MOLECULAR DOCKING OF HSPs (HEAT SHOCK PROTEINS) FROM RICE SEED

LEE KE VIN

A dissertation submitted in partial fulfillment
of the requirements for the award of the degree of
Master of Science specialized in Biotechnology

Faculty of Biosciences and Medical Engineering
Universiti Teknologi Malaysia

JUNE 2018

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisors, Dr Zaidah Rahmat who accepted me as one of her mixed mode master students. She always guides me during accomplishment of my project as well as sharing her professional knowledge with me.

I would also like to express my appreciation to the faculty officer, Pn Linda who assigned my working place in Bioinformatics Computer Lab and prepared the facilities for my convenience. Besides that, I would like to express my special thanks to post graduate students especially Ms Ng Mei Ling who always show her patience in guiding me which helps me to finish the project work on time and smoothly.

Finally, I would like to thank my family for their understanding and support to me. Last but not least, I wish to show my appreciation to all my coursemates who helped me during this period. Thank You.

ABSTRACT

The expression of heat shock proteins (HSPs) is an essential part of the heat shock response in plants where HSPs act as chaperones that help in protein folding and unfolding mechanism. This project focuses on determining the docking mechanism and conformation stability of HSPs from rice seed at several simulated temperatures. HSP20 was selected from previously isolated rice seed proteins and the sequence was used for homology modelling. The suitable model of this sequence was then selected to undergo molecular dynamics simulation and docking procedure with the targeted proteins. The simulated environmental temperature was set as 37ºC and 100ºC during the simulation process. Model for the selected HSP20 protein was generated successfully via I-Tasser server. The results of 50ns simulation at 310K and 373K for selected model were plotted in the graphs of RMSD, RMSF, hydrogen bonding number and radius of gyration. The RMSD result showed that HSP20 model was more stable at simulated temperature of 37ºC as compared to 100ºC. Whereas for the RMSF graph visualization, two significant loops of this model were found in the same position of its 3D structure as corresponded in both 37ºC and 100ºC results. For the results of hydrogen bonding number and radius of gyration, the mean numbers were around 102 and 1.6 respectively for 37ºC and 100ºC simulations. These findings indicated the responsible stability and flexibility of the model's 3D structure in terms of its secondary structure, folding pattern and loops location. Molecular docking of this HSP20 model with the selected proteins: TPR and SGT1 was carried out successfully with ZDock server. The docked structures generated were used to understand the docking mechanism and protein-protein interactions between these proteins. Further study of these protein models is required for understanding on the roles of binding conformation between them as well as possible protein-protein interaction of HSP20 with other co-chaperones.

ABSTRAK

Ekspresi heat shock proteins (HSPs) atau protein tahan haba adalah suatu tindak balas tahan haba oleh tumbuh-tumbuhan. HSPs berperanan sebagai ‘chaperones’ dengan membantu dalam mekanisme penglipatan protein. Kajian ini menumpu pada mekanisme dok dan kestabilan struktur HSPs daripada beras padi dalam beberapa suhu simulasi. HSP20 dipilih daripada kandungan protein beras padi terpencil terdahulu dan digunakan untuk pemodelan homologi. Model yang sesuai seterusnya dipilih demi penjalanan simulasi dinamik molekul. Suhu simulasi ditetapkan pada 37^{o}C dan 100^{o}C sepanjang proses simulasi. Model protein HSP20 telah berjaya dihasilkan dengan server I-Tasser. Hasil simulasi bagi 50ns pada suhu 310K dan 373K telah diplotkan dalam graf-graf RMSD, RMSF, jumlah ikatan hidrogen serta radius gyration. Graf RMSD menunjukkan bahawa model HSP20 adalah lebih stabil pada suhu simulasi 37^{o}C berbanding dengan 100^{o}C. Manakala dalam graf RMSF, didapati kedua-dua gelung dalam model ini berada pada kedudukan yang sama seperti 3D strukturnya sebagaimana dalam kehasilan simulasi pada 37^{o}C and 100^{o}C. Dalam graf-graf jumlah ikatan hidrogen dan radius gyration, nombor mean adalah kira-kira 102 dan 1.6 masing-masing bagi kedua-dua simulasi pada 37^{o}C dan 100^{o}C. Penemuan ini juga bersesuaian dengan keputusan struktur 3D bagi model HSP20 dari segi struktur sekunder, bentuk lipatan dan lokasi gelung yang berkenaan. Proses dok molekul bagi model HSP20 dengan protein TPR dan SGT1 telah berjaya dijalankan dalam server ZDock. Struktur-struktur hasilan proses dok telah digunakan untuk memahami mekanisme dok serta interaksi protein-protein antara mereka. Kajian lanjut bagi model-model ini adalah diperlukan bagi pemahaman berhubung peranan ikatan dalam konformasi mereka serta kemungkinan interaksi protein-protein antara HSP20 dengan protein ‘co-chaperone’ yang lain.

