RECONSTRUCTION OF *Nannochloropsis oculata* PROTEIN-PROTEIN INTERACTION NETWORK FOR GROWTH AND TRIACYLGLYCEROL PRODUCTION

KATIJAH BINTI MAHAT

UNIVERSITI TEKNOLOGI MALAYSIA
RECONSTRUCTION OF *Nannochloropsis oculata* PROTEIN-PROTEIN INTERACTION NETWORK FOR GROWTH AND TRIACYLGLYCEROL PRODUCTION

KATIJAH BINTI MAHAT

A thesis submitted in fulfilment of the requirements for the award of the degree of Doctor of Philosophy (Bioscience)

Faculty of Science
Universiti Teknologi Malaysia

JANUARY 2019
SPECIALY DEDICATED TO:

My Abah, Mahat bin Senon

My Emak, Habibah binti Hussein

My siblings:

Khamaliah, Kamaliana, Kamarudi, Kamisah dan Khairil Nizam

Brother and sister-in-laws:

Nazmi, Nur Hasanah, Mohd Zalman
In the name of Allah, the Most Beneficent, the Most Merciful’

Alhamdulillah, the greatest thankful to Allah S.W.T. for His blessing, I have been given health and inspiration to complete my PhD project and writing this thesis from the first page to the last page. I wish to express my deepest appreciation to my supervisor, Dr. Haryati binti Jamaluddin for her tremendous encouragement, professional advice, guidance and all the time had spent. Without her continuous support, this thesis would not be successful presented as it was. It was a great honour to finish this work under her supervision. Not forgetting many thanks to Professor Dr. Mohd Shahir Shamsir Bin Omar, Dr. Nor Azimah Bt Mohd Zain, Dr. Shaza Eva Binti Mohamad and Dr Goh Kian Mau for sharing their invaluable expertise, ideas and comments on the research conduct. It is also a pleasure to record a million thanks to my best friends; Badzilah, Qistina, Syafiqah, Kak Ana, Su, Mun, Kak Fareh, Karim, Haruna, Mia, Kak Lin, Balqis, Laila and colleagues at FBME for being such a wonderful friend against all the challenging experiences that we went through together.

To them I say “Although there is no perfect life, but we are able to fill it with lots of perfect moments together.” Not to forget the lab management team Puan Amalina, En Hairul Anuar and En Ruzaini Ramli for their dedication to their work that ease my PhD journey. Finally, special thanks to my beloved family for being my backbone, strength and supporting me all my way to complete the project, which made all this possible. I won’t be this stronger without you all as my inspiration. I love you all so much!
The aim of this study was to reconstruct protein-protein interaction map to allow researcher to identify pathway, biological roles and dynamic assembly of microalgae network pathways contributing to lipid production. The constructed PPI network was then characterised, and potential modifications were proposed based on this network to improve lipid production for mass biodiesel production. This bioinformatic approach is used to avoid the time consuming and expensive modification using high throughput approaches. In this study, literature search and data mining of rich information were interpreted into visualized outputs using Cytoscape. The outputs were then integrated and analysed to provide meaningful and reliable data. The study found that the constructed network consists of 153 nodes (proteins) and 1073 edges (interaction between proteins) with node degree $R^2$ coefficient distribution of 0.193 to exponent -0.387 denoting that the protein distributes randomly in the network and has clustering coefficient of 0.641 which was characterised by a highly connected node. The acyl-lipid (7-3)-desaturase (EDP09855) was identified as a bottleneck protein and a putative perturbation target for high quality biodiesel production. Cytoscape simulation and integration showed that alkaline growth conditions and phosphate and iron supplementation influenced PPI-directed pathway of *Chlamydomonas* sp. (model species) to glycerol lipid metabolism and fatty acid biosynthesis important for achieving high lipid productivity. In actual experimentation, microalgae *Nannochloropsis oculata* achieved highest percent of lipid yield when cultured in pH 8 growth conditions (3.45 %) or when growth was supplemented with 0.1 g/L phosphate (5.70 %) and 9.44 mg/L iron (6.21 %). Even though Cytoscape integration of three different vitamin B supplementation showed only minor possibility of pathway direction towards fatty acid biosynthesis in model species, maximum lipid productivity was achieved for *N. oculata* cultured using 500 μg/L biotin (4.98%), 50 μg/L thiamine (4.82%) and 0.5 μg/L cobalamin (7.59%). Altogether, computational biological pathway reconstruction positively supported network perturbations to optimize microalgae lipid productivity and actual experimentation validated the protein-protein network.
ABSTRAK

