603-608.

103.   Huda, M. Z. N.; Zuki, A. B. Z.; Azhar, K.; Goh, Y. M.; Suhaimi, H.; Hazmi, A. J. A.; Zairi, M. S., Proximate, elemental and fatty acid analysis of pre-processed edible birds' nest (Aerodramus fuciphagus): A comparison between regions and type of nest. *J. Food Technol.* **2008**, *6*, 39-44.

104.   Tran, B. N.; Richard, O.; Robin, S.; Robert, J.; Aldous, K. M., Use of methanol for the efficient extraction and analysis of melamine and cyanuric acid residues in dairy products and pet foods. *J. Agric. Food Chem.* **2010**, *58*, 101-107.

105.   Lin, J. R.; Zhou, H.; Lai, X. P.; Hou, Y.; Xian, X. M.; Chen, J. N.; Wang, P. X.; Zhou, L.; Dong, Y., Genetic identification of edible birds' nest based on mitochondrial DNA sequences. *Food Res. Int.* **2009**, *42*, 1053-1061.

106.   Zhang, S.; Lai, X.; Liu, X.; Li, Y.; Li, B.; Huang, X.; Zhang, Q.; Chen, W.; Lin, L.; Yang, G., Development of monoclonal antibodies and quantitative sandwich enzyme linked immunosorbent assay for the characteristic sialoglycoprotein of edible bird's nest. *J. Immunoassay Immunochem.* **2013**, *34*, 49-60.

107.   Liu, X. Q.; Lai, X. T.; Zhang, S. W.; Huang, X. L.; Lan, Q. X.; Li, Y.; Li, B. F.; Chen, W.; Zhang, Q. L.; Hong, D. Z.; Yang, G. W., Proteomic

profile of edible bird's nest proteins. *J. Agric. Food Chem.* **2012**, *60*, 12477-12481.

108. Pozsgay, V.; Jennings, H.; Kasper, D. L., 4,8-anhydro-N-acetylneuraminic acid. Isolation from edible bird's nest and structure determination. *Eur J Biochem* **1987**, *162*, 445-450.

109. Nakagawa, H.; Hama, Y.; Sumi, T.; Li, S. C.; Maskos, K.; Kalayanamitra, K.; Mizumoto, S.; Sugahara, K.; Li, Y. T., Occurrence of a nonsulfated chondroitin proteoglycan in the dried saliva of Collocalia swiftlets (edible bird's-nest). *Glycobiology* **2007**, *17*, 157-164.

110. Yagi, H.; Yasukawa, N.; Yu, S. Y.; Guo, C. T.; Takahashi, N.; Takahashi, T.; Bukawa, W.; Suzuki, T.; Khoo, K. H.; Suzuki, Y.; Kato, K., The expression of sialylated high-antennary N-glyeans in edible bird's nest. *Carbohydr. Res.* **2008**, *343*, 1373-1377.

111. Oda, M.; Ohta, S.; Suga, T.; Aoki, T., Study on food components: the structure of N-Linked asialo carbohydrate from the edible bird's nest built by collocalia fuciphaga. *J. Agric. Food Chem.* **1998**, *46*, 3047-3053.

112. Wieruszeski, J. M.; Michalski, J. C.; Montreuil, J.; Strecker, G.; Peterkatalinic, J.; Egge, H.; Vanhalbeek, H.; Mutsaers, J.; Vliegenthart, J. F. G., Structure of the monosialyl oligosaccharides derived from salivary gland mucin glycoproteins of the Chinese swiftlet (genus collocalia) - characterization of novel types of extended core structure, Gal-Beta(1- 3) GlcNAc-Beta(1- 6) GalNAc-Alpha1- 3)GalNAc(1-OL), and of chain termination, Gal-Alpha(1- 4) 0-1 Gal-Beta(1- 4) 2GlcNAc-Beta(1- .)E. *J. Biol. Chem.* **1987**, *262*, 6650-6657.

113.   Saengkrajang, W.; Matan, N.; Matan, N., Nutritional composition of the farmed edible bird's nest (Collocalia fuciphaga) in Thailand. *Journal of Food Composition and Analysis* **2013**, *31*, 41-45.

114.   Marcone, M. F., Characterization of the edible bird's nest the "Caviar of the East". *Food Res. Int.* **2005**, *38*, 1125-1134.

115.   Ma, F.; Liu, D., Extraction and determination of hormones in the edible bird's nest. *Asian J. Chem.* **2012**, *24*, 117-120.

116.   Norhayati, M. K.; Azman, O.; Wan Nazaimoon, W. M., Preliminary study of the nutritional content of Malaysian edible bird's nest. *Malaysian Journal of Nutrition* **2010**, *16*, 389-396.

117.   Ekeanyanwu, R. C., Evaluation of the crude protein and amino acid composition of Nigerian Monodora myristica (Ehuru). *Pakistan Journal of Nutrition* **2013**, *12*, 219-223.

118.   Sun, Y.; Lige, B.; Van Huystee, R. B., HPLC determination of the sugar compositions of the glycans on the cationic peanut peroxidase. *J. Agric. Food Chem.* **1997**, *45*, 4196-4200.

119.   Jong, C. H.; Tay, K. M.; Lim, C. P., Application of the fuzzy failure mode and effect analysis methodology to edible bird nest processing. *Comput Electron Agric* **2013**, *96*, 90-108.

120.   Fountoulakis, M.; Lahm, H. W., Hydrolysis and amino acid composition analysis of proteins. *J. Chromatogr.* **1998**, *826*, 109-134.

121.   Weiss, M.; Manneberg, M.; Juranville, J. F.; Lahm, H. W.; Fountoulakis, M., Effect of the hydrolysis method on the determination of the amino acid composition of proteins. *J. Chromatogr.* **1998**, *795*, 263-275.

122.   Kwon, H.; Kim, J., Determination of monosaccharides in glycoproteins by reverse-phase high-performance liquid chromatography. *Anal Biochem* **1993**, *215*, 243-252.

123.   Watson, J. D.; Crick, F. H. C., Genetical implications of the structure of deoxyribonucleic acid. *Nature* **1953**, *171*, 964-967.

124.   Shama, N.; Bai, S. W.; Chung, B. C.; Jung, B. H., Quantitative analysis of 17 amino acids in the connective tissue of patients with pelvic organ prolapse using capillary electrophoresis-tandem mass spectrometry. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* **2008**, *865*, 18-24.

125.   Kirdar, A. O.; Green, K. D.; Rathore, A. S., Application of multivariate data analysis for identification and successful resolution of a root cause for a bioprocessing application. *Biotechnol. Prog.* **2008**, *24*, 720-726.

126.   George, J. P.; Chen, Z.; Shaw, P., Fault detection of drinking water treatment process using PCA and hotelling's T2 chart. *World Academy of Science, Engineering and Technology* **2009**, *38*, 1061-1066.

127.   Nozal, M. J.; Bernal, J. L.; Toribio, M. L.; Diego, J. C.; Ruiz, A., Rapid and sensitive method for determining free amino acids in honey by gas chromatography with flame ionization or mass spectrometric detection. *J. Chromatogr.* **2004**, *1047*, 137-146.

128.   Zhang, P.; Wang, Z.; Xie, M.; Nie, W.; Huang, L., Detection of carbohydrates using a pre-column derivatization reagent 1-(4-isopropyl) phenyl-3-methyl-5-pyrazolone by high-performance liquid chromatography coupled with electrospray ionization mass spectrometry. *Journal of*

Chromatography B: Analytical Technologies in the Biomedical and Life Sciences **2010**, *878*, 1135-1144.

129.   Honda, S.; Akao, E.; Suzuki, S.; Okuda, M.; Kakehi, K.; Nakamura, J., High performance liquid chromatography of reducing carbohydrates as strongly ultraviolet absorbing and electrochemically sensitive 1-phenyl-3-methyl-5-pyrazolone derivatives. *Anal Biochem* **1989**, *180*, 351-357.

130.   De Ruiter, G. A.; Schols, H. A.; Voragen, A. G. J.; Rombouts, F. M., Carbohydrate analysis of water-soluble uronic acid-containing polysaccharides with high-performance anion-exchange chromatography using methanolysis combined with TFA hydrolysis is superior to four other methods. *Anal Biochem* **1992**, *207*, 176-185.

131.   Lv, Y.; Yang, X.; Zhao, Y.; Ruan, Y.; Yang, Y.; Wang, Z., Separation and quantification of component monosaccharides of the tea polysaccharides from Gynostemma pentaphyllum by HPLC with indirect UV detection. *Food Chem.* **2009**, *112*, 742-746.

132.   Mason, B. S.; Slover, H. T., A gas chromatographic method for the determination of sugars in foods. *J. Agric. Food Chem.* **1958**, *6*, 551-554.

133.   Widmer, W., Analysis of biomass sugars and galacturonic acid by gradient anion exchange chromatography and pulsed amperometric detection without post-column addition. *BiotL* **2011**, *33*, 365-368.

134.   McRae, G.; Monreal, C. M., LC-MS/MS quantitative analysis of reducing carbohydrates in soil solutions extracted from crop rhizospheres. *Anal. Bioanal. Chem.* **2011**, *400*, 2205-2215.

135.   Rühmann, B.; Schmid, J.; Sieber, V., Fast carbohydrate analysis via liquid chromatography coupled with ultra violet and electrospray ionization ion trap detection in 96-well format. *J. Chromatogr.* **2014**, 1350, 44-50.

136.   Wu, X.; Jiang, W.; Lu, J.; Yu, Y.; Wu, B., Analysis of the monosaccharide composition of water-soluble polysaccharides from Sargassum fusiforme by high performance liquid chromatography/electrospray ionisation mass spectrometry. *Food Chem.* **2014**, *145*, 976-983.

137.   Hammad, L. A.; Saleh, M. M.; Novotny, M. V.; Mechref, Y., Multiple-reaction monitoring liquid chromatography mass spectrometry for monosaccharide compositional analysis of glycoproteins. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1224-1234.

138.   Sun, Z.; Song, C.; Xia, L.; Wang, X.; Suo, Y.; You, J., Comprehensive comparisons between 1-Phenyl-3-methyl-5-pyrazolones, 1-(4-Methoxyphenyl)-3-methyl-5-pyrazolones and 1-(2-Naphthyl)-3-methyl-5-pyrazolones as labeling reagents used in LC-DAD-ESI-MS-MS analysis of neutral aldoses and uronic acids. *Chromatographia* **2010**, *71*, 789-797.

139.   Yet, M. G.; Chin, C. C. Q.; Wold, F., The covalent structure of individual N-linked glycopeptides from ovomucoid and asialofetuin. *J. Biol. Chem.* **1988**, *263*, 111-117.

140.   Wu, W.; Hamase, K.; Kigucki, M.; Yamamoto, K.; Zaitsu, K., Reversed-phase HPLC of monosaccharides in glycoproteins derivatized with aminopyrazine with fluorescence detection. *Anal. Sci.* **2000**, *16*, 919-922.

141.   Salchert, K.; Pompe, T.; Sperling, C.; Werner, C., Quantitative analysis of immobilized proteins and protein mixtures by amino acid analysis. *J. Chromatogr.* **2003**, *1005*, 113-122.

142.   Ceballos, L. S.; Morales, E. R.; Adarve, G. D. T.; Castro, J. D.; Martinez, L. P.; Sampelayo, M. R. S., Composition of goat and cow milk produced under similar conditions and analyzed by identical methodology. *Journal of Food Composition and Analysis* **2009**, *22*, 322-329.

143.   Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C., Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *AnaCh* **2006**, *78*, 2262-2267.

144.   Chaturvedi, P.; Warren, C. D.; RuizPalacios, G. M.; Pickering, L. K.; Newburg, D. S., Milk oligosaccharide profiles by reversed-phase HPLC of their perbenzoylated derivatives. *Anal Biochem* **1997**, *251*, 89-97.

145.   Lunven, P.; Le Clement De St Marcq, C.; Carnovale, E.; Fratoni, A., Amino acid composition of hen's egg. *Br J Nutr* **1973**, *30*, 189-194.

146.   Liu, F.; Ong, E. S.; Li, S. F. Y., A green and effective approach for characterisation and quality control of Chrysanthemum by pressurized hot water extraction in combination with HPLC with UV absorbance detection. *Food Chem.* **2013**, *141*, 1807-1813.

147.   Agostoni, C.; Carratu, B.; Boniglia, C.; Riva, E.; Sanzini, E., Free amino acid content in standard infant formulas: Comparison with human milk. *J. Am. Coll. Nutr.* **2000**, *19*, 434-438.

148.   Ang, H. G.; Kwik, W. L.; Theng, C. Y., Total protein and amino acids contents of some common foods in South East Asia. *Food Chem.* **1984**, *15*, 243-255.

149.   Wu, G. Y., Amino acids: metabolism, functions, and nutrition. *Amino Acids* **2009**, *37*, 1-17.

150. Neuman, R. E.; Logan, M. A., The determination of collagen and elastin in tissues. *The Journal of Biological Chemistry* **1950**, *186*, 549-556.

151. Katsanos, C. S.; Kobayashi, H.; Sheffield-Moore, M.; Aarsland, A.; Wolfe, R. R., A high proportion of leucine is required for optimal stimulation of the rate of muscle protein synthesis by essential amino acids in the elderly. *American Journal of Physiology-Endocrinology and Metabolism* **2006**, *291*, E381-E387.

152. Lowe, J. B.; Marth, J. D., A genetic approach to mammalian glycan function. *Annu Rev Biochem.* **2003**, 72, 643-691.

153. Foye, O. T.; Uni, Z.; Ferket, P. R., Effect of in ovo feeding egg white protein, beta-hydroxy-beta-methylbutyrate, and carbohydrates on glycogen status and neonatal growth of turkeys. *Poult. Sci.* **2006**, *85*, 1185-1192.

154. Klein, C. J., Nutrient requirements for preterm infant formulas. *J. Nutr.* **2002**, *132*, 1395S-1577S.

155. Hascoët, J. M.; Hubert, C.; Rochat, F.; Legagneur, H.; Gaga, S.; Emady-Azar, S.; Steenhout, P. G., Effect of formula composition on the development of infant gut microbiota. *J. Pediatr. Gastroenterol. Nutr.* **2011**, *52*, 756-762.

156. Yang, M.; Cheung, S. H.; Li, S. C.; Cheung, H. Y., Establishment of a holistic and scientific protocol for the authentication and quality assurance of edible bird's nest. *Food Chem.* **2014**, *151*, 271-278.

157. Griffin, J. L., Metabolic profiles to define the genome: Can we hear the phenotypes? *Philosophical Transactions of the Royal Society B: Biological Sciences* **2004**, *359*, 857-871.

158.   Liu, F.; Gan, P. P.; Wu, H.; Woo, W. S.; Ong, E. S.; Li, S. F. Y., A combination of metabolomics and metallomics studies of urine and serum from hypercholesterolaemic rats after berberine injection. *Anal. Bioanal. Chem.* **2012**, *403*, 847-856.

159.   Chan, E. C. Y.; Yap, S. L.; Lau, A. J.; Leow, P. C.; Toh, D. F.; Koh, H. L., Ultra-performance liquid chromatography/time-of-flight mass spectrometry based metabolomics of raw and steamed Panax notoginseng. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 519-528.

160.   Xu, C. J.; Liang, Y. Z.; Chau, F. T.; Heyden, Y. V., Pretreatments of chromatographic fingerprints for quality control of herbal medicines. *J. Chromatogr.* **2006**, *1134*, 253-259.

161.   Fang, G.; Goh, J. Y.; Tay, M.; Lau, H. F.; Li, S. F. Y., Characterization of oils and fats by 1H NMR and GC/MS fingerprinting: Classification, prediction and detection of adulteration. *Food Chem.* **2013**, *138*, 1461-1469.

162.   Ong, E. S.; Chor, C. F.; Zou, L.; Ong, C. N., A multi-analytical approach for metabolomic profiling of zebrafish (Danio rerio) livers. *Molecular BioSystems* **2009**, *5*, 288-298.

163.   Issaq, H. J.; Van, Q. N.; Waybright, T. J.; Muschik, G. M.; Veenstra, T. D., Analytical and statistical approaches to metabolomics research. *J. Sep. Sci.* **2009**, *32*, 2183-2199.

164.   Zhang, W.; Tan, N. G. J.; Li, S. F. Y., NMR-based metabolomics and LC-MS/MS quantification reveal metal-specific tolerance and redox homeostasis in Chlorella vulgaris. *Molecular BioSystems* **2014**, *10*, 149-160.

165.   Pasikanti, K. K.; Ho, P. C.; Chan, E. C. Y., Development and validation of a gas chromatography/mass spectrometry metabonomic platform

for the global profiling of urinary metabolites. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 2984-2992.

166.   Liang, H.; Bao, F.; Dong, X.; Tan, R.; Zhang, C.; Lu, Q.; Cheng, Y., Antibacterial thymol derivatives isolated from Centipeda minima. *Molecules* **2007**, *12*, 1606-1613.

167.   Divito, E. B.; Davic, A. P.; Johnson, M. E.; Cascio, M., Electrospray ionization and collision induced dissociation mass spectrometry of primary fatty acid amides. *AnaCh* **2012**, *84*, 2388-2394.

168.   Van Den Ouweland, J. M. W.; Vogeser, M.; Bächer, S., Vitamin D and metabolites measurement by tandem mass spectrometry. *Reviews in Endocrine and Metabolic Disorders* **2013**, *14*, 159-184.

169.   Johnson, D. W.; Ten Brink, H. J.; Jakobs, C., A rapid screening procedure for cholesterol and dehydrocholesterol by electrospray ionization tandem mass spectrometry. *J. Lipid Res.* **2001**, *42*, 1699-1705.

170.   Bisogno, T.; Sepe, N.; De Petrocellis, L.; Mechoulam, R.; Di Marzo, V., The sleep inducing factor oleamide is produced by mouse neuroblastoma cells. *Biochem Biophys Res Commun* **1997**, *239*, 473-479.

171.   Martínez-González, D.; Bonilla-Jaime, H.; Morales-Otal, A.; Henriksen, S. J.; Velázquez-Moctezuma, J.; Prospéro-García, O., Oleamide and anandamide effects on food intake and sexual behavior of rats. *Neurosci. Lett.* **2004**, *364*, 1-6.

172.   Wakamatsu, K.; Masaki, T.; Itoh, F.; Kondo, K.; Sudo, K., Isolation of fatty acid amide as an angiogenic principle from bovine mesentery. *Biochem Biophys Res Commun* **1990**, *168*, 423-429.

173.   Mitchell, C. A.; Davies, M. J.; Grounds, M. D.; McGeachie, J. K.; Crawford, G. J.; Hong, Y.; Chirila, T. V., Enhancement of neovascularization in regenerating skeletal muscle by the sustained release of erucamide from a polymer matrix. *J. Biomater. Appl.* **1996**, *10*, 230-249.

174.   Lips, P.; Duong, T.; Oleksik, A.; Black, D.; Cummings, S.; Cox, D.; Nickelsen, T., A global study of vitamin D status and parathyroid function in postmenopausal women with osteoporosis: Baseline data from the multiple outcomes of raloxifene evaluation clinical trial. *J. Clin. Endocrinol. Metab.* **2001**, *86*, 1212-1221.

175.   Kikuchi, K.; Matahira, Y., Oral N-acetylglucosamine supplementation improves skin conditions of female volunteers: Clinical evaluation by a microscopic three-dimensional skin surface analyzer. *Journal of Applied Cosmetology* **2002**, *20*, 143-152.

176.   Parmar, K.; Suthar, B.; Prajapati, S.; Suthar, A., Synthesis and biological activity of novel 1,3,5-trisubstituted 1,2,4-triazole derivatives. *J. Heterocycl. Chem.* **2010**, *47*, 156-161.

177.   Pereira, G. E.; Gaudillere, J. P.; Van Leeuwen, C.; Hilbert, G.; Lavialle, O.; Maucourt, M.; Deborde, C.; Moing, A.; Rolin, D., 1H NMR and chemometrics to characterize mature grape berries in four wine-growing areas in Bordeaux, France. *J. Agric. Food Chem.* **2005**, *53*, 6382-6389.

178.   Lee, E.-J.; Rustem, S.; Weljie, A. M.; Vogel, H. J.; Facchini, P. J.; Park, S.-U.; Kim, Y.-K.; Yang, T.-J., Quality assessment of ginseng by1H NMR metabolite fingerprinting and profiling analysis. *J. Agric. Food Chem.* **2009**, *57*, 7513-7522.

179.   Abdullah, S. A.; Hezri, A. A., From forest landscape to agricultural landscape in the developing tropical country of Malaysia: Pattern, process, and their significance on policy. *Environ Manage* **2008**, *42*, 907-917.

180.   Seeboonruang, U., A statistical assessment of the impact of land uses on surface water quality indexes. *J. Environ. Manage.* **2012**, *101*, 134-142.

181.   Lourie, S. A.; Tompkins, D. M., The diets of Malaysian swiftlets. *Ibis* **2000**, *142*, 596-602.

182.   Tres, A.; Ruiz-Samblas, C.; Van Der Veer, G.; Van Ruth, S. M., Geographical provenance of palm oil by fatty acid and volatile compound fingerprinting techniques. *Food Chem.* **2013**, *137*, 142-150.

183.   Zhou, Z.; Blanchard, C.; Helliwell, S.; Robards, K., Fatty acid composition of three rice varieties following storage. *J. Cereal Sci.* **2003**, *37*, 327-335.

184.   Koh, L. P.; Wilcove, D. S., Oil palm: Disinformation enables deforestation. *Trends in Ecology and Evolution* **2009**, *24*, 67-68.

185.   Divito, E. B.; Cascio, M., Metabolism, physiology, and analyses of primary fatty acid amides. *Chem. Rev.* **2013**, *113*, 7343-7353.

186.   Viruhpintu, S.; Thirakhupt, K.; Pradatsundarasar, A.; Poonswad, P., Nest-site characteristics of the edible-nest swiftlet Aerodramus fuciphagus (Thunberg, 1812) at Si-Ha Islands, Phattalung Province, Thailand. *The Natural History Journal of Chulalongkorn University* **2002**, *2*, 31-35.

187.   Saxena, J.; Munimbazi, C.; Bullerman, L. B., Relationship of mould count, ergosterol and ochratoxin A production. *Int. J. Food Microbiol.* **2001**, *71*, 29-34.

188.   Thevissen, K.; Hillaert, U.; Meert, E. M. K.; Chow, K. K.; Cammue, B. P. A.; Van Calenbergh, S.; François, I. E. J. A., Fungicidal activity of truncated analogues of dihydrosphingosine. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 3728-3730.

189. Newsweek New program seeks out food fraud. http://www.newsweek.com/2010/02/07/the-fake-food-detectives.html (3th March 2014).

190.   Wu, Y. J.; Chen, Y.; Wang, B.; Bai, L. Q.; Han, W. R.; Ge, Y. Q.; Yuan, F., Application of SYBRgreen PCR and 2DGE methods to authenticate edible bird's nest food. *Food Res. Int.* **2010**, *43*, 2020-2026.

191.   Asensio, L.; González, I.; García, T.; Martín, R., Determination of food authenticity by enzyme-linked immunosorbent assay (ELISA). *Food Control* **2008**, *19*, 1-8.

192.   Zhang, S. W.; Lai, X. T.; Liu, X. Q.; Li, Y.; Li, B. F.; Huang, X. L.; Zhang, Q. L.; Chen, W.; Lin, L.; Yang, G. W., Competitive enzyme-linked immunoassay for sialoglycoprotein of edible bird's nest in food and cosmetics. *J. Agric. Food Chem.* **2012**, *60*, 3580-3585.

193.   Fajardo, V.; González Isabel, I.; Rojas, M.; García, T.; Martín, R., A review of current PCR-based methodologies for the authentication of meats from game animal species. *Trends Food Sci. and Tech.* **2010**, *21*, 408-421.

194.   Lüthy, J., Detection strategies for food authenticity and genetically modified foods. *Food Control* **1999**, *10*, 359-361.

195.   Hamzah, Z.; Jeyaraman, S.; Ibrahim, N. H.; Hashim, O.; Lee, B. B.; Hussin, K., A rapid technique to determine purity of edible bird nest. *Adv. Environ. Biol.* **2013**, *7*, 3758-3765.

196.   Teo, P.; Ma, F.; Liu, D., Evaluation of taurine by HPTLC reveals the mask of adulterated edible bird's nest. *J. of Chem.* **2013**, art no 325372

197.   Yu-Qin, Y.; Liang, X.; Hua, W.; Hui-Xing, Z.; Xin-Fang, Z.; Bu-Sen, L., Determination of edible bird's nest and its products by gas chromatography. *J. Chromtogr. Sci.* **2000**, *38*, 27-32.

198.   Tung, C. H.; Pan, J. Q.; Chang, H. M.; Chou, S. S., Authentic determination of bird's nests by saccharides profile. *J. Food Drug Anal.* **2008**, *16*, 86-91.

199.   Kvasnička, F., Capillary electrophoresis in food authenticity. *J. Sep. Sci.* **2005**, *28*, 813-825.

200.   Cha, M. H.; Heo, J. Y.; Lee, C.; Lo, Y. M.; Moon, B., Quality and sensory characterization of white jelly mushroom (Tremella fuciformis) as a meat substitute in pork patty formulation. *J. Food Process. Preserv.* **2012**.

201.   Dixon, P. S., The structure and development of the thallus in the British species of Gelidium and Pterocladia. *Ann. Bot.* **1958**, *22*, 353-368.

202.   Hickman, D.; Sims, T. J.; Miles, C. A.; Bailey, A. J.; De Mari, M.; Koopmans, M., Isinglass/collagen: Denaturation and functionality. *J. Biotechnol.* **2000**, *79*, 245-257.

203.   Munck, L.; Nørgaard, L.; Engelsen, S. B.; Bro, R.; Andersson, C. A., Chemometrics in food science - A demonstration of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance. *Chemometrics Intellig. Lab. Syst.* **1998**, *44*, 31-60.

204.   Bertoncelj, J.; Polak, T.; Kropf, U.; Korošec, M.; Golob, T., LC-DAD-ESI/MS analysis of flavonoids and abscisic acid with chemometric approach for the classification of Slovenian honey. *Food Chem.* **2011**, *127*, 296-302.

205. Lutter, P.; Savoy-Perroud, M. C.; Campos-Gimenez, E.; Meyer, L.; Goldmann, T.; Bertholet, M. C.; Mottier, P.; Desmarchelier, A.; Monard, F.; Perrin, C.; Robert, F.; Delatour, T., Screening and confirmatory methods for the determination of melamine in cow's milk and milk-based powdered infant formula: Validation and proficiency-tests of ELISA, HPLC-UV, GC-MS and LC-MS/MS. *Food Control* **2011**, *22*, 903-913.

206. Devreese, M.; De Baere, S.; De Backer, P.; Croubels, S., Quantitative determination of several toxicological important mycotoxins in pig plasma using multi-mycotoxin and analyte-specific high performance liquid chromatography-tandem mass spectrometric methods. *J. Chromatogr.* **2012**, *1257*, 74-80.

207. Waters, N. J.; Holmes, E.; Williams, A.; Waterfield, C. J.; Duncan Farrant, R.; Nicholson, J. K., NMR and pattern recognition studies on the time-related metabolic effects of α-naphthylisothiocyanate on liver, urine, and plasma in the rat: An integrative metabonomic approach. *Chem. Res. Toxicol.* **2001**, *14*, 1401-1412.

208. Constantinou, M. A.; Papakonstantinou, E.; Spraul, M.; Sevastiadou, S.; Costalos, C.; Koupparis, M. A.; Shulpis, K.; Tsantili-Kakoulidou, A.; Mikros, E., 1H NMR-based metabonomics for the diagnosis of inborn errors of metabolism in urine. *Anal. Chim. Acta* **2005**, *542*, 169-177.