TABLE OF CONTENTS

CHAPTER	TITLE		PAGE
	DECLARATION		ii
	ACKNOWLEDGEMENTS		iv
	ABSTRACT		v
	ABSTRAK		vi
	TABLE OF CONTENTS		vii
	LIST OF TABLES		x
	LIST OF FIGURES		xi
	LIST OF APPENDICES		xiv
	LIST OF SYMBOLS		xv
	LIST OF ABBREVIATIONS		xvi
1	**INTRODUCTION**		
	1.1	Background Information	1
	1.2	Problem Statement	2
	1.3	Objectives	3
	1.4	Significance of Work	3
	1.5	Scope of Work	3
2	**LITERATURE REVIEW**		
	2.1	Rice	5
	2.2	Heat Shock Proteins (HSPs)	6

	2.3	Protein-protein Interaction of Chaperone and Co-chaperone	8
		2.3.1 SGT1	8
		2.3.2 TPR	9
	2.4	Proteins Structure Simulation	10
3	**MATERIALS AND METHODS**		
	3.1	Experimental Design	12
	3.2	HSP20 Sequence Selection	13
	3.3	Homology Modelling	13
	3.4	Molecular Dynamic Simulation	13
		3.4.1 Topology generation	14
		3.4.2 Setting Simulation Box and Solvating the Box	15
		3.4.3 Neutralization Stage	16
		3.4.4 Energy Minimization Stage	16
		3.4.5 Equilibration Stage	17
		3.4.6 Production Stage	18
	3.5	Molecular Docking	18
4	**RESULTS AND DISCUSSION**		
	4.1	Homology Modelling	19
		4.1.1 HSP20	19
	4.2	Molecular Dynamics (MD) Simulation	23
		4.2.1 Root Mean Square Deviation (RMSD)	23
		4.2.2 Root Mean Square Fluctuation (RMSF)	25
		4.2.3 Hydrogen Bonds	26
		4.2.4 Radius of Gyration	28
	4.3	Protein-protein Docking	29
		4.3.1 HSP20-TPR	29
		4.3.2 HSP20-SGT1	33

4.3.3 Overall comparison for Simulated Models 37

5 CONCLUSIONS AND FUTURE WORK

5.1 Conclusions 40

5.2 Future work 41

REFERENCES

APPENDIX

LIST OF TABLES

TABLE	TITLE	PAGE
4.1	Model validation summary using different bioinformatics tools	22
4.2	MD simulation comparison summary between HSP1, TPR1 and SGTA with setting of 50ns at 37^{o}C	38

LIST OF FIGURES

FIGURE	TITLE	PAGE
2.1	Crystal structure of HSP16.9 monomer from wheat	7
2.2	Crystal structures of SGT1 CS domain and N-terminal of HSP90	9
2.3	Crystal structure of TPR domain in Hop protein	10
3.1	Flow chart of summarized methodology in this study	12
4.1	Full sequence of HSP20 (Uniprot ID: B0FFN6) where the related MS sequences from 2 cultivars were highlighted in yellow	19
4.2	Sequence alignment of HSP20 (Uniprot ID: B0FFN6) with selected template (PDB ID: 1GME) by using ClustalW and presented via ESPript. The conserved regions were highlighted in red and loop locations were shown with yellow boxes	20
4.3	The 3D structure of HSP1 model with alpha helices shown in light blue ribbons and beta strands in purple ribbons	21
4.4	Superimposition of HSP1 model and template, 1GME. The red ribbon represents HSP1 structure, while the blue ribbon represents 1GME model. The loops were highlighted in yellow circles	21
4.5	RMSD graph of HSP20 model (HSP1) for 50ns simulation at 310K	24
4.6	RMSD graph of HSP20 model (HSP1) for 50ns simulation at 373K	24

