Faculty Resource Science and Technology

Hunting for Second Copy of Proline-rich EVH1 Ligand (PREL) Family Genes in *Danio rerio* Genome.

Tay Ai Peng (20110)

Bachelor of Science with Honours

(Resource Biotechnology)

2010

HUNTING FOR A SECOND COPY OF PROLINE-RICH EVH1 LIGAND (PREL) FAMILY GENES IN *DANIO RERIO* GENOME

TAY AI PENG

SUPERVISOR: DR LEE KUI SOON

This project is submitted in partial fulfillment of the requirement for the degree of Bachelor of Science with Honors

Resource Biotechnology
Department of Molecular Biology

Faculty of Resource Science and Technology
Universiti Malaysia Sarawak
19 May 2010

Declaration

I declare that this thesis entitled "**HUNTING FOR A SECOND COPY OF PROLINE-RICH EVH1 LIGAND (PREL) FAMILY GENES IN *DANIO RERIO* GENOME**" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : .....

Name : .....

Date : .....

ACKNOWLEDGEMENT

First of all, I would like to express my thousand of thanks you to my supervisor, Dr Lee Kui Soon for willing to accept me under his supervise and his advice, opinions, and guidance throughout my final year project and finally complete it. Besides that, I would like to acknowledge Dr Leaw Chiu Pin and Dr Ramlah Zainudin for their guideline on bioinformatics part especially in phylogenetic analysis during Basic Computational Phylogenetic Analysis Workshop 2010. Other than that, a special thanks to my course mates for willing giving opinions and advice to me during this project. Apart from that, I would like to express my sincere appreciation to Faculty Resource Science and Technology for providing facilities for me to complete my project. Finally, deeply gratitude I would like reflects to my beloved family members who are always beside me and giving support to me in term of mentality and monetary throughout my undergraduate study life in Universiti Malaysia Sarawak (UNIMAS).

TABLE OF CONTENT

LIST OF ABBREVIATIONS

APBB1-interacting protein 1	Amyloid beta A4 precursor protein-binding family B member 1-interacting protein
AIC	Akaike Information Criterion
BLAST	Basic Local Alignment Search Tool
EBI	European Bioinformatics Institute
EMBL	European Molecular Biology Lab
EVL	Ena-VASP-like
EVH1	Ena/VASP Homology 1
EVH2	Ena/VASP Homology 2
hLRTs	Hierarchical Likelihood Ratio Tests
Lpd	Lamellipodin
ML	Maximum Likelihood
MP	Maximum Parsimony
MRL	Mig10/RIAM/Lpd
NCBI	National Centre for Biotechnology Information
NJ	Neighbor-joining
PAUP*	Phylogentic Analysis Using Parsimony* and other methods
PP	Posterior Probability
PH	Pleckstrin homology
PHYLIP	Phylogenetic Inference Package
PREL 1	Proline-rich EVH1 Ligand 1
PREL 2	Proline-rich EVH1 Ligand 2

PSRF	Potential Scale Reduction Factor
RA	Ras-association
RAPH1	Ras-associated and pleckstrin domain-containing protein 1
RARP-1	Retinoic acid-responsive proline-rich protein 1
RIAM	Rap 1 Interacting Adaptor Molecule
VASP	Vasodilator-Stimulated Phosphoprotein

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Hunting for a Second Copy of Proline-rich EVH1 Ligand (PREL) Family Genes in *Danio rerio* Genome

Tay Ai Peng

Resource Biotechnology Programme
Faculty of Resource Science and Technology
University Malaysia Sarawak

ABSTRACT

Understanding the mechanism of cells migration and proliferation within tissue becomes bottleneck in studying the cell motility and action dynamics. Remodeling of the actin cytoskeleton provides the driving force for cell migration. Profilin and Ena/VASP proteins family were found to be important regulators of cell motility and actin polymerization. Previous studies found that Proline-rich EVH1 Ligand 1(PREL1) as the link between Ras signaling and action cytoskeleton via Ena.VASP proteins while Proline-rich EVH1 Ligand 2 (PREL2) as the ligand for Ena/VASP proteins and profilin which interacts with Rap1-GTP to mediates the Rap1-Induced adhesion. However, the exact mechanism on how profilin and Ena/VASP proteins regulate those processes and details of the molecular interaction involved are remains largely unclear. Hence, in this study, the characteristics and functions of the PREL genes in *Danio rerio* genome were studied though bioinformatics analysis approach. Second copy of both PREL1 and PREL2 genes were targeted in *Danio rerio* genome. This study was conducted by performing multiple sequence alignment, phylogeny analysis and three dimensional (3D) structural modeling to confirm the identity of selected genes. CLUSTAL X was used to perform multiple sequence alignment while PAUP* version 4.0b10 and Mr Bayers software were selected for phylogeny analysis. I-TASSER server was employed to construct the 3D models of selected PREL1 and PRE2 sequences. We found that it was two copies of PREL2 genes in *Danio rerio* genome located in linkage 9 and linkage 1 respectively. However, the present second copy of PREL1 gene still under investigated.

