ISOLATION, CHARACTERIZATION AND IDENTIFICATION OF ANTI-AMYLASE PEPTIDES FROM CUMIN SEEDS (*CUMINUM CYMINUM*)

SIOW HWEE LENG

UNIVERSITI SAINS MALAYSIA

2018

ISOLATION, CHARACTERIZATION AND IDENTIFICATION OF ANTI-AMYLASE PEPTIDES FROM CUMIN SEEDS (*CUMINUM CYMINUM*)

by

SIOW HWEE LENG

Thesis submitted in fulfillment of the requirements
for the degree of
Doctor of Philosophy

May 2018

ACKNOWLEDGEMENT

First and foremost, I would like to take this opportunity to express my gratitude to my main supervisor, Assoc. Prof. Dr. Gan Chee Yuen for his continued guidance, advice and encouragement. During my research tenure, he contributed to a rewarding research experience by giving me intellectual freedom in my work, engaging me with new ideas, and demanding a high quality of work in all my endeavours. Additionally, I would like to thank my co-supervisor, Dr. Lim Theam Soon for his insightful comment, guidance and constructive feedback on my research work. Special thanks also go to Dr. Choi Sy Bing and Dr. Tye Gee Jun for their kind advice and help. Without their helping hands, this work would have been impossible and things have not progressed as planned.

I would like to convey my sincere appreciation to all the staff and member in the Analytical Biochemistry Research Centre (ABrC), Universiti Sains Malaysia (USM), for their assistance and cooperation throughout the completion of this project. I am heartily thankful to my fellow labmates, especially Ms. Ngoh Ying Yuan and Ms. Nurul Hidayatul Afifah, for their friendship, kindness, and unyielding supports. I have also indebted to many friends who had always given their endless moral support and encouragement during critical periods of my doctoral studies.

I am grateful for the funding sources that allow me to undertake this research: The MyPhD scholarship from Ministry of Education Malaysia, The RUI (1001/CAATS/814257) and ERGS (203/CAATS/6730103) grants from USM. Last but not least, I thank my beloved family members for their unconditional love, support and encouragement. Their faith in me has kept me motivated and enthused to complete my doctoral degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS / ABBREVIATIONS

2-D	two-dimensional
3-D	three-dimensional
3^k	three-level factorial design
AABA	l-α-amino-n-butyric acid
ABTS	2,2'-azino-bis-(3-ethylbenzothiazoline-6-sulfonic acid)
ACE	angiotensin-converting enzyme
ALC	average local confidence
ANOVA	analysis of variance
AQC	6-aminoquinolyl-n-hydroxysuccinimidyl carbamate
ATP	adenosine triphosphate
ATR	attenuated total reflectance
BAA	*Bacillus amyloliquefaciens* α-amylase
BBD	Box-Behnken design
BLAST	basic local alignment search tool
BSA	bovine serum albumin
C=O	carbon-oxygen double bond
Ca^{2+}	calcium ion
C-C	carbon-carbon bond
CCD	central composite design
CCKR	cholecystokinin receptor
CDK	cyclin-dependent kinase
CDL	curved desolvation line
cDNA	complementary deoxyribonucleic acid

CH_2	methylene
CH_3	methyl
CID	collision-induced dissociation
Cl^-	chloride ion
C-N	carbon-nitrogen bond
C-O	carbon-oxygen bond
CSP	cumin seed peptide
Cu^{2+}	cupric ion
CV	coefficient of variation
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNS	3,5-dinitrosalicylic acid
DPPH	2,2-diphenyl-1-picrylhydrazyl
E/TN	ratio of essential to total amino acids
EDC	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
E/S	enzyme-to-substrate
ESI	electrospray ionization
ETA	ethanolamine
F	F-statistics
FFD	full factorial design
FSG	fish skin gelatine
FTIR	Fourier transform infrared spectroscopy
G1	gap 1 phase

GELFREE	gel-eluted liquid fraction entrapment electrophoresis
GK	glucokinase
GLUT2	glucose transporter 2
H_2	hydrogen gas
HCl	hydrochloric acid
HMG-CoA	3-hydroxy-3-methylglutaryl-coenzyme A
HPLC	high-performance liquid chromatography
HRP	horseradish peroxidise
HSA	human salivary α-amylase
IC_{50}	half maximal inhibitory concentration
IR	infrared radiation
ITMS	ion trap mass spectrometry
K_m	Michaelis-Menten constant
LC/MS	liquid chromatography-mass spectrometry
LTQ	linear trap quadrupole
MAP	mitogen-activated protein
MES	2-morpholin-4-ylethanesulfonic acid
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
MW	molecular weight
n	number in a trial or sample
NaCl	sodium chloride

NaOH	sodium hydroxide
NCD	non-communicable disease
N-H	nitrogen-hydrogen bond
NHMS	national health and morbidity survey
nLC	nanoflow liquid chromatography
OD	optical density
p	probability
PBS	phosphate-buffered saline
PBS-T	phosphate-buffered saline containing tween 20
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque-forming units
pI	isoelectric point
PPA	porcine pancreatic α-amylase
PYK	pyruvate kinase
q	radio frequency collision quadrupole
R^2	coefficient of determination
R^2_{adj}	adjusted coefficient of determination
RSM	response surface methodology
[S]	substrate concentration
SAR	structure-activity relationship
SD	standard deviation
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SGLT1	sodium-glucose cotransporter-1
sPLA2	secreted phospholipase A2

S-PSSM	spatial position-specific scoring matrix
TOF	time of flight
V_0	initial velocity
V_{max}	maximum velocity
WHO	World Health Organization
X	independent variable
x	coded independent invariable
Y	response variable
α	alpha
β	beta

PEMISAHAN, PENCIRIAN DAN IDENTIFIKASI PEPTIDA ANTI-AMILASE DARIPADA JINTAN PUTIH (*CUMINUM CYMINUM*)

ABSTRAK

Objektif utama tesis ini adalah untuk menerokai peptida anti-amilase baharu daripada jintan putih (*Cuminum cyminum*). Biji rempah ini mengandungi protein yang tinggi dan kaya penggunaannya dalam perubatan tradisi, terutamanya dalam perawatan kencing manis. Maka, jintan putih boleh dijadikan sebagai protein prekursor untuk penghasilan peptida bioaktif yang dapat memberikan manfaat kesihatan selain daripada nilai pemakanannya. Fasa pertama penyelidikan ini merangkumi kajian asas tentang pengekstrakan dan pencirian protein terlarut daripada jintan putih. Melalui kaedah metodologi tindak balas permukaan (RSM), hasil maksimum protein terlarut sebanyak 44.98 mg protein terlarut/g sampel telah diperolehi dengan menggunakan 0.6 jam, 26.3 ˚C dan nisbah penampan-kepada-sampel 10 mL/g. Berdasarkan kandungan asid amino, protein mengandungi 37.4% asid amino yang bersifat hidrofobik dan 62.6% asid amino yang bersifat hidrofilik. Maka, protein tersebut dicadangkan berpotensi untuk merencatkan aktiviti α-amilase. Walau bagaimanapun, protein terlarut hanya memberikan aktiviti perencatan lebih kurang 6.7% terhadap α-amilase. Kekurangan aktiviti perencatan ini disebabkan oleh kekangan struktur tertiari protein. Oleh demikian, enzim Protamex® digunakan untuk membebaskan jujukan peptida bioaktif daripada rantai protein. Sekali lagi, RSM telah digunakan untuk mengoptimumkan proses hidrolisis. Keadaan optimum ialah 42.6 ˚C, 1.83 jam, dan nisbah enzim-kepada-substrat sebanyak 70 (U/mg) telah mengakibatkan 2.5 kali lebih tinggi perencatan activitinya berbanding dengan protein utuh. Kemudian, pemeringkatan hidrolisat protein dilakukan, diikuti dengan

