IDENTIFICATION OF PROTEIN FROM *EURYCOMA LONGIFOLIA* ROOT EXTRACT

NURULAINI BT ABD. RAHMAN

UNIVERSITI TEKNOLOGI MALAYSIA
IDENTIFICATION OF PROTEIN FROM *EURYCOMA LONGIFOLIA* ROOT EXTRACT

NURULAINI BT ABD. RAHMAN

A thesis submitted in fulfillment of the requirements for the award of the degree of Master of Engineering (Bioprocess)

Faculty of Chemical Engineering
Universiti Teknologi Malaysia

AUGUST 2014
Specially dedicated to my beloved umi, abah, siblings, supervisor and caring friend

Thank you for your support

“Jika mahu cemerlang, mesti sanggup menempuh semua halangan. Halangan terbesar ialah MINDA kita sendiri, yang hanya sukakan sesuatu yang selesa dan mudah.”
ACKNOWLEDGEMENT

By the Name of Allah, the Most Merciful, the Most Beneficent

First and foremost, I am thankful to Allah s.w.t, as I have completed this study fulfilling my master degree successfully. In preparing this thesis, I was helped and assisted by many inspiring people. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation and gratitude to my supervisor, Dr. Chua Lee Suan and Professor Dr Mohamad Roji bin Sarmidi for their encouragement, patient guidance and criticism throughout this project. It is my honour to work under their supervision and all of their efforts are greatly appreciated.

Acknowledgment and appreciation also goes to all my laboratory friends and staff at Institute of Bioproduct Development in Universiti Teknologi Malaysia for their help and guidance which enabled me to complete my work.

Appreciation also especially goes to my beloved ‘umi’, ‘abah’ and my siblings for their precious love, time and countless support that was an extraordinary help towards the success of this project.

Last but not least, I would like to specially thank my beloved friends especially Nurul ‘Amirah who always besides me to guide and support me. Million thanks to all.
ABSTRACT