Keywords: Cell motility, Ena/VASP proteins, profilin, actin polymerization, PREL1, PREL2, bioinformatics

ABSTRAK

Pengetahuan tentang mekanisma bagaimana sel bergerak dan membahagi dalam tisu telah menghalang penyelidikan lanjutan dalam pergerakan sel dan aktin dinamika. Pengubahsuaian dalam bentuk aktin sitoskeleton telah memberi kekuatan untuk pergerakan sel. Profilin dan keluarga protein Ena/VASP dikatakan merupakan pengawal penting dalam pergerakan sel dan polimerisasi aktin. Pneyelidikan sebelum ini telah membuktikan bahawa Proline-rich EVH1 Ligand (PREL1) sebagai penyambung antara pemberi isyarat Ras dan aktin sitoskeleton melalui Ena/VASP.Proline-rich EVH1Ligand 2(PREL2) merupakan ligand untuk Ena/VASP dan profilin dimana bertindak dengan Rap1-GTP untuk menyansangkan adhesi yang dicetuskan oleh Rap1. Walau bagaimanapun, mekanisma sebenar bagaimana profilin dan protein Ena/VASP mengawal interaksi ini masih lagi tidak diketahui. Oleh itu, dalam project ini, ciri-ciri dan peranan gen PREL dalam genome Danio rerio telah dibelajari melalui pendekatan bioinformatik. Salinin kedua bagi kedua-dua gen ini dipercayai wujut dalam genome Danio rerio. Penyelidikan ini dijalankan dengan melakukan penjajaran urutan multiple, filogeni analisis dan 3D reka bentuk untuk memastikan identiti gen yang dipilih.. CLUSTAL X digunakan untuk penjajaran urutan multiple, PAUP 4.0b10 dan Mr Bayers digunakan untuk filogeni analisis. Selain itu, I-TASSER telah digunakan untuk reka bentuk protein 3D.Salinan kedua bagi PREL2 telah dijumpai(masing-masing berada di linkage 9 and linkage1) tetapi salinan kedua bagi PREL1 masih dalam proses penyelidikan.*

Kata kunci: Pergerakan sel, protein Ena/VASP, profilin, polimerisasi aktin, PREL1, PREL2, bioinformatik

CHAPTER 1.0

INTRODUCTION

The actin cytoskeleton dynamic represents fundamental of molecular machinery in the regulation of cell adhesion, migration and polarity in response to extracellular signaling. Remodeling of actin cytoskeleton provides the force required for cell motility, structural changes needed for cell shape modulation, and intracellular anchoring support for adhesion. Axon guidance and T-cell polarization are examples of process with activated in response to motility changes due to environmental signaling (Bailly, 2004; Krause *et al.*, 2003).

Ena/ Vasodilator-stimulated phosphoprotein (VASP) family protein are conversed family of actin regulatory proteins. It acts as important regulators of cell migration by assembly and undergoes rearrangement during lamellipodin and filopodia formation (Bailly, 2004; Jenzora *et al*., 2005; Krause *et al*., 2003). In addition, Ena/VASP proteins family also involves in modulate morphology and behavior of membrane protusions (Bear *et al*., 2002; Gertler *et al*., 2006; Lacayo *et al*., 2007).

In 2004, Lafuente and his colleagues showed that RIAM interacts with Rap 1 and links it to integrin activation. In addition, the study of Jenzora *et al*. (2005) reported that PREL1 as the direct link between Ras signaling and cytoskeleton remodeling via Ena/VASP proteins (RIAM actually is PREL1, different name given due to different findings from different researcher). Activation of cellular integrin is the central to most physiological process including cell migration, assembly of the extracellular matrix, the immune response and homeostasis. Recent study from Lee and his associates (2009) found that RIAM (PREL1) plays important role in

activation of integrins by linking talin to Ras GTPase membrane-targeting sequences. Also in PREL2 or Lamellipodin was grouped into Mig-10/RIAM/Lamellipodin (MRL) family and confirmed as an EVH1 ligand and hence showed that it interacts with Ena/VASP proteins (Krause *et al*., 2004).

