



UNIVERSITI PUTRA MALAYSIA

***MOLECULAR CHARACTERIZATION AND HOMOLGY MODELING OF
SHORT-CHAIN DEHYDROGENASE (GcSDR) AND NDP-SUGAR
EPIMERASES (GcNSEs) cDNAs FROM Gracilaria changii***

NUR ANISZA HANOUM BINTI NASERON

FBSB 2013 40



**MOLECULAR CHARACTERIZATION AND HOMOLGY MODELING OF
SHORT-CHAIN DEHYDROGENASE (GcSDR) AND NDP-SUGAR
EPIMERASES (GcNSEs) cDNAs FROM *Gracilaria changii***

By

NUR ANISZA HANOUM BINTI NASERON

**Thesis submitted to the School of Graduate Studies, Universiti Putra Malaysia,
in Fulfilment of the Requirements for the Degree of Master of Science**

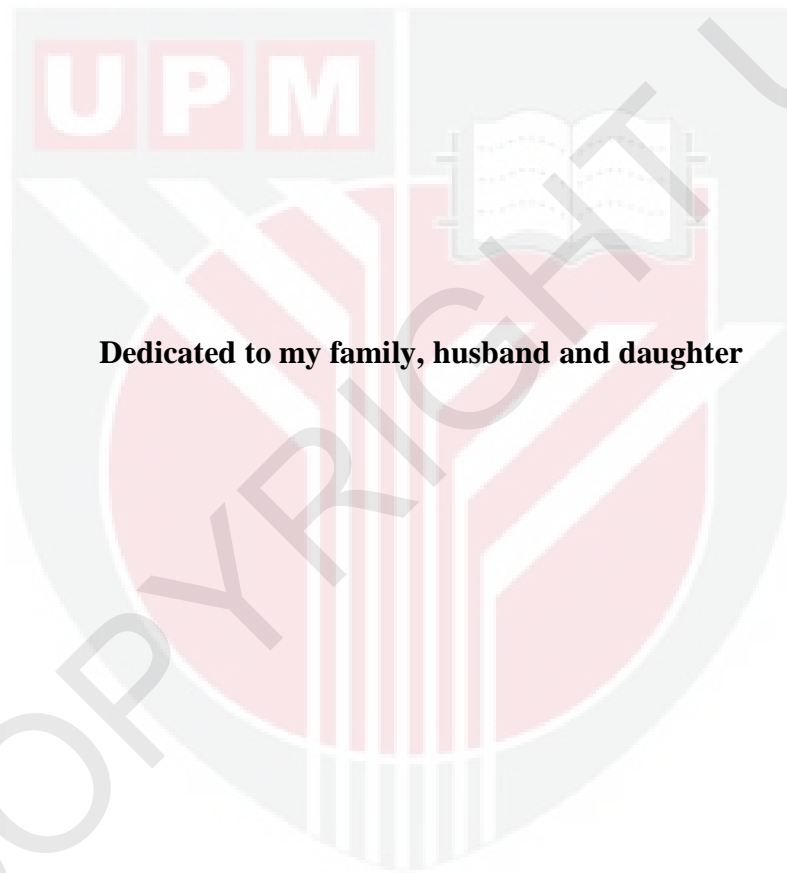
2014

COPYRIGHT

All material contained within the thesis, including without limitation text, logos, icons, photographs, and all other artwork, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial use of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright© Universiti Putra Malaysia





Dedicated to my family, husband and daughter

© COPYRIGHT UPM

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment
of the requirement for the degree of Master of Science

**MOLECULAR CHARACTERIZATION AND HOMOLGY MODELING OF
SHORT-CHAIN DEHYDROGENASE (GcSDR) AND NDP-SUGAR
EPIMERASES (GcNSEs) cDNAs FROM *Gracilaria changii***