This study was carried out to identify plant protein from the root samples of *Eurycoma longifolia* harvested from Pahang and Perak, Malaysia. Protein usually presents in small quantity which is only 0.001% in plant, therefore it is critical to determine the extraction method for high yield and good quality of protein for the subsequent process of protein identification spectrometrically. Four extraction methods, namely water, Triton X-100 (non-ionic detergent), phenol-SDS precipitation and TCA-acetone precipitation were investigated for plant protein extraction. The yield of protein extracted from the plant samples was determined using Bradford assay. Both water extracts (water and Triton X-100) methods contained slightly higher protein content (0.2-0.53 mg protein/ml crude protein) than the extracts of precipitation (phenol and TCA) methods (0.12-0.29 mg protein/ml crude protein). The water extraction method also produced the highest resolution of 15% polyacrylamide gel with six and five protein bands for *E. longifolia* Pahang and Perak, respectively. However, the number of protein bands decreased from five to three for the extraction method of Triton X-100, phenol-SDS and TCA-acetone, respectively. After trypsin digestion of the protein bands, the presence of protein was analyzed by liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). The mass spectra matched with the databases showed that the addition of Triton X-100 could assist the extraction of mitochondrial protein, namely superoxide dismutase. Proteins that involved in energy metabolism such as phosphoenolpyruvate carboxykinase and plant protein inhibitor were also detected in the phenol-SDS buffer extraction because SDS acted as ionic detergent for cell lysis in this method. In line with previous studies, TCA-acetone did not exhibit clear gel image and subsequently less peptides were detected spectrometrically. In the present study, the addition of detergent (Triton X-100 and SDS) could enhance plant protein extraction from *E. longifolia* root from 46.7 to 72.7% w/w, but the use of TCA did not improve protein precipitation effectively.
Kajian ini dijalankan untuk mengenal pasti protein tumbuhan daripada sampel akar *Eurycoma longifolia* yang dituai dari Pahang dan Perak, Malaysia. Protein biasanya dalam kuantiti yang sedikit dalam tumbuhan, maka adalah penting untuk menentukan kaedah pengekstrakan bagi mendapatkan kandungan protein yang tinggi dan berkualiti supaya proses pengenalpastian protein dapat dilakukan secara spektrometrik. Empat kaedah pengekstrakan, iaitu air, Triton X-100 (detergen bukan ionik), pemendakan phenol-SDS dan pemendakan TCA-aseton telah dikaji untuk mengekstrak protein tumbuhan. Hasil protein daripada sampel tumbuhan telah ditentukan dengan menggunakan ujian Bradford. Hasilnya didapati pengekstrakan menggunakan air dapat mengekstrak hasil protein yang lebih tinggi (0.2-0.53 mg protein/ml protein mentah) daripada kedua-dua kaedah pemendakan (0.12-0.29 mg protein/ml protein mentah). Kaedah pengekstrakan air juga menghasilkan 15% poliakrilamida gel beresolusi tinggi dengan 6 dan 5 jalur protein bagi *E. longifolia* Pahang dan Perak masing-masing. Walau bagaimanapun, bilangan jalur protein menurun dari 5 hingga 3 jalur bagi kaedah pengekstrakan Triton X-100, phenol-SDS dan TCA-aseton masing-masing. Selepas dicerna oleh trypsin, kehadiran protein dianalisis dengan menggunakan ujian kromatografi cecair diintegrasikan dengan jisim spektrometri sejajar (LC-MS/MS). Spektra jisim dipadankan dengan pangkalan data dan didapati bahawa penambahan Triton X-100 membantu pengekstrakan protein mitokondria iaitu superosida dismutase. Protein terlibat dalam metabolisma tenaga seperti fosfenolpiruvate karboxykinase dan pemangkin protein dikesan dalam kaedah pemendakan phenol-SDS kerana SDS bertindak sebagai detergen ionik yang boleh memecahkan sel tumbuhan. Sebagaimana kajian sebelumnya, TCA-aseton tidak menunjukkan imej gel jelas dan kurang peptida dapat dikesan oleh spektrometer. Dalam kajian ini, penambahan detergen (Triton X-100 dan SDS) boleh meningkatkan proses pengekstrakan protein daripada tumbuhan sebanyak 46.7 hingga 72.7% w/w, tetapi penggunaan TCA tidak dapat memberikan hasil pemendakan protein secara berkesan.
# TABLE OF CONTENT

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DECLARATION</td>
<td>ii</td>
</tr>
<tr>
<td></td>
<td>DEDICATION</td>
<td>iii</td>
</tr>
<tr>
<td></td>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td></td>
<td>ABSTRAK</td>
<td>vi</td>
</tr>
<tr>
<td></td>
<td>TABLE OF CONTENT</td>
<td>vii</td>
</tr>
<tr>
<td></td>
<td>LIST OF TABLES</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td></td>
<td>LIST OF ABBREVIATIONS</td>
<td>xii</td>
</tr>
<tr>
<td></td>
<td>LIST OF APPENDICES</td>
<td>xiv</td>
</tr>
<tr>
<td>1</td>
<td>RESEARCH BACKGROUND</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Problem Statement</td>
<td>2</td>
</tr>
<tr>
<td>1.3</td>
<td>Objective</td>
<td>4</td>
</tr>
<tr>
<td>1.4</td>
<td>Scope</td>
<td>4</td>
</tr>
</tbody>
</table>
## LITERATURE REVIEW

2.1 Introduction to *Eurycoma longifolia*  
2.2 Chemical Composition of *Eurycoma longifolia*  
2.3 Pharmacological Properties of *Eurycoma longifolia*  
2.4 Protein Application of Plant Protein  
2.5 Protein Extraction from Plant Samples  
2.6 Plant Extraction from *Eurycoma longifolia*  
2.7 Protein Extraction by Detergent Lysis  
2.7.1 Common Detergent for Extraction  
2.7.1.1 Triton X-100  
2.7.1.2 Sodium Dodecyl Sulfate  
2.8 Protein Extraction by Precipitation  
2.9 Measurement of Protein Concentration  
2.10 Polyacrylamide Gel-based Electrophoresis  
2.11 Protein Identification from *Eurycoma longifolia*  
2.12 Protein Database Search