pengenalpastian jujukan peptida. Hanya pecahan yang mengandungi saiz peptida yang terkecil (<15 kDa) memaparkan perencatan activiti α-amilase, dan dalam pecahan ini, 56 jujukan peptida baru (ALC> 80%) telah dikenalpasti dengan menggunakan spektrometri jisim LTQ-Orbitrap. Dengan tujuan menangani masalah kaedah konvensional yang melibatkan kerja yang banyak dan memakan masa, suatu sistem penapisan yang menggabungkan teknologi Bioinformatik-paparan faj telah digunakan untuk mengenalpasti calon peptida yang bersifat anti-kencing manis dalam kajian seterusnya. CSP4 (RCMAFLLSDGAAAAQQLLPQYW) and CSP6 (DPAQPNYPWTAVLVFRH) telah dikenalpasti berpotensi sebagai perencat enzim α-amilase. Keselamatan dan kecekapan dalam perencatan yang berkaitan dengan penggunan CSPs tersebut telah dinilai dengan meggunakan sel model AR42J yang dapat merembeskan α-amilase. Kedua-dua CSPs menunjukkan bahawa aktiviti perencatan dan keracunan terhadap sel bergantung kepada dos, dan CSPs tersebut dapat mengekalkan aktiviti perencatan (> 50%) sehingga 2 jam. Analisis kinetik menunjukkan bahawa aktiviti perencatan yang dipamerkan oleh CSPs ialah jenis perencatan tidak kompetitif bercampur. Keputusan ini menyokong hipotesis bahawa pelbagai tapak pengikatan oleh peptida telah berlaku dalam α-amilase yang dapat menyumbangkan tindakan perencatan tersebut. Kesimpulannya, kajian ini telah berjaya mengenalpasti perencat α-amilase yang baharu daripada benih jintan putih dan mereka berpotensi untuk rawatan kencing manis.

ISOLATION, CHARACTERIZATION AND IDENTIFICATION OF ANTI-AMYLASE PEPTIDES FROM CUMIN SEEDS (*CUMINUM CYMINUM*)

ABSTRACT

The main objective of this thesis was to discover novel anti-amylase peptides from cumin (*Cuminum cyminum*) seeds. This spice seed is high in protein and has a long history of use in traditional medicine, especially in treating diabetes, making it potential precursor proteins for producing peptides with potent bioactivities that can impart health-promoting effects beyond their nutritional values. The first phase of this study comprises fundamental studies on the extraction and characterization of soluble protein from cumin seeds. Through response surface methodology (RSM), a maximum yield of 44.98 mg soluble protein/g defatted sample was obtained under a condition of 0.6 h, 26.3 ˚C, and buffer-to-sample ratio of 10 mL/g. Based on its amino acid composition, the hydrophobic (37.4%) and hydrophilic (62.6%) characters of the protein suggest a high potential for causing inhibitory interactions with α-amylase. However, the soluble proteins only gave an approximately 6.7% inhibitory activity against α-amylase. The lack of inhibition was likely related to the structural constraints of the protein tertiary structure, and thus, Protamex®-assisted hydrolysis was applied to liberate bioactive peptide sequences from the protein chain where they are encrypted. Again, RSM was adopted to optimize the hydrolysis process. The optimum condition was 42.6 ˚C, 1.83 h, and enzyme-to-substrate ratio of 70 (U/mg), resulting in a 2.5-fold higher inhibitory activity compared to the intact proteins. The optimized protein hydrolysates were subsequently subjected to fractionation, followed by peptide sequence identification. The smallest peptide

fraction (< 15 kDa) was the only fraction to display α-amylase inhibitory activity, and in this fraction, 56 peptide sequences (ALC > 80%) were identified using tandem mass spectrometry LTQ-Orbitrap. To address labour-intensive and time-consuming problems in conventional methods, an effective screening system that comprises an integrated Bioinformatics-phage display approach was developed for identifying potential candidates in the following study. CSP4 (RCMAFLLSDGAAAAQQLLPQYW) and CSP6 (DPAQPNYPWTAVLVFRH) were identified as the potent α-amylase inhibitors. Their safety and inhibitory efficacies were assessed by adopting models of AR42J α-amylase-secreting cell. Both CSPs exhibited a dose-dependent α-amylase inhibition and cytotoxicity against the cells, and were able to retain their initial inhibition (>50%) up to 2 h. Kinetic analysis indicated a mixed-type non-competitive inhibition, supporting the hypothesis that multiple peptide-binding sites within the α-amylase exist, each of which may substantially contribute to the inhibitory action of the CSPs. In conclusion, this study has successfully identified novel and potent α-amylase inhibitors from cumin seeds, which are potentially useful for diabetes treatment.

CHAPTER 1

INTRODUCTION

1.1 Background and rationale

Decades of research have uncovered the intimate relationship among diet, nutrition and health (Danquah & Agyei, 2012). Noted that foods are not intended to only satisfy hunger and offer essential nutrients for humans, but also to have beneficial physiological effects beyond the provision of the basic nutritional values (Menrad, 2003). There is a mounting scientific evidence suggesting that the therapeutic properties of these functional foods have been accredited to the presence of biologically active components, which could confer health benefits or desirable bioactivities by modulating one or more metabolic processes or pathways in the body (Swanson, 2003). Hence, these bioactive components may play an outstanding role in various stages of disease control that involve multiple progressive steps, from initiation to development, thus making them to be effectively applied in the treatment and prevention of numerous diseases (Wildman, 2001). This has led to a growing interest in attaining a clear understanding and a strong confidence level in the scientific criteria that are utilized to document physiological effects and health claims of these dietary components (Abuajah *et al.*, 2015), which makes it an exciting research topic in this dissertation.

Peptides are small protein fragments that appear as the most widely studied and most functionally diverse food-derived biomolecules (Crowe & Francis, 2013). The term “bioactive peptide” refers to specific protein fragments containing amino acid sequences of between 2-30 amino acids in length that can confer a

positive, “hormone-like” response to the consumer with profound health-promoting effects *in vivo* (Korhonen & Pihlanto, 2003). However, their relevance depends on the bioactivities they possess once they are liberated from their native protein. These functional peptides are typically inactive within the sequence of protein chain due to structural constraints, and they can be liberated by digestive enzymes during gastrointestinal transit or by fermentation or ripening during food processing (Dziuba & Dziuba, 2014). Once released as independent entities, these peptides may play a crucial role in the living body system by coordinating intra and inter-cellular communications and cellular functions (Korhonen & Pihlanto, 2003). Hence, the concept of food-borne bioactive peptide as a new and promising therapeutic strategy has received growing interest due to its advantageous properties, such as high affinity, strong specificity for targets, and low toxicity profile over protein and other small-molecule drugs (Agyei & Danquah, 2011). Depending on the structure and sequence of amino acids, these food-derived peptides have been reported to exert a diverse range of physiological properties, including antidiabetic, anticancer, antihypertensive, antioxidative, and cardio protective effects, which is notably useful for the treatment of non-communicable diseases (NCDs), such as diabetes, cancer, as well as cardiovascular and respiratory diseases (Zhang & Beynen, 1993; Zhang *et al.*, 2010; Stuknyte *et al.*, 2011; Najafian & Babji, 2012; Suarez-Jimenez *et al.*, 2012; Zambrowicz *et al.*, 2013b).