Tujuan kajian ini adalah untuk merangka semula peta interaksi protein-protein untuk membolehkan penyelidik mengenal pasti laluan, fungsi biologi dan rangkaian dinamik jejaring mikroalga yang menyumbang kepada pengeluaran lipid. Rangkaian PPI yang dibina kemudiannya dicirikan, dan pengubahsuaian yang berpotensi dicadangkan berdasarkan rangkaian ini untuk meningkatkan pengeluaran lipid bagi penghasilan biodiesel. Pendekatan bioinformatik ini digunakan untuk mengelakkan proses yang memakan masa yang lama dan kos yang tinggi dengan penggunaan pendekatan ‘high throughput’. Dalam kajian ini, pencarian literatur dan perlombongan maklumat yang kaya ditafsirkan ke dalam hasil visual menggunakan ‘Cytoscape’. Hasil ini kemudiannya diintegrasikan dan dianalisa untuk menghasilkan data yang bermakna dan boleh dipercayai. Kajian mendapati rangkaian yang dibina terdiri daripada 153 nod (protein) dan 1073 rangkaian (interaksi antara protein) dengan taburan koefisien darjah nod $R^2$ pada 0.193 hingga eksponen -0.387 menunjukkan bahawa protein tersusun secara rawak dalam rangkaian dan mempunyai pekali kluster 0.641 yang dicirikan oleh nod yang mempunyai hubungan tinggi. ‘Acyl-lipid (7-3) -desaturase’ (EDP09855) telah dikenal pasti sebagai protein cerutan dan sasaran pengubahsuaian untuk pengeluaran biodiesel yang berkualiti tinggi. Simulasi dan integrasi ‘Cytoscape’ menunjukkan bahawa keadaan pertumbuhan alkali dan suplemen fosfat dan zat besi mempengaruhi laluan yang diarahkan oleh PPI *Chlamydomonas* sp. (spesies model) ke metabolisme lipid gliserol dan biosintesis asid lemak yang penting untuk mencapai produktiviti lipid yang tinggi. Dalam eksperimen sebenar, mikroalga *Nannochloropsis oculata* mencapai peratusan tertinggi hasil lipid apabila dikulturkan dalam keadaan pertumbuhan pH 8 (3.45%) atau apabila pertumbuhan ditambah dengan 0.1 g/L fosfat (5.70%) dan 9.44 mg/L zat besi (6.21%). Walaupun integrasi ‘Cytoscape’ dari tiga jenis vitamin B yang berlainan menunjukkan hanya kemungkinan kecil arah jalur terarah ke biosintesis asid lemak dalam spesies model, produktiviti lipid maksimum *N. oculata* tercapai dengan menggunakan 500 μg/L biotin (4.98%), 50 μg/L tiamin (4.82%) dan 0.5 μg/L cobalamin (7.59%). Secara keseluruhannya, pembinaan semula laluan biologi komputasi menyokong pengubahsuaian rangkaian secara positif untuk mengoptimakan produktiviti lipid mikroalga dan eksperimen sebenar membuktikan rangkaian protein yang telah dibina adalah sah.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DECLARATION</td>
<td>ii</td>
<td></td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iii</td>
<td></td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>iv</td>
<td></td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>v</td>
<td></td>
</tr>
<tr>
<td>ABSTRAK</td>
<td>vi</td>
<td></td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>vii</td>
<td></td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xi</td>
<td></td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiv</td>
<td></td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xviii</td>
<td></td>
</tr>
<tr>
<td>LIST OF SYMBOLS</td>
<td>xx</td>
<td></td>
</tr>
<tr>
<td>LIST OF APPENDICES</td>
<td>xii</td>
<td></td>
</tr>
</tbody>
</table>

1 INTRODUCTION 1

1.1 Background information 1

1.2 Problem statement 4

1.3 Objectives 5

1.4 Scope of study 5

1.5 Significances of study 6

2 LITERATURE REVIEW 7

2.1 Introduction 7

2.2 Microalgae the photosynthetic microorganism 8
2.2.1 The marine microalgae *Nannochloropsis oculata*  

2.3 Alternative biofuels displacing petroleum derived transport fuels  

2.3.1 Biodiesel production from microalgae  

2.3.2 Microalgae lipid extraction and biodiesel conversion  

2.3.3 The classification of high-quality biodiesel  

2.3.4 Transferring the rich information of the predicted cellular pathway in the biological system into actual experimentation  

2.3.5 Microalgae cultivation process for biodiesel production  

2.3.6 Experimental limitation on optimizing biodiesel production form microalgae that bring out the biological pathway network protein-protein interaction network prediction  

2.4 The protein-protein interaction  

2.4.1 The protein-protein interaction network  

2.4.2 The computational method for protein-protein interaction network prediction  

2.4.3 Proteomic and metabolomic database  

2.4.4 Subcellular localization by the PredAlgo program  

2.4.5 Cytoscape: the visualization tools for mapping and integration of protein-protein interactions  

2.4.6 Biological pathway protein-protein interaction network reconstruction of microalgae growth and lipid productivity for high quality biodiesel production.  