## MATERIALS AND METHODS

3.1 Introduction to the Overall Research Methodology  
3.2 Chemicals and reagents  
3.3 Plant Protein Extraction  
3.3.1 Water Extraction Method  
3.3.2 Detergent Lysis using Triton X-100  
3.3.3 Phenol-SDS Buffer Precipitation  
3.3.4 TCA-Acetone Precipitation  
3.4 Bradford Assay for Protein Quantification  
3.5 Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)  
3.6 Staining and Destaining Solution Preparation  
3.7 In-Gel Digestion of Plant Protein  
3.8 Identification of Plant Peptide by Mass Spectrometry  
3.9 Mass Spectra Databases Search
4 RESULTS AND DISCUSSION
  4.1 Introduction to the Overall Study 53
  4.2 Protein Quantification by Bradford Assay 54
  4.3 Protein Separation by Electrophoresis 56
  4.4 Protein Identification by LC-MS/MS 62

5 CONCLUSION AND RECOMMENDATION 72
  5.1 Conclusion 72
  5.2 Recommendation 74

REFERENCES 75

Appendices A - G 89-104


LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Major compounds identified in <em>E. longifolia</em> extracts using Solid Phase Microextraction (SPME) coupled with GC–MS</td>
<td>8</td>
</tr>
<tr>
<td>2.2</td>
<td>The reported pharmacological properties of <em>E. longifolia</em></td>
<td>12</td>
</tr>
<tr>
<td>2.3</td>
<td>Three classes of plant peptides</td>
<td>15</td>
</tr>
<tr>
<td>2.4</td>
<td>Protein extraction of different plant roots for protein analysis</td>
<td>28</td>
</tr>
<tr>
<td>2.5</td>
<td>Methods for protein measurement</td>
<td>31</td>
</tr>
<tr>
<td>4.1</td>
<td>Protein concentration of <em>E. longifolia</em> protein via different extraction methods</td>
<td>55</td>
</tr>
<tr>
<td>4.2</td>
<td>Molecular weight of <em>E. longifolia</em> proteins extracted from different extraction methods using SDS-PAGE</td>
<td>60</td>
</tr>
<tr>
<td>4.3</td>
<td>Peptides identified from water, Triton X-100, phenol-SDS buffer and TCA-acetone buffer methods</td>
<td>62</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td><em>E. longifolia</em> roots and its common phytochemicals</td>
<td>7</td>
</tr>
<tr>
<td>2.2</td>
<td>Chemical structure of Triton X-100</td>
<td>22</td>
</tr>
<tr>
<td>2.3</td>
<td>Chemical structure of Sodium dodecyl sulfate (SDS)</td>
<td>25</td>
</tr>
<tr>
<td>2.4</td>
<td>Schematic diagram of polyacrylamide gel electrophoresis</td>
<td>34</td>
</tr>
<tr>
<td>2.5</td>
<td>Chemical formation of SDS-PAGE</td>
<td>35</td>
</tr>
<tr>
<td>3.1</td>
<td>Experimental flow chart of this study</td>
<td>41</td>
</tr>
<tr>
<td>3.2</td>
<td>Reduction and alkylation of disulfide bonds</td>
<td>49</td>
</tr>
<tr>
<td>3.3</td>
<td>In-gel digestion of plant protein</td>
<td>50</td>
</tr>
<tr>
<td>3.4</td>
<td>Flow of the plant protein identification</td>
<td>52</td>
</tr>
<tr>
<td>4.1</td>
<td>SDS-PAGE based protein separation for <em>E. longifolia</em> Pahang and Perak samples from four types of extraction methods</td>
<td>57</td>
</tr>
<tr>
<td>4.2</td>