209. Lavine, B. K., Pattern recognition. *Crit. Rev. Anal. Chem.* **2006**, *36*, 153-161.

210. Parsons, H. M.; Ludwig, C.; Günther, U. L.; Viant, M. R., Improved classification accuracy in 1- and 2-dimensional NMR metabolomics data using

the variance stabilising generalised logarithm transformation. *BMC Bioinformatics* **2007**, *8*.

211.   Defernez, M.; Le Gall, G., Strategies for data handling and statistical analysis in metabolomics studies. In 2013; Vol. 67, pp 493-555.

212.   Chan, P. H.; Zheng, K. Y. Z.; Tsim, K. W. K.; Lam, H., Metabonomic analysis of water extracts from Chinese and American ginsengs by 1H nuclear magnetic resonance: identification of chemical profile for quality control. *Chinese Medicine (United Kingdom)* **2012**, *7*.

213.   Tay, M.; Fang, G.; Chia, P. L.; Li, S. F. Y., Rapid screening for detection and differentiation of detergent powder adulteration in infant milk formula by LC-MS. *Forensic Sci. Int.* **2013**, *232*, 32-39.

214.   Casale, M.; Casolino, C.; Oliveri, P.; Forina, M., The potential of coupling information using three analytical techniques for identifying the geographical origin of Liguria extra virgin olive oil. *Food Chem.* **2010**, *118*, 163-170.

215.   Jiang, Z.; Liu, F.; Ong, E. S.; Li, S. F. Y., Metabolic profile associated with glucose and cholesterol lowering effects of berberine in Sprague-Dawley rats. *Metabolomics* **2012**, *8*, 1052-1068.

216.   Xiao, C.; Dai, H.; Liu, H.; Wang, Y.; Tang, H., Revealing the metabonomic variation of rosemary extracts using 1H NMR spectroscopy and multivariate data analysis. *J. Agric. Food Chem.* **2008**, *56*, 10142-10153.

217.   Pasikanti, K. K.; Esuvaranathan, K.; Hong, Y.; Ho, P. C.; Mahendran, R.; Raman Nee Mani, L.; Chiong, E.; Chan, E. C. Y., Urinary metabotyping of bladder cancer using two-dimensional gas chromatography time-of-flight mass spectrometry. *J. Proteome Res.* **2013**, *12*, 3865-3873.

218. Tarachiwin, L.; Koichi, U.; Kobayashi, A.; Fukusaki, E., 1H NMR based metabolic profiling in the evaluation of Japanese green tea quality. *J. Agric. Food Chem.* **2007**, *55*, 9330-9336.

219. Ochi, H.; Naito, H.; Iwatsuki, K.; Bamba, T.; Fukusaki, E., Metabolomics-based component profiling of hard and semi-hard natural cheeses with gas chromatography/time-of-flight-mass spectrometry, and its application to sensory predictive modeling. *J. Biosci. Bioeng.* **2012**, *113*, 751-758.

220. Cozzolino, D.; Cynkar, W. U.; Shah, N.; Smith, P., Multivariate data analysis applied to spectroscopy: Potential application to juice and fruit quality. *Food Res. Int.* **2011**, *44*, 1888-1896.

221. Ganti, S.; Taylor, S. L.; Aboud, O. A.; Yang, J.; Evans, C.; Osier, M. V.; Alexander, D. C.; Kim, K.; Weiss, R. H., Kidney tumor biomarkers revealed by simultaneous multiple matrix metabolomics analysis. *Cancer Res.* **2012**, *72*, 3471-3479.

222. Eriksson, L.; Hermens, J. L.; Johansson, E.; Verhaar, H. J.; Wold, S., Multivariate analysis of aquatic toxicity data with PLS. *Aquat Sci* **1995**, *57*, 217-241.

223. Udenigwe, C. C.; Aluko, R. E., Chemometric analysis of the amino acid requirements of antioxidant food protein Hydrolysates. *Int.l J.l Mol. Sci.* **2011**, *12*, 3148-3161.

224. Beveridge, J. M. R.; Lucas, C. C., Amino acids of isinglass. *J. Biol. Chem.* **1944**, *155*, 547-556.

225. Lai, D. M.; Deng, L.; Li, J.; Liao, B.; Guo, Q. X.; Fu, Y., Hydrolysis of cellulose into glucose by magnetic solid acid. *ChemSusChem* **2011**, *4*, 55-58.

226.   Weber, P.; Steinhart, H.; Paschke, A., Characterization, antigenicity and detection of fish gelatine and isinglass used as processing aids in wines. *Food Addit. Contam. Part A Chem. Ana. Control Expo. Risk Assess.* **2010**, *27*, 273-282.

227.   Andrade, P. B.; Barbosa, M.; Matos, R. P.; Lopes, G.; Vinholes, J.; Mouga, T.; Valentão, P., Valuable compounds in macroalgae extracts. *Food Chem.* **2013**, *138*, 1819-1828.

228.   Lu, C. Y.; Deng, Y.; Mei, L.; Guo, D. L., Studies on chemical constituents of Gracilaria lemaneiformis. *Chin. Tradi. Herbal Drugs* **2011**, *42*, 1069-1071.

229.   Gessner, M. O.; Schmitt, A. L., Use of solid-phase extraction to determine ergosterol concentrations in plant tissue colonized by fungi. *Appl Environ Microbiol* **1996**, *62*, 415-419.

230.   Kim, J.; Kim, D. N.; Lee, S. H.; Yoo, S. H.; Lee, S., Correlation of fatty acid composition of vegetable oils with rheological behaviour and oil uptake. *Food Chem.* **2010**, *118*, 398-402.

231.   Aparicio, R.; Aparicio-Ruíz, R., Authentication of vegetable oils by chromatographic techniques. *J. Chromatogr.* **2000**, *881*, 93-104.

APPENDICES

Appendix 1 Summary of the mass spectrometer parameter values of MRM transitions

Compound Name	Precursor ion (m/z)	Product ion (m/z)	DP (V)	CE (V)	CXP (V)	EP (V)	FP (V)
Mannose	511.4	175.1	95	34	2	10	400
		187.1	95	50	3	10	400
Glucose	511.4	175.1	95	34	2	10	400
		187.1	95	50	3	10	400
Galactose	511.4	175.1	95	34	2	10	400
		187.1	95	50	3	10	400
Rhamnose	495.3	175.1	100	33	2	8	400
		217.2	100	42	3	8	400
Fucose	495.3	175.1	100	33	2	8	400
		217.2	100	42	3	8	400
Ribose	481.2	175.0	100	34	2	9	400
		187.2	100	46	3	9	400
Xylose	481.2	175.0	100	34	2	9	400
		187.2	100	46	3	9	400
D-glucose-$^{13}C_6$	517.2	175.1	90	38	12	10	400
		188.1	90	48	12	10	400

Appendix 2 PCA score plots subjected to different scaling methods for the identification of EBN

(A) Amino acid
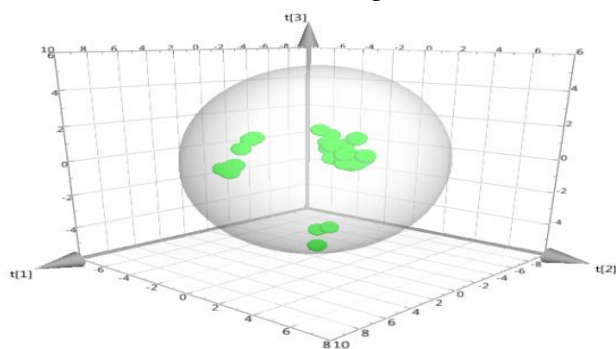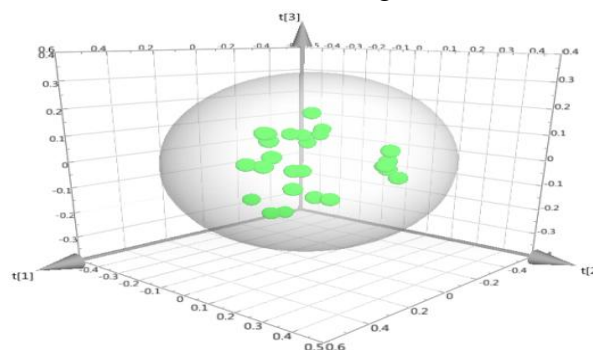
| (I) Uv scaling | (II) Par scaling | (III) No scaling |
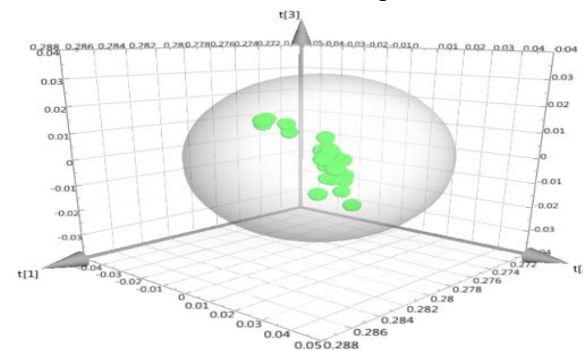
(B) Monosaccharide
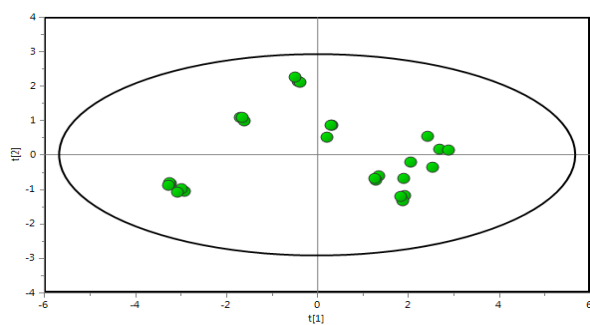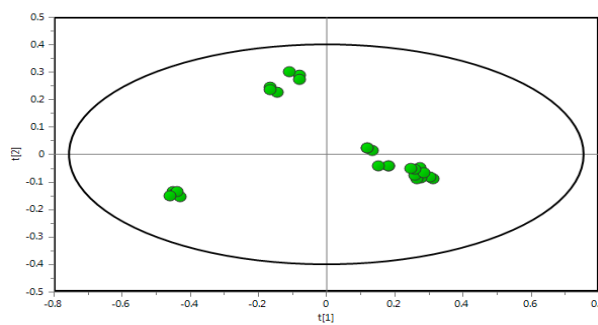
| (I) Uv scaling | (II) Par scaling | (III) No scaling |
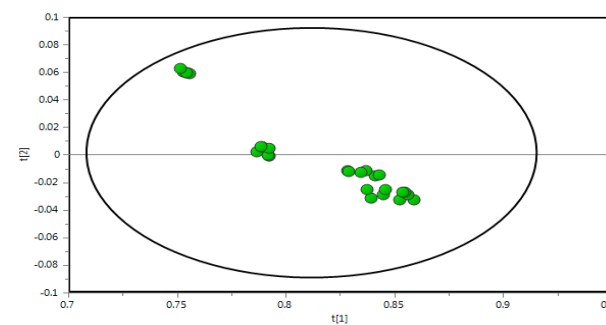
Appendix 3 Hotelling T2 range plots for the identification of EBN
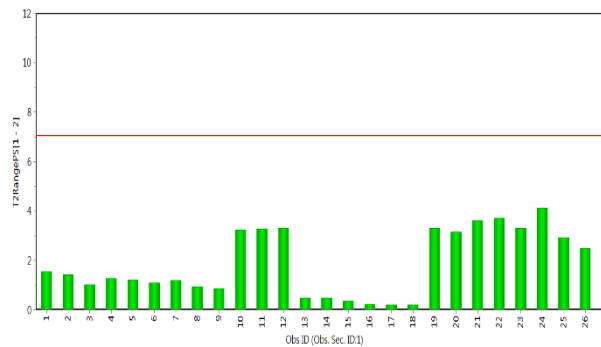
(A) Amino acid
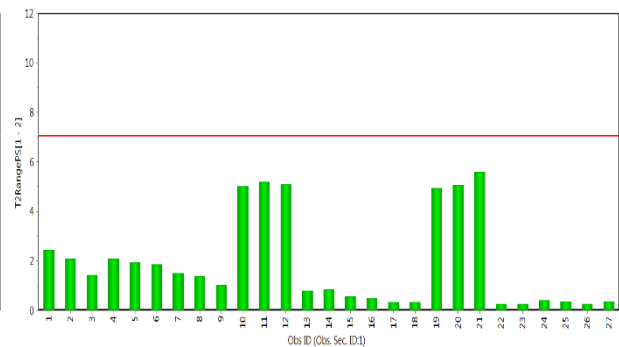
(B) Monosaccharide

Appendix 4 Contribution plots based on the amino acid data

(A) Chicken egg yolk
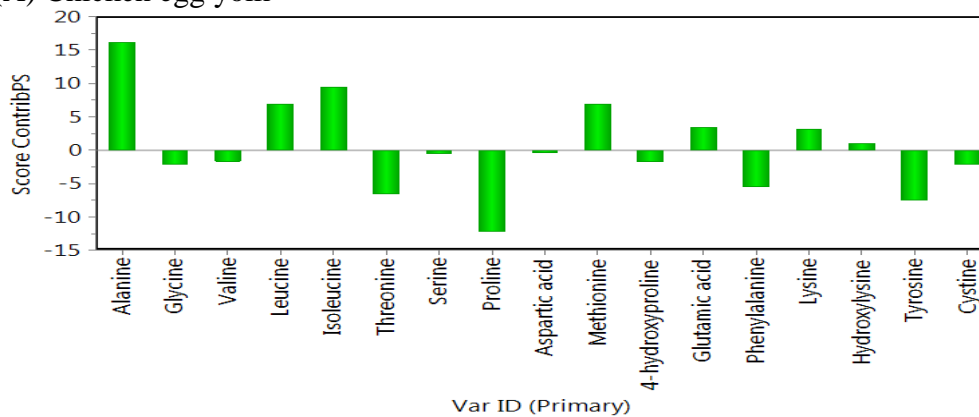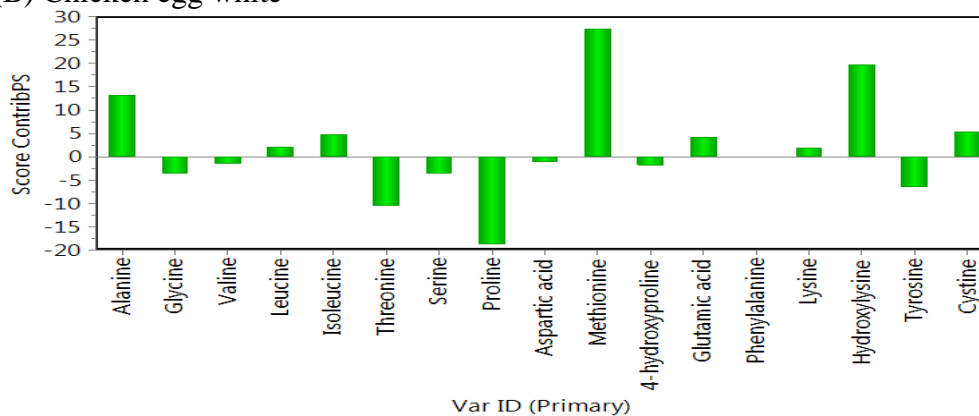
(B) Chicken egg white
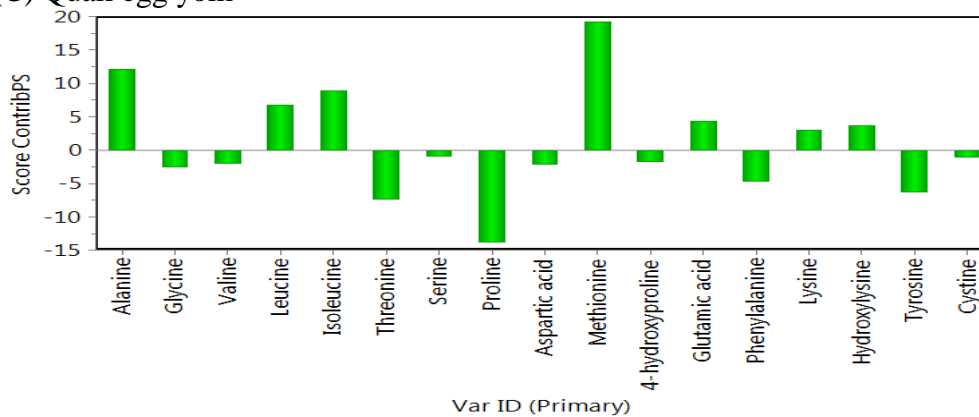
(C) Quail egg yolk
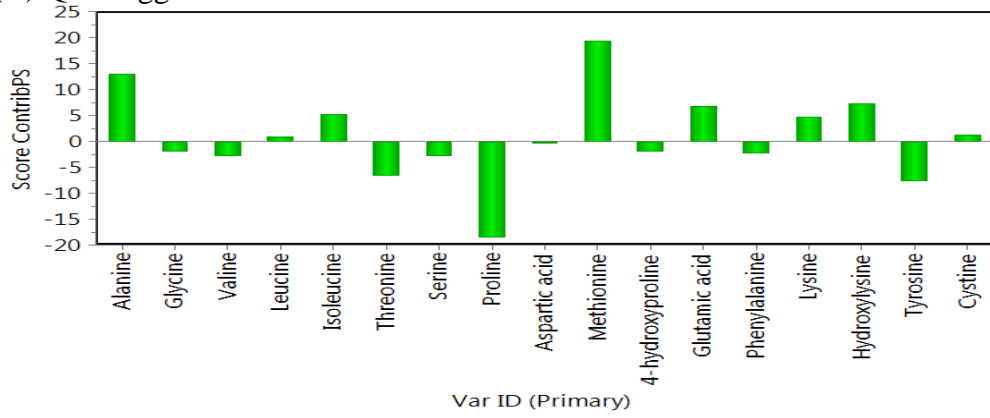
(D) Quail egg white
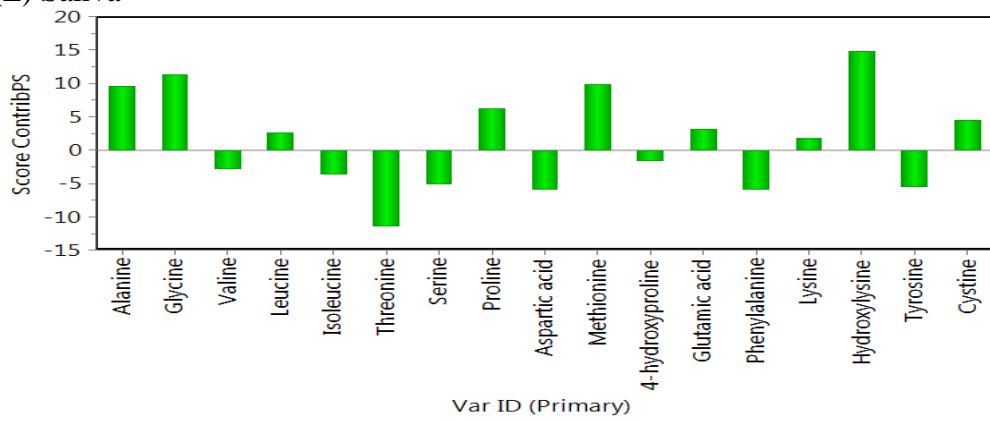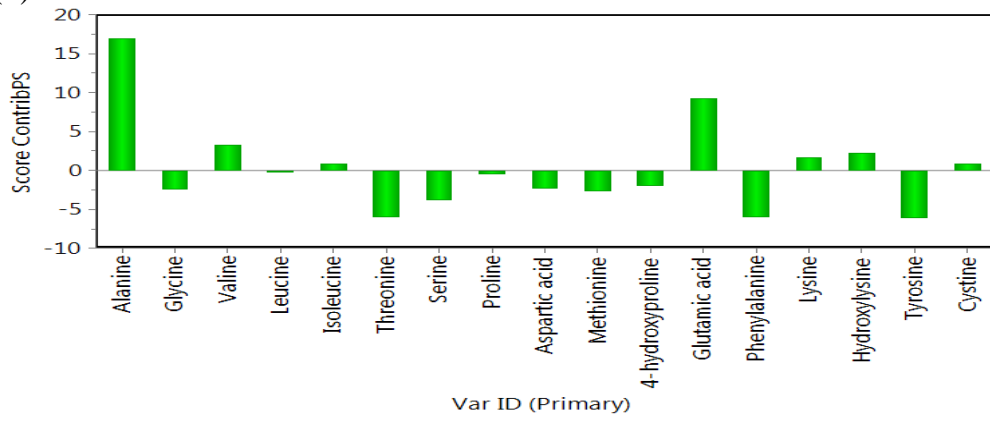
(E) Saliva

(F) Fetuin

Appendix 5 Contribution plots based on the monosaccharide data

(A) Chicken egg yolk
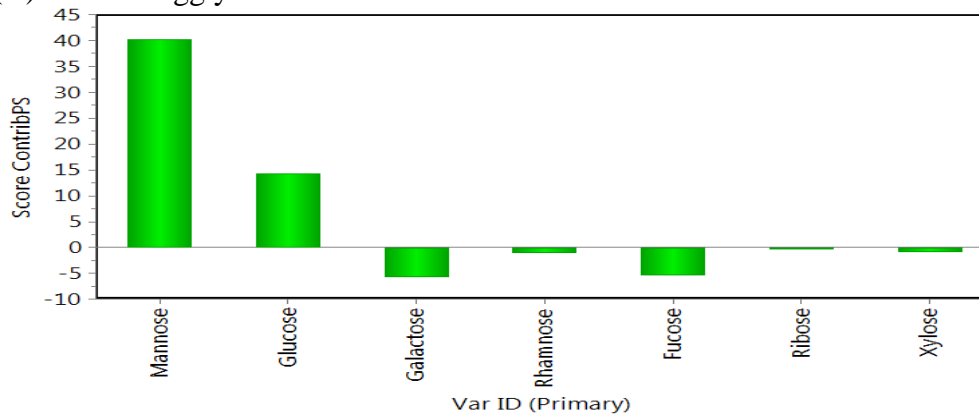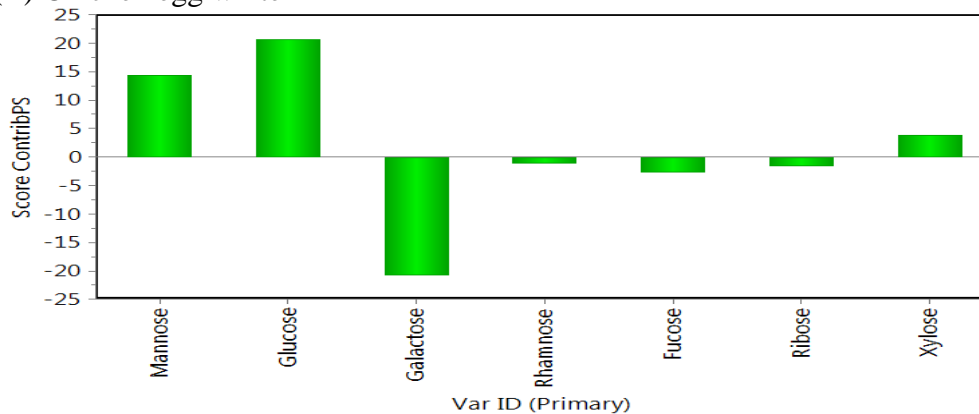
(B) Chicken egg white

(C) Quail egg yolk
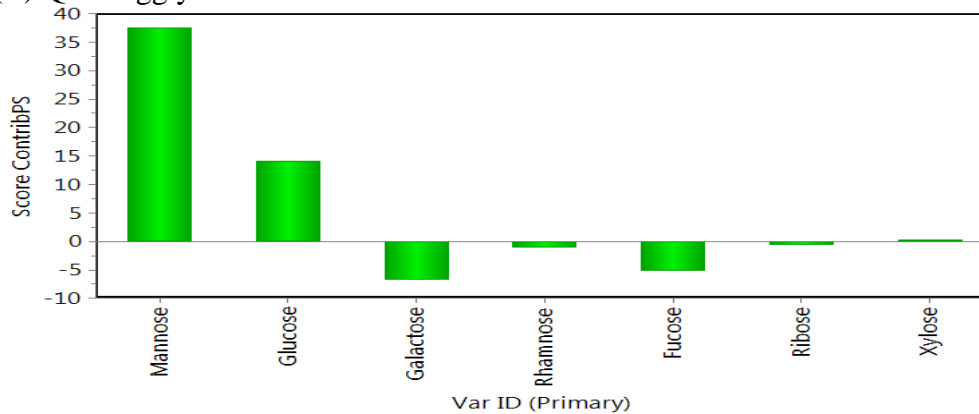
(D) Quail egg white
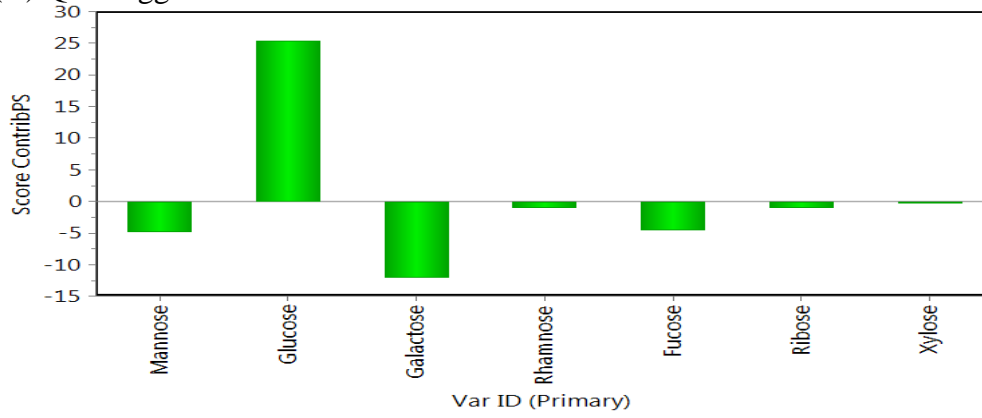
(E) Saliva
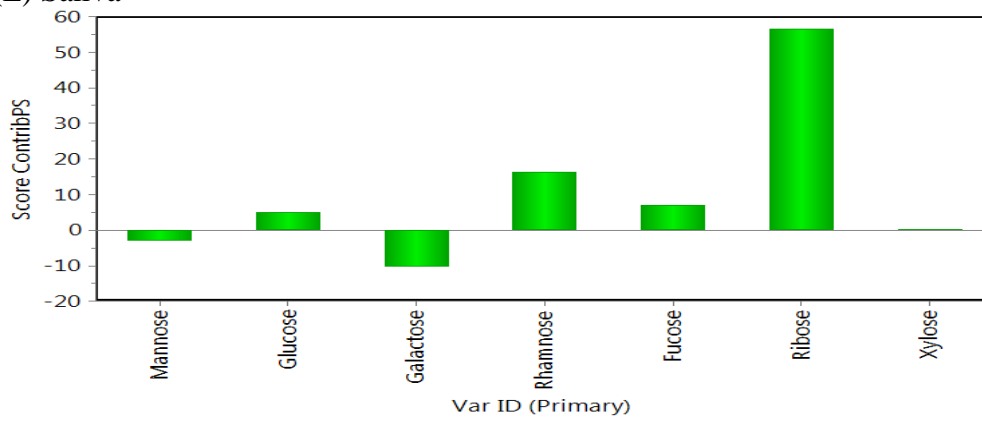
(F) Fetuin
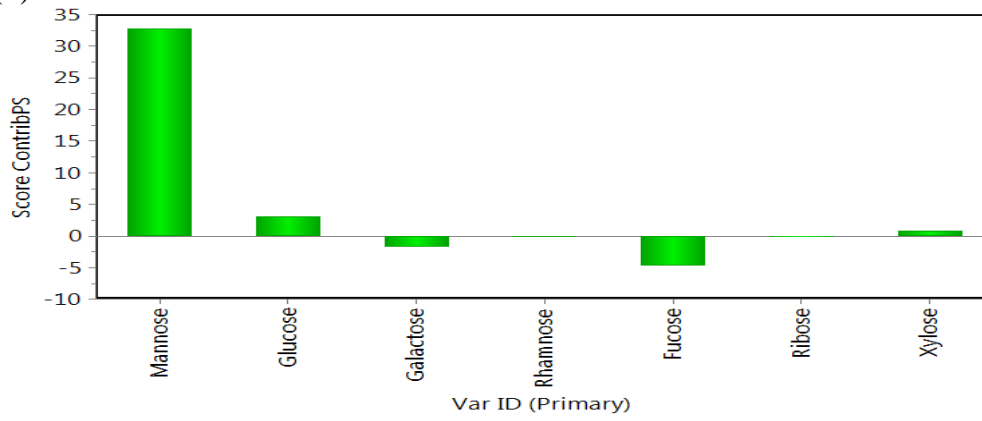
Appendix 6 Concentration of the amino acids of EBN and different types of food items