In this case of study, peptides with antidiabetic properties are the central research focus. Diabetes, one of the most common NCDs globally, is growing at an alarming rate around the globe. According to the WHO global report on diabetes, the number of adults living with diabetes has nearly doubled since 1980, with an estimated 422 million adults compared to 108 million in 1980 (NCD Risk Factor

Collaboration, 2016). In Malaysia, the fourth Malaysian National Health and Morbidity Survey (NHMS IV) reported that the prevalence of Type II diabetes has risen from 14.9% in 2006 to 20.8% in 2011 (Feisul, 2012). Previous findings have indicated the possible pharmacological actions of food-derived peptides on the carbohydrate metabolism, among the most interesting of which are inhibition of α-amylase responsible for carbohydrate digestion and absorption (Picot *et al.*, 2014), preventing the hyperglycaemia symptoms that characterize Type II diabetes. However, the underlying mechanism of the α-amylase inhibitory effect of the peptide is yet to be clarified. Hence, this study aims to shed light on the inhibitory mechanisms and structural basis that lead to antidiabetic nature of food-derived bioactive peptides.

Protein-rich food sources have been the most explored resource of bioactive peptides, especially those of animal source, such as milk, egg, and meat proteins, as well as plant protein sources, notably legume. Nevertheless, very little attention has been paid to the spice seeds as a potential precursor of bioactive peptides, which can render health benefits beyond their basis functions as flavour enhancers. This makes them attractive candidates to be explored and developed as a new source of bioactive peptides in this study. Cumin (*Cuminum cyminum*) seed is one of the popular spices with major importance worldwide due to its distinctive aroma effect (Rebey *et al.*, 2012). In addition to its common use as a flavour enhancer, cumin seed has been used as folklore therapy, particularly for the treatment of mild digestive disorders, toothache and epilepsy (Jalali–Heravi *et al.*, 2007; Hashemi *et al.*, 2009). Previous studies further proved that cumin seed could be an effective alternative to many modern medical treatments with its multitude pharmacological actions, including antimicrobial, anticarcinogenic, and

immunodulatory activities, which were mainly attributed to its non-protein components, including phenolic constituents and volatile oils (Dhandapani *et al.,* 2002; Gagandeep *et al.,* 2003; Allahghadri *et al.,* 2010; Singh *et al.,* 2017). Importantly, cumin seed was reported to have a protein level of 18%, which is substantially higher compared to other reported spice seeds, such as fennel (9.38%), black pepper (9.28%), basil (9.58%), and anise seeds (15%) (Bukhari *et al.,* 2014; Nwinuka *et al.,* 2005; Nandi & Ghosh, 2014; Nwofia *et al.,* 2013). Despite its abundance in protein, there have only been limited attempts to exploit the health and medicinal benefits of the protein and/or peptide entities of cumin seed. Therefore, the present study is an attempt to highlight the therapeutic potentials of these proteinaceous components.

The central hypothesis of this research study was that cumin seed is a rich source of antidiabetic peptides for its probable inhibitory acitivty against α-amylase. Herein, strategies were employed in an attempt to discover and unlock the biological role of peptides encrypted within the sequence of the native protein. However, the path of bioactive peptide discovery has never been a straightforward process. The time and the cost associated with the classical, conventional methods often result in a lower rate of discovery of new bioactive peptides (Sharma *et al.,* 2011). In some cases, bioactive peptides may fail to be discovered after an extensive processing, especially when bioactivity is associated with additive or synergistic effects of various components in the hydrolysates (Udenigwe, 2014). It was thereby hypothesized that the combination of Bioinformatics-aided prediction and phage display selection could provide an effective screening system to facilitate the identification of α-amylase inhibitor peptide from cumin seeds. Another hypothesis proposed that bioactive peptide may exert their inhibiotry actions by establishing

direct bindings to regions of biological interaction on potential targets, thereby affecting the activities and metabolism of the targets. This assumption is due to the fact that peptide has remarkable molecular recognition that renders its nature of intermolecular complementarity to interact with a wide variety of biomolecules through noncovalent bondings, such as hydrophilic, hydrophobic and ionic interactions (Gfeller, 2012). Therefore, the structural basis for binding of specific CSPs towards potential targets was assessed with the aims to decipher the mechanism underlying their α-amylase inhibitory actions. Overall, by addressing these gaps, the present dissertation seeks to broaden our knowledge on the possible role of CSPs in management of diabetes.

1.2 Objectives:

The main objective of this project was to explore and discover novel biologically active peptides from cumin seeds with health-promoting properties. However, there is a major focus on the antidiabetic effect of the pre-screened cumin seed peptides by assessing their ability to inhibit α-amylase from breaking down carbohydrate into simple sugar, which may be useful in reducing post-prandial blood glucose levels. The measurable objectives of this study are listed as follows:

1. To optimize the extraction of soluble protein from cumin seeds and to characterize the protein isolate.
2. To optimize the enzymatic hydrolysis process for producing α-amylase inhibitor peptides from cumin seed protein.
3. To identify and characterize the cumin seed-derived α-amylase inhibitor peptides.
4. To provide pre-clinical evidences for the efficacy and safety of the cumin seed-derived α-amylase inhibitor peptides.

1.3 Thesis outline

The bioactive peptide discovery from soluble protein of cumin seeds for potential therapeutic applications, notably potent antidiabetic agent, is presented in this thesis. The main body of this dissertation consists of a general introduction and background, literature reviews, material and methods, results and discussion, general conclusions as well as recommendations for future study.

CHAPTER ONE is a general introduction on the background of the current research topic, stating the scope, context, and significance of the research being conducted by summarizing the current situations, understanding and challenges encountered in the discovery of bioactive peptides from cumin seed. It also explains briefly the framework established for the discovery and identification of cumin seed-derived bioactive peptides with potential antidiabetic effects. In addition, the rationales and objectives of this research project are briefly discussed.

CHAPTER TWO illustrates the general literature review of the fundamental knowledge relevant to the subject matter of this research. Notably, it uncovers the historical background of the cumin seed, the current scientific knowledge on biologically active proteins and peptides, as well as the state-of-art technologies in the bioactive peptide discovery. The viewpoint of the physiological role of a bioactive peptide in tackling several major health issues, particularly diabetes mellitus, is presented in this chapter.

CHAPTER THREE details and lists down all the applied materials and methodologies for every single assay conducted throughout the whole study.

The experimental results with discussions are presented in CHAPTER FOUR. This chapter is generally divided into two major parts. The first part of this chapter focuses on the optimization of soluble protein from cumin seeds using

response surface methodology (RSM; section 4.1.1), and physicochemical and bioactivity characterization of cumin seed protein (section 4.1.2). On the other hand, the second part of this chapter starts with optimization (section 4.2.1) and fractionation (section 4.2.2) studies. In section 4.2.3, it describes the analytical strategies that involve an integrated Bioinformatics-phage display approach for screening and identifying peptide candidates with α-amylase inhibitory potentials. The finding of the structure-activity relationship between cumin seed peptides (CSP) and α-amylase are subsequently interpreted in section 4.2.4. In order to provide pre-clinical evidence, the α-amylase inhibitory efficacy of the pre-screened CSPs in AR42J cell model is explicated in the last section (4.2.5). Each sub-section describes and summarizes the results and the statistical analysis was used to evaluate the result.