<td>Spectra of (a) phosphoenolpyruvate carboxykinase [GTP], (b) sec7 domain containing protein and (c) anthranilate phosphoribosyltransferase-like protein</td>
<td>66</td>
</tr>
<tr>
<td>4.3</td>
<td>Comparison of total protein number identified from different extraction methods</td>
<td>67</td>
</tr>
</tbody>
</table>
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>µl</td>
<td>Microliter</td>
</tr>
<tr>
<td>2DE</td>
<td>Two-Dimensional Electrophoresis</td>
</tr>
<tr>
<td>ACN</td>
<td>acetonitrile</td>
</tr>
<tr>
<td>AIDS</td>
<td>acquired immunodeficiency syndrome</td>
</tr>
<tr>
<td>APS</td>
<td>ammonium persulfate</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>BCA</td>
<td>Bicinchoninic acid assay</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>BPB</td>
<td>bromophenol blue</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>CBB</td>
<td>Coomassie Brilliant Blue</td>
</tr>
<tr>
<td>cm</td>
<td>centimeter</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>ELISA</td>
<td>enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>EPF</td>
<td>Epidermal Patterning Factor</td>
</tr>
<tr>
<td>FDA</td>
<td>US Food and Drug Administration</td>
</tr>
<tr>
<td>FRIM</td>
<td>Forest Research Institute Malaysia</td>
</tr>
<tr>
<td>FTIR</td>
<td>Fourier transform infrared spectroscopy</td>
</tr>
<tr>
<td>GC–MS</td>
<td>Gas chromatography–mass spectrometry</td>
</tr>
<tr>
<td>HCl</td>
<td>hydrogen chloride</td>
</tr>
</tbody>
</table>
HIV - human immunodeficiency virus
HPLC - High-Performance Liquid Chromatography
IAA - iodoacetamide
KCl - potassium chloride
kD - kilo Dalton
LC-MS/MS - Liquid Chromatography Coupled with Tandem Mass
MALDI-TOF - Matrix-assisted laser desorption/ionization - time-of-flight
mg - Miligram
ml - Mililiter
mM - Mili Molar
Mn - Manganese
MS - Mass Spectrometry
NCBI - National Center for Biotechnology Information
nm - nanometer
PCA - Principal component analysis
PMF - Peptide Mass Fingerprinting
pmol - picomoles
PMSF - phenylmethylsulfonyl fluoride
RALF - Rapid Alkalization Factor; 49 amino acid residues
RIA - Radioimmunoassay
rpm - round per minute
SDS-PAGE - Sulphate Polyacrylamide Gel Electrophoresis
SELDI-MS - Surface Enhance Laser Desorption Ionization Mass Spectrometer
SPME - Solid Phase Microextraction
SRM - selected reaction monitoring
TCA - Trichloroacetic acid
TDIF - Tracheary Element Differentiation Inhibitory Factor
TEMED - tetramethylethylenediamine
TMV - Tobacco Mosaic Virus
TPD1 - Tapetum Determinant 1
UV-Vis - Ultraviolet-visible
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>List of Chemicals</td>
<td>89</td>
</tr>
<tr>
<td>B</td>
<td>Bradford Assay</td>
<td>91</td>
</tr>
<tr>
<td>C</td>
<td>Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)</td>
<td>95</td>
</tr>
<tr>
<td>D</td>
<td>The Pictures of SDS-PAGE</td>
<td>98</td>
</tr>
<tr>
<td>E</td>
<td>Molecular Weight of Protein Sample</td>
<td>99</td>
</tr>
<tr>
<td>F</td>
<td>Picture of SDS Gel</td>
<td>101</td>
</tr>
<tr>
<td>G</td>
<td>Amino Acid Sequence</td>
<td>103</td>
</tr>
</tbody>
</table>
CHAPTER 1