Compound	Concentration of individual amino acid for different food items (mg / g of sample)						
	EBN	Milk	Chicken Egg Yolk	Chicken Egg White	Quail Egg Yolk	Quail Egg White	Infant formula
Alanine	14.70 ± 0.93	0.97 ± 0.02	11.24 ± 0.93	7.15 ± 0.99	7.52 ± 0.80	6.67 ± 0.33	3.59 ± 0.15
Glycine	19.95 ± 1.39	0.51 ± 0.03	6.38 ± 0.46	3.91 ± 0.38	4.73 ± 0.40	4.25 ± 0.11	2.72 ± 0.16
Valine	36.43 ± 4.56	1.92 ± 0.04	12.08 ± 0.84	8.39 ± 0.82	9.03 ± 0.77	7.56 ± 0.08	6.25 ± 0.29
Leucine	35.53 ± 2.30	2.86 ± 0.03	16.96 ± 1.26	10.17 ± 0.78	12.94 ± 1.04	9.27 ± 0.16	0.00 ± 0.00
Isoleucine	15.23 ± 1.76	1.38 ± 0.04	9.97 ± 0.72	5.58 ± 0.50	7.35 ± 0.55	5.19 ± 0.11	10.18 ± 0.64
Threonine	41.01 ± 2.96	1.01 ± 0.06	9.96 ± 1.81	4.47 ± 1.02	6.91 ± 1.35	6.34 ± 0.72	5.40 ± 0.26
Serine	37.20 ± 4.16	0.94 ± 0.04	13.67 ± 3.90	5.83 ± 2.22	9.78 ± 2.33	6.60 ± 1.23	6.16 ± 0.35
Proline	43.56 ± 3.74	3.58 ± 0.12	9.39 ± 0.85	3.91 ± 0.41	6.41 ± 0.53	3.62 ± 0.07	5.22 ± 0.30
Aspartic acid	53.86 ± 5.09	2.53 ± 0.10	20.32 ± 2.82	13.65 ± 2.10	13.98 ± 2.37	13.32 ± 1.20	9.11 ± 0.33
Methionine	3.86 ± 0.54	0.82 ± 0.03	2.91 ± 0.16	4.94 ± 0.41	4.19 ± 0.31*	3.69 ± 0.08*	10.51 ± 0.33
4-Hydroxyproline	4.57 ± 2.86	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	2.04 ± 0.13
Glutamic acid	34.15 ± 2.73	8.28 ± 0.20	23.87 ± 4.04	19.08 ± 4.70	19.82 ± 4.01	19.68 ± 2.92	0.00 ± 0.00
Phenylalanine	29.46 ± 2.34	1.52 ± 0.04	8.57 ± 0.76	7.89 ± 0.63	6.79 ± 0.69	6.59 ± 0.28	16.82 ± 0.60
Lysine	16.22 ± 2.54	1.44 ± 0.09	9.74 ± 2.94	5.99 ± 0.94	7.04 ± 0.80	6.67 ± 1.28	4.60 ± 0.35
Hydroxylysine	5.57 ± 0.57	0.00 ± 0.00	2.37 ± 0.25	4.59 ± 0.58	2.27 ± 0.19	2.46 ± 0.20	8.05 ± 0.52
Tyrosine	31.41 ± 2.38	1.31 ± 0.07	7.27 ± 0.83	5.57 ± 0.85	6.05 ± 0.52	4.26 ± 0.27	0.00 ± 0.00
Cystine	16.42 ± 1.93	0.00 ± 0.00	4.96 ± 0.82	6.77 ± 0.71	4.30 ± 0.51	4.74 ± 0.17	5.27 ± 0.47

* represents no significance difference, $p > 0.05$

Appendix 7 GC/MS chromatograms of different catogories of EBN–coloration, countries and production sites

(A) Coloration - white EBN
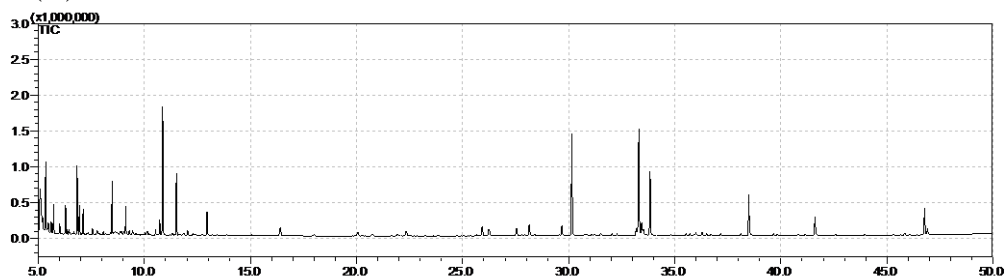
(B) Coloration - orange EBN
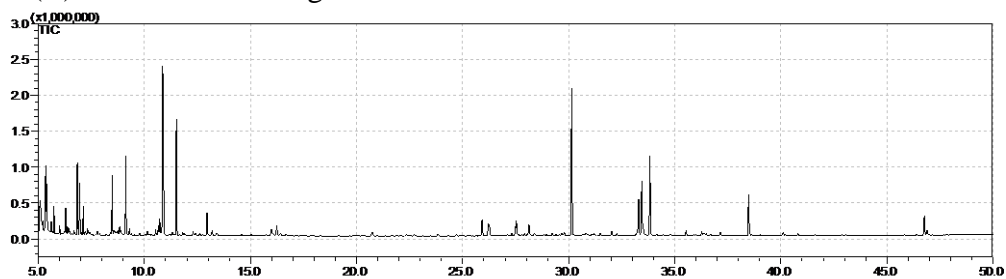
(C) Coloration - red EBN
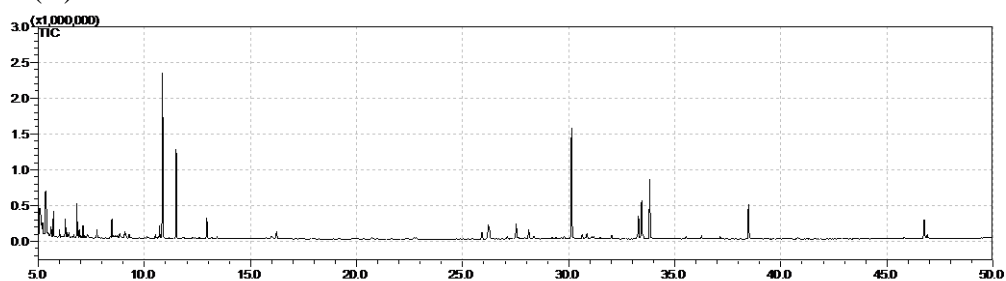
(D) Countries - Malaysian EBN
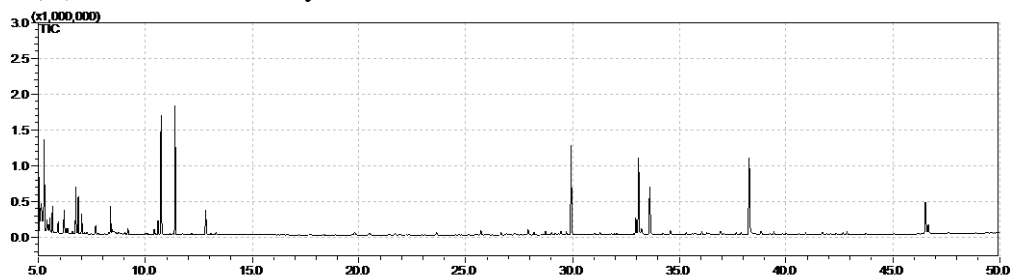
(E) Countries - Indonesia EBN
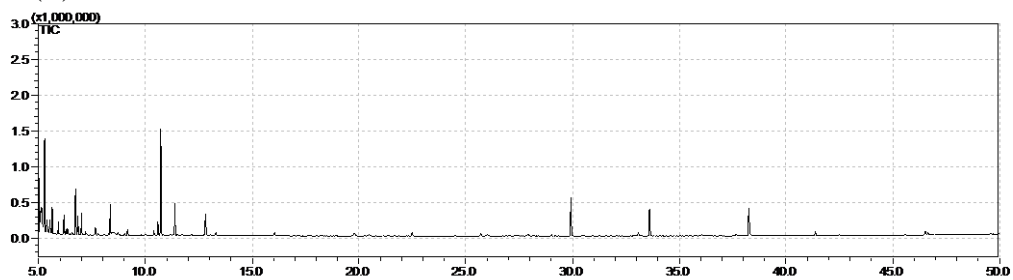
(F) Countries - Thai EBN
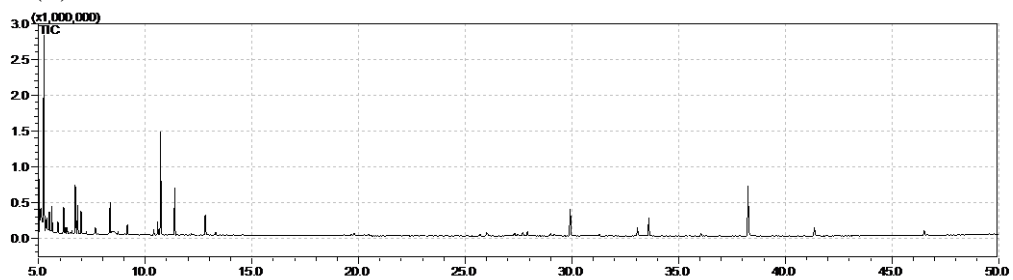
(G) Production sites - farm EBN
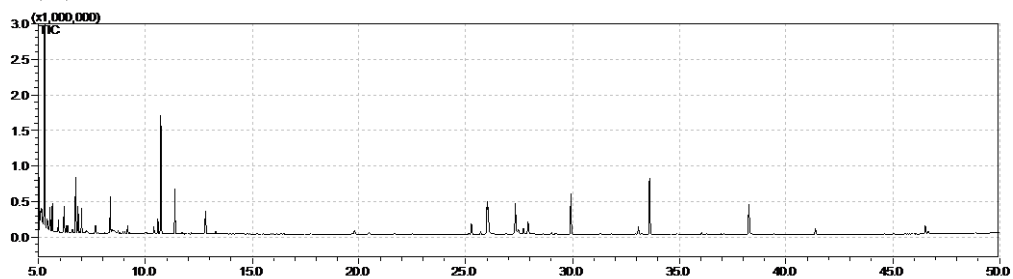
(H) Production sites - cave EBN
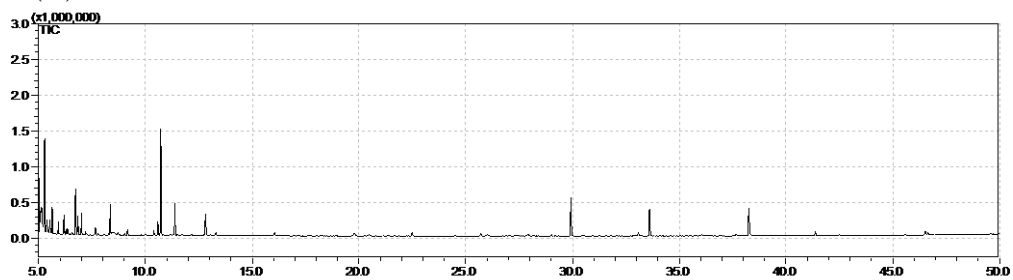
Appendix 8 LC/MS chromatograms of different catogories of EBN–coloration, countries and production sites

(A) Coloration - white EBN in positive mode
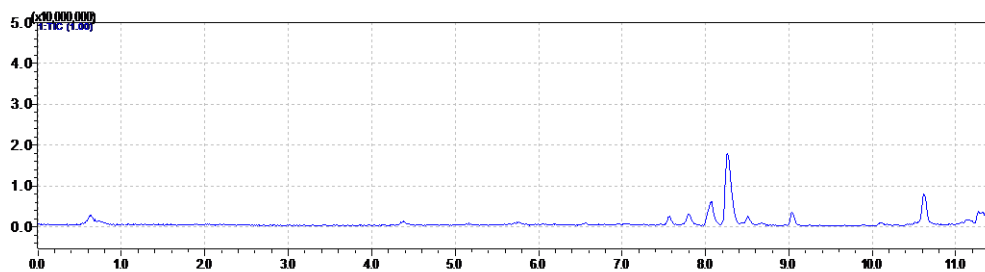
(A) Coloration - white EBN in negative mode
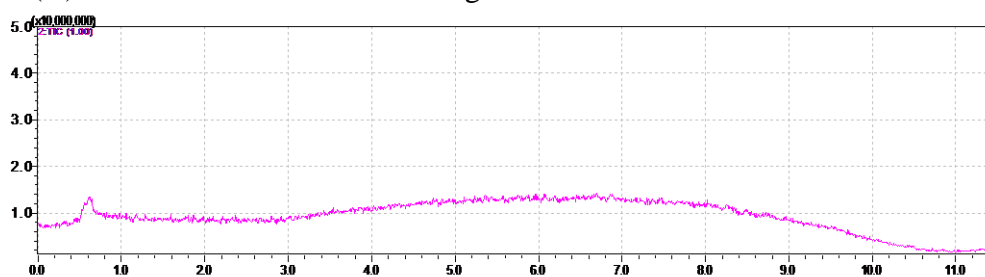
(B) Coloration - orange EBN in positive mode
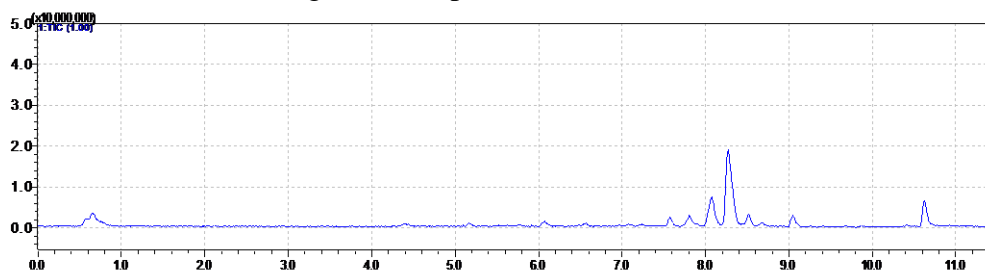
(B) Coloration - orange EBN in negative mode
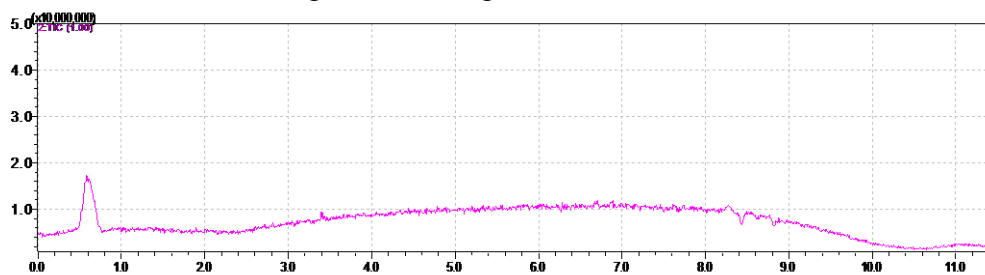
(C) Coloration - red EBN in positive mode
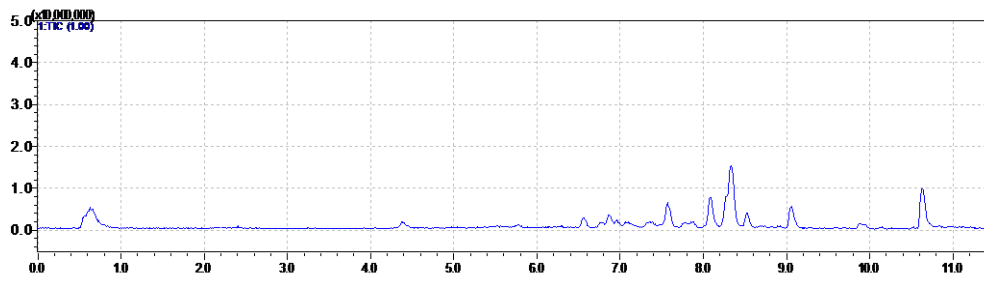
(C) Coloration - red EBN in negative mode
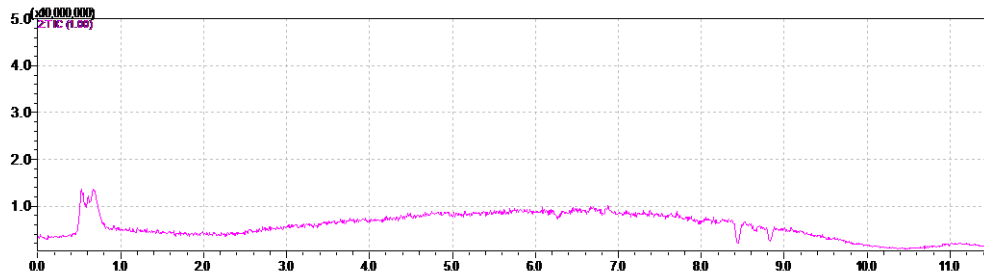
(D) Countries - Malaysian EBN in positive mode
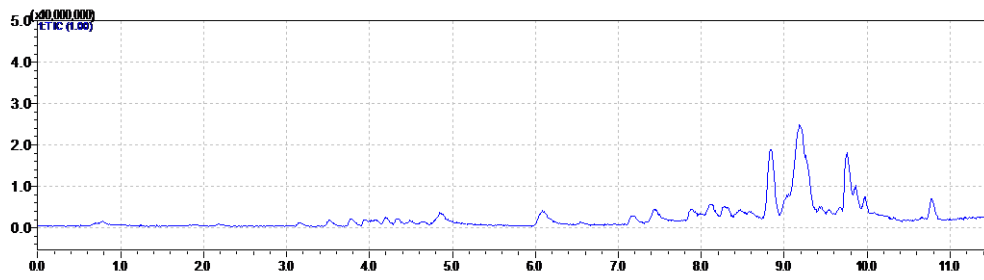
(D) Countries - Malaysian EBN in negative mode
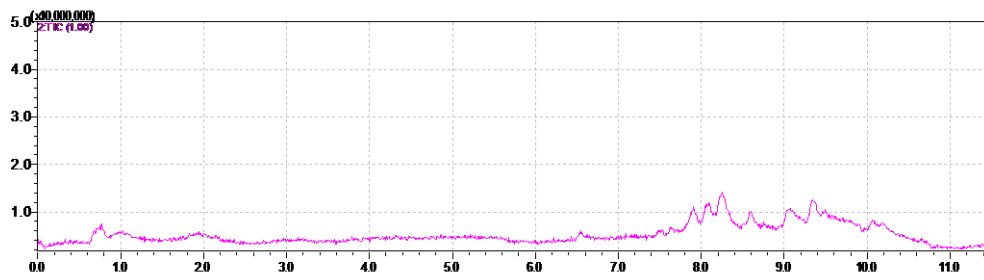
(E) Countries - Indonsian EBN in positive mode
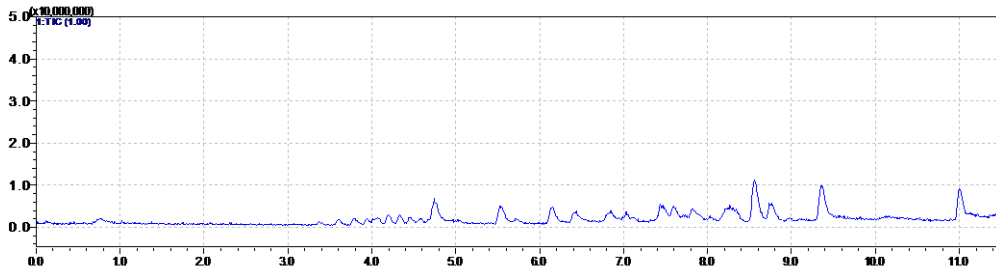
(E) Countries - Indonsian EBN in positive mode
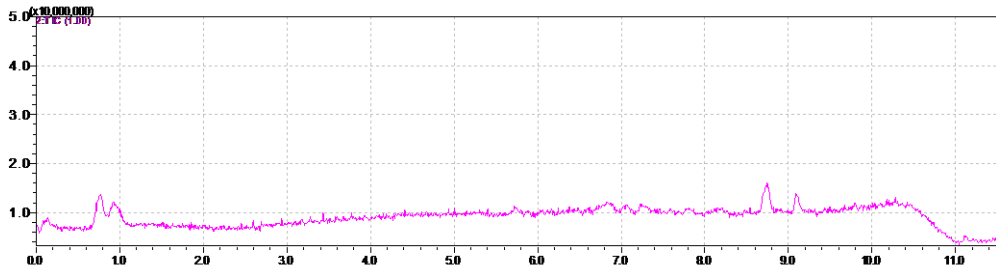
(F) Countries - Thai EBN in positive mode
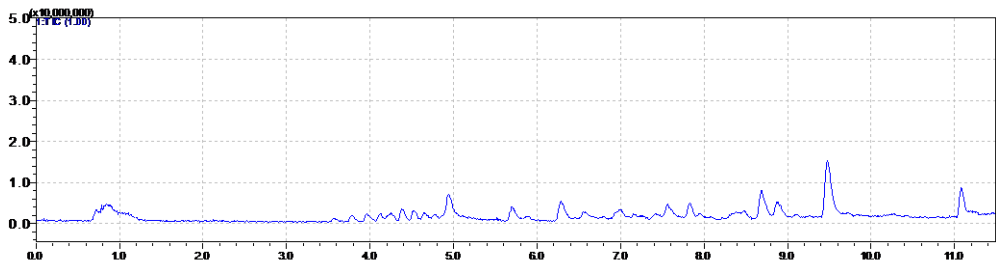
(F) Countries - Thai EBN in negative mode
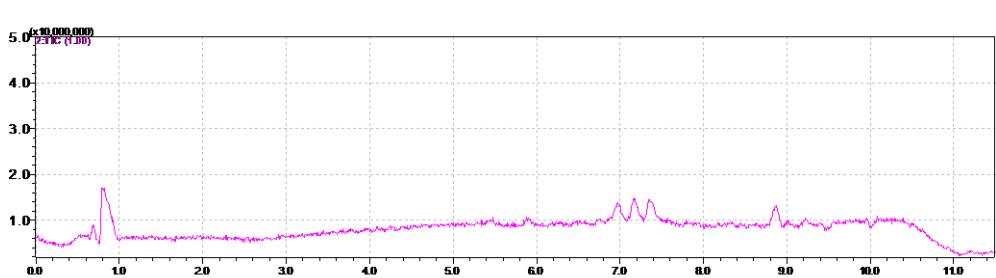
(G) Farm EBN in positive mode
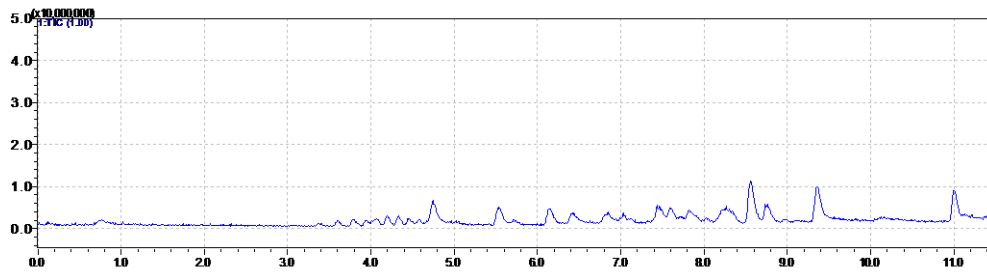
(G) Farm EBN in negative mode
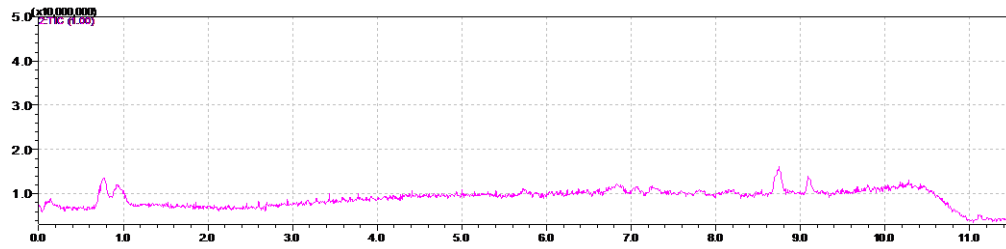
(H) Cave EBN in positive mode
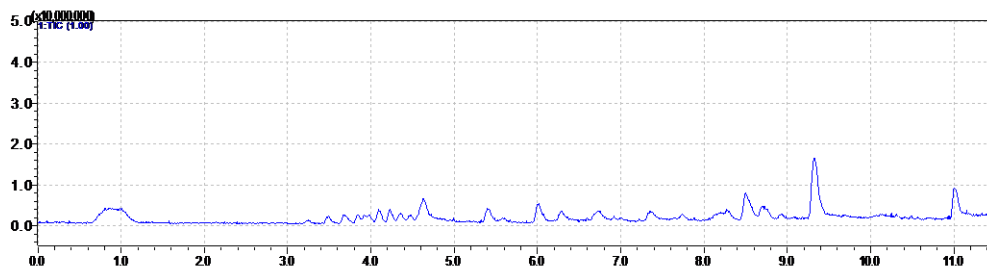
(H) Cave EBN in negative mode
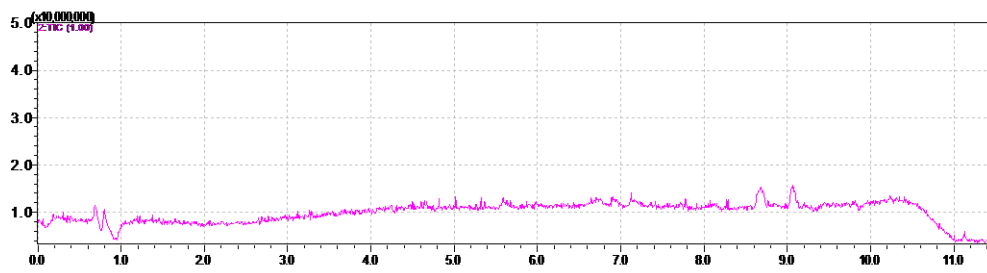
Appendix 9 GC/MS chromatograms with the main region of difference being
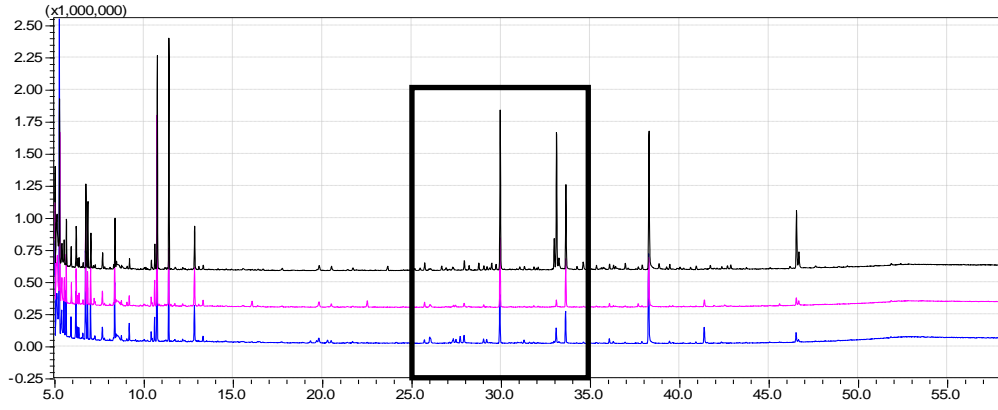
hightlighted