2.4.7 The biological pathway involves in lipid and triacylglycerol production in microalgae  

2.4.8 The effect of nutrient requirements and pH of growth conditions on microalgae cellular pathway for nutrients uptake, accumulation, metabolism and regulation
3 RESEARCH METHODOLOGY

3.1 Introduction

3.2 Computational method of protein–protein interaction network construction

3.2.1 Chlamydomonas reinhardtii as a model organism for the protein–protein interaction network build up

3.2.2 The data collection and protein–protein interaction network mapping

3.2.3 The protein–protein interaction network filtering using STRING database confidence score

3.2.4 The construction of protein–protein interaction network prediction

3.2.5 The analysis and integration of protein–protein interaction network prediction

3.2.6 The protein of established network sub-cellular location prediction

3.3 Network validation by actual experimentation

3.3.1 Microalgae growth medium and nutrients

3.3.2 Autotrophic growth of microalgae

3.3.3 The microalgae growth evaluation

3.3.4 Microalgae harvest and cell disruption

3.3.5 Lipid Extraction

3.3.6 Transesterification process

3.3.7 Determination of fatty acid profile

3.3.8 Statistical analysis

4 RESULTS AND DISCUSSIONS

4.1 Introduction
4.1.1 Fix parameters and variable in each of the experimental set up 67

4.2 Enhancing microalgae growth and lipid productivity through identification of selected nutrient requirement and pH of growth condition factors based on the protein–protein interaction networks reconstruction 69