By

NUR ANISZA HANOUM BINTI NASERON

October 2014

Chairman : Ho Chai Ling, PhD

Faculty : Biotechnology and Biomolecular Sciences

Gracilaria changii is a source of agar in Malaysia. Despite its economic importance, many of its genes are unknown. The objective of this study was to characterize the sequence and functions of three novel proteins from *G. changii* using an approach that combines molecular biology and bioinformatics. Amino acid sequence analysis of *Gracilaria changii* short-chain dehydrogenase (GcSDR) showed that the GxxxGxG co-enzyme binding motif and the YxxxK active-site motif are well conserved. GcSDR transcript encoded 282 amino acids with a predicted molecular mass of 48 kDa. The open reading frame (ORF) of GcSDR was successfully cloned into pET32a(+) and expressed as a soluble protein in *Escherichia coli* strain BL21 (DE3) pLysS. Recombinant GcSDR showed the highest activities when pyruvaldehyde dimethyl acetal was used as a substrate compared to alcohols and steroid, with a K_m value of 174.1 mM and a V_{max} value of 0.87 μmol of product formed $\text{min}^{-1} \text{mg}^{-1}$. GcSDR favoured NADPH rather than NADH, NADP^+ and NAD^+ as its primary co-factor. The three-dimensional structure of GcSDR was predicted using the crystal structure of a short-chain dehydrogenase (SDR) from *Streptomyces avermitilis* which shares 34.5 % identities. Virtual screening showed that CMP-N-acetyl-beta-neuraminate(2-) has the highest possibility to be the substrate of GcSDR with a total binding energy of -157.336 kJ/mol. The *Gracilaria changii* NDP-sugar epimerase 1 (GcNSE1) transcript encoded 576 amino acids (82 kDa) while *Gracilaria changii* NDP-sugar epimerase 2 (GcNSE2) encoded 401 amino acids (62 kDa). GcNSE1 shared high identities with the NADH: ubiquinone oxidoreductase complex I from *Fischerella* sp. JSC-11 while GcNSE2 has high identities with several NAD dependent epimerases/dehydratases. GcNSE1 possesses conserved co-enzyme binding pattern GxxxGxG and active site motif YxxxK for SDRs, while GcNSE2 has an atypical SDR co-enzyme binding site (GGxxxxG) and active site motif (YxxxN). The ORFs of GcNSE1 and GcNSE2 were cloned into pET32a(+) and transformed into *E. coli* strain BL21 (DE3) pLysS. Both recombinant GcNSE1 and GcNSE2 were produced as insoluble proteins. The insoluble fractions of both GcNSE1 and GcNSE2 were solubilized and refolded to produce soluble

proteins. Threading using i-TASSER was performed for GcNSE1 using the X-ray structure of a tyrosine-phosphorylated protein from *Arabidopsis thaliana* (At5g02240) as template. The 3D structure of GcNSE2 was predicted using the crystal structure of *Thermus thermophilus* Hb8 UDP-glucose 4-epimerase complex with NAD. Virtual screening showed that UDP-D-glucose with the lowest total binding energy (-180.755 kJ/mol) could be the possible substrate of GcNSE2. The 5' flanking regions of GcSDR, GcNSE1 and GcNSE2 were analyzed and putative *cis*-regulatory elements (CREs) were predicted. The light responsive elements were frequently found in the 5'-flanking regions of GcSDR, GcNSE1 and GcNSE2 indicating that these three proteins might be regulated by light. The full length gene sequences encoding these three cDNAs have no introns. In conclusion, the approach used in this study has shed lights on the amino acid sequences and putative functions of three unknown seaweed proteins.



Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk Ijazah Master Sains

**PENCIRIAN MOLEKUL DAN PERMODELAN HOMOLOGI cDNA
DEHIDROGENASE RANTAI PENDEK (GcSDR) DAN NDP-GULA
EPIMERASE (GcNSEs) DARI *Gracilaria changii***