Appendix 10 LC/MS chromatograms with the main region of difference being

hightlighted

(A) Chromatogram A in positive mode
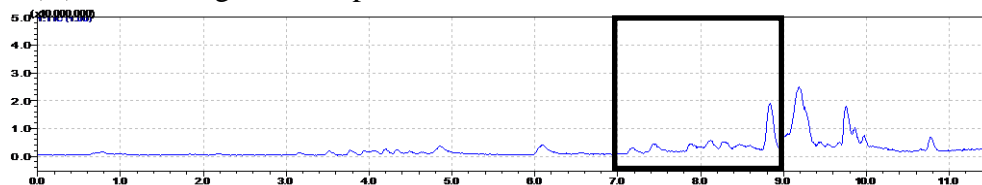
(A) Chromatogram A in negative mode
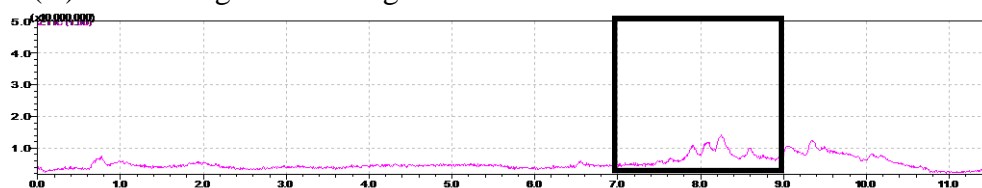
(B) Chromatogram B in positive mode
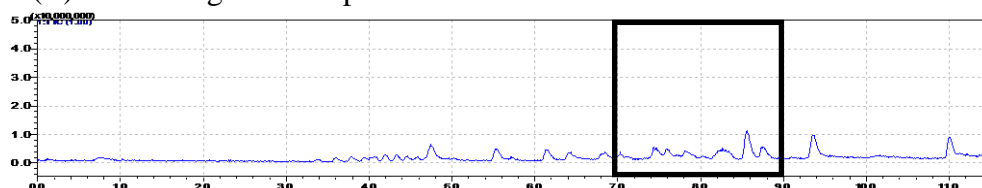
(B) Chromatogram B in negative mode
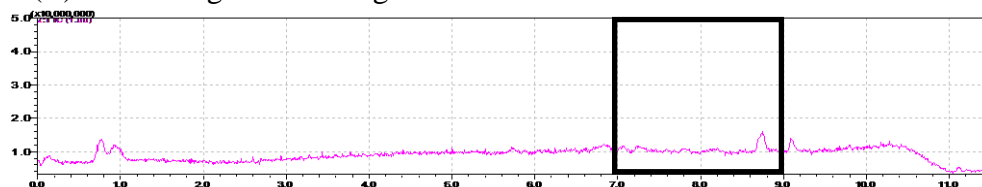
(C) Chromatogram C in positive mode
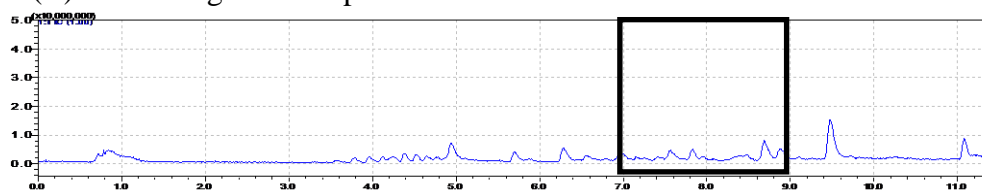
(C) Chromatogram C in negative mode
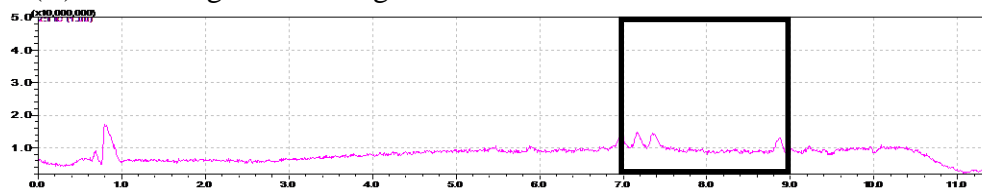
Appendix 11 Amino acid chromatograms of samples spiked different amount of adulterant.
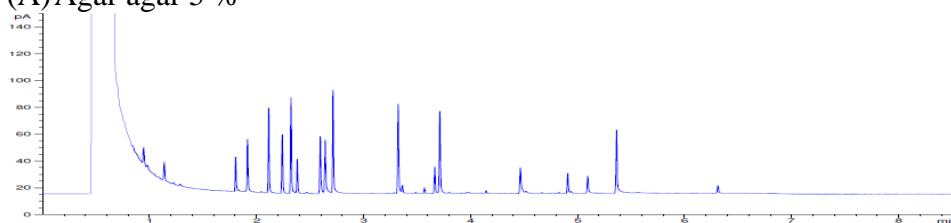
(A) Agar agar 5 %
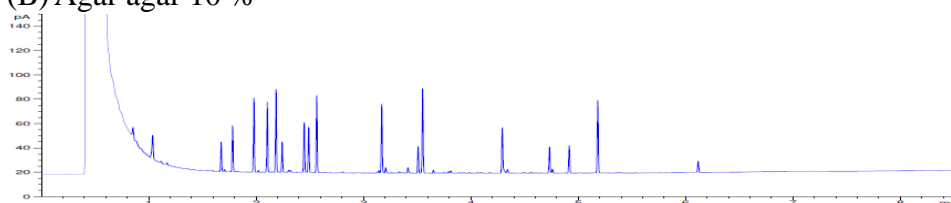
(B) Agar agar 10 %
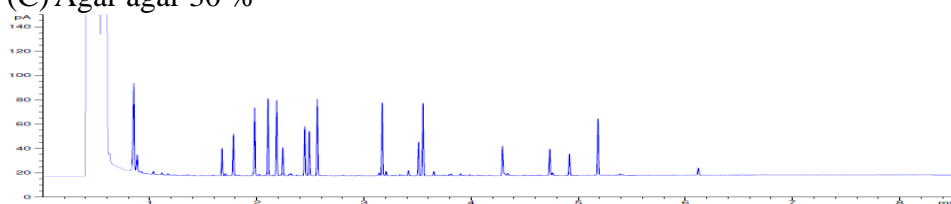
(C) Agar agar 30 %
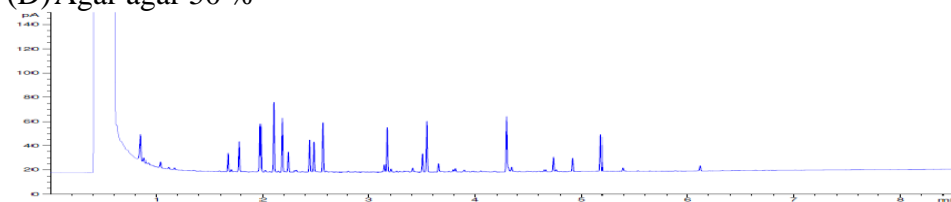
(D) Agar agar 50 %
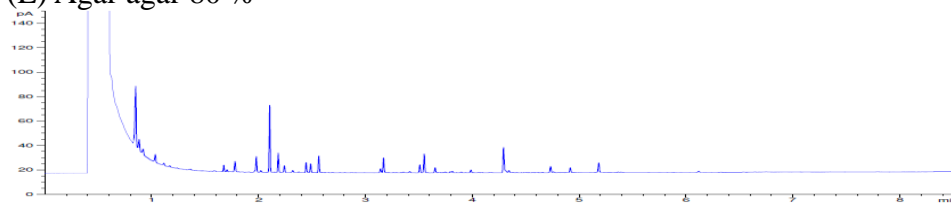
(E) Agar agar 80 %

(F) Agar agar 100 %
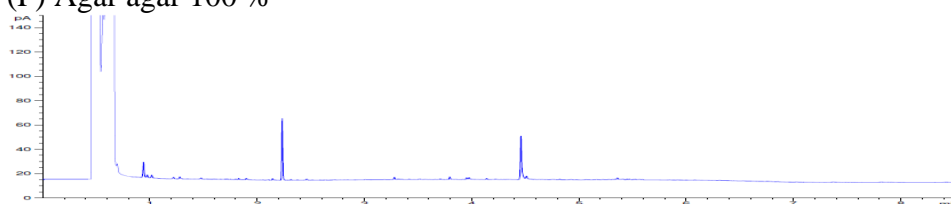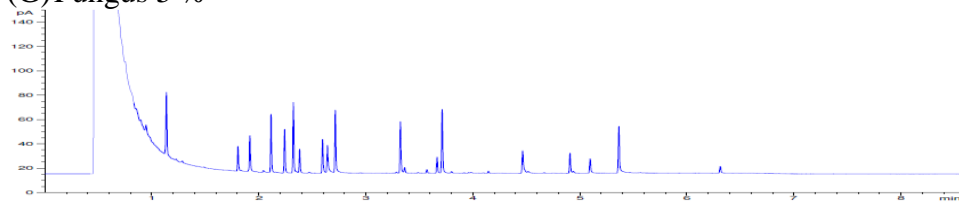
(G) Fungus 5 %
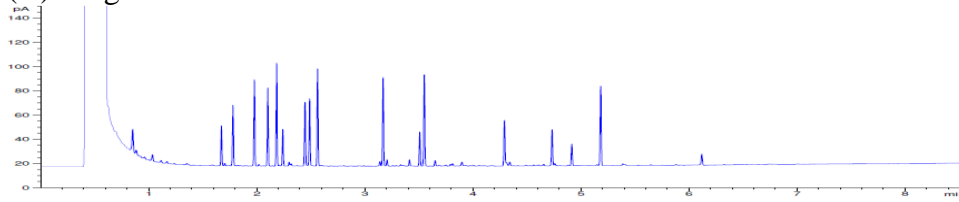
(H) Fungus 10 %

(I) Fungus 30 %
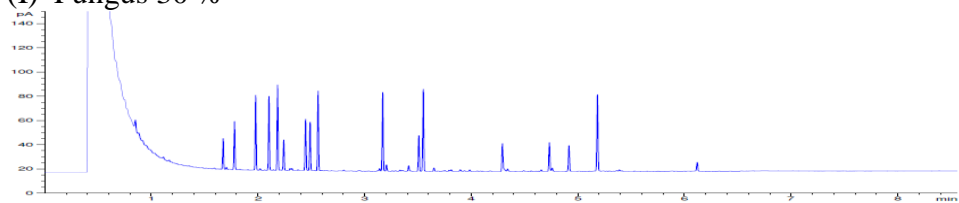
(J) Fungus 50 %
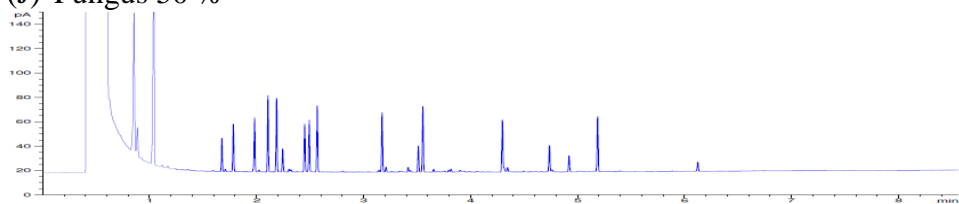
(K) Fungus 80 %
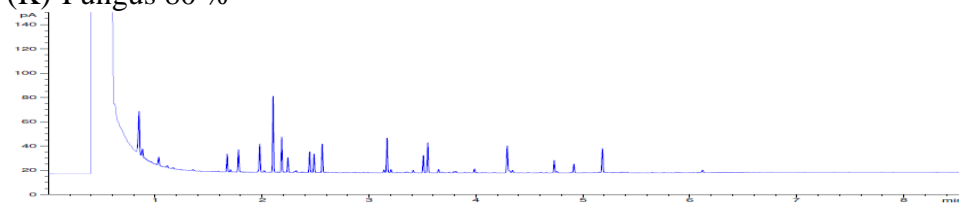
(L) Fungus 100 %
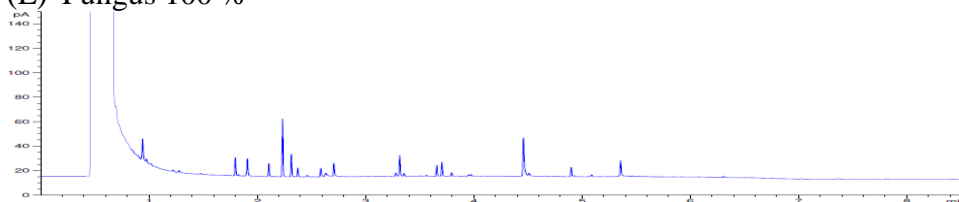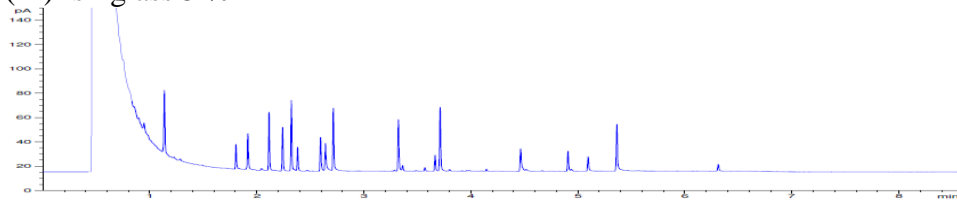
(M) Isinglass 5 %
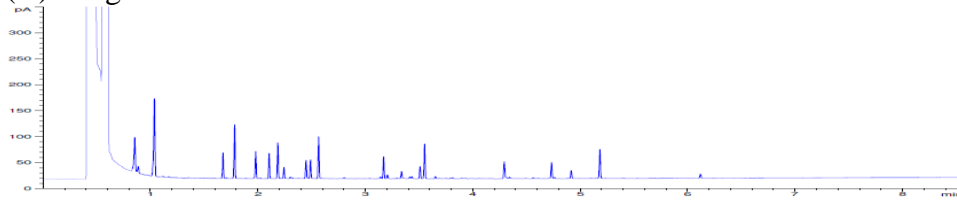
(N) Isinglass 10 %
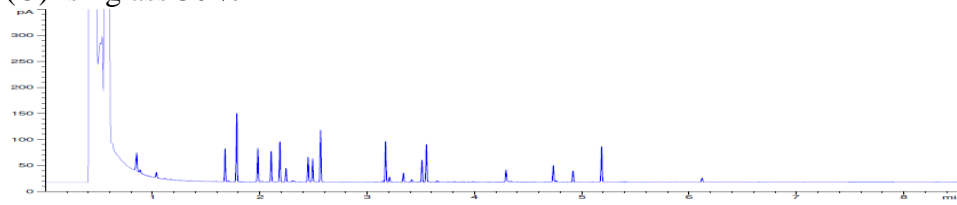
(O) Isinglass 30 %
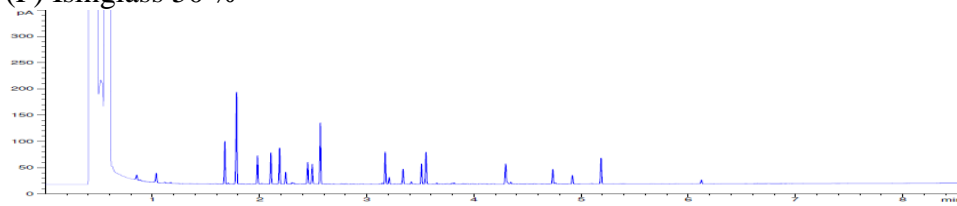
(P) Isinglass 50 %
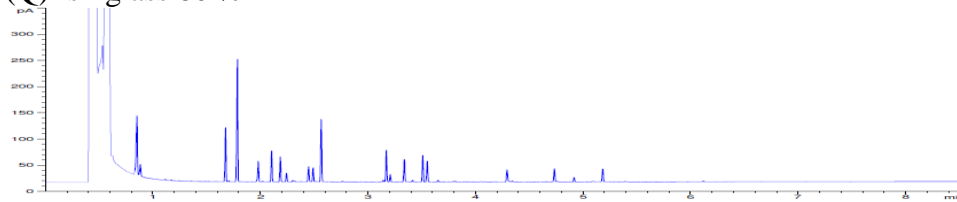
(Q) Isinglass 80 %

(R) Isinglass 100 %
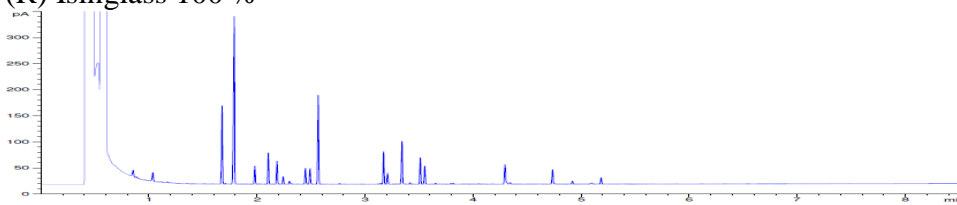
Appendix 12 Monosaccharide chromatograms of samples spiked different amount of adulterant
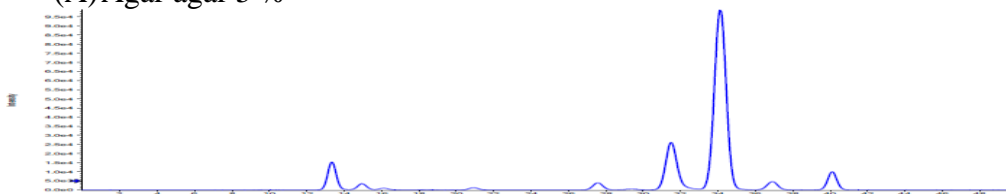
(A) Agar agar 5 %
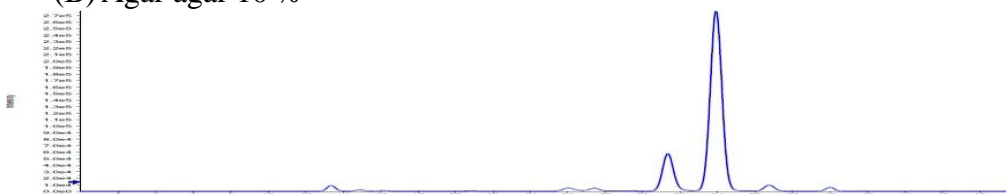
(B) Agar agar 10 %
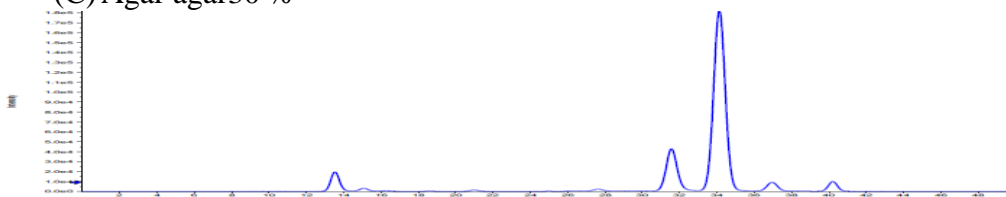
(C) Agar agar30 %
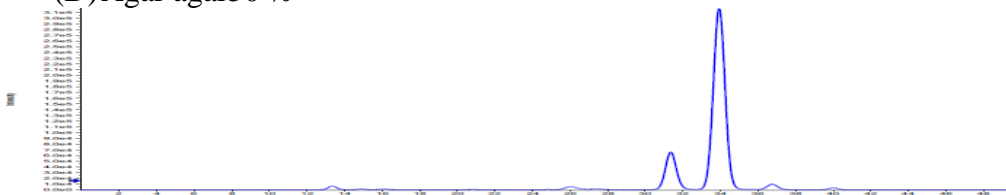
(D) Agar agar50 %
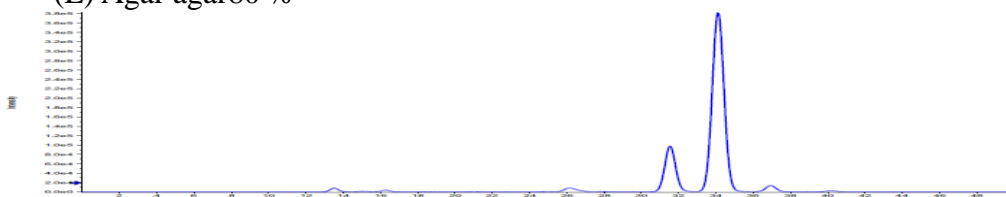
(E) Agar agar80 %

(F) Agar agar 100 %
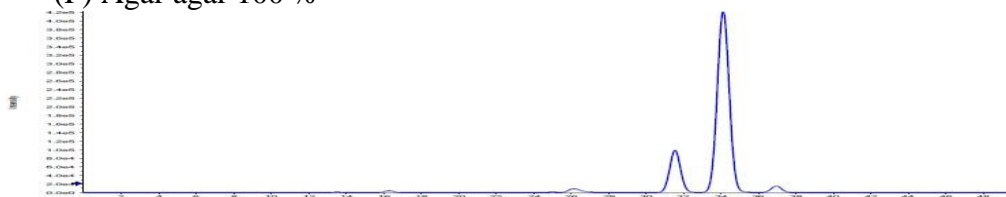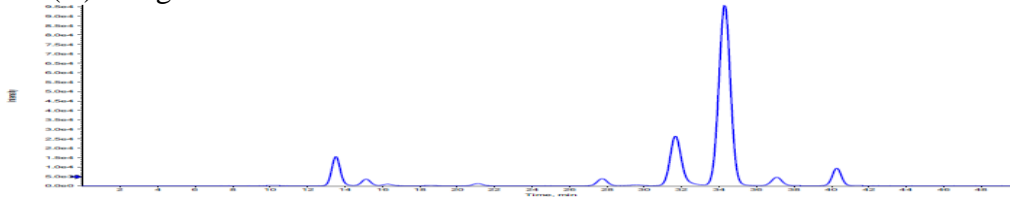
(G) Fungus 5 %
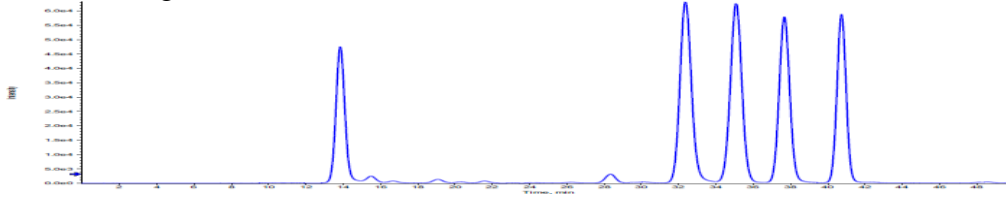
(H) Fungus 10 %

(I) Fungus 30 %
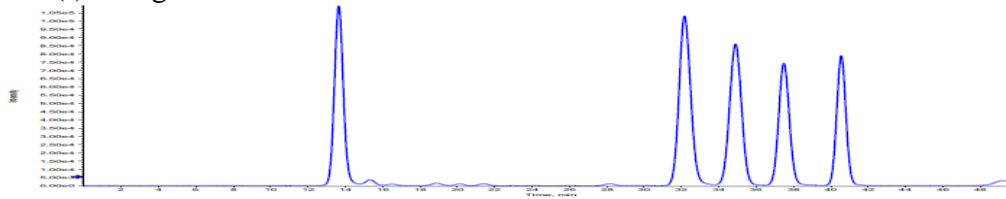
(J) Fungus 50 %
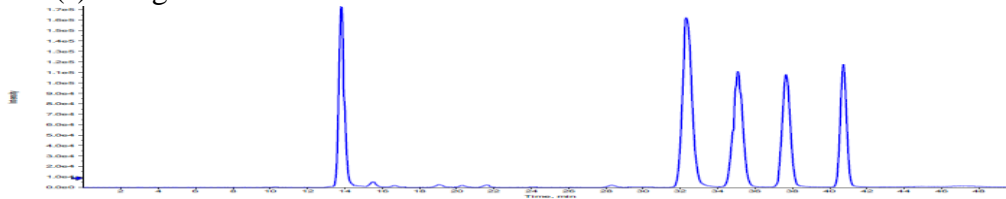
(K) Fungus 80 %
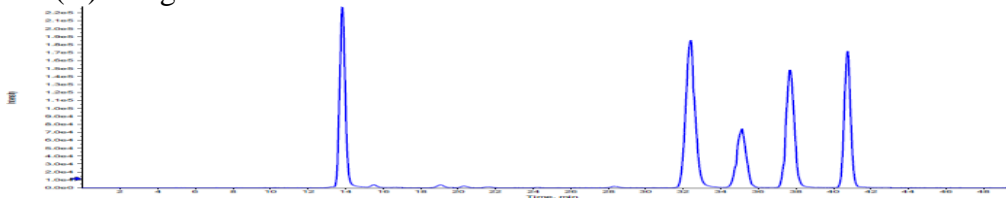
(L) Fungus 100 %
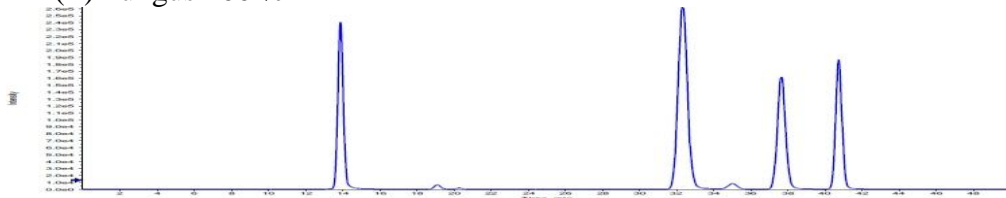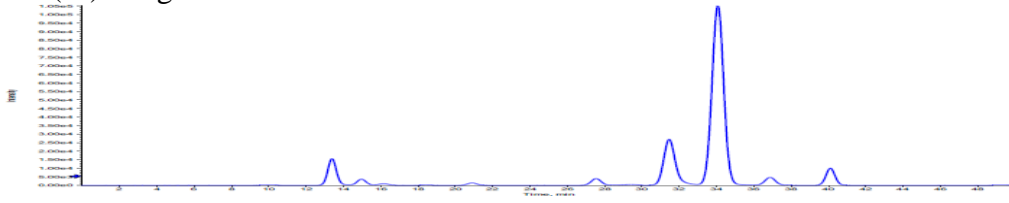
(M) Isinglass 5 %
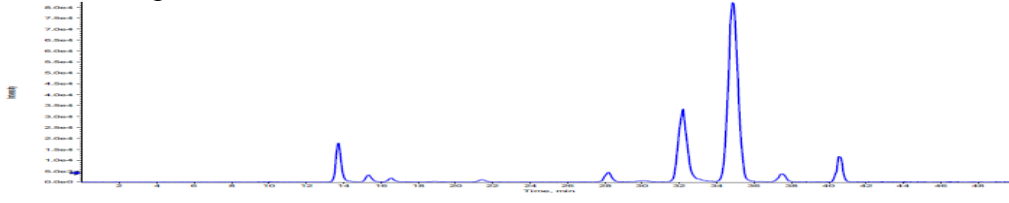
(N) Isinglass 10 %
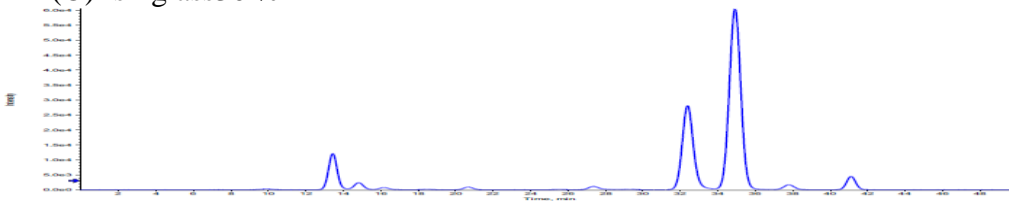
(O) Isinglass30 %
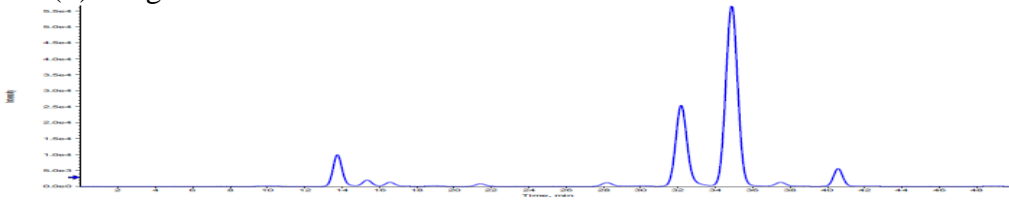
(P) Isinglass50 %
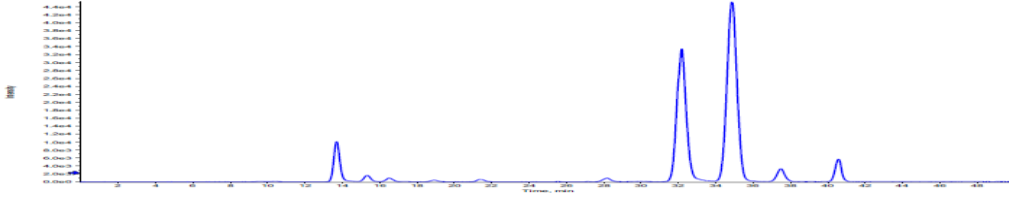
(Q) Isinglass80 %