4.2.1 The construction of protein–protein interaction network 69

4.2.2 Effect of phosphate on microalgae growth and triacylglycerol production 91

4.2.3 Effect of iron on microalgae growth and triacylglycerol production 104

4.2.4 Effect of alkaline conditions on microalgae growth and triacylglycerol biosynthesis 115

4.2.5 Effect of three vitamins B concentration on microalgae growth and triacylglycerol biosynthesis 128

5 CONCLUSION AND FUTURE WORKS 149

5.1 Conclusions 149

5.2 Future works 150

REFERENCES 152

Appendices A-D 185-259
# 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>Effects of different nutrients and growth conditions on microalgae growth and lipid productivity</td>
<td>18</td>
</tr>
<tr>
<td>2.2</td>
<td>Lists of proteins involved in microalgae lipid and triacylglycerol synthesis</td>
<td>35</td>
</tr>
<tr>
<td>2.3</td>
<td>Lists of proteins involves in phosphate supplementation during microalgae growth</td>
<td>38</td>
</tr>
<tr>
<td>2.4</td>
<td>Lists of proteins involves in iron supplementation during microalgae growth</td>
<td>40</td>
</tr>
<tr>
<td>2.5</td>
<td>Lists of proteins involves in alkaline growth environment</td>
<td>43</td>
</tr>
<tr>
<td>2.6</td>
<td>Lists of proteins involves in biotin, thiamine and cobalamin supplementation during microalgae growth</td>
<td>47</td>
</tr>
<tr>
<td>3.1</td>
<td>The summary of software and database utilise in the study</td>
<td>51</td>
</tr>
<tr>
<td>4.1</td>
<td>Protein–protein interaction subcellular location prediction based on PredAlgo database</td>
<td>74</td>
</tr>
<tr>
<td>4.2</td>
<td>High degree nodes with evaluated betweenness and centrality and closeness and centrality in microalgae growth and triacylglycerol metabolic protein–protein interaction network</td>
<td>80</td>
</tr>
<tr>
<td>4.3</td>
<td>The names of genes and their respective classes in the eight cluster</td>
<td>82</td>
</tr>
<tr>
<td>4.4</td>
<td>The overrepresented gene ontology of biological process for each cluster</td>
<td>90</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4.5</td>
<td>Function associated to gene cluster related to phosphate supplementation of highly connected nodes</td>
<td>94</td>
</tr>
<tr>
<td>4.6</td>
<td>Individual functions associated with the highly connected node according to literature</td>
<td>95</td>
</tr>
<tr>
<td>4.7</td>
<td>The generation time of microalgae cultured in different phosphate concentration</td>
<td>101</td>
</tr>
<tr>
<td>4.8</td>
<td>The percentage of lipid productivity of microalgae cultured in different phosphate concentration</td>
<td>103</td>
</tr>
<tr>
<td>4.9</td>
<td>Function associated to gene cluster related to iron supplementation of highly connected node</td>
<td>107</td>
</tr>
<tr>
<td>4.10</td>
<td>Individual functions associated with iron supplementation the highly connected nodes according to literature</td>
<td>107</td>
</tr>
<tr>
<td>4.11</td>
<td>The generation time of microalgae cultured in different iron concentration</td>
<td>112</td>
</tr>
<tr>
<td>4.12</td>
<td>The percentage of lipid productivity of microalgae cultured in different iron concentration</td>
<td>114</td>
</tr>
<tr>
<td>4.13</td>
<td>Functions associated to gene cluster related to alkaline environment of highly connected node</td>
<td>118</td>
</tr>
<tr>
<td>4.14</td>
<td>Individual functions associated with alkaline environment the highly connected node according to literature</td>
<td>118</td>
</tr>
<tr>
<td>4.15</td>
<td>The generation time of microalgae cultured in different pH conditions</td>
<td>124</td>
</tr>
<tr>
<td>4.16</td>
<td>The percentage of lipid productivity of microalgae cultured in different pH conditions</td>
<td>127</td>
</tr>
<tr>
<td>4.17</td>
<td>Function associated to gene cluster related to biotin, thiamine, and cobalamin supplementation of highly connected nodes</td>
<td>130</td>
</tr>
<tr>
<td>4.18</td>
<td>Individual functions associated with biotin, thiamine, and cobalamin supplementation the highly connected nodes according to literature</td>
<td>130</td>
</tr>
<tr>
<td>4.19</td>
<td>The generation time of microalgae cultured in different biotin concentration</td>
<td>136</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>4.20</td>
<td>The percentage of lipid productivity of microalgae cultured in different biotin concentration</td>
<td>137</td>
</tr>
<tr>
<td>4.21</td>
<td>The generation time of microalgae cultured in different thiamine concentration</td>
<td>140</td>
</tr>
<tr>
<td>4.22</td>
<td>The percentage of lipid productivity of microalgae cultured in different thiamine concentration</td>
<td>142</td>
</tr>
<tr>
<td>4.23</td>
<td>The generation time of microalgae cultured in different cobalamin concentration</td>
<td>145</td>
</tr>
<tr>
<td>4.24</td>
<td>The percentage of lipid productivity of microalgae cultured in different cobalamin concentration</td>
<td>147</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>Predicted rates of global fossil fuel depletion (Stephens <em>et al.</em>, 2010)</td>
<td>11</td>
</tr>
<tr>
<td>2.2</td>
<td>General equation for biodiesel production (Fukuda <em>et al.</em>, 2001).</td>
<td>14</td>
</tr>
<tr>
<td>2.3</td>
<td>The microalgae growth phase (Moazami <em>et al.</em>, 2012)</td>
<td>16</td>
</tr>
<tr>
<td>2.4</td>
<td>The biochemical pathway of lipid and triacylglycerol production in microalgae with enzyme marked by yellow box</td>
<td>34</td>
</tr>
<tr>
<td>3.1</td>
<td>The overview of the study</td>
<td>51</td>
</tr>
<tr>
<td>4.1</td>
<td>The <em>Nannochloropsis oculata</em> morphology under electron microscope 1000X magnification</td>
<td>68</td>
</tr>
<tr>
<td>4.2</td>
<td>The phylogenetic relationship of 18S sequence in between <em>Nannochloropsis oculata</em> and <em>Chlamydomonas reinhardtii</em></td>
<td>70</td>
</tr>
<tr>
<td>4.3</td>
<td>The microalgae growth and triacylglycerol metabolic protein–protein interaction network (created by cystoscope)</td>
<td>71</td>
</tr>
<tr>
<td>4.4</td>
<td>The microalgae growth and triacylglycerol metabolic protein–protein interaction network with high confidence score created by Cytoscape.</td>
<td>72</td>
</tr>
<tr>
<td>4.5</td>
<td>Histogram showing distribution of the shortest path</td>
<td>76</td>
</tr>
<tr>
<td>4.6</td>
<td>Microalgae growth and triacylglycerol metabolic protein–protein interaction network is colour-filtered by node degree. The highly connected nodes are labelled as dark</td>
<td></td>
</tr>
</tbody>
</table>
blue-coloured, big-sized circles and highly connected edges are labelled with dark blue-coloured thick sticks.