The last chapter (CHAPTER FIVE) includes overall conclusions on the whole study, acknowledgement of the limitation of the current study as well as recommendations for the future direction of this research.

CHAPTER 2

LITERATURE REVIEWS

2.1 Cumin (*Cuminum cyminum*)

2.1.1 Origin and distribution of cumin

Cumin (*Cuminum cyminum*), one of the earliest known minor spices used by mankind, has a significant demand all around the globe (Thamaraikannan & Sengottuvel, 2012). Cumin has been used since ancient time, which its origins date back almost 5000 years ago. In the ancient Egyptian civilization, it has been used both as a spice and as a preservative in mummification. The cumin was then introduced as a spice to the western world from Iran, and the name of cumin originates from the word *Kerman*, a city in Iran, which was being used to cultivate cumin. The *Kerman*, locally called as *Kermun*, would have become *Kumun*, and then cumin in the European languages. Cumin has been widely cultivated in India, Egypt, Turkey, Syria, Mexico, Morocco, and Chile (Azeez, 2008). Cumin is also known by different names in different countries: Kreuzkümmel in German; Cumin de Maroc in French; Cumino in Italian; Romai Köminyi in Hungarian; Komijin in Belgium and The Netherlands; Spisskummen in Norwegian; Kimyon in Turkish; Xiāo Huí Xiāng in China; and Jintan Putih in Malaysia.

The estimated world production of cumin is approximately 300,000 tons, which is mainly produced in Central and South Asia. India represents the world's largest producer (i.e., 70% of world production), with annual production ranging from 0.1 to 0.2 million tons (Azees, 2008). India consumes almost 90% of its total production, and the remaining 10% is exported to other countries (Azeez, 2008).

Other major consumers' countries of cumin seeds are Bangladesh, China, Indonesia, Nepal, Malaysia, Singapore, United Kingdom and United States of America (Azeez, 2008). Over the years, cumin has gained a renewed global recognition due to newfound appreciation of its culinary and medicinal properties.

2.1.2 Botany and plant description of cumin

Cumin (*Cuminum cyminum*) belongs to the family of the Apiaceae with taxonomic classification shown in Table 2.1 and the morphology illustrated in Figure 2.1. Cumin is a small, slender herbaceous annual of the parsley family, which can grow to a height of 30-45 cm (Azeez, 2008). It has a main stem that branches up to five secondary branches and each branch may have 2-3 sub-branches. These branches attain the same height, rendering the plant a uniform canopy of inflorescence. The leaves are arranged pinnately or bi-pinnately with thread-like leaflets, while the flowers are in small sizes and either pink or white colored. Typical of the Apiaceae family, the flowers of cumin are borne in umbels (arrangement of flowers in a shape of an umbrella), and each umbel consists of 5 to 7 umbellets. The fruit is nonfleshy, elongated and oval shaped schizocarp of two mericarps, and each fruit contains a single seed. The seed is oblong in shape with vertical ridges on its outer surface. In addition, cumin seed has a similar appearance to a few members of Apiaceae, particularly fennel seeds (or known as jintan manis in Malaysia), however, their differences lie in their physicochemical characteristics (Singh *et al*., 2017). The seeds of cumin are brownish in color whereas the seeds of fennel are greenish in color. They also have significant different flavor and aroma profiles.

Table 2.1. Taxonomic classification of cumin (Adapted from https://en.wikipedia.org/wiki/Cumin).

Kingdom	**Plantae**
(Unranked)	Angiosperms
(Unranked)	Eudicots
(Unranked)	Asterids
Order	Apiales
Family	Apiaceae
Genus	*Cuminum*
Species	*C. cyminum*

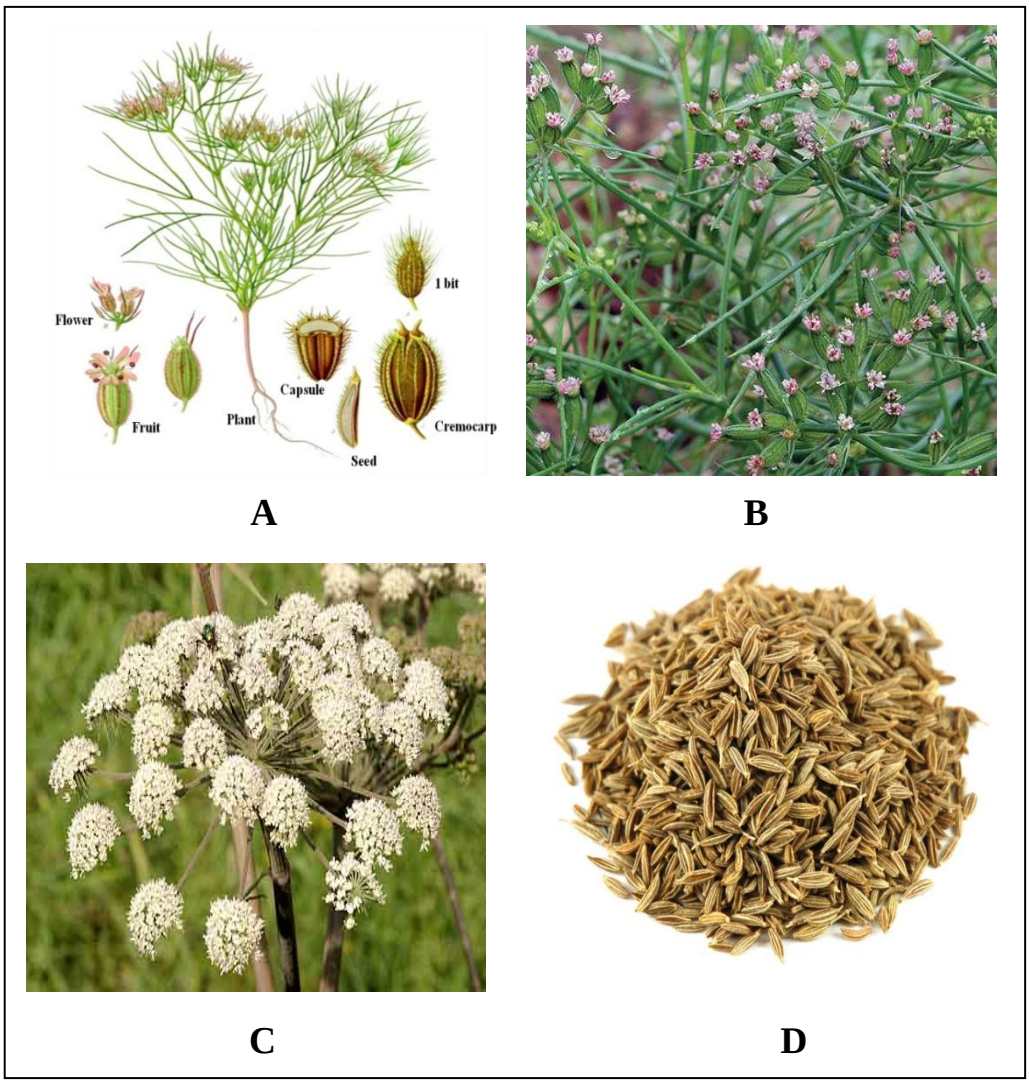

Figure 2.1 Morphology of cumin (*Cuminum cyminun*). A: whole structures of cumin (Adapted from https://en.wikipedia.org/wiki/Cumin; 16/2/2017); B: cumin plants (Adapted from http://www.johnnyseeds.com/herbs/herbs-for-salad-mix/cumin-seed-2931.html; 16/2/2017); C: cumin flowers (Adapted from http://www.floralencounters.com/Seeds/seed_detail.jsp?productid=92650; 16/2/2017); D: cumin seeds (Adapted from http://www.savoryspiceshop.com/spices/cumin-seeds; 16/2/2017)