(R) Isinglass 100 %
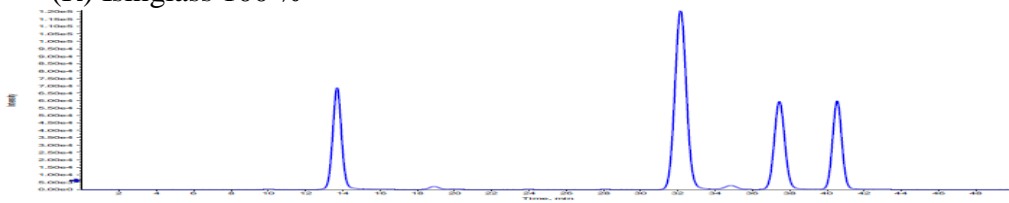
Appendix 13 GC/MS chromatograms of samples spiked with different amount of adulterant
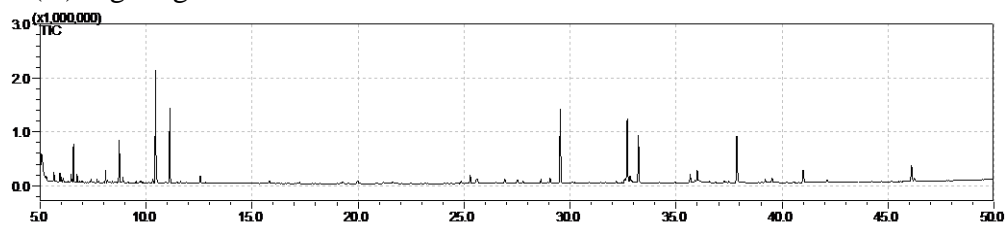
(A) Agar agar 1 %

(B) Agar agar 5 %
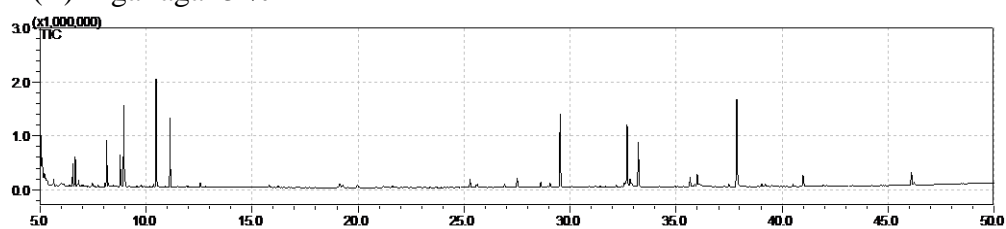
(C) Agar agar 10 %
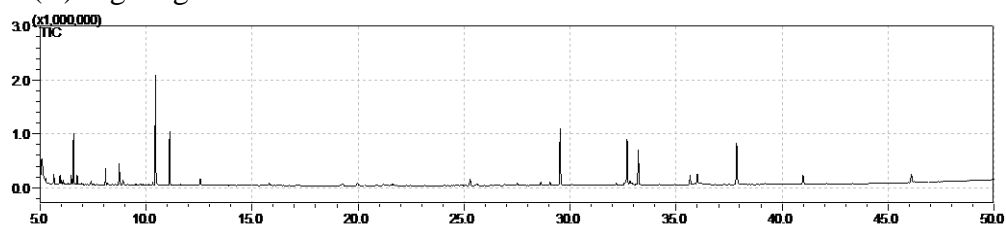
(D) Agar agar 30 %
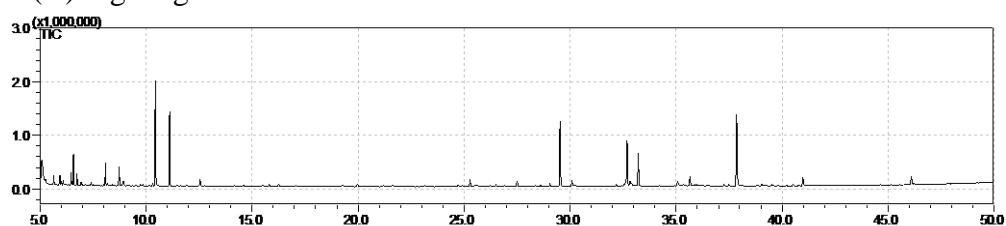
(E) Agar agar 50 %
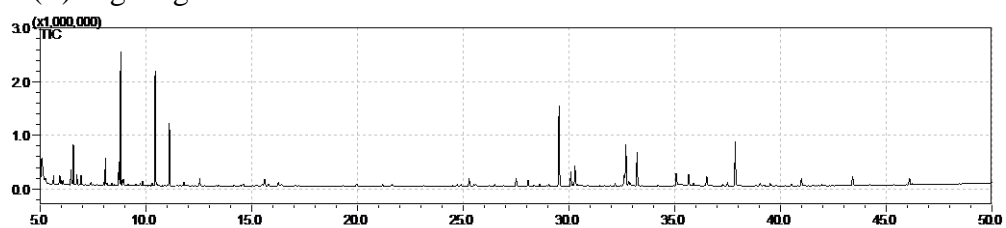
(F) Agar agar 80 %
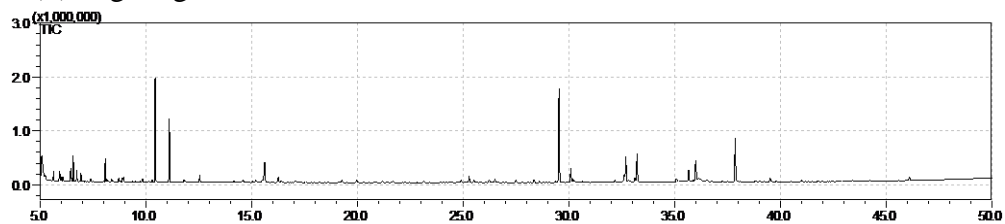
(G) Agar agar 100 %
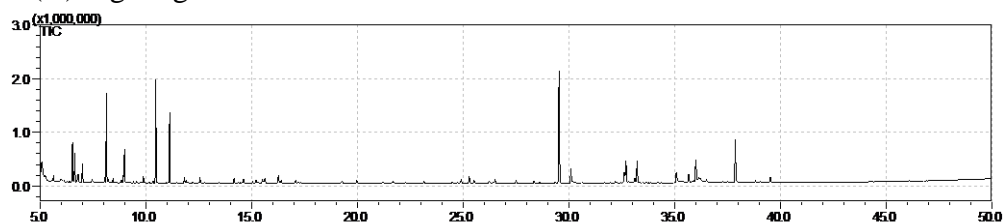
(H) Fungus 0.1 %
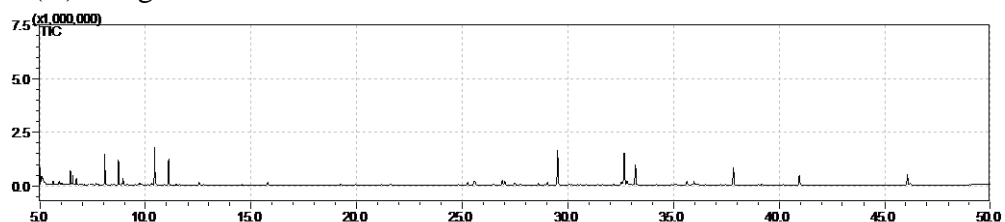
(I) Fungus 0.5 %
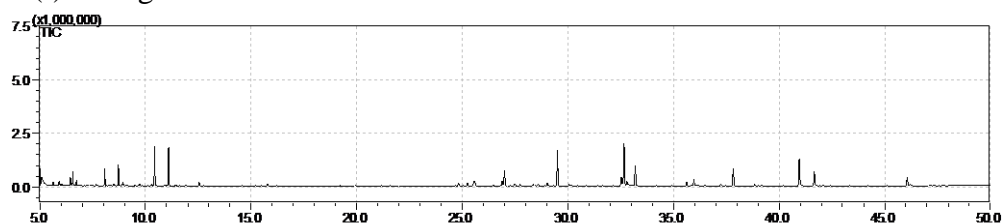
(J) Fungus 1 %
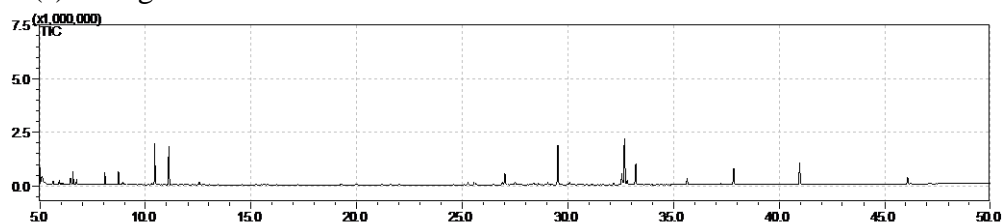
(K) Fungus 5 %
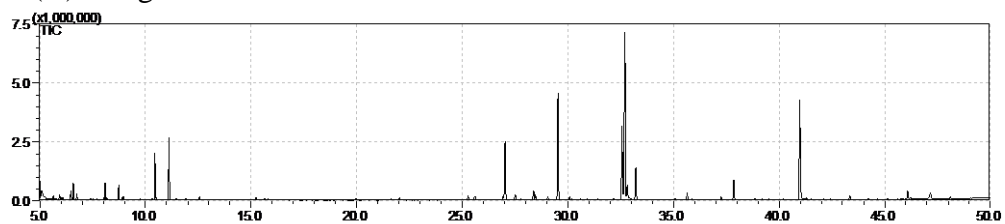
(L) Fungus 10 %
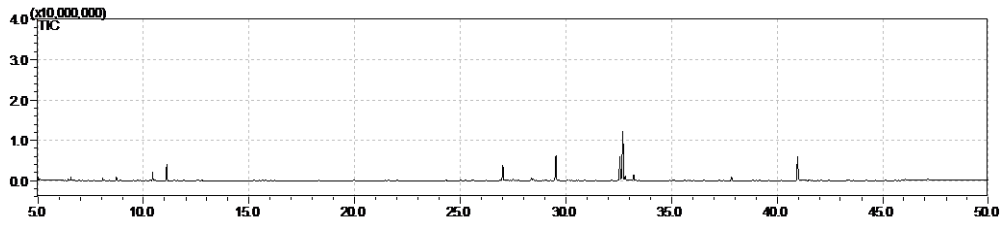
(M) Fungus 30 %
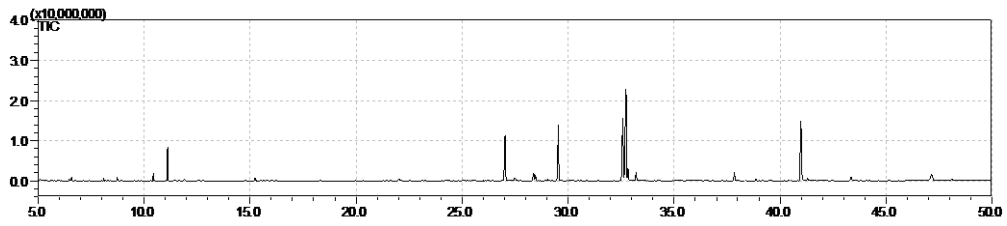
(N) Fungus 50 %
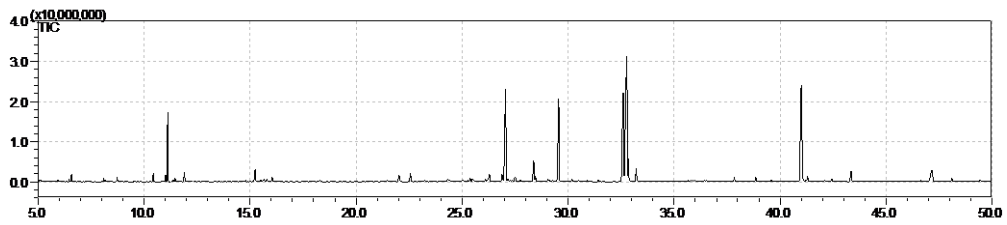
(O) Fungus 80 %
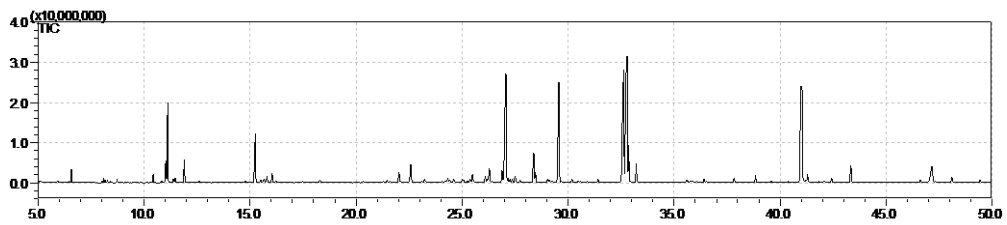
(P) Fungus 100 %
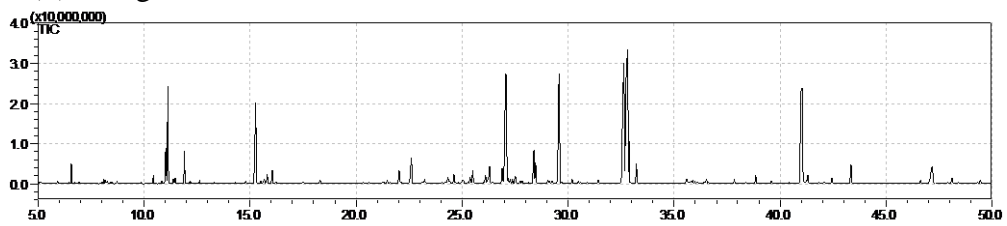
(Q) Isinglass 0.1 %
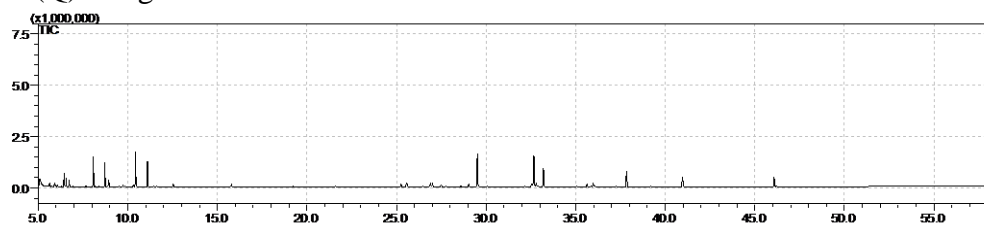
(R) Isinglass 0.5 %
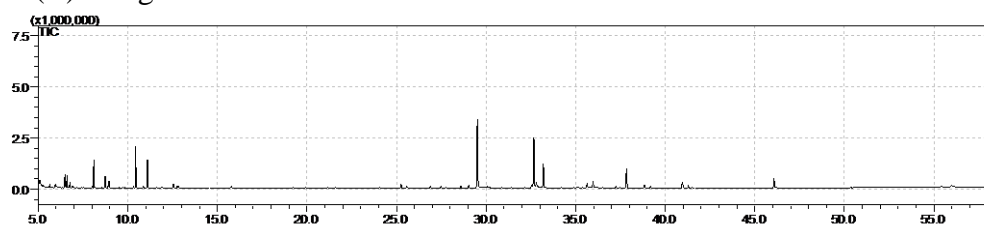
(S) Isinglass 1 %
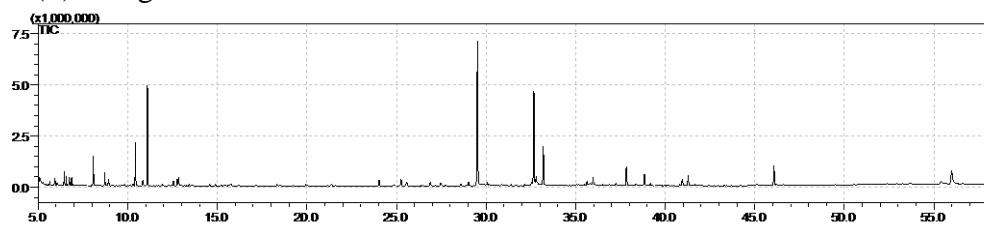
(T) Isinglass 5 %
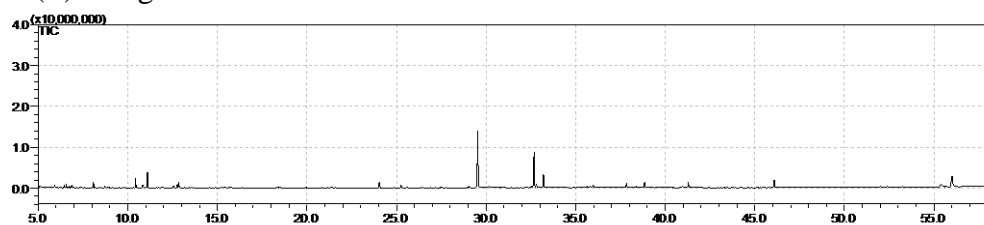
(U) Isinglass 10 %
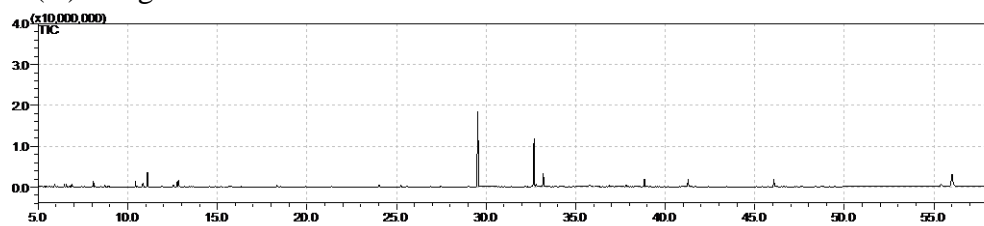
(V) Isinglass 30 %
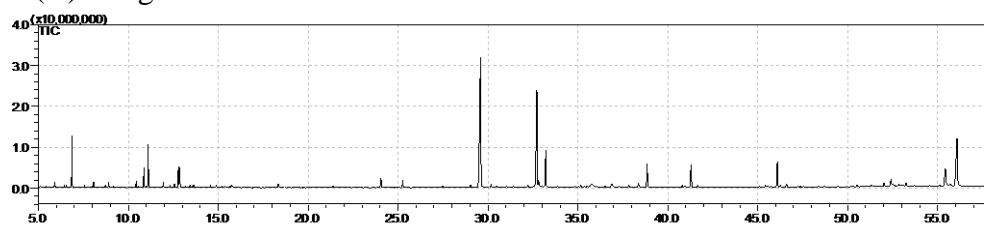
(W) Isinglass 50 %
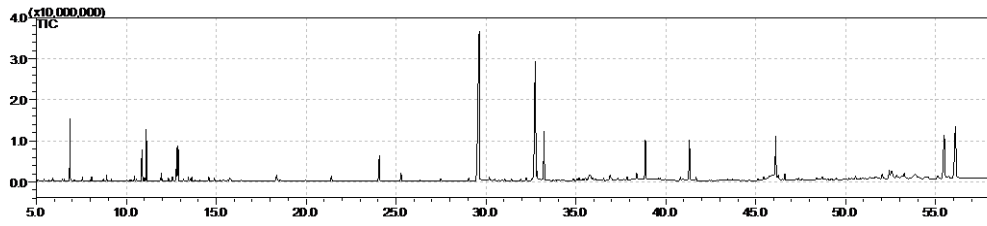
(X) Isinglass 80 %
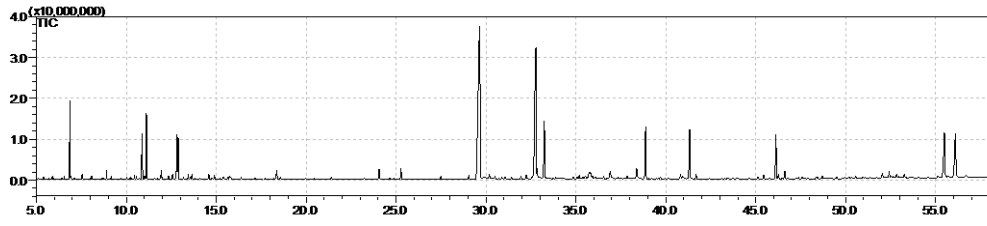
(Y) Isinglass 100 %
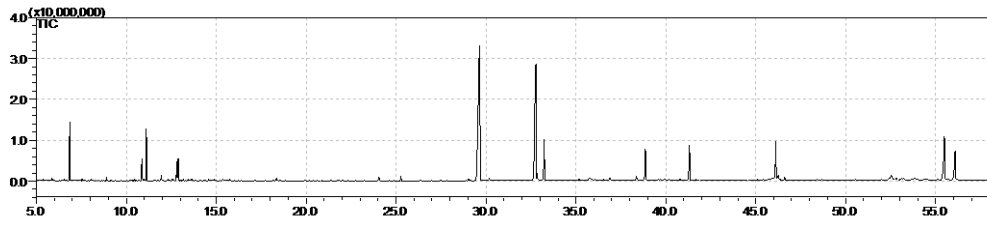
Appendix 14 LC/MS chromatograms of samples spiked with different amount of adulterant