The Ras signaling pathway has attracted considerable attention as a target for anticancer therapy due to its significance role in carcinogenesis. Oncogenic mutations in Ras gene are present in approximately 30% of all human cancers (Adjei, 2001). Novel cancer therapeutic approaches based on the inhibition of Ras-mediated signaling pathway have been started to practicing.

Hence, the function of PREL gene family as link of Ras signaling and binding partners for Ena/VASP protein has attracted our attention to find a second copy of this gene family in *Danio rerio* genome for further study on how is the relationship between Ras signaling pathway, actin polymerization and cell migration.

Bioinformatics approach was employed as major tool of analysis in this study. CLUSTAL X (Thompson *et al*., 1997) was used for sequence alignment while PAUP* version 4.0beta 10 (Swofford, 2000) was used for phylogenetic analysis.

The main aim of my study is to identify a second copy of PREL family gene which might exists in *Danio rerio* genome. The possibility of present of this second copy is considered high due to gene duplication as major mechanism of evolving new gene function over the evolution. The evidence of gene duplication is the widespread existence of gene families (Hurles, 2004).

The objectives of this study are:

- To perform multiple sequence alignment to identify the conserved domains of PREL family genes among *Danio rerio* and selected species

- To perform the synteny study to confirm the identity of the candidate gene

- To construct the phylogenetic tree to study the evolutionary relationships of PREL family genes between *Danio rerio* and selected species.

- To generate the 3D protein model to predict the structural function and physiological roles of PREL family genes

CHAPTER 2.0

LITERATURE REVIEW

2.1 Mig-10/RIAM/Lamellipodin (MRL)

The study of Lafuente *et al*. in 2004 found that RIAM shares some similarities with Grb 7 protein family through screening of database. However, further research by sequence and domain structure comparisons of these proteins indicated that they comprise of two distinct families. In contrast, RIAM shares higher domains similarity with Lamellipodin (Lpd), Mig-10 in *C.elegans* and protein CG11940 found in *D. melanogaster*. Hence, the name MRL was proposed by Lafluente *et al*. (2004) for RIAM-related adaptor molecules. The similarity between MRL family and Grb 7 family are sharing of conserved Pleckstrin Homology (PH) domain and Ras-association (RA) domain. Beside these domains, MRL family also contains potential binding sites for profilin and Ena/VASP proteins and a highly conserved region with 27 residues that are predicted to form a coiled-coil region immediately N-terminal of the RA domain (Lafuente *et al*., 2004; Legg & Machesky, 2004). In 2008, *Drosophila pico* (CG11940), an ortholog to MRL, was first time found required in the cell proliferation and tissue growth. Decreasing in the *pico* levels lowered the rates of cell division, resulting in growth retardation and affecting G: F-acin ratio. On the other hand, the rate of tissue growth increased as the levels of *pico* increased (Lyulcheva *et al*., 2008). Research paper published by Lee and his colleagues in 2009 explains that MRL proteins play role as scaffolds that link the membrane targeting sequences in Ras GTPases to talin, hence recruiting talin to the plasma membrane and activation of integrins. RIAM and talin were found in the integrin activation complex. This protein family interacts with both talin and Ras GTPases to activate integrins. (Lee *et al*., 2009).

2.2 Proline-rich EVH1 Ligand 1 (PREL1) or RIAM

Proline-rich EVH1 Ligand 1 (PREL1) (Jenzora *et al*., 2005) or Rap1-GTP-interacting adaptor molecule (RIAM) (Lafuente *et al.*, 2004) as one of the member on MRL family was the first Rap-1 interacting protein found to have a PH domain. It contains six putative profiling-binding sites (XPPPPP) and six putative EVH1 binding motifs (D/E)(F/L/W/Y)PPPPPX(D/E0(D/E) (Lafuente *et al*., 2004; Legg & Machesky, 2004). These motifs give the function of RIAM as profilin and Ena/VASP ligand. Human PREL-1 is located on chromosome 10p12.1 with 1998bps of open reading frame and encodes a protein with 665aa (Lafuente *et al*., 2004). Lafuente *et al*. (2004) proposed that RIAM interacts with Rap1-GTP lead to integrin activation and functions as regulatory in actin dynamics. In 2005, Jenzora and his associates determined that PREL1 as direct link between Ras signaling and cytoskeleton remodeling via Ena/VASP proteins and interacts with GTP-loaded Ras in a lipid dependent manner during cell migration and spreading. PREL-1 interacts with Rap1-GTP and Ena/VASP both *in vitro* and *in vivo*. Ménasché *et al*. (2007) identified RIAM as the key to link ADAP/SKAP-55 signaling module to the small GTPase Rap1, thus facilitating TCR- mediated integrin activation. Further research by Lee and his associates in 2009 found that activation of RIAM-induced integrin activation was mediated by forming complex containing activated Rap1, talin and integrin. RIAM functions as scaffolds that connect the membrane targeting sequences in Ras GTPases to talin, recruiting talin to the plasma membrane thereby avtivate integrins.