2.1.3 Compositional profile of cumin seeds

The proximate composition of the cumin seeds, which has been reported in a study of Al-Snafi (2016), is demonstrated in Table 2.2. The major constituents of the seeds are carbohydrate (44%), followed by fat (22%), protein (18%), dietary fibre (11%), moisture (8%) and ash (8%). Surprisingly, the protein content of the cumin seeds is approximately double than that of fennel seeds (9.38%) despite them having similar morphological features (Bukhari *et al.,* 2014). Likewise, the protein contents of other common spices, including anise seeds (15%), basil seeds (9.58%), ashanti pepper seeds (12.5%), black pepper seeds (9.28%) and hot pepper seeds (4.94%) are also relatively lower than that of cumin seeds (Nwinuka *et al.,* 2005; Nandi & Ghosh, 2014; Nwinuka, 2005; Nwofia *et al.,* 2013; Zou *et al.,* 2015). Therefore, cumin seed protein is suggested as a good source of precursor that might have extraordinary advantages to the discovery of novel and promising bioactive peptides in this study. It is also important to note that the cumin seeds have a high level of total fat content. Appropriate defatting techniques are thereby indispensable to carry out in order to avoid possible interference in the subsequent biochemical analyses.

Table 2.2. Proximate composition of cumin seeds (Adapted from Al-Snafi, 2016).

Types of nutrients	**Nutrient value(g)***	**Percentage (%)**
Protein	0.36	18
Carbohydrate	0.88	44
Dietary fibre	0.22	11
Total fat	0.44	22
Saturated fat	0.04	2
Monounsaturated fat	0.28	14
Polyunsaturated fat	0.06	3
Moisture	0.16	8
Ash	0.16	8

* means the nutrient contents of cumin seeds in 2 g of seeds.

2.1.4 Uses of cumin seeds

2.1.4(a) Culinary uses

Due to its distinctive aroma and warmly bitter flavour, cumin seed can be proudly claimed as the most popular spice in the world after black pepper (Thamaraikannan & Sengottuvel, 2012). In fact, the warm, spicy, heavy and curry-like flavours of the seed are credited to its flavour component, cuminaldehyde (Azeez, 2008). High content of volatile oil provides the seed with a peppery astringent taste with slightly citrus overtones. The cumin seed is undoubtedly an essential flavouring for many food preparations, which is commercially sold as seeds or powders. To have a better aroma, the whole seeds are often fried in oil or dry-roasted prior to being added into dishes. Ground form of cumin seeds is one of the main ingredients of curry powder, which are used to add pungent flavours to the dishes and cuisines. In Malaysia, a fragrant combination of coriander and cumin seeds is fundamental to many cultural cuisines, especially serves as a basis of many Malay curries. As of today, it seems to become an integral part of every culture around the globe.

2.1.4(b) Traditional uses

The seed is an excellent source of dietary fibre and minerals such as iron, copper, calcium, potassium, manganese, selenium, zinc and magnesium (Thamaraikannan & Sengottuvel, 2012). It is also high in vitamins A, C, E and B-complex, such as thiamin, pyridoxal, niacin, and riboflavin. In addition to essential nutrients, cumin seed comprises numerous bioactive compounds, such as phytochemicals and phenolics, which are known to exhibit antioxidant, anticancer, antibacterial and anti-inflammatory effects (Thamaraikannan & Sengottuvel, 2012).

Therefore, it has been widely used in traditional medicine, especially for the treatment of hyperlipidemia, cancer and diabetes (Mnif & Aifa, 2015). It has also been long utilized as an aid for indigestion, jaundice, diarrhoea, flatulence and other inflammatory diseases. The seeds were also documented as stimulant, carminative and astringent (Zargari, 2001). The medicinal usage of cumin seeds has been widespread in different ethnomedical systems from Northern Europe to the Mediterranean regions, Russia, Iran, Indonesia, and North America, indicating its popularity to be used as folk medicines for their health care needs (Johri, 2011).

2.1.4(c) Modern therapeutic uses

The literature indicates that cumin has moved from the kitchen shelf into the clinic. According to the study of Derakhshan *et al.* (2008), the essential oil of cumin seeds possessed a significant inhibition against the growth of *Klebsiella pneumonia* by improving cell morphology, suppressing capsule expression and decreasing urease activity. The following finding of these authors concluded the usefulness of cumin seed essential oils as potent antimicrobial agents through their inhibitory actions against biofilm formation of *Klebsiella pneumoniae* (Derakhshan *et al.,* 2010). Cuminaldehyde was suggested as the main active constituent for this property (Derakhshan *et al.*, 2010). The antioxidative activity of cumin seed oil has been attributed to the presence of monoterpene alcohols, linalool, carvacrol, anethole, estragol, flavonoids, and other polyphenolic compounds (Bettaieb *et al.,* 2010). These antioxidant compounds have also been highlighted for their anticarcinogenic potentials (Aruna & Sivaramakrishnan, 1996). With a rich content of α-pinene (29.2%), cumin seed oil has been noted to exert remarkable inhibitory effects on the fungal growth of *Aspergillus* strains (Mohammadpour *et al.,* 2012). Cumin seed has

also been acknowledged for its antiosteoporotic effect by reducing urinary calcium excretion while, at the same time, increasing calcium content and mechanical strength of bones in animal models (Shirke *et al.,* 2008). The presence of phytoestrogens is responsible for this bioactivity. It has been reported that essential oil in cumin seed can act as a potent analgesic agent by ameliorating the morphine tolerance and dependence by formalin test in rats (Haghparast *et al.,* 2008). Linolool and β-D-glucopyranoside derivatives in the seed oil were suggested as the main contributors for these activities (Haghparast *et al.,* 2008). In addition, cumin seed has been examined for a possible wound healing property due to its ability to show improved epithelization in open, restructured incision, and granuloma wound models (Patil *et al.,* 2009). The wound healing property was closely related to the presence of terpenoids that predominantly found in alcoholic extract and petroleum-ether fraction of cumin seed (Patil *et al.,* 2009). In a study of Mekaway *et al.,* (2009), antitumor compound, 1-(2-ethyl, 6-heptyl) phenol extracted from cumin seeds had shown a significant inhibitory activity against six types of tumor cell lines (CACO2, HEPG2, HELA, HEP2, HCT116, and MCF7). A study by Lee (2005) demonstrated *in vivo* inhibitory activity of cumin seed oil against aldose, reductase and alpha-glucosidase isolated from rats, indicating its potential use in the treatment of diabetes. All these aforementioned medicinal values of cumin seeds were summarized in the Table 2.3.

Table 2.3 Documented medicinal benefits of cumin seeds.