RESEARCH BACKGROUND

1.1 Introduction

_Eurycoma longifolia_ or commonly known as Tongkat Ali in Malaysia is a tropical herbal plant found in several parts of South East Asia countries. It is also known as Payung Ali, Penawar Pahit, Setunjang Bumi, Bedara Pahit, Tongkat Baginda, Pokok Syurga, Tongkat Ali Hitam, Pokok Jelas and Jelaih. There are four common species of Tongkat Ali, namely _Eurycoma longifolia_, _Eurycoma apiculata_, _Polyathiabullata_ and _Goniothalamus_ sp. (Aziz et al., 2003). Among these species, _Eurycoma longifolia_ is commonly used in traditional medication for various diseases. Traditionally, _E. longifolia_ is used for its aphrodisiac, anti-pyretic and anti-malarial effects as well as a general tonic (Kuo et al., 2003). A decoction of its long woody tap root is taken orally. The benefits of the roots of _E. longifolia_ include restoring energy and vitality, enhancing blood flow and functioning as herbal ingredient for women after child birth (Ismail et al., 1999). This plant bears fruit after two and a half years of cultivation, while the root is generally taken to be processed after 4 years of cultivation (Athimulam et al., 2006).
To our knowledge, plant protein from *E. longifolia* is seldom studied by researchers. It is believed that besides small metabolites, plant protein from the plant could be the bioactive constituent that contributed to the reported pharmacological properties (Chua *et al*., 2011). The detection of bioactive peptide (4.3 kD) which had been reported to have aphrodisiac activity from *E. longifolia* (Asiah *et al*., 2007) has encouraged more intensive research on protein or peptide in the plant. In the following year, another two plant proteins from *E. longifolia* with the molecular weight of 7.5 and 6.0 kD had been identified by Farouk *et al*. (2008). They reported the proteins showed antibacterial activity against human pathogenic bacteria. Antimicrobial peptides are ribosomally synthesised as natural antibiotics by almost all living organisms including bacteria, animals and plants. The plant still contains many proteins to be discovered, particularly for pharmacological applications.

1.2 Problem Statement

The hard cell walls of the plant have been reported to be the limitation factor for plant protein extraction and separation (Granier, 1988). Therefore, plant protein identification is a challenge since plant protein usually presents in small quantity. Plant proteins need to be extracted efficiently without degradation to ensure maximal yield of protein extraction. This study investigated the method of protein extraction that can give the highest number of proteins and the highest protein content. Usually, detergent is used to break the cell wall for proteins extraction. However, the type of detergent chosen must contribute to the lowest protein loss during extraction.
Recent technical advancement in sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and mass spectrometry (MS) have made the effort of protein separation and identification to be faster and higher in accuracy and sensitivity. There are limited studies on proteins from *E. longifolia* from gel electrophoresis to LC-MS/MS technique. This is mainly due to the minute amount of plant protein. Fortunately, this technique only requires small amount of sample for analysis. Somehow, extra precaution in sampling handling such as protein alkylation and reduction must take into consideration before mass spectrometric analysis.

There are still many types of proteins need to be explored, mainly for pharmacological application. For instance, phytohormones which are used in intercellular responses for plant growth and development might be useful for drug development. Phytohormones involve in plant defense mechanism in response to wound signal transduction by pests (Lindsey *et al.*, 2002). Plant protein could also be used as a carrier such as plant non-specific lipid-transfer proteins in drug delivery (Cheng *et al.*, 2004). However, the number of plant protein in the database is lesser than protein from bacteria and animals (Kim *et al.* and Park *et al.*, 2002). Only 0.001% of the total 961,019 number of total protein sequence belongs to plant protein (National Center for Biotechnology Information (NCBI) database, assessed on 26th April 2014).
1.3 Objective

The objective of this study was to identify plant proteins from *Eurycoma longifolia* (Tongkat Ali) root extracts collected from Pahang and Perak in Malaysia.

1.4 Scope

In order to achieve the objective of this study, the scopes were:

1. To investigate the protein extraction methods for *E. longifolia* roots using four different extraction methods, namely water, Triton X-100 (non-ionic detergent), phenol-SDS buffer and TCA-acetone precipitation methods.

2. To fingerprint plant proteins extracted from *E. longifolia* roots using one dimensional gel electrophoresis and mass spectrometric approaches.
REFERENCES


Christer, E., and Monica, N. (2003). Protein Extraction From Frozen Brain Samples. *Karolinska Institutet, Cancer Center Karolinska, Department of Oncology - Pathology*, R8:05, 171 76 Stockholm, Sweden