2.3 Proline-rich EVH1 Ligand 2 (PREL2) or Lamellipodin (Lpd)

PREL2 or lamellipodin (Lpd) is another member on MRL protein family. The term lamellipodin was named by Krause and her associates in their finding in 2004. The N-terminal of Lpd comprises of 50 amino acid is highly charged, followed by a putative coiled-coil motif, RA and PH domains. The C-terminal of Lpd is 500 amino acids longer than RIAM. It is rich in proline with eight potential SH3 binding sites, three potential Profilin binding sites and six putative EVH1 binding sites. Each motif of Lpd can bind to EVL directly, different to RIAM which requires two motifs for Mena binding site (Jenzora *et al*., 2005; Krause *et al*., 2004). In 2009, Lee and his colleagues identify a short talin binding sequence in PREL2. Binding of talin to PREL2 and links to Rap1 membrane targeting sequence is sufficient to recruit talin and induces integrins activation. Both PREL1 and PREL2 mediate activation of integrin by a common scaffolding mechanism; in which contains amphipathic helices that mediate direct binding to talin and RA domains that bind Ras superfamily GTPases (Lee *et al*., 2009). This finding further confirms that PREL genes family is contributed to the actin polymerization pathway and cell migration. In certain aspect, PREL1 and PREL2 have some opposing effects. In cell adhesion, PREL1 is required for Rap1 to stimulate β1 and β2 integrin-mediated adhesion. In contrast, PREL2 seem to be negatively regulates cell adhesion (Krause *et al*., 2004; Lafuente *et al*., 2004).

2.4 Ena/Vasodilator-Stimulated Phosphoprotein (VASP) Family Proteins

Ena/VASP proteins are conversed family of multi-functional actin-modulating proteins (Scott *et al*., 2006) and actin polymerization nucleating proteins share conversed domains: an N-terminal EVH1, Proline-rich central region and C-terminal EVH2 domain (Krause *et al*., 2003). The EVH1 domain binds to the specific proline-rich motif via recognizes the consensus site (D/E)-FPPPP-X(D/E)(D/E) and localizes the Ena/VASP proteins to receptor signaling complex sich as focal adhesion complexes. On the other hand, the EVH2 domain mediates tetramerization of Ena/VASP proteins and has binding sites for G-actin and F-actin and probably important for the elongation of actin filaments. The proline-rich central region contains binding sites for SH3 and WW domain-containing protein and profilin (Bear & Gertler, 2009; Legg & Machesky, 2004; Krause *et al*., 2003). The members of this protein family consist of *Drosophila* Ena, *C.elegans* Unc-34, *Dictyostelium* DdVASP and three mammalian members: VASP, Mena and EVL (Krause *et al*., 2003). This protein family is mainly involved in a range of processes dependent on cytoskeleton remodeling and cell polarity (Krause *et al*., 2003). It also play roles in several physiological processes such as morphogenesis, axon guidance, cancer cell invasion, and endothelial barrier function (Bear *et al*., 2002; Bear & Gertler, 2009; Krause *et al*., 2003). Ena/VASP proteins are frequently found in various cell type including fibroblasts (Bear *et al*., 2000) , endothelial cells and tips of the lamellipodia (Bear & Gertler, 2009; Krause *et al*., 2003; Trichet *et al*., 2008). However, the biochemistry mechanism on how Ena/VASP regulates the actin dynamic is still under debate. In 2009, Bear and Gertler presented their paper discussed on possible mechanism of Ena/VASP including anti-capping hypothesis, inhibition of branching and bundling and profilin-actin recruitment.