Health benefits	Active constituent	Activity	Reference
Antimicrobial	Essential oil	Growth inhibition of *K. pneumoniae* by causing cell elongation, repression of capsule expression and inhibition of urease activity.	Derakhshan *et al.*, 2008
	Cuminaldehyde	Inhibition of biofilm formation.	Derakhshan *et al.*, 2010
Antifungal	α-Pinene	Inhibition of the fungal growth of *Aspergillus* strains.	Mohammad-pour *et al.*, 2012
Antioxidant	Monoterpene alcohols, linalool, carvacrol, anethole, estragol, flavonoids, polyphenolic compounds	Scavenging effects of hydroxyl radicals, 1,1-diphenyl-2-picrylhydrazyl (DPPH) radicals and lipid peroxides, reducing and chelating power assays.	Bettaieb *et al.*, 2010
Anti-osteoporotic	Phytoestrogens	Decreased urinary calcium excretion, increased calcium content and mechanical strength of bones in animals.	Shirke *et al.*, 2008
Analgesic	Linolool, β-D-Glucopyranoside derivaties	Ameliorated morphine tolerance and dependence by formalin test in rats.	Hagpparast *et al.*, 2008
Wound healing	Terpenoids	Improved epithelization in open and restructured incision and granuloma wound models	Patil *et al.*, 2009
Antitumor	2-ethyl-6-heptlphenol	Growth inhibition of tumor cell lines, including CACO2, HEPG2, HELA, HEP2, HCT116, and MCF7	Mekawey *et al.*, 2009
Anti-carcinogenic	Monoterpene alcohols, linalool, carvacrol, anethole, estragol, flavonoids, polyphenolic compounds	Decreased glutathione-S-transferase by more than 78% in stomach, liver, and oesophagus of mice	Aruna & Sivaramakrish-nan, 1996
Antidiabetic	Cuminaldehyde	Inhibited aldose reductase and α-glucosidase isolated from rat	Lee, 2005

In conclusion, cumin is an aromatic plant that widely used for a variety of culinary and pharmacological purposes. The seeds of cumin have been long used in traditional medicine for the treatment of several health disorders and diseases. However, being a spice, cumin seed is typically consumed in low quantities which may not be sufficient to render the medicinal and health benefits. The increasing awareness and demand for functional foods have encouraged the consumption of cumin in different forms, particularly, as active ingredients in medicine or dietary supplements. As evidenced by the past and recent findings, these medicinal properties and health benefits of cumin seeds have been generally accredited to their rich content and potent action of phytochemicals and essential oils. To the best of the author's knowledge, little data are currently available on the impacts of protein components in functional perspectives. Therefore, it would be a great deal of interest to explore the functional benefits of protein and peptide entities derived from cumin seeds.

2.2 Protein as precursor of bioactive peptides

2.2.1 Fundamental of protein function and structure

The term of a protein is defined as molecules with more than 50 amino acids, whereas in the presence of less than 50 amino acids are generally called as oligopeptides or simply peptides. This polymer is linked by the α-carboxyl group of one amino acid to the α-amino group of another amino acid with an amide bond (or known as peptide bond), followed by a subsequent loss of a water molecule. A series of amino acids linked by peptide bonds, construct a polypeptide chain, and each amino acid unit in a polypeptide is termed as a residue. Each residue consists of a central carbon atom (C_α), which is adjacent to the carboxyl group, is bonded to four

varying chemical groups: an amino (NH_2) group, a carboxyl (COOH) group, a hydrogen (H) atom, and one variable group (side chain or R group) (Fig 2.2).

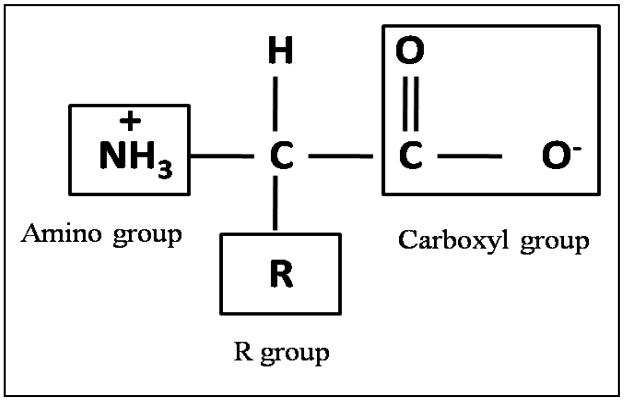

Figure 2.2 General structures of the amino acid.

The aforementioned features of a protein represent the general structures adopted by the 20 different amino acids, however, their side-chain groups may vary in terms of size, shape, charge, hydrophobicity and reactivity. In general, the amino acids are typically grouped into three main categories according to the nature of their R groups: hydrophobic amino acids, with an aliphatic or bulky and aromatic R group; hydrophilic amino acids, with a charged or polar and uncharged R group; amino acids with a special group, composing of cysteine, glycine, and proline (Fig. 2.3). Such variation in the properties of different R groups will give amino acid individual characteristics, and a sequence made up of different amino acids will then render each protein unique characteristics. In addition, the amino group represents an excellent hydrogen-bond donor, whereas the carboxyl group is a good hydrogen-bond acceptor. These groups interact with each other, and with the functional groups from side chains are critically important to stabilize the protein or peptide structure.