Oleh

NUR ANISZA HANOUM BINTI NASERON

Oktober 2014

Pengerusi : Ho Chai Ling, PhD

Fakulti : Bioteknologi dan Sains Biomolekul

Gracilaria changii merupakan satu sumber agar di Malaysia. Walaupun ia mempunyai kepentingan ekonomi, kebanyakan gennya tidak diketahui. Objektif kajian ini adalah untuk mencirikan jujukan dan fungsi bagi tiga protein baru daripada *G. changii* menggunakan kaedah yang menggabungkan biologi molekul dan bioinformatik. Analisis jujukan asid amino *Gracilaria changii* dehidrogenase rantai pendek (GcSDR) menunjukkan bahawa motif pengikatan ko-enzim GxxxGxG dan motif tapak aktif YxxxK adalah terpelihara. GcSDR mengkod 282 asid amino dengan anggaran jisim molekulnya 48 kDa. Rangka bacaan terbuka (ORF) bagi GcSDR berjaya diklon ke dalam pET32a(+) dan dihasilkan sebagai protein larut di dalam *Escherichia coli* BL21 (DE3) pLysS. Rekombinan GcSDR menunjukkan aktiviti tertinggi apabila piruvaldehid dimetil asetal digunakan sebagai substrat berbanding alkohol dan steroid, dengan nilai K_m 174.1 mM dan nilai V_{max} 0.87 μmol produk terbentuk $\text{min}^{-1} \text{mg}^{-1}$. GcSDR lebih cenderung kepada NADPH berbanding NADH, NADP⁺ dan NAD⁺ sebagai ko-faktor utama. Struktur tiga dimensi GcSDR diramal menggunakan struktur kristal dehidrogenase rantai pendek (SDR) daripada *Streptomyces avermitilis* yang mempunyai identiti sebanyak 34.5 %. Saringan maya menunjukkan CMP-N-asetil-beta-neuraminat (2-) mempunyai kebarangkalian tertinggi sebagai substrat bagi GcSDR dengan jumlah tenaga pengikatan terendah iaitu -157.336 kJ/mol. Transkrip *Gracilaria changii* NDP-gula epimerase 1 (GcNSE1) mengkod 576 asid amino (82 kDa) sementara *Gracilaria changii* NDP-gula epimerase 2 (GcNSE2) mengkod 401 asid amino (62 kDa). GcNSE1 berkongsi identiti tertinggi dengan NADH: ubikuinon oksidoreduktase kompleks I daripada *Fischerella* sp. JSC-11 manakala GcNSE2 mempunyai identiti tertinggi dengan beberapa epimerase/dehidratase berpandukan NAD. GcNSE1 memiliki corak pengikatan ko-enzim terpelihara GxxxGxG dan motif tapak aktif YxxxK untuk SDR, sementara GcNSE2 mempunyai tapak pengikatan ko-enzim SDR atipikal (GGxxxxG) dan motif tapak aktif (YxxxN). ORF bagi GcNSE1 dan GcNSE2 telah diklon ke dalam pET32a(+) dan ditransformasi ke dalam *E. coli* BL21 (DE3) pLysS. Kedua-dua rekombinan GcNSE1 dan GcNSE2 dihasilkan sebagai protein tidak larut.

Fraksi tidak larut GcNSE1 dan GcNSE2 telah dilarut dan dilipat semula untuk menghasilkan protein larut. Kaedah bebenang menggunakan i-TASSER telah dilakukan untuk GcNSE1 menggunakan struktur sinar-X protein tirosin-terfosfat daripada *Arabidopsis thaliana* (At5g02240) sebagai templat. Struktur 3D bagi GcNSE2 diramalkan menggunakan struktur kristal dari *Thermus thermophilus* Hb8 UDP-glukosa 4-epimerase yang berkompleks dengan NAD. Penyaringan maya menunjukkan bahawa UDP-D-glucosa dengan tenaga pengikatan terendah (-180.755 kJ/mol) mungkin merupakan substrat yang sesuai untuk GcNSE2. Bahagian pengapit 5' bagi GcSDR, GcNSE1 dan GcNSE2 telah dianalisis dan elemen pengawalaturan-*cis* (CREs) telah diramal. Elemen pengawalan cahaya kerap dijumpai di bahagian pengapit 5' bagi GcSDR, GcNSE1 dan GcNSE2 menunjukkan bahawa ketiga-tiga protein ini mungkin dikawal oleh cahaya. Jujukan gen penuh yang mengkodkan ketiga-tiga cDNA tidak mempunyai sebarang intron. Kesimpulannya, kaedah yang digunakan dalam kajian ini telah memberikan maklumat tambahan mengenai jujukan dan fungsi putatif bagi tiga protein rumpai laut yang tidak diketahui.

ACKNOWLEDGEMENTS

Alhamdulillah, praised to Allah, the Almighty who gave me the opportunity to breath in this beautiful earth, I was able to complete this wonderful journey in my life. I would like to express the highest gratitude to the person who gave me the chance to work on this enlightened research, Associate Professor Dr. Ho Chai Ling, who is also my supervisor throughout my study periods. No words can describe how helpful she is in passionately guiding me, tolerating with my childish acts, great flow of ideas and positive criticism as well as endless amount of moral supports. My deepest appreciation also directed towards my supervisory committee, Associate Professor Dr. Parameswari Namasivayam, who always talks to me in a supportive ways.