(A) Agar agar 1 % in positive mode
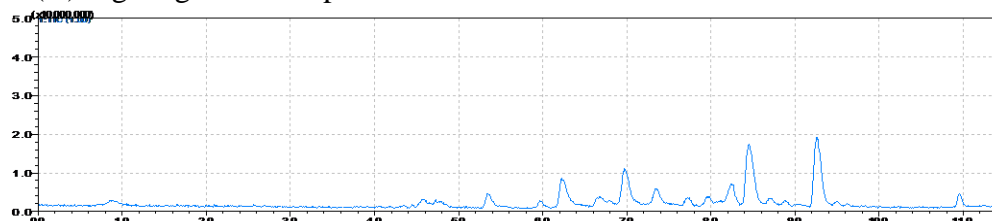
(A) Agar agar 1 % in negative mode
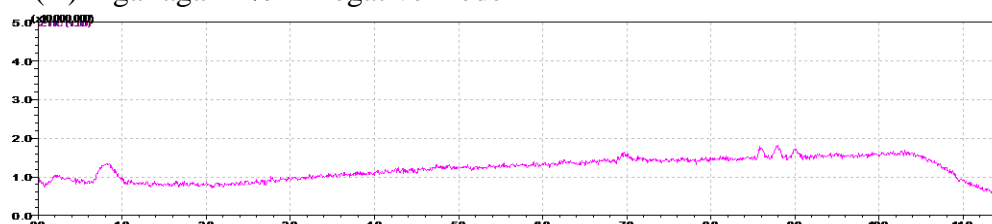
(B) Agar agar 5 % in positive mode
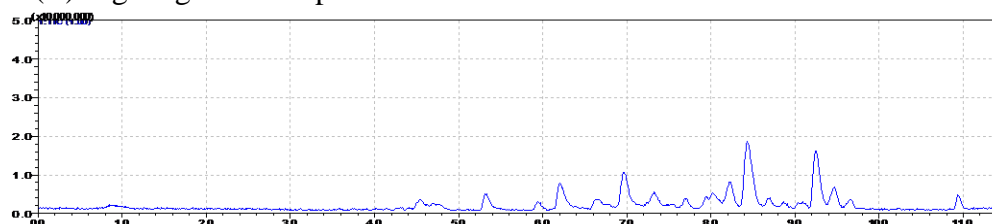
(B) Agar agar 5 % in negative mode
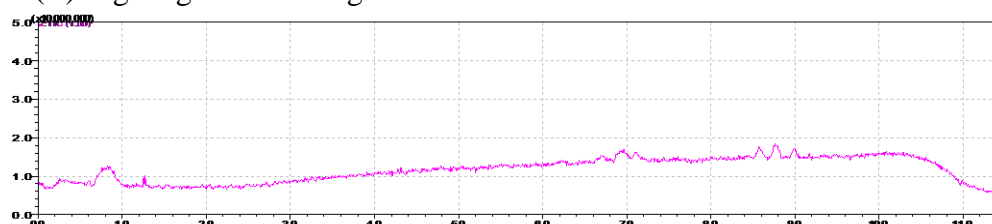
(C) Agar agar 10 % in positive mode
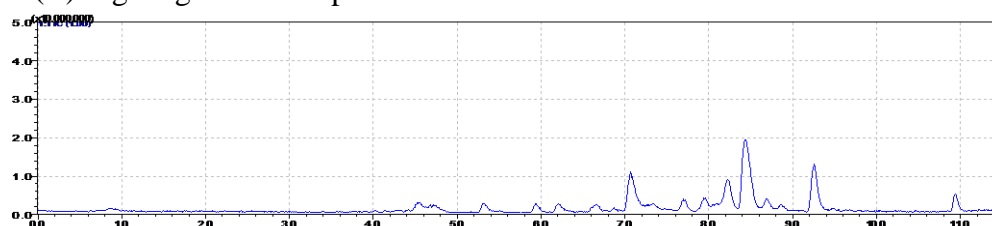
(C) Agar agar 10 % in negative mode
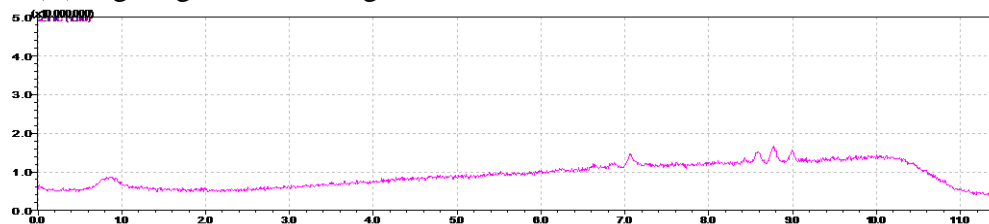
(D) Agar agar 30 % in positive mode
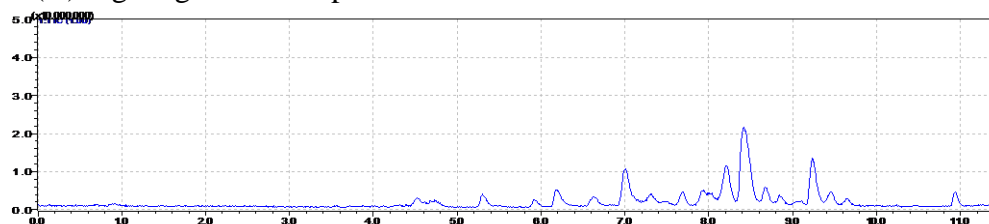
(D) Agar agar 30 % in negative mode
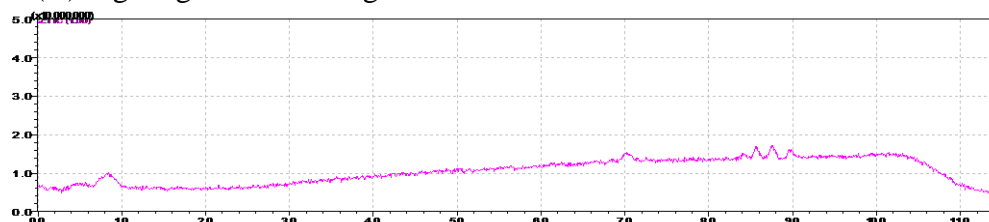
(E) Agar agar 50 % in positive mode
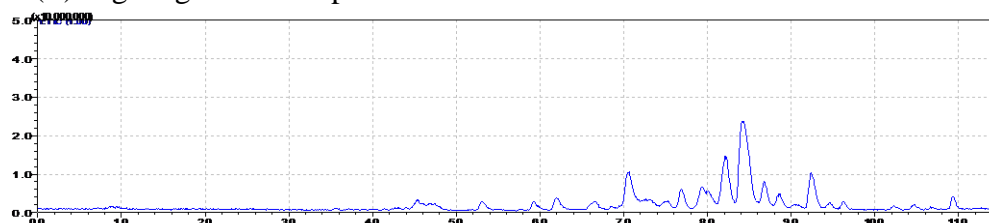
(E) Agar agar 50 % in negative mode
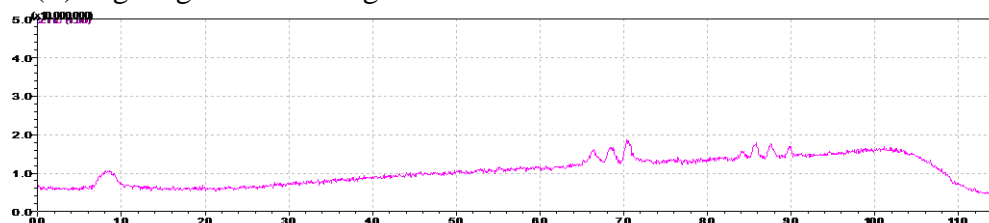
(F) Agar agar 80 % in positive mode
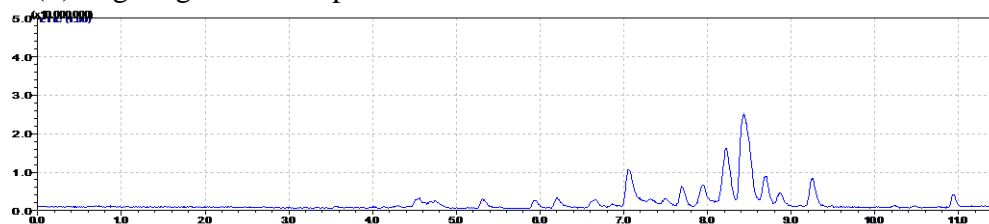
(F) Agar agar 80 % in negative mode
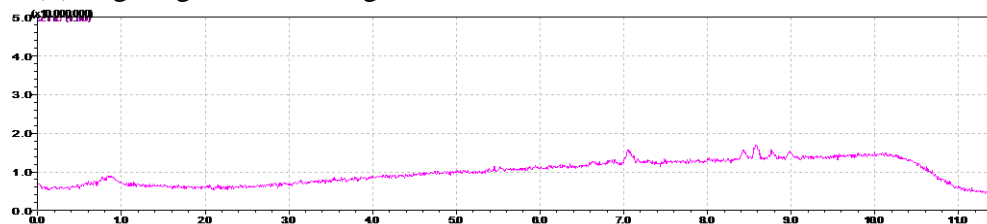
(G) Agar agar 100 % in positive mode
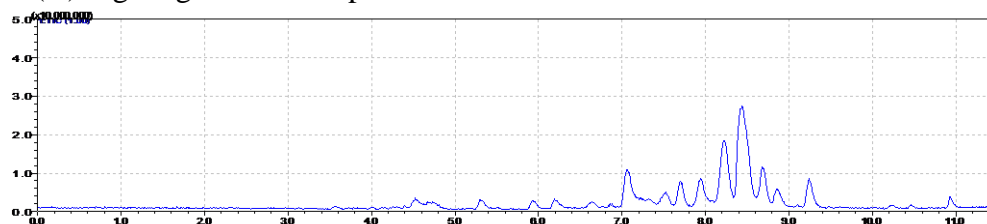
(G) Agar agar 100 % in negative mode
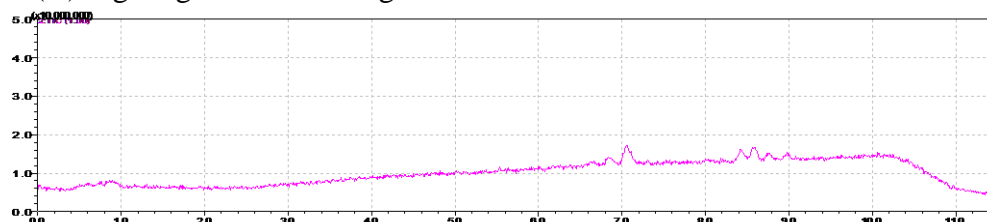
(H) Fungus 0.1 % in positive mode
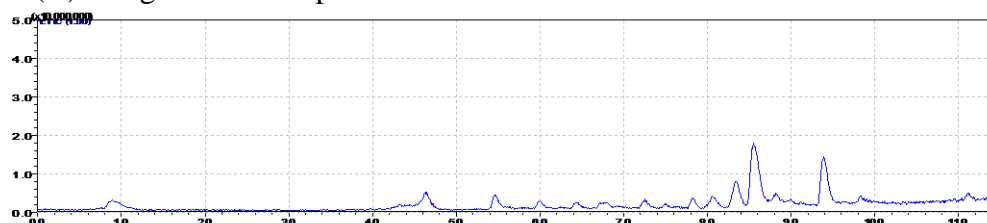
(H) Fungus 0.1 % in negative mode
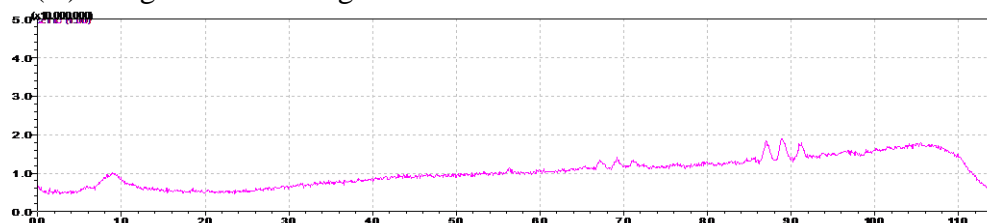
(I) Fungus 0.5 % in positive mode
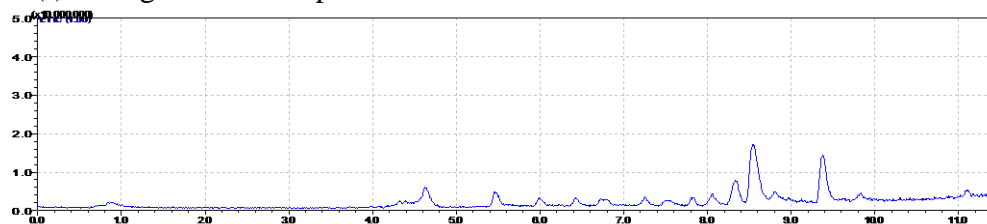
(I) Fungus 0.5 % in negative mode
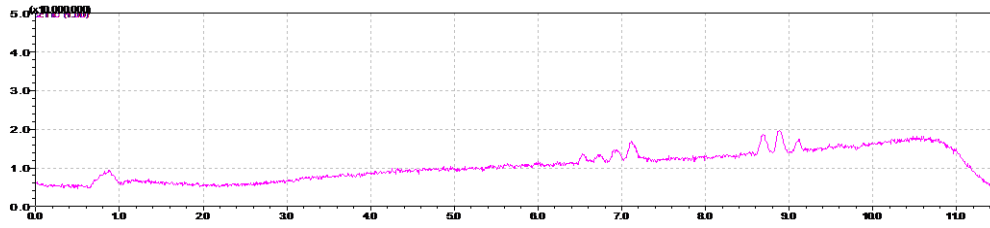
(J) Fungus 1 % in positive mode
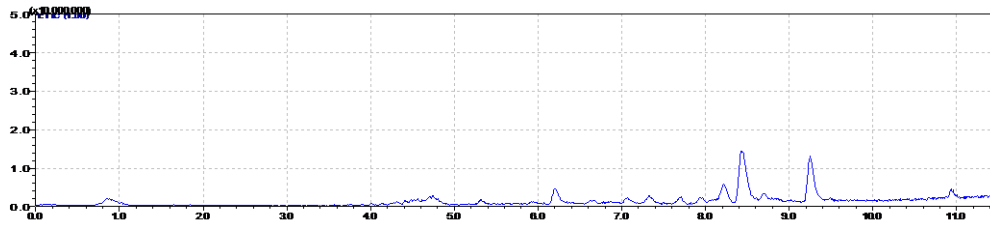
(J) Fungus 1 % in negative mode
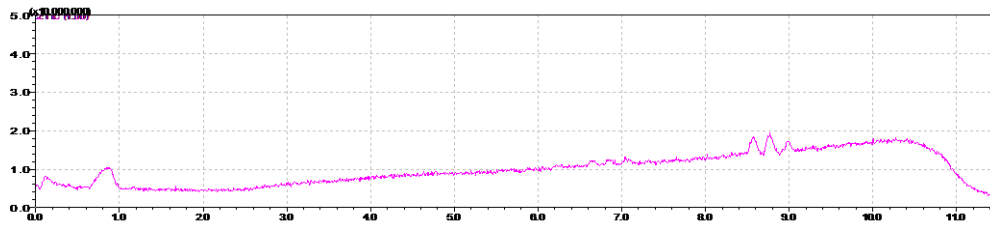
(K) Fungus 5 % in positive mode
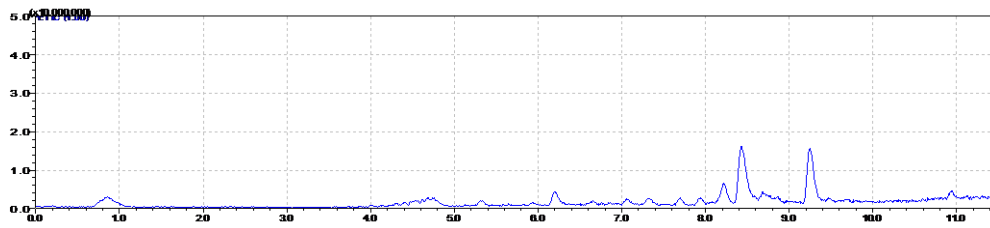
(K) Fungus 5 % in negative mode
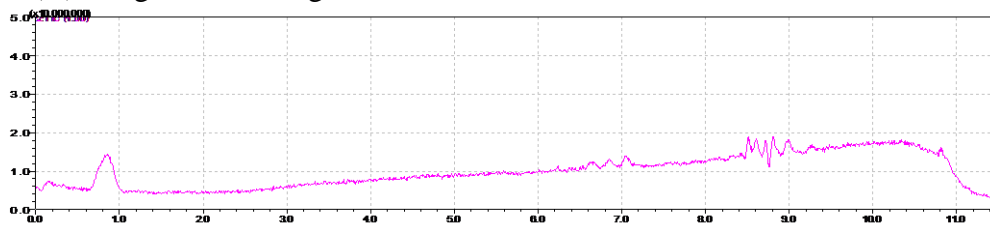
(L) Fungus 10 % in positive mode
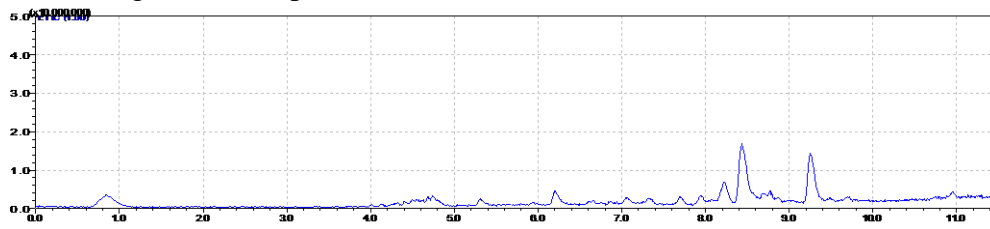
(L) Fungus 10 % in negative mode
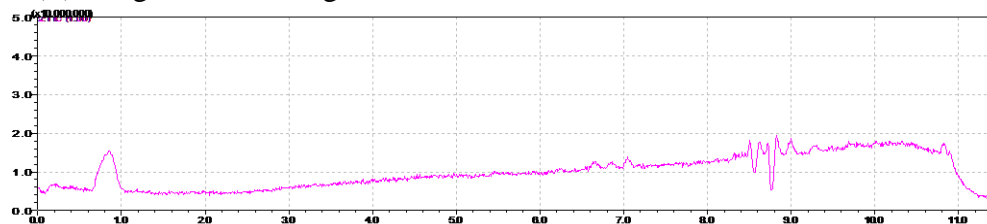
(M) Fungus 30 % in positive mode
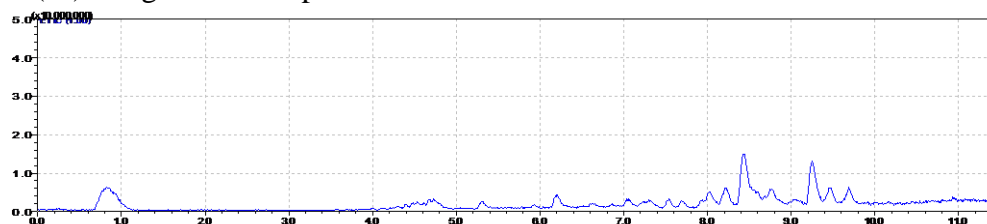
(M) Fungus 30 % in negative mode
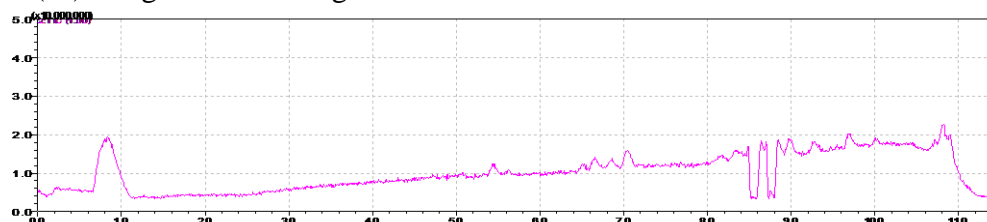
(N) Fungus 50 % in positive mode
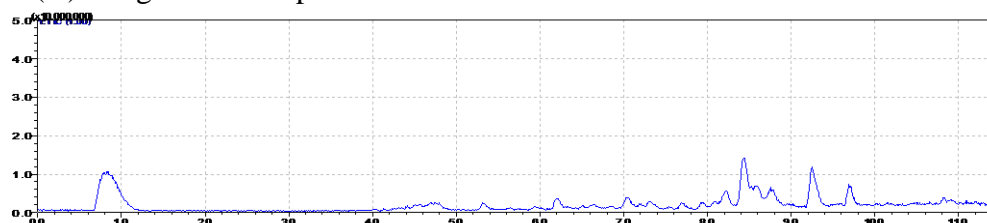
(N) Fungus 50 % in negative mode
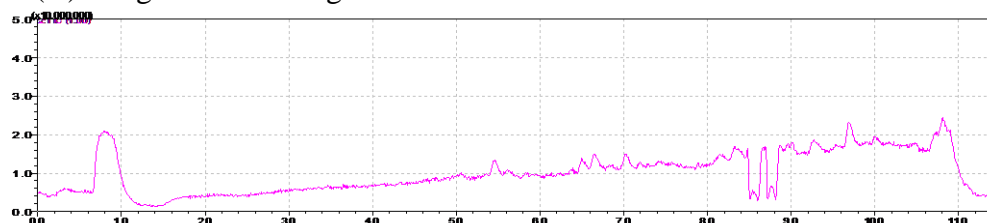
(O) Fungus 80 % in positive mode
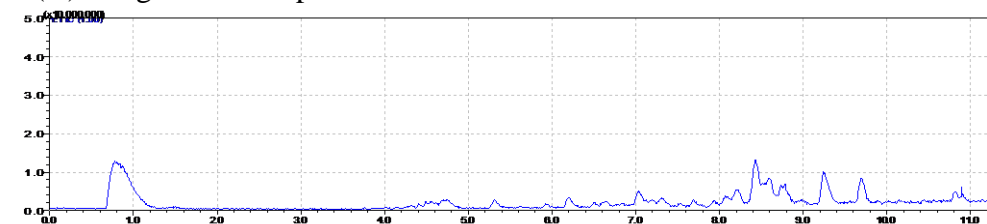
(O) Fungus 80 % in negative mode
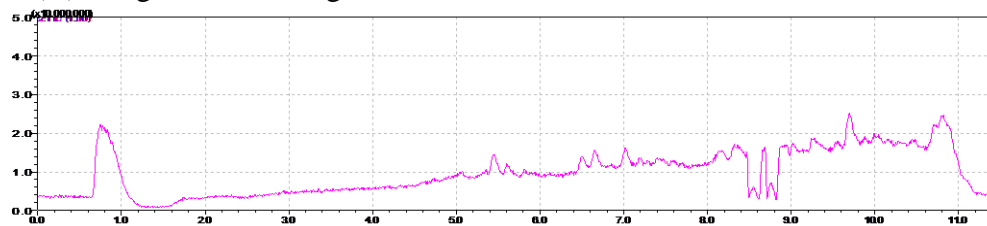
(P) Fungus 100 % in positive mode
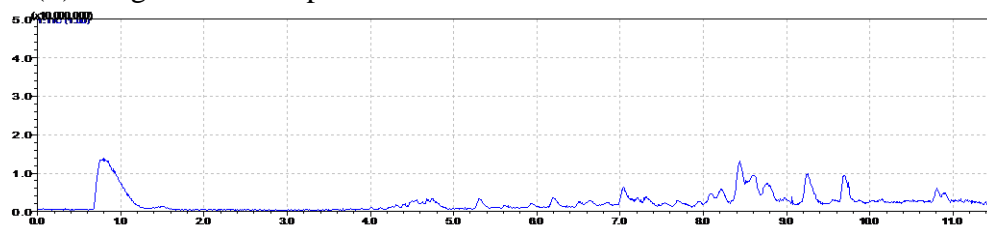
(P) Fungus 100 % in negative mode
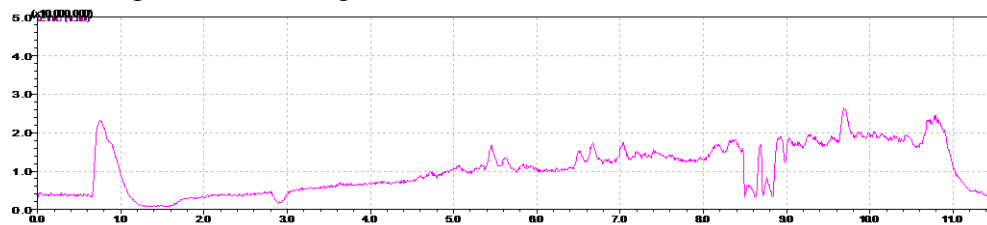
(Q) Isinglass 0.1 % in positive mode
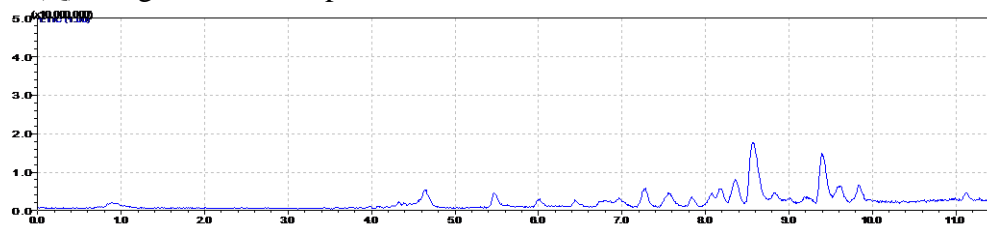
(Q) Isinglass 0.1 % in negative mode
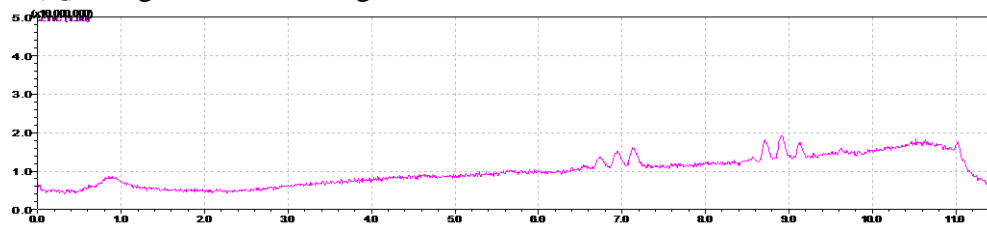
(R) Isinglass 0.5 % in positive mode
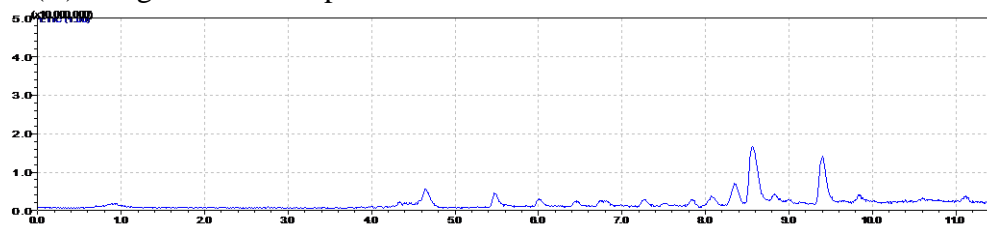
(R) Isinglass 0.5 % in negative mode
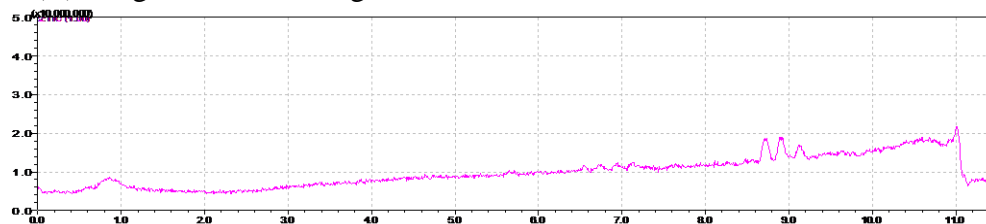
(S) Isinglass 1 % in positive mode
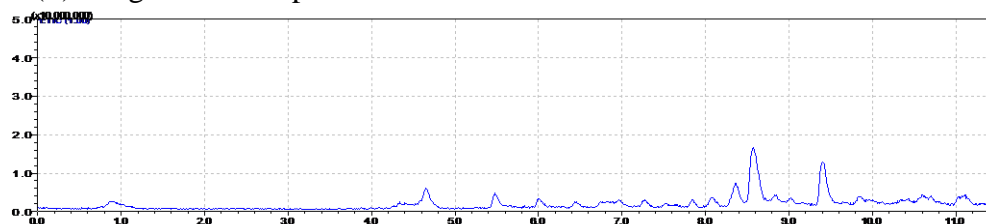
(S) Isinglass 1 % in negative mode
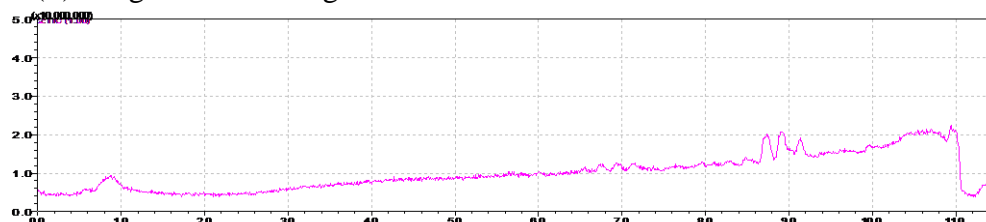
(T) Isinglass 5 % in positive mode
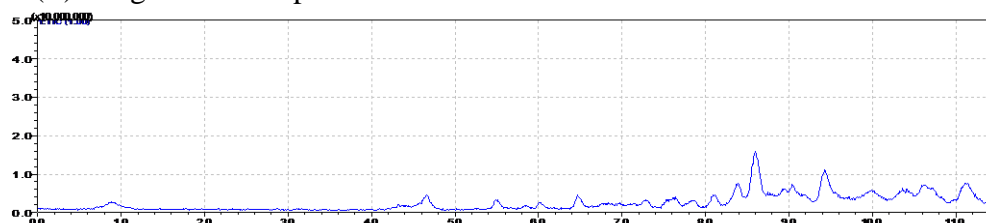
(T) Isinglass 5 % in negative mode
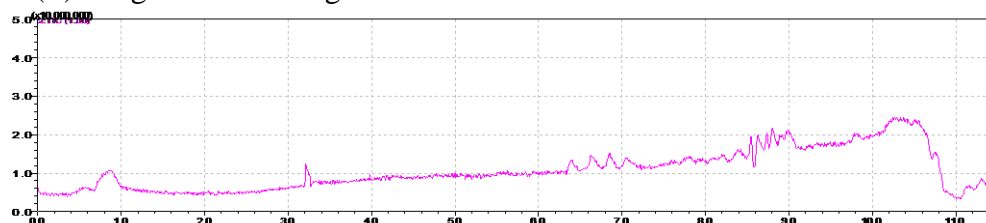
(U) Isinglass 10 % in positive mode
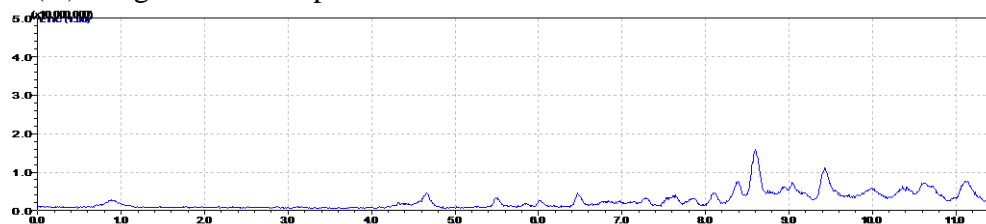
(U) Isinglass 10 % in negative mode
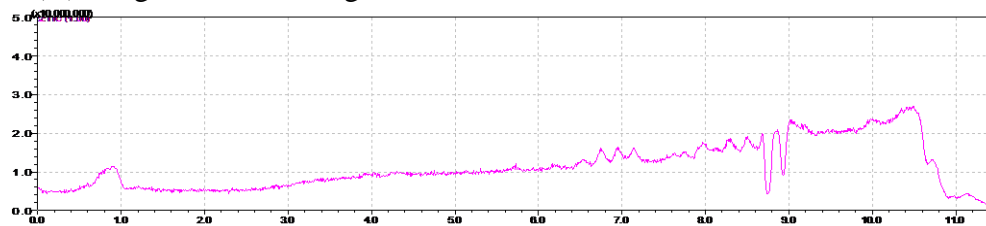
(V) Isinglass 30 % in positive mode
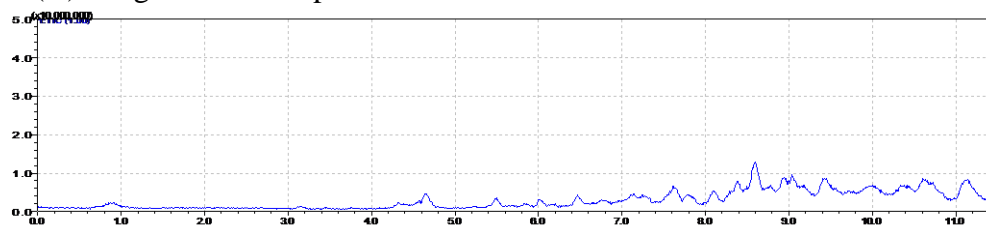
(V) Isinglass 30 % in negative mode
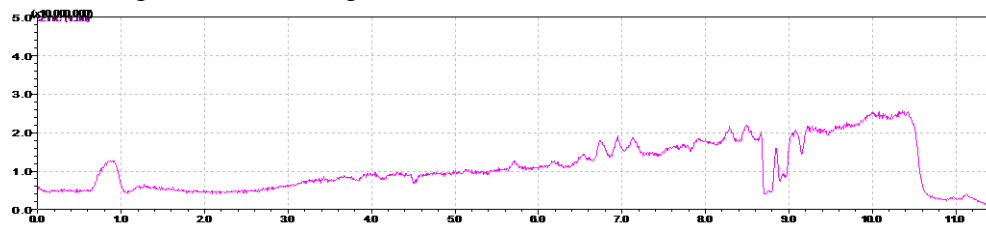
(W) Isinglass 50 % in positive mode
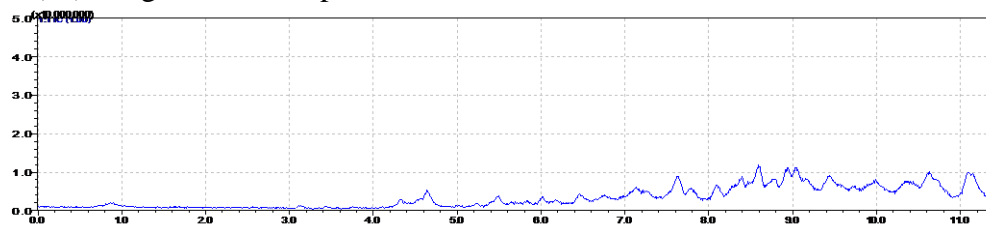
(W) Isinglass 50 % in negative mode
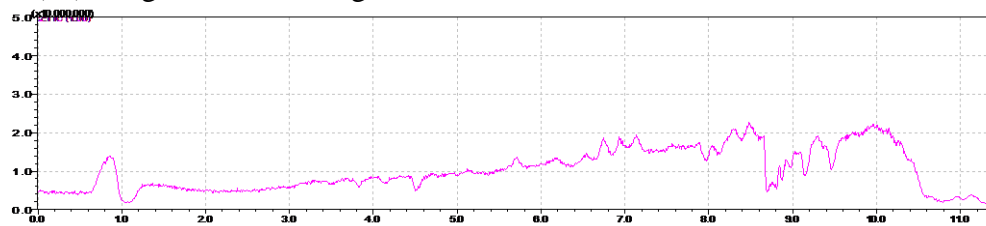
(X) Isinglass 80 % in positive mode
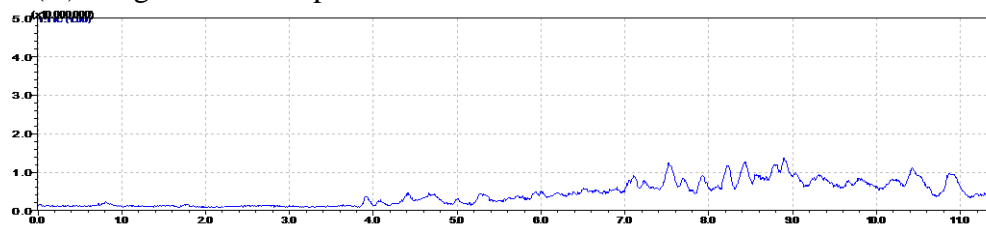
(X) Isinglass 80 % in negative mode
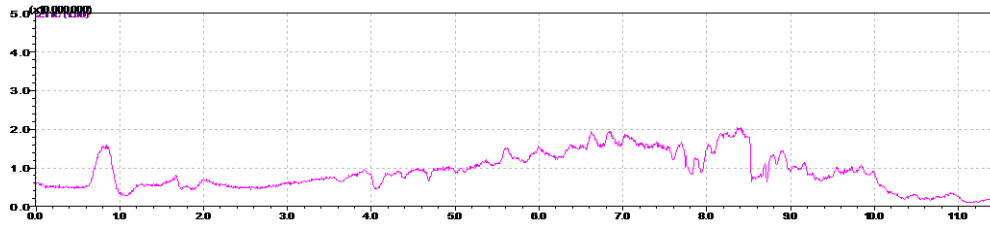
(Y) Isinglass 100 % in positive mode
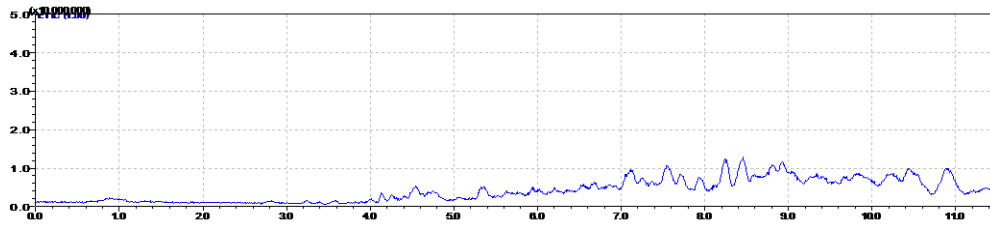
(Y) Isinglass 100 % in negative mode
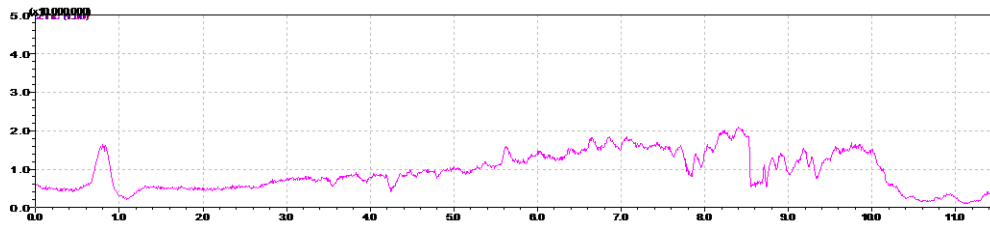
Appendix 15 PCA score plots which are subjected to different normalization methods