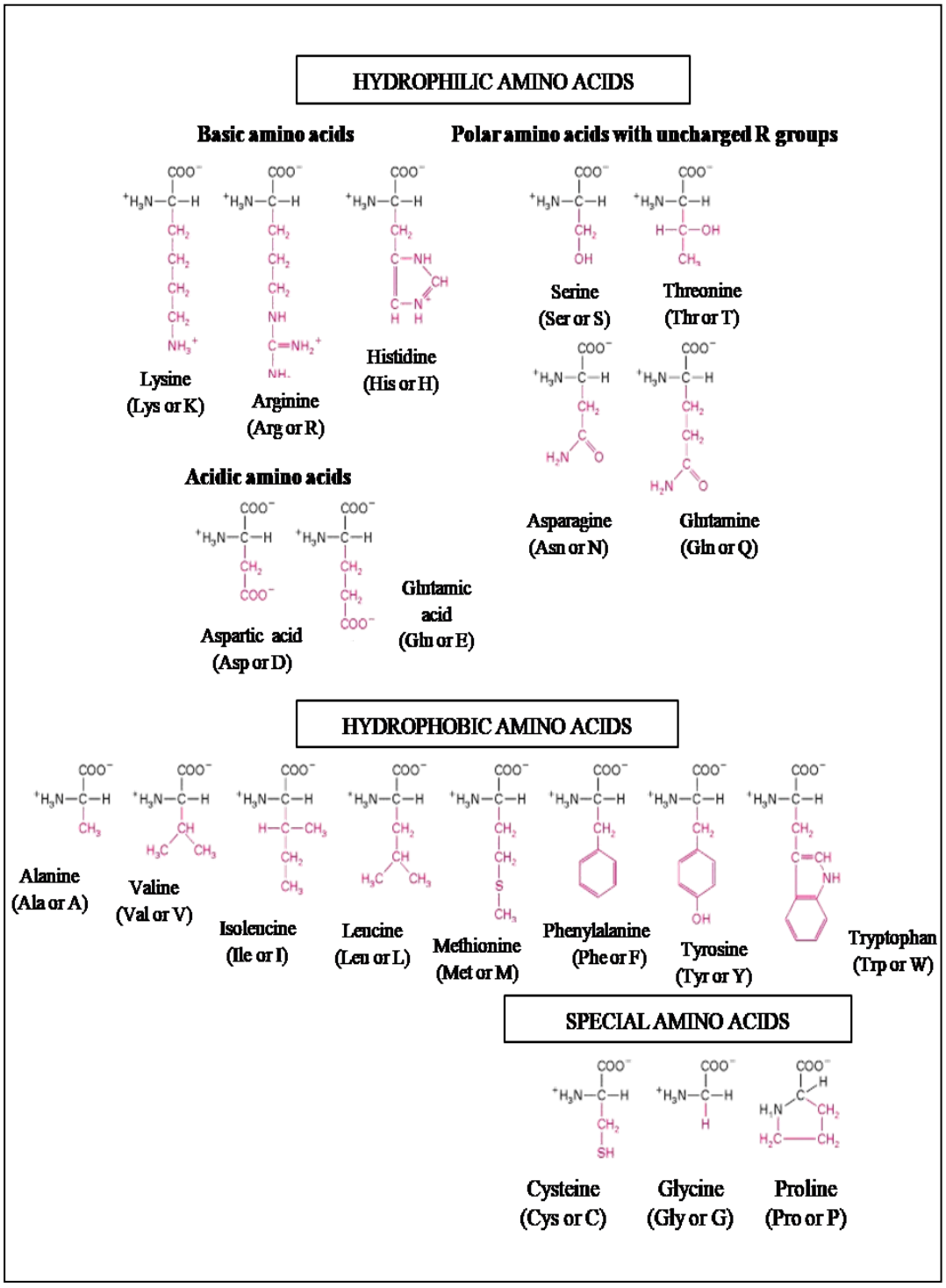

Figure 2.3 The structures of the 20 standard amino acids grouped into three categories: hydrophilic, hydrophobic and special amino acids. All amino acid structures are presented in the zwitterions form, which exist at the pH of the cytosol. In parentheses are the three-letter and one-letter abbreviations for each amino acid (Adapted from Lodish *et al.*, 2000).

2.2.2 Types of protein extraction

Extraction is a separation process widely used for the recovery of active constituents or for the removal of undesirable constituents from raw materials, while components of interest may be separated from natural products with many applications of food industries. The ideal extraction method for plant proteins is highly challenging due to the metabolic and structural characteristics of plant tissues. The conventional routes to produce plant-derived protein flours, concentrates, and isolates involve dry or wet extraction processes. Each of these approaches has their pros and cons. It is therefore of paramount importance in choosing suitable techniques and conditions for protein extraction.

Milling and air classification are categorized into the dry extraction process. Air classification is a separation process which utilized for fractionation of powder particles (i.e., cereals and grain legumes) into light fine fraction (protein) and heavy coarse fraction (starch) according to their specific sizes and densities which result from the milling step (Boye *et al*., 2010). The dry extraction process is commonly applied to the high-starch containing crops. Advantages of this technique include: (1) the water usage is relatively low or no water required; (2) does not require a stabilization process after separation (i.e., no or less heating/drying); (3) more emphasis on functionality instead of molecular purity; and (4) easily scalable (Schutyser & van der Goot, 2011). However, a major drawback of dry extraction is that the yields, concentrations and purities of protein fractions are often lower in relative to wet extraction (Gao *et al.*, 2009).

Classical wet extraction processes involve alkaline extraction/isoelectric precipitation, acid extraction and salt solubilization. The basis of such methods lies in the application of diverse solubility and precipitation profiles of proteins (Jayasena

et al., 2011). Higher solubility of proteins is typically observed at the alkaline and acid pH range while they have the lowest solubility at their respective isoelectric points. In the case of seed proteins, they often show a minimum solubility at pH values ranging from four to five (Klupšaitė & Juodeikienė, 2015). As shown in Fig. 2.4, the procedures of wet extraction involve: (1) preparing a flour suspension, (2) isolating protein using alkaline or acidic conditions, (3) separating insoluble by isoelectric precipitation or ultrafiltration, and (4) spray-drying or freeze-drying (Schutyser & van der Goot, 2011). This method has the advantage of obtaining relatively pure (>90%) isolates. However, its major disadvantages include the usage of huge amounts of water and chemicals, especially for neutralization and acidification purposes (Schutyser & van der Goot, 2011).

In particular, alkaline extraction followed by isoelectric precipitation has been reported to produce a protein content of 90-96% and a protein recovery yield in a range of 60-65% (Klupšaitė & Juodeikienė, 2015). Typically, the raw material is subjected to an alkaline pH level between pH 8 and 11 where the protein is found to be highly soluble (Jayasena *et al.,* 2011). The pH of the soluble protein fraction is subsequently adjusted to isoelectric points of the tested protein for precipitation (Jayasena *et al.,* 2011). A similar principle is established by acidic extraction to that of alkaline extraction with the exception that the initial protein extraction is performed under acidic conditions. The salt extraction process, or known as micellization, is primarily relied on the salting-in and salting-out phenomenon of food proteins (Boye *et al.,* 2010). During this process, the protein is extracted using an appropriate salt solution at desired ionic strength, followed by dilution and centrifugation to recover from the protein precipitation, and lastly subjected to drying (Boye *et al.,* 2010).

In this research project, the wet extraction technique was selected over the dry extraction in anticipating to produce cumin seed protein isolates with high yield and purity. Among the wet extraction techniques, a vast number of research findings have shown that protein isolates resulted from alkaline extraction possesses better nutritional profiles and higher essential amino acid content compared with acidic and salt extractions (Chen *et al.*, 2009). Therefore, alkaline extraction followed by acidic precipitation is an ideal approach for isolating proteins from cumin seed.