4.7 Protein–protein interaction network associated with phosphate supplementation is colour-filtered by node degree. Highly connected nodes are labelled as dark blur-coloured, big-sized circles and highly connected edges are labelled with dark blue-coloured thick sticks.

4.8 ClueGO analysis of protein–protein interaction network associated with phosphate supplementation.

4.9 The optical density measurements of microalgae cultured in different phosphate concentration.

4.10 The dry cell weight measurements of microalgae cultured in different phosphate concentration.

4.11 The summary of microalgae subnetwork reconstruction analysis with phosphate supplementation.

4.12 The fatty acid methyl ester profile from biodiesel convertible lipid extracted from microalgae cultured in different phosphate concentration.

4.13 Protein–protein interaction network associated with iron supplementation is colour-filtered by node degree. Highly connected nodes are labelled as dark blue-coloured big-sized circles and highly connected edges are labelled with dark blue-coloured thick sticks.

4.14 ClueGO analysis of protein–protein interaction network associated with iron supplementation.

4.15 The optical density measurements of microalgae cultured in different iron concentration.

4.16 The dry cell weight measurements of microalgae cultured in different iron concentration.

4.17 The summary of the microalgae subnetwork reconstruction analysis with iron supplementation.
4.18 The fatty acid methyl ester profile from biodiesel convertible lipid extracted from microalgae cultured in different iron concentration

4.19 Protein–protein interaction network associated with alkaline environment is colour-filtered by node degree. Highly connected nodes are labelled with dark blue-coloured big-sized circles and highly connected edges are labelled with dark blue-coloured thick sticks

4.20 ClueGo analysis of protein-protein interaction network associated with alkaline growth environment

4.21 The optical density measurements of microalgae cultured in different pH conditions

4.22 The dry cell weight measurements of microalgae cultured in different pH conditions

4.23 The summary of the microalgae subnetwork reconstruction analysis in alkaline growth environment

4.24 The fatty acid methyl ester profile from biodiesel convertible lipid extracted from microalgae cultured in different pH conditions

4.25 Protein–protein interaction network associated with biotin, thiamine, and cobalamin supplementation is colour-filtered by node degree. Highly connected nodes are labelled as a dark blue-coloured, big-sized circles and highly connected edges are labelled with dark blue-coloured thick sticks

4.26 ClueGO analysis of protein–protein interaction network associated with biotin, thiamine, and cobalamin supplementation.