(A) Amino acid
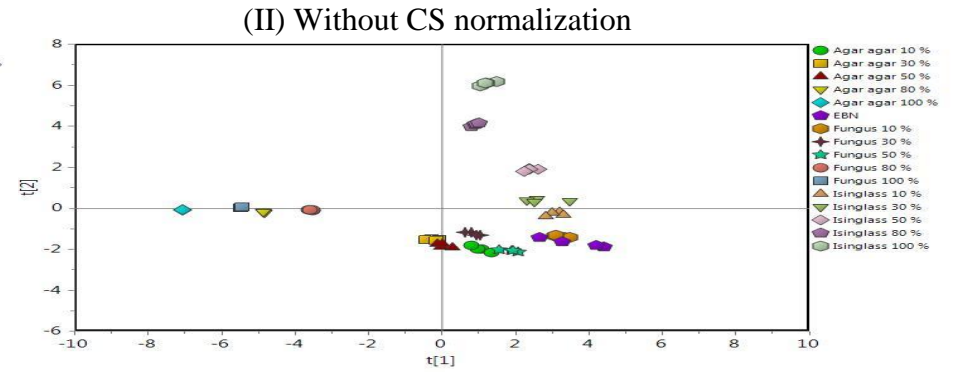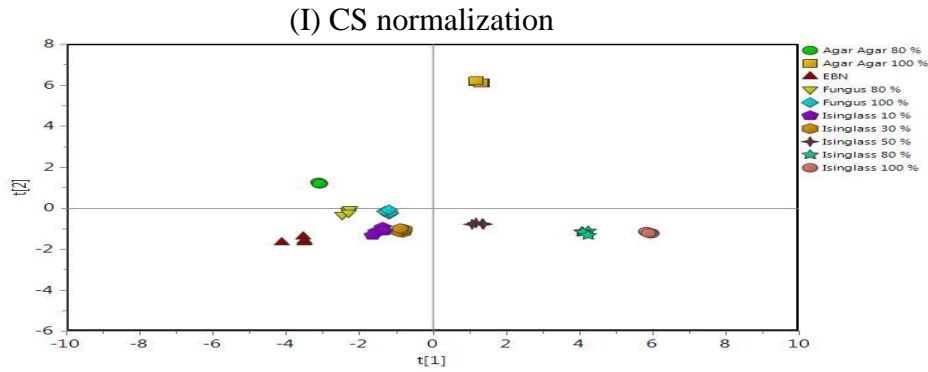
(B) Monosaccharide
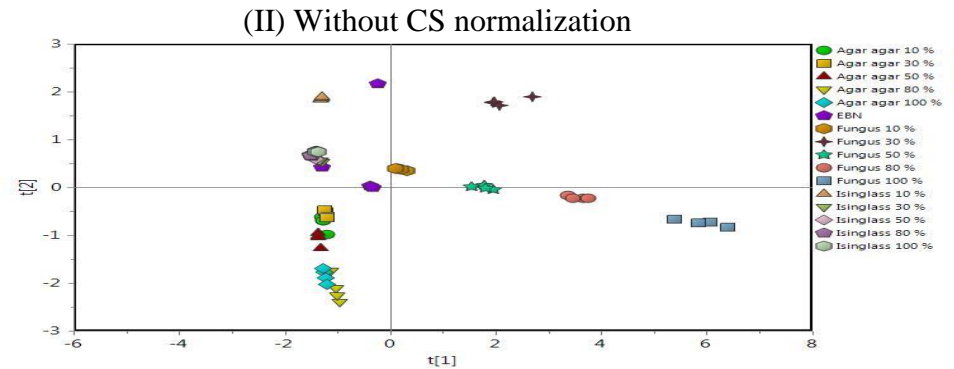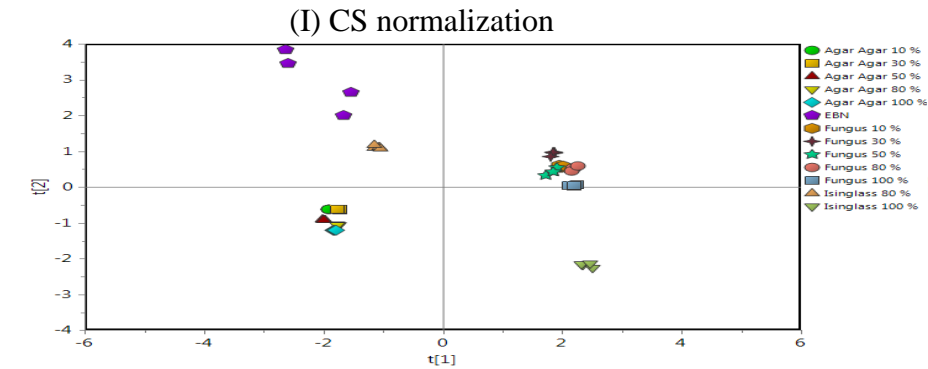
(C) Metabolite fingerprinting with GC/MS

(I) CS normalization
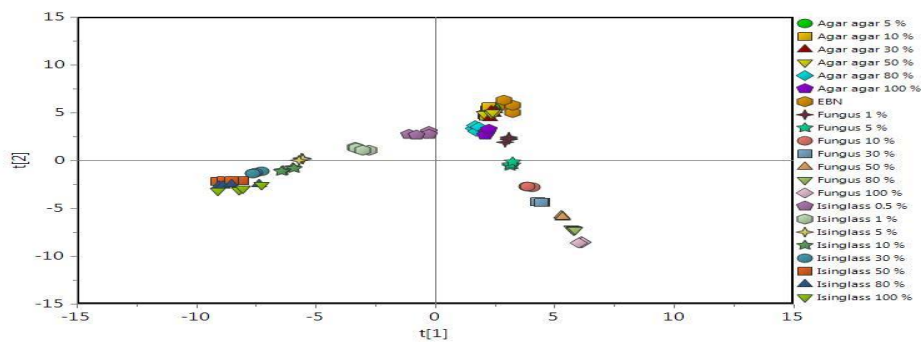
(II) Without CS normalization
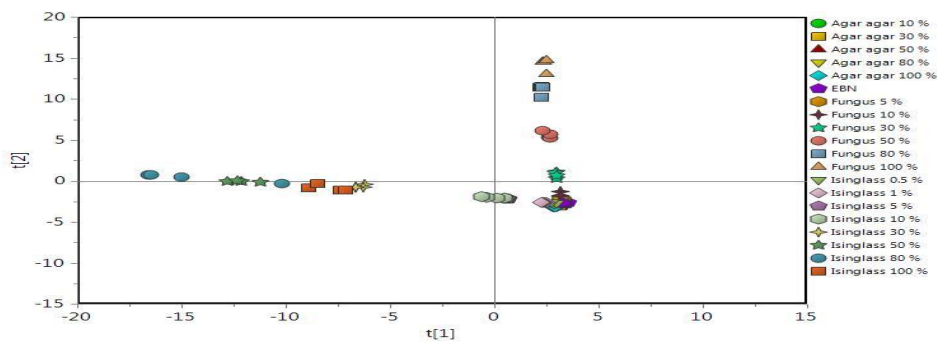
(D) Metabolite fingerprinting with LC/MS

(I) CS normalization
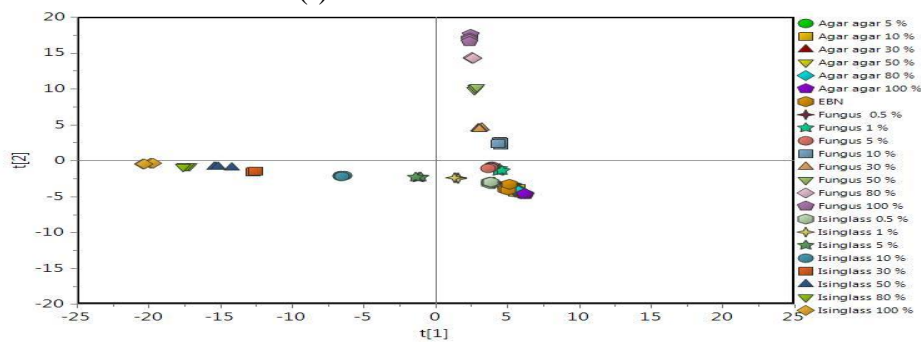
(II) Without CS normalization
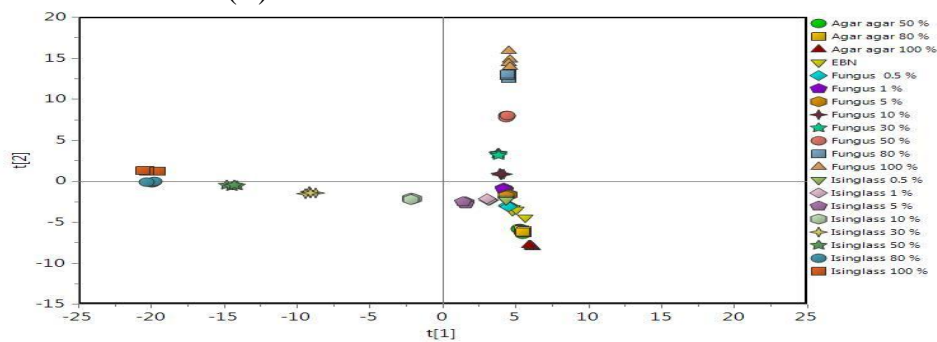
Appendix 16 PCA score plots which are subjected to different scaling methods

(A) Amino acid
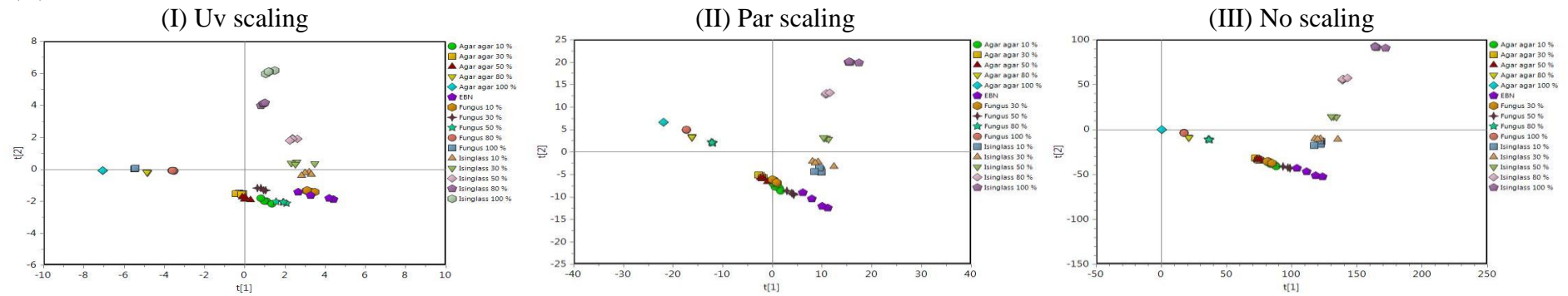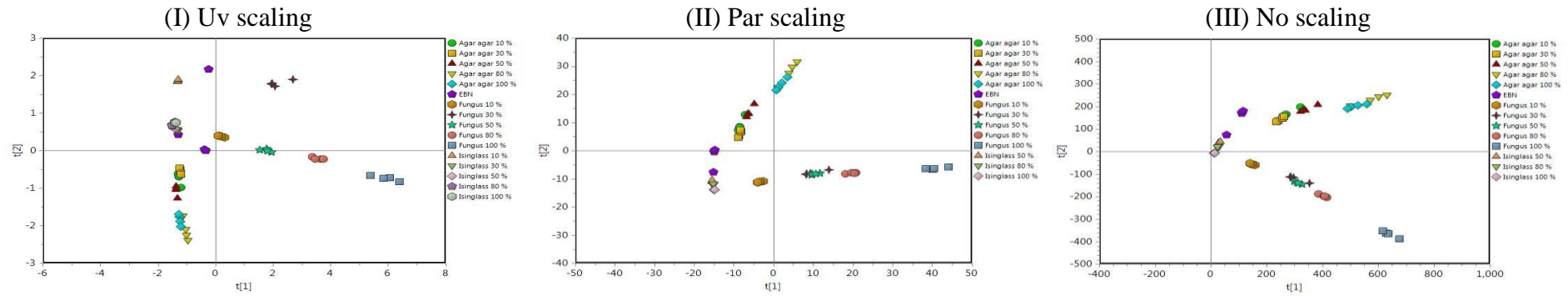
(B) Monosaccharide

(C) Metabolite fingerprinting with GC/MS

(I) Uv scaling   (II) Par scaling   (III) No scaling

(D) Metabolite fingerprinting with LC/MS

(I) Uv scaling   (II) Par scaling   (III) No scaling
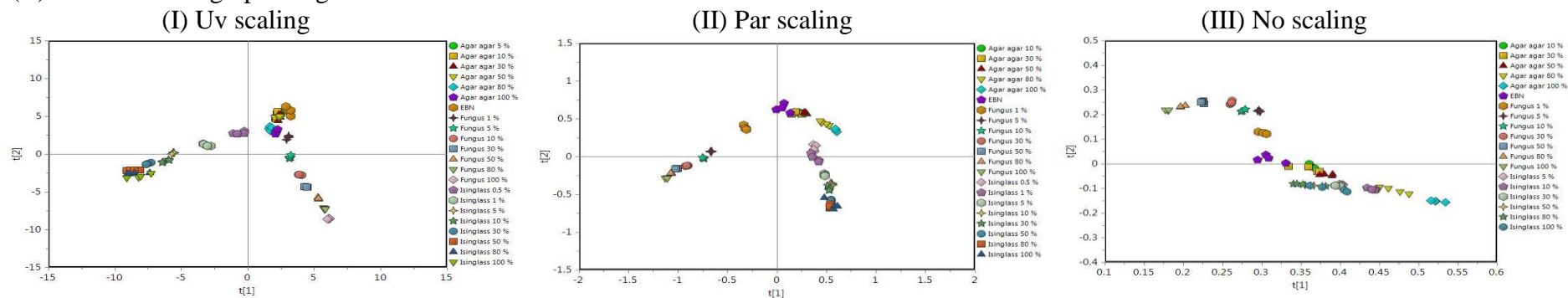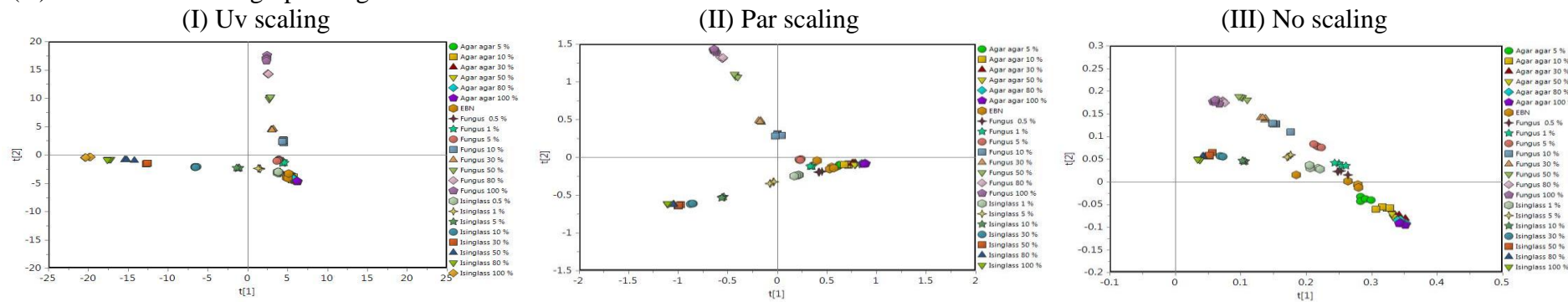
Appendix 17 OPLS-DA and PLS-DA score plots for the determiantion of multivariate analysis in qualitative discrimination
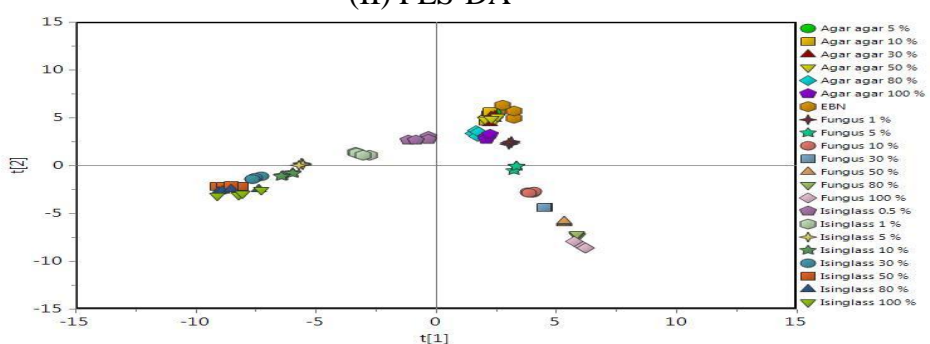
(A) Amino acid
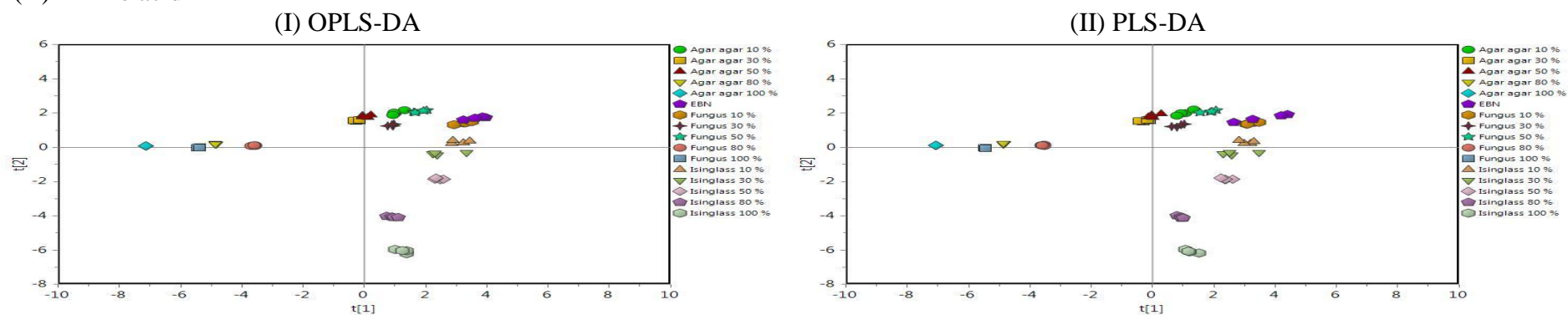
(B) Monosaccharide
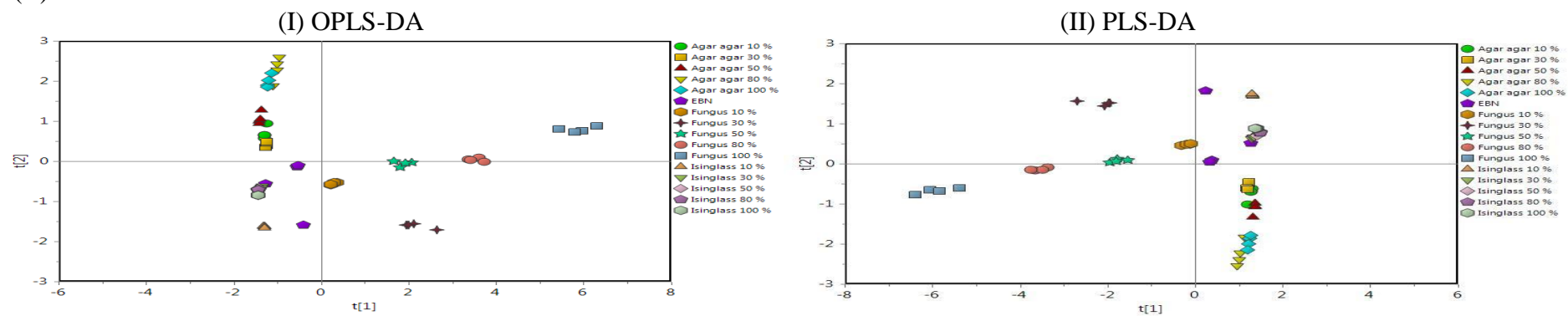
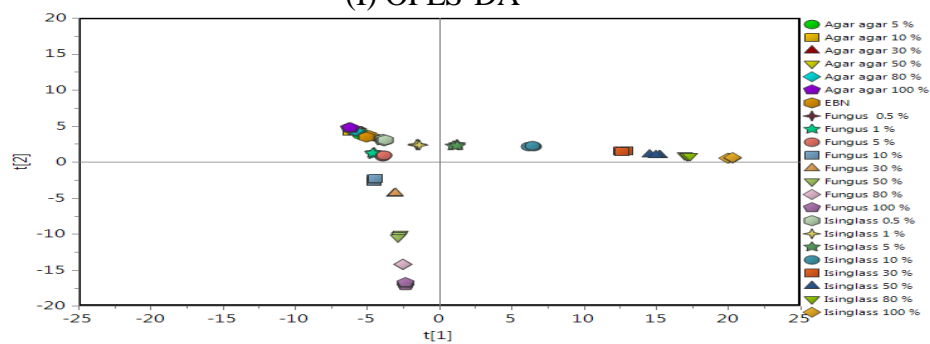(C) Metabolite fingerprinting with GC/MS
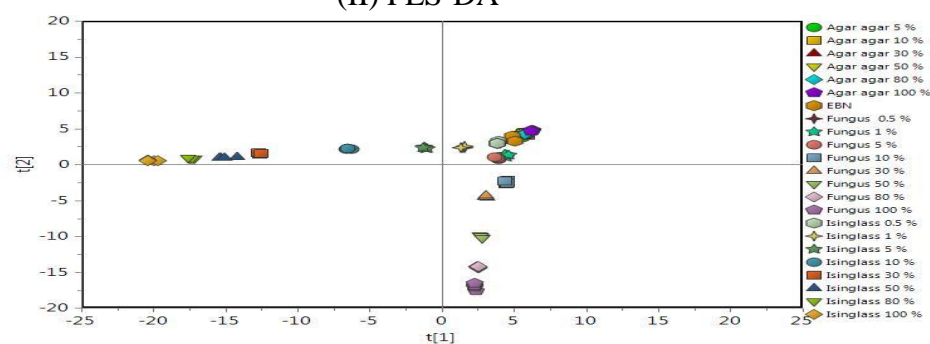
(I) OPLS-DA
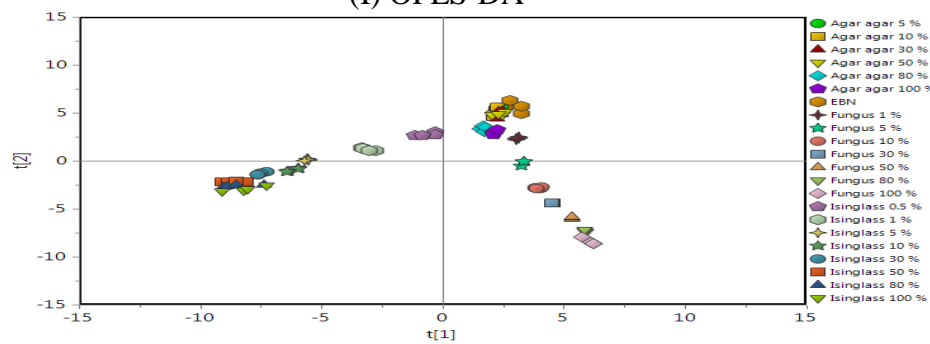
(II) PLS-DA