4.7	RMSF graph of HSP20 model (HSP1) for 50ns simulation at 310K	25
4.8	RMSF graph of HSP20 model (HSP1) for 50ns simulation at 373K	26
4.9	Hydrogen bonding number graph of HSP20 model (HSP1) for 50ns simulation at 310K	27
4.10	Hydrogen bonding number graph of HSP20 model (HSP1) for 50ns simulation at 373K	27
4.11	Rg graph of HSP20 model (HSP1) for 50ns simulation at 310K	28
4.12	Rg graph of HSP20 model (HSP1) for 50ns simulation at 373K	29
4.13	Docked structure of HSP1 model with the selected TPR, 4YVO. The red ribbon represents HSP1 model, while the blue ribbon represents 4VYO structure	30
4.14	Protein-protein interaction between the HSP1 N-terminal (red sticks) and TPR structure of 4YVO (light blue sticks) in TPR1 complex as highlighted in orange circles, where the yellow dotted lines represented the hydrogen bonds	30
4.15	RMSD graph of docked model (TPR1) for 50ns simulation at 310K	31
4.16	RMSF graph of docked model (TPR1) for 50ns simulation at 310K where the red line separated the regions of HSP20 (left) and TPR (right)	32
4.17	Hydrogen bonding number graph of docked model (TPR1) for 50ns simulation at 310K	32
4.18	Rg graph of docked model (TPR1) for 50ns simulation at 310K	33
4.19	Docked structure of HSP1 model with the chosen SGT1, 2JKI. The red ribbon represents HSP1 model, while the blue ribbon represents SGT1 CS domain	34
4.20	Protein-protein interactions in SGTA between the residues of HSP1 N-terminal (red sticks) and SGT1 CS domain (light blue	34

	sticks), where the major binding regions were highlighted in orange circles	
4.21	RMSD graph of docked model (SGTA) for 50ns simulation at 310K	35
4.22	RMSF graph of docked model (SGTA) for 50ns simulation at 310K where the red line separated the regions of HSP20 (left) and SGT1 (right)	36
4.23	Hydrogen bonding number graph of docked model (SGTA) for 50ns simulation at 310K	36
4.24	Rg graph of docked model (SGTA) for 50ns simulation at 310K	37

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	Em.mdp file for energy minimization of the system	46
B	Pr.mdp file for equilibration of the system	48
C	Md.mdp file for 50 ns MD simulation	50

LIST OF SYMBOLS

α	-	Alpha
β	-	Beta
^{o}C	-	Degree Celcius
K	-	Kelvin
kDa	-	Kilo Dalton
nm	-	Nanometer
ns	-	Nanosecond
%	-	Percentage

LIST OF ABBREVIATIONS

CHORD	-	Cysteine and histidine-rich domain
et al.	-	And friends
HSP	-	Heat Shock Protein
kDa	-	Kilo Dalton
Mb	-	Megabases
MD	-	Molecular Dynamics
NMR	-	Nuclear Magnetic Resonance
PDB	-	Protein Data Bank
Rg	-	Radius of gyration
RMSD	-	Root Mean Square Deviation
RMSF	-	Root Mean Square Fluctuation
SGT1	-	Suppressor of G2 allele of SKP1
SKP1	-	S-Phase Kinase Associated Protein 1
TPR	-	Tetratricopeptide Repeat

CHAPTER 1

INTRODUCTION

1.1 Background Information

One common environmental variable encountered by organisms is the surrounding temperature that influences homeostasis maintenance. High temperature will cause negative impacts that interrupt cellular homeostasis and physiology at various growth stages of plants. A complex signaling network is therefore constructed by plants to minimize the effects of heat stress. This involves rapid induction of genes with the expression of heat shock proteins (HSPs), which acts as an essential part of the heat shock response (Zhang *et al*., 2016). In general, HSPs are categorized in terms of their molecular weight and have been grouped in families including HSP90, HSP70, HSP60, HSP40, and small HSPs (sHSPs). Basically, HSPs act as chaperones which help in protein folding and unfolding mechanism. They also served as stress-response factors that have vital role in a range of cell processes, such as apoptosis, cell cycle and division, development as well as differentiation (Martín-Folgar *et al*., 2015).