I would like to extend my gratitude to my laboratory senior, Mr. Lim Kok Ang and Mr. Teoh Seddon who provided me with the cloned GcSDR and GcNSE2; and Miss Hariyanti Baharum who guided me with the enzymatic analysis. To Dr. Adam Leow Thean Chor from the Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, many thanks for your help in guiding me with the virtual screening processes. To Miss Siow Rouh San, thank you very much for being my most trusted research companion with her endless help, thoughtful acts and wonderful smile during our 3 years of research together. Also to Mr. Putra Taufik Gutawa Lok, for his supportive and brilliant ideas in solving the promoter isolation problems, I thanked you for it.

To all my other labmates from the Plant Molecular Biology Laboratory, Miss Kalai Vani d/o Maniam, Miss Erica Khew, Miss Tee Syin Ying, Miss Lim Ee Leen and Mr. Tan Yung Chie; from the Cell and Molecular Biology Laboratory, Miss Vasagi d/o Ramachandran, Miss Conie Toh and Miss Kanagamalar d/o Silvarajoo, I appreciate the friendship, laughter and joy that we all share together.

Lastly, I also want to thank my beloved mother, Madam Azizah Mohd Dom, my siblings and my lovely husband, Mr. Aizat Syafiee Abu Zahid who put their strong beliefs in my abilities, continuously pray, love and support me that help me to complete this study. I would like to dedicate this thesis to my beloved daughter, Aryssa Reina Aizat Syafiee for being part of the journey in completing my studies.

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master of Science. The members of the Supervisory Committee are as follows:

Ho Chai Ling, Ph.D.

Senior Lecturer

Faculty of Biotechnology and Biomolecular Sciences

Universiti Putra Malaysia

(Chairman)

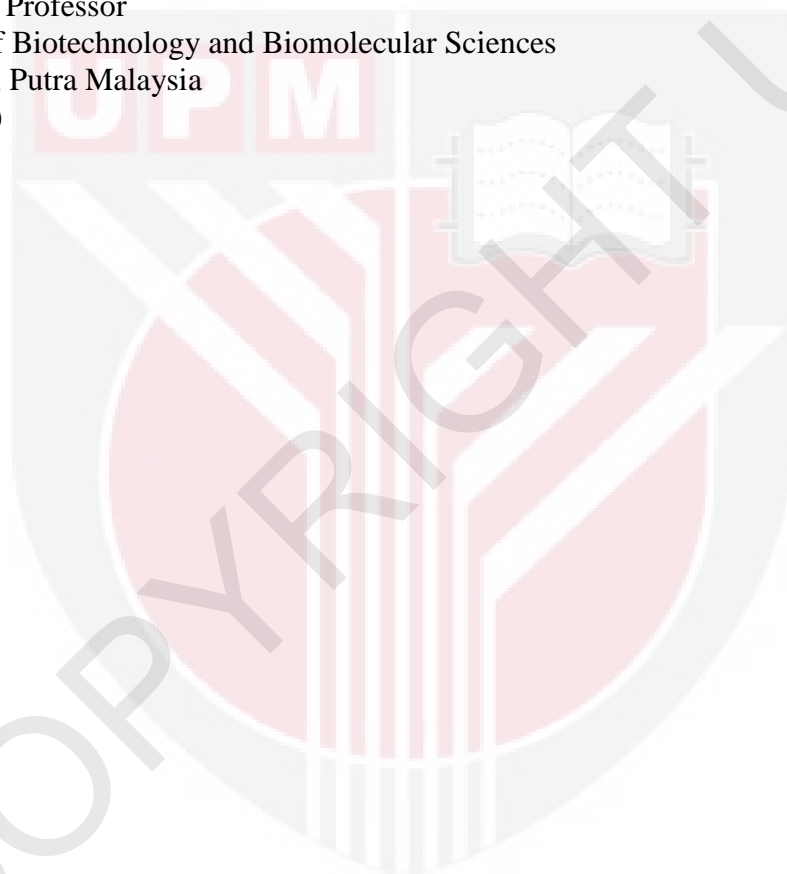
Parameswari d/o Namasivayam, Ph.D.