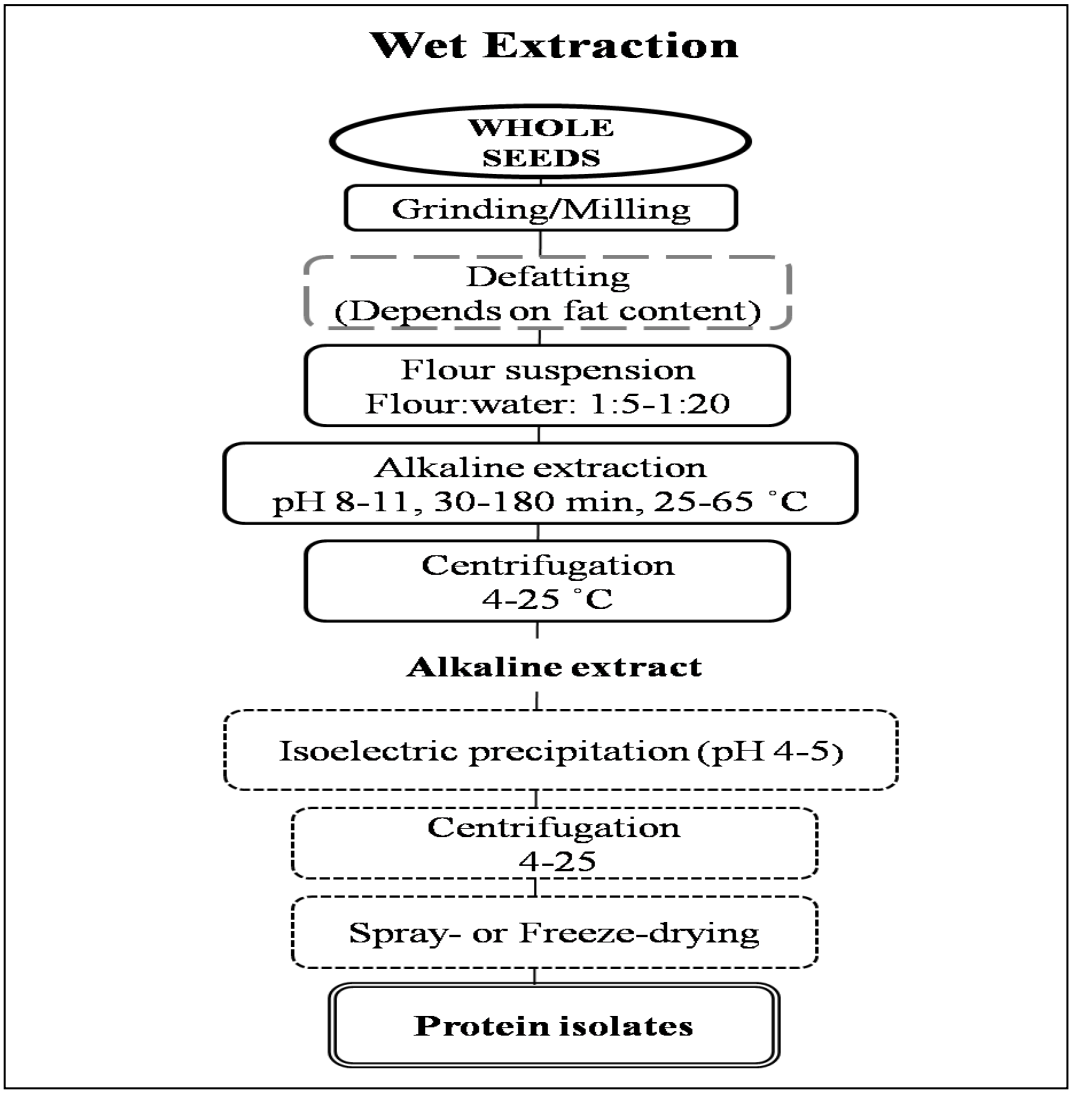

Figure 2.4 Schematic illustration of wet extraction process (Modified from Schutyser & van der Goot, 2011).

2.2.3 Critical parameters affecting the protein extractability

The extraction of food-derived proteins is a complex phenomenon that is influenced by not only the parameters of the extraction process itself but also by the diversity of the proteins present and the intricately organized structure of the tissues (Eady *et al.*, 2014). Usually only about one-half of the available protein could be successfully extracted and the rest may largely lose during the preparation process. It is thereby posing a need to improve the extraction yield of protein production from natural sources. In laboratory settings, extraction parameters, such as buffer pH, ionic strength, type of buffers, buffer-to-sample ratio, extraction time and temperature, have been extensively studied for improving or maximizing the yield of protein production (Munasinghe & Sakai, 2004; Thomas *et al.*, 2007; Eady *et al.*, 2014).

The pH of a solution influences protein solubility mainly in a way in which it modifies the charge distribution over the protein surface. Note that a protein attains a net positive charged at pH values below its isoelectric point (pI), whereas it will be negatively charged at pH values above its pI. Several studies reported that the concentration and protein profiles of salt-soluble and water-soluble protein extracts are greatly influenced by the extraction pH (Xiong & Blanchard, 1994; Lesiow & Xiong; 2003; Eady *et al.*, 2014). In many cases, the extractability of protein was influenced by extraction at slightly acidic to neutral pH levels (pH 5–7), while there is a minimal impact on both the total amount and composition of protein that could be extracted using buffer at higher pH levels (7-9). According to the study of Eady *et al.* (2014), the influence of pH on extracted protein concentration indicates that the buffer pH altered the surface properties of proteins and the protein to protein interactions sufficiently to affect overall extractability. As the pH of the extraction buffer approached the pI of the target protein, the overall number of reactive groups

on the protein for binding water diminished. The diminished capacity for the proteins to bind water likely resulted in the reduced protein solubility and extractability observed at its pI. Since most biochemical experiments take place at or near neutral pH between 6 and 8, ideal buffer should have pK_a values (i.e., dissociation constant for the weak acid component of the buffer) in this region to provide maximum buffering capacity and maintain the protein stability (Will *et al.*, 2011).

Although pH control is inevitable for the protein extraction, ionic strength that affects the partitioning behaviour of proteins should not be ignored (Chang *et al.*, 2001). It is well-documented that protein solubility is increased at low ionic strength condition, which is usually attributed to the colloidal stability of a protein (Shire, 2015). At any given pH, a protein has a specific net charge. Therefore, the protein molecules repel each other and predominate over any attractive interactions. As the ionic strength is increased the net charge repulsion is decreased, resulting in potential attractive protein–protein interactions, and hence, leading to a decrease in solubility and increase in protein aggregation (Shire, 2015). In addition, a higher ionic strength may aid in the adsorption of proteins with hydrophobic surfaces, making the extraction process more difficult (Mitchell *et al.*, 2002). To improve protein solubility and reduce its aggregation tendency, additives such as salt and chaotropic agent, are added to maintain the ionic strength of the extraction solution. It was also found that phosphate buffer could be a good choice to maintain ionic strength of a solution due to the stabilization effect provided by the binding of phosphate ion to phosphate binding sites in protein structures (Ugwu & Apte, 2004). In addition, increasing in phosphate concentration from 50 mM to 0.1 M caused a 3- and 10-fold increase in the thermal stability of a target protein (pI~4.4) at pH 6 and 8, respectively (Ugwu & Apte, 2004). Results suggested that electrostatic interactions

are stabilizing in the native protein, and these interactions are reduced at a low ionic strength. Therefore, buffer molarity was usually set at 0.1-0.5 M, as a lower molarity value reduced the effectiveness of the buffer, affecting the pH of the final crude extract and thereby the stability of the protein (Gutarra *et al.*, 2009).

During extraction, it is important to provide proteins with a well-buffered environment to cope up with sudden changes in pH. Without a well-buffered solution, the proteins can be easily and irreversibly denatured once they are released from the cells. Therefore, the choice of extraction buffer is pivotal to maintain stability and solubility of proteins from both a pH and ionic-strength standpoint (Ugwu & Apte, 2004). A good protein extraction buffer possess a wide range of buffering capacity with constant ionic strength, have high water solubility and chemical stability, as well as compatible to downstream applications of proteins (Ugwu & Apte, 2004). Sodium phosphate buffer is one of the commonly used broad-range buffers that have pKa (6.86) near to physiological pH, making it compatible for many biological applications (Kondratiuk *et al.*, 2015). Several studies have also provided evidences that a higher yield of proteins can be extracted using sodium phosphate buffer, as compared to other buffer systems, due to its capability to reduce the influence of phenolics and derivatives in the extract, as well as to prevent the formation of complexes involving proteins (Nahar *et al.*, 2013; Eady *et al.*, 2014; Kondratiuk *et al.*, 2015). In addition, the toxic effects of a buffer system are generally lacking, which make it ideal as a working solvent in various biological works (Kondratiuk *et al.*, 2015). However, it has been reported that phosphate buffer may pose a particular difficulty in extracting out certain protein bodies that bound strongly to other biomolecules, such as carbohydrates, polyphenols, lipids and starch granules. These cross-linked structures with numerous disulfide and hydrogen bonds, as well as