Rice (*Oryza sativa* L.) can be referred as one of the critical crops in eastern Asia. Numerous rice cultivars and wild species of rice have been widely grown for identification

of their genetic and molecular makeup. According to the study of Bernier and co-workers, rice ecosystems can be divided into four major categories: deep-water, irrigated, rainfed lowland and rainfed upland (Bernier *et al*., 2008). On the other hand, rice is also considered as model plant in monocot category due to its small size genome which consists of about 430 million base pairs and around 30,000 genes responsible for proteins production (Komatsu *et al*., 2003). When it comes to proteomics in rice, it basically involved the study of proteomes in various parts of rice. These proteins are resulted from specific gene expression under various growth and developmental conditions, as well as under different abiotic and biotic stresses. Kim and co-workers mentioned about technological development and improvement in both gel-based and gel-free approaches, especially MS analysis and bioinformatics resources in plant proteomics have contributed greatly to the global advance of rice proteomics analyses (Kim *et al*., 2014).

1.2 Problem Statement

HSPs expression can be referred as the strategy of plants to cope with the several biotic and abiotic stresses as well as environmental changes due to their sessile properties. Study regarding the mechanism of HSPs in plants is particularly important in understanding of their functions and discover more of their potential roles. Although currently there are some studies on HSPs from rice plants, but molecular docking of HSPs sequences isolated from rice seed is still not available. Moreover, there is also a lack of research on binding mechanism of HSP20 with co-chaperones or other proteins, which is crucial for HSPs role in plant immunity. Thus, a study on comparison of selected HSP20 conformation under different simulated temperature would provide preliminary information before proceeding to the determination of HSP20 stability and its docking mechanism followed by the prediction on protein-protein interactions of HSP20 with the chosen proteins.

1.3 Objectives

1. To determine the stability of HSP20 at 37ºC and 100ºC
2. To determine the docking mechanism of HSP20
3. To predict the protein-protein interaction of HSP20 with coat protein, TPR (Tetratricopeptide Repeat) and SGT1 (Suppressor of G2 allele of SKP1)

1.4 Significance of Work

This project focuses on determining the docking mechanism and conformation stability of HSPs from rice seed at several simulated temperatures. Current studies on HSPs are mainly based on animal sources rather than plants for eukaryotes and study regarding rice HSPs is very few. There is still absent of work related to molecular docking of HSPs from rice seed. Hence, this study could provide information for structure stability of selected HSPs at different simulated temperatures as well as possible interaction between HSPs and the chosen ligands.

1.5 Scope of Work

HSP20 was selected from previously isolated rice seed proteins (Lee, 2015) and these HSP20 sequences were used for homology modelling. The suitable model of the chosen HSP20 sequence was proceeded with molecular dynamics simulation. The simulated environmental temperatures were set as 37ºC and 100ºC during the 50ns MD

run. The stability and flexibility of the selected HSP20 model can therefore be identified. Molecular docking of HSP20 model with the targeted TPR and SGT1 proteins was also be carried out as for the prediction of protein-protein interactions between these docked structures could be achieved successfully.

REFERENCES

Agrawal, G. K. & Rakwal, R. (2011). Rice proteomics: a move toward expanded proteome coverage to comparative and functional proteomics uncovers the mysteries of rice and plant biology. *Proteomics,* 11**,** 1630-1649.

Assimon, V. A., Southworth, D. R., & Gestwicki, J. E. (2015). Specific binding of tetratricopeptide repeat proteins to heat shock protein 70 (Hsp70) and heat shock protein 90 (Hsp90) is regulated by affinity and phosphorylation. *Biochemistry*, *54*(48), 7120-7131.

Bernier, J., Atlin, G. N., Serraj, R., Kumar, A. & Spaner, D. (2008). Breeding upland rice for drought resistance. *Journal of the Science of Food and Agriculture,* 88**,** 927-939.

Chen, X., Lin, S., Liu, Q., Huang, J., Zhang, W., Lin, J., ... & He, H. (2014). Expression and interaction of small heat shock proteins (sHsps) in rice in response to heat stress. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1844*(4), 818-828.

Ding, F., Furukawa, Y., Nukina, N. & Dokholyan, N. V. (2012). Local unfolding of Cu, Zn superoxide dismutase monomer determines the morphology of fibrillar aggregates. *Journal of molecular biology,* 421, 548-560.

Enany, S. (2014). Structural and functional analysis of hypothetical and conserved proteins of *Clostridium tetani*. *Journal of infection and public health*.

Fu, X. (2014). Chaperone function and mechanism of small heat-shock proteins. *Acta Biochim Biophys Sin*, *46*(5), 347-356.

Hartl, F. U., Bracher, A., & Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature*, *475*(7356), 324-332.

Kalman, M., & Ben-Tal, N. (2010). Quality assessment of protein model-structures using evolutionary conservation. *Bioinformatics*, *26*(10), 1299-1307.