4.27 The optical density measurements of microalgae cultured in different biotin concentration

4.28 The dry cell weight measurements of microalgae cultured in different biotin concentration
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.29</td>
<td>The fatty acid methyl ester profile from biodiesel convertible lipid extracted from microalgae cultured in different biotin concentration</td>
<td>138</td>
</tr>
<tr>
<td>4.30</td>
<td>The optical density measurements of microalgae cultured in different thiamine concentration</td>
<td>139</td>
</tr>
<tr>
<td>4.31</td>
<td>The dry cell weight measurements of microalgae cultured in different thiamine concentration</td>
<td>140</td>
</tr>
<tr>
<td>4.32</td>
<td>The fatty acid methyl ester profile from biodiesel convertible lipid extracted from microalgae cultured in different thiamine concentration</td>
<td>142</td>
</tr>
<tr>
<td>4.33</td>
<td>The optical density measurements of microalgae cultured in different cobalamin concentration</td>
<td>144</td>
</tr>
<tr>
<td>4.34</td>
<td>The dry cell weight measurements of microalgae cultured in different cobalamin concentration</td>
<td>144</td>
</tr>
<tr>
<td>4.35</td>
<td>The summary of the microalgae subnetwork reconstruction analysis with biotin, thiamine and cobalamin supplementation</td>
<td>145</td>
</tr>
<tr>
<td>4.36</td>
<td>The fatty acid methyl ester profile from biodiesel convertible lipid extracted from microalgae cultured in different cobalamin concentration</td>
<td>148</td>
</tr>
</tbody>
</table>
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPI</td>
<td>Protein-protein interaction</td>
</tr>
<tr>
<td>KEGG</td>
<td>Kyoto Encyclopedia of Genes and Genomes</td>
</tr>
<tr>
<td>BLASTP</td>
<td>Basic Local Alignment Search Tool Protein</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>Tsv</td>
<td>Tab-separated values</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>UniProt</td>
<td>Universal Protein Knowledgebase</td>
</tr>
<tr>
<td>App</td>
<td>Application</td>
</tr>
<tr>
<td>API</td>
<td>Application programming interface</td>
</tr>
<tr>
<td>GO</td>
<td>Gene ontology</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>NADPH</td>
<td>Nicotinamide adenine dinucleotide phosphate</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>FAME</td>
<td>Fatty acid methyl ester</td>
</tr>
<tr>
<td>TAG</td>
<td>Triacylglycerol</td>
</tr>
<tr>
<td>BC</td>
<td>Betweenness Centrality</td>
</tr>
<tr>
<td>CC</td>
<td>Closeness centrality</td>
</tr>
<tr>
<td>GC</td>
<td>Gas chromatography</td>
</tr>
<tr>
<td>CoA</td>
<td>Coenzyme A</td>
</tr>
<tr>
<td>ACP</td>
<td>Acyl carrier protein</td>
</tr>
<tr>
<td>DHAP</td>
<td>Dihydroxyacetone phosphate</td>
</tr>
<tr>
<td>GPAT</td>
<td>Glycerol-3-phosphate acyltransferase</td>
</tr>
<tr>
<td>PAP</td>
<td>Phosphatidic acid phosphatase</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>DAG</td>
<td>Diacylglycerol</td>
</tr>
<tr>
<td>LPAT</td>
<td>Lysophosphatidic acid acyl-transferase</td>
</tr>
<tr>
<td>DAGAT</td>
<td>Diacylglycerol acyltransferase</td>
</tr>
<tr>
<td>PDAT</td>
<td>Phospholipid: diacylglycerol transferase</td>
</tr>
<tr>
<td>TCA</td>
<td>Tricarboxylic acid</td>
</tr>
<tr>
<td>C</td>
<td>Carbon</td>
</tr>
</tbody>
</table>
# LIST OF SYMBOLS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>Percent</td>
</tr>
<tr>
<td>≥</td>
<td>Less than and equal to</td>
</tr>
<tr>
<td>≤</td>
<td>Higher than and equal to</td>
</tr>
<tr>
<td>&lt;</td>
<td>Less than</td>
</tr>
<tr>
<td>&gt;</td>
<td>Higher than</td>
</tr>
<tr>
<td>G</td>
<td>Gram</td>
</tr>
<tr>
<td>Mg</td>
<td>Milligram</td>
</tr>
<tr>
<td>L</td>
<td>Liter</td>
</tr>
<tr>
<td>g/L</td>
<td>Gram per liter</td>
</tr>
<tr>
<td>mg/L</td>
<td>Milligram per liter</td>
</tr>
<tr>
<td>µg/L</td>
<td>Microgram per liter</td>
</tr>
<tr>
<td>Ml</td>
<td>Milliliter</td>
</tr>
<tr>
<td>µM</td>
<td>Micrometer</td>
</tr>
<tr>
<td>v/v</td>
<td>Volume per volume</td>
</tr>
<tr>
<td>°C</td>
<td>Degree Celsius</td>
</tr>
<tr>
<td>Rpm</td>
<td>Rotation per minute</td>
</tr>
<tr>
<td>Ppt</td>
<td>Parts per thousand</td>
</tr>
<tr>
<td>R²</td>
<td>R-squared</td>
</tr>
<tr>
<td>CO²</td>
<td>Carbon dioxide</td>
</tr>
<tr>
<td>Ω</td>
<td>Omega</td>
</tr>
<tr>
<td>Vitamin B₁</td>
<td>Thiamine</td>
</tr>
<tr>
<td>Vitamin B₇</td>
<td>Biotin</td>
</tr>
<tr>
<td>Vitamin B₁₂</td>
<td>Cobalamin</td>
</tr>
<tr>
<td>Symbol</td>
<td>Unit</td>
</tr>
<tr>
<td>--------------</td>
<td>------</td>
</tr>
<tr>
<td>$\text{div.day}^{-1}$</td>
<td>-</td>
</tr>
<tr>
<td>$\text{div/day}$</td>
<td>-</td>
</tr>
</tbody>
</table>
# 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 proteins involve in protein-protein interaction network and</td>
<td>185</td>
</tr>
<tr>
<td></td>
<td>subnetwork and key evidence from literature and database</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>SPSS software outputs</td>
<td>201</td>
</tr>
<tr>
<td>C</td>
<td>Standard curve for fatty acid methyl ester profile using gas</td>
<td>223</td>
</tr>
<tr>
<td></td>
<td>chromatography machine</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Gas chromatography outputs</td>
<td>228</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Background information

The functions of proteins are complicated because there is a hierarchal order in which proteins work with each other to form all sorts of biological processes and pathways. Understanding protein–protein interaction (PPI) can ease the discovery of existing interactions in organism as it is crucial for the revelation of protein functions and pathways in biological processes (Zahiri et al., 2013). Such understanding can be achieved by high throughput methods such as the yeast two-hybrid method (Xing et al., 2016). The results from PPI analysis can provide predictions of undiscovered protein complexes, hence enhancing the understanding of the underlying principles of the organisms’ cellular organisation and also predicting plausible functions of a protein in a particular biological pathway (King et al., 2004). However, genomics- and proteomics-based experimental methods for protein–protein interaction studies are often expensive and tedious and can be plagued with uncertainties and problems in reproducibility (Zahiri et al., 2013).