(D) Metabolite fingerprinting with LC/MS
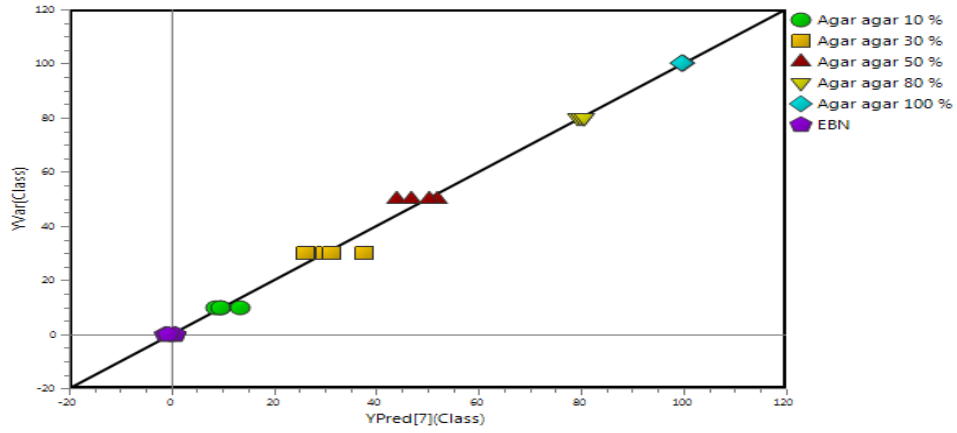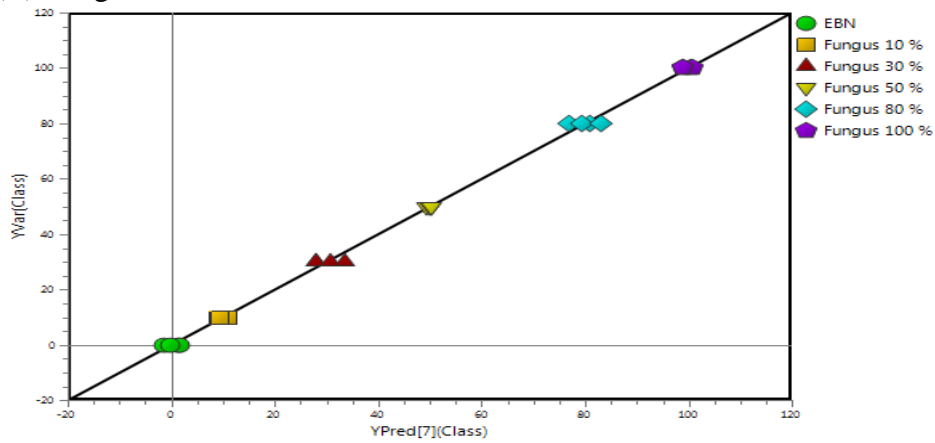
(I) OPLS-DA

(II) PLS-DA

Appendix 18 PLS regressions for the qualitative discrimination of EBN and

spiked samples
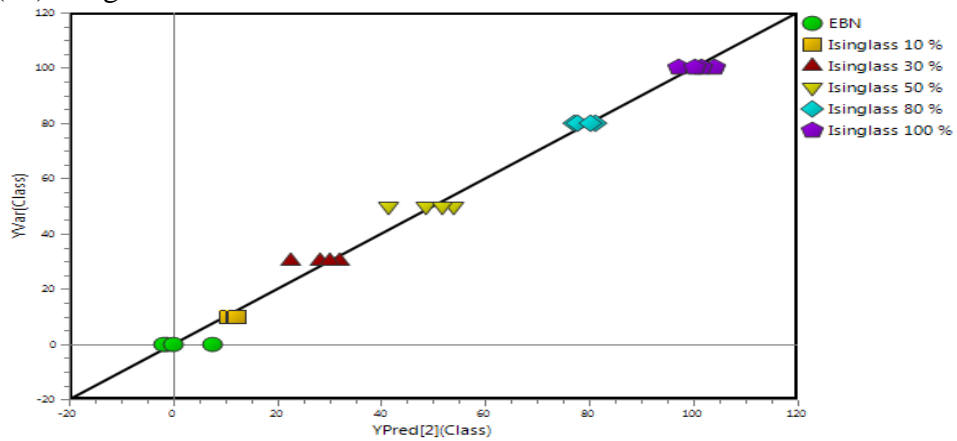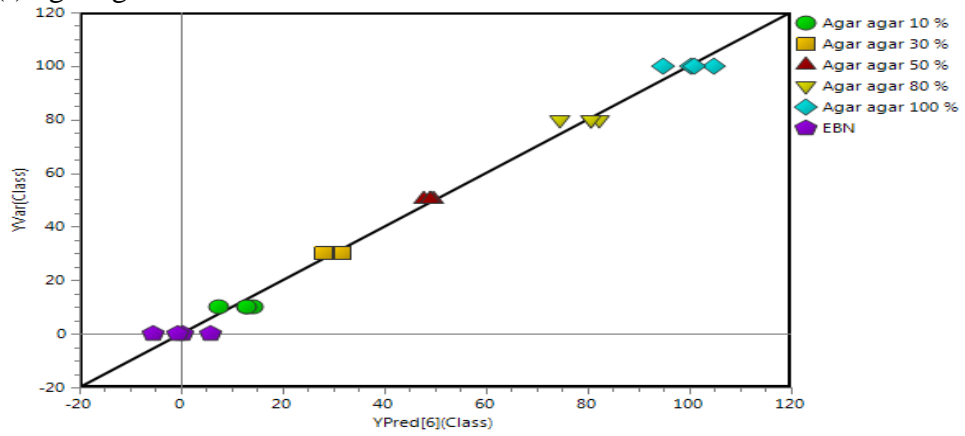
(A) Amino acid
 (I) Agar agar

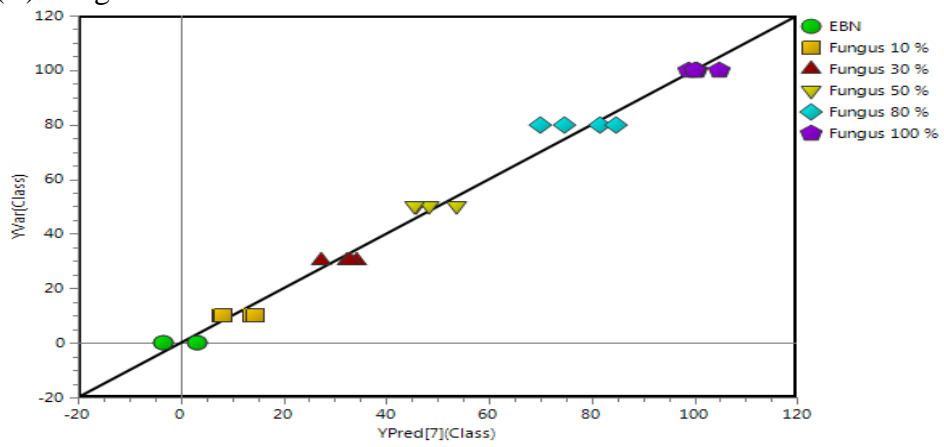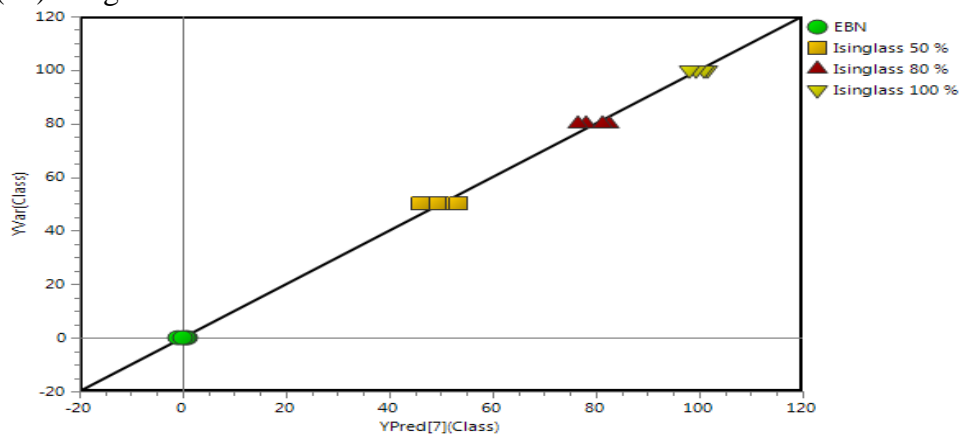(II) Fungus

(III) Isinglass

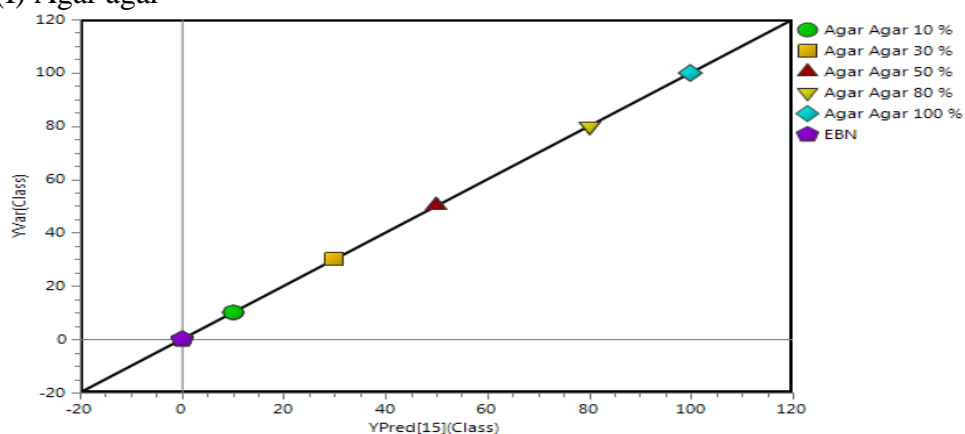(B) Monosaccharide

(I) Agar agar
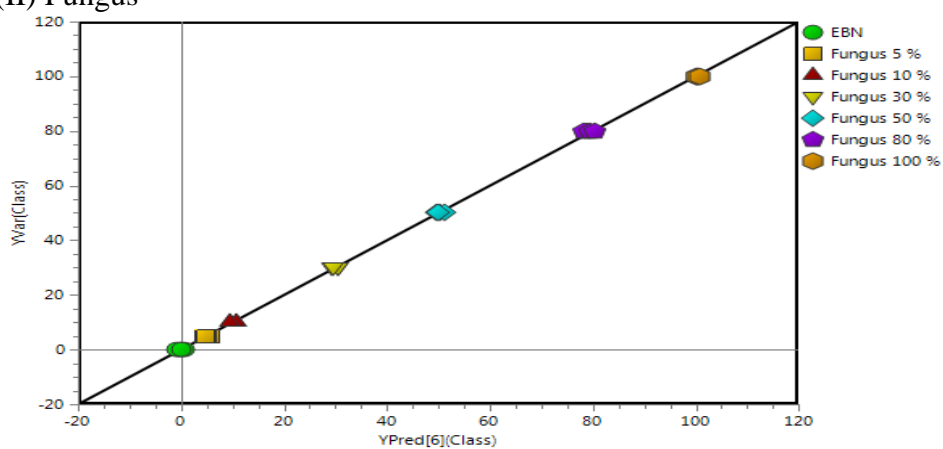

(II) Fungus

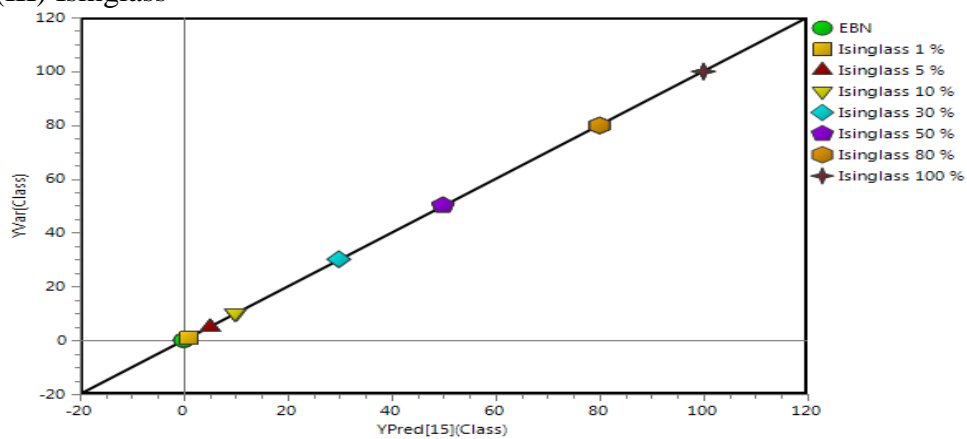(III) Isinglass

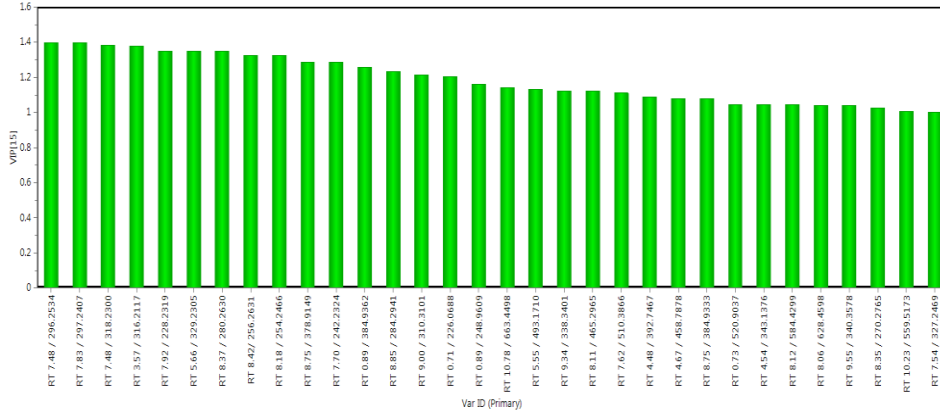(C) Metabolite fingerprinting with LC/MS

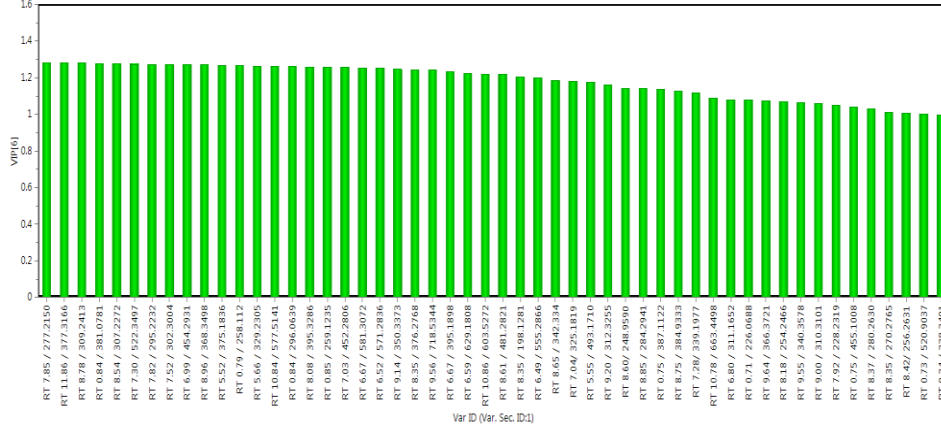(I) Agar agar

(II) Fungus

(III) Isinglass

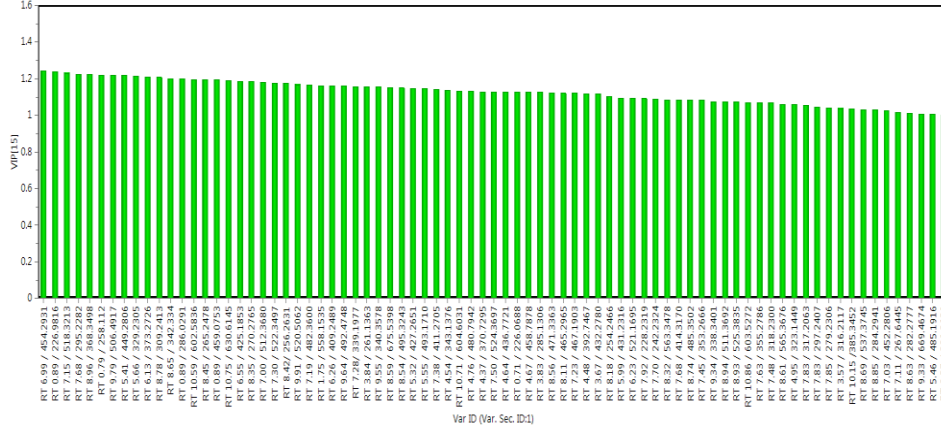**Appendix 19** VIP plots based on metabolite fingerprinting with LC/MS

(A) Metabolite fingerprinting with LC/MS
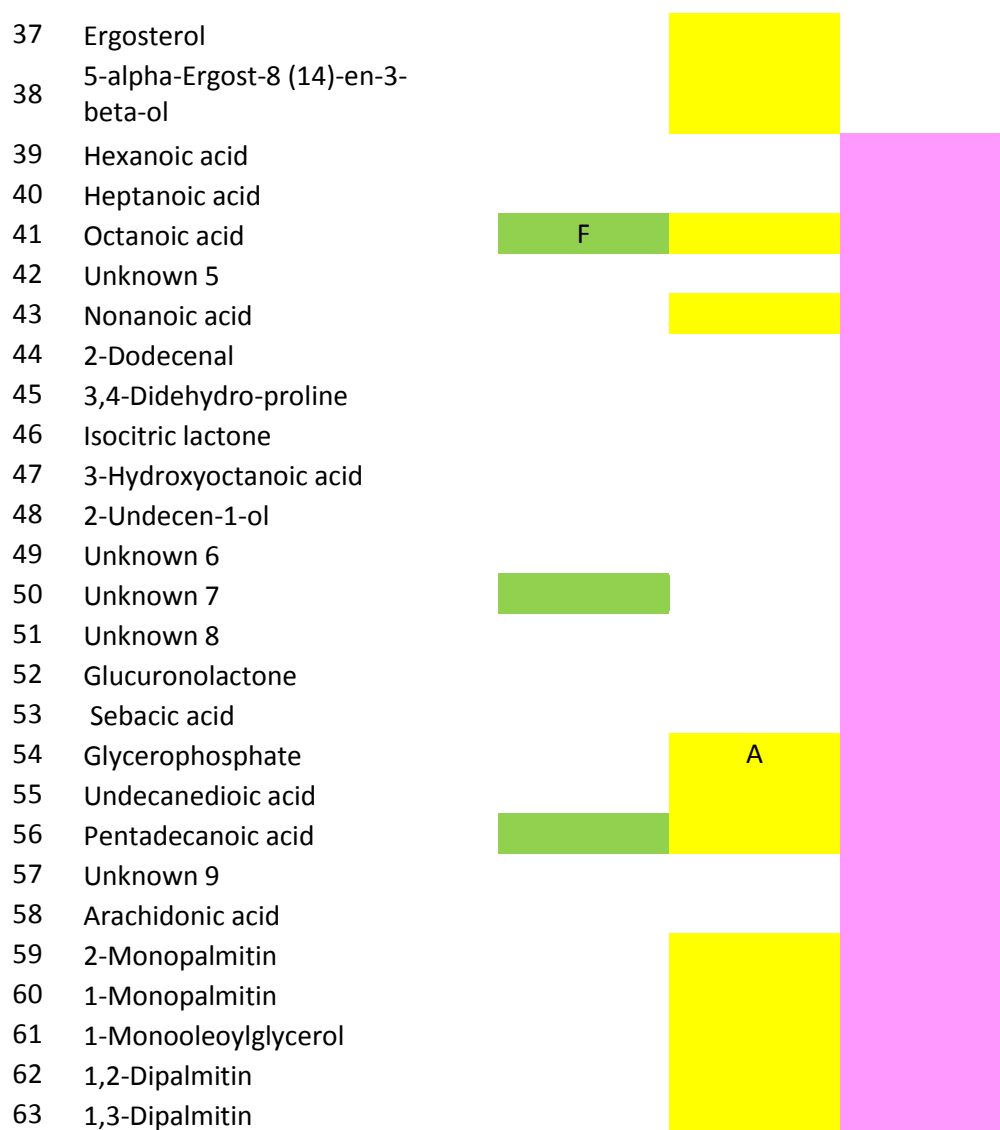

(II) Fungus

(III) Isinglass

Appendix 20 Compound table on the metabolites found in agar agar, fungus and isinglass. Metabolites determined in agar agar, fungus and isinglass is highlighted by green, yellow and pink respectively.

No.	Compound Name	Agar Agar	Fungus	Isinglass
1	Hydroxybutyric acid	F		
2	Decanoic acid			
3	5-Oxoproline			
4	Dodecanoic acid			
5	Tetradecanoic acid			
6	Methyl 2-acetamido-2-deoxy-beta-D-glucopyranoside			
7	Unknown 1			
8	Methyl 2-acetamido-2-deoxy-alpha-D-glucopyranoside			
10	(Z)-9-Hexadecenoic acid	F		
11	(E)-9-Hexadecenoic acid	F		
12	Hexadecanoic acid	F	A	
13	Heptadecanoic acid	F		
14	(Z)-9,12-Octadecadienoic acid			
15	E-9-Octadecadienoic acid	F	A	
16	E-11-Octadecadienoic acid		A	
17	Octadecanoic acid	F		
18	Eicosanoic acid			
19	Thymol-beta-D-glucopyranoside	F		
20	Cholesterol			
21	(Z)-9-Octadecenenitrile			
22	9-Octadecenamide			
23	Propanetriol			
24	Succinic acid	F	A	
25	Hydroxybutanedioic acid			
26	4-Aminobutanoic acid			
27	Xylitol	F	A	
28	Unknown 2			
29	Unknown 3			
30	D-Glucoheptono-1,4-lactone	F	A	
31	D-Glucitol			
32	D-Galactitol			
33	Myo-Inositol		A	
34	Unknown 4			
35	Tetracosanoic acid			
36	D-Turanose			

#	Metabolite		
37	Ergosterol		
38	5-alpha-Ergost-8 (14)-en-3-beta-ol		
39	Hexanoic acid		
40	Heptanoic acid		
41	Octanoic acid	F	
42	Unknown 5		
43	Nonanoic acid		
44	2-Dodecenal		
45	3,4-Didehydro-proline		
46	Isocitric lactone		
47	3-Hydroxyoctanoic acid		
48	2-Undecen-1-ol		
49	Unknown 6		
50	Unknown 7		
51	Unknown 8		
52	Glucuronolactone		
53	Sebacic acid		
54	Glycerophosphate	A	
55	Undecanedioic acid		
56	Pentadecanoic acid		
57	Unknown 9		
58	Arachidonic acid		
59	2-Monopalmitin		
60	1-Monopalmitin		
61	1-Monooleoylglycerol		
62	1,2-Dipalmitin		
63	1,3-Dipalmitin		

F represents the common metabolites between agar agar and fungus and has a VIP value greater than one in the fungus's PLS regression plot.

A represents the common metabolites between fungus and agar agar and has a VIP value greater than one in the agar agar's PLS regression plot.

LIST OF PUBLICATIONS AND MANUSCRIPTS

Journal Papers (Published)

1. Metabolite profiling of edible bird nest using gas chromatography/mass spectrometry and liquid chromatography/mass spectrometry

Chua, Y.G., Bloodworth, B.C.,Leong, L.P. and Li, F.Y.S, Journal of Mass Spectrometry, 2014, 28, 1-14.

2. Identification of edible bird's nest with amino acid and monosaccharide analysis

Chua, Y.G., Chan, S.H., Bloodworth, B.C., Leong, L.P. and Li, F.Y.S, Food chemistry

Journal Papers (submitted and under peer review)

1. Authentication of edible bird's nest with different analytical methods and multivariate analysis

Chua, Y.G., Chan, S.H., Bloodworth, B.C., Leong, L.P. and Li, F.Y.S, Food chemistry

Conference Papers

1. The use of principal component analysis for the identification of edible bird's nest

Chua, Y.G., Chan, S.H., Bloodworth, B.C., Leong, L.P. and Li, F.Y.S International Student Academic Conference, Taiwan, Hsinchu, 2012 (Oral Presentation)

2. The use of principal component analysis for identification of edible bird's nest

Chua, Y.G., Chan, S.H., Bloodworth, B.C., Leong, L.P. and Li, F.Y.S, 7th Singapore International Chemistry Conference, Singapore, 2012 (Poster Presentation)

3. The use of principal component analysis for identification of edible bird's nest

Chua, Y.G., Chan, S.H., Bloodworth, B.C., Leong, L.P. and Li, F.Y.S, XVIIth European Conference on Analytical Chemistry, Poland, Warsaw, 2013 (Poster Presentation)

4. Metabolite profiling of edible bird nest using gas chromatography mass spectrometry and liquid chromatography mass spectrometry

Chua, Y.G., Chan, S.H., Bloodworth, B.C., Leong, L.P. and Li, F.Y.S, 62nd ASMS Conference on Mass Spectrometry and Allied Topics, United States of America, 2014 (Poster Presentation)