Kampinga, H. H., & Craig, E. A. (2010). The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nature reviews Molecular cell biology*, *11*(8), 579-592.

Kim, S. T., Kim, S. G., Agrawal, G. K., Kikuchi, S. & Rakwal, R. (2014). Rice proteomics: A model system for crop improvement and food security. *Proteomics,* 14**,** 593-610.

Komatsu, S., Konishi, H., Shen, S. & Yang, G. (2003). Rice Proteomics A Step Toward Functional Analysis of the Rice Genome. *Molecular & Cellular Proteomics,* 2**,** 2-10.

Li, J., Xiang, C. Y., Yang, J., Chen, J. P., & Zhang, H. M. (2015). Interaction of HSP20 with a viral RdRp changes its sub-cellular localization and distribution pattern in plants. *Scientific reports*, *5*, 14016.

Likitwattanasade, T. & Hongsprabhas, P. (2010). Effect of storage proteins on pasting properties and microstructure of Thai rice. *Food research international,* 43**,** 1402-1409.

Lobanov, M. Y., Bogatyreva, N. S., & Galzitskaya, O. V. (2008). Radius of gyration as an indicator of protein structure compactness. *Molecular Biology*, *42*(4), 623-628.

Martín-Folgar, R., De La Fuente, M., Morcillo, G., & Martínez-Guitarte, J. L. (2015). Characterization of six small HSP genes from *Chironomus riparius* (Diptera, Chironomidae): Differential expression under conditions of normal growth and heat-induced stress. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *188*, 76-86.

Masulli, F. & Mitra, S. (2009). Natural computing methods in bioinformatics: A survey. *Information Fusion,* 10**,** 211-216.

Morozov, A. V., Kortemme, T., Tsemekhman, K., & Baker, D. (2004). Close agreement between the orientation dependence of hydrogen bonds observed in protein

structures and quantum mechanical calculations. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(18), 6946-6951.

Nanni, L., Brahnam, S. & Lumini, A. (2014). Prediction of protein structure classes by incorporating different protein descriptors into general Chou's pseudo amino acid composition. *Journal of theoretical biology,* 360**,** 109-116.

Ouyang, Y., Chen, J., Xie, W., Wang, L., & Zhang, Q. (2009). Comprehensive sequence and expression profile analysis of Hsp20 gene family in rice. *Plant molecular biology*, *70*(3), 341-357.

Oszvald, M., Tömösközi, S., Larroque, O., Keresztenyi, E., Tamás, L. & Békés, F. (2008). Characterization of rice storage proteins by SE-HPLC and micro z-arm mixer. *Journal of cereal science,* 48**,** 68-76.

Park, C. J., & Seo, Y. S. (2015). Heat shock proteins: a review of the molecular chaperones for plant immunity. *The plant pathology journal*, *31*(4), 323.

Reza, M., Chowdhury, A. & Pasha, M. (2005). Characterization of proteins of brown, bran and endosperm of raw and parboiled rice. *Research Journal of Agriculture and Biological Sciences,* 1**,** 184-189.

Sohrabi, M., Rafii, M., Hanafi, M., Siti Nor Akmar, A. & Latif, M. (2012). Genetic Diversity of Upland Rice Germplasm in Malaysia Based on Quantitative Traits. *The Scientific World Journal,* 2012.

Sudha, G., Nussinov, R. & Srinivasan, N. (2014). An overview of recent advances in structural bioinformatics of protein–protein interactions and a guide to their principles. *Progress in biophysics and molecular biology*.

Ul-Haq, Z., Khan, W., Zarina, S., Sattar, R. & Moin, S. T. (2010). Template-based structure prediction and molecular dynamics simulation study of two mammalian Aspartyl-tRNA synthetases. *Journal of Molecular Graphics and Modelling,* 28**,** 401-412.

Waters, E. R. (2012). The evolution, function, structure, and expression of the plant sHSPs. *Journal of experimental botany*, *64*(2), 391-403.

Zhang, M., Boter, M., Li, K., Kadota, Y., Panaretou, B., Prodromou, C., ... & Pearl, L. H. (2008). Structural and functional coupling of Hsp90-and Sgt1-centred multi-protein complexes. *The EMBO journal*, *27*(20), 2789-2798.

Zhang, Y., Zou, B., Lu, S., Ding, Y., Liu, H., & Hua, J. (2016). Expression and promoter analysis of the OsHSP16. 9C gene in rice. *Biochemical and biophysical research communications*, *479*(2), 260-265.