Associate Professor

Faculty of Biotechnology and Biomolecular Sciences

Universiti Putra Malaysia

(Member)



BUJANG BIN KIM HUAT, PhD

Professor and Dean

School of Graduate Studies

Universiti Putra Malaysia

Date:

DECLARATION

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012 – 2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: _____ Date: _____

Name and Matric No.: _____

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012 – 2013) are adhered to.

Signature: _____
Name of
Chairmain of
Supervisory
Committee: _____

Signature: _____
Name of
Chairmain of
Supervisory
Committee: _____

Signature: _____
Name of
Chairmain of
Supervisory
Committee: _____

Signature: _____
Name of
Chairmain of
Supervisory
Committee: _____

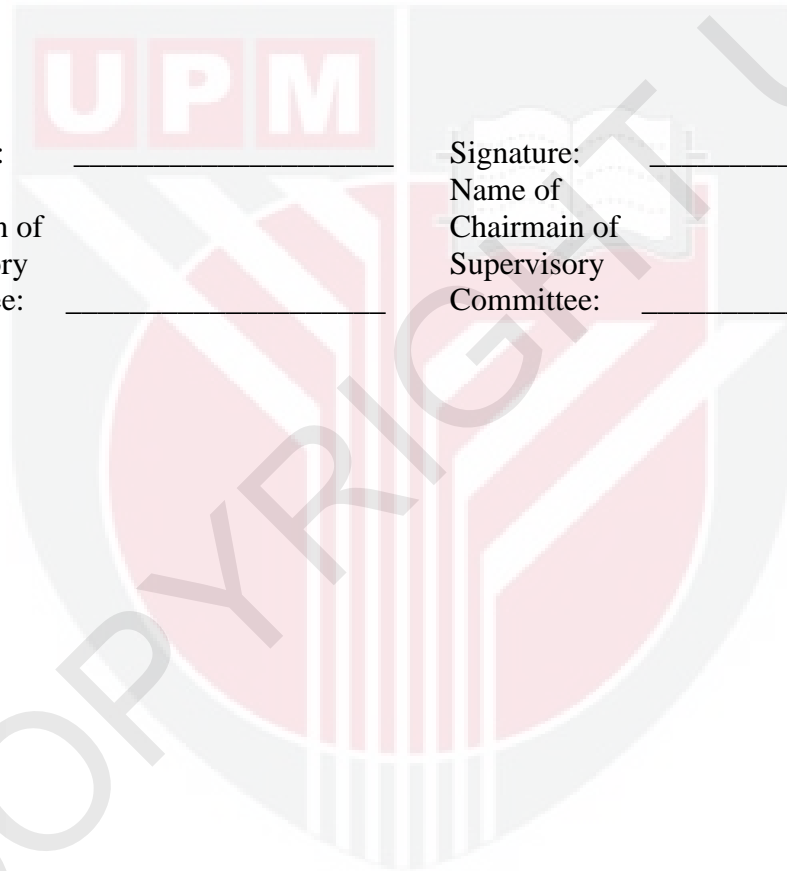


TABLE OF CONTENTS

	Page
ABSTRACT	i
ABSTRAK	iii
ACKNOWLEDGEMENTS	v
APPROVAL	vi
DECLARATION	vii
LIST OF TABLES	xiii
LIST OF FIGURES	xiv
LIST OF ABBREVIATIONS	xvii

CHAPTER

1	INTRODUCTION	1
2	LITERATURE REVIEW	3
2.1	Algae	3
2.2	Benthic macroalgae	3
2.2.1	Seaweeds	3
2.2.2	Red seaweeds	6
2.2.3	<i>Gracilaria changii</i>	8
2.3	Short-chain dehydrogenases	11
2.3.1	Structural characteristics of short-chain dehydrogenases	11
2.3.2	Classification of SDRs	12
2.3.2.1	Classical SDRs	12
2.3.2.2	Extended SDRs	12
2.3.2.3	Intermediate SDRs	13
2.3.2.4	Divergent SDRs	13
2.3.2.5	Complex SDRs	14
2.3.3	NAD (H)- and NADP (H)-dependent SDRs	14
2.3.4	Functions of SDRs in plants	17
2.4	NDP-sugar epimerases	19
2.4.1	Definition and mechanisms of epimerization	19
2.4.2	Functions of NDP-sugar epimerases in plants	21
2.5	Prediction of secondary structure and identification of substrate	23
2.5.1	Homology modeling	23
2.5.2	Threading or fold-recognition	24
2.5.3	Assessing the quality of structure prediction	24
2.5.4	Determination of enzyme substrate using molecular docking	24
3	MATERIALS AND METHODS	26
3.1	Experimental materials	26
3.1.1	Collection of <i>Gracilaria changii</i>	26
3.1.2	Recombinant clones of short-chain dehydrogenase (GcSDR) and NDP-sugar epimerase 2 (GcNSE2)	26
3.1.3	Preparation of bacterial competent cell	27