Direct experimental method approaches for optimisation of biodiesel production are unfeasible due to the complex regulatory network and mutual interaction of the physiological processes involved that require longer time to achieve desirable outcome (Rupprecht, 2009). As alternatives, several computational methods based on gene neighbourhood, gene fusion, phylogeny, docking, co-expression, and
interologs are developed to understand the protein–protein interaction, causing the recent rise of the systems biology discipline in the field of bioinformatics (Rupprecht, 2009). These computational methods have emerged as powerful platforms for protein–protein interaction network build-up, integration, and analysis that can also be employed to validate experimental results (Zahiri et al., 2013). Other than that, the ability to obtain information over a wide range of study framework and information gathered across species based on conserved protein sequences (Shen et al., 2007) facilitates the protein–protein interaction network build-up. Protein–protein interaction analysis can give insights on how to enhance high quality microalgae lipid production by understanding the rich information of the protein–protein interaction involved in specific pathways inside the microalgae (Blatti et al., 2012) because the biological pathway reconstruction can be done through wide range of physical and functional interaction (Muley and Ranjan, 2013) with possible alteration over different growth parameter implementation during cultivation. In addition, the protein-protein interaction dynamics and assembly causing the interaction to be more specific and for instance, one enzyme might favour a specific outcome meanwhile interactions between the enzyme with different partner might change the protein activity (Cesa et al., 2015). The protein–protein interaction network prediction is plausible but there is a lack of evidence on protein–protein interaction network study on microalgae, which increases the chances of getting false positive results. This problem can be resolved by the employment of gene ontology annotation (Mahdavi and Lin, 2007) that mark the biological process of each protein in the network to identify the targeted protein.

Microalgae are the most potential feedstock for biodiesel production as their oil can exceed 80% of the dry cell weight (Chisti, 2007). *Nannochloropsis oculata*, which belongs to the Eustigmatophyte group, has received so much attention because this species is rich with fatty acid and can achieve lipid content of up to 60% of dry cell weight (Doan et al., 2011), topping the chart as the most reliable species for biodiesel production. *Nannochloropsis oculata* lipid constituents consist of mainly neutral lipids (Huang et al., 2014) that can be readily converted into high quality biodiesel. Biodiesel from microalgae is produce by the transesterification of triacylglycerol (microalgae storage lipid) to produce fatty acid methyl ester
High quality biodiesel is characterised by a high cetane number value which indicates that the biodiesel has high ignition quality (Knothe et al., 2003). Cetane number increases as the degree of the fatty acid saturation decreases and the number of carbon increases, but the number of carbon needs to be kept low enough to avoid increase in viscosity that will lead to biodiesel deposition during cold temperatures (Knothe, 2005). Highly saturated fatty acids such as palmitic acid methyl ester, myristic acid methyl ester, stearic acid methyl ester, and undecanoic acid methyl ester are favoured components in biodiesel production as it will produce biodiesel with high cetane number (Rasoul-Amini et al., 2011) thus this had shown that microalgae should produce lipid that contain saturated or non-unsaturated fatty acid to ensure that the lipid being converted to fatty acid methyl ester (biodiesel) during transesterification is favourable lipid for high-quality biodiesel production.

Microalgae lipid content can be controlled by modifying certain aspects such as their growth conditions or nutrient requirements, which will result in the increase of triglyceride production (Goncalves et al., 2016) for instance alkaline environment, iron (Huang et al., 2014) and vitamins (Neumara et al., 2014) supplementation during growth causes increased accumulation of neutral lipid mainly saturated and monounsaturated fatty acids. Since algae lipid is expressed as the percentage of dry cell weight of the starting material, microalgae growth rate needs to be taken into consideration as lipid productivity will increase as the microalgae biomass increases. Studies have shown that iron (Sasireka and Muthuvelayudham, 2015) and vitamin supplementation (Li-xin et al., 2013) and alkaline conditions (Difusa et al., 2015) play important roles in triggering the increase of microalgae biomass. Reconstruction of biological pathway through wide range of physical or functional protein-protein interaction network prediction were made possible nowadays (Muley and Ranjan, 2013). There is growing body of knowledge stated that protein-protein interaction network were able to highlight the importance of microalgae pathway reconstruction for accelerating the development of commercially viable biodiesel production from algae biomass (Blatti et al., 2012). Therefore, this study attempts to enhance the mass productivity of high-quality biodiesel by manipulating microalgae growth aspects and lipid enhancing parameter in silico by reconstruction of the microalgae biological
pathway with the establishment of protein-protein interaction network prediction that will be further validated by laboratory-scale experiments. This research will extend our knowledge with two main strategies which are to increase the desired lipid per unit biomass and to increase the biomass density to maximise biomass per culture volume or area with the assistance of the established protein–protein interaction network reconstruction.