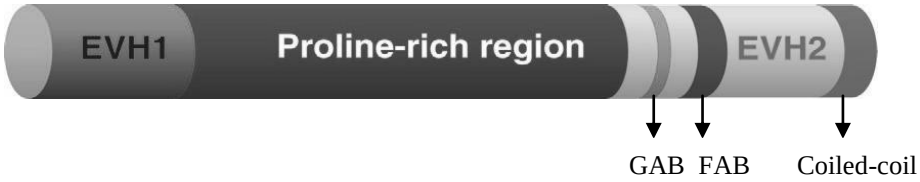

Figure 1: Domains structure of Ena/VASP Protein (Bear & Gertler, 2009)

GAB = G-actin binding site; FAB = F-actin binding site

2.5 Pleckstrin Homology (PH) Domain

The 'Pleckstrin Homology' domain is term proposed by Haslam *et al*. and Mayer *et al.* (1993) for a homology sequence consists of approximately 100 to 120 amino acid residues which repeatedly in both N- and C-terminals of pleckstrin. This domain consists of two nearly antiparallel β-sheets, comprised from four and three β-strands to form a β-sandwich, a C-terminal α-helix and three inter-strand loops (Hirata *et al*., 1998). PH domain was found that interacts with variety of proteins involved in signal transduction pathways and cytoskeleton organization (Lemmon *et al*., 2000; Musacchio *et al*., 1993). This leads to the investigation that this PH domain maybe involve in membrane recruitment of PH domain-containing proteins by targeting them to appropriate cellular compartment through interacting with PH domain-binding-ligand or non-protein molecule such as src homology (SH) domain (Hirata *et al*., 1998; Lemmon, 2004). Inositol 1,4,5-triphosphate[Ins (1,4,5)P_3]/ phosphatidylinositol 4,5-bisphosphate [PtdIns (4,5)P_2], the βγ-subunits of heterotrimeric G proteins (G βγ) and protein kinase C are some of the identified potential PH domain-binding ligand (Hirata *et al*., 1998). The PH domain of PREL family shows different binding affinity toward phosphoinositides (Jenzora *et al*., 2005; Krause *et al*., 2004).

2.6 Ras-association (RA) Domain

The Ras-association (RA) domain of PREL family shared approximately 36% of homology with RA domain of Grb7/10/14 (Jenzora *et al*., 2005). RA domain for PREL1 and PREL2 seem like showing different binding affinity toward Rap1. The RA domain of PREL1 interacts specifically with activated Rap1 to stimulate integrin-mediated adhesion in a lipid dependent manner. Besides Rap1, RA domain for PREL1 also found bound to Ras superfamily protein including H-, N- and K- Ras (Holt & Daly, 2005; Jenzora *et al*., 2005; Lafuente *et al*., 2004). On the other hand, PREL2 was reported fail to interact directly with Rap1 and Ras superfamily protein. This might be due to different in specificity of its RA domain (Holt & Daly, 2005; Krause *et al*., 2004).

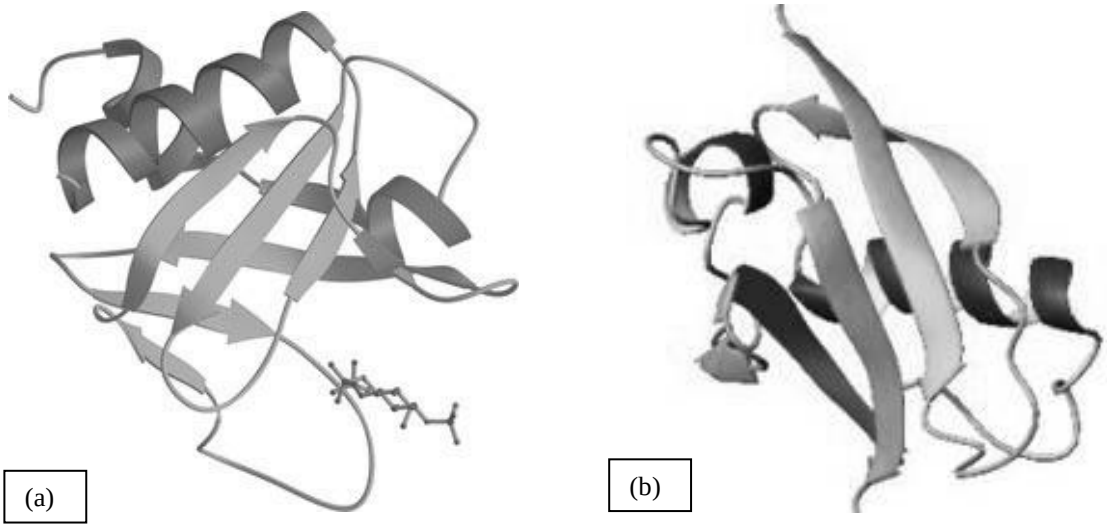

Figure 2: Domain Topology of (a) PH Domain; (b) RA Domain

Figure adapted from (a) http://www.cellsignal.com/reference/domain/ph.html; (b) http://www.fmp-berlin.de/proteinstructuressectionstructur.html