3.2	Cloning and transformation of NDP-sugar epimerase 1 (GcNSE1)	27
3.3	Recombinant protein expression and purification	28
3.3.1	GcSDR	28
3.3.2	GcNSE1 and GcNSE2	29
3.4	Solubilization and refolding of recombinant proteins	29
3.5	Western blot analysis	30
3.6	Homology modeling for secondary structure prediction	31
3.6.1	GcSDR	31
3.6.2	GcNSE1	31
3.6.3	GcNSE2	32
3.7	Virtual screening and molecular docking of possible substrates for GcSDR, GcNSE1 and GcNSE2	32
3.7.1	Preparation of the active sites	32
3.7.2	Virtual screening and molecular docking of substrates	33
3.8	Biochemical analysis of recombinant proteins	33
3.8.1	GcSDR	33
3.8.2	GcNSE1	35
3.8.3	GcNSE2	35
3.9	Isolation of genomic sequences	35
3.9.1	Genomic DNA extraction	35
3.9.2	Preparation of genomic DNA libraries	36
3.9.3	Isolation of the 5' flanking region of GcSDR	37
3.9.4	Isolation of the 5' flanking region of GcNSE1	38
3.9.4.1	GenomeWalker approach	38
3.9.4.2	Inverse PCR approach	39
3.9.5	Isolation of the 5' flanking region of GcNSE2	40
3.9.5.1	GenomeWalker approach	40
3.9.5.2	Inverse PCR approach	40
3.10	Sequence analyses and prediction of <i>cis</i> -regulatory elements (CREs) at the 5' flanking sequences	41
3.10.1	GcSDR	41
3.10.2	GcNSE1 and GcNSE2	41
4	RESULTS	42
4.1	Sequence analysis and phylogenetic analysis of GcSDR transcript	42
4.2	Sequence analysis of GcNSE2 transcript	42
4.3	Cloning and sequence analysis of GcNSE1 transcript	49
4.4	Expression of recombinant proteins	53
4.4.1	GcSDR	53
4.4.2	GcNSE1 and GcNSE2	53
4.5	Functional analysis of recombinant proteins from <i>Gracilaria changii</i>	57
4.5.1	GcSDR	57
4.5.2	GcNSE1	59
4.5.3	GcNSE2	59
4.6	Homology modeling and virtual screening	60
4.6.1	GcSDR	60
4.6.2	GcNSE1	63
4.6.3	GcNSE2	66

4.7	The isolation and analyses of 5' flanking sequence of GcSDR, GcNSE1 and GcNSE2	70
4.7.1	Genomic DNA isolation from <i>G. changii</i>	70
4.7.2	GcSDR	72
4.7.3	GcNSE1	75
4.7.4	GcNSE2	77
4.8	Isolation of the full length structural sequences	81
4.8.1	GcSDR	81
4.8.2	GcNSE1	82
4.8.3	GcNSE2	82
5	DISCUSSIONS	83
5.1	GcSDR	83
5.2	GcNSE1	86
5.3	GcNSE2	88
6	SUMMARY, CONCLUSION AND FUTURE RECOMMENDATIONS	92
	REFERENCES	94
	APPENDICES	118
	Appendix 1 : Virtual screening results of GcSDR	
	Appendix 2 : Virtual screening results of GcNSE2	
	Appendix 3 : Summary of <i>cis</i> -acting regulatory elements (CREs)	
	Appendix 4 : Pairwise alignment between full length structural sequence of GcSDR and its corresponding cDNA sequence	
	Appendix 5 : Pairwise alignment between full length structural sequence of GcNSE1 and its corresponding cDNA sequence	
	Appendix 6 : Pairwise alignment between full length structural sequence of GcNSE2 and its corresponding cDNA sequence	
	BIODATA OF STUDENT PUBLICATIONS	140
		141