1.2 Problem statement

Since 1970, microalgae have been shown to be a promising third generation source for biodiesel production that can be sustainably developed (John S. et al., 1998). However, a lot of studies still need to be conducted to optimise the production of high-quality biodiesel to meet the global demand on energy consumption. The understanding of all cellular processes is possible with protein–protein interaction but understanding the protein–protein interaction by high-throughput methods is time consuming and far from being cost effective (Raman, 2010). Thus, in this study, the protein–protein interaction networks were reconstructed by using bioinformatics techniques since they are more economical, rapid, and convenient as proven by the study conducted by Han et al. (Han et al., 2016). These in silico analyses provide data to help us understand the pathways involved in the growth and lipid production of microalgae so that these pathways can be targeted for modifications to enhance the microalgae growth and lipid production. As there is lack of experimental evidence to confirm the microalgae network reconstruction, the gene ontology annotation was employed to overcome the problem with the false positive outcomes (Ji et al., 2014).
1.3 Objectives

The objectives of this study are:

1. To mine and map protein–protein interaction network based on microalgae growth and triacylglycerol production and find the important node for putative perturbation target.

2. To identify the role of selected parameters in enhancing microalgae growth and triacylglycerol productivity through the established subnetwork.

3. To enhance the growth and microalgae triacylglycerol production and validate the established subnetwork through experimentation.

1.4 Scope of study

The study scope comprised the generation of candidate pathway of the network reconstruction in which various data sources were pooled together by using efficient text mining and literature mining procedure (Zahiri et al., 2013) with the aid of tools such as Cytoscape plugins and open source databases. The data were then compiled, curated and filtered by using the STRING database (string-db.org) combined score (Szklarczyk et al., 2017). The created protein–protein interaction network was visualised and integrated by using Cytoscape. The predicted protein–protein interaction network provide insight on the parameters for optimisation of Nannochloropsis oculata high quality biodiesel production. Nannochloropsis oculata was cultured under different growth parameter. Microalgae lipid was extracted and converted to biodiesel. Fatty acid profile was characterised by using gas chromatography.
1.5 **Significances of study**

This study was carried out to promote microalgae growth and lipid productivity in which the protein-protein interaction map had reveals the possible protein interaction and major molecular functions association of genes involve in microalgae growth and triacylglycerol production. This study facilitates the optimization of microalgae growth and lipid productivity based on the reconstruction of biological pathway in the protein-protein interaction network with alteration on the pH of growth environment and nutrient supplementation such as iron, phosphate, biotin, thiamin and cobalamin supplementation on the microalgae growth media. This study indicates the feasibility of using protein–protein interaction network to reconstruct the biological pathway and guides specific experiments for researcher.


evolution in the supramolecular organization of photosystem I. *Biochimica et Biophysica Acta (BBA) - Bioenergetics.* 1837(2), 306-314.


Dhup, S., Kannan, D. C., and Dhawan, V. Growth, lipid productivity and cellular mechanism of lipid accumulation in microalgae *Monoraphidium* sp. following different phosphorous concentrations for biofuel production. *Current Science.* 112(3), 539-548.


Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U., Read, N. D., Jaffe, D., FitzHugh, W., Ma, L.-J., Smirnov, S., Purcell, S., Rehman, B., Elkins, T., Engels, R., Wang, S., Nielsen, C. B., Butler, J., Endrizzi, M., Qui, D., Ianakiev,


complexes retaining the PsbO, P and Q proteins from *Euglena gracilis*. *Plant and Cell Physiology*. 45(9), 1168-1175.


Xin, L., Hong-ying, H., Ke, G., and Ying-xue, S. (2010). Effects of different nitrogen and phosphorus concentrations on the growth, nutrient uptake, and lipid
accumulation of a freshwater microalga Scenedesmus sp. Bioresource Technology. 101(14), 5494-5500.


comparison with high-throughput experimental data. *Genome Research*. 17(4), 